arXiv:2601.06173v1 [math.HO] 7 Jan 2026

稀疏优化入门课程

Jun Lu
JUN.LU.LOCKY@GMAIL.COM

序言

在这个由数据定义的时代——数据的海量、速度与复杂度——我们面临着一个悖论：收集的信息越多，就越难从中提取意义。高维数据集如今遍布科学、工程和工业的每个角落，从医学成像和基因组学，到无线通信和机器学习。然而，在这场数字洪流之下，蕴藏着一个重要洞见：我们收集的大多数数据并没有乍看之下那么复杂。事实上，许多现实世界信号都是稀疏的——换个恰当的视角，只需少数几个关键成分，就能准确描述它们。

这一观察催生了一个变革性领域，它位于应用数学、信号处理、统计学和优化的交汇处：稀疏优化。从本质上说，稀疏优化旨在恢复或构造尽可能简单的解——其中大多数元素为零或可忽略不计——同时忠实地表征底层现象。无论目标是根据少量测量值重建一幅像、识别与某种疾病相关的关键遗传标记，还是构建具有可解释特征的预测模型，稀疏性都兼具计算效率与概念清晰性。

这一领域已经形成两种主要范式。第一种植根于压缩感知，它彻底重新构想了数据的获取方式。通过认识到许多信号在某些定义域中是稀疏的，压缩感知表明，我们可以绕过传统的采样极限，用远少于以往认为可能的测量值重建信号。第二种范式以 LASSO 及相关正则化技术为代表，将稀疏性直接融入建模过程，促使模型不仅准确，而且简约，只从海量候选变量中选取最相关的变量。

尽管这些方法在动机和应用上各不相同，但它们共享深厚的数学基础和算法策略。本书旨在邀请读者探索这些联系。它既可作为进入这一领域的学生和研究人员易于理解的入门读物，也可作为希望在工作中应用稀疏优化方法的实践者所需的严谨参考书。我们从基本的数学和算法构件入手，随后系统地发展稀疏信号恢复与稀疏正则化的理论与实践。在此过程中，我们在不牺牲精度的前提下强调直觉，同时也不因重视理论而忽视应用。

本书将带领我们了解高维空间几何、能够保持稀疏性的测量系统设计，以及使大规模稀疏恢复成为可能的算法。我们将看到如何利用随机性来保证恢复，如何通过凸松弛将难以处理的问题转化为可求解的问题，以及现代扩展——例如结构化稀疏性和非凸惩罚——如何拓展可能性的边界。

关键词。 稀疏优化、LASSO、广义 LASSO、组 LASSO、稀疏恢复、凸优化、凸松弛、设计矩阵、零空间性质、受限特征值性质、受限等距性质、精确恢复、高斯随机矩阵、次高斯随机矩阵、原始-对偶算法、梯度下降。

目录

1 引言　　论 2

1.1 导论与背景.31.2 线性代数.51.2.1 向量范数、对偶范数、球.141.2.2 矩阵范数：弗罗贝尼乌斯范数与谱范数.171.2.3 正交矩阵与投影矩阵.201.2.4 奇异值分解（SVD）.231.2.5 伪逆.291.2.6 著名不等式.341.3 概率论与常见概率分布.351.3.1 一些集中不等式.421.3.2 高斯分布及其性质.48第1章习题.57

I 理论 62

2 优化基础 64

2.1 可微函数与微分学.652.1.1 二阶偏导数.692.1.2 多元实函数.702.1.3 微积分基本定理.712.2 凸组合与凸集.722.3 凸函数.762.3.1 凸函数的性质.782.3.2 凸函数的刻画.802.4 连续函数与闭函数，以及魏尔斯特拉斯定理.822.5 次梯度与共轭函数.862.6 利普希茨函数.912.7 强凸、强光滑函数.932.8 投影、近端、布雷格曼近端与分离.1002.8.1 投影算子的性质.1032.8.2 分离定理.106

2.8.3 近端算子的性质.1062.8.4 莫罗分解.1102.9 凸优化与最优性条件.1112.9.1 凸集上的约束优化.1132.9.2 等式/不等式约束.1142.9.3 对偶问题.1142.9.4 复合函数.118第2章习题.119

3 稀疏优化问题与最优性条件 122

3.1 稀疏优化问题.123

3.1.1 稀疏回归问题.1243.1.2 稀疏恢复问题.1273.1.3 某些问题的等价性.1323.1.4 ℓ_1-分析模型.133

3.2 最小二乘问题的最优性条件.134

3.2.1 凸优化中的最小二乘.1363.2.2 加权最小二乘问题.1373.2.3 回归及其他形式.138

3.3 最优性条件 ℓ_1正则化问题.140

3.3.1 LASSO 对偶.1423.3.2 对偶多面体投影与筛选规则.1443.3.3 安全特征消除（SAFE）.1473.4 最优性条件 ℓ_1最小化问题.1503.4.1 充要条件.1513.4.2 对偶证书与唯一性.1523.4.3 ℓ_1最小化与含噪测量.1563.5 最优性条件 ℓ_1分析模型*.1603.5.1 无噪声情形下的唯一恢复.1613.5.2 含噪信号中的唯一恢复.1643.5.3 噪声信号中的稳定恢复.165第3章习题.168

II 设计 矩阵 172

4 设计 gn 矩阵与性质 174

4.1 零空间、受限特征值及其他设计性质 . . . . . 1774.1.1 零空间性质-II.1784.1.2 稳定零空间性质.1804.1.3 鲁棒零空间性质.1824.2 火花、互相干性与稀疏恢复的唯一性 . . . . . 1844.2.1 火花 .1844.2.2 互相干性 .1864.2.3 ℓ_1- 相干性 .1904.2.4 互相干性与 ℓ_1- 相干性的界 . . . . . 1934.3 受限等距性质（RIP）.1964.3.1 受限正交性性质（ROP）.1994.3.2 受限等距常数、受限正交常数及其他常数之间的关系 201

4.3.3 受限等距性质的界.204 4.3.4 限制等距性质与稳定性.207 4.3.5 限制等距性质、零空间性质及其变体.209 第4章习题.217

5 稀疏优化分析 **220**

5.1 零空间性质下的稀疏恢复.222 5.1.1 无噪声信号恢复.222 5.1.2 含噪信号恢复.224 5.2 相干性下的稀疏恢复.226 5.3 RIP 下的稀疏恢复.227 5.3.1 对偶证书：再探.227 5.3.2 无噪声信号恢复.232 5.3.3 实例最优保证.234 5.3.4 含噪信号恢复.235 5.3.5 使用 Dantzig 选择器进行信号恢复.238 5.4 稀疏线性回归.239 5.4.1 LASSO 与稳健线性回归.241 5.4.2 LASSO 误差界.244 5.4.3 高斯噪声与预言机不等式.247 第5章习题.249

6 设计 高斯系综下感知矩阵的设计 252

6.1 覆盖数.253 6.2 高斯向量的集中性.255 6.3 高斯随机矩阵下的随机投影 . . . . . 260 6.4 随机投影与 Johnson–Lindenstrauss 引理 261 6.5 高斯随机矩阵的奇异值.266 6.6 高斯随机矩阵下的稀疏优化.268 第6章习题.271

7 次高斯系综下的传感矩阵设计 **274**

7.1 次高斯（SG）随机变量.275 7.2 严格次高斯（SSG）随机变量.279 7.3 各向同性次高斯随机向量.283 7.4 SSG随机变量的测度集中.285 7.5 次高斯随机矩阵下的设计性质. 289 7.5.1 各向同性与次高斯随机系综下的限制等距性质 . . . . . 290 7.5.2 SSG 随机系综下的限制等距性质.293 7.5.3 其他问题.295 7.6 次高斯系综下的稀疏优化.295 7.6.1 正交归一基下的稀疏性.296 7.6.2 严格次高斯条件下的稀疏恢复 297 第7章习题.298

III 算法 **300**

8 统一 稀疏优化问题算法 302

8.1 Fenchel 对偶定理：再探.303

8.2 原始-对偶算法.3048.3 原始-对偶框架下的不同问题.3068.3.1 拉格朗日 LASSO.3068.3.2 约束 LASSO.3078.3.3 ℓ_1-最小化.3088.3.4 ℓ_1-含噪测量下的最小化.3098.4 原始-对偶算法的最优性条件.3098.5 原始-对偶算法的收敛性.3108.6 原始-对偶算法的收敛速度.315第8章 习题.316

9 算法 稀疏回归问题的算法318

9.1 ℓ_0约束问题的算法.319 9.1.1 L驻点条件下的最优性条件.320 9.1.2 迭代硬阈值（IHT）方法.322 9.2 约束 LASSO 问题的算法.329 9.2.1 投影梯度下降（PGD）.329 9.2.2 近端梯度法.332 9.2.3 条件梯度（Frank–Wolfe）法.336 9.2.4 最小角回归（LARS）.339 9.3 拉格朗日 LASSO 问题的算法.353 9.3.1 Huber损失平滑.354 9.3.2 近端梯度法与主化最小化方法 . . . . . 355 9.3.3 快速近端梯度法—FISTA.357 9.3.4 罚函数法.361 9.3.5 交替方向乘子法（ADMM） . . . . . 365 9.3.6 块坐标下降.371 9.3.7 Barzilai–Borwein 梯度投影算法 . . . . . 372 9.4 其他稀疏优化建模形式.374 9.4.1 广义 LASSO 问题.374 9.4.2 组 LASSO 与对偶多面体投影 375 9.4.3 稀疏噪声.380 第9章习题.381

10 稀疏恢复问题算法386

10.1 ℓ_0-最小化.387

10.1.1 贪心算法.38710.1.2 正交匹配追踪算法.39010.1.3 基于阈值的算法.394

10.2 ℓ_1-最小化算法.39810.2.1 同伦算法.39810.2.2 迭代重加权最小二乘法（IRLS）.40210.2.3 一阶增广拉格朗日算法.41210.2.4 布雷格曼迭代算法.423第10章 问题.429

参考文献 431

索引 441

符号说明

本节简要介绍本书中通用的符号。如果您不熟悉相关的数学概念，本书会在第1章（第2页）对其中大多数概念进行介绍。

数字与数组

a 标量（整数或实数，斜体）$\boldsymbol{a}$ 向量（斜体）$\boldsymbol{A}$ 矩阵（斜体）$\boldsymbol{I}_n$ 具有 n 行和 n 列的单位矩阵$\boldsymbol{I}$ 维数由上下文确定的单位矩阵$\boldsymbol{e}_i$ 标准基向量 $[0, \dots, 0, 1, 0, \dots, 0]$ 在位置 i 处取值为 1 的向量$\boldsymbol{a}$diag($\boldsymbol{a}$) 对角元素由给出的方阵、对角矩阵标量随机变量（正体） $\mathbf{a}$ 向量值随机变量（正体） $\mathbf{A}$ 矩阵值随机变量（正体）

Sets

$\mathbb{A}$ 集合$\varnothing$ 空集$\mathbb{R}$、$\mathbb{C}$、$\mathbb{F} \equiv \{\mathbb{R}$ 或 $\mathbb{C}\}$ 实数、复数或实数与复数的集合$\mathbb{N}$、$\mathbb{N}_0$ 自然数集合 $\mathbb{N} = \{1, 2, \ldots\}$、$\mathbb{N}_0 = \{0\} \cup \mathbb{N}\{0, 1\}$ 包含 0 和 1 的集合$\{0, 1, \ldots, n\}$ 包含 0 到 n 之间所有整数的集合$[a, b]$ 包含 a 和 b 的实区间（a, $b]$ 不包含 a 但包含 b 的实区间$\mathbb{A} \backslash \mathbb{B}$ 集合差，即包含 $\mathbb{A}$ 中不属于 $\mathbb{B}$ 的元素的集合$\mathbb{B}_s$（$\boldsymbol{c}$, r）开球：$\mathbb{B}_s$（$\boldsymbol{c}$, r）$= \{\boldsymbol{x} \mid \|\boldsymbol{c} - \boldsymbol{x}\|_s < r\}\mathbb{B}_s[\boldsymbol{c}, r]$ 闭球：$\mathbb{B}_s[\boldsymbol{c}, r] = \{\boldsymbol{x} \mid \|\boldsymbol{c} - \boldsymbol{x}\|_s \leq r\}\mathbb{B}$（$\boldsymbol{c}$, r）ℓ_2 开球：$\mathbb{B}$（$\boldsymbol{c}$, r）$= \{\boldsymbol{x} \mid \|\boldsymbol{c} - \boldsymbol{x}\|_2 < r\}\mathbb{B}[\boldsymbol{c}, r]$ ℓ_2 闭球：$\mathbb{B}[\boldsymbol{c}, r] = \{\boldsymbol{x} \mid \|\boldsymbol{c} - \boldsymbol{x}\|_2 \leq r\}$

索引

x_i 向量 $\boldsymbol{x}$的第 i 个元素，索引从 1 开始$\boldsymbol{x}_{-i}$ 向量 $\boldsymbol{x}$中除第 i个元素以外的所有元素x_{ij} 矩阵 $\boldsymbol{X}$的第 i个元素，j $\boldsymbol{X}_{i,:} = \boldsymbol{X}[i,:]$，$\boldsymbol{x}^{(i)}$ 矩阵 $\boldsymbol{X}$的第 i 行$\boldsymbol{X}_{:,i} = \boldsymbol{X}[:, i]$，$\boldsymbol{x}_i$ 矩阵 $\boldsymbol{X}$的第 i 列$\boldsymbol{X}_{\mathbb{S}} = \boldsymbol{X}[:, \mathbb{S}]$，$\boldsymbol{X}(\mathbb{S})$ $\boldsymbol{X}_{\mathbb{S}} \in \mathbb{R}^{n \times |\mathbb{S}|}$和 $\boldsymbol{X}(\mathbb{S}) \in \mathbb{R}^{n \times p}$如果 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$

线性代数运算

$\boldsymbol{X}^\top$ 矩阵的转置 $\boldsymbol{X}\boldsymbol{X}^+$ $\boldsymbol{X}$的摩尔-彭若斯伪逆$\boldsymbol{X} \circ \boldsymbol{Y}$ $\boldsymbol{X}$ 与 $\boldsymbol{Y}$的逐元素（Hadamard）乘积$\det(\boldsymbol{X})$ $\boldsymbol{X}$的行列式$\mathcal{C}(\boldsymbol{X})$ $\mathcal{N}(\boldsymbol{X})$ $\mathcal{V}$列、$\boldsymbol{X}$的零空间以及一般空间$\text{rank}(\boldsymbol{X})$ $\boldsymbol{X}$的秩$\text{tr}(\boldsymbol{X})$ $\boldsymbol{X}$的迹$\lambda_{\max}(\boldsymbol{X})$ $\lambda_{\min}(\boldsymbol{X})$ $\boldsymbol{X}$的最大和最小特征值$\sigma_{\max}(\boldsymbol{X})$ $\sigma_{\min}(\boldsymbol{X})$ $\boldsymbol{X}$的最大和最小奇异值$\text{diag}(\boldsymbol{\beta})$ 对角元素为 $\boldsymbol{\beta}$的对角矩阵

微积分$\frac{dy}{dx}$ y 关于 x的导数$\frac{\partial y}{\partial x}$ y 关于 x的偏导数$\nabla_{\boldsymbol{x}} y$ y 关于 $\boldsymbol{x}$的梯度$\nabla_{\boldsymbol{X}} y$ y 关于 $\boldsymbol{X}$的矩阵导数$\frac{\partial f}{\partial \boldsymbol{x}}$ $\boldsymbol{J} \in \mathbb{R}^{m \times n}$ 的雅可比矩阵（关于 f）：$\mathbb{R}^n \xrightarrow{\boldsymbol{x}} \mathbb{R}^m \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$或 $\boldsymbol{H}(f)(\boldsymbol{x})$ f在输入点 $\boldsymbol{x}$处的海森矩阵$\int f(\boldsymbol{x})d\boldsymbol{x}$ 在整个定义域上的定积分（ ）关于的定积分（在集合 $\mathbb{S}$上）$\int_{\mathbb{S}} f\ \boldsymbol{x}\ d\boldsymbol{x}$

概率论与信息论 a⊥b 随机变量 a 和 b 相互独立 a⊥b |c 给定 c 时，它们条件独立 Pr(***x***) 离散变量上的概率分布$p(\boldsymbol{x})$ $p_{\mathrm{x}}(\boldsymbol{x})$ $f(\boldsymbol{x})$ $f_{\mathrm{x}}(\boldsymbol{x})$ 连续变量上的概率分布，或类型未指定的变量上的概率分布 a $\sim P$ 随机变量 a 服从分布 $P\mathbb{E}_{\mathrm{x}\sim P}[f(x)]$或 $\mathbb{E}[f(x)]$关于 P(x)$\mathbb{V}$的 $f(x)$的期望 ar$[f(x)]$在 P(x)$\mathbb{C}$下 $f(x)$的方差 ov$[f(x)\ g(x)]$在 P(x)$\mathbb{C}$下 $f(x)$与 $g(x)$的协方差 orr$[f(x)\ g(x)]$在 P(x)$\mathcal{N}$下 $f(x)$与 $g(x)$的相关性 μ, σ^2) 均值为 μ、方差为 σ^2的高斯分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 均值为 $\boldsymbol{\mu}$、协方差为$\boldsymbol{\Sigma}$的多元高斯分布 Bern(p) 均值为 p的伯努利分布$\mathcal{E}(\lambda)$ 尺度为 λ的指数分布$\chi^2_{(s)}$ 自由度为 s的卡方分布 (df)$\mathcal{SG}(\sigma^2)$ 次高斯分布$\mathcal{SSG}(\sigma^2)$ 严格次高斯分布

函数

$f:\ \mathbb{A}\to\mathbb{B}$	函数 f，定义域为 $\mathbb{A}$，值域为 $\mathbb{B}$
$f\circ g$	函数 f 和 g 的复合
$f(\boldsymbol{x};\ \boldsymbol{\theta})$	由 $\boldsymbol{\theta}$ 参数化的 $\boldsymbol{x}$ 函数。（有时我们写作 $f(\boldsymbol{x})$ 并省略参数 $\boldsymbol{\theta}$ 以简化符号表示）
$\ln(x)$，$\log(x)$	x 的自然对数
x 的符号函数	符号函数 x，取值为 $+1$、-1 和 0
$\sigma(x)$，Sigmoid函数(x)	逻辑Sigmoid函数，即，$\dfrac{1}{1+\exp\{-x\}}$
$\text{logit}(\pi)$	Logit函数，即 $\text{logit}(\pi)=\ln(\pi/(1-\pi))$，其中 $\pi\in(0,\ 1)$
$\zeta(x)$	Softplus函数，$\log(1+\exp\{x\})$
$\|\boldsymbol{x}\|_p$ $\|\boldsymbol{x}\|_s$	ℓ_p $\boldsymbol{x}$的范数
$\|\boldsymbol{x}\|=\|\boldsymbol{x}\|_2$	$\boldsymbol{x}$的ℓ_2-范数
$\|\boldsymbol{x}\|=\|\boldsymbol{x}\|_1$	$\boldsymbol{x}$的ℓ_1-范数
$\|\boldsymbol{x}\|=\|\boldsymbol{x}\|_\infty$	$\boldsymbol{x}$的ℓ_∞ 范数
$[x]_+$	x的正部，即 $\max(0,\ x)$
$u(x)$	当 $x\geq 0$ 时取值为 1，否则取值为 0 的阶跃函数
$\mathbb{1}\{\text{条件}\}$	若条件成立则为 1，否则为 0
$\delta_{\mathbb{S}}(\boldsymbol{\alpha})$	若 $\boldsymbol{\alpha}\in\mathbb{S}$，$+\infty$ 则为 0
$\Phi(x)$ $\Phi^{-1}(\pi)$	标准高斯累积分布函数和概率单位函数，其中 $\pi\in(0,\ 1)$
负二项分布$(\alpha,\ x)$	$\eta=\ln(x/(x+1/\alpha))$
$\mathcal{T}_\lambda(\boldsymbol{\beta})$	软阈值算子
$\sigma_k(\boldsymbol{\beta})_s$	ℓ_s-误差：最佳 k-稀疏近似的 $\min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]}\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|_s$

有时，我们会使用一个函数 f，其参数是标量，但将其应用于向量或矩阵：$f(\boldsymbol{x})$、$f(\boldsymbol{X})$。这表示对数组逐元素应用 f。例如，若$\boldsymbol{C}=\sigma(\boldsymbol{X})$，则对于 i 和 j 的所有有效取值，$c_{ij}=\sigma(x_{ij})$。

其他一般记号

$\triangleq$ 按定义相等：$=$，$\leftarrow$ 按赋值相等

$\equiv$ 按等价性相等π 概率值或 3.141592.....

e，exp 2.71828.....

缩略语

PD	正定
PSD	半正定
独立同分布	独立同分布
概率密度函数，概率密度函数	概率密度函数
概率质量函数，概率质量函数	概率质量函数
MGF	矩母函数
MVT	中值定理
LSC	下半连续性
LP	线性规划
最小二乘、普通最小二乘	普通最小二乘
GD	梯度下降
SGD	随机梯度下降
MSE	均方误差
CLT	中心极限定理
SVD	奇异值分解
BP	基追踪
BPDN	基追踪去噪
LASSO	最小绝对收缩与选择算子
DPP	对偶多胞体投影
SAFE	安全特征消除
RIP	受限等距性质
RIC	限制等距常数
ROP	受限正交性
ROC	限制正交常数
NSP	零空间性质
JL引理	Johnson–Lindenstrauss 引理
SVM	支持向量机
FW	Frank–Wolfe 算法

CS 压缩感知 LASSO 最小绝对收缩与选择算子

r

LARS 最小角回归 IHT 迭代硬阈值 ADMM 交替方向乘子法 BB rule Barzilai–Borwein 法则 BCD 块坐标下降 BCQP 有界约束二次规划 PGD 投影梯度下降 PGM 近端梯度法 SG 次高斯 SSG 严格次高斯 ISTA 迭代收缩阈值算法 FISTA 快速近端梯度法 MP 匹配追踪 ROMP 正则化正交匹配追踪 OMP 正交匹配追踪 StOMP 逐步正交匹配追踪 CoSaMP 压缩采样匹配追踪 SP 子空间基追踪 BT 基本阈值法 HTP 硬阈值追踪 IRLS 迭代重加权最小二乘 ALF 增广拉格朗日函数 FAL 一阶增广拉格朗日

第1章

引言

目录

1.1 导论与背景.3 1.2 线性代数.5 1.2.1 向量范数、对偶范数、球.14 1.2.2 矩阵范数：弗罗贝尼乌斯范数与谱范数.17 1.2.3 正交矩阵与投影矩阵.20 1.2.4 奇异值分解（SVD）.23 1.2.5 伪逆.29 1.2.6 著名不等式.34

1.3 概率论与常见概率分布 . . 35

1.3.1 一些集中不等式.42 1.3.2 高斯分布及其性质.48 **第1章习题.57**

1.1. 导论与背景



I在当今数据驱动的世界中，我们在收集、存储和分析海量信息的方式上正经历着深刻的转变。随着传感技术和计算方法不断发展，一个关键挑战随之出现：如何要高效地管理本质上稀疏或可压缩的高维数据。这一挑战正是稀疏优化的核心，该领域既关注稀疏信号的恢复，也关注构建能够促使解具有稀疏性的优化问题。

得益于数字传感技术的快速发展，我们如今能够以空前的分辨率采集数据。然而，这些系统往往会产生海量信息，给存储、处理和传输带来巨大挑战。幸运的是，许多现实世界信号——从图像和音频录音到基因组序列和医学扫描图像——在某个变换域中都是稀疏或可压缩的。换言之，尽管这类信号可能处于高维空间中，但通常只需使用少量非零系数就能对其进行准确表示。认识并利用这种内在稀疏性，对于开发能够降低维度而不损失关键信息的高效算法至关重要。

稀疏信号恢复：压缩感知的作用。稀疏优化技术的主要应用之一是稀疏信号恢复。这项任务旨在利用信号在某个基或变换域中具有稀疏表示这一事实，从数量较少的测量值中恢复信号。

由奈奎斯特–香农采样定理所支配的传统信号采集方法要求以信号中最高频率的两倍进行采样，以确保完美重建（Nyquist，2002年；Shannon，2006年）。虽然该理论对带限信号效果良好，但当信号具有稀疏性时，其效率非常低。压缩感知（*CS*，即压缩式传感）这一领域于21世纪初兴起，从根本上重新思考了这一范式。CS并非先以高速率采集信号，再对其进行压缩（如 JPEG、MP3 或 MPEG），而是主张以压缩形式采集数据，即直接以较低速率进行采样。通过利用稀疏性，可以从数量大幅减少的非自适应线性测量值中重建信号，其数量远少于环境维数所暗示的测量数量。这并非单纯的数据压缩，而是在采集点进行压缩。这既减少了所需的测量数量，也降低了存储和处理数据的计算成本（Candes 和 Tao，2006年；Candes 等人，2011年；Davenport 等人，2012）。

压缩感知的核心思想是：如果一个信号在已知基（例如小波、傅里叶基或学习字典）下是稀疏的，我们就能用远少于传统方法所需的测量值来重建它。压缩感知中的恢复过程通常表述为一个优化问题，旨在找到与现有测量值相符的最稀疏信号。形式上，给定测量向量 $\boldsymbol{y}$，目标是恢复满足下式的信号 $\boldsymbol{\beta}$

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

其中，$\|\boldsymbol{\beta}\|_0$ 表示 ℓ_0-范数（即 $\boldsymbol{\beta}$中非零系数的数量），$\boldsymbol{X}$ 是测量矩阵，而 $\boldsymbol{\epsilon}$ 表示噪声。

在实践中，最小化 ℓ_0-范数在计算上难以实现，因为该问题是非凸的和组合的。为克服这一点，压缩感知通常将 ℓ_0-范数

替换为凸的 ℓ_1-范数，从而得到著名的基追踪 公式：

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

这种松弛通过凸优化算法实现了对稀疏信号的高效、稳健恢复——这是现代压缩感知理论与实践的基石。

使用 LASSO 进行稀疏优化：在稀疏性约束下最小化损失。稀疏信号恢复侧重于从压缩测量中重构信号，*LASSO*（最小绝对收缩与选择算子）则处理优化中的一个相关但有所区别的问题：它寻求在对稀疏性约束或稀疏正则化解施加Tibshirani，1996年）的同时，最小化损失函数。LASSO 广泛应用于机器学习和统计学中的回归等任务，其目标是识别出能够最佳解释观测数据的一小组特征。

在线性回归的标准设置中，目标函数是最小化观测响应向量 $\boldsymbol{y}$ 与模型预测值 $\boldsymbol{X}\boldsymbol{\beta}$之间的平方误差。LASSO 通过添加一个 ℓ_1-范数正则化项来修改该目标函数，从而促进估计系数的稀疏性。LASSO 优化问题可以写为：

$$\min_{\boldsymbol{\beta}} \frac{1}{2} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 .$$

这里，$\frac{1}{2} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 表示最小二乘损失，$\|\boldsymbol{\beta}\|_1$ 是 ℓ_1 惩罚项 或 ℓ_1 正则化，通过惩罚系数绝对值之和来促进稀疏性。超参数 λ控制数据拟合与施加稀疏性之间的权衡。λ 的值越大，被压缩至零的系数就越多，从而促进更高的稀疏性。

LASSO 是一种用途广泛且计算上易于处理的方法，已广泛应用于回归、分类和信号去噪等领域。在这些领域中，稀疏性有助于提升可解释性、泛化能力或效率。虽然 LASSO 并非专门针对从有限测量值中恢复信号，但它提供了一个强大且通用的框架，可将稀疏性直接纳入基于优化的建模中。

稀疏优化。稀疏信号恢复（如压缩感知）与稀疏正则化 或稀疏约束（如 LASSO）之间的关键区别在于它们的目标以及底层优化问题的性质：

- 稀疏信号恢复（*CS*）。目标是从有限数量的测量值中恢复稀疏（或近似稀疏）的信号。该问题通常被表述为：寻找与现有数据一致的最稀疏解。CS 通常利用基追踪等优化技术（使用 ℓ_1-范数）从压缩数据中恢复信号。稀疏正则化（*LASSO*）。在 LASSO 等稀疏优化问题中，目标是在促使模型参数具有稀疏性的同时，最小化数据
- 拟合损失函数（例如最小二乘）。具体做法是在目标函数中加入 ℓ_1-范数惩罚项，使不太重要的系数趋近于零，从而有效执行特征选择。

尽管这两种范式密切相关，但其目的各不相同，所需的算法方法也有区别。压缩感知主要关注如何从有限的测量值中采集和重建信号，而 LASSO 则侧重于构建具有可解释性和稀疏参数集的预测模型，通常应用于统计学习和回归领域。

稀疏优化领域对众多领域产生了深远影响，尤其是涉及高维数据的领域。在医学成像中，压缩感知能够缩短磁共振成像扫描时间并减少辐射暴露；在通信领域，它有助于提高数据传输效率（Lustig 等人，2006年、2007年）。与此同时，LASSO 及其扩展已成为机器学习、基因组学和计量经济学领域的标准工具。在这些领域中，稀疏模型对于处理大规模数据集并生成可解释的结果至关重要（Tibshirani，1996年；Hastie 等人，2015年；Puig 等人，2011年）。

本书旨在全面概述稀疏优化，重点介绍稀疏信号恢复 和稀疏正则化 技术。我们将从探索稀疏优化的基础开始，深入讲解支撑稀疏信号恢复和 LASSO 的数学工具与模型。随后，我们将讨论稀疏恢复 （如基追踪、匹配追踪）和稀疏正则化 （如 LASSO、弹性网）中的关键算法，以及它们在实际问题中的应用。全书兼顾直观阐释与严谨的数学表述，为该领域的初学者和专家提供全面的参考资料。本书有两个目标：为初涉该领域的学生和研究人员提供一份内容自足的入门读物，并为希望在科学和工程领域应用稀疏优化的从业者提供严谨的参考资料。

本书的组织结构如下。第 2 章和 3 章通过回顾（凸）优化的基本概念及其与稀疏问题的关联，奠定基础。第 4∼7 章讨论如何设计能够保证稀疏信号精确恢复或稳定恢复的测量矩阵，包括基于随机矩阵理论和受限等距性质（RIP）的构造方法。第 8∼10 章介绍求解各种稀疏优化问题的算法，涵盖从贪心方法到凸松弛的方法。

在本章余下部分，我们将简要回顾基本数学符号和基础概念。为确保清晰，文中会根据需要进一步引入定义。

1.2. 线性代数

在所有情况下，标量都用非粗体字体表示，也可以带下标（例如， a、α、 α_i）。向量用**粗体**小写字母表示，也可以带下标（例如， $\boldsymbol{\mu}$、 $\boldsymbol{x}$、 $\boldsymbol{x}_n$、 $\boldsymbol{z}$）；矩阵则用**粗体**大写字母表示，也可以带下标（例如， $\boldsymbol{X}$、 $\boldsymbol{L}_j$）。向量 $\boldsymbol{z}$ 的第 i个元素将用非粗体字体 z_i 表示。为区分确定性量和随机量，我们采用以下约定：

- 标量随机变量用正体（直立体）表示（例如，a 和 b_1是随机变量，而斜体 a和 b_1 表示确定性标量）。随机向量用正体粗体小写字母表示，也可以带下标（例如，$\mathbf{a}$ 和 $\mathbf{b}_1$ 是随机向量，而粗斜体 $\boldsymbol{a}$和 $\boldsymbol{b}_1$ 表示确定性向量）。随机矩阵用正体粗体大写字母表示，也可以带下标（例如，$\mathbf{A}$ 和$\mathbf{B}_1$ 是随机矩阵，而粗斜体 $\boldsymbol{A}$ 和 $\boldsymbol{B}_1$表示确定性矩阵）。

子数组由固定矩阵的一个索引子集构成。矩阵 $\boldsymbol{X}$ 中位于第 i 行、第 j 列的元素（即 (i,j) 处的元素）记作 x_{ij}。因此，矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 可表示为 $\boldsymbol{X} = \{x_{ij}\}^{n,p}_{i,j=1} = [x_{ij}]$。为方便符号表示，我们采用**Matlab 风格索引**：

- 由 $\boldsymbol{X}$ 的第 i 行至第 j 行，以及第 k 列至第 m 列组成的子矩阵记作 $\boldsymbol{X}_{i:j,k:m} \equiv \boldsymbol{X}[i:j, k:m]$。冒号用于表示某一维度上的所有元素。例如，$\boldsymbol{X}_{:,k:m} \equiv \boldsymbol{X}[:, k:m]$ 表示由第 k 列至第 m 列构成的子矩阵；而 $\boldsymbol{X}_{:,k} \equiv \boldsymbol{X}[:, k]$ 表示 $\boldsymbol{X}$ 的第 k 列。此外，$\boldsymbol{X}$ 的第 k 列也可更简洁地记作 $\boldsymbol{x}_k$；$\boldsymbol{X}$ 的第 k 行可记作 $\boldsymbol{x}^{(k)}$。
-

当所选索引不连续时，令 $\mathbb{I}$ 和 $\mathbb{J}$ 为有序索引集。则：

- $\boldsymbol{X}[\mathbb{I}, \mathbb{J}]$ 表示从 $\boldsymbol{X}$ 中提取由 $\mathbb{I}$ 和 $\mathbb{J}$ 分别索引的 $\boldsymbol{X}$ 的行和列所得到的子矩阵。$\boldsymbol{X}[:, \mathbb{J}]$ 表示从 $\boldsymbol{X}$ 中提取由 $\mathbb{J}$ 索引的 $\boldsymbol{X}$ 的列所得到的子矩阵，其中冒号同样表示“所有行”。

定义 1.1（MATLAB表示法）。令 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，并令 $\mathbb{I} = \{i_1, i_2, \ldots, i_k\}$ 和 $\mathbb{J} = \{j_1, j_2, \ldots, j_l\}$ 为两个指标集。则 $\boldsymbol{X}[\mathbb{I}, \mathbb{J}]$ 表示 $k \times l$ 子矩阵

$$\boldsymbol{X}[\mathbb{I}, \mathbb{J}] = \begin{bmatrix} x_{i_1,j_1} & x_{i_1,j_2} & \cdots & x_{i_1,j_l} \\ x_{i_2,j_1} & x_{i_2,j_2} & \cdots & x_{i_2,j_l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i_k,j_1} & x_{i_k,j_2} & \cdots & x_{i_k,j_l} \end{bmatrix}.$$

类似地，$\boldsymbol{X}[\mathbb{I},:]$ 表示 $k \times p$ 子矩阵，而 $\boldsymbol{X}[:, \mathbb{J}]$ 则以类似方式表示 $n \times l$ 子矩阵。我们还应注意，索引的值域满足：

$$\begin{cases} 0 \leq \min(\mathbb{I}) \leq \max(\mathbb{I}) \leq n; \\ 0 \leq \min(\mathbb{J}) \leq \max(\mathbb{J}) \leq p. \end{cases}$$

更紧凑地，对于 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 和 $\boldsymbol{\beta} \in \mathbb{R}^p$，令

$$\boldsymbol{X}_{\mathbb{I}} \triangleq \boldsymbol{X}[:, \mathbb{I}] \in \mathbb{R}^{n \times k} \qquad \text{and} \qquad \boldsymbol{\beta}_{\mathbb{I}} \triangleq \boldsymbol{\beta}[\mathbb{I}] \in \mathbb{R}^k. \tag{1.1}$$

注意，$\boldsymbol{X}_{\mathbb{I}}^{\top}$表示子矩阵 $\boldsymbol{X}_{\mathbb{I}}$的转置，而不是 $\boldsymbol{X}^{\top}$的子矩阵。

另一方面，当指标集 $\mathbb{I}$ 作为函数的参数出现时——写作 $\boldsymbol{X}(\mathbb{I})$或 $\boldsymbol{\beta}(\mathbb{I})$——它表示一个投影算子，将所有索引不属于 $\mathbb{I}$的元素置零。具体来说：

- 对于 $\boldsymbol{\beta} \in \mathbb{R}^p$，向量 $\boldsymbol{\beta}(\mathbb{I}) \in \mathbb{R}^p$ 在 $\mathbb{I}$ 对应的元素上等于 $\boldsymbol{\beta}$，在其他位置为零；也就是说，若 $i \in \mathbb{I}$，则 $\boldsymbol{\beta}(\mathbb{I})_i = \beta_i$，否则为 0。对于 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，矩阵 $\boldsymbol{X}(\mathbb{I}) \in \mathbb{R}^{n \times p}$
- 保留 $\boldsymbol{X}$ 中由 $\mathbb{I}$索引的列，并将所有其他列置零。

采用这种符号，矩阵向量乘积可分解为

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}(\mathbb{I})\boldsymbol{\beta}(\mathbb{I}) + \boldsymbol{X}(\mathbb{I}^c)\boldsymbol{\beta}(\mathbb{I}^c) = \boldsymbol{X}[:, \mathbb{I}]\boldsymbol{\beta}[\mathbb{I}] + \boldsymbol{X}[:, \mathbb{I}^c]\boldsymbol{\beta}[\mathbb{I}^c] = \boldsymbol{X}_{\mathbb{I}}\boldsymbol{\beta}_{\mathbb{I}} + \boldsymbol{X}_{\mathbb{I}^c}\boldsymbol{\beta}_{\mathbb{I}^c},$$

其中，$\mathbb{I}^c$ 表示 $\mathbb{I}$的补集。

本书中，除非明确转置，否则所有向量均假定为列向量。行向量表示为列向量的转置，例如 $\boldsymbol{x}^{\top}$。具体向量采用 MATLAB风格语法书写。列向量中的元素以分号分隔

“;”用于分隔元素，例如，

$$\boldsymbol{x} = [1; 2; 3] \qquad \text{(column vector)}$$

是 $\mathbb{R}^3$ 中的列向量。类似地，行向量使用逗号“，”表示，例如，

$$\boldsymbol{y} = [1, 2, 3] \qquad \text{(row vector)}$$

是一个含有 3 个值的行向量。等价地，列向量可以表示为行向量的转置： $\boldsymbol{y} = [1, 2, 3]^\top$ 是一个列向量。

矩阵 $\boldsymbol{X}$ 的转置记为 $\boldsymbol{X}^\top$，其逆矩阵（如果存在）记为 $\boldsymbol{X}^{-1}$。 $p \times p$ 单位矩阵记为 $\boldsymbol{I}_p$，或简记为 $\boldsymbol{I}$。全由零组成的向量或矩阵记为**粗体**符号 $\mathbf{0}$；其维数可根据上下文推断，也可明确表示为 $\mathbf{0}_p$ ，表示一个 p 维零向量。类似地，全由 1 组成的向量或矩阵记为**粗体**数字 $\mathbf{1}$，其大小可根据上下文确定；或者，我们将 $\mathbf{1}_p$ 记作一个含有 p 个元素的全 1 向量。如果维数可根据上下文明确判断，通常会省略下标。

定义 1.2（特征值、特征向量）。设 $\boldsymbol{X} \in \mathbb{C}^{p \times p}$。标量 $\lambda \in \mathbb{C}$ 称为 $\boldsymbol{X}$ 的（右）特征值（也称为本征值或特征值），若存在非零向量 $\boldsymbol{u} \in \mathbb{C}^p$ 使得

$$\boldsymbol{X}\boldsymbol{u} = \lambda \boldsymbol{u}.$$

在这种情况下， $\boldsymbol{u}$ 称为与 λ 对应的 $\boldsymbol{X}$ 的（右）特征向量。

除非另有说明，我们只讨论实矩阵。除非明确指出，所有讨论的特征值也都假定为实数。

直观地说，特征向量 $\boldsymbol{u}$ 表示矩阵 $\boldsymbol{X}$ 在 $\mathbb{R}^p$ 中的一个方向，该方向在由 $\boldsymbol{X}$定义的线性变换下保持不变：将 $\boldsymbol{X}$ 作用于 $\boldsymbol{u}$不会改变其方向，只会按相应的特征值 λ缩放它。请注意，实矩阵一般可能具有复特征值，但实对称矩阵的特征值始终为实数（见定理1.46）。

这对（λ， $\boldsymbol{u}$）通常称为特征对。需要注意的是，特征向量并不唯一：如果 $\boldsymbol{u}$ 是一个特征向量，那么它的任何非零标量倍数 $\eta\boldsymbol{u}$ （其中$\eta \in \mathbb{R} \setminus \{\mathbf{0}\}$）也都是特征向量。为消除这种不确定性，通常会对特征向量进行归一化，例如要求 $\|\boldsymbol{u}\|_2 = 1$ （单位 ℓ_2范数；参见第1.2.1节）。此外，由于 $\boldsymbol{u}$ 和$-\boldsymbol{u}$ 对应于同一特征值，且除符号外表示相同方向，因此通常按照惯例固定符号（例如，要求 $\boldsymbol{u}$ 的第一个非零元素为正）。

关于复矩阵特征值的一个重要结果是格尔什戈林圆盘定理 （也称为*Gershgorin* 圆盘定理），由苏联数学家 Semen Aronovich Gershgorin 于 1931年提出（格尔什戈林，1931年）。

定理 1.3：（格尔什戈林圆盘定理（格尔什戈林，1931年）） 设 $\boldsymbol{X} = [x_{ij}]$ 是一个 $p \times p$ 复矩阵。对于每一行 i，定义格尔什戈林半径 R_i 为

$$R_i \triangleq \sum_{j=1, j \neq i}^{p} |x_{ij}| .$$

那么， λ 的每个特征值都位于以对角元素 x_{ii} 为中心、半径为 R_i的至少一个格尔什戈林圆盘内。换言之，对于每个特征值 λ，都存在

索引 i 使得:

$$|\lambda - x_{ii}| \leq R_i.$$

这意味着， $\boldsymbol{X}$ 的所有特征值都位于复平面上这些 p 个圆盘的并集内。每个圆盘以对角元素 x_{ii}为中心，其半径等于对应行中非对角元素绝对值之和。格尔什戈林圆盘定理提供了一种简单有效的方法，无需显式计算特征值即可估计其位置。它在分析矩阵的性质、稳定性以及判断矩阵是否可逆时尤其有用。例如，如果格尔什戈林圆盘都不包含原点，那么 0 就不可能是 $\boldsymbol{X}$的特征值，因此$\boldsymbol{X}$ 可逆。

在线性代数中，每个向量空间都有一个基，并且该空间中的每个向量都可以唯一地表示为基向量的线性组合。基于这一思想，我们用基的概念来定义子空间的张成空间和维数。

定义 1.4（子空间与张成空间）。非空子集 $\mathcal{V}$ of $\mathbb{R}^n$ 若对线性组合封闭，则称为子空间；也就是说，对于每个 $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}$以及每个 $x, y \in \mathbb{R}$，均满足 $x\boldsymbol{a} + y\boldsymbol{b} \in \mathcal{V}$。此外，如果子空间 $\mathcal{V}$ 中的每个向量 $\boldsymbol{v}$ 都可以写成向量 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$的线性组合，那么我们就说这些向量张成子空间 $\mathcal{V}$。

向量集的线性无关性是线性代数的核心概念。下面给出两种等价的定义。

定义 1.5（线性无关）。如果不存在一种组合可以得到 $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \ldots + a_n\boldsymbol{x}_n = 0$, 除非所有 a_i 都为零，那么向量集 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ 就称为线性无关。另一种等价的定义是 $\boldsymbol{x}_1 \neq \boldsymbol{0}$, 并且对于每个 $k > 1$，向量 $\boldsymbol{x}_k$ 都不属于 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{k-1}\}$ 的张成空间。

定义 1.6（基与维数）。向量集 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ 称为子空间 $\mathcal{V}$ 的基， $\mathcal{V}$ 当且仅当这些向量线性无关且张成 $\mathcal{V}$。给定子空间的每个基都包含相同数量的向量。这个数量称为子空间的维数 $\mathcal{V}$。按照惯例，零子空间 $\{\boldsymbol{0}\}$ 的维数为零。此外，任何正维数的子空间都存在一组正交基——也就是说，总能选取一组由彼此正交的向量构成的基（参见定义 1.39）。

定义 1.7（列空间（值域）与行空间）。设 $\boldsymbol{X}$ 是一个 $n \times p$ 实矩阵。其列空间（或值域）定义为可表示为其各列的线性组合的所有向量构成的集合:

$$\mathcal{C}(\boldsymbol{X}) = \{\boldsymbol{v} \in \mathbb{R}^n \mid \exists\, \boldsymbol{u} \in \mathbb{R}^p,\, \boldsymbol{v} = \boldsymbol{X}\boldsymbol{u}\}.$$

同样， $\boldsymbol{X}$ 的行空间是由其各行张成的所有向量组成的集合。等价地，它是 $\boldsymbol{X}^\top$的列空间:

$$\mathcal{C}(\boldsymbol{X}^\top) = \{\boldsymbol{u} \in \mathbb{R}^p \mid \exists\, \boldsymbol{v} \in \mathbb{R}^n,\, \boldsymbol{u} = \boldsymbol{X}^\top\boldsymbol{v}\}.$$

定义 1.8（零空间（nullspace，核））。设 $\boldsymbol{X}$ 是一个 $n \times p$ 实矩阵。 $\boldsymbol{X}$ 的零空间（或核，或零空间）定义为集合：

$$\mathcal{N}(\boldsymbol{X}) = \{\boldsymbol{v} \in \mathbb{R}^p \mid \boldsymbol{X}\boldsymbol{v} = \boldsymbol{0}\}.$$

在某些情况下， $\boldsymbol{X}$ 的零空间也称为 $\boldsymbol{X}$的右零空间 。 $\boldsymbol{X}^\top$ 的零空间为

$$\mathcal{N}(\boldsymbol{X}^\top) = \{\boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{X}^\top\boldsymbol{u} = \boldsymbol{0}\},$$

通常称为 $\boldsymbol{X}$的左零空间 。

定义 1.9（秩）。矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的 $rank$ 定义为其列空间的维数。也就是说， $\boldsymbol{X}$ 的秩等于 $\boldsymbol{X}$中线性无关列的最大数量，也等于 $\boldsymbol{X}$中线性无关行的最大数量。矩阵 $\boldsymbol{X}$ 与其转置 $\boldsymbol{X}^\top$ 的秩总是相同的。如果 $\boldsymbol{X}$ 的秩等于 $\min\{n, p\}$，则称其为满秩 。具体而言，给定向量 $\boldsymbol{u} \in \mathbb{R}^n$ 和向量 $\boldsymbol{v} \in \mathbb{R}^p$，由它们的外积得到的 $\boldsymbol{u}\boldsymbol{v}^\top$ 矩阵 $n \times p$ 的秩为 1。总之，矩阵的秩可概括为：

- 线性无关列的数量；线性无关行的数量；值得注意的是，这两个数量总
- 是相等（见引理 1.12）。
-

$\boldsymbol{X}$ 的列空间和 $\boldsymbol{X}^\top$ 的零空间都是 $\mathbb{R}^n$的子空间。事实上， $\mathcal{N}(\boldsymbol{X}^\top)$中的每个向量都与 $\mathcal{C}(\boldsymbol{X})$中的每个向量正交，反之亦然。同样， $\mathcal{N}(\boldsymbol{X})$中的每个向量也都与 $\mathcal{C}(\boldsymbol{X}^\top)$中的每个向量正交，反之亦然。这引出了正交补子空间的概念。

定义 1.10（一般的正交补）。子空间 $\mathcal{V} \subseteq \mathbb{R}^n$ 的正交补 $\mathcal{V}^\perp \subseteq \mathbb{R}^n$ 由所有与 $\mathcal{V}$中的每个向量都垂直的向量组成。形式化地说，

$$\mathcal{V}^\perp = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}^\top\boldsymbol{u} = 0,\ \forall\, \boldsymbol{u} \in \mathcal{V}\}.$$

这两个子空间互不相交，且它们共同张成整个空间 $\mathbb{R}^n$。 $\mathcal{V}$ 和 $\mathcal{V}^\perp$ 的维数之和等于整个空间的维数：$\dim(\mathcal{V}) + \dim(\mathcal{V}^\perp) = n$。此外，$(\mathcal{V}^\perp)^\perp = \mathcal{V}$。

定义 1.11（列空间的正交补）。设 $\boldsymbol{X}$是一个 $n \times p$实矩阵。列空间 $\mathcal{C}(\boldsymbol{X})$ 的正交补记为 $\mathcal{C}^\perp(\boldsymbol{X})$，它是由下式定义的子空间：

$$\begin{aligned}\mathcal{C}^\perp(\boldsymbol{X}) &= \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}^\top\boldsymbol{X}\boldsymbol{u} = \boldsymbol{0}, \forall\, \boldsymbol{u} \in \mathbb{R}^p\} \\ &= \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}^\top\boldsymbol{w} = \boldsymbol{0}, \forall\, \boldsymbol{w} \in \mathcal{C}(\boldsymbol{X})\}.\end{aligned}$$

对于任意秩为 r 的矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，我们得到四个基本子空间：

- $\mathcal{C}(\boldsymbol{X})$： $\boldsymbol{X}$ 的列空间，即列的线性组合，维数为 r。$\mathcal{N}(\boldsymbol{X})$： $\boldsymbol{X}$ 的（右）零空间，即满
- 足 $\boldsymbol{X}\boldsymbol{u} = \boldsymbol{0}$ 的所有 $\boldsymbol{u}$，维数为 $p - r$。$\mathcal{C}(\boldsymbol{X}^\top)$： $\boldsymbol{X}$ 的行空间，即行的线性组合，维数
- 为 r。

- $\mathcal{N}(\boldsymbol{X}^\top)$: $\boldsymbol{X}$ 的左零空间，即满足 $\boldsymbol{X}^\top\boldsymbol{v}=\boldsymbol{0}$ 的所有 $\boldsymbol{v}$，维数为 $n-r$。

此外，$\mathcal{N}(\boldsymbol{X})$ 是 $\mathcal{C}(\boldsymbol{X}^\top)$ 的正交补，而 $\mathcal{C}(\boldsymbol{X})$ 是 $\mathcal{N}(\boldsymbol{X}^\top)$ 的正交补。这称为线性代数基本定理。证明见Lu(2022年。

现在我们来证明定义1.9中所述的列空间维数与行空间维数相等。

引理 1.12：（列空间与行空间的维数）矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 的列空间维数等于其行空间维数，即矩阵 $\boldsymbol{X}$ 的行秩和列秩相同。

引理 1.12] 的证明 [首先注意到，$\boldsymbol{X}$ 的零空间是 $\boldsymbol{X}$ 的行空间的正交补：$\mathcal{N}(\boldsymbol{X})\perp\mathcal{C}(\boldsymbol{X}^\top)$（其中，$\boldsymbol{X}$ 的行空间等价于 $\boldsymbol{X}^\top$ 的列空间）。也就是说，$\boldsymbol{X}$ 零空间中的向量与 $\boldsymbol{X}$ 行空间中的向量正交。为说明这一点，设 $\boldsymbol{X}$ 的行分块为 $\boldsymbol{X}=[\boldsymbol{x}_1^\top;\ \boldsymbol{x}_2^\top;\ ...;\ \boldsymbol{x}_n^\top]$。对于任意向量 $\boldsymbol{\beta}\in\mathcal{N}(\boldsymbol{X})$，有 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{0}$；即 $[\boldsymbol{x}_1^\top\boldsymbol{\beta};\ \boldsymbol{x}_2^\top\boldsymbol{\beta};\ ...;\ \boldsymbol{x}_n^\top\boldsymbol{\beta}]=\boldsymbol{0}$。由于 $\boldsymbol{X}$ 的行空间由 $\{\boldsymbol{x}_1^\top,\boldsymbol{x}_2^\top,\ldots,\boldsymbol{x}_n^\top\}$ 张成，因此 $\boldsymbol{\beta}$ 与来自 $\mathcal{C}(\boldsymbol{X}^\top)$ 的任意向量垂直。这表明 $\mathcal{N}(\boldsymbol{X})\perp\mathcal{C}(\boldsymbol{X}^\top)$。

现在，假设 $\boldsymbol{X}$ 的行空间维数为 r，令 $\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_r$ 为 $\mathbb{R}^p$ 中的一组向量，并构成行空间的一组基。因此，r 个向量 $\boldsymbol{X}\boldsymbol{r}_1,\boldsymbol{X}\boldsymbol{r}_2,\ldots,\boldsymbol{X}\boldsymbol{r}_r$属于 $\boldsymbol{X}$的列空间；此外，它们线性无关。为说明这一点，假设我们有这些 r 个向量的一个线性组合：$\beta_1\boldsymbol{X}\boldsymbol{r}_1+\beta_2\boldsymbol{X}\boldsymbol{r}_2+\ldots+\beta_r\boldsymbol{X}\boldsymbol{r}_r=0$，也就是说，$\boldsymbol{X}(\beta_1\boldsymbol{r}_1+\beta_2\boldsymbol{r}_2+\ldots+\beta_r\boldsymbol{r}_r)=0$，且向量 $\boldsymbol{v}=\beta_1\boldsymbol{r}_1+\beta_2\boldsymbol{r}_2+\ldots+\beta_r\boldsymbol{r}_r$属于 $\boldsymbol{X}$的零空间。但是，由于 $\{\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_r\}$ 是 $\boldsymbol{X}$ 行空间的一组基，$\boldsymbol{v}$也属于 $\boldsymbol{X}$的行空间。我们已经证明，来自 $\boldsymbol{X}$ 零空间的向量与来自 $\boldsymbol{X}$ 行空间的向量正交，因此 $\boldsymbol{v}^\top\boldsymbol{v}=0$和 $\beta_1=\beta_2=\ldots=\beta_r=0$。于是，$\boldsymbol{X}\boldsymbol{r}_1,\boldsymbol{X}\boldsymbol{r}_2,\ldots,\boldsymbol{X}\boldsymbol{r}_r$ 属于 $\boldsymbol{X}$的列空间，且它们线性无关。这意味着 $\boldsymbol{X}$ 列空间的维数大于 r。这一结果表明，$\boldsymbol{X}\leq$ 的行秩 $\boldsymbol{X}$列秩。

对 $\boldsymbol{X}^\top$ 应用相同的论证，可得反向不等式：**列秩$\boldsymbol{X}\leq$ 行秩** $\boldsymbol{X}$。因此，两个秩相等，证明完毕。 ■

从前面的证明中，我们还可以得出：如果 $\{\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_r\}$ 是 $\boldsymbol{X}$的行空间的一组基，那么 $\{\boldsymbol{X}\boldsymbol{r}_1,\boldsymbol{X}\boldsymbol{r}_2,\ldots,\boldsymbol{X}\boldsymbol{r}_r\}$ 构成 $\boldsymbol{X}$的列空间的一组基。我们将在下面的引理中对这一观察作出正式表述。

引理 1.13：（由行基得到列基）对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$，设$\{\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_r\}$ 是 $\mathbb{R}^p$中的一个向量集，它构成 $\boldsymbol{X}$的行空间的一组基。那么，集合 $\{\boldsymbol{X}\boldsymbol{r}_1,\boldsymbol{X}\boldsymbol{r}_2,\ldots,\boldsymbol{X}\boldsymbol{r}_r\}$ 构成 $\boldsymbol{X}$的列空间的一组基。

定义 1.14（置换矩阵）。置换矩阵置换矩阵 $\boldsymbol{P}\in\mathbb{R}^{p\times p}$ 是一个方形二进制矩阵，每一行和每一列中恰有一个元素为 1，其余元素均为 0。

行置换点。也就是说，置换矩阵 $\boldsymbol{P}$ 的各行以任意顺序排列于单位矩阵 $\boldsymbol{I}$ 中，而排列顺序决定了行置换的序列。若要置换矩阵 $\boldsymbol{X}$ 的各行，我们在左侧乘以 $\boldsymbol{P}\boldsymbol{X}$。

列置换点。或者，等价地，置换矩阵 $\boldsymbol{P}$ 的各列以任意顺序排列于单位矩阵 $\boldsymbol{I}$ 中，而排列顺序决定了列置换的序列。要对 $\boldsymbol{X}$ 应用列置换，我们在右侧乘以 $\boldsymbol{XP}$。

置换矩阵 $\boldsymbol{P}$ 可以用指标集 $\mathbb{J} \in \mathbb{Z}_+^p$ 表示得更简洁，其中 $\boldsymbol{P} = \boldsymbol{I}[:, \mathbb{J}]$，$\boldsymbol{I}$ 是单位矩阵。值得注意的是，$\mathbb{J}$ 的元素是 $\{1, 2, \ldots, p\}$的一个排列，因此它们的和为 $1 + 2 + \ldots + p = \frac{p^2+p}{2}$。 ——

定义 1.15（行列式：按子式展开的拉普拉斯展开）。设 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$为一个方阵，令 $\boldsymbol{X}_{ij} \in \mathbb{R}^{(p-1)\times(p-1)}$ 表示从 $\boldsymbol{X}$ 中删除第 i行和第 j列后得到的子矩阵。$\boldsymbol{X}$ 的行列式可以通过以下公式递归计算：

$$\det(\boldsymbol{X}) = \sum_{k=1}^{p} (-1)^{i+k} x_{ik} \det(\boldsymbol{X}_{ik}) = \sum_{k=1}^{p} (-1)^{k+j} x_{kj} \det(\boldsymbol{X}_{kj}), \tag{1.2}$$

其中，第一个方程是沿第 i行的按子式展开的拉普拉斯展开，第二个方程是沿第 j列的按子式展开的拉普拉斯展开。等价地，对于任意满足 $1 \leq r \leq p$的整数 r，令 $\mathbb{J} \subseteq \{1, 2, \ldots, p\}$为基数为 r($|\mathbb{J}| = r$)的指标集，并令 $\mathbb{J}^c = \{1, 2, \ldots, n\} \backslash \mathbb{J}$ 表示其补集。于是

$$\det(\boldsymbol{X}) = \sum_{\mathbb{I}} (-1)^{\gamma} \det(\boldsymbol{X}[\mathbb{I}, \mathbb{J}]) \det(\boldsymbol{X}[\mathbb{I}^c, \mathbb{J}^c]) = \sum_{\mathbb{I}} (-1)^{\gamma} \det(\boldsymbol{X}[\mathbb{J}, \mathbb{I}]) \det(\boldsymbol{X}[\mathbb{J}^c, \mathbb{I}^c]),$$

其中 $\gamma = \sum_{i\in\mathbb{I}} i + \sum_{j\in\mathbb{J}} j$，求和遍历所有基数为 $\mathbb{I} \subseteq \{1, 2, \ldots, p\}$ 的指标集 r。当 $r = 1$ 时，这就简化为（1.2）。

行列式为每个方阵赋予一个标量值，用来编码相关线性变换的几何和代数性质。对于一个 2×2 矩阵，行列式的绝对值等于其列向量张成的平行四边形的面积。符号表示定向：如果各列构成逆时针基，则行列式为正；如果构成顺时针基，则为负。对于一个 3×3 矩阵，行列式给出由其三个列向量确定的平行六面体的带符号体积，其符号同样反映定向。一般而言，对于一个由 $p \times p$ 矩阵表示的线性变换，行列式的绝对值衡量该变换对 p维体积进行缩放的倍数。行列式为正意味着空间的定向（或“手性”）得以保留；行列式为负则意味着定向被反转。方阵可逆，当且仅当其行列式非零。从几何上看，这意味着该变换不会将空间压缩到低维子空间或单个点——而这两种情况恰好都发生在行列式为零时。从代数上看，矩阵的行列式等于其特征值的乘积（按重数计）。因此，它反映了变换沿其特征向量方向产生的累积缩放效应。

现在我们总结行列式的关键性质。

引理 1.16：（行列式的性质）行列式满足以下性质：

- 两个矩阵乘积的行列式为 $\det(\boldsymbol{XY}) = \det(\boldsymbol{X})\det(\boldsymbol{Y})$；转置矩阵的行列式为
- $\det(\boldsymbol{X}^\top) = \det(\boldsymbol{X})$；如果 λ 是 $\boldsymbol{X}$的一个特征值，则 $\det(\boldsymbol{X} - \lambda\boldsymbol{I}) = 0$；任意
- 单位矩阵的行列式为 1；
-

- 对于任意正交矩阵 $\boldsymbol{Q}$（即 $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$），我们有

$$\det(\boldsymbol{Q}) = \det(\boldsymbol{Q}^\top) = \pm 1, \qquad \text{since } \det(\boldsymbol{Q}^\top)\det(\boldsymbol{Q}) = \det(\boldsymbol{Q}^\top\boldsymbol{Q}) = \det(\boldsymbol{I}) = 1;$$

- 对于任意方阵 $\boldsymbol{X}$ 和正交矩阵 $\boldsymbol{Q}$，有

$$\det(\boldsymbol{X}) = \det(\boldsymbol{Q}^\top)\det(\boldsymbol{X})\det(\boldsymbol{Q}) = \det(\boldsymbol{Q}^\top\boldsymbol{X}\boldsymbol{Q});$$

- 假设 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$，则 $\det(-\boldsymbol{X}) = (-1)^p\det(\boldsymbol{X})$。

正定性和半正定性是矩阵可能具有的最重要性质之一。

定义 1.17（正定和半正定）。如果对所有非零 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都有 $\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} > 0$，则称对称矩阵 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为正定（*PD*），记作 $\boldsymbol{X} \succ \boldsymbol{0}$。如果对所有 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都有 $\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} \geq 0$，则称其为半正定（*PSD*），记作 $\boldsymbol{X} \succeq \boldsymbol{0}$。 [a] [b]

[a]。在本书中，正（半）定矩阵始终假设为对称矩阵。正定性和半正定性的概念仅对对称矩阵有意义。[b]。如果对所有非零 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都有 $\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} < 0$，则称对称矩阵 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为 *negative definite*（ND）；如果对所有 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都有 $\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} \leq 0$，则称对称矩阵 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为 *negative semidefinite*（NSD）；如果存在 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha} \in \mathbb{R}^p$，使得 $\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} < 0$ 和 $\boldsymbol{\alpha}^\top\boldsymbol{X}\boldsymbol{\alpha} > 0$，则称对称矩阵 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为 *indefinite*（ID）。

线性代数中的一个基本结果指出，对称矩阵 $\boldsymbol{X}$ 是正定的，当且仅当它的所有特征值都严格为正。同样，$\boldsymbol{X}$ 是半正定的，当且仅当它的所有特征值均非负。由此得到以下定理：

定理 1.18：（特征值刻画定理）矩阵 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 是正定的，当且仅当它的所有特征值均为正。同样，矩阵 $\boldsymbol{X}$ 是半正定的，当且仅当它的所有特征值均为非负。 [a] 此外，我们有以下推论：

- $\boldsymbol{X} - \gamma\boldsymbol{I} \succeq \boldsymbol{0}$ 当且仅当 $\lambda_{\min}(\boldsymbol{X}) \geq \gamma$；
- $\boldsymbol{X} - \gamma\boldsymbol{I} \succ \boldsymbol{0}$ 当且仅当 $\lambda_{\min}(\boldsymbol{X}) > \gamma$；
- $\boldsymbol{X} - \gamma\boldsymbol{I} \preceq \boldsymbol{0}$ 当且仅当 $\lambda_{\max}(\boldsymbol{X}) \leq \gamma$；
- $\boldsymbol{X} - \gamma\boldsymbol{I} \prec \boldsymbol{0}$ 当且仅当 $\lambda_{\max}(\boldsymbol{X}) < \gamma$；$\lambda_{\min}$
- $(\boldsymbol{X})\ \boldsymbol{I} \preceq \boldsymbol{X} \preceq \lambda_{\max}(\boldsymbol{X})\ \boldsymbol{I}$,

其中，$\lambda_{\min}(\boldsymbol{X})$ 和 $\lambda_{\max}(\boldsymbol{X})$ 分别表示 $\boldsymbol{X}$ 的最小特征值和最大特征值，而 $\boldsymbol{B} \prec \boldsymbol{C}$ 表示 $\boldsymbol{C} - \boldsymbol{B}$ 是半正定的。

[a]。正（半）定矩阵的迹、行列式和主子式将在问题1.1中讨论。

该定理通过特征值给出了正定性和半正定性的另一种刻画。给定一个特征对（λ，$\boldsymbol{\beta}$），其所属的 $\boldsymbol{X}$，可以验证正向推论：$\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} = \lambda\boldsymbol{\beta}^\top\boldsymbol{\beta} > 0$ 使得 $\lambda = (\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta}) / (\boldsymbol{\beta}^\top\boldsymbol{\beta}) > 0$（分别为$\geq 0$）当 $\boldsymbol{X}$ 为正定（分别为半正定）时成立。借助谱定理（定理1.46），可以证明完整的等价性。

半正定矩阵中一个值得注意的子类是某些对角占优矩阵，我们将在下文给出其严格定义。

定义 1.19（对角占优矩阵）。设 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 是一个对称矩阵。那么 $\boldsymbol{X}$ 称为对角占优，如果

$$|x_{ii}| \geq \sum_{j\neq i} |x_{ij}|, \qquad \forall\, i \in \{1,2,\ldots,p\},$$

且严格对角占优当且仅当

$$|x_{ii}| > \sum_{j\neq i} |x_{ij}|, \qquad \forall\, i \in \{1,2,\ldots,p\}.$$

现在我们证明：对角元素非负的对角占优矩阵 是半正定的，而对角元素为正的严格对角占优矩阵 是正定的。

定理 1.20：（对角占优矩阵的正定性）设$\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为对称矩阵。则：

（i）如果 $\boldsymbol{X}$ 是对角占优且对角元素非负的，那么 $\boldsymbol{X}$ 是半正定的。（ii）如果 $\boldsymbol{X}$ 是严格对角占优且对角元素为正的，那么 $\boldsymbol{X}$ 是正定的。

该结果可直接由格尔什戈林圆盘定理（定理 1.3）得到。不过，下面我们将基于特征值分析给出一个初等证明。

证明 [定理 1.20](i). 反设 $\boldsymbol{X}$ 不是半正定矩阵，且有一个负特征值 λ 对应于特征向量 $\boldsymbol{v}$，满足 $\boldsymbol{X}\boldsymbol{v} = \lambda\boldsymbol{v}$。令 i 为一个索引，使得 $|v_i| = \max_{1\leq j\leq p} |v_j|$。考虑 i-th 个分量 $(\boldsymbol{X} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$，可得

$$|x_{ii} - \lambda| \cdot |v_i| = \left|\sum_{j\neq i} x_{ij}v_j\right| \leq \left(\sum_{j\neq i} |x_{ij}|\right) |v_i| \leq |x_{ii}| \cdot |v_i|.$$

这意味着 $|x_{ii} - \lambda| \leq |x_{ii}|$。由于对角元素非负，因此有 $x_{ii} < x_{ii} - \lambda \leq x_{ii}$，从而导致矛盾。

(ii). 由 (i) 可知，$\boldsymbol{X}$ 是半正定的。反设 $\boldsymbol{X}$ 不是正定的，且有一个特征值为 0 的零特征值，对应于非零特征向量$\boldsymbol{v}$，满足 $\boldsymbol{X}\boldsymbol{v} = \boldsymbol{0}$。同理，可得

$$|x_{ii}| \cdot |v_i| = \left|\sum_{j\neq i} x_{ij}v_j\right| \leq \left(\sum_{j\neq i} |x_{ii}|\right) |v_i| < |x_{ii}| \cdot |v_i|,$$

这是不可能的，因此结论成立。 ■

习题1.21（半正定矩阵的幂）。设 $\boldsymbol{X}$ 为半正定矩阵。证明 $\boldsymbol{X}^k$ 对$k = 1$，2，.....也为半正定矩阵。

回顾入门线性代数课程，我们知道以下这些众所周知的等价条件可用于刻画非奇异（即可逆）方阵。

注1.22（矩阵非奇异性的等价条件列表）。设 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 为一个方阵。以下陈述彼此等价：

- $\boldsymbol{X}$ 是非奇异的；[a]$\boldsymbol{X}$ 可逆，即 $\boldsymbol{X}^{-1}$ 存在；$\boldsymbol{X}\boldsymbol{u} = \boldsymbol{y}$ 有唯一解 $\boldsymbol{u} = \boldsymbol{X}^{-1}\boldsymbol{y}$；
- $\boldsymbol{X}\boldsymbol{u} = \boldsymbol{0}$ 有唯一的平凡解：$\boldsymbol{u} = \boldsymbol{0}$；$\boldsymbol{X}$ 的列线性无关；$\boldsymbol{X}$ 的行线性无关；
- $\det(\boldsymbol{X}) \neq 0$；$\dim(\mathcal{N}(\boldsymbol{X})) = 0$；$\mathcal{N}(\boldsymbol{X}) = \{\boldsymbol{0}\}$，即零空间是平凡的；$\mathcal{C}(\boldsymbol{X})$ $= \mathcal{C}(\boldsymbol{X}^\top) = \mathbb{R}^p$，即列空间或行空间张成整个 $\mathbb{R}^p$；$\boldsymbol{X}$ 满秩 $r = p$；$\boldsymbol{X}^\top\boldsymbol{X}$ 是对称正定（PD）的；$\boldsymbol{X}$ 有 p 个非零（正）奇异值；所有特征值均非零。

[a]。其名称来源于奇异值分解（SVD；参见第 1.2.4节）。

牢记这些等价关系至关重要，因为它们在理论和应用中都经常出现。同样，下面的注释汇总了奇异矩阵的等价刻画，尤其是在特征值的背景下。

注1.23（矩阵奇异性的等价条件列表）。设 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$是一个方阵，且 $(\lambda, \boldsymbol{u})$ 是 $\boldsymbol{X}$的一个特征对。以下陈述等价：

- $(\boldsymbol{X} - \lambda\boldsymbol{I})$ 是奇异的；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 不可逆；$(\boldsymbol{X} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$ 有非零的 $\boldsymbol{v} \neq \boldsymbol{0}$ 解，且 $\boldsymbol{v} = \boldsymbol{u}$ 是其中一个解；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的列线性相关；$\det(\boldsymbol{X} - \lambda\boldsymbol{I}) = 0$；$\dim(\mathcal{N}($ $\boldsymbol{X} - \lambda\boldsymbol{I})) > 0$；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的零空间非平凡；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的列线性相关；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的行线性相关；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的秩为 $r < p$；列空间 $=$ 的维数 $= r < p$ 行空间的维数；$(\boldsymbol{X} - \lambda\boldsymbol{I})^\top(\boldsymbol{X} - \lambda\boldsymbol{I})$ 是对称半正定的；$(\boldsymbol{X} - \lambda\boldsymbol{I})$ 有 $r < p$ 个非零（正）奇异值；零是 $(\boldsymbol{X} - \lambda\boldsymbol{I})$ 的一个特征值。

1.2.1 向量范数、对偶范数、球

范数的概念对于衡量向量的大小至关重要，因此也使我们能够在赋范线性空间上定义度量。范数定量地衡量向量或矩阵的“大小”，这在许多应用中都不可或缺，例如计算向量在欧几里得空间中的长度，或评估多维情形下矩阵的尺度。

此外，范数使我们能够定义向量或矩阵之间的距离。具体而言，向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的距离可以计算为它们之差 $\|\boldsymbol{u}-\boldsymbol{v}\|$ 的范数。这一概念对于依赖邻近度量的任务至关重要，包括机器学习中的聚类算法。[1]

对于任意向量或矩阵，范数必须满足以下三个性质；例如，参见 Lu (2021a) 以了解更详细的介绍。

定义 1.24（向量范数和矩阵范数）。令 $\|\cdot\|$表示定义在向量或矩阵上的范数。那么，对于任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和任意向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，以下性质必须成立：

- 非负性。 $\|\boldsymbol{X}\| \geq 0$ 和 $\|\boldsymbol{\beta}\| \geq 0$，当且仅当 $\boldsymbol{X}=\boldsymbol{0}$ 或 $\boldsymbol{\beta}=\boldsymbol{0}$时等号成立。正齐次性。对于任意尺度 $\lambda \in \mathbb{R}$， $\|\lambda \boldsymbol{X}\| = |\lambda| \cdot \|\boldsymbol{X}\|$ 和 $\|\lambda \boldsymbol{\beta}\| = |\lambda| \cdot \|\boldsymbol{\beta}\|$。三角不等式。
- 对于所有矩阵 $\boldsymbol{X}$， $\boldsymbol{Y} \in \mathbb{R}^{n\times p}$， $\|\boldsymbol{X}+\boldsymbol{Y}\| \leq \|\boldsymbol{X}\| + \|\boldsymbol{Y}\|$，以及对于所有向量
- $\boldsymbol{\alpha}$， $\boldsymbol{\beta} \in \mathbb{R}^p$， $\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| \leq \|\boldsymbol{\alpha}\| + \|\boldsymbol{\beta}\|$。

基于这个定义，我们现在介绍向量的几种特定范数： ℓ_1、 ℓ_2 和 ℓ_∞-范数。

定义 1.25（向量 ℓ_1、 ℓ_2、 ℓ_∞、 ℓ_s-范数）。对于向量 $\boldsymbol{\beta} \in \mathbb{R}^p$， ℓ_2 向量范数定义为 $\|\boldsymbol{\beta}\|_2 = \sqrt{\beta_1^2+\beta_2^2+\ldots+\beta_p^2}$。类似地， ℓ_1-范数可由 $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^p |\beta_i|$. 得到。而 ℓ_∞-范数可由 $\|\boldsymbol{\beta}\|_\infty = \max_{i=1、2、\ldots\ldots、p} |\beta_i|$. 得到。更一般地，$\ell_s$-范数定义为 $\|\boldsymbol{\beta}\|_s = \sqrt[s]{\sum_{i=1}^p |\beta_i|^s}$，其中 $s \geq 1$。[a]

[a]。在本书中，为避免与维数 p 混淆，我们使用“ℓ_s-范数”一词，而不使用“ℓ_p-范数”。

习题 1.26（向量范数的标准界）。设 $\boldsymbol{\beta} \in \mathbb{R}^p$。利用柯西–施瓦茨不等式（见定理1.59），证明以下关系式：

$$\begin{aligned}\|\boldsymbol{\beta}\|_\infty \leq \|\boldsymbol{\beta}\|_1 \leq p\,\|\boldsymbol{\beta}\|_\infty;\\ \|\boldsymbol{\beta}\|_\infty \leq \|\boldsymbol{\beta}\|_2 \leq \sqrt{p}\,\|\boldsymbol{\beta}\|_\infty;\\ \|\boldsymbol{\beta}\|_2 \leq \|\boldsymbol{\beta}\|_1 \leq \sqrt{p}\,\|\boldsymbol{\beta}\|_2.\end{aligned}$$

上述习题表明，在有限维向量空间上， ℓ_1、 ℓ_2和 ℓ_∞ 向量范数是等价的。给定正常数 c、$C>0$，对于两个范数 $\|\cdot\|_a$ 和$\|\cdot\|_b$，若不等式 $c\cdot\|\boldsymbol{\beta}\|_a \leq \|\boldsymbol{\beta}\|_b \leq C\cdot\|\boldsymbol{\beta}\|_a$ 对所有向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都成立，则意味着这些范数诱导出相同的“接近”概念或拓扑——它们定义了相同的收敛序列、开集、连续性等。

对偶范数

考虑 ℓ_s 向量范数。根据 *H¨older* 不等式（见定理 1.60），我们有 $\boldsymbol{\beta}^\top\boldsymbol{\alpha} \leq \|\boldsymbol{\beta}\|_s \|\boldsymbol{\alpha}\|_t$，其中 s, $t>1$ 满足 $\frac{1}{s}+\frac{1}{t}=1$，且 $\boldsymbol{\alpha}$， $\boldsymbol{\beta} \in \mathbb{R}^p$。当两个序列 $\{|\beta_i|^s\}$和 $\{|\alpha_i|^t\}$线性相关时，等号成立。这意味着

$$\max_{\|\boldsymbol{\alpha}\|_t=1} \boldsymbol{\beta}^\top\boldsymbol{\alpha} = \|\boldsymbol{\beta}\|_s. \tag{1.3}$$

1. 我们仅讨论实向量空间和矩阵空间上的范数（及内积）。大多数结果都可直接推广到复数情形。

因此， $\|\cdot\|_t$ 称为 $\|\cdot\|_s$的对偶范数。另一方面，对于每个满足 $\|\boldsymbol{\beta}\|_s = 1$的 $\boldsymbol{\beta} \in \mathbb{R}^p$，都存在一个向量 $\boldsymbol{\alpha} \in \mathbb{R}^p$，使得 $\|\boldsymbol{\alpha}\|_t = 1$ 且 $\boldsymbol{\beta}^\top \boldsymbol{\alpha} = 1$。值得注意的是， ℓ_2-范数是自对偶的，而 ℓ_1-范数和 ℓ_∞-范数互为对偶范数。

定义 1.27（原始对应集合）。设 $\|\cdot\|$ 是 $\mathbb{R}^p$上的任意范数。那么 的原始对应集合定义为

$$\Lambda_{\boldsymbol{a}} \triangleq \arg\max_{\boldsymbol{u}\in\mathbb{R}^p} \{\langle \boldsymbol{a}, \boldsymbol{u}\rangle \mid \|\boldsymbol{u}\| \leq 1\}. \tag{1.4}$$

也就是说，对于任意 $\boldsymbol{a}^\dagger \in \Lambda_{\boldsymbol{a}}$， $\langle \boldsymbol{a}, \boldsymbol{a}^\dagger\rangle = \|\boldsymbol{a}\|_*$，其中 $\|\cdot\|_*$ 表示 $\|\cdot\|$的对偶范数。由此可得

(i) 若 $\boldsymbol{a} \neq \boldsymbol{0}$，则对于任意 $\boldsymbol{a}^\dagger \in \Lambda_{\boldsymbol{a}}$， $\|\boldsymbol{a}^\dagger\| = 1$。
(ii) 若 $\boldsymbol{a} = \boldsymbol{0}$，则 $\Lambda_{\boldsymbol{a}} = \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}\| \leq 1\}$。

例 1.28（原始对应集合）。以下给出几个原始对应集合的例子：

- 若范数为 ℓ_2-范数，则对于任意 $\boldsymbol{a} \neq \boldsymbol{0}$， $\Lambda_{\boldsymbol{a}} = \{\boldsymbol{a}/\|\boldsymbol{a}\|_2\}$。若范数
- 为 ℓ_1-范数，则对于任意 $\boldsymbol{a} \neq \boldsymbol{0}$，

$$\Lambda_{\boldsymbol{a}} = \left\{ \sum_{i\in\mathbb{I}(\boldsymbol{a})} \lambda_i \operatorname{sign}(a_i)\boldsymbol{e}_i \mid \sum_{i\in\mathbb{I}(\boldsymbol{a})} \lambda_i = 1, \lambda_j \geq 0, j \in \mathbb{I}(\boldsymbol{a}) \right\},$$

  其中 $\mathbb{I}(\boldsymbol{a}) \triangleq \arg\max_{i=1,\ 2,\ \ldots,\ p} |a_i|$。
- 如果范数是 ℓ_∞-范数，那么对于任意 $\boldsymbol{a} \neq \boldsymbol{0}$，

$$\Lambda_{\boldsymbol{a}} = \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \beta_i = \operatorname{sign}(a_i), i \in \mathbb{I}_{\neq}(\boldsymbol{a}), |\beta_j| \leq 1, j \in \mathbb{I}_0(\boldsymbol{a})\},$$

  where $\mathbb{I}_{\neq}(\boldsymbol{a}) \triangleq \{i \in \{1, 2, \ldots, p\} \mid a_i \neq 0\}$ and $\mathbb{I}_0(\boldsymbol{a}) \triangleq \{i \in \{1, 2, \ldots, p\} \mid a_i = 0\}$.

这些例子在非欧几里得梯度下降方法的发展中发挥着至关重要的作用Lu(2025年)。□

球

给定一个特定的范数，我们引入开球和闭球的概念如下：

定义 1.29（开球、闭球）。令 $\|\cdot\|_s : \mathbb{R}^p \to \mathbb{R}_+$ 表示 ℓ_s-范数。以 $\boldsymbol{c} \in \mathbb{R}^p$ 为中心、半径为 $r > 0$ 的开球定义为

$$\mathbb{B}_s(\boldsymbol{c}, r) \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta} - \boldsymbol{c}\|_s < r\}.$$

类似地，以 $\boldsymbol{c} \in \mathbb{R}^p$ 为中心、半径为 $r > 0$ 的闭球 定义为

$$\mathbb{B}_s[\boldsymbol{c}, r] \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta} - \boldsymbol{c}\|_s \leq r\}.$$

例如， $\mathbb{B}_2[\boldsymbol{0}, 1]$ 表示关于 ℓ_2-范数的单位闭球。为简化符号说明，对于 ℓ_2-范数，我们省略下标“2”；对于以原点为中心的球，则省略中心$\boldsymbol{0}$，例如，

$$\mathbb{B}[1] \triangleq \mathbb{B}_2[\boldsymbol{0}, 1]. \tag{1.5}$$

作为一个特例，记号

$$\mathbb{B}_0[k] \triangleq \mathbb{B}_0[\mathbf{0}, k] \tag{1.6}$$

用于表示 k-稀疏向量的集合，即包含 $\mathbb{R}^p$ 中至多有k 个非零项的向量。更一般地，设 $\|\cdot\|$ 为任意范数，相应的开球与闭球记为

$$\mathbb{B}_{\|\cdot\|}(\boldsymbol{c}, r) \qquad \text{and} \qquad \mathbb{B}_{\|\cdot\|}[\boldsymbol{c}, r]. \tag{1.7}$$

1.2.2 矩阵范数：弗罗贝尼乌斯范数与谱范数

向量范数衡量向量的“大小”或“长度”，而矩阵范数将这一概念扩展到矩阵——但关键区别在于，矩阵范数还必须考虑矩阵作为线性算子的作用，通常通过缩放或拉伸等性质来刻画其对向量的影响。对于矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，我们定义（矩阵）弗罗贝尼乌斯范数如下。

定义 1.30（矩阵弗罗贝尼乌斯范数）。矩阵弗罗贝尼乌斯范数定义为 $\boldsymbol{X} \in\mathbb{R}^{n\times p}$：

$$\|\boldsymbol{X}\|_F = \sqrt{\sum_{i=1,j=1}^{n,p} x_{ij}^2} = \sqrt{\operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^\top)} = \sqrt{\operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{X})} = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2},$$

其中，σ_1、σ_2、.....、σ_r是 $\boldsymbol{X}$的非零奇异值，而 $\operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{X})$ 表示 $\boldsymbol{X}^\top\boldsymbol{X}$的迹，即 $\boldsymbol{X}^\top\boldsymbol{X}$对角元素之和。

谱范数定义如下。

定义 1.31（矩阵谱范数）。矩阵的谱范数 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$定义为

$$\|\boldsymbol{X}\|_2 = \max_{\boldsymbol{\beta}\neq\mathbf{0}} \frac{\|\boldsymbol{X}\boldsymbol{\beta}\|_2}{\|\boldsymbol{\beta}\|_2} = \max_{\boldsymbol{u}\in\mathbb{R}^p:\|\boldsymbol{u}\|_2=1} \|\boldsymbol{X}\boldsymbol{u}\|_2 . \tag{1.8}$$

[a] 该量等于 $\boldsymbol{X}$的最大奇异值，即$\|\boldsymbol{X}\|_2 = \sigma_{\max}(\boldsymbol{X})$；参见第 1.2.4节。

[a]。请注意，本书不会使用上确界/下确界记号，而只使用最小值/最大值记号；使用这种记号并不意味着最大值或最小值一定能够取到。

如果 $\boldsymbol{X}$ 是**对称**矩阵，那么它的谱范数也等于最大绝对特征值 e:

$$\|\boldsymbol{X}\|_2 = \max_{\boldsymbol{u}\in\mathbb{R}^p:\|\boldsymbol{u}\|_2=1} \left|\boldsymbol{u}^\top\boldsymbol{X}\boldsymbol{u}\right|, \quad \text{if } \boldsymbol{X} = \boldsymbol{X}^\top. \tag{1.9}$$

注 1.32（矩阵范数的标准上界）。与向量范数的标准界类似（习题 1.26），谱范数有弗罗贝尼乌斯范数作为上界

$$\|\boldsymbol{X}\|_2 \leq \|\boldsymbol{X}\|_F . \tag{1.10}$$

这可直接由定义 $\|\boldsymbol{X}\|_2 = \sigma_{\max}(\boldsymbol{X})$ 和 $\|\boldsymbol{X}\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}$ 得出。或者，也可以利用

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \leq \sum_{i=1}^{n} \left(\sum_{j=1}^{p} x_{ij}^2 \right) \left(\sum_{j=1}^{p} \beta_j^2 \right) = \|\boldsymbol{X}\|_F \, \|\boldsymbol{\beta}\|_2 \,,$$

where 该不等式由柯西–施瓦茨不等式得出；参见定理 1.59 .

我们注意到，弗罗贝尼乌斯范数相当于向量 ℓ_2-范数在矩阵上的对应形式。为简便起见，当上下文已能明确所指范数时，我们省略向量 ℓ_2-范数或弗罗贝尼乌斯范数的完整下标：

$$\|\boldsymbol{X}\| = \|\boldsymbol{X}\|_F \qquad \text{and} \qquad \|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 \,.$$

然而，对于谱范数，$\|\boldsymbol{X}\|_2$ 下标**不能** 省略。

赋予给定范数 $\|\cdot\|$的向量空间 $\mathbb{R}^p$称为赋范向量空间。另一方面，定义矩阵范数的另一种方式是将矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$视为 $\mathbb{R}^{np}$中的向量，例如通过向量化。矩阵范数与矩阵上的一般向量范数的区别在于一个称为次乘性的关键性质：$\|\boldsymbol{X}\boldsymbol{Y}\| \leq \|\boldsymbol{X}\| \, \|\boldsymbol{Y}\|$ 若$\|\cdot\|$ 是次乘性矩阵范数（见下文讨论）。

在一些文献中，不具次乘性 的矩阵范数称为矩阵上的向量范数 或广义矩阵范数。虽然矩阵范数的定义适用于方阵和矩形矩阵，但次乘性在方阵分析中尤为重要。对于满足 $\|\boldsymbol{X}\boldsymbol{Y}\| \leq \|\boldsymbol{X}\| \, \|\boldsymbol{Y}\|$的次乘性矩阵范数 $\|\cdot\|$ ，以及方阵 $\boldsymbol{X} \in \mathbb{R}^{n\times n}$，由归纳法可得

$$\|\boldsymbol{X}^2\| \leq \|\boldsymbol{X}\|^2 \quad \Longrightarrow \quad \left\|\boldsymbol{X}^k\right\| \leq \|\boldsymbol{X}\|^k \,, \; \forall\, k \in \{1, 2, \ldots, \}. \tag{1.11}$$

因此，若 $\boldsymbol{X}$ 是幂等的（即 $\boldsymbol{X}^2 = \boldsymbol{X}$），则 $\|\boldsymbol{X}\| \geq 1$。特别地，这意味着

$$\|\boldsymbol{I}\| \geq 1, \quad \text{if} \quad \|\cdot\| \text{ is submultiplicative.} \tag{1.12}$$

O另一方面，若 $\boldsymbol{X}$ 是非奇异的，则利用次乘性可得 :

$$1 \leq \|\boldsymbol{I}\| = \left\|\boldsymbol{X}\boldsymbol{X}^{-1}\right\| \leq \|\boldsymbol{X}\| \left\|\boldsymbol{X}^{-1}\right\| .$$

也就是说，次乘性范数满足 $\|\boldsymbol{I}\| \geq 1$，且当且仅当 $\|\boldsymbol{I}\| = 1$时，称其为规范化的 。

定理 1.33：（次乘性与正交不变性）Frobenius 范数和谱范数是次乘性的。也就是说，

$$\|\boldsymbol{X}\boldsymbol{Y}\|_F \leq \|\boldsymbol{X}\|_F \, \|\boldsymbol{Y}\|_F \qquad \text{and} \qquad \|\boldsymbol{X}\boldsymbol{Y}\|_2 \leq \|\boldsymbol{X}\|_2 \, \|\boldsymbol{Y}\|_2 \,.$$

此外，这两种范数都具有正交不变性。也就是说，对于任意正交矩阵$\boldsymbol{U} \in \mathbb{R}^{n\times n}$ 和 $\boldsymbol{V} \in \mathbb{R}^{p\times p}$，以及任意 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，

$$\|\boldsymbol{X}\|_F = \|\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\|_F \qquad \text{and} \qquad \|\boldsymbol{X}\|_2 = \|\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}\|_2 \,.$$

证明 参见Lu (2021a)。 ■

诱导矩阵范数：推广谱范数

更多推广通常，许多矩阵范数都可以通过诱导范数的概念推导得出。

定义1.34（诱导矩阵范数：一般矩阵范数）。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为一个矩阵，且 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 分别为定义在 $\mathbb{R}^p$ 和 $\mathbb{R}^n$ 上的向量范数。诱导矩阵范数$\|\boldsymbol{X}\|_{a,\ b}$ 定义为

$$\|\boldsymbol{X}\|_{a,b} = \max_{\|\boldsymbol{\beta}\|_a \neq 0} \frac{\|\boldsymbol{X}\boldsymbol{\beta}\|_b}{\|\boldsymbol{\beta}\|_a} = \max_{\|\boldsymbol{\beta}\|_a = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_b \,.\ ^{a}$$

由此定义可立即得到矩阵向量乘积不等式：

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_b \leq \|\boldsymbol{X}\|_{a,b} \cdot \|\boldsymbol{\beta}\|_a \,. \tag{1.13}$$

诱导矩阵范数也可称为（a，b）-范数。当 $a = b$ 时，它简称为 a-范数，我们用 $\|\boldsymbol{X}\|_a$ 代替$\|\boldsymbol{X}\|_{a,\ a}$。特别地，当 $a = b = 2$ 时，我们得到经典的谱范数。

[a] 虽然范数 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 有多种选择，但在此语境下最常讨论的是 ℓ_s-范数。在这种情况下，诱导范数有时也称为 *Hölder's induced matrix norm*。

直观地说，诱导范数衡量矩阵 $\boldsymbol{X}$ 拉伸非零向量的最大倍数：输入向量 $\boldsymbol{\beta}$ 使用范数 $\|\cdot\|_a$ 衡量，而输出向量 $\boldsymbol{X}\boldsymbol{\beta}$ 使用范数 $\|\cdot\|_b$ 衡量。

根据上述定义可知，谱范数是诱导范数的一种特殊情况，其中 $a = b = 2$，即 $\|\boldsymbol{X}\|_2 = \|\boldsymbol{X}\|_{2,\ 2}$。类似地，矩阵*1*-范数可表示为

$$\textbf{Matrix 1-norm:} \quad \|\boldsymbol{X}\|_1 = \max_{j=1,2,\ldots,p} \sum_{i=1}^{n} |x_{ij}| = \max_{j=1,2,\ldots,p} \|\boldsymbol{x}_j\|_1 \,, \tag{1.14}$$

这也称为最大绝对列和范数。为验证这一点，考虑 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。对于任意向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，我们有

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_1 = \left\| \sum_{i=1}^{p} \beta_i \boldsymbol{x}_i \right\|_1 \leq \sum_{i=1}^{p} |\beta_i| \, \|\boldsymbol{x}_i\|_1 \leq \sum_{i=1}^{p} |\beta_i| \left(\max_{j=1,2,\ldots,p} \|\boldsymbol{x}_j\|_1 \right) = \|\boldsymbol{\beta}\|_1 \, \|\boldsymbol{X}\|_1 \,.$$

因此，$\max_{\|\boldsymbol{\beta}\|_1 = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_1 \leq \|\boldsymbol{X}\|_1$。另一方面，考虑 $\max_{\|\boldsymbol{\beta}\|_1 = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_1$，对于$\boldsymbol{\beta} = \boldsymbol{e}_j$，$\forall\, j \in \{1, 2, \ldots, p\}$，我们有 $\max_{\|\boldsymbol{\beta}\|_1 = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_1 \geq \max_{j=1,\ 2,\ \ldots\ldots,\ p} \|\boldsymbol{x}_j\|_1 = \|\boldsymbol{X}\|_1$。通过“夹逼”，我们有 $\max_{\|\boldsymbol{\beta}\|_1 = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_1 = \|\boldsymbol{X}\|_1$。同样，矩阵 ∞范数 定义为

$$\textbf{Matrix } \infty\textbf{-norm:} \quad \|\boldsymbol{X}\|_\infty = \max_{i=1,2,\ldots,n} \sum_{j=1}^{p} |x_{ij}| \,, \tag{1.15}$$

这也称为最大绝对行和范数。

练习 1.35（最大绝对行和范数）。证明 $\|\boldsymbol{X}\|_\infty$ 是由向量 ℓ_∞范数诱导的矩阵范数。

练习 1.36（等诱导范数与归一化范数）。设 $a=b$ 且令 $\boldsymbol{X}\in\mathbb{R}^{p\times p}$为定义 1.34中的一个方阵。在这种情况下，诱导范数称为等诱导范数。证明等诱导范数是归一化的：$\|\boldsymbol{I}_p\|_a=1$ 且 $\|\boldsymbol{X}\|_a\geq 1$。

我们已经看到，谱范数——诱导范数的一个重要例子——是次可乘的。事实上，所有诱导范数都是次可乘的。

定理 1.37：（诱导范数的次可乘性）对于 a，b，$c\geq 1$ 以及矩阵 $\boldsymbol{X}$，$\boldsymbol{Y}$（前提是矩阵乘积 $\boldsymbol{XY}$ 有定义），我们有

$$\|\boldsymbol{XY}\|_{a,b}\leq\|\boldsymbol{X}\|_{c,b}\|\boldsymbol{Y}\|_{a,c}.$$

也就是说，乘积的范数不超过各范数之积。

证明 [定理 1.37] 根据诱导范数的定义，我们有

$$\|\boldsymbol{XY}\|_{a,b}=\max_{\|\boldsymbol{\beta}\|_a=1}\|\boldsymbol{XY\beta}\|_b\leq\max_{\|\boldsymbol{\beta}\|_a=1}\|\boldsymbol{X}\|_{c,b}\|\boldsymbol{Y\beta}\|_c\leq\max_{\|\boldsymbol{\beta}\|_a=1}\|\boldsymbol{X}\|_{c,b}\|\boldsymbol{Y}\|_{a,c}\|\boldsymbol{\beta}\|_a.$$

证明完毕。 ■

注 1.38（条件数）。对于非奇异矩阵 $\boldsymbol{X}\in\mathbb{R}^{p\times p}$，我们注意到$1=\|\boldsymbol{I}\|_{b,\,b}\leq\|\boldsymbol{X}\|_{a,\,b}\left\|\boldsymbol{X}^{-1}\right\|_{b,\,a}$。数值$\kappa_{a,\,b}=\|\boldsymbol{X}\|_{a,\,b}\left\|\boldsymbol{X}^{-1}\right\|_{b,\,a}$通常称为相对于诱导向量范数 $\|\cdot\|_a$ 和 $\|\cdot\|_b$的条件数。常见的情形是

$$\kappa(\boldsymbol{X})=\|\boldsymbol{X}\|_2\left\|\boldsymbol{X}^{-1}\right\|_2=\frac{\sigma_{\max}(\boldsymbol{X})}{\sigma_{\min}(\boldsymbol{X})},$$

其中 $\sigma_{\max}$ 和 $\sigma_{\min}$ 分别表示 $\boldsymbol{X}$的最大和最小奇异值（节1.2.4）。更多详情请参见 Lu (2021a)。

1.2.3 正交矩阵与投影矩阵

我们常用字母 $\boldsymbol{Q}$，$\boldsymbol{U}$，$\boldsymbol{V}\in\mathbb{R}^{n\times p}$表示列为正交归一向量的矩阵（当 $p\neq n$时称为半正交矩阵）。如果这样的矩阵是方阵（即 $p=n$），则称为正交矩阵。正交矩阵表示保持任意向量长度（大小）的线性变换——它们可以改变向量的方向，但不会改变其范数。

定义 1.39（正交矩阵、半正交矩阵）。实方阵 $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ 是正交矩阵，当且仅当它的逆矩阵等于它的转置；即$\boldsymbol{Q}^{-1}=\boldsymbol{Q}^\top$ 且 $\boldsymbol{QQ}^\top=\boldsymbol{Q}^\top\boldsymbol{Q}=\boldsymbol{I}$。换句话说，设 $\boldsymbol{Q}=[\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n]$，其中$\boldsymbol{q}_i\in\mathbb{R}^n$对所有 $i\in\{1,2,\ldots,n\}$均成立，则 $\boldsymbol{q}_i^\top\boldsymbol{q}_j=\delta(i,j)$，其中 $\delta(i,j)$为克罗内克δ函数。对于任意向量 $\boldsymbol{x}$，正交矩阵保持欧几里得长度不变：$\|\boldsymbol{Qx}\|_2=\|\boldsymbol{x}\|_2$，其中 $\|\cdot\|_2$ 表示 ℓ_2-范数（定义 1.25）。注意，正交矩阵的各列都是单位向量，且彼此正交，因此构成一个正交归一集。不过，按照历史惯例，不使用“正交归一矩阵”这一术语；这类矩阵始终称为正交矩阵，尽管它们的列是正交归一的。

另一方面，如果 $\boldsymbol{Q} \in \mathbb{R}^{n\times\gamma}$ 只包含这些列中的 $\gamma < n$ 列，那么 $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_\gamma$ 仍然成立，其中 $\boldsymbol{I}_\gamma$ 是 $\gamma\times\gamma$ 单位矩阵。但 $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}$ 不成立。在这种情况下， $\boldsymbol{Q}$ 称为半正交矩阵。

向量 $\boldsymbol{q}_1$、$\boldsymbol{q}_2$、……、$\boldsymbol{q}_\gamma \in \mathbb{R}^n$ 若其点积 $\boldsymbol{q}_i^\top\boldsymbol{q}_j$满足 $\boldsymbol{q}_i^\top\boldsymbol{q}_j = 0$ ，则称它们两两正交。此外，如果每个向量的长度都为单位长度（即，对于所有 i，$\|\boldsymbol{q}_i\|_2 = 1$ ），则称这些向量两两正交归一。通常，我们将这类正交归一向量作为矩阵 $\boldsymbol{Q}$ 的列。

- 当 $n \neq \gamma$ 时：矩阵 $\boldsymbol{Q}$ 易于处理，因为 $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I} \in \mathbb{R}^{\gamma\times\gamma}$。当 $n = \gamma$ 时：矩阵 $\boldsymbol{Q}$ 是
- 方阵， $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$ 意味着 $\boldsymbol{Q}^\top = \boldsymbol{Q}^{-1}$，即 $\boldsymbol{Q}$ 的转置是 $\boldsymbol{Q}$的逆矩阵。于是我们还有 $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}$，也就是说， $\boldsymbol{Q}^\top$ 是 $\boldsymbol{Q}$的双侧逆矩阵。这样的矩阵 $\boldsymbol{Q}$ 称为正交矩阵。

根据定义，任意正交矩阵的列 $\boldsymbol{q}_1$、$\boldsymbol{q}_2$、……、$\boldsymbol{q}_n$都具有单位范数，且彼此正交，因此构成一个标准正交基。将 $\boldsymbol{Q}^\top$ 作用于向量 $\boldsymbol{x} \in \mathbb{R}^n$ ，等价于计算该向量在由 $\boldsymbol{Q}$ 的各列构成的基下的表示系数。将 $\boldsymbol{Q}$作用于 $\boldsymbol{Q}^\top\boldsymbol{x}$ ，则可通过用对应的系数缩放各基向量来恢复 $\boldsymbol{x}$ ：

$$\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{x} = \sum_{i=1}^{n}\langle\boldsymbol{q}_i, \boldsymbol{x}\rangle\boldsymbol{q}_i.$$

由于正交矩阵表示纯旋转（也可能包含反射），两个正交矩阵的乘积也为正交矩阵，这一点很直观。事实上，若 $\boldsymbol{Q}$、$\boldsymbol{U} \in \mathbb{R}^{n\times n}$是正交矩阵，那么它们的乘积 $\boldsymbol{Q}\boldsymbol{U}$ 也是正交矩阵。

具体而言，以下引理表明，正交矩阵保持任意向量的 ℓ_2-范数不变。

引理 1.40 设 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ 为正交矩阵。那么，对于任意向量 $\boldsymbol{x} \in \mathbb{R}^n$，

$$\|\boldsymbol{Q}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2 .$$

证明 [引理 1.40] $\boldsymbol{x}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x} = \|\boldsymbol{x}\|_2^2$。根据正交矩阵的定义，我们有。对两边取平方根即可得到所需结果。 $\|\boldsymbol{Q}\boldsymbol{x}\|_2^2 =$ ■

半正交矩阵与正交投影矩阵的概念密切相关。形式化地，我们定义投影矩阵如下：

定义 1.41（投影矩阵）。如果且仅如果矩阵 $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ 满足以下性质，则称其为到子空间 $\mathcal{V} \in \mathbb{R}^n$ 上的投影矩阵或投影算子：

- （P1）。 $\boldsymbol{H}\boldsymbol{y} \in \mathcal{V}$ 对所有 $\boldsymbol{y} \in \mathbb{R}^n$成立：任意向量都可以投影到子空间 $\mathcal{V}$上。（
- P2）。 $\boldsymbol{H}\boldsymbol{y} = \boldsymbol{y}$ 对所有 $\boldsymbol{y} \in \mathcal{V}$成立：已经位于 $\mathcal{V}$ 中的向量在投影下保持不变。
- （P3）。 $\boldsymbol{H}^2 = \boldsymbol{H}$：应用两次投影与应用一次具有相同的效果。这一性质称为幂等性。

由于我们将向量从 $\mathbb{R}^n$ 投影到 $\mathbb{R}^n$的一个子空间上，因此任何投影矩阵都必须是方阵。如果 $\boldsymbol{H}$ 是 $k\times n$ 且满足 $k \neq n$，那么其像将位于 $\mathbb{R}^k$中，而不是 $\mathbb{R}^n$的一个子空间中。

注意，$\boldsymbol{H}\boldsymbol{y}$ 始终位于 $\boldsymbol{H}$的列空间中，记作 $\mathcal{C}(\boldsymbol{H})$。接下来可以问：目标子空间 $\mathcal{V}$ 与列空间 $\mathcal{C}(\boldsymbol{H})$之间是什么关系？事实上，它们是相同的（提示见引理 1.44）。

假设 $\mathcal{V}=\mathcal{C}(\boldsymbol{H})$，并令 $\boldsymbol{y}\in\mathcal{V}$。那么存在某个向量 $\boldsymbol{\alpha}$ 使得 $\boldsymbol{y}=\boldsymbol{H}\boldsymbol{\alpha}$。仅利用条件（P3），我们得到

$$\boldsymbol{H}\boldsymbol{y}=\boldsymbol{H}\boldsymbol{H}\boldsymbol{\alpha}=\boldsymbol{H}\boldsymbol{\alpha}=\boldsymbol{y}.$$

因此，条件（P3）蕴含条件（P1）和（P2）。于是，矩阵 $\boldsymbol{H}$ 是投影矩阵，当且仅当它是幂等的（$\boldsymbol{H}^2=\boldsymbol{H}$）。

直观地说，我们通常要求的不只是幂等性：我们希望投影向量$\hat{\boldsymbol{y}}=\boldsymbol{H}\boldsymbol{y}$ 是 $\mathcal{V}$ 中距离原向量 $\boldsymbol{y}$最近的点，这里的距离指欧几里得距离。当且仅当残差 $\boldsymbol{y}-\hat{\boldsymbol{y}}$ 与 $\hat{\boldsymbol{y}}$ 正交（更一般地说，与整个子空间 $\mathcal{V}$正交）时，这一条件才成立。这种投影称为正交投影，符合最小二乘误差最小化原则。

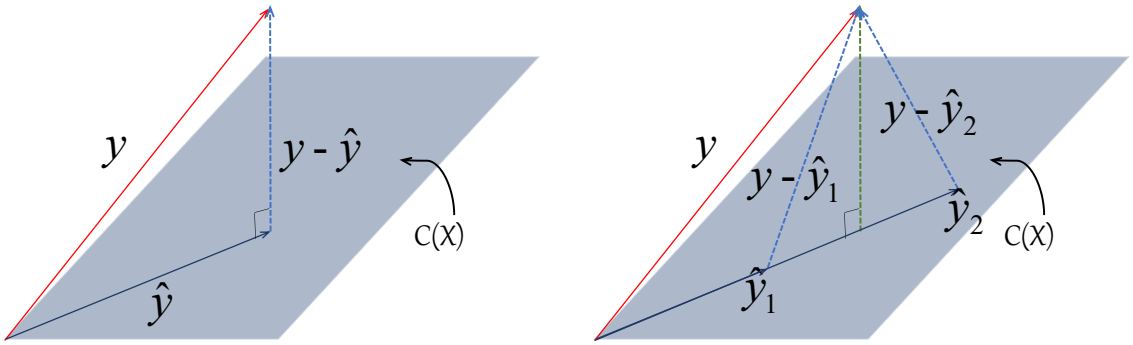
定义 1.42（正交投影矩阵与斜投影矩阵）。矩阵 $\boldsymbol{H}$ 称为子空间 $\mathcal{V}\subseteq\mathbb{R}^n$上的正交投影矩阵或正交投影算子，当且仅当 $\boldsymbol{H}$ 是投影矩阵，且任意向量 $\boldsymbol{y}\in\mathbb{R}^n$ 的投影 $\hat{\boldsymbol{y}}$ 都与 $\boldsymbol{y}-\hat{\boldsymbol{y}}$正交；也就是说，$\boldsymbol{H}$ 投影到 $\mathcal{V}$ 上，并沿着 $\mathcal{V}$的正交补 $\mathcal{V}^\perp$投影。

否则，如果 $\hat{\boldsymbol{y}}$ 与 $\boldsymbol{y}-\hat{\boldsymbol{y}}$不正交，那么该投影矩阵称为斜投影矩阵或斜投影算子。正交投影与斜投影的几何对比见图1.1。

习题 1.43. 证明：投影矩阵 $\boldsymbol{H}$ 是正交投影矩阵，当且仅当它是对称矩阵。

直观地说，设 $\mathcal{V}$ 是 $\mathbb{R}^n$的一个子空间，且设 $\boldsymbol{H}\in\mathbb{R}^{n\times n}$ 是到 $\mathcal{V}$上的正交投影矩阵。如图1.1所示，对于任意向量 $\boldsymbol{y}\in\mathbb{R}^n$，正交投影使其到 $\mathcal{V}$的欧几里得距离最小：

$$\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{y}\|_2^2\le\|\boldsymbol{y}-\boldsymbol{v}\|_2^2,\qquad\forall\boldsymbol{v}\in\mathcal{V}. \tag{1.16}$$



(a) 正交投影：将 $\boldsymbol{y}$ 投影到 $\hat{\boldsymbol{y}}$。　(b) 斜投影：将 $\boldsymbol{y}$ 投影到 $\hat{\boldsymbol{y}}_1$ 或 $\hat{\boldsymbol{y}}_2$。

F图 1.1： 投影到超平面 $\mathcal{C}(\boldsymbol{X})$，即 $\boldsymbol{X}$的列空间。

需要注意的是，术语“正交投影”并不意味着投影矩阵 $\boldsymbol{H}$ 本身是正交矩阵（按照定义1.39的含义，其中

$\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}$）。相反，“正交”指的是被投影向量 $\hat{\boldsymbol{y}}$ 与残差向量 $\boldsymbol{y}-\hat{\boldsymbol{y}}$ 之间的几何正交性。除非另有说明，本书后文中的“投影矩阵”均指正交投影矩阵。

以下结果表明，对由一组线性无关向量张成空间进行正交投影时，该投影具有特定的闭式表达式，并且与普通最小二乘（OLS）解密切相关（见第3.2节）。

引理 1.44：（由一组向量得到的投影矩阵）设 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，……，$\boldsymbol{x}_p \in \mathbb{R}^n$ 是满足 $\mathcal{C}([\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]) = \mathcal{V}$ 的线性无关向量，并假设 $n \geq p$。则到子空间 $\mathcal{V}$ 的正交投影为

$$\boldsymbol{H} \triangleq \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top,$$

其中，$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是其列为 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、……、$\boldsymbol{x}_p$ 的矩阵。

引理 [1.44] 的证明：容易验证，$\boldsymbol{H}$ 是对称且幂等的，这表明它是一个正交投影矩阵。考虑 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ 的奇异值分解（见第 1.2.4 节），可得 $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{U}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top$。令 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 为 $\boldsymbol{U}$ 的列分块。根据定理1.50，$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\}$ 是 $\mathcal{C}(\boldsymbol{X})$ 的标准正交基。而 $\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^\top$ 在 $\boldsymbol{H}$ 中是一个 $n \times n$ 矩阵，其中左上部分是 $p \times p$ 单位矩阵，其余部分为零。因此，$\boldsymbol{H} = \sum_{i=1}^{p} \boldsymbol{u}_i\boldsymbol{u}_i^\top$，这意味着 $\mathcal{C}(\boldsymbol{H})$ = 张成 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\} = \mathcal{C}(\boldsymbol{X})$。所以，$\boldsymbol{H}$ 是到 $\mathcal{C}(\boldsymbol{X}) = \mathcal{V}$ 的正交投影算子，如所要求。

■

引理 1.44 表明 $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ 是投影到 $\mathcal{C}(\boldsymbol{X})$ 的正交投影算子，前提是 $\boldsymbol{X}$ 具有满列秩。更一般地，具有相互正交归一列的矩阵（半正交矩阵）也可用于构造正交投影矩阵。

引理 1.45：（一般子空间上的正交投影算子）设 $\mathcal{V}$ 是 $\mathbb{R}^n$ 的一个子空间，其维数为 r。设 $\boldsymbol{Q}_1 \in \mathbb{R}^{n\times r}$ 和 $\boldsymbol{Q}_2 \in \mathbb{R}^{n\times(n-r)}$ 是半正交矩阵（即其列相互正交归一；定义 1.39），满足 $\mathcal{C}(\boldsymbol{Q}_1) = \mathcal{V}$ 且 $\mathcal{C}(\boldsymbol{Q}_2) = \mathcal{V}^\perp$，其中 $\mathcal{V}^\perp$ 表示 $\mathcal{V}$的正交补。那么，投影到 $\mathcal{V}$ 和 $\mathcal{V}^\perp$ 上的正交投影算子分别为

$$\boldsymbol{H}_1 \triangleq \boldsymbol{Q}_1\boldsymbol{Q}_1^\top \qquad \text{and} \qquad \boldsymbol{H}_2 \triangleq \boldsymbol{Q}_2\boldsymbol{Q}_2^\top, \tag{1.17}$$

分别。

证明 [引理 1.45] 有 $\boldsymbol{H}_1^2 = \boldsymbol{Q}_1\boldsymbol{Q}_1^\top\boldsymbol{Q}_1\boldsymbol{Q}_1^\top = \boldsymbol{Q}_1\boldsymbol{Q}_1^\top = \boldsymbol{H}_1$，因为 $\boldsymbol{Q}_1^\top\boldsymbol{Q}_1 = \boldsymbol{I}_r$。这表明 $\boldsymbol{H}_1$ 是投影到 $\mathcal{V}$上的投影算子。由于 $\boldsymbol{H}_1$ 是对称矩阵，第一部分得证。对 $\boldsymbol{H}_2 = \boldsymbol{Q}_2\boldsymbol{Q}_2^\top$ 采用相同的论证，可知它是投影到 $\mathcal{V}^\perp$上的正交投影算子。

■

1.2.4 奇异值分解（SVD）

本节介绍矩阵的奇异值分解（SVD）。在讲解 SVD 的一般形式之前，我们先从对称矩阵的谱分解开始。谱定理——也称为对称矩阵的谱分解

指出，任何实对称矩阵都具有实特征值，并且可以使用实正交归一基将其对角化。[2]

定理 1.46：（谱分解）实矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ 是对称矩阵，当且仅当存在一个正交矩阵 $\boldsymbol{Q}$ 和一个对角矩阵 $\boldsymbol{\Lambda}$，使得

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top,$$

其中，$\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n]$ 的列是 $\boldsymbol{X}$的两两正交归一特征向量，而 $\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ 的元素是 $\boldsymbol{X}$对应的特征值，这些特征值均为实数。特别地，以下性质成立：

1. 对称矩阵的所有特征值都是**实数**。2. 特征向量可以选取为构成一个**正交归**一集合。3. $\boldsymbol{X}$ 的秩等于其非零特征值的个数。4. 如果所有特征值互不相同，那么对应的特征向量线性无关。

证明参见Lu (2022年)。 ■

利用谱分解，我们可以将对称矩阵分解为对角形式。然而，这种方法并非总是适用：当 $\boldsymbol{X}$ 不是对称矩阵或不是方阵时，这种对角化通常不可能实现。奇异值分解（SVD）解决了这一局限。SVD 不再依赖单个由特征向量构成的正交矩阵，而是将任意矩阵表示为两个正交矩阵与一个奇异值对角矩阵（或类似对角矩阵）的乘积。现在给出奇异值分解定理。

定理 1.47：（完全奇异值分解）每个秩为 r 的实 $n \times p$ 矩阵 $\boldsymbol{X}$ 都存在如下形式的分解

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top,$$

其中 $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times p}$ 具有块结构 $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$，且 $\boldsymbol{\Sigma}_r = \text{diag}(\sigma_1, \sigma_2 \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$，奇异值满足 $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$。

- 标量 σ_i 是 $\boldsymbol{X}$的非零奇异值。每个 σ_i 都等于 $\boldsymbol{X}^\top\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^\top$的一个非零特征值的正平方根。$\boldsymbol{U} \in \mathbb{R}^{n\times n}$ 是一个正交矩阵，其前 r 列是 $\boldsymbol{X}\boldsymbol{X}^\top$对应于其 r 个非零特征
- 值的特征向量；其余 $n - r$ 列构成 $\mathcal{N}(\boldsymbol{X}^\top)$的标准正交基。$\boldsymbol{V} \in \mathbb{R}^{p\times p}$ 是一个正交矩阵，其前 r 列是 $\boldsymbol{X}^\top\boldsymbol{X}$对应于其 r 个非零特征值的特征向量；其余 $p - r$ 列构成
- $\mathcal{N}(\boldsymbol{X})$的标准正交基。$\boldsymbol{U}$ 和 $\boldsymbol{V}$的列分别称为 $\boldsymbol{X}$的左、右奇异向量。

此外，这种分解可以表示为秩一矩阵之和：$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$。

2。注意，*Hermitian matrices* 同样保证了实特征值以及通过复正交归一基进行对角化。

如果 $\boldsymbol{X}$ 的秩为 r，那么它的奇异值满足 $\sigma_1, \sigma_2, \ldots, \sigma_r > 0$，而 $\sigma_{r+1} = \sigma_{r+2} = \ldots = 0$。在许多应用中，使用约化奇异值分解（约化 SVD）更为方便。对于秩为 r 且（完整）SVD 为 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$的 $\boldsymbol{X}$，我们取子矩阵 $\boldsymbol{U}_r \in \mathbb{R}^{n\times r}$, $\boldsymbol{V}_r \in \mathbb{R}^{p\times r}$，使得 $\boldsymbol{U} = [\boldsymbol{U}_r, \boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$, $\boldsymbol{V} = [\boldsymbol{V}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_p]$, 并且 $\boldsymbol{\Sigma}_r = \mathrm{diag}([\sigma_1, \ldots, \sigma_r]) \in \mathbb{R}^{r\times r}$。因此，我们得到 $\boldsymbol{X}$的约化奇异值分解（约化 SVD）： $\boldsymbol{X} = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^\top$。图1.2展示了约化 SVD 与完整 SVD 的对比，其中白色元素为零，蓝色元素不一定为零。

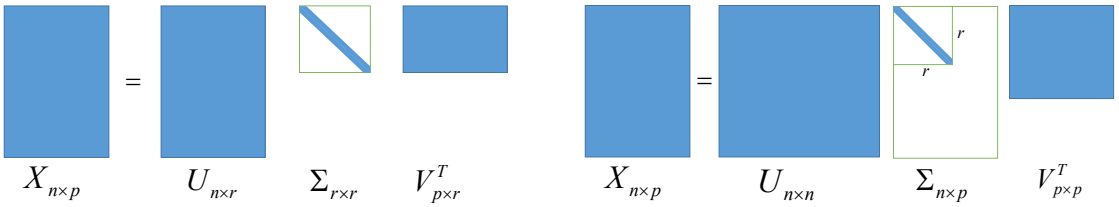
给定 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的约化奇异值分解 $\boldsymbol{X} = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^\top$，我们注意到

$$\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{V}_r\boldsymbol{\Sigma}_r^\top\boldsymbol{U}_r^\top\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^\top = \boldsymbol{V}_r\boldsymbol{\Sigma}_r^2\boldsymbol{V}_r^\top;$$
$$\boldsymbol{X}\boldsymbol{X}^\top = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^\top\boldsymbol{V}_r\boldsymbol{\Sigma}_r^\top\boldsymbol{U}_r^\top = \boldsymbol{U}_r\boldsymbol{\Sigma}_r^2\boldsymbol{U}_r^\top.$$

因此，我们分别得到对称半正定矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^\top$的（约化）谱分解。特别地，奇异值 $\sigma_i = \sigma_i(\boldsymbol{X})$ 满足

$$\sigma_i(\boldsymbol{X}) = \sqrt{\lambda_i(\boldsymbol{X}^\top\boldsymbol{X})} = \sqrt{\lambda_i(\boldsymbol{X}\boldsymbol{X}^\top)}, \quad i = 1, \ldots, \min\{n, p\}, \tag{1.18}$$

其中，$\lambda_1(\boldsymbol{X}^\top\boldsymbol{X}) \geq \lambda_2(\boldsymbol{X}^\top\boldsymbol{X}) \geq \ldots$是 $\boldsymbol{X}^\top\boldsymbol{X}$的特征值，按降序排列。此外，$\boldsymbol{U}, \boldsymbol{V}$中列出的左右奇异向量可以通过对半正定矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^\top$进行谱分解得到。因此，可以直接根据 $\boldsymbol{X}^\top\boldsymbol{X}$和 $\boldsymbol{X}\boldsymbol{X}^\top$的谱分解构造 $\boldsymbol{X}$ 的奇异值分解——这也为奇异值分解的存在性提供了另一种证明。



(a) 约化奇异值分解。　　(b) 完整奇异值分解。

图1.2： 约化 SVD 与完整 SVD 的比较。白色元素为零，蓝色元素不一定为零

练习 1.48（SVD 的证明）。 利用上述讨论和谱分解，证明矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的奇异值分解存在。

SVD 中的四组正交归一基

对于任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，以下性质成立：

- 零空间 $\mathcal{N}(\boldsymbol{X})$ 是 $\mathbb{R}^p$ 中行空间 $\mathcal{C}(\boldsymbol{X}^\top)$ 的正交补：$\dim(\mathcal{N}(\boldsymbol{X})) + \dim(\mathcal{C}(\boldsymbol{X}^\top)) = p$。
- 左零空间 $\mathcal{N}(\boldsymbol{X}^\top)$ 是 $\mathbb{R}^n$ 中列空间 $\mathcal{C}(\boldsymbol{X})$ 的正交补：$\dim(\mathcal{N}(\boldsymbol{X}^\top)) + \dim(\mathcal{C}(\boldsymbol{X})) = n$。

这一结果称为线性代数基本定理（参见引理 1.12前的讨论）。特别是，奇异值分解的构造为该定理描述的四个基本子空间都提供了标准正交基。为了证明这一点，我们首先需要下面的引理。

引理 1.49：（$\boldsymbol{X}^\top\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^\top$的子空间）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 已给定。那么，

- $\boldsymbol{X}^\top\boldsymbol{X}$ 的列空间与 $\boldsymbol{X}^\top$ 的列空间相同（即 $\boldsymbol{X}$的行空间）： $\mathcal{C}(\boldsymbol{X}^\top\boldsymbol{X}) = \mathcal{C}(\boldsymbol{X}^\top)$; 根据线性代数基本定理，这也表明 $\mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X}) = \mathcal{N}(\boldsymbol{X})$（参见引理 1.12 前的讨论）。
- $\boldsymbol{X}\boldsymbol{X}^\top$ 的列空间与 $\boldsymbol{X}$的列空间相同： $\mathcal{C}(\boldsymbol{X}\boldsymbol{X}^\top) = \mathcal{C}(\boldsymbol{X})$。因此，这也表明 $\mathcal{N}(\boldsymbol{X}\boldsymbol{X}^\top) = \mathcal{N}(\boldsymbol{X}^\top)$。

证明 [引理 1.49] 令 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$，我们有 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \implies \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$，即 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \implies \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X})$。因此，$\mathcal{N}(\boldsymbol{X}) \subseteq \mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X})$。此外，令 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X})$，我们有

$$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \implies \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = 0 \implies \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 = 0 \implies \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0},$$

即，$\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X}) \implies \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$。因此，$\mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X}) \subseteq \mathcal{N}(\boldsymbol{X})$。结合两个包含关系可得 $\mathcal{N}(\boldsymbol{X}) = \mathcal{N}(\boldsymbol{X}^\top\boldsymbol{X})$。根据线性代数基本定理（参见引理1.12前的讨论），可知

$$\mathcal{C}(\boldsymbol{X}^\top) = \mathcal{C}(\boldsymbol{X}^\top\boldsymbol{X}).$$

对 $\boldsymbol{X}^\top$ 应用相同的论证，即可证明第二个结论。 ■

定理1.50：（SVD 中的四组正交归一基）设 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ 是 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$的完整奇异值分解，其中 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 和 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p]$ 分别是 $\boldsymbol{U}$ 和 $\boldsymbol{V}$的列分块。则：

- $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$ 是行空间 $\mathcal{C}(\boldsymbol{X}^\top)$的标准正交基；$\{\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots, \boldsymbol{v}_p\}$ 是零空
- 间 $\mathcal{N}(\boldsymbol{X})$的标准正交基；$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ 是列空间 $\mathcal{C}(\boldsymbol{X})$的标准正交基；
- $\{\boldsymbol{u}_{r+1}, \boldsymbol{u}_{r+2}, \ldots, \boldsymbol{u}_n\}$ 是左零空间 $\mathcal{N}(\boldsymbol{X}^\top)$的标准正交基。

证明 [定理1。50] 根据谱分解，对于对称矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$，其列空间 $\mathcal{C}(\boldsymbol{X}^\top\boldsymbol{X})$由特征向量张成。因此，集合 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$构成 $\mathcal{C}(\boldsymbol{X}^\top\boldsymbol{X}$的标准正交基。因此，根据引理 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$，$\mathcal{C}(\boldsymbol{X}^\top$也构成1.49的标准正交基。

由于 $\boldsymbol{V}$是正交的，剩余向量 $\{\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots, \boldsymbol{v}_n\}$张成的空间是向量 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$张成空间的正交补，也就是$\mathcal{N}(\boldsymbol{X})$。因此，$\{\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots, \boldsymbol{v}_n\}$ 构成 $\mathcal{N}(\boldsymbol{X})$的标准正交基。

对 $\boldsymbol{X}\boldsymbol{X}^\top$ 应用类似的论证可知，$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ 张成 $\mathcal{C}(\boldsymbol{X})$，而 $\{\boldsymbol{u}_{r+1}, \boldsymbol{u}_{r+2}, \ldots, \boldsymbol{u}_n\}$ 张成 $\mathcal{N}(\boldsymbol{X}^\top)$。另一种方法是，根据引理 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ 可知，$\boldsymbol{X}$ 的列空间的基为1.13 [3]，因为 $\boldsymbol{u}_i = \frac{\boldsymbol{X}\boldsymbol{v}_i}{\sigma_i}$，$\forall i \in \{1, 2, \ldots, r\}$（这是 SVD 的一个关键性质）。 ■

四个子空间之间的关系如图 1.3 所示。具体而言，对于每个 $i \in \{1, 2, \ldots, r\}$，矩阵 $\boldsymbol{X}$ 按照关系式 $\sigma_i\boldsymbol{u}_i = \boldsymbol{X}\boldsymbol{v}_i$，将行空间基向量 $\boldsymbol{v}_i$ 映射为列空间基向量 $\boldsymbol{u}_i$。

3。回忆一下，对于任意矩阵 $\boldsymbol{X}$，令 $\{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_r\}$ 为 $\mathbb{R}^p$中的一个向量集，该向量集构成行空间的一个基，那么 $\{\boldsymbol{X}\boldsymbol{r}_1, \boldsymbol{X}\boldsymbol{r}_2, \ldots, \boldsymbol{X}\boldsymbol{r}_r\}$ 是 $\boldsymbol{X}$的列空间的一个基。

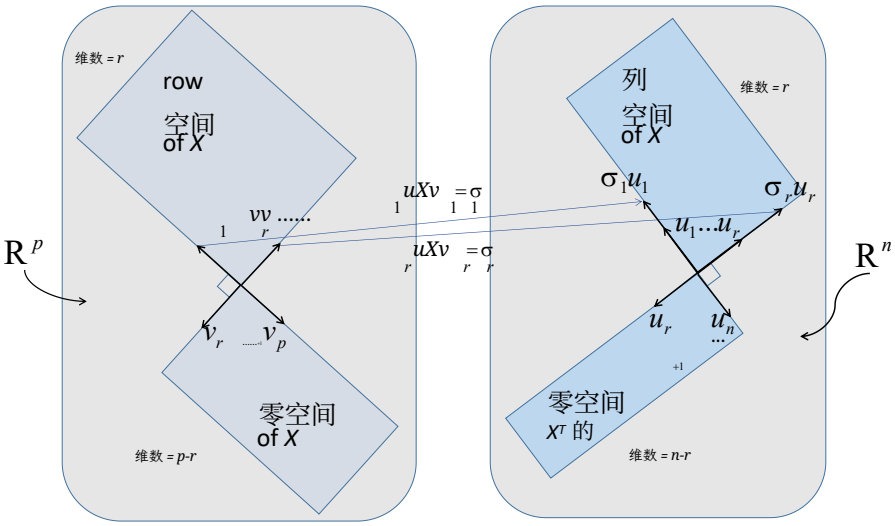


图 1.3： 通过奇异值分解对角化 $\boldsymbol{X}$ 的标准正交基。集合 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$构成行空间 $\mathcal{C}(\boldsymbol{X}^\top)$的标准正交基，而 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ 构成列空间 $\mathcal{C}(\boldsymbol{X})$的标准正交基。 $\boldsymbol{X}$ 的作用将这些基联系起来：对于每个$i \in \{1, 2, \ldots, r\}$，它将行空间基向量 $\boldsymbol{v}_i$ 变换为经过奇异值 σ_i缩放的列空间基向量 $\boldsymbol{u}_i$ ，即 $\boldsymbol{X}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$。

奇异值界

在线性映射的奇异值分解形式中，可以将其解释为对向量执行三种操作：先旋转，再按照奇异值沿正交方向进行缩放，最后再次旋转。这意味着可能的最大缩放因子等于最大的奇异值 σ_1，并且发生在对应右奇异向量 $\boldsymbol{v}_1$ 的方向上。下面的定理将这一直观认识形式化。定理表明，如果我们将注意范围限制在 $\boldsymbol{v}_1$ 的正交补上，那么由于奇异向量的正交性，最大缩放量由第二大奇异值 σ_2 决定。更一般地，与前 $i-1$ 个右奇异向量正交的最大缩放方向，对应于第 i 个奇异值 σ_i，并与第 i 个右奇异向量$\boldsymbol{v}_i$ 对齐。

定理 1.51：（奇异值的 Courant–Fischer 型刻画）设$\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 的奇异值分解为 $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$。则奇异值满足

$$\sigma_1 = \max_{\{\|\boldsymbol{\beta}\|_2 = 1 \mid \boldsymbol{\beta} \in \mathbb{R}^p\}} \|\boldsymbol{X}\boldsymbol{\beta}\|_2 , \tag{1.19}$$

$$= \max_{\{\|\boldsymbol{\alpha}\|_2 = 1 \mid \boldsymbol{\alpha} \in \mathbb{R}^n\}} \left\|\boldsymbol{X}^\top \boldsymbol{\alpha}\right\|_2 , \tag{1.20}$$

$$\sigma_i = \max_{\{\|\boldsymbol{\beta}\|_2 = 1 \mid \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\beta} \perp \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}\}} \|\boldsymbol{X}\boldsymbol{\beta}\|_2 , \tag{1.21}$$

$$= \max_{\{\|\boldsymbol{\alpha}\|_2 = 1 \mid \boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\alpha} \perp \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}\}} \left\|\boldsymbol{X}^\top \boldsymbol{\alpha}\right\|_2 , \qquad 2 \le i \le \min\{n, p\}, \tag{1.22}$$

右奇异向量满足

$$\boldsymbol{v}_1 = \underset{\{\|\boldsymbol{\beta}\|_2=1|\boldsymbol{\beta}\in\mathbb{R}^p\}}{\arg\max} \|\boldsymbol{X}\boldsymbol{\beta}\|_2, \tag{1.23}$$

$$\boldsymbol{v}_i = \underset{\{\|\boldsymbol{\beta}\|_2=1|\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\beta}\perp\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{i-1}\}}{\arg\max} \|\boldsymbol{X}\boldsymbol{\beta}\|_2, \qquad 2\le i\le p, \tag{1.24}$$

且左奇异向量满足

$$\boldsymbol{u}_1 = \underset{\{\|\boldsymbol{\alpha}\|_2=1|\boldsymbol{\alpha}\in\mathbb{R}^n\}}{\arg\max} \left\|\boldsymbol{X}^\top\boldsymbol{\alpha}\right\|_2, \tag{1.25}$$

$$\boldsymbol{u}_i = \underset{\{\|\boldsymbol{\alpha}\|_2=1|\boldsymbol{\alpha}\in\mathbb{R}^n, \boldsymbol{\alpha}\perp\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{i-1}\}}{\arg\max} \left\|\boldsymbol{X}^\top\boldsymbol{\alpha}\right\|_2, \qquad 2\le i\le n. \tag{1.26}$$

证明 [定理 1.51] 取一个单位向量 $\boldsymbol{\beta}\in\mathbb{R}^p$（即 $\|\boldsymbol{\beta}\|_2=1$），它与$\boldsymbol{v}_1$、……、$\boldsymbol{v}_{i-1}$正交，其中 $1\le i\le p$（若 $i=1$，则 $\boldsymbol{\beta}$可在单位向量中任意选取）。我们将$\boldsymbol{\beta}$按 $\boldsymbol{X}$的右奇异向量基展开：$\boldsymbol{\beta}=\sum_{j=i}^p\gamma_j\boldsymbol{v}_j$，其中 $1=\|\boldsymbol{\beta}\|_2^2=\sum_{j=i}^p\gamma_j^2$。根据定理1.47中奇异值的排列顺序，

$$\begin{aligned}\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 &= \left\langle\sum_{k=1}^p\sigma_k\langle\boldsymbol{v}_k,\boldsymbol{\beta}\rangle\boldsymbol{u}_k, \sum_{k=1}^p\sigma_k\langle\boldsymbol{v}_k,\boldsymbol{\beta}\rangle\boldsymbol{u}_k\right\rangle = \sum_{k=1}^p\sigma_k^2\langle\boldsymbol{v}_k,\boldsymbol{\beta}\rangle^2\\ &= \sum_{k=1}^p\sigma_k^2\left\langle\boldsymbol{v}_k,\sum_{j=i}^p\gamma_j\boldsymbol{v}_j\right\rangle^2 = \sum_{j=i}^p\sigma_j^2\gamma_j^2\le\sigma_i^2\sum_{j=i}^p\gamma_j^2=\sigma_i^2.\end{aligned}$$

这就证明了（1.19）和（1.21）。为了验证（1.23）和（1.24），注意到 $\boldsymbol{v}_i$ 达到了这个上界：

$$\|\boldsymbol{X}\boldsymbol{v}_i\|_2^2=\sum_{k=1}^p\sigma_k^2\langle\boldsymbol{v}_k,\boldsymbol{v}_i\rangle^2=\sigma_i^2.$$

对 $\boldsymbol{X}^\top$ 运用相同的推理可得（1.20）、（1.22）、（1.25）和（1.26）。 ■

由上述证明还可得，矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$（假设 $p\le n$）的最大和最小奇异值满足

$$\sigma_{\max}(\boldsymbol{X})=\sigma_1(\boldsymbol{X})=\|\boldsymbol{X}\|_2=\max_{\|\boldsymbol{\beta}\|_2=1}\|\boldsymbol{X}\boldsymbol{\beta}\|_2; \tag{1.27}$$

$$\sigma_{\min}(\boldsymbol{X})=\sigma_p(\boldsymbol{X})=\min_{\|\boldsymbol{\beta}\|_2=1}\|\boldsymbol{X}\boldsymbol{\beta}\|_2. \tag{1.28}$$

[4] 这表明

$$\sigma_{\min}(\boldsymbol{X})\|\boldsymbol{\beta}\|_2\le\|\boldsymbol{X}\boldsymbol{\beta}\|_2\le\sigma_{\max}(\boldsymbol{X})\|\boldsymbol{\beta}\|_2, \quad \text{for any } \boldsymbol{\beta}\in\mathbb{R}^p. \tag{1.29}$$

对于本书的目的而言，以下观察尤为有用。

定理1.52：（近似正交条件下的奇异值界）设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$满足 $n\ge p$。假设存在某个 $\delta\in[0,1]$，

$$\left\|\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I}\right\|_2\le\delta. \tag{1.30}$$

4. 参见Lu (2021a)，其中给出了详细证明，该证明依赖于对称矩阵的瑞利–里兹定理。

那么， $\boldsymbol{X}$ 的最大和最小奇异值满足

$$\sigma_{\max}(\boldsymbol{X}) \leq \sqrt{1+\delta} \qquad \text{and} \qquad \sigma_{\min}(\boldsymbol{X}) \geq \sqrt{1-\delta}. \tag{1.31}$$

反之，如果（1.31）中的两个不等式都成立，那么（1.30）便成立。

定理 [的证明1。52] 回想一下，根据（1.18）， $\boldsymbol{X}^\top \boldsymbol{X}$ 的特征值恰好是 $\boldsymbol{X}$的奇异值平方，λ_i（$\boldsymbol{X}^\top \boldsymbol{X}$）$= \sigma_i^2$（$\boldsymbol{X}$）， $i = 1$、2、.....、p。因此， $\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}$的特征值为$\sigma_i^2$（$\boldsymbol{X}$）。根据（1.27）和假设（1.30），我们有

$$\max\{\sigma_{\max}^2(\boldsymbol{X}) - 1,\, 1 - \sigma_{\min}^2(\boldsymbol{X})\} = \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I} \right\|_2 \leq \delta.$$

由此得证。 ■

极端奇异值——最大值和最小值——是关于谱范数和弗罗贝尼乌斯范数的 1-Lipschitz 函数（见定义 2.71）。换言之，两个矩阵的极端奇异值之差，不超过它们在任一范数下的距离。

定理1.53：（极端奇异值的利普希茨连续性）最小和最大奇异值 $\sigma_{\min}$， $\sigma_{\max}$ 对任意维数相同的矩阵 $\boldsymbol{X}$， $\boldsymbol{Y}$ 均满足：

$$|\sigma_{\max}(\boldsymbol{X}) - \sigma_{\max}(\boldsymbol{Y})| \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_2 \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_F; \tag{1.32}$$

$$|\sigma_{\min}(\boldsymbol{X}) - \sigma_{\min}(\boldsymbol{Y})| \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_2 \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_F. \tag{1.33}$$

证明 [定理 1.53] 根据最大奇异值与算子范数的对应关系，我们有

$$|\sigma_{\max}(\boldsymbol{X}) - \sigma_{\max}(\boldsymbol{Y})| = \big|\, \|\boldsymbol{X}\|_2 - \|\boldsymbol{Y}\|_2 \,\big| \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_2 .$$

由（1.28），对于最小奇异值，我们有

$$\begin{aligned} \sigma_{\min}(\boldsymbol{X}) &= \min_{\|\boldsymbol{\beta}\|_2 = 1} \|\boldsymbol{X}\boldsymbol{\beta}\|_2 \leq \min_{\|\boldsymbol{\beta}\|_2 = 1} (\|\boldsymbol{Y}\boldsymbol{\beta}\|_2 + \|(\boldsymbol{X} - \boldsymbol{Y})\boldsymbol{\beta}\|_2) \\ &\leq \min_{\|\boldsymbol{\beta}\|_2 = 1} (\|\boldsymbol{Y}\boldsymbol{\beta}\|_2 + \|\boldsymbol{X} - \boldsymbol{Y}\|_2) = \sigma_{\min}(\boldsymbol{Y}) + \|\boldsymbol{X} - \boldsymbol{Y}\|_2 . \end{aligned}$$

因此， $\sigma_{\min}$（$\boldsymbol{X}$） $- \sigma_{\min}$（$\boldsymbol{Y}$）$\leq \|\boldsymbol{X} - \boldsymbol{Y}\|_2$，由对称性可得（1.33）。最后，涉及弗罗贝尼乌斯范数的不等式可由标准关系推出 $\|\cdot\|_2 \leq \|\cdot\|_F$（见注 1.32）。 ■

1.2.5 伪逆

对于矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，我们可以定义其伪逆，记为 $\boldsymbol{X}^+$，它是一个$p \times n$ 矩阵。直观地说，当 $\boldsymbol{X}$ 乘以一个属于其行空间的向量 $\boldsymbol{\beta}$ （即$\boldsymbol{\beta} \in \mathcal{C}(\boldsymbol{X}^\top)$）时，结果 $\boldsymbol{X}\boldsymbol{\beta}$ 属于 $\boldsymbol{X}$ 的列空间。这两个空间的维数相同，均为 r，也就是 $\boldsymbol{X}$的秩。在这些子空间上， $\boldsymbol{X}$ 的作用类似于可逆线性变换，而 $\boldsymbol{X}^+$ 则充当其逆矩阵。更准确地说：

- 如果 $\boldsymbol{\beta}$ 属于 $\boldsymbol{X}$的行空间，那么 $\boldsymbol{X}^+\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$。如果 $\boldsymbol{y}$ 属于 $\boldsymbol{X}$的列空间，
- 那么 $\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{y} = \boldsymbol{y}$ （见图 1.4）。

$\boldsymbol{X}^+$ 的零空间与 $\boldsymbol{X}^\top$的零空间重合。它由满足 $\boldsymbol{y} \in \mathbb{R}^n$ 的向量 $\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{0}$组成。这些向量 $\boldsymbol{y}$ 与 $\boldsymbol{X}$的列空间中的每个向量正交，也就是说，与形如 $\boldsymbol{X}\boldsymbol{\beta}$的每个向量正交。定理 1.56 中将给出这一性质的证明。

更正式地说，伪逆，也称为摩尔-彭若斯伪逆，$\boldsymbol{X}^+$是由满足以下四个条件的**唯一的**（见命题1.54） $p \times n$ 矩阵定义的，这四个条件通常称为彭罗斯条件（Penrose，1955年）：

$$
\begin{array}{lllll}
(C1) & \boldsymbol{X}\boldsymbol{X}^+\boldsymbol{X} & =\boldsymbol{X} & & \\
(C2) & \boldsymbol{X}^+\boldsymbol{X}\boldsymbol{X}^+ & =\boldsymbol{X}^+ & & \\
(C3) & (\boldsymbol{X}\boldsymbol{X}^+)^\top & =\boldsymbol{X}\boldsymbol{X}^+ & & (\boldsymbol{X}\boldsymbol{X}^+ \text{ 为对称矩阵}) \\
(C4) & (\boldsymbol{X}^+\boldsymbol{X})^\top & =\boldsymbol{X}^+\boldsymbol{X} & & (\boldsymbol{X}\boldsymbol{X}^+ \text{ 为对称矩阵})
\end{array}
\tag{1.34}
$$

条件（$C1$）和（$C2$）意味着 $\boldsymbol{X}\boldsymbol{X}^+$ 和 $\boldsymbol{X}^+\boldsymbol{X}$ 是幂等矩阵，因此它们是投影矩阵（参见定义 1.41）。条件（$C3$）和（$C4$）进一步确保这些投影矩阵是对称的，这意味着它们不仅是斜投影，而且是正交投影矩阵（参见习题1.43）。

任意矩阵的伪逆的存在性和唯一性可由其秩分解推出。

命题 1.54：（伪逆的存在唯一性）每个矩阵$\boldsymbol{X}$ 都有唯一的伪逆。

命题 [的证明 1.54] **存在性。** 设 $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{Z} \in \mathbb{R}^{n\times p}$ 是$\boldsymbol{X}$的一个秩分解，其中 $\boldsymbol{W} \in \mathbb{R}^{n\times r}$ 和 $\boldsymbol{Z} \in \mathbb{R}^{r\times p}$ 都满秩。定义

$$
\boldsymbol{X}^+ = \boldsymbol{Z}^+\boldsymbol{W}^+ = \boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top,
$$

其中 $\boldsymbol{Z}^+ = \boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}$且 $\boldsymbol{W}^+ = (\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top$。[5]值得注意的是，$\boldsymbol{Z}\boldsymbol{Z}^\top$和 $\boldsymbol{W}^\top\boldsymbol{W}$可逆，因为 $\boldsymbol{W} \in \mathbb{R}^{n\times r}$ 和 $\boldsymbol{Z} \in \mathbb{R}^{r\times p}$ 满秩 r。

现在验证 $\boldsymbol{X}^+$ 满足四个彭罗斯条件：

$$
\begin{aligned}
(C1)\quad \boldsymbol{X}\boldsymbol{X}^+\boldsymbol{X} &= \boldsymbol{W}\boldsymbol{Z}\left(\boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right)\boldsymbol{W}\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{Z} = \boldsymbol{X}, \\
(C2)\quad \boldsymbol{X}^+\boldsymbol{X}\boldsymbol{X}^+ &= \left(\boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right)\boldsymbol{W}\boldsymbol{Z}\left(\boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right) \\
&= \boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top = \boldsymbol{X}^+, \\
(C3)\quad (\boldsymbol{X}\boldsymbol{X}^+)^\top &= \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{W}^\top = \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top \\
&= \boldsymbol{W}\boldsymbol{Z}\boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top = \boldsymbol{X}\boldsymbol{X}^+, \\
(C4)\quad (\boldsymbol{X}^+\boldsymbol{X})^\top &= \boldsymbol{Z}^\top\boldsymbol{W}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}\boldsymbol{Z} = \boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}\boldsymbol{Z} \\
&= \boldsymbol{Z}^\top(\boldsymbol{Z}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{Z} = \boldsymbol{X}^+\boldsymbol{X}.
\end{aligned}
$$

这表明伪逆的存在性。

5。不难验证，$\boldsymbol{Z}^+$是 $\boldsymbol{Z}$的伪逆，而 $\boldsymbol{W}^+$是 $\boldsymbol{W}$的伪逆。

唯一性。 假设 $\boldsymbol{X}_1^+$和 $\boldsymbol{X}_2^+$是满足 $\boldsymbol{X}$的四个彭罗斯条件的两个矩阵。我们将证明 $\boldsymbol{X}_1^+ = \boldsymbol{X}_2^+$。从 $\boldsymbol{X}_1^+$出发，并反复应用彭罗斯条件，可得：

$$\begin{aligned}
\boldsymbol{X}_1^+ &= \boldsymbol{X}_1^+\boldsymbol{X}\boldsymbol{X}_1^+ = \boldsymbol{X}_1^+(\boldsymbol{X}\boldsymbol{X}_2^+\boldsymbol{X})\boldsymbol{X}_1^+ = \boldsymbol{X}_1^+(\boldsymbol{X}\boldsymbol{X}_2^+)(\boldsymbol{X}\boldsymbol{X}_1^+) && \text{(by } (C2),(C1)\text{)} \\
&= \boldsymbol{X}_1^+(\boldsymbol{X}\boldsymbol{X}_2^+)^\top(\boldsymbol{X}\boldsymbol{X}_1^+)^\top = \boldsymbol{X}_1^+\boldsymbol{X}_2^{+\top}\boldsymbol{X}^\top\boldsymbol{X}_1^{+\top}\boldsymbol{X}^\top && \text{(by } (C3)\text{)} \\
&= \boldsymbol{X}_1^+\boldsymbol{X}_2^{+\top}(\boldsymbol{X}\boldsymbol{X}_1^+\boldsymbol{X})^\top = \boldsymbol{X}_1^+\boldsymbol{X}_2^{+\top}\boldsymbol{X}^\top && \text{(by } (C1)\text{)} \\
&= \boldsymbol{X}_1^+(\boldsymbol{X}\boldsymbol{X}_2^+)^\top = \boldsymbol{X}_1^+\boldsymbol{X}\boldsymbol{X}_2^+ = \boldsymbol{X}_1^+(\boldsymbol{X}\boldsymbol{X}_2^+\boldsymbol{X})\boldsymbol{X}_2^+ && \text{(by } (C3),(C1)\text{)} \\
&= (\boldsymbol{X}_1^+\boldsymbol{X})(\boldsymbol{X}_2^+\boldsymbol{X})\boldsymbol{X}_2^+ = (\boldsymbol{X}_1^+\boldsymbol{X})^\top(\boldsymbol{X}_2^+\boldsymbol{X})^\top\boldsymbol{X}_2^+ && \text{(by } (C4)\text{)} \\
&= (\boldsymbol{X}\boldsymbol{X}_1^+\boldsymbol{X})^\top\boldsymbol{X}_2^{+\top}\boldsymbol{X}_2^+ = \boldsymbol{X}^\top\boldsymbol{X}_2^{+\top}\boldsymbol{X}_2^+ && \text{(by } (C1)\text{)} \\
&= (\boldsymbol{X}_2^+\boldsymbol{X})^\top\boldsymbol{X}_2^+ = \boldsymbol{X}_2^+\boldsymbol{X}\boldsymbol{X}_2^+ = \boldsymbol{X}_2^+. && \text{(by } (C4),(C2)\text{)}
\end{aligned}$$

因此， $\boldsymbol{X}_1^+ = \boldsymbol{X}_2^+$，这证明了伪逆是唯一的。 ■

引理1.55：（伪逆的性质）对于任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，有

$$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^+ = \boldsymbol{X}^\top.$$

证明 [引理 1.55] 根据条件（$C1$） $\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{X} = \boldsymbol{X}$，对等式两边取转置可得：

$$(\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{X})^\top = \boldsymbol{X}^\top \qquad \Longrightarrow \qquad \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^+)^\top = \boldsymbol{X}^\top.$$

将条件 $(C3)$ $(\boldsymbol{X}\boldsymbol{X}^+)^\top = \boldsymbol{X}\boldsymbol{X}^+$ 代入上面的方程，我们得到$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^+ = \boldsymbol{X}^\top$. 证明完毕。 ■

等式 $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^+ = \boldsymbol{X}^\top$ 表达了这样一个思想：对 $\boldsymbol{X}$ 应用 $\boldsymbol{X}^+$ 不会改变 $\boldsymbol{X}^\top$ 对向量的作用。

我们已经在定理 1.50中给出了与矩阵的奇异值分解相关的四个基本子空间。类似地，现在我们来描述与伪逆相关的四个基本子空间。

定理 1.56：（伪逆的四个子空间）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为任意矩阵， $\boldsymbol{X}^+$ 表示其伪逆。那么以下关系式成立：

- $\boldsymbol{X}^+$ 的列空间与 $\boldsymbol{X}$的行空间相同； $\boldsymbol{X}^+$ 的行空间与 $\boldsymbol{X}$的列
- 空间相同； $\boldsymbol{X}^+$ 的零空间与 $\boldsymbol{X}^\top$的零空间相同； $\boldsymbol{X}^{+\top}$ 的零
- 空间与 $\boldsymbol{X}$的零空间相同。

四个基本子空间之间的这些关系式如图 1.4所示。

定理 [的证明 1。56] 仿照引理 1.49的证明中的论证，我们可以证明

$$\begin{aligned}
\mathcal{C}(\boldsymbol{X}\boldsymbol{X}^+) &= \mathcal{C}(\boldsymbol{X}) \qquad \text{and} \qquad \mathcal{N}(\boldsymbol{X}^+\boldsymbol{X}) = \mathcal{N}(\boldsymbol{X}) \\
\mathcal{C}(\boldsymbol{X}^+\boldsymbol{X}) &= \mathcal{C}(\boldsymbol{X}^+) \qquad \text{and} \qquad \mathcal{N}(\boldsymbol{X}\boldsymbol{X}^+) = \mathcal{N}(\boldsymbol{X}^+).
\end{aligned}$$

此外，根据伪逆定义中的条件 $(C3)$ 和 $(C4)$，我们知道：

$$(\boldsymbol{X}^+\boldsymbol{X})^\top = \boldsymbol{X}^+\boldsymbol{X} \qquad \text{and} \qquad (\boldsymbol{X}\boldsymbol{X}^+)^\top = \boldsymbol{X}\boldsymbol{X}^+.$$

利用线性代数基本定理（参见引理1.12前面的讨论），我们知道 $\mathcal{C}(\boldsymbol{X}\boldsymbol{X}^+)$ 是 $\mathcal{N}((\boldsymbol{X}\boldsymbol{X}^+)^\top)$ 的正交补，而 $\mathcal{C}(\boldsymbol{X}^+\boldsymbol{X})$ 是 $\mathcal{N}((\boldsymbol{X}^+\boldsymbol{X})^\top)$ 的正交补：

$$\begin{aligned}\mathcal{C}(\boldsymbol{X}\boldsymbol{X}^+) \perp \mathcal{N}((\boldsymbol{X}\boldsymbol{X}^+)^\top) \quad &\underrightarrow{\text{leads to}} \quad \mathcal{C}(\boldsymbol{X}\boldsymbol{X}^+) \perp \mathcal{N}(\boldsymbol{X}\boldsymbol{X}^+) \\ \mathcal{C}(\boldsymbol{X}^+\boldsymbol{X}) \perp \mathcal{N}((\boldsymbol{X}^+\boldsymbol{X})^\top) \quad &\underrightarrow{\text{leads to}} \quad \mathcal{C}(\boldsymbol{X}^+\boldsymbol{X}) \perp \mathcal{N}(\boldsymbol{X}^+\boldsymbol{X}).\end{aligned}$$

这意味着

$$\mathcal{C}(\boldsymbol{X}) \perp \mathcal{N}(\boldsymbol{X}^+) \quad \text{and} \quad \mathcal{C}(\boldsymbol{X}^+) \perp \mathcal{N}(\boldsymbol{X}).$$

也就是说，$\mathcal{N}(\boldsymbol{X}^+) = \mathcal{N}(\boldsymbol{X}^\top)$，且 $\mathcal{C}(\boldsymbol{X}^+) = \mathcal{C}(\boldsymbol{X}^\top)$。再次根据线性代数基本定理，这还意味着：$\mathcal{C}(\boldsymbol{X}^{+\top}) = \mathcal{C}(\boldsymbol{X})$，且 $\mathcal{N}(\boldsymbol{X}^{+\top}) = \mathcal{N}(\boldsymbol{X})$。 ■

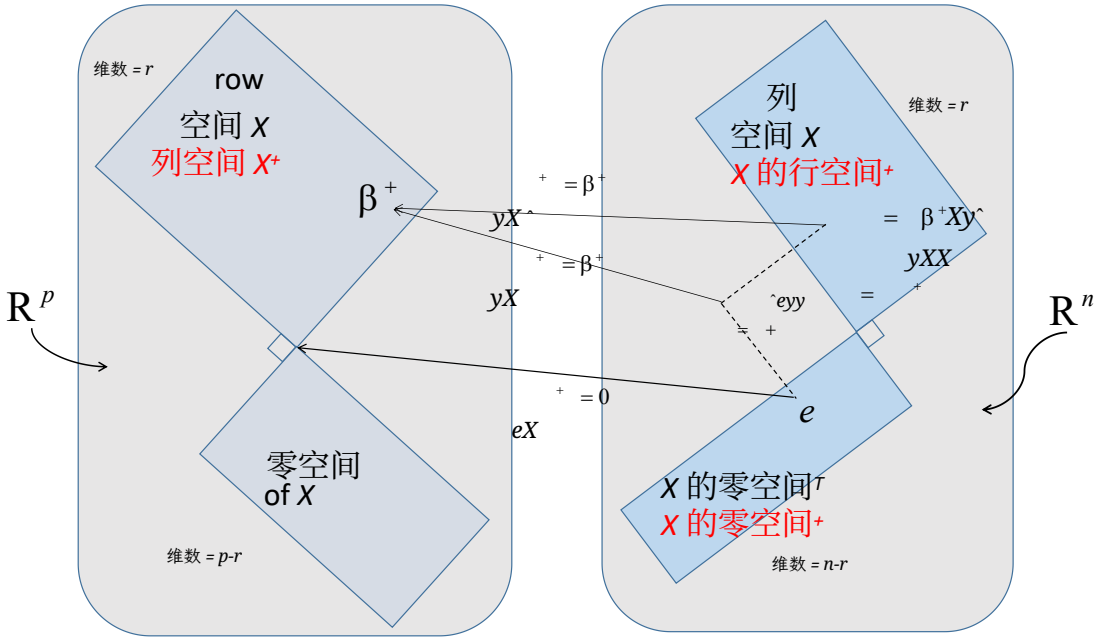


图1.4： 伪逆 $\boldsymbol{X}^+$的列空间和行空间。 $\boldsymbol{X}$ 将向量从行空间映射到列空间。 $\boldsymbol{X}^+$ 将向量从列空间映射到行空间。

奇异值分解中的伪逆

给定矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的奇异值分解，我们将进一步讨论其在不同情况下的伪逆。对于 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ 的完整奇异值分解，我们考虑以下几种情形：

情形 $n > p = r$。 由于矩阵 $\boldsymbol{X}$ 的列线性无关，因此可以通过以下方式求得左伪逆：

$$\begin{aligned}\text{left-pseudo-inverse} = \boldsymbol{X}^+ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = (\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top \\ &= \boldsymbol{V}(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})^{-1}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top = \boldsymbol{V}[(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^\top]\boldsymbol{U}^\top = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^\top,\end{aligned} \tag{1.35}$$

其中最后一个等式成立是因为 $\boldsymbol{\Sigma}^+ = (\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^\top$。

情况 $p > n = r$。由于矩阵 $\boldsymbol{X}$ 具有独立行，右伪逆可通过以下方式获得

$$\begin{aligned}\text{right-pseudo-inverse} = \boldsymbol{X}^+ &= \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top[(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top]^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top)^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}^{-\top}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}\boldsymbol{U}^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Sigma}^\top(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}\boldsymbol{U}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^\top,\end{aligned} \tag{1.36}$$

其中，最后一个等式成立是因为 $\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma}^\top(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}$。

秩亏情形。即使 $\boldsymbol{X}$ 是秩亏的，伪逆仍由 $\boldsymbol{X}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^\top$给出，其中 $\boldsymbol{\Sigma}^+ \in \mathbb{R}^{p\times n}$的左上角具有对角块 $\text{diag}(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_r})$，其余位置均为零。很容易验证，该定义满足(1.34)中的四个彭罗斯条件。

在所有情况下，$\boldsymbol{\Sigma}^+$ 都可以通过取$\boldsymbol{\Sigma}$中每个非零奇异值的倒数得到。因此，通过奇异值分解得到伪逆可概括为：

$$\boldsymbol{X}^+ = \boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}_1^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U}^\top, \quad \text{with } \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \text{ and } \boldsymbol{\Sigma}_1 \in \mathbb{R}^{r\times r}. \tag{1.37}$$

如果 $\boldsymbol{X}$ 是非奇异的，则 $\boldsymbol{X}^+ = \boldsymbol{X}^{-1}$，因此（1.37）是通常逆矩阵的推广。标量的伪逆为

$$\sigma^+ = \begin{cases}1/\sigma, & \text{if } \sigma \neq 0;\\ 0, & \text{if } \sigma = 0.\end{cases}$$

这说明了一个重要事实：伪逆 $\boldsymbol{X}^+$ 不是 $\boldsymbol{X}$的连续函数，除非我们只允许不改变 $\boldsymbol{X}$秩的扰动。伪逆具有以下性质

$$\boldsymbol{X}^+ = \lim_{\delta\to 0}(\boldsymbol{X}^\top\boldsymbol{X} + \delta\boldsymbol{I})^{-1}\boldsymbol{X}^\top.$$

利用（1.37）中的伪逆奇异值分解表示，我们可以推导出矩阵 $\boldsymbol{X}$ 的伪逆的其他性质。

引理 1.57：（利用奇异值分解的伪逆性质）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。伪逆的以下性质可由（1.37）推出：

(i) $\boldsymbol{X}^+ = (\boldsymbol{X}^\top\boldsymbol{X})^+\boldsymbol{X}^\top = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^+$。(ii) 当 $\text{rank}(\boldsymbol{X}) = p$ 时，这变为 $\boldsymbol{X}^+ = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$，$(\boldsymbol{X}^\top)^+ = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$。(iii) $(\boldsymbol{X}^+)^+ = \boldsymbol{X}$。(iv) $(\alpha\boldsymbol{X})^+ = \alpha^+\boldsymbol{X}^+$。(v) $(\boldsymbol{X}^+)^\top = (\boldsymbol{X}^\top)^+$。(vi) $(\boldsymbol{X}^\top\boldsymbol{X})^+ = \boldsymbol{X}^+(\boldsymbol{X}^\top)^+$。(vii) $\boldsymbol{X}$、$\boldsymbol{X}^\top$、$\boldsymbol{X}^+$ 和 $\boldsymbol{X}^+\boldsymbol{X}$ 的秩都相同，且等于 $\text{tr}(\boldsymbol{X}^+\boldsymbol{X})$。(viii) 若 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 正交，则 $(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\top)^+ = \boldsymbol{V}\boldsymbol{X}^+\boldsymbol{U}^\top$。(ix) 若 $\boldsymbol{X} = \sum_i \boldsymbol{X}_i$，其中 $\boldsymbol{X}_i\boldsymbol{X}_j^\top = \boldsymbol{0}$、$\boldsymbol{X}_i^\top\boldsymbol{X}_j = \boldsymbol{0}$、$i \neq j$，则 $\boldsymbol{X}^+ = \sum_i \boldsymbol{X}_i^+$。(x) 若 $\boldsymbol{X}$ 是正规（$\boldsymbol{X}\boldsymbol{X}^\top = \boldsymbol{X}^\top\boldsymbol{X}$），则 $\boldsymbol{X}^+\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^+$ 且 $(\boldsymbol{X}^n)^+ = (\boldsymbol{X}^+)^n$。

需要注意的是，与标准矩阵逆不同，伪逆通常不满足 $\boldsymbol{X}\boldsymbol{X}^+ = \boldsymbol{X}^+\boldsymbol{X}$ 或 $(\boldsymbol{X}\boldsymbol{Y})^+ = \boldsymbol{Y}^+\boldsymbol{X}^+$。例如，设 $\boldsymbol{X} = [1, 0]$且 $\boldsymbol{Y} = [1, 1]^\top$。那么 $\boldsymbol{X}\boldsymbol{Y} = 1$，但 $\boldsymbol{Y}^+\boldsymbol{X}^+ = \frac{1}{2}[1, 1]\begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{2}$. 恒等式 $(\boldsymbol{X}\boldsymbol{Y})^+ = \boldsymbol{Y}^+\boldsymbol{X}^+$ 成立的充要条件由Greville (1966年)给出。

下述定理给出了使该等式成立的一个重要充分条件。

命题 1.58：（（$\boldsymbol{XY}$）$^{+}=\boldsymbol{Y}^{+}\boldsymbol{X}^{+}$的充分条件）如果 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$，$\boldsymbol{Y}\in\mathbb{R}^{p\times m}$，且秩（$\boldsymbol{X}$）$=$秩（$\boldsymbol{Y}$）$=p$，那么（$\boldsymbol{XY}$）$^{+}=\boldsymbol{Y}^{+}\boldsymbol{X}^{+}=\boldsymbol{Y}^{\top}(\boldsymbol{YY}^{\top})^{-1}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}$.

证明 [命题1.58] 最后一个等式由引理1.57得出。(ii)。第一个等式由命题1.54 的证明得出，可通过验证四个彭罗斯条件均成立来确认。

■

1.2.6 著名不等式

本节将介绍若干经典不等式，这些不等式将在全文中频繁使用。柯西–施瓦茨不等式被认为是数学中最重要、应用最广泛的不等式之一。

定理 1.59：（柯西–施瓦茨矩阵（向量）不等式）对于任意矩阵$\boldsymbol{X}$，$\boldsymbol{Y}\in\mathbb{R}^{n\times p}$，以下不等式成立：

$$\left\|\boldsymbol{X}^{\top}\boldsymbol{Y}\right\|_{F}\leq\|\boldsymbol{X}\|_{F}\cdot\|\boldsymbol{Y}\|_{F}.$$

这是将柯西–施瓦茨不等式应用于弗罗贝尼乌斯内积 $\langle\boldsymbol{X},\boldsymbol{Y}\rangle=\operatorname{tr}(\boldsymbol{X}^{\top}\boldsymbol{Y})$ 后直接得到的结果。

类似地，对于任意向量 $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^{n}$，有

$$\left|\boldsymbol{u}^{\top}\boldsymbol{v}\right|\leq\|\boldsymbol{u}\|_{2}\cdot\|\boldsymbol{v}\|_{2}. \tag{1.38}$$

在二维情形下，这便是熟悉的代数恒等式：

$$(ac+bd)^{2}\leq(a^{2}+b^{2})(c^{2}+d^{2}).$$

柯西–施瓦茨不等式的向量形式在现代数学的多个分支中发挥着重要作用，包括希尔伯特空间理论和数值分析（WuandWu，2009年）。为完整起见，我们给出向量情形的一个简单证明。设 $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^{n}$。考虑以下非负和

$$\begin{aligned}0\leq\sum_{i=1}^{n}\sum_{j=1}^{n}(u_iv_j-u_jv_i)^{2}&=\sum_{i=1}^{n}\sum_{j=1}^{n}u_i^{2}v_j^{2}+\sum_{i=1}^{n}\sum_{j=1}^{n}v_i^{2}u_j^{2}-2\sum_{i=1}^{n}\sum_{j=1}^{n}u_iu_jv_iv_j\\&=\left(\sum_{i=1}^{n}u_i^{2}\right)\left(\sum_{j=1}^{n}v_j^{2}\right)+\left(\sum_{i=1}^{n}v_i^{2}\right)\left(\sum_{j=1}^{n}u_j^{2}\right)-2\left(\sum_{i=1}^{n}u_iv_i\right)^{2}\\&=2\|\boldsymbol{u}\|_{2}^{2}\cdot\|\boldsymbol{v}\|_{2}^{2}-2\left\|\boldsymbol{u}^{\top}\boldsymbol{v}\right\|_{2}^{2},\end{aligned}$$

由此即可得到结论。当且仅当 $\boldsymbol{u}=\eta\boldsymbol{v}$ 对于某个常数$\eta\in\mathbb{R}$成立时，等号成立，也就是说，$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 线性相关。

H¨older 不等式以 *Otto H¨older* 命名，是优化、机器学习、概率和泛函分析等领域广泛应用的另一项基石性成果。它推广了柯西–施瓦茨不等式（对应特殊情形 $s=t=2$）。

定理 1.60：（H¨older 不等式）令 s，$t > 1$ 满足 $\frac{1}{s} + \frac{1}{t} = 1$。-则对于任意向量 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathbb{R}^n$，有

$$\sum_{i=1}^{n} x_i y_i \leq \left| \sum_{i=1}^{n} x_i y_i \right| \leq \sum_{i=1}^{n} |x_i| \, |y_i| \leq \left(\sum_{i=1}^{n} |x_i|^s \right)^{1/s} \left(\sum_{i=1}^{n} |y_i|^t \right)^{1/t} = \|\boldsymbol{x}\|_s \, \|\boldsymbol{y}\|_t ,$$

其中 $\|\boldsymbol{x}\|_s = (\sum_{i=1}^{n} |x_i|^s)^{1/s}$ 表示 ℓ_s-范数 $\boldsymbol{x}$。等号成立当且仅当两个序列 $\{|x_i|^s\}$ 和 $\{|y_i|^t\}$ 线性相关 [a]。当 $s = t = 2$时，这就化为向量柯西–施瓦茨不等式（1.38）。

[a]。更具体地说，当且仅当 $|\boldsymbol{x}^\top \boldsymbol{y}| = |\boldsymbol{x}|^\top |\boldsymbol{y}|$ 和

$$\begin{cases} |\boldsymbol{x}| \circ |\boldsymbol{y}| = \|\boldsymbol{y}\|_\infty |\boldsymbol{x}| , & \text{if } s = 1; \\ |\boldsymbol{x}| \circ |\boldsymbol{y}| = \|\boldsymbol{x}\|_\infty |\boldsymbol{y}| , & \text{if } s = \infty; \\ \|\boldsymbol{y}\|_t^{1/s} |\boldsymbol{x}|^{\circ 1/t} = \|\boldsymbol{x}\|_s^{1/t} |\boldsymbol{y}|^{\circ 1/s} , & \text{if } 1 < s < \infty, \end{cases}$$

其中 $\circ$ 表示哈达玛积/幂；参见 Bernstein(2008年) 及其中的参考文献 .

证明 证明见Lu (2021a)。 ■

1.3. 概率论与常见概率分布

随机变量是一个根据某种潜在随机性取不同值的变量，用于对不确定结果或事件进行建模。我们用小写的正体（非斜体）字体表示随机变量本身，用斜体字体的小写字母表示其可能的实现值（已观察到的或潜在的值）。例如，y_1 和 y_2 是随机变量 y 的可能取值（实现值）。对于向量值随机变量，我们使用粗正体符号：随机向量写作$\mathbf{y}$，特定的实现值则用 $\boldsymbol{y}$ 以斜体粗体表示——不过为避免混淆，通常会通过排版加以区分。在本文中，我们约定随机向量表示为$\mathbf{y}$（粗正体），实现值表示为 $\boldsymbol{y}$ （粗斜体）。同样，对于矩阵值随机变量，我们将随机矩阵表示为$\mathbf{Y}$ （粗正体），将特定实现值表示为$\boldsymbol{Y}$ （粗斜体）。

然而，随机变量本身只能描述可能结果的集合——它必须与概率分布结合使用，由概率分布规定每个结果发生的可能性。在概率论中，随机现象的结果被建模为随机变量，而该现象本身则由其累积分布函数（*c.d.f.*、*CDF*）描述，定义如下：

$$F(x) \equiv \Pr[\mathrm{x} \leq x].$$

现在假设该现象具有某种特征 $\boldsymbol{\theta}$，它会影响 x 的结果的概率。这种特征称为参数。由于事件 $\{\mathrm{x} \leq x\}$ 的概率取决于 $\boldsymbol{\theta}$，累积分布函数也必然取决于 $\boldsymbol{\theta}$。因此我们写作：

$$F(x; \boldsymbol{\theta}) \equiv \Pr_{\boldsymbol{\theta}}[\mathrm{x} \leq x].$$

在概率论中，如果已知函数形式 $F(x; \boldsymbol{\theta})$ 和 $\boldsymbol{\theta}$ 的真实值，那么对于任何可能的结果 x，我们都可以计算其概率：$\Pr_{\boldsymbol{\theta}}[\mathrm{x} \leq x] = F(x; \boldsymbol{\theta})$。

概率论中的一个有用结果是并集界，它给出了一组事件之并集的概率上界。

定理 1.61：（并集界）设 $\mathbb{A}_1, \mathbb{A}_2, \ldots, \mathbb{A}_n$ 是概率空间中的一组事件。那么

$$\Pr\left(\cup_i \mathbb{A}_i\right) \leq \sum_{i=1}^{n} \Pr(\mathbb{A}_i).$$

证明 [定理 1.61] 对每个 $i \in \{1, 2, \ldots, n\}$，细化集合 $\widetilde{\mathbb{A}}_i \triangleq \mathbb{A}_i \cap \bigcap_{j=1}^{i-1} \mathbb{A}_j^c$ 容易证明 $\bigcup_{j=1}^{k} \widetilde{\mathbb{A}}_j = \bigcup_{j=1}^{k} \mathbb{A}_j$ 对任意 k，由归纳法可得 $\bigcup_i \mathbb{A}_i = \bigcup_i \widetilde{\mathbb{A}}_i$。~~根据构造，集合 $\mathbb{A}_1$，$\mathbb{A}_2$，..... 两两不相交，因此

$$\Pr\left(\bigcup_i \mathbb{A}_i\right) = \Pr\left(\bigcup_i \widetilde{\mathbb{A}}_i\right) = \sum_i \Pr\left(\widetilde{\mathbb{A}}_i\right) \leq \sum_i \Pr(\mathbb{A}_i)$$

~~其中，最后一个不等式成立是因为 $\mathbb{A}_i \subseteq \mathbb{A}_i$。 ■

► **离散随机变量**。 随机变量可以是离散的，也可以是连续的。离散随机变量 取有限个或可数无穷多个不同的状态。这些状态不必是数值；它们可以是符号或类别状态，不具有数值（例如“红色”“蓝色”“成功”“失败”）。相比之下，连续随机变量 取值于不可数集合中，通常是实数集的子集。

概率分布描述了如何为一个或多个随机变量可能出现的结果分配概率。表示这种分布的方式取决于变量是离散的还是连续的。

对于离散随机变量，我们使用概率质量函数（*p.m.f.*，*PMF*），通常用大写的 Pr 表示（有时也用 f）。概率质量函数将随机变量的每个可能状态映射为该变量取该状态的概率。例如，记号

$$\Pr(\mathrm{y} = y) \quad (\text{or equivalently } \Pr(y),\ f_{\mathrm{y}}(y),\ \Pr_{\mathrm{y}}(y))$$

表示随机变量 y 等于值 y 的概率。根据定义，概率取值范围为 $[0, 1]$，，其中 1 表示必然发生，0 表示不可能发生。另一种常见的约定是先定义一个随机变量，再用“∼”符号指定其分布：

$$\mathrm{y} \sim \Pr(\mathrm{y}).$$

概率质量函数也可以同时定义在多个随机变量上。这称为联合概率质量分布函数 或联合频率函数。例如，$\Pr(\mathsf{x} = x, \mathsf{y} = y)$ 表示 $x = x$ 和 $y = y$ 同时发生的概率。我们常使用简写记号，例如 $\Pr(x, y)$ 或 $\Pr_{\mathrm{x},\,\mathrm{y}}(x, y)$。此外，如果概率质量函数依赖于某些已知参数 $\boldsymbol{\theta}$，为简洁起见，我们可以将其写作 $\Pr(x, y \mid \boldsymbol{\theta})$（或 $\Pr_{\boldsymbol{\theta}}(x, y)$，$f(x, y; \boldsymbol{\theta})$）。

在许多应用中，我们关心在另一个事件已发生的条件下某个事件发生的概率。这称为条件概率。在 $\mathsf{y} = y$ 发生的条件下 $\mathsf{x} = x$ 的条件概率记作 $\Pr(\mathsf{x} = x \mid \mathsf{y} = y)$，其计算公式为：

$$\Pr(\mathrm{x} = x \mid \mathrm{y} = y) = \frac{\Pr(\mathrm{x} = x, \mathrm{y} = y)}{\Pr(\mathrm{y} = y)},$$

前提是 $\Pr(\mathsf{y} = y) > 0$。这一关系是贝叶斯定理(贝叶斯，1958年）的基础。

反过来，当已知一组变量的联合分布时，我们可能希望求出其中某个子集的分布。这称为边缘概率质量分布（或简称边缘分布）。例如，若 x 和 y 是已知联合 PMF Pr(x，y) 的离散随机变量，则 x 的边缘分布可通过对 y 的所有可能取值求和得到：

$$\Pr(\mathrm{x} = x) = \sum_{y} \Pr(\mathrm{x} = x, \mathrm{y} = y).$$

▶ **连续随机变量**。处理连续随机变量时，我们使用概率密度函数（概率密度函数，*PDF*）而不是概率质量函数（PMF）来描述概率分布。函数 p 要成为有效的概率密度函数，必须满足以下性质：

- p 的定义域必须是随机变量 y 的所有可能取值（或状态）组成的集合。与概率质量函数不同，概率密度函数不必满足 $p(y) \leq 1$。但是，它在任何地方都必须非负：
- $\forall y \in \mathrm{y}\ p(y) \geq 0$。它在整个定义域上的积分必须为 1：$\int p(y)dy = 1$。
-

重要的是，数值 $p(y)$（也记作 $f_{\mathrm{y}}(y)$ 或 $p_{\mathrm{y}}(y)$）并不表示概率，而是表示密度。随机变量落在 y 附近宽度为 δy 的无穷小区间内的概率近似为 $p(y)\delta y$。更正式地说，对于任意可测集 $\mathbb{A}$，$y \in \mathbb{A}$ 的概率为

$$\Pr(y \in \mathbb{A}) = \int_{\mathbb{A}} p(y)dy.$$

如果概率密度函数（PDF）取决于已知参数 $\boldsymbol{\theta}$，通常写作

$$p(x \mid \boldsymbol{\theta}), \quad f_{\mathrm{x}}(x; \boldsymbol{\theta}), \quad \text{or simply } f(x; \boldsymbol{\theta}).$$

其中，分号或竖线表示对参数的依赖关系 .

▶ **分布函数**。另一方面，任何概率分布也可以通过其累积分布函数（c.d.f.、CDF）来表征。设 x 为一个随机变量，其 CDF 定义为 $F(x) = \Pr[\mathrm{x} \leq x]$，其中 $F(x)$ 是非递减的、右连续的，并且满足

$$0 \leq F(x) \leq 1, \qquad \lim_{x \to -\infty} F(x) = 0, \qquad \text{and} \qquad \lim_{x \to \infty} F(x) = 1.$$

连续情形下，x 的期望值（均值）μ 和方差 ω^2 定义为

$$\mu = \mathbb{E}[\mathrm{x}] = \int_{-\infty}^{\infty} x dF(x), \qquad \omega^2 = \mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}[(\mathrm{x} - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x). \tag{1.39a}$$

类似地，在离散情形下，其中 $\Pr(\mathrm{x} = x)$ 表示概率质量函数，它们定义为

$$\mu = \mathbb{E}[\mathrm{x}] = \sum_{x} x \Pr(x), \qquad \omega^2 = \mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}[(\mathrm{x} - \mu)^2] = \sum_{x} (x - \mu)^2 \Pr(x). \tag{1.39b}$$

不难验证以下有用的恒等式：

$$\mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}[\mathrm{x}^2] - (\mathbb{E}[\mathrm{x}])^2. \tag{1.39c}$$

随机变量 x 的中位数是满足以下条件的任意数 M

$$\Pr(\mathrm{x} \geq M) \geq \frac{1}{2} \qquad \text{and} \qquad \Pr(\mathrm{x} \leq M) \geq \frac{1}{2}. \tag{1.39d}$$

练习 1.62（均匀分布）。设 $\mathrm{x} \sim \mathrm{Uniform}(x \mid a, b)$ 是一个均匀分布变量，使得 $f_{\mathrm{x}}(x) = \frac{1}{b-a}$ 当 $a \leq x \leq b$ 时成立，而 $f_{\mathrm{x}}(x) = 0$ 在其他情况下成立。证明

$$\mathbb{E}[\mathrm{x}] = \frac{a+b}{2} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}[\mathrm{x}] = \frac{(b-a)^2}{12}.$$

现在考虑一个随机向量 $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_n]^\top$。其联合概率分布函数（也简称为概率分布函数）记作 $F_{\mathbf{x}}(\boldsymbol{x})$,$F(x_1, x_2, \ldots, x_n)$，或 $F(\boldsymbol{x})$，定义如下：

$$F_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = \Pr[\mathrm{x}_1 \leq x_1, \mathrm{x}_2, \leq x_2, \ldots, \mathrm{x}_n \leq x_n].$$

根据各分量是离散还是连续，我们将联合分布描述如下：

- 当所有变量都是离散变量时，如上所述，联合概率质量函数或联合频率函数可表示为

$$f_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = \Pr[\mathrm{x}_1 = x_1, \mathrm{x}_2 = x_2, \ldots, \mathrm{x}_n = x_n].$$

- 相反，如果变量是连续的，则联合概率密度函数是一个函数 $f_{\mathbf{x}}\colon \mathbb{R}^n \to [0, \infty)$（或简称为 $f(\boldsymbol{x})$、$p_{\mathbf{x}}(\boldsymbol{x})$、$p(\boldsymbol{x})$），使得

$$F_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{x}}(y_1, y_2, \ldots, y_n) dy_1 \ldots dy_n.$$

在这种情况下，当 $f_{\mathbf{x}}$ 在 $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top$处连续时，我们有

$$f_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\mathbf{x}}(x_1, x_2, \ldots, x_n).$$

当且仅当以下条件成立时，这些随机变量相互独立

$$\begin{aligned} & F_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = F_{\mathrm{x}_1}(x_1) \cdot F_{\mathrm{x}_2}(x_2) \cdot \ldots \cdot F_{\mathrm{x}_n}(x_n); \\ \text{or} \quad & f_{\mathbf{x}}(x_1, x_2, \ldots, x_n) = f_{\mathrm{x}_1}(x_1) \cdot f_{\mathrm{x}_2}(x_2) \cdot \ldots \cdot f_{\mathrm{x}_n}(x_n). \end{aligned}$$

连续情形下的 条件或边缘概率密度函数 与离散情形类似，只是以积分代替求和。例如，$f_{\mathrm{x}}(x) = \int f_{\mathrm{x},\,\mathrm{y}}(x,\ y) dy$。

具体而言，两个随机变量 x_i 和 x_j 的联合分布由联合累积分布函数 $F(x_i,\ x_j) \equiv \Pr[\mathrm{x}_i \leq x_i,\ \mathrm{x}_j \leq x_j]$给出。然后，$\mathrm{x}_i$ 和 x_j之间的协方差记为 ω_{ij}，定义为

$$\omega_{ij} \triangleq \mathbb{C}\mathrm{ov}[\mathrm{x}_i, \mathrm{x}_j] \triangleq \mathbb{E}[(\mathrm{x}_i - \mu_i)(\mathrm{x}_j - \mu_j)] = \int_{x_i, x_j = -\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) dF(x_i, x_j),$$

其中 $\mu_i = \mathbb{E}[\mathrm{x}_i]$ 等价地，$\omega_{ij} = \mathbb{E}[\mathrm{x}_i \mathrm{x}_j] - \mu_i \mu_j$。对于随机向量 $\mathbf{x} \in \mathbb{R}$，协方差矩阵 $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$ 定义为

$$\mathbb{C}\mathrm{ov}[\mathbf{x}] \triangleq \boldsymbol{\Omega} = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] = \mathbb{E}(\mathbf{x}\mathbf{x}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top, \tag{1.40}$$

其中 $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = [\mu_1, \mu_2, \ldots, \mu_n]^\top$。

两个随机变量的方差量化了二者之间线性相关性的程度。一个密切相关的度量是**相关性**，其定义为

$$\mathbb{C}\mathrm{orr}[\mathrm{x}_1, \mathrm{x}_2] \triangleq \frac{\mathbb{C}\mathrm{ov}[\mathrm{x}_1, \mathrm{x}_2]}{\sqrt{\mathbb{V}\mathrm{ar}[\mathrm{x}_1]\mathbb{V}\mathrm{ar}[\mathrm{x}_2]}}. \tag{1.41}$$

虽然协方差和相关性都能反映线性相关性，但相关性与尺度无关，并且始终位于区间 $[-1, 1]$ 内。此外，还有

$$|\mathbb{C}\mathrm{orr}[\mathrm{x}_1, \mathrm{x}_2]| \leq \sqrt{\mathbb{V}\mathrm{ar}[\mathrm{x}_1]\mathbb{V}\mathrm{ar}[\mathrm{x}_2]}. \tag{1.42}$$

引理 1.63：（线性变换）令 $\mathbf{y} = \boldsymbol{A}\mathbf{x} + \boldsymbol{b}$，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 和 $\boldsymbol{b} \in \mathbb{R}^m$ 为确定性的；设 $\mathbf{x} \in \mathbb{R}^n$ 为均值为 $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} \in \mathbb{R}^n$、协方差为 $\boldsymbol{\Omega} \in \mathbb{R}^{n\times n}$ 的随机向量。则

$$\mathbb{E}[\mathbf{y}] = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} \qquad \text{and} \qquad \mathbb{C}\mathrm{ov}[\mathbf{y}] = \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^\top.$$

证明 [引理 1.63] 第一个等式由期望的线性性得到。对于第二个等式，注意到

$$\mathbb{C}\mathrm{ov}[\mathbf{y}] = \mathbb{C}\mathrm{ov}[\boldsymbol{A}\mathbf{x}] = \mathbb{E}[\boldsymbol{A}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{A}^\top] = \boldsymbol{A}\mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top]\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^\top.$$

证明完毕。 ■

在 $\boldsymbol{A} = \boldsymbol{a}^\top$ 为行向量的特殊情况下，$\mathbf{y} = \boldsymbol{a}^\top\mathbf{x}$ 是关于 $\mathbf{x}$ 的线性泛函。因此，若 $\mathbb{C}\mathrm{ov}[\mathbf{x}] = \sigma^2\boldsymbol{I}$，则 $\mathbb{C}\mathrm{ov}[\mathbf{y}] = \sigma^2\boldsymbol{a}^\top\boldsymbol{a}$。

引理 1.64：（二次变换）设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 为对称矩阵，且 $\mathbf{x} \in \mathbb{R}^n$ 为均值为 $\boldsymbol{\mu} = [\mu_i] \in \mathbb{R}^n$、协方差为 $\boldsymbol{\Omega} = [\omega_{ij}] \in \mathbb{R}^{n\times n}$ 的随机向量。那么

$$\mathbb{E}[\mathbf{x}^\top\boldsymbol{A}\mathbf{x}] = \boldsymbol{\mu}^\top\boldsymbol{A}\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}),$$

其中，$\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega})$ 表示 $\boldsymbol{A}\boldsymbol{\Omega}$ 的迹，即该矩阵对角元素之和。

证明 [引理 1.64] $\mathbf{x}^\top\boldsymbol{A}\mathbf{x} = \sum_{i=1}^n\sum_{j=1}^n a_{ij}\mathrm{x}_i\mathrm{x}_j$。由于 $\mathbb{E}[\mathrm{x}_i\mathrm{x}_j] = \mu_i\mu_j + \omega_{ij}$，因此 $\mathbb{E}[\mathbf{x}^\top\boldsymbol{A}\mathbf{x}] = \sum_{i=1}^n\sum_{j=1}^n a_{ij}\mathbb{E}[\mathrm{x}_i\mathrm{x}_j]$。代入期望 $\mathrm{x}_i\mathrm{x}_j$：

$$\mathbb{E}[\mathbf{x}^\top\boldsymbol{A}\mathbf{x}] = \sum_{i=1}^n\sum_{j=1}^n a_{ij}\mu_i\mu_j + \sum_{i=1}^n\sum_{j=1}^n a_{ij}\omega_{ij} = \boldsymbol{\mu}^\top\boldsymbol{A}\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}).$$

证明至此完成。 ■

在许多应用中，我们将随机向量进行中心化，使得 $\boldsymbol{\mu} = \mathbf{0}$。在这种情况下，二次型期望可简化为 $\mathbb{E}[\mathbf{x}^\top\boldsymbol{A}\mathbf{x}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega})$。进一步地，如果 $\boldsymbol{\Omega} = \sigma^2\boldsymbol{\Sigma}$，那么 $\mathbb{E}[\mathbf{x}^\top\boldsymbol{A}\mathbf{x}] = \sigma^2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})$。为了得到 σ^2 的无偏估计量，我们可以选择一个矩阵 $\boldsymbol{A}$，使得 $\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) = 1$。

▶ **其他矩。** $\mathbb{E}[\mathrm{x}^s]$ $s>0$ 称为随机变量 x 的矩，而 $\mathbb{E}[|\mathrm{x}|^s]$ 称为其绝对矩。特别地，均值 $\mu=\mathbb{E}[\mathrm{x}]$和方差 $\mathbb{E}[(\mathrm{x}-\mathbb{E}\mathrm{x})^2]=\mathbb{E}[\mathrm{x}^2]-(\mathbb{E}[\mathrm{x}])^2$ 是 x 的矩的特殊情形。对于$1\le s<\infty$, $(\mathbb{E}[|\mathrm{x}|^s])^{1/s}$ 定义了一个范数；特别地，三角不等式成立：

$$(\mathbb{E}[|\mathrm{x}+\mathrm{y}|^s])^{1/s}\le(\mathbb{E}[|\mathrm{x}|^s])^{1/s}+(\mathbb{E}[|\mathrm{y}|^s])^{1/s}\quad\text{for all x, y.}\tag{1.43}$$

类似地，*H¨older's inequality*（霍尔德不等式）对于随机变量的表述是：对于同一概率空间上的随机变量 x、y，有

$$\text{Hölder's inequality:}\qquad |\mathbb{E}[\mathrm{xy}]|\le\big(\mathbb{E}[|\mathrm{x}|^s]\big)^{1/s}\big(\mathbb{E}[|\mathrm{y}|^t]\big)^{1/t},\quad s,t\ge1,\ \frac{1}{s}+\frac{1}{t}=1.\tag{1.44a}$$

特殊情形 $s=t=2$ 给出了随机变量的 柯西–施瓦茨不等式：

$$\text{Cauchy–Schwarz inequality:}\qquad |\mathbb{E}[\mathrm{xy}]|\le\sqrt{\mathbb{E}[|\mathrm{x}|^2]\mathbb{E}[|\mathrm{y}|^2]}.\tag{1.44b}$$

这自然可以推广到随机矩阵。对于任意随机 $n\times p$ 矩阵$\mathbf{X}$ 和$\mathbf{Y}$，弗罗贝尼乌斯范数形式的柯西–施瓦茨不等式为

$$\text{Cauchy–Schwarz inequality:}\quad \mathbb{E}\left[\left\|\mathbf{X}^\top\mathbf{Y}\right\|_F\right]\le\mathbb{E}\left[\|\mathbf{X}\|_F^2\right]^{1/2}\mathbb{E}\left[\|\mathbf{Y}\|_F^2\right]^{1/2},\tag{1.44c}$$

其中，内积定义为 $\langle\mathbf{X},\ \mathbf{Y}\rangle=\mathbb{E}\left[\left\|\mathbf{X}^\top\mathbf{Y}\right\|_F\right]$。类似的矩阵推广也适用于 H¨older's inequality（霍尔德不等式）。由于常数（确定性）随机变量 1 的期望为 $\mathbb{E}[1]=1$，H¨older's inequality（霍尔德不等式）表明

$$\mathbb{E}[|\mathrm{x}|^a]=\mathbb{E}[1\cdot|\mathrm{x}|^a]\le(\mathbb{E}[|\mathrm{x}|^b])^{a/b},\quad\text{for all } a>0,\frac{b}{a}\ge1.\tag{1.44d}$$

因此，

$$\big(\mathbb{E}[|\mathrm{x}|^s]\big)^{1/s}\le\big(\mathbb{E}[|\mathrm{x}|^t]\big)^{1/t},\quad\text{for all } 0<s\le t<\infty.\tag{1.44e}$$

绝对矩可以用尾概率表示，如下列引理所示。

引理 1.65：（绝对矩引理）对于任意随机变量 x 以及 $s>0$,

$$\mathbb{E}[|\mathrm{x}|^s]=s\int_0^\infty\Pr(|\mathrm{x}|\ge\tau)\tau^{s-1}d\tau,\quad s>0.$$

证明 [引理 1.65] 回忆一下，示性函数 $\mathbb{1}(|\mathrm{x}|^s\ge x)$ 是一个随机变量：在事件 $\{|\mathrm{x}|^s\ge x\}$ 上取值为 1，否则取值为 0。利用富比尼定理交换积分次序，计算得：

$$\begin{aligned}\mathbb{E}[|\mathrm{x}|^s]&=\int_\Omega|x|^s\,dF(x)=\int_\Omega\int_0^{|x|^s}1dydF(x)=\int_\Omega\int_0^\infty\mathbb{1}(|x|^s\ge y)dydF(x)\\&=\int_0^\infty\int_\Omega\mathbb{1}(|x|^s\ge y)dF(x)dy=\int_0^\infty\Pr(|\mathrm{x}|^s\ge y)dy\\&\xrightarrow{y\triangleq\tau^s}s\int_0^\infty\Pr(|\mathrm{x}|^s\ge\tau^s)\tau^{s-1}d\tau=s\int_0^\infty\Pr(|\mathrm{x}|\ge\tau)\tau^{s-1}d\tau,\end{aligned}$$

最后一个等式由变量替换得出。证明至此完成。 ■

习题 1.66。 利用引理 1.65 证明随机变量 x 的期望满足

$$\mathbb{E}[\mathrm{x}] = \int_0^\infty \Pr(\mathrm{x} \geq \tau) d\tau - \int_0^\infty \Pr(\mathrm{x} \leq -\tau) d\tau.$$

提示：将随机变量x 表示为$\mathrm{x} = \mathrm{x} \cdot \mathbb{1}\big(\mathrm{x} \in [0, \infty)\big) + \mathrm{x} \cdot \mathbb{1}\big(-\mathrm{x} \in (0, \infty)\big)$。

▶ **矩母函数（MGF）。** 函数 $\theta \mapsto \mathbb{E}[\exp(\theta \mathrm{x})]$ 通常称为随机变量 x 的拉普拉斯变换或矩母函数（*MGF*）。它是概率论和统计学中的有力工具，因为它编码了随机变量（或随机向量）的矩信息，并能简化独立随机变量之和的分析。

定义 1.67（矩母函数（MGF））。 设 x 是一个实值随机变量。其矩母函数（*MGF*）定义为

$$\mathbb{M}_{\mathrm{x}}(\theta) = \mathbb{E}\left[\exp(\theta \mathrm{x})\right] : \mathbb{R} \to \mathbb{R} \cup \{\infty\}, \quad \theta \in \mathbb{R} \tag{1.45a}$$

类似地，对于随机向量$\mathbf{x} \in \mathbb{R}^n$，MGF 定义为

$$\mathbb{M}_{\mathbf{x}}(\boldsymbol{\theta}) = \mathbb{E}\left[\exp(\boldsymbol{\theta}^\top \mathbf{x})\right] : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}, \quad \boldsymbol{\theta} \in \mathbb{R}^n. \tag{1.45b}$$

另一方面，x 的累积量生成函数 *(CGF)* 是其矩母函数的对数：

$$\mathbb{K}_{\mathrm{x}}(\theta) = \ln \mathbb{M}_{\mathrm{x}}(\theta). \tag{1.45c}$$

以下是矩母函数的一些重要性质：

1. 由矩母函数求矩。 如果矩母函数在零附近的某个开区间内存在，那么 x 的所有矩都存在，并且可以通过对矩母函数求导并在 $\theta = 0$ 处取值来计算：

$$\mathbb{E}[\mathrm{x}^k] = \frac{d^k}{d\theta^k} \mathbb{M}_{\mathrm{x}}(0),$$

其中 k 是表示矩的阶的正整数。

2. 唯一性。 如果两个随机变量在包含零的某个开区间内具有相同的矩母函数，那么它们具有相同分布。也就是说，如果对于包含零的某个开区间内的所有 θ，$\mathbb{M}_{\mathrm{x}}(\theta) = \mathbb{M}_{\mathrm{y}}(\theta)$，那么 x 和 y 的累积分布函数相同（$F_{\mathrm{x}} = F_{\mathrm{y}}$）。

3. 独立变量的乘积。 对于独立随机变量 x 和 y，它们的和 $\mathrm{x} + \mathrm{y}$ 的矩母函数等于各自矩母函数的乘积：

$$\mathbb{M}_{\mathrm{x}+\mathrm{y}}(\theta) = \mathbb{M}_{\mathrm{x}}(\theta) \mathbb{M}_{\mathrm{y}}(\theta), \quad \text{if x and y are independent.}$$

4. 另一方面，累积量生成函数对独立随机变量具有可加性

$$\mathbb{K}_{\mathrm{x}+\mathrm{y}}(\theta) = \mathbb{K}_{\mathrm{x}}(\theta) + \mathbb{K}_{\mathrm{y}}(\theta), \quad \text{if x and y are independent.}$$

这使累积量在极限定理和渐近分析中尤其有用。累积量生成函数常常能简化对独立随机变量之和的分析。

这些性质使矩母函数成为分析独立随机变量之和、求分布的矩以及证明分布等价性的有用工具。不过，需要注意的是，并非所有分布都有矩

母函数；例如，柯西分布没有矩母函数，因为它的矩不存在。在这种情况下，可以改用特征函数等其他工具。

1.3.1 一些集中不等式

马尔可夫不等式和切比雪夫不等式是概率论中的基本集中不等式。它们为随机变量偏离其期望值的概率提供上界，其中马尔可夫不等式适用于非负变量，而切比雪夫不等式则通过考虑任意随机变量的方差扩展了这一概念。

更具体地说，函数 $\tau \mapsto \Pr(|\mathrm{x}| \geq \tau)$ 称为随机变量 x 的尾部 （或尾部概率）。马尔可夫不等式提供了一种简单但往往实用的方法，可以利用期望来界定这一尾部。

定理 1.68：（马尔可夫不等式）设 x 是一个**非负**随机变量。那么，对于任意 $\tau > 0$，有

$$\Pr(\mathrm{x} \geq \tau) \leq \frac{\mathbb{E}[\mathrm{x}]}{\tau}.$$

证明 [定理1.68] 由于 x 非负，我们有逐点不等式 $0 \leq \tau \mathbb{1}\{\mathrm{x} \geq \tau\} \leq \mathrm{x}$。这意味着 $\mathbb{E}[\tau \mathbb{1}\{\mathrm{x} \geq \tau\}] \leq \mathbb{E}[\mathrm{x}]$。我们还有

$$\mathbb{E}\left[\tau \mathbb{1}\{\mathrm{x} \geq \tau\}\right] = \tau \mathbb{E}\left[\mathbb{1}\{\mathrm{x} \geq \tau\}\right] = \tau\left(1 \cdot \Pr(\mathrm{x} \geq \tau) + 0 \cdot \Pr(\mathrm{x} < \tau)\right) = \tau \cdot \Pr(\mathrm{x} \geq \tau) \leq \mathbb{E}[\mathrm{x}].$$

证明完毕。

或者，令 $f(x)$ 表示 x 的概率密度函数。那么，

$$\mathbb{E}[\mathrm{x}] = \int_0^\infty x f(x)\, dx \geq \int_\tau^\infty x f(x)\, dx \geq \int_\tau^\infty \tau f(x)\, dx = \tau \Pr(\mathrm{x} \geq \tau),$$

由此可得该结果。 ■

设 x 为任意随机变量（不一定非负）。将马尔可夫不等式应用于非负随机变量 $|\mathrm{x}|$，我们得到 x 的尾界：

$$\Pr(|\mathrm{x}| \geq \tau) \leq \frac{\mathbb{E}[|\mathrm{x}|]}{\tau}, \quad \text{for all } \tau > 0. \tag{1.46a}$$

类似地，对于 $\theta > 0$，我们得到

$$\Pr(\mathrm{x} \geq \tau) = \Pr\left(\exp(\theta \mathrm{x}) \geq \exp(\theta\tau)\right) \leq \frac{\mathbb{E}[\exp(\theta \mathrm{x})]}{\exp(\theta\tau)}, \quad \text{for all } \tau \in \mathbb{R}. \tag{1.46b}$$

一个重要的结论是以下广义马尔可夫型界：对于任意 $s > 0$,

$$\Pr(|\mathrm{x}| \geq \tau) = \Pr(|\mathrm{x}|^s \geq \tau^s) \leq \tau^{-s} \mathbb{E}[|\mathrm{x}|^s], \quad \text{for all } t > 0. \tag{1.46c}$$

特殊情况 $s = 2$ 可得切比雪夫不等式，前提是方差存在。

定理 1.69：（切比雪夫不等式）设 x 是一个均值有限的随机变量 $\mathbb{E}[\mathrm{x}] < \infty$。那么，对于任意 $\tau > 0$，我们有

$$\Pr(|\mathrm{x} - \mathbb{E}[\mathrm{x}]| \geq \tau) \leq \frac{\mathbb{V}\mathrm{ar}[\mathrm{x}]}{\tau^2}.$$

这可直接通过对非负随机变量 $\mathrm{y} \triangleq (\mathrm{x} - \mathbb{E}[\mathrm{x}])^2$ 应用马尔可夫不等式得出。

切尔诺夫方法利用矩母函数推导出尾概率的指数衰减上界。由此得到的不等式称为*Chernoff*界。

定理 1.70：（Chernoff 界）设 x 为随机变量。那么，对于任意 $\theta > 0$

$$\Pr(\mathrm{x} \geq \tau) \leq \min_{\theta > 0} \left(\exp(-\theta\tau)\mathbb{M}_{\mathrm{x}}(\theta)\right).$$

这可通过对 $\exp(\theta\mathrm{x})$ 应用马尔可夫不等式得到：$\Pr(\mathrm{x} \geq \tau) = \Pr(\exp(\theta\mathrm{x}) \geq \exp(\theta\tau)) \leq \mathbb{M}_{\mathrm{x}}(\theta)\exp(-\theta\tau)$，并取下确界。

这一思想的一个强大推广是克拉 ´默定理，它为独立随机变量之和提供指数尾界。

定理 1.71：（克拉 ´默定理（克拉 ´默，1994年）设 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ 是一列独立随机变量，其累积量生成函数为 $\mathbb{K}_{\mathrm{x}_i}$、$i \in \{1, 2, \ldots, n\}$。那么，对于任意 $\tau > 0$，

$$\Pr\left(\sum_{i=1}^{n} \mathrm{x}_i \geq \tau\right) \leq \exp\left(\min_{\theta > 0}\left\{-\theta\tau + \sum_{i=1}^{n} \mathbb{K}_{\mathrm{x}_i}(\theta)\right\}\right).$$

定理 [的证明：1.71] 我们使用切尔诺夫方法。对于任意 $\theta > 0$，由马尔可夫不等式（定理 1.68）和独立性可得

$$\begin{aligned}
\Pr\left(\sum_{i=1}^{n} \mathrm{x}_i \geq \tau\right) &= \Pr\left(\exp\left(\theta\sum_{i=1}^{n} \mathrm{x}_i\right) \geq \exp(\theta\tau)\right) \leq \exp(-\theta\tau)\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^{n} \mathrm{x}_i\right)\right] \\
&= \exp(-\theta\tau)\mathbb{E}\left[\prod_{i=1}^{n} \exp(\theta\mathrm{x}_i)\right] = \exp(-\theta\tau)\prod_{i=1}^{n} \mathbb{E}\left[\exp(\theta\mathrm{x}_i)\right] \\
&= \exp(-\theta\tau)\prod_{i=1}^{n} \exp(\mathbb{K}_{\mathrm{x}_i}(\theta)) = \exp\left(-\theta\tau + \sum_{i=1}^{n} \mathbb{K}_{\mathrm{x}_i}(\theta)\right).
\end{aligned}$$

由于这对每个 $\theta > 0$ 都成立，对 $\theta > 0$ 取下确界即可得到所需的上界。 ■

Cram ´er 定理为独立随机变量之和的尾部概率提供了指数上界。相比之下，当各项几乎必然有界时，霍夫丁不等式给出的上界更紧——这在经验过程理论和学习理论中很常见。

引理1.72：（霍夫丁引理）设 x 是一个实值随机变量，使得$\mathbb{E}[\mathrm{x}]=0$ 且 $\mathrm{x}\in[a,b]$ 几乎必然成立。那么对于任意 $\theta\in\mathbb{R}$

$$\mathbb{M}_{\mathrm{x}}(\theta)\equiv\mathbb{E}\big[\exp(\theta\mathrm{x})\big]\le\exp\Big(\frac{\theta^2(b-a)^2}{8}\Big).$$

引理 [的证明1.72] 设 $t\triangleq\theta(b-a)$ 且 $\mathrm{y}\triangleq\frac{\mathrm{x}-a}{b-a}\in[0,1]$。则 $\mathrm{x}=a+(b-a)\mathrm{y}$ 且$\mathbb{E}[\mathrm{y}]=\frac{\mathbb{E}[\mathrm{x}]-a}{b-a}=\frac{-a}{b-a}\triangleq\mu$，因为 $\mathbb{E}[\mathrm{x}]=0$，其中 $\mu\in[0,1]$ 因为 $a\le\mathbb{E}[\mathrm{x}]\le b$。指数函数是凸函数，因此对任意 $\lambda\in[0,1]$ 和任意实数 s，

$$\exp(\lambda s)\le(1-\lambda)+\lambda e^{s}.$$

在 $\lambda=\mathrm{y}$ 和 $s=t$ 中应用此不等式，其中 $\mathrm{y}=\frac{\mathrm{x}-a}{b-a}\in[0,1]$ 且 $t=\theta(b-a)$，可得

$$\exp(\theta\mathrm{x})=\exp(\theta a)\exp(t\mathrm{y})\le\exp(\theta a)\big(1-\mathrm{y}+\mathrm{y}e^{t}\big).$$

对两边取期望可得

$$\mathbb{E}[\exp(\theta\mathrm{x})]\le\exp(\theta a)\big(1-\mathbb{E}[\mathrm{y}]+\mathbb{E}[\mathrm{y}]e^{t}\big)=\exp(\theta a)\big(1-\mu+\mu e^{t}\big).$$

取对数并定义 $g(t)\triangleq\ln\big(1-\mu+\mu e^{t}\big).$，则累积量生成函数满足

$$\mathbb{K}_{\mathrm{x}}(\theta)\equiv\ln\mathbb{E}[\exp(\theta\mathrm{x})]\le\theta a+g(t),\qquad t=\theta(b-a).\tag{1.47}$$

我们通过其二阶导数分析 $g(t)$：

$$g''(t)=\frac{\mu(1-\mu)e^{t}}{\big(1-\mu+\mu e^{t}\big)^{2}}.$$

为得到其上界，注意到基本不等式 $4uv\le(u+v)^2$。令 $u=\sqrt{\mu e^{t}}$ 和$v=\sqrt{1-\mu}$，可得

$$\mu(1-\mu)e^{t}=(uv)^2\le\Big(\frac{u^2+v^2}{2}\Big)^2=\frac{\big(\mu e^{t}+1-\mu\big)^2}{4}.$$

将两边同时除以 $(1-\mu+\mu e^{t})^2$，得到

$$g''(t)\le\frac{1}{4}.$$

由于 $g''(t)\le\frac{1}{4}$ 对所有 t成立，且 $g(0)=\ln(1)=0$，$g'(0)=\frac{\mu e^{0}}{1-\mu+\mu e^{0}}=\mu$，我们可以进行两次积分。此外，注意右侧在 θ 处关于 θ 式（1.47）的导数在 $\theta=0$ 处为零：

$$\frac{d}{d\theta}\big(\theta a+g(\theta(b-a))\big)\Big|_{\theta=0}=a+g'(0)\,(b-a)=a+\mu(b-a)=0.$$

因此，式（1.47）右侧关于 $t=0$在 $g''(t)\le1/4$处的泰勒展开为

$$\ln\mathbb{E}[\exp(\theta\mathrm{x})]\le\theta a+g(t)\le\frac{1}{2}\cdot\frac{1}{4}t^2=\frac{t^2}{8}=\frac{\theta^2(b-a)^2}{8}.$$

对两边取指数，得到所需的上界： ■

利用上述引理，我们证明霍夫丁不等式，该不等式量化了有界独立随机变量之和在其均值附近的集中程度。

定理 1.73：（霍夫丁不等式（霍夫丁，1963年）） 令 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ 为独立随机变量，使得每个 x_i 几乎处处有界于区间$[a_i, b_i]$内，其中 a_i 和 b_i 为已知常数。令 $\mathrm{s}_n \triangleq \sum_{i=1}^n \mathrm{x}_i$ 为这些随机变量之和，并令 $\mathbb{E}[\mathrm{s}_n]$ 为其均值。则对于任意 $\tau > 0$，霍夫丁不等式表明：

$$\Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \geq \tau) \leq \exp\left(-\frac{2\tau^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

类似地，对于下尾：

$$\Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \leq -\tau) \leq \exp\left(-\frac{2\tau^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

因此，双侧形式成立：

$$\Pr(|\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n]| \geq \tau) \leq 2\exp\left(-\frac{2\tau^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

该上界关于 τ^2 呈指数衰减，这意味着偏离均值的大偏差极不可能发生。分母 $\sum(b_i - a_i)^2$ 刻画了随机变量的总变异性：各个变量的界 $[a_i, b_i]$ 越宽，均值附近的集中就越弱（即界越松）。

证明 [定理1。73] 我们使用切尔诺夫方法。对于任意 $\theta > 0$，根据马尔可夫不等式（定理 1.68）以及 x_i的独立性：

$$\Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \geq \tau) = \Pr(e^{\theta(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n])} \geq \exp(\theta\tau)) \leq \frac{\mathbb{E}[e^{\theta(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n])}]}{\exp(\theta\tau)} = \exp(-\theta\tau) \prod_{i=1}^n \mathbb{E}[e^{\theta(\mathrm{x}_i - \mathbb{E}[\mathrm{x}_i])}].$$

接下来，我们对 $\mathrm{x}_i - \mathbb{E}[\mathrm{x}_i]$ 的矩母函数分别进行上界估计。令 $\mathrm{y}_i \triangleq \mathrm{x}_i - \mathbb{E}[\mathrm{x}_i]$。则 $\mathrm{y}_i \in [a_i - \mathbb{E}[\mathrm{x}_i], b_i - \mathbb{E}[\mathrm{x}_i]] \triangleq [c_i, d_i]$，其中 $c_i \triangleq a_i - \mathbb{E}[\mathrm{x}_i]$，$d_i \triangleq b_i - \mathbb{E}[\mathrm{x}_i]$，$d_i - c_i = b_i - a_i$，以及 $\mathbb{E}[\mathrm{y}_i] = 0$。对每个 y 应用霍夫丁引理（引理 1.72）：

$$\mathbb{E}\big[\exp(\theta(\mathrm{x}_i - \mathbb{E}[\mathrm{x}_i]))\big] \leq \exp\left(\frac{\theta^2(b_i - a_i)^2}{8}\right),$$

由此可得

$$\Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \geq \tau) \leq \exp(-\theta\tau) \prod_{i=1}^n \mathbb{E}\big[\exp(\theta(\mathrm{x}_i - \mathbb{E}[\mathrm{x}_i]))\big] \leq \exp\left(-\theta\tau + \frac{\theta^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right).$$

在上述推导中取最优的 $\theta = \frac{4\tau}{\sum_{i=1}^n (b_i - a_i)^2}$ 即可得到上尾界。对 $-\mathrm{x}_i$应用相同的论证，可类似地得到下尾不等式。最后，由并集界可得双侧上界：

$$\Pr(|\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n]| \geq \tau) \leq \Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \geq \tau) + \Pr(\mathrm{s}_n - \mathbb{E}[\mathrm{s}_n] \leq -\tau) \leq 2\exp\left(-\frac{2\tau^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

证明至此完成。 ■

习题 1.74. 设 $\mathbb{E}[\mathrm{x}_i]=0$ 对于所有 $i\in\{1,2,\ldots,n\}$ 定理 1.73中。使用克拉默尔定理（定理 1.71）完成另一种证明。

若变量 $\mathbf{x}_i\in\{0,1\}$ 是伯努利随机变量，则 $a_i=0$，$b_i=1$，因此$b_i-a_i=1$，且 $\sum_{i=1}^{n}(b_i-a_i)^2=n$。令 $\mathbf{x}\triangleq\frac{1}{n}\sum\mathbf{x}_i$ 且 $\mu\triangleq\mathbb{E}[\mathbf{x}]$。那么，霍夫丁不等式给出

$$\Pr(|\bar{\mathrm{x}}-\mu|\geq\varepsilon)\leq 2\exp(-2n\varepsilon^2).$$

这种形式在学习理论中应用广泛，例如用于证明经验风险最小化的一致收敛性。

拉德马赫变量（也称为对称伯努利变量）是伯努利随机变量的一种变体：它是一个随机变量 ϵ，以相同的概率取值 $+1$ 和 -1。拉德马赫序列 $\boldsymbol{\epsilon}$ 是由独立拉德马赫随机变量组成的向量。对于这样的序列，我们得到以下形式的霍夫丁不等式。

推论 1.75：（Rademacher 序列的霍夫丁不等式）设 $\boldsymbol{a}\in\mathbb{R}^n$和 $\boldsymbol{\epsilon}=\{\epsilon_1,\epsilon_2,\ldots,\epsilon_n\}$是一个 Rademacher 序列。那么，对于 $\mu>0$,

$$\Pr\left(\left|\sum_{i=1}^{n}\epsilon_i a_i\right|\geq\|\boldsymbol{a}\|_2\,\mu\right)\leq 2\exp(-\frac{\mu^2}{2}).$$

证明 [推论 1.75] 随机变量 $\epsilon_i a_i$ 的均值为零，且其绝对值不超过 $|a_i|$。因此，所述不等式可直接由霍夫丁不等式推出。

■

伯恩斯坦不等式 为霍夫丁不等式提供了一个实用的推广，可用于有界甚至无界的独立随机变量之和，同时还纳入了有关其方差（或高阶矩）的信息。我们先介绍下述版本，然后推导出若干有用的推论形式。

定理 1.76：（伯恩斯坦不等式（Bernstein，1924）设 $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$ be 独立零均值随机变量（即对于所有 i，$\mathbb{E}[\mathbf{x}_i]=0$），且对于所有整数$k\geq 2$,

$$\mathbb{E}[|\mathrm{x}_i|^k]\leq k!C^{k-2}\sigma_i^2/2,\quad\text{for all } i\in\{1,2,\ldots,n\}$$

对于某些常数 $C>0$和 $\sigma_i>0$, $i\in\{1,2,\ldots,n\}$。那么，对于所有 $\tau>0$,

$$\Pr\left(\left|\sum_{i=1}^{n}\mathrm{x}_i\right|\geq\tau\right)\leq 2\exp\left(-\frac{\tau^2/2}{\sigma^2+C\tau}\right),$$

其中 $\sigma^2\triangleq\sum_{i=1}^{n}\sigma_i^2$。

证明 [定理 1。76] 利用众所周知的泰勒展开 $\exp(z)=\sum_{i=0}^{\infty}\frac{z^i}{i}!$ 对所有$z\in\mathbb{R}$ 在 0附近成立这一事实，以及 $\mathbb{E}[\mathbf{x}_i]=0$ 这一事实，我们得到

$$\mathbb{E}[\exp(\theta\mathrm{x}_i)]=1+\theta\mathbb{E}[\mathrm{x}_i]+\sum_{k=2}^{\infty}\frac{\theta^k\mathbb{E}[\mathrm{x}_i^k]}{k!}=1+\frac{\theta^2\sigma_i^2}{2}\sum_{k=2}^{\infty}\frac{\theta^{k-2}\mathbb{E}[\mathrm{x}_i^k]}{k!\sigma_i^2/2},$$

对于每个 $i\in\{1,2,\ldots,n\}$，定义 $F_i(\theta)\triangleq\sum_{k=2}^{\infty}\theta^{k-2}\mathbb{E}[\mathrm{x}^k\ \ \frac{k]}{i}!\ \sigma_i^2/2$。那么，

$$\mathbb{E}[\exp(\theta\mathrm{x}_i)]=1+\frac{\theta^2\sigma_i^2F_i(\theta)}{2}\leq\exp(\frac{\theta^2\sigma_i^2F_i(\theta)}{2}),$$

这里用到了不等式 $1+x \le \exp(x)$，对所有 $x \in \mathbb{R}$都成立。现在令 $F(\theta) \triangleq \max_{i\in\{1,2,\dots,n\}} F_i(\theta$。那么，根据 $\sigma^2 = \sum_{i=1}^n \sigma_i^2$和 Cramér 定理（定理 1.71），我们得到

$$\Pr\left(\sum_{i=1}^n \mathrm{x}_i \ge \tau\right) \le \min_{\theta>0} \exp\left(-\theta\tau + \frac{\theta^2\sigma^2 F(\theta)}{2}\right) \le \min_{0<C\theta<1} \exp\left(-\theta\tau + \frac{\theta^2\sigma^2 F(\theta)}{2}\right),$$

其中我们假设 $C\theta < 1$。结合不等式 $\mathbb{E}[\mathrm{x}_i^k] \le \mathbb{E}[|\mathrm{x}_i|^k]$ 以及关于 $\mathbb{E}[|\mathrm{x}_i|^k]$的上界假设，可得

$$F_i(\theta) \le \sum_{k=2}^{\infty} \frac{\theta^{k-2}\mathbb{E}[|\mathrm{x}_i|^k]}{k!\sigma_i^2/2} \le \sum_{k=2}^{\infty} \frac{(C\theta)^{k-2}}{1-C\theta} = \frac{1}{1-C\theta}.$$

因此，$F(\theta) \le (1-C\theta)^{-1}$，从而

$$\Pr\left(\sum_{i=1}^n \mathrm{x}_i \ge \tau\right) \le \min_{0<C\theta<1} \exp\left(-\theta\tau + \frac{\theta^2\sigma^2}{2(1-C\theta)}\right).$$

选择 $\theta = \tau/(\sigma^2 + C\tau)$ 满足 $C\theta < 1$，由此可得

$$\Pr\left(\sum_{i=1}^n \mathrm{x}_i \ge \tau\right) \le \exp\left(-\frac{\tau^2/2}{\sigma^2 + C\tau}\right).$$

将相同的论证应用于 $-\mathrm{x}_i$，即可得到 $\Pr\left(\sum_{i=1}^n \mathrm{x}_i \le -\tau\right)$ 的相同上界。通过并集界合并两侧尾部，即完成证明。 ■

我们给出伯恩斯坦不等式的两个重要推论。第一个是有界随机变量的伯恩斯坦不等式，这种形式在文献中更为常见。

推论1.77：（有界随机变量的伯恩斯坦不等式）设 $\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n$ 为相互独立的零均值随机变量，且对于 $i \in \{1,2,\dots,n\}$，$|\mathrm{x}_i| \le B$ 几乎必然成立，其中 $B > 0$ 为常数。进一步假设，对于某些常数 $\sigma_i > 0$、$i \in \{1,2,\dots,n\}$，有 $\mathbb{E}[\mathrm{x}_i^2] \le \sigma_i^2$。那么，对于所有 $\tau > 0$，

$$\Pr\left(\left|\sum_{i=1}^n \mathrm{x}_i\right| \ge \tau\right) \le 2\exp\left(-\frac{\tau^2/2}{\sigma^2 + B\tau/3}\right),$$

其中 $\sigma^2 \triangleq \sum_{i=1}^n \sigma_i^2$。

证明 [推论1.77] 对于 $k = 2$，根据假设，定理1.76 中的矩条件成立。因此，令 $k \in \mathbb{N}$，$k \ge 3$。由于于是 $k! \ge 3 \cdot 2^{k-2}$，我们得到

$$\mathbb{E}[|\mathrm{x}_i|^k] = \mathbb{E}[|\mathrm{x}_i|^{k-2}\,\mathrm{x}_i^2] \le B^{k-2}\sigma_i^2 \le \frac{k!B^{k-2}}{k!}\sigma_i^2 \le \frac{k!B^{k-2}}{2\cdot 3^{k-2}}\sigma_i^2.$$

因此，对于所有1.76 ，定理 $k \ge 2$ 在 $C = B/3$ 且具有相同的 σ_i时，其矩条件均成立。所需不等式直接由定理1.76得出。 ■

作为第二个推论，我们给出次指数随机变量的伯恩斯坦不等式（详见定义7.1 ）。

推论 1.78：（次指数随机变量的伯恩斯坦不等式）设 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ 为独立的零均值次指数随机变量；也就是说，对于某些常数 Φ，$\xi > 0$，对所有 $\tau > 0$, $i \in \{1, 2, \ldots, n\}$，均有 $\Pr(|\mathrm{x}_i| \geq \tau) \leq \Phi \exp(-\xi\tau)$。那么，对所有 $\tau > 0$,

$$\Pr\left(\left|\sum_{i=1}^{n} \mathrm{x}_i\right| \geq \tau\right) \leq 2\exp\left(-\frac{(\xi\tau)^2/2}{2\Phi n + \xi\tau}\right).$$

证明 [推论 1.78] 利用绝对矩引理（引理 1.65），对于 $k \in \mathbb{N}, k \geq 2$,

$$\mathbb{E}[|\mathrm{x}_i|^k] = k\int_0^\infty \Pr(|\mathrm{x}_i| \geq x)x^{k-1}dx \leq \Phi k \int_0^\infty \exp(-\xi x)x^{k-1}dx$$
$$\underline{\underline{u=\xi x}}\ \Phi k\xi^{-k}\int_0^\infty e^{-u}u^{k-1}du = \Phi k!\xi^{-k} = k!\xi^{-(k-2)}\frac{2\Phi\xi^{-2}}{2},$$

其中我们用到了伽马函数 $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)dt$ 以及以下事实：$\Gamma(k) = (k-1)!$ 对任意整数 k均成立。因此，定理 1.76 中的矩条件成立，其中$C = \xi^{-1}$ 且 $\sigma_i^2 = 2\Phi\xi^{-2}$。再次由定理 1.76可得结论。 ■

回顾一下，伽马函数 对 $y \geq 0$ 定义如下：

$$\Gamma(y) = \int_0^\infty t^{y-1}e^{-t}dt, \qquad y \geq 0. \tag{1.48}$$

利用分部积分 $\int_a^b u(t)v'(t)dt = u(t)v(t)|_a^b - \int_a^b u'(t)v(t)dt$，其中 $u(t) = t^{y-1}$且 $v(t) = -e^{-t}$，我们推导得

$$\Gamma(y) = -t^{y-1}e^{-t}|_0^\infty - \int_0^\infty (y-1)t^{y-2}(-e^{-t})dt$$
$$= 0 + (y-1)\int_0^\infty t^{y-2}e^{-t}dt = (y-1)\Gamma(y-1).$$

这个递推关系表明，当 y 是正整数时，$\Gamma(y) = (y-1)!$。

1.3.2 高斯分布及其性质

高斯分布（或正态）随机变量是统计建模和信号处理中的一种广泛首选。之所以如此，是因为一项基本原理：独立随机变量之和趋于服从高斯分布——这一结果由中心极限定理（见下文定理1.80 ）形式化地阐述。因此，由许多互不相关的独立分量相加形成的任何测量值或量，都很可能近似服从高斯分布。正因如此，在信号处理和工程等领域，噪声通常采用高斯分布建模，因为它具有理想的数学性质。

高斯随机变量

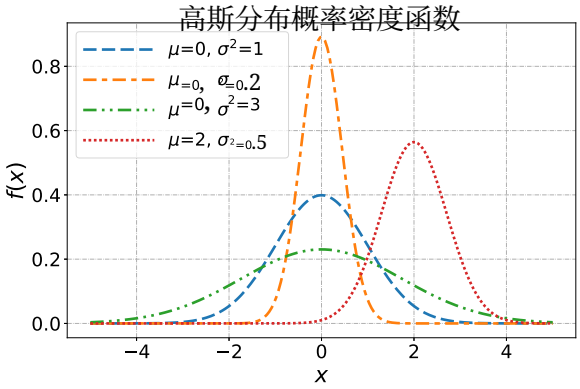


图 1.5： 均值和方差参数 μ 和 σ^2取不同值时的高斯概率密度函数。

定义 1.79（高斯分布或正态分布）。若随机变量 x 服从高斯分布（又称正态分布），其均值为 μ，方差为 $\sigma^2 > 0$，记作 $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$ [a]，当且仅当

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} = \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}(x-\mu)^2\right\},$$

其中 τ 也称为该分布的精度。$\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$ 的均值、方差和矩母函数为

$$\mathbb{E}[\mathrm{x}] = \mu, \qquad \mathbb{V}\mathrm{ar}[\mathrm{x}] = \sigma^2 = \tau^{-1}, \qquad \mathbb{M}_{\mathrm{x}}(\theta) = \exp\{\theta\mu + \theta^2\sigma^2/2\},$$

x 的累积分布函数（c.d.f.，CDF）为

$$F(x;\mu,\sigma^2) = \Pr(\mathrm{x} < x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2\sigma^2}(z-\mu)^2\right\} dz.$$

特别地，我们将 $\Phi(y) = \int_{-\infty}^{y} \mathcal{N}(u \mid 0,1)du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-\frac{u^2}{2})du$ 记为 $\mathcal{N}(0,1)$ 的累积分布函数，即标准正态分布。

[a] 注意，如果两个随机变量 a 和 b 具有相同分布，则记作 $\mathrm{a} \sim \mathrm{b}$。

图1.5 展示了参数的不同取值 μ, σ^2 如何影响高斯分布的形状。

如上所述，中心极限定理指出，大量具有有限方差的随机变量的平均值的分布会收敛于一个高斯分布。

定理 1.80：（中心极限定理（CLT））设 $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3$，.....为一个均值为 μ 且方差 $\sigma^2 < \infty$ 有限的独立同分布随机变量序列。我们将平均值序列 $\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3$，.....定义为

$$\mathrm{y}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i.$$

那么，标准化序列 $\mathrm{z}_1, \mathrm{z}_2, \mathrm{z}_3$，.....

$$\mathrm{z}_n \triangleq \sqrt{n}(\mathrm{y}_n - \mu)$$

依分布收敛于均值为 0、方差为 σ^2的高斯随机变量。也就是说，对于任意 $x \in \mathbb{R}$,

$$\lim_{n\to\infty} f_{z_n}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

当 n 足够大时，上述定理还意味着样本均值 y_n 近似服从高斯分布，其均值为 μ，方差为 σ^2/n。

两个高斯变量的乘积在解析处理上仍颇具挑战，而高斯变量之和本身仍服从高斯分布，并可由其均值和方差完全刻画。

注 1.81（高斯变量之和）。设 $\mathrm{x} \sim \mathcal{N}\mu_x, \sigma_x^2$ 和 $\mathrm{y} \sim \mathcal{N}\mu_y, \sigma_y^2$ 是两个高斯随机变量。

- 如果 x 和 y 不相关，那么
$$\mathrm{x} + \mathrm{y} \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$
- 如果 x 和 y 的相关系数为 ρ，那么
$$\mathrm{x} + \mathrm{y} \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y).$$

高斯随机变量的一项重要性质是，仿射变换（即缩放和平移）能够保持高斯形式。

引理 1.82：（高斯变量的线性变换）设 x 是一个均值为 μ、方差为 σ^2 的高斯随机变量。那么，对于任意 a，$b \in \mathbb{R}$，随机变量

$$\mathrm{y} \triangleq a\mathrm{x} + b \tag{1.49}$$

也服从高斯分布，其均值为 $a\mu + b$，方差为 $a^2\sigma^2$。

证明可由线性变换引理（引理1.63 ）.

高斯分布具有很强的集中性质。在此背景下，一个基本结果是关于中心高斯随机变量的如下尾界。

练习 1.83（中心高斯分布的 Chernoff 界）。对于 $\mathrm{x} \sim \mathcal{N}(0, \sigma^2)$，证明

$$\Pr(|\mathrm{x}| \geq \tau) \leq 2e^{-\frac{\tau^2}{2\sigma^2}}, \quad \text{for all } \tau > 0.$$

提示：应用 *Chernoff* 界（定理*1.70*）

尽管 Chernoff 界为中心高斯分布的 $\Pr(|\mathrm{x}| \geq \tau)$ 提供了有效的上界，但该上界并不紧。可以如下推导出一个更精确的上界。

引理 1.84：（中心高斯分布的更紧尾界）对于 $\mathrm{x} \sim \mathcal{N}(0, \sigma^2)$，有

$$\Pr(|\mathrm{x}| \geq \tau) \leq \begin{cases} e^{-\frac{\tau^2}{2\sigma^2}}; \\ \frac{\sigma\sqrt{2}}{\tau\sqrt{\pi}} \exp\left(-\frac{\tau^2}{2\sigma^2}\right), \end{cases} \quad \text{for all } \tau > 0.$$

请注意，后一个不等式称为*Mill* 不等式。

证明 [引理 1。84] 根据定义，$\Pr(|\mathrm{x}| \geq \tau) = \frac{2}{\sqrt{2\pi\sigma^2}} \int_\tau^\infty e^{-x^2/2\sigma^2} dx$。应用变量替换：

$$\begin{aligned}\int_\tau^\infty e^{-x^2/2\sigma^2} dx &= \int_0^\infty e^{-(x+\tau)^2/2\sigma^2} dx = e^{-\tau^2/2\sigma^2} \int_0^\infty e^{-x\tau/\sigma^2} e^{-x^2/2\sigma^2} dx \\ &\leq e^{-\tau^2/2\sigma^2} \int_0^\infty e^{-x^2/2\sigma^2} dx = \sqrt{\frac{\pi\sigma^2}{2}} e^{-\tau^2/2\sigma^2},\end{aligned}$$

其中，不等式由 $e^{-x\tau/\sigma^2} \leq 1$ for $x\ \tau \geq 0$ $|\ |\geq$, 得出。这表明 $\Pr(\mathrm{x})$。$\tau \leq e^{-\tau^2/2\sigma^2}$ 对于第二部分，根据对称性和并集界，$\Pr(|\mathrm{x}| \geq \tau) \leq 2\Pr(\mathrm{x} \geq \tau)$:

$$\begin{aligned}2\Pr(\mathrm{x} \geq \tau) &= \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \int_\tau^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \leq \frac{\sigma^2\sqrt{2}}{\sqrt{\pi}\sigma^2} \int_\tau^{+\infty} \frac{x}{\sigma^2\tau} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \frac{\sigma\sqrt{2}}{\tau\sqrt{\pi}} \int_\tau^{+\infty} -\frac{\partial}{\partial x} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{\sigma\sqrt{2}}{\tau\sqrt{\pi}} \exp\left(-\frac{\tau^2}{2\sigma^2}\right).\end{aligned}$$

证明完毕。 ■

根据标准高斯随机变量 $\mathrm{x} \sim \mathcal{N}(0,1)$ 的定义可知，$\mathbb{M}_\mathrm{x}(\theta) = \mathbb{E}[\exp(\theta\mathrm{x})] = \exp(\theta^2/2)$。对于涉及 x 的二次函数的期望，也有更一般的结果。

引理 1.85：（标准高斯变量的矩）设 $\mathrm{x} \sim \mathcal{N}(0,1)$。那么，对于 $a,\ b,\ c \in \mathbb{R}$ 且满足 $a < 1/2$,

$$\mathbb{E}[\exp(a\mathrm{x}^2 + b\mathrm{x} + c)] = \frac{1}{\sqrt{1-2a}} \exp\left(\frac{b^2}{2(1-2a)} + c\right).$$

证明 [引理 1。85] 根据标准正态密度下期望的定义，

$$\mathbb{E}[\exp(a\mathrm{x}^2 + b\mathrm{x} + c)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(ax^2 + bx + c) \exp\left(-\frac{x^2}{2}\right) dx.$$

注意到单位矩阵

$$ax^2 - \frac{x^2}{2} + bx = -\frac{1-2a}{2}\left(x - \frac{b}{1-2a}\right)^2 + \frac{b^2}{2(1-2a)} + c.$$

经过变量替换，后一个积分化简为方差为 $1/(1-2a)$的高斯概率密度函数的积分，因此等于一。证明完毕。 ■

注意，上述引理给出了标准高斯分布变量 x 的高阶矩的一个上界，因为

$$\mathbb{E}[\exp(\theta\mathrm{x}^2)] = \mathbb{E}\left[\sum_{i=0}^{\infty} \frac{(\theta\mathrm{x}^2)^i}{i!}\right] = \sum_{i=0}^{\infty} \frac{\mathbb{E}[(\theta\mathrm{x}^2)^i]}{i!},$$

其中，第一个等式来自众所周知的泰勒展开 $\exp(z) = \sum_{i=0}^{\infty} \frac{z^i}{i}!$ 对于所有 $z \in \mathbb{R}$ 在 0 附近成立。利用高阶矩的性质，通过指数函数的期望来控制尾概率的界，通常称为 Chernoff 界或 Chernoff 方法；另请参见定理1.70。

高斯随机向量

多元高斯分布（也称为多元正态分布，或简称高斯分布）是一种连续概率分布，用于描述一组联合服从正态分布的随机变量。它完全由其均值向量和协方差矩阵确定。协方差矩阵刻画了所有变量之间的两两协方差，从而编码了各变量的方差（位于对角线上）及其线性依赖关系（位于非对角元素处）。由于具有良好的数学可处理性和灵活性，多元高斯分布广泛应用于机器学习、统计学和信号处理等领域。现在我们给出多元高斯分布的正式定义。

定义1.86（多元高斯分布）。随机向量$\mathbf{x} \in \mathbb{R}^p$被称为服从参数为 $\boldsymbol{\mu} \in \mathbb{R}^p$ 和 $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ 的多元高斯分布（多元正态分布，*MVN*），记作$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，如果

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \ ^{a}$$

其中，$\boldsymbol{\mu} \in \mathbb{R}^p$ 称为均值向量，而 $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ 是正定的，称为协方差矩阵。$|\boldsymbol{\Sigma}| = \det(\boldsymbol{\Sigma})$表示 $\boldsymbol{\Sigma}$的行列式。多元高斯分布的均值、众数和协方差分别为

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}, \qquad \text{Mode}[\mathbf{x}] = \boldsymbol{\mu}, \qquad \text{and} \qquad \mathbb{C}\text{ov}[\mathbf{x}] = \boldsymbol{\Sigma}.$$

协方差矩阵也可以表示为：

$$\mathbb{C}\text{ov}[\mathbf{x}] = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top] = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top.$$

[a] 该形式可以利用 p 独立同分布的一元标准高斯随机变量的矩母函数加以证明。

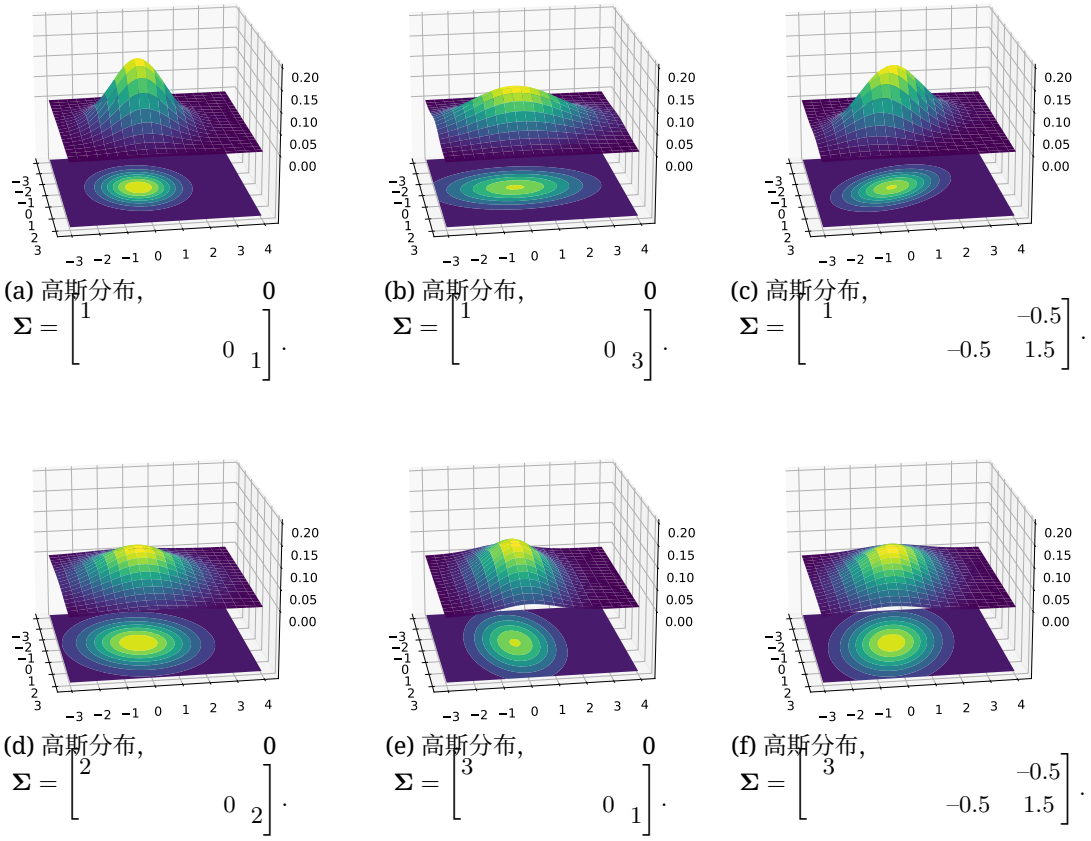
图 1.6展示了不同协方差结构下的高斯分布密度图。当高斯向量的协方差矩阵为单位矩阵且均值为零时，其元素是均值为零、方差为单位方差的独立同分布标准高斯分布。我们将这样的向量称为独立同分布的标准高斯向量。此外，多元高斯随机向量可以由一元高斯样本生成；参见习题1.7。

与多元高斯分布相关的两个重要分布是伽马分布及其特例卡方分布。下面给出这两个分布的定义。

定义 1.87（伽马分布）。若随机变量 $\mathbf{x}$ 服从伽马分布，其形状参数为 $r > 0$，率参数为 $\lambda > 0$，记作 $\mathbf{x} \sim \mathcal{G}(r, \lambda)$，其概率密度函数为

$$f(x; r, \lambda) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} \exp(-\lambda x), & \text{if } x \geq 0. \\ 0, & \text{if } x < 0, \end{cases}$$

其中 $\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt$ 是伽马函数（见（1.48）），它将分布归一化，使其积分为 1。$\mathbf{x} \sim \mathcal{G}(r, \lambda)$ 的均值和方差为



(a) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
(b) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$.
(c) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$.
(d) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.
(e) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$.
(f) 高斯分布，$\boldsymbol{\Sigma} = \begin{bmatrix} 3 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$.

图1.6： 多元高斯分布的密度和等高线图（蓝色=低，黄色=高），展示了在 $\mathbb{R}^2$ 空间中，零均值向量下不同协方差／尺度矩阵对应的情形。图 1.6(a) 和 1.6(d)：球形协方差矩阵呈圆形；图 1.6(b) 和 1.6(e)：对角协方差矩阵产生轴对齐椭圆等高线；图 1.6(c) 和 1.6(f)：满协方差矩阵产生旋转椭圆等高线。

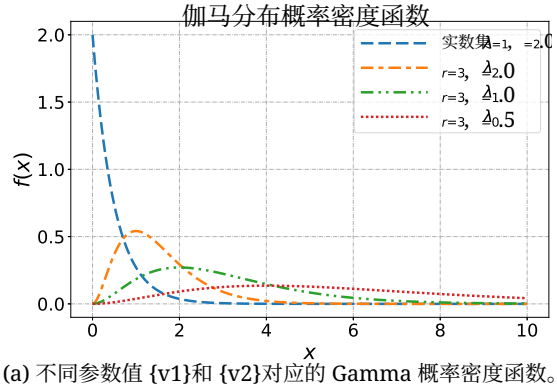
给定为

$$\mathbb{E}[\mathrm{x}] = \frac{r}{\lambda}, \qquad \mathbb{V}\mathrm{ar}[\mathrm{x}] = \frac{r}{\lambda^2}.$$

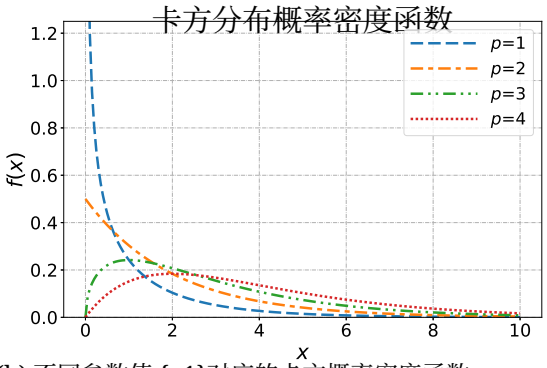
图1.7(a)展示了伽马分布的不同参数。值得注意的是，伽马分布中的形状参数 r 不受自然数的限制，而是可以取任意正值。不过，当 r 为正整数时，伽马分布可以解释为 r 个独立指数变量之和，每个变量的速率为 λ（Lu，2023）。这一求和性质更普遍地适用于具有相同率参数的伽马分布变量。若 x_1 和 x_2 分别是从 $\mathcal{G}(r_1, \lambda)$ 和 $\mathcal{G}(r_2, \lambda)$ 中抽取的随机变量，那么它们的和 $\mathrm{x}_1 + \mathrm{x}_2$ 服从伽马分布 $\mathcal{G}(r_1 + r_2, \lambda)$。

虽然卡方分布是伽马分布的一种特殊情况，但它在统计理论中发挥着尤为重要的作用。

定义 1.88（卡方分布，χ^2-分布）。 令 $\mathrm{a}_i \sim \mathcal{N}(0,\ 1)$，其中 $i \in \{1, 2, \ldots, p\}$（等价地，$\mathbf{a} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{I}_p)$；见定义 1.86）。则随机



(a) 不同参数值 {v1}和 {v2}对应的 Gamma 概率密度函数。
不同参数值 r 和 λ对应的 Gamma 概率密度函数。



(b) 不同参数值 {v1}对应的卡方概率密度函数。
不同参数值 p对应的卡方概率密度函数。

图 1.7： 伽马分布与卡方分布的比较。

变量 $\mathrm{x} = \sum_{i=1}^{p} \mathrm{a}_i^2$ 服从卡方分布（或卡方分布，χ^2-分布），其 p 自由度为 $\sim \chi^2_{(p)}$，这等价于 $\mathrm{x} \sim \mathcal{G}(p/2,\ 1/2)$ 在定义 1.87中所述的情形。x 的概率密度函数为

$$f(x;p) = \begin{cases} \dfrac{1}{2^{p/2}\Gamma(\frac{p}{2})} x^{\frac{p}{2}-1} \exp(-\dfrac{x}{2}), & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

$\mathrm{x} \sim \chi^2_{(p)}$ 的均值和方差为

$$\mathbb{E}[\mathrm{x}] = p, \qquad \mathbb{V}\mathrm{ar}[\mathrm{x}] = 2p.$$

图 1.7(b) 展示了自由度 p 变化对卡方分布形状的影响。由定义可知，如果 $\mathbf{a} = [\mathrm{a}_1, \mathrm{a}2, \ldots, \mathrm{a}_p]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$（多元高斯分布；参见定义 1.86），那么 $\mathrm{x} = \mathbf{a}^\top \mathbf{a} \sim \chi^2_{(p)}$。更一般地，设 $\boldsymbol{H}$ 为秩为 $r < p$ 的正交投影矩阵（参见定义 1.42）。于是可得

$$\mathbf{a}^\top \boldsymbol{H} \mathbf{a} \sim \chi^2_{(r)}, \text{ with orthogonal projector } \mathrm{rank}(\boldsymbol{H}) = r < p. \tag{1.50a}$$

F此外，假设 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中 $\boldsymbol{\mu} \in \mathbb{R}^p$ 且 $\boldsymbol{\Sigma}^{-1}$ 为非奇异。 n,

$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2_{(p)}. \tag{1.50b}$$

要看出这一点，注意到由于 $\boldsymbol{\Sigma}$是对称正定的，因此存在谱分解 $\boldsymbol{\Sigma} = \boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 是一个 $p \times p$ 正交矩阵，而 $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$，且 $\lambda_j > 0$（参见定理 1.46）。定义标准化向量 $\mathbf{y} \triangleq \boldsymbol{\Lambda}^{-1/2} \boldsymbol{Q}(\mathbf{x} - \boldsymbol{\mu})$。于是向量$\mathbf{y}$ 服从正态分布（事后看来，参见引理1.90）：

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^{-1/2} \boldsymbol{Q} \boldsymbol{\Sigma} \boldsymbol{Q}^\top \boldsymbol{\Lambda}^{-1/2}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^{-1/2} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{\Lambda}^{-1/2}) = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p).$$

由此可知，$\mathbf{y}$ 的各个分量相互独立（参见引理1.92），且 $\mathbf{y}_i \sim \mathcal{N}(0,\ \mathbf{1})$。因此，$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}^\top \mathbf{y} \sim \chi^2_{(p)}$。

我们还可能关注$\mathbf{x}^\top \boldsymbol{A} \mathbf{x}$ 的二次型，其中 $\boldsymbol{A}$ 为对称矩阵。下面的定理给出了此类二次型服从卡方分布的关键刻画。

命题 1.89：（高斯随机变量的二次型）对于高斯分布的二次型，我们有以下结果：

- 给定$\mathbf{x} \sim \mathcal{N}(\mathbf{0},\ \lambda \boldsymbol{I})$（长度为 p）以及对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{p\times p}$。则可得

$$\frac{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}{\lambda} \sim \chi^2_{(r)},$$

  当且仅当 $\boldsymbol{A}$ 是秩为 $r < p$的幂等矩阵（$\boldsymbol{A}^2 = \boldsymbol{A}$）时。
- 给定$\mathbf{x} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}) \in \mathbb{R}^p$ 以及对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{p\times p}$。则可得

$$\mathbf{x}^\top \boldsymbol{A}\mathbf{x} \sim \chi^2_{(r)},$$

  当且仅当 $\boldsymbol{A\Sigma}$ 是秩为 $r < p$的幂等矩阵时。

多元高斯分布的性质

高斯随机向量的一个基本性质是仿射变换保持高斯性。这是引理 1.82 的多维推广，可以利用概率分布的线性变换法则（引理1.63）或通过多元高斯分布的矩生成函数推导得到。

引理 1.90：（多元高斯分布的仿射变换）给定固定矩阵和向量，$\boldsymbol{A},\ \boldsymbol{B} \in \mathbb{R}^{p\times d}$和 $\boldsymbol{c} \in \mathbb{R}^p$，设 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x,\ \boldsymbol{\Sigma}_x)$ 与$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y,\ \boldsymbol{\Sigma}_y)$ 为独立随机向量（长度为 d）。则，

$$\mathbf{z} = \boldsymbol{A}\mathbf{x} + \boldsymbol{B}\mathbf{y} + \boldsymbol{c} \sim \mathcal{N}(\boldsymbol{A\mu}_x + \boldsymbol{B\mu}_y + \boldsymbol{c}, \boldsymbol{A\Sigma}_x\boldsymbol{A}^\top + \boldsymbol{B\Sigma}_y\boldsymbol{B}^\top).$$

此外，对任意固定向量 $\boldsymbol{d} \in \mathbb{R}^d$，线性组合 $\boldsymbol{d}^\top\mathbf{x}$ 服从单变量高斯分布：

$$\boldsymbol{d}^\top\mathbf{x} \sim \mathcal{N}(\boldsymbol{d}^\top\boldsymbol{\mu}_x, \boldsymbol{d}^\top\boldsymbol{\Sigma}_x\boldsymbol{d}).$$

该结果还依赖于独立高斯分布之和：

$$\sum_{i=1}^n \mathbf{x}_i \sim \mathcal{N}\Big(\sum_{i=1}^n \boldsymbol{\mu}_i, \sum_{i=1}^n \boldsymbol{\Sigma}_i\Big) \quad \text{if } \mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),\ \forall i \in \{1, 2, \ldots, n\}.$$

作为特例，设 $\boldsymbol{A} = \boldsymbol{e}_i^\top$为 $\mathbb{R}^d$ 中的第 i 个标准基向量。则 $\mathbf{x}_i = \boldsymbol{e}_i^\top\mathbf{x}$ $\sim\mathcal{N}(\mu_{x,\ i},\ \sigma^2_{x,\ ii})$，其中 $\mu_{x,\ i}$ 表示 $\boldsymbol{\mu}_x$ 的第 i 个分量，$\sigma^2_{x,\ ii}$ 表示 $\boldsymbol{\Sigma}_x$ 的第 i 个对角元。

引理 1.90 的一个直接推论是，多元高斯向量的任意子向量本身也是高斯向量。引理1.90的另一个推论是，独立同分布标准高斯向量是各向同性的，这意味着其分布在旋转下保持不变。形式上，对于任意正交矩阵 $\boldsymbol{Q}$，若 $\mathbf{x}$ 是独立同分布标准高斯向量，则根据引理1.90，$\boldsymbol{Q}\mathbf{x}$ 具有相同分布，因为其均值为 $\boldsymbol{Q}\mathbf{0} = \mathbf{0}$，协方差矩阵为 $\boldsymbol{QIQ}^\top = \boldsymbol{QQ}^\top = \boldsymbol{I}$。这种旋转不变性比仅仅在所有方向上具有相同方差更强（任何分量不相关且同分布的向量都满足这一点）。各向同性意味着整个分布——而不仅仅是二阶矩——在正交变换下保持不变。

引理 1.91：（多元高斯分布的旋转）对高斯分布进行旋转不会改变其分布。也就是说，对于任意正交矩阵$\boldsymbol{Q}$，若 $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$，则当 $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ 时，$\boldsymbol{Q}\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$。

▶ **“标准化与去相关。”**分布 $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ 称为独立同分布标准高斯向量。给定$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则对 $\mathbf{x}$ 进行去相关可得

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \Longrightarrow \quad \mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{1.51}$$

这也表明，若$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{1.52}$$

假设 $\{\boldsymbol{x}_i, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ 是来自 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 n 个随机样本，并令 $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$。于是可得

$$\sqrt{n}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \tag{1.53}$$

引理 1.92：（对于高斯随机向量，不相关意味着相互独立）如果高斯随机向量$\mathbf{x}$ 的所有分量都不相关，那么它们相互独立。

直接观察可知，当协方差矩阵为对角矩阵时（即各分量不相关时），$\mathbf{x}$ 的联合概率密度函数可分解为其边缘概率密度函数的乘积，因此该结果显然成立。

现在我们来讨论高斯向量的分块。设 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中 $\boldsymbol{\mu} \in \mathbb{R}^p$，并将 $\mathbf{x}$ 分为两个子向量：

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right).$$

当且仅当 $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ 时，$\mathbf{x}_1$ 与 $\mathbf{x}_2$ 独立。更一般地，设 $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_1, \ldots\ldots, \mathrm{x}_p]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。那么，

$$\text{the } \mathrm{x}_i\text{'s are mutually independent if and only if } \boldsymbol{\Sigma} \text{ is diagonal.} \tag{1.54}$$

下面给出严格证明：

证明 [引理1.92和方程 (1.54)] 假设这些 x_i彼此独立。那么 $\mathrm{x}_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ 对某个 $\sigma_i > 0$ 成立。因此，$\mathbf{x}$ 的密度为 $\mathbf{x}$ 是

$$\begin{aligned} p_{\mathbf{x}}(\boldsymbol{x}) &= \prod_{i=1}^{p} p_{\mathrm{x}_i}(x_i) = \prod_{i=1}^{p} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\frac{(x_i - \mu_i)^2}{\sigma_i^2} \right\} \\ &= \frac{1}{(2\pi)^{p/2}\left|\operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2)\right|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \operatorname{diag}(\sigma_1^{-2}, \ldots, \sigma_p^{-2})(\boldsymbol{x} - \boldsymbol{\mu}) \right\}. \end{aligned}$$

因此，$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \operatorname{diag}(\sigma_1^2, \ldots\ldots, \sigma_p^2))$，也就是说，协方差 $\boldsymbol{\Sigma}$ 是对角的。

反过来，假设 $\boldsymbol{\Sigma}$ 是对角的，记作 $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots\ldots, \sigma_p^2)$。那么，我们可以逆向进行第一部分中的推导，得出联合密度 $p_{\mathbf{x}}(\boldsymbol{x})$可以写成边缘密度 $p_{\mathrm{x}_i}(x_i)$ 的乘积，从而证明它们独立。■

上述结果立即推出以下推论。

推论1.93：（高斯分布中线性组合的独立性）设 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 是 $\mathbb{R}^p$ 中的一个随机向量，并设 $\boldsymbol{A} \in \mathbb{R}^{m\times p}$，$\boldsymbol{B} \in \mathbb{R}^{d\times p}$ 为固定矩阵。那么，

$$\boldsymbol{A}\mathbf{x} \text{ is independent of } \boldsymbol{B}\mathbf{x} \iff \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^\top = \mathbf{0}. \tag{1.55}$$

利用多元高斯分布的矩母函数的性质，可以得到另一种证明；此处省略细节。

根据基本概率论，可测函数作用于独立随机向量后仍然相互独立。也就是说，如果$\mathbf{x}_1$和$\mathbf{x}_2$ 相互独立，那么对于任意可测函数 $g_1(\mathbf{x}_1)$和 $g_2(\mathbf{x}_2)$，$g_1(\mathbf{x}_1)$和 $g_2(\mathbf{x}_2)$也相互独立。

作为一个重要示例，设变量x_i是独立同分布的 $\mathcal{N}(\mu, \sigma^2)$，其中 $i \in \{1, 2, \ldots, p\}$。那么，我们可以构造一个由$\bar{\mathrm{x}}$和$\mathrm{x}_i - \bar{\mathrm{x}}$组成的向量：

$$\begin{bmatrix} \bar{\mathrm{x}} \\ \mathrm{x}_1 - \bar{\mathrm{x}} \\ \vdots \\ \mathrm{x}_p - \bar{\mathrm{x}} \end{bmatrix} = \begin{bmatrix} \frac{1}{p} & \frac{1}{p} & \cdots & \frac{1}{p} \\ & \boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p & & \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_p \end{bmatrix} \quad \text{where} \quad \boldsymbol{J}_p = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & \ddots & & 1 \\ 1 & & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix} \in \mathbb{R}^{p\times p}.$$

由此可得

$$\begin{bmatrix} \bar{\mathrm{x}} \\ \mathrm{x}_1 - \bar{\mathrm{x}} \\ \vdots \\ \mathrm{x}_p - \bar{\mathrm{x}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{p} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p \end{bmatrix} \right).$$

因此，我们发现$\bar{\mathrm{x}}$与$\mathrm{x}_1 - \bar{\mathrm{x}}$，$\mathrm{x}_2 - \bar{\mathrm{x}}$.....，$\mathrm{x}_p - \bar{\mathrm{x}}$相互独立。在许多应用中，我们可以构造一个随机变量：

$$t \triangleq \sqrt{p}\frac{(\bar{\mathrm{x}} - \mu)/\sigma}{\sqrt{\frac{1}{p-1}\sum_i (x_i - \bar{\mathrm{x}})^2/\sigma^2}},$$

在这种情况下，分子和分母是独立变量。

第1章习题

1. 正定/半正定/负定矩阵的迹和行列式。设 $\boldsymbol{X}$ 为正定（分别为半正定）矩阵。证明 $\mathrm{tr}(\boldsymbol{X})$ 和 $\det(\boldsymbol{X})$ 均为正（分别为非负）。此外，证明 $\mathrm{tr}(\boldsymbol{X}) = 0$ 当且仅当 $\boldsymbol{X} = \mathbf{0}$。设 $\boldsymbol{Y} \in \mathbb{R}^{p\times p}$ 为负定矩阵。证明 $\mathrm{tr}(\boldsymbol{Y})$ 为负；当 p 为奇数时，$\det(\boldsymbol{Y})$ 为负；当 p为偶数时，$\det(\boldsymbol{Y})$ 为正。提示：使用定理 1.18。2. 证明注1.22和注1.23。3. 证明向量 ℓ_2范数、矩阵弗罗贝尼乌斯范数和矩阵谱范数满足定义 1.24 中列出的三个条件。4. 诱导范数性质（Lu，2021a）。考虑定义1.34中的诱导矩阵范数。

(a). 给定 $\boldsymbol{\alpha} \in \mathbb{R}^n$、$\boldsymbol{\beta} \in \mathbb{R}^p$ 和 a，$b \geq 1$，其中 $\frac{1}{a}+\frac{1}{a^\star}=1$ 且 $\frac{1}{b}+\frac{1}{b^\star}=1$，证明

$$\left\|\boldsymbol{\alpha}\boldsymbol{\beta}^\top\right\|_{a,b} = \|\boldsymbol{\alpha}\|_b \|\boldsymbol{\beta}\|_{a^\star}.$$

(b). 给定 $\boldsymbol{\alpha} \in \mathbb{R}^n$ 和 $\boldsymbol{\beta} \in \mathbb{R}^p$，且 $\|\boldsymbol{\beta}\|_a = \|\boldsymbol{\alpha}\|_b = 1$，证明存在一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，使得$\|\boldsymbol{X}\|_{a,\ b}=1$ 且 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{\alpha}$。(c). 特别地，证明$\left\|\boldsymbol{X}^\top\right\|_{a,\ b} = \|\boldsymbol{X}\|_{b^\star,\ a^\star}$。

5. **矩阵范数的等价性**。与习题 1.26中的向量范数类似，矩阵范数也满足等价性结论。设 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 是两个不同的矩阵范数：$\mathbb{R}^{n\times p} \to \mathbb{R}$。证明存在正标量 α 和 β，使得对所有 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$,

$$\alpha \|\boldsymbol{X}\|_a \leq \|\boldsymbol{X}\|_b \leq \beta \|\boldsymbol{X}\|_a.$$

等价定理指出，如果一个矩阵在某个范数下很小，那么它在其他范数下也很小，反之亦然。

6. **由诱导范数构造诱导范数**。设 $\|\cdot\|$ 是 $\mathbb{R}^{p\times p}$上的矩阵范数（分别地，次乘性矩阵范数），且 $\boldsymbol{S} \in \mathbb{R}^{p\times p}$ 非奇异。证明，对 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 定义的下列函数也是矩阵范数（分别地，次乘性矩阵范数）：

$$\|\boldsymbol{X}\|_{\boldsymbol{S}} = \left\|\boldsymbol{S}\boldsymbol{X}\boldsymbol{S}^{-1}\right\|.$$

此外，若 $\|\cdot\|$是由 $\mathbb{R}^p$上的向量范数 $\|\cdot\|_v$ 诱导的，证明矩阵范数 $\|\boldsymbol{X}\|_{\boldsymbol{S}}$ 是由 $\mathbb{R}^p$上的向量范数 $\|\boldsymbol{\beta}\|_{\boldsymbol{S}} = \|\boldsymbol{S}\boldsymbol{\beta}\|_v$ 诱导的。

7. 假设我们能够生成单变量高斯随机变量 $\mathcal{N}(0,\ 1)$。请给出一种生成多变量高斯随机变量 $\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$的方法，其中 $\boldsymbol{\Sigma}=\boldsymbol{C}\boldsymbol{C}^\top$、$\boldsymbol{\mu} \in \mathbb{R}^n$和$\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$。提示：若$\mathrm{x}_1, \mathrm{x}_2, \ldots\ldots, \mathrm{x}_n$是来自 $\mathcal{N}(0,\ 1)$的独立同分布变量，并令$\mathbf{x}=[\mathrm{x}_1, \mathrm{x}_2, \ldots\ldots, \mathrm{x}_n]^\top$，则可得 $\boldsymbol{C}\mathbf{x}+\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$。

8. 严谨地证明引理1.90 和引理1.91 。

9. **高斯分布的 KL 散度**。给定两个概率分布函数 $P(\boldsymbol{x})$ 和 $Q(\boldsymbol{x})$，我们用 $D_{\mathrm{KL}}[P \parallel Q] = \int P(\boldsymbol{x}) \ln\left(\frac{P}{Q}(\boldsymbol{x})(\boldsymbol{x})\right) d\boldsymbol{x} \geq 0$ 表示 P 与 Q之间的*Kullback–Leibler*（*KL*）散度，其中等号仅在 $P=Q$ 时成立。给定$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu},\ \mathrm{diag}(\boldsymbol{\sigma}^2))$ 和 $p(\boldsymbol{x}) = \mathcal{N}(\mathbf{0}_D,\ \boldsymbol{I}_D)$，其中 $\boldsymbol{\mu},\ \boldsymbol{\sigma}^2 \in \mathbb{R}^D$。证明 $D_{\mathrm{KL}}[q \parallel p] = \frac{1}{2}\sum_{i=1}^D(\mu_i^2+\sigma_i^2-\ln\sigma_i^2-1)$。

10. **高斯分布的 KL 散度**。设 $p(x) = \mathcal{N}(\mu_1,\ \sigma_1^2)$ 和 $q(x) = \mathcal{N}(\mu_2,\ \sigma_2^2)$。证明

$$D_{\mathrm{KL}}[p \parallel q] = \ln\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2+(\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

考虑多变量情形。设 $\mathcal{N}_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}_1,\ \boldsymbol{\Sigma}_1)$ 且 $\mathcal{N}_2(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_2)$。证明

$$D_{\mathrm{KL}}[\mathcal{N}_1 \parallel \mathcal{N}_2] = \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\right| + \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}_2^{-1}\left((\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top + \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\right).$$

更一般地，考虑一般分布 $p(\boldsymbol{x})$ 和多元高斯分布$\mathcal{N}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，其中 $\boldsymbol{x} \in \mathbb{R}^D$。证明

$$D_{\mathrm{KL}}[p \parallel \mathcal{N}] = \int \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})d\boldsymbol{x} + \frac{1}{2}\ln|\boldsymbol{\Sigma}| + \frac{D}{2}\ln 2\pi + \int p(\boldsymbol{x})\ln p(\boldsymbol{x})d\boldsymbol{x}.$$

11. **高斯分布的熵**。与 KL散度密切相关的一个量是熵。分布 $p(\boldsymbol{x})$ 的熵 $\mathrm{H}[p(\boldsymbol{x})]$ 定义为

$$\mathrm{H}[p(\boldsymbol{x})] \triangleq -\int p(\boldsymbol{x}) \ln p(\boldsymbol{x}) d\boldsymbol{x}.$$

给定一个多元高斯分布 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^D$，证明

$$\mathrm{H}[\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})] = \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{D}{2} \ln(2\pi e).$$

12. **联合高斯密度（Lu，2023）**。设 $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ 是从高斯分布中独立同分布抽取的，参数为 $\mathcal{N}\left(x \mid \mu, \sigma^2\right)$。证明 $\mathcal{X}$ 的联合概率分布可以写成

$$p(\mathcal{X} \mid \mu, \sigma^2) = (2\pi)^{-N/2} (\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left[N(\overline{x} - \mu)^2 + N S_{\overline{x}}\right]\right\}, \tag{1.56}$$

其中 $S_x \triangleq \sum_{n=1}^{N} (x_n - x)^2$ 且 $x \triangleq \frac{1}{N} \sum_{i=1}^{N} x_i$。类似于单变量高斯分布下的似然（1.56），证明由均值向量为 N、协方差矩阵为 $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$的多元高斯分布生成的随机观测 $\boldsymbol{\mu}$，其似然为 $\boldsymbol{\Sigma}$

$$\begin{aligned} p(\mathcal{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-ND/2} |\boldsymbol{\Sigma}|^{-N/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{S}_{\boldsymbol{\mu}})\right\} \\ &= (2\pi)^{-ND/2} |\boldsymbol{\Sigma}|^{-N/2} \exp\left\{-\frac{N}{2}(\boldsymbol{\mu} - \overline{\boldsymbol{x}})^{\top} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \overline{\boldsymbol{x}})\right\} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{S}_{\overline{x}})\right\}, \end{aligned}$$

其中

$$\boldsymbol{S}_{\boldsymbol{\mu}} \triangleq \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^{\top}, \quad \boldsymbol{S}_{\overline{x}} \triangleq \sum_{n=1}^{N} (\boldsymbol{x}_n - \overline{\boldsymbol{x}})(\boldsymbol{x}_n - \overline{\boldsymbol{x}})^{\top}, \quad \overline{\boldsymbol{x}} \triangleq \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n.$$

矩阵 $\boldsymbol{S}_x$ 是平方和矩阵，也称为散布矩阵。

13. **多元正态分布的边缘分布与条件分布（Lu，2023）**。设$\mathbf{x}$和$\mathbf{y}$ 为联合高斯随机向量，满足

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}^{\top} & \boldsymbol{B} \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{C}} \\ \widetilde{\boldsymbol{C}}^{\top} & \widetilde{\boldsymbol{B}} \end{bmatrix}^{-1}\right). \text{[6]}$$

6. 给定非奇异 $\boldsymbol{M}$ 及其逆矩阵 $\boldsymbol{M}^{-1}$；并假设下列分块的维度适当（Williams 和 Rasmussen，2006年）：

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \quad \boldsymbol{M}^{-1} = \begin{bmatrix} \widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{C}} & \widetilde{\boldsymbol{D}} \end{bmatrix}.$$

我们有

$$\begin{aligned} \widetilde{\boldsymbol{A}} &= \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1} \boldsymbol{B} \widetilde{\boldsymbol{D}} \boldsymbol{C} \boldsymbol{A}^{-1} &&= (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}, \\ \widetilde{\boldsymbol{B}} &= -\boldsymbol{A}^{-1} \boldsymbol{B} \widetilde{\boldsymbol{D}} &&= -\widetilde{\boldsymbol{A}} \boldsymbol{B} \boldsymbol{D}^{-1}, \\ \widetilde{\boldsymbol{C}} &= -\widetilde{\boldsymbol{D}} \boldsymbol{C} \boldsymbol{A}^{-1} &&= -\boldsymbol{D}^{-1} \boldsymbol{C} \widetilde{\boldsymbol{A}}, \\ \widetilde{\boldsymbol{D}} &= (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} &&= \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1} \boldsymbol{C} \widetilde{\boldsymbol{A}} \boldsymbol{B} \boldsymbol{D}^{-1}, \end{aligned} \tag{1.57}$$

其中，$\mathbf{x}$ 和$\mathbf{y}$ 当且仅当 $\mathbb{C}\text{ov}[\mathbf{x},\mathbf{y}] = \boldsymbol{C} = \mathbf{0}$时相互独立。证明：多元高斯分布的每个边缘分布本身都是多元高斯分布，且条件分布 $\mathbf{x} \mid \mathbf{y}$ 也服从多元高斯分布：

$$\begin{aligned}
\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{A}), \quad \mathbf{x} \mid \mathbf{y} = \boldsymbol{y} &\sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{C}\boldsymbol{B}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y), \boldsymbol{A} - \boldsymbol{C}\boldsymbol{B}^{-1}\boldsymbol{C}^\top) \\
&= \mathcal{N}(\boldsymbol{\mu}_x - \widetilde{\boldsymbol{A}}^{-1}\widetilde{\boldsymbol{C}}(\boldsymbol{y} - \boldsymbol{\mu}_y), \widetilde{\boldsymbol{A}}^{-1}); \\
\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{B}), \quad \mathbf{y} \mid \mathbf{x} = \boldsymbol{x} &\sim \mathcal{N}(\boldsymbol{\mu}_y + \boldsymbol{C}^\top\boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \boldsymbol{B} - \boldsymbol{C}^\top\boldsymbol{A}^{-1}\boldsymbol{C}) \\
&= \mathcal{N}(\boldsymbol{\mu}_y - \widetilde{\boldsymbol{B}}^{-1}\widetilde{\boldsymbol{C}}^\top(\boldsymbol{x} - \boldsymbol{\mu}_x), \widetilde{\boldsymbol{B}}^{-1}).
\end{aligned} \tag{1.58}$$

14. **高斯变量的仿射依赖**。假设随机向量 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 和 $\mathbf{y} \mid \mathbf{x} = \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}, \boldsymbol{M})$。注意，$\mathbf{y}$ 并非简单地等于仿射变换 $\boldsymbol{A}\mathbf{x}+\boldsymbol{b}$，但可以推出 $\mathbf{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{\epsilon}$，其中 $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{M})$。证明

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{M} + \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top), \quad \mathbf{x} \mid \mathbf{y} \sim \mathcal{N}\big(\boldsymbol{L}\big\{\boldsymbol{A}^\top\boldsymbol{M}^{-1}(\mathbf{y} - \boldsymbol{b}) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\big\}, \boldsymbol{L}\big),$$

其中 $\boldsymbol{L} = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^\top\boldsymbol{M}^{-1}\boldsymbol{A})^{-1}$。提示：使用习题 *1.13*，通过 $\mathbb{C}ov[x, y] = \mathbb{E}[(x - \boldsymbol{\mu}_x)(y - \boldsymbol{\mu}_y)^\top] = \boldsymbol{\Sigma}\boldsymbol{A}^\top$ 计算 x 和 y 的交叉协方差，其中 $\boldsymbol{\mu}_x = \boldsymbol{\mu}$ 和 $\boldsymbol{\mu}_y = \mathbb{E}[\ ] = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}$

$\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\ \boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\ ^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$ $\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C}\ ^{-1} = y$，
$\boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{C}\qquad \boldsymbol{D}$
并对适当的矩阵，，，使用伍德伯里矩阵恒等式：$(\ )(\ )$；例如参见 *Lu (2021a)*。

15. **期望的性质**。设 x 和 y 是两个随机变量，a，b 是标量。证明 $\mathbb{E}[a\mathrm{x} + b\mathrm{y}] = a \cdot \mathbb{E}[\mathrm{x}] + b \cdot \mathbb{E}[\mathrm{y}]$。进一步给定一个函数 h，证明

$$\mathbb{E}[h(\mathrm{x})] = \sum_x h(x)\Pr(x) \qquad \text{and} \qquad \mathbb{E}[h(\mathrm{x})] = \int_{-\infty}^{\infty} h(x)dF(x),$$

分别在离散和连续情形下成立。

16. **条件期望的性质**。设 x 和 y 是两个随机变量，h是一个函数。证明

(a)。注意，$\mathbb{E}[\mathrm{x} \mid \mathrm{y}]$ 是 y 的函数。然而，如果 x 与 y 独立，则$\mathbb{E}[\mathrm{x} \mid \mathrm{y}] = \mathrm{x}$。(b)。$\mathbb{E}[c \mid \mathrm{x}] = \mathrm{x}$，其中 c 是一个常数。(c)。线性性：$\mathbb{E}[a\mathrm{x}_1 + b\mathrm{x}_2 \mid \mathrm{y}] = a \cdot \mathbb{E}[\mathrm{x}_1 \mid \mathrm{y}] + b \cdot \mathbb{E}[\mathrm{x}_2 \mid \mathrm{y}]$。(d)。条件常数：$\mathbb{E}[h(\mathrm{y})\mathrm{x} \mid \mathrm{y}] = h(\mathrm{y})\mathbb{E}[\mathrm{x} \mid \mathrm{y}]$，，其中 $h(\mathrm{y})$ 称为关于 y 的条件常数。(e)。单调性：如果 $\mathrm{x}_1 \leq \mathrm{x}_2$，，则 $\mathbb{E}[\mathrm{x}_1 \mid \mathrm{y}] \leq \mathbb{E}[\mathrm{x}_2 \mid \mathrm{y}]$。(f)。塔式性质：$\mathbb{E}\left[\mathbb{E}[\mathrm{x} \mid \mathrm{y}] \mid h(\mathrm{y})\right] = \mathbb{E}[\mathrm{x} \mid h(\mathrm{y})]$；也就是说，$h(\mathrm{y})$至多传达与 y 一样多的信息。(g)。无偏性：$\mathbb{E}\big\{\mathbb{E}[h(\mathrm{x}, \mathrm{y}) \mid \mathrm{y}]\big\} = \mathbb{E}[h(\mathrm{x}, \mathrm{y})]$；特别地，$\mathbb{E}\big[\mathbb{E}[\mathrm{x} \mid \mathrm{y}]\big] = \mathbb{E}[\mathrm{x}]$。(h)。最小二乘：对于任意函数 h，$\mathbb{E}\big[(\mathrm{y} - \mathbb{E}[\mathrm{y} \mid \mathrm{x}])^2\big] \leq \mathbb{E}\big[(\mathrm{y} - h(\mathrm{x}))^2\big]$。这也意味着，在最小二乘意义下，$g(\mathrm{x}) \triangleq \mathbb{E}[\mathrm{y} \mid \mathrm{x}]$ 是最佳估计值。

17. **随机变量的卷积求和**。设 x 和 y 是连续随机变量，其概率密度函数为 f_x 和 f_y。证明，$\mathrm{x} + \mathrm{y}$ 的密度函数是 f_x 与 f_y的卷积：

$$f_{\mathrm{x}+\mathrm{y}}(u) = \int_{-\infty}^{+\infty} f_\mathrm{x}(u - v) f_\mathrm{y}(v)\, dv.$$

18. **方差与相关性的性质。** 设 $\mathrm{x}, \mathrm{x}_1, \mathrm{x}_2, \mathrm{y}, \mathbf{x}$ 是随机变量或向量，且 a, b 是常数。证明：

- 设 Ω 为实对称矩阵。当且仅当 Ω 是某个随机向量 $\mathbf{x}$ 的协方差矩阵时，Ω 为半正定矩阵（定义 1.17）。$\mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}[\mathrm{x}^2] - (\mathbb{E}[\mathrm{x}])^2 = \mathbb{C}\mathrm{ov}[\mathrm{x}, \mathrm{x}]$.$\mathbb{V}\mathrm{ar}[a\mathrm{x} + b] = a^2\mathbb{V}\mathrm{ar}[\mathrm{x}]$.$\mathbb{V}\mathrm{ar}\big[\sum_i \mathrm{x}_i\big] = \sum_i \mathbb{V}\mathrm{ar}[\mathrm{x}_i] + \sum_{i \neq j} \mathbb{C}\mathrm{ov}[\mathrm{x}_i, \mathrm{x}_j]$. $\mathbb{C}\mathrm{ov}[\mathrm{x}_1, \mathrm{x}_2] = \mathbb{E}[\mathrm{x}_1\mathrm{x}_2] - \mathbb{E}[\mathrm{x}_1]\mathbb{E}[\mathrm{x}_2]$.$\mathbb{C}\mathrm{ov}[a\mathrm{x}_1 + b\mathrm{x}_2, \mathrm{y}] = a \cdot \mathbb{C}\mathrm{ov}[\mathrm{x}_1, \mathrm{y}] + b \cdot \mathbb{C}\mathrm{ov}[\mathrm{x}_2, \mathrm{y}]$；也就是说，协方差对其中一个变量是线性的。若 $\mathbb{E}[\mathrm{x}_1^2] + \mathbb{E}[\mathrm{x}_2^2] < \infty$，则以下条件等价：(i) $\mathbb{E}[\mathrm{x}_1\mathrm{x}_2] = \mathbb{E}[\mathrm{x}_1]\mathbb{E}[\mathrm{x}_2]$；(ii) $\mathbb{C}\mathrm{ov}[\mathrm{x}_1, \mathrm{x}_2] = 0$；(iii) $\mathbb{V}\mathrm{ar}[\mathrm{x}_1 \pm \mathrm{x}_2] = \mathbb{V}\mathrm{ar}[\mathrm{x}_1] + \mathbb{V}\mathrm{ar}[\mathrm{x}_2]$。注意，独立性会推出这最后三个性质，但这些性质均不能推出独立性。
- 设 h 为非递减函数，满足 $\mathbb{E}[\mathrm{x}^2] < \infty$ 且 $\mathbb{E}[h(\mathrm{x})^2] < \infty$。则$\mathbb{C}\mathrm{ov}[\mathrm{x}, h(\mathrm{x})] > 0$。

19. **条件方差与全方差公式。** 设 x 和 y 是两个随机变量。给定 y 时 x 的条件方差 定义为

$$\mathbb{V}\mathrm{ar}[\mathrm{x} \mid \mathrm{y}] \triangleq \mathbb{E}\big[(\mathrm{x} - \mathbb{E}[\mathrm{x} \mid \mathrm{y}])^2 \mid \mathrm{y}\big] = \mathbb{E}[\mathrm{x}^2 \mid \mathrm{y}] - (\mathbb{E}[\mathrm{x} \mid \mathrm{y}])^2. \tag{1.59}$$

条件方差告诉我们，使用 $\mathbb{E}[\mathrm{x} \mid \mathrm{y}]$ 来“预测”x 后，还剩下多少方差。证明全方差公式：

$$\mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}\big[\mathbb{V}\mathrm{ar}[\mathrm{x} \mid \mathrm{y}]\big] + \mathbb{V}\mathrm{ar}\big[\mathbb{E}[\mathrm{x} \mid \mathrm{y}]\big]. \tag{1.60}$$

提示：$\mathbb{V}\mathrm{ar}[\mathrm{x}] = \mathbb{E}[\mathrm{x}^2] - (\mathbb{E}[\mathrm{x}])^2 = \mathbb{E}[\mathbb{E}[\mathrm{x}^2 \mid \mathrm{y}]] - (\mathbb{E}[\mathbb{E}[\mathrm{x} \mid \mathrm{y}]])^2$ 根据问题中的无偏性 *1.16*。

第一部分

理论

第2章

优化基础

目录

2.1 可微函数与微分学 65

2.1.1 二阶偏导数.69 2.1.2 多元实函数.70 2.1.3 微积分基本定理.71 **2.2 凸组合与凸集.72 2.3 凸函数.76** 2.3.1 凸函数的性质.78 2.3.2 凸函数的刻画.80 **2.4 连续函数与闭函数，以及魏尔斯特拉斯定理 . 82 2.5 次梯度与共轭函数.86 2.6 利普希茨函数.91 2.7 强凸、强光滑函数.93 2.8 投影、近端、布雷格曼近端与分离 . . . 100** 2.8.1 投影算子的性质.103 2.8.2 分离定理.106 2.8.3 近端算子的性质.106 2.8.4 莫罗分解.110

2.9 凸优化与最优性条件 111

2.9.1 凸集上的约束优化.113 2.9.2 等式/不等式约束.114 2.9.3 对偶问题.114 2.9.4 复合函数.118 **第2章习题.119**



T本章简明而系统地介绍凸分析和凸优化的基本概念，涵盖凸集与凸函数、次微分以及最优性条件等重要主题。

本书旨在为读者介绍统计学、机器学习和信号处理应用所需的核心工具，但并不追求面面俱到。若希望更深入、更全面地学习相关内容——包括高级理论结果、对偶理论、约束优化和专用算法——建议读者参考该领域的标准文献，例如 Bertsekas 等人；Boyd and Vandenberghe（2004年）；Bertsekas （2009年）；Foucart 和 Rauhut（2013年）；Rockafellar（1970年）；Lu （2025年）。

2.1. 可微函数与微分学

可微性与微分学构成数学分析的基础，尤其是在研究定义于多维空间上的函数时更是如此。本节介绍一些基本概念，用于量化函数相对于其输入变量的变化方式。

本节讨论的核心是方向导数，它用于衡量函数 f 在点 $\boldsymbol{\alpha}$沿给定向量 $\boldsymbol{d}$方向的瞬时变化率。

定义 2.1（方向导数、偏导数）。设 f 是定义在集合 $\mathbb{S} \subseteq \mathbb{R}^p$上的实值函数，且 $\boldsymbol{\alpha} \in \mathbb{R}^p$。对于任意非零向量 $\boldsymbol{d} \in \mathbb{R}^p$， f 在 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 沿 $\boldsymbol{d}$ 方向的方向导数定义为——如果该极限存在——

$$f'(\boldsymbol{\alpha}; \boldsymbol{d}) \triangleq \lim_{\mu \to 0} \frac{f(\boldsymbol{\alpha} + \mu \boldsymbol{d}) - f(\boldsymbol{\alpha})}{\mu}.$$

这个量有时也称为盖托导数。

对于每个 $i \in \{1, 2, \ldots, p\}$， 在 $\boldsymbol{\alpha}$ 沿第 i 个标准基向量 $\boldsymbol{e}_i$ 方向的方向导数（如果存在）称为第 i 个偏导数。它通常记作 $\frac{\partial f}{\partial \alpha_i}(\boldsymbol{\alpha})$、 $D_{\boldsymbol{e}_i} f(\boldsymbol{\alpha})$ 或 $\partial_i f(\boldsymbol{\alpha})$。

需要注意的是，某点沿所有非零方向的方向导数都存在，并不能保证该点的偏导数存在——更一般地说，也不能保证该点可微。一个经典例子是函数 $f(\alpha, \beta) = \sqrt{\alpha^2 + \beta^2}$（见图2.1）。考虑方向 $\boldsymbol{d} = [a, b]^\top$。在 $[0, 0]^\top$ 处的偏导数为

$$\frac{\partial f}{\partial \alpha}(0,0) = \lim_{h \to 0} \frac{\sqrt{(0+h)^2 + 0^2}}{h} = \lim_{h \to 0} \frac{|h|}{h},$$
$$\frac{\partial f}{\partial \beta}(0,0) = \lim_{h \to 0} \frac{\sqrt{0^2 + (0+h)^2}}{h} = \lim_{h \to 0} \frac{|h|}{h}.$$

当 $h > 0$ 时，偏导数为 1；当 $h < 0$ 时，偏导数为 -1。由于左极限和右极限不同，因此极限不存在。

定义 2.2（梯度）。如果函数 f 在点 $\boldsymbol{\alpha} \in \mathbb{R}^p$处的所有偏导数都存在，那么 f 在 $\boldsymbol{\alpha}$ 处的梯度定义为由这些偏导数组成的列向量：

$$\nabla f(\boldsymbol{\alpha}) = \left[\frac{\partial f}{\partial \alpha_1}(\boldsymbol{\alpha}), \frac{\partial f}{\partial \alpha_2}(\boldsymbol{\alpha}), \ldots, \frac{\partial f}{\partial \alpha_p}(\boldsymbol{\alpha}) \right]^\top \in \mathbb{R}^p.$$

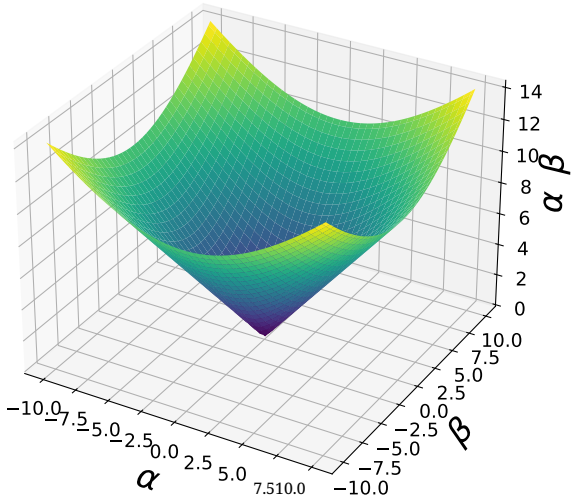


图 2.1： 函数 $f(\alpha,\ \beta) = \sqrt{\alpha^2+\beta^2}$ 的图像。在 $[0,0]^\top$ 处，偏导数不存在。

习题 2.3。证明函数在某点处可能存在所有偏导数，却在该点不连续（参见定义2.48）。反过来，证明函数可能处处连续，却仍在某些点处缺少偏导数。

定义在开集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上的函数 f (见定义 2.19) 称为在点 $\boldsymbol{\alpha} \in \mathbb{S}$ 处可微，如果它在 $\boldsymbol{\alpha}$ 附近存在一个线性近似。这一概念通过弗雷歇´可微性形式化，它比仅要求偏导数存在更强。

定义 2.4（（弗雷歇´）可微性）。设 $f : \mathbb{R}^p \to \mathbb{R}$。函数 f 称为在 $\boldsymbol{\alpha}$ 处可微或弗雷歇´可微，如果存在一个向量 $\boldsymbol{g} \in \mathbb{R}^p$ 使得

$$\lim_{\boldsymbol{d}\to\mathbf{0}} \frac{f(\boldsymbol{\alpha}+\boldsymbol{d}) - f(\boldsymbol{\alpha}) - \boldsymbol{g}^\top \boldsymbol{d}}{\|\boldsymbol{d}\|_2} = 0.$$

向量 $\boldsymbol{g}$ 是唯一的，且等于梯度 $\nabla f(\boldsymbol{\alpha})$。

此外，如果函数 f 定义在开集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上，且它的所有偏导数都存在并且在 $\mathbb{S}$上连续，那么称它在 $\mathbb{S}$ 上连续可微。

请注意，函数可以是可微的，但不连续可微。一个经典例子是单变量函数：

$$f(\alpha) = \begin{cases} \alpha^2 \sin(\frac{1}{\alpha}), & \text{if } \alpha \neq 0, \\ 0, & \text{if } \alpha = 0. \end{cases}$$

该函数在 $\mathbb{R}$ 上处处可微，包括 $\alpha = 0$ 处。为此，对于 $\alpha \neq 0$，$f(\alpha)$ 是两个可微函数（α^2 和 $\sin(1/\alpha)$）的乘积，因此它是可微的。在 $\alpha = 0$ 处，利用极限定义计算导数：

$$f'(0) = \lim_{\mu\to 0} \frac{f(\mu) - f(0)}{\mu} = \lim_{\mu\to 0} \frac{\mu^2 \sin(\frac{1}{\mu}) - 0}{\mu} = \mu \sin(\frac{1}{\mu})$$

由于对所有 $\mu \neq 0$ 都有 $|\sin(1/\mu)| \leq 1$，因此极限存在且等于 0。因此，$f(\alpha)$ 在 $\alpha = 0$ 处可微，且 $f'(0) = 0$。然而，$f(\alpha)$ 在 $\alpha = 0$ 处并非连续可微。当 $\alpha \neq 0$ 时，$f(\alpha)$ 的导数为

$$f'(\alpha) = 2\alpha \sin(\frac{1}{\alpha}) - \cos(\frac{1}{\alpha}).$$

极限 $\lim_{\alpha \to 0} f'(\alpha)$ 不存在，因为当 α 趋近于 0 时，正弦函数和余弦函数会振荡。

引理 2.5：（方向导数的梯度公式）若 f 在 $\boldsymbol{\alpha} \in \mathbb{S} \subseteq \mathbb{R}^p$处可微，则对任意方向 $\boldsymbol{d} \in \mathbb{R}^p$，$f$ 在 $\boldsymbol{\alpha}$ 处沿方向 $\boldsymbol{d}$ 的方向导数满足

$$f'(\boldsymbol{\alpha}; \boldsymbol{d}) = \nabla f(\boldsymbol{\alpha})^\top \boldsymbol{d}, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{S} \text{ and } \boldsymbol{d} \in \mathbb{R}^p. \tag{2.1}$$

证明 [引理2.5] 当 $\boldsymbol{d} = \mathbf{0}$时，该公式显然成立。下面假设$\boldsymbol{d} \neq \mathbf{0}$。的可微性意味着

$$0 = \lim_{\mu \to 0^+} \frac{f(\boldsymbol{\alpha} + \mu \boldsymbol{d}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \mu \boldsymbol{d} \rangle}{\|\mu \boldsymbol{d}\|_2} = \lim_{\mu \to 0^+} \left[\frac{f(\boldsymbol{\alpha} + \mu \boldsymbol{d}) - f(\boldsymbol{\alpha})}{\mu \|\boldsymbol{d}\|_2} - \frac{\langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{d} \rangle}{\|\boldsymbol{d}\|_2} \right].$$

因此，

$$\begin{aligned} f'(\boldsymbol{\alpha}; \boldsymbol{d}) &= \lim_{\mu \to 0^+} \frac{f(\boldsymbol{\alpha} + \mu \boldsymbol{d}) - f(\boldsymbol{\alpha})}{\mu} \\ &= \lim_{\mu \to 0^+} \left\{ \|\boldsymbol{d}\|_2 \left[\frac{f(\boldsymbol{\alpha} + \mu \boldsymbol{d}) - f(\boldsymbol{\alpha})}{\mu \|\boldsymbol{d}\|_2} - \frac{\langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{d} \rangle}{\|\boldsymbol{d}\|_2} \right] + \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{d} \rangle \right\} = \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{d} \rangle . \end{aligned}$$

这就证明了（2.1）。 ■

回顾 f 在 $\boldsymbol{\alpha}$处的可微性，我们有

$$\lim_{\boldsymbol{d} \to \mathbf{0}} \frac{f(\boldsymbol{\alpha} + \boldsymbol{d}) - f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\alpha})^\top \boldsymbol{d}}{\|\boldsymbol{d}\|_2} = 0, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{S}, \tag{2.2}$$

or

$$f(\boldsymbol{\beta}) = f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top (\boldsymbol{\beta} - \boldsymbol{\alpha}) + o(\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2), \tag{2.3}$$

其中小 o 函数 $o(\cdot)$：$\mathbb{R}_+ \to \mathbb{R}$ 是一个一维函数，满足 $o(\mu)/\mu \to 0$当 $\mu \to 0^+$时。[1]因此，任何可微函数都满足 $\lim_{\boldsymbol{\alpha} \to \boldsymbol{a}} f(\boldsymbol{\alpha}) = f(\boldsymbol{a})$。因此，**任何可微函数都是连续的**。

▶ **下降方向与最陡方向**。梯度与方向导数之间的关系式（2.1）的一个重要推论是，梯度编码了最速上升和下降的方向。

1. 注意，我们也使用标准的 *big-Oh* 记号来描述函数的渐近行为。具体而言，记号 $g(\boldsymbol{d}) = \mathcal{O}(\|\boldsymbol{d}\|_2^p)$ 表示存在正数 C_1 和 δ，使得对所有 $\|\boldsymbol{d}\|_2 \leq \delta$，都有 $|g(\boldsymbol{d})| \leq C_1 \|\boldsymbol{d}\|_2^p$。实践中，当 $\boldsymbol{d}$充分小时，它通常等价于 $|g(\boldsymbol{d})| \approx C_2 \|\boldsymbol{d}\|_2^p$，其中 C_2 是另一个正常数。*soft-Oh* 记号用于隐藏多对数因子，即 $f = \widetilde{\mathcal{O}}(g)$ 将意味着对某个绝对常数 c，有 $f = \mathcal{O}(g \log^c(g))$。

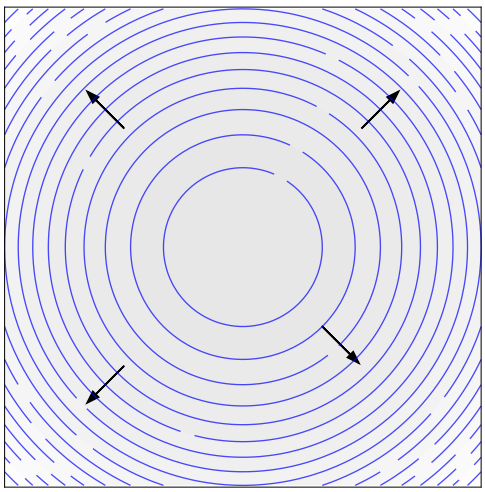


图 2.2： 函数$f:\mathbb{R}^2\to\mathbb{R}$的等高线。不同点处的梯度以黑色箭头表示，它们与等高线正交。

引理 2.6：（最陡方向）设 $f:\mathbb{R}^p\to\mathbb{R}$ 在 $\boldsymbol{\alpha}$处可微。在所有单位方向 $\boldsymbol{d}\in\mathbb{R}^p$（即 $\|\boldsymbol{d}\|_2=1$）中，当 $\boldsymbol{d}$ 指向 $\nabla f(\boldsymbol{\alpha})$ 的方向时，方向导数 $f'(\boldsymbol{\alpha};\boldsymbol{d})$ 取得最大值；当 $\boldsymbol{d}$ 指向相反方向时，方向导数取得最小值（最负）。

证明 [引理 2.6] 对于所有 $\boldsymbol{d}\in\mathbb{R}^p$ 满足 $\|\boldsymbol{d}\|_2=1$的情形，根据柯西–施瓦茨不等式（定理 1.59），可得

$$\left|f'(\boldsymbol{\alpha};\boldsymbol{d})\right|=\left|\nabla f(\boldsymbol{\alpha})^\top\boldsymbol{d}\right|\le\|\nabla f(\boldsymbol{\alpha})\|_2\,\|\boldsymbol{d}\|_2=\|\nabla f(\boldsymbol{\alpha})\|_2\,.$$

当且仅当 $\boldsymbol{d}$ 是 ∇f 的标量倍数时，等号成立（$\boldsymbol{\alpha}$）。由于 $\|\boldsymbol{d}\|_2=1$，最大化方向为 $\boldsymbol{d}=\pm\frac{\nabla f}{(\boldsymbol{\alpha})\,\|\nabla f(\boldsymbol{\alpha})\|_2}$ ■

图 2.2 展示了函数 $f:\mathbb{R}^2\to\mathbb{R}$ 的梯度场。梯度向量（以黑色箭头表示）与等高线正交。这种正交性是因为函数沿每条等高线保持常数，因此沿任意切向方向的方向导数都为零。由（2.1），可得 $\nabla f(\boldsymbol{\alpha})^\top\boldsymbol{d}=0$ 对所有切向量$\boldsymbol{d}$均成立，这意味着梯度必定垂直于水平集。

注意，$\boldsymbol{d}=-\nabla f(\boldsymbol{\alpha})$称为最速下降方向。更一般地，任何满足以下条件的方向 $\boldsymbol{d}$

$$f'(\boldsymbol{\alpha};\boldsymbol{d})<0$$

称为在 $\boldsymbol{\alpha}$处的下降方向。

引理 2.7：（无约束与有约束下降性质）设 f 是定义在 $\mathbb{S}\subseteq\mathbb{R}^p$（分别为 $\mathbb{R}^p$）上的连续可微函数，且 $\boldsymbol{\alpha}\in\mathbb{S}$（分别为 $\boldsymbol{\alpha}\in\mathbb{R}^p$）。假设 $\boldsymbol{d}$ 是 f 在 $\boldsymbol{\alpha}$处的下降方向。那么，存在 $\varepsilon>0$ 使得

$$f(\boldsymbol{\alpha}+\eta\boldsymbol{d})<f(\boldsymbol{\alpha}),\quad\text{for any }\eta\in(0,\varepsilon].$$

引理 [的证明 2。7] 由于 $f'(\boldsymbol{\alpha};\ \boldsymbol{d})=\nabla f(\boldsymbol{\alpha})^\top\boldsymbol{d}<0$，根据方向导数的定义（定义2.1），可得

$$\lim_{\eta\to0^+}\frac{f(\boldsymbol{\alpha}+\eta\boldsymbol{d})-f(\boldsymbol{\alpha})}{\eta}=f'(\boldsymbol{\alpha};\boldsymbol{d})<0.$$

因此，存在一个 $\varepsilon > 0$ 使得 $\frac{f(\boldsymbol{\alpha}+\eta\boldsymbol{d})-f(\boldsymbol{\alpha})}{\eta} < 0$ 对于任意 $\eta \in (0, \varepsilon]$，由此完成证明。 ■

2.1.1 二阶偏导数

偏导数 $\frac{\partial f}{\partial \alpha_i}(\boldsymbol{\alpha})$ 本身是关于 $\boldsymbol{\alpha} \in \mathbb{S}$ 的实值函数，也可以再次求偏导。$\frac{\partial f}{\partial \alpha_i}(\boldsymbol{\alpha})$ 的 j 阶偏导数定义为

$$\frac{\partial^2 f}{\partial \alpha_j \partial \alpha_i}(\boldsymbol{\alpha}) = \frac{\partial \left(\frac{\partial f}{\partial \alpha_i}(\boldsymbol{\alpha})\right)}{\partial \alpha_j}(\boldsymbol{\alpha}).$$

这称为函数 f 的 (j, i) 阶二阶偏导数。

在开集 $\mathbb{S} \subseteq$ 上定义的函数 f，若其所有二阶偏导数都存在且在 $\mathbb{S}$上连续，则称其在$\mathbb{S}$上二阶连续可微。在二阶连续可微的条件下，混合偏导数是对称的：

$$\frac{\partial^2 f}{\partial \alpha_j \partial \alpha_i}(\boldsymbol{\alpha}) = \frac{\partial^2 f}{\partial \alpha_i \partial \alpha_j}(\boldsymbol{\alpha}).$$

函数 f 在点 $\boldsymbol{\alpha} \in \mathbb{S}$ 处的海森矩阵是由二阶偏导数组成的 $p \times p$ 矩阵：

$$\nabla^2 f(\boldsymbol{\alpha}) = \begin{bmatrix} \frac{\partial^2 f}{\partial \alpha_1^2}(\boldsymbol{\alpha}) & \frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_2}(\boldsymbol{\alpha}) & \cdots & \frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_p}(\boldsymbol{\alpha}) \\ \frac{\partial^2 f}{\partial \alpha_2 \partial \alpha_1}(\boldsymbol{\alpha}) & \frac{\partial^2 f}{\partial \alpha_2 \partial \alpha_2}(\boldsymbol{\alpha}) & \cdots & \frac{\partial^2 f}{\partial \alpha_2 \partial \alpha_p}(\boldsymbol{\alpha}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \alpha_p \partial \alpha_1}(\boldsymbol{\alpha}) & \frac{\partial^2 f}{\partial \alpha_p \partial \alpha_2}(\boldsymbol{\alpha}) & \cdots & \frac{\partial^2 f}{\partial \alpha_p^2}(\boldsymbol{\alpha}) \end{bmatrix}.$$

当 f 二阶连续可微时，由于混合偏导数相等，该矩阵是对称的。

定义 2.8（二阶可微）。根据定义，可微函数f： $\mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$ 在点 $\boldsymbol{\alpha} \in \mathbb{S}$ 处是二阶可微的，当且仅当存在一个矩阵 $\nabla^2 f(\boldsymbol{\alpha}) \in \mathbb{R}^{p \times p}$ 使得

$$\lim_{\boldsymbol{d} \to \mathbf{0}} \frac{\left\| \nabla f(\boldsymbol{\alpha}+\boldsymbol{d}) - \nabla f(\boldsymbol{\alpha}) - \nabla^2 f(\boldsymbol{\alpha})\boldsymbol{d} \right\|_2}{\|\boldsymbol{d}\|_2} = 0.$$

矩阵 $\nabla^2 f(\boldsymbol{\alpha}) \in \mathbb{R}^{p \times p}$ 是唯一的，且与函数 f 在 $\boldsymbol{\alpha}$处的海森矩阵一致。

从几何角度看，函数的曲率描述了其斜率变化的快慢；在一维情形下，这由二阶导数刻画。在高维情形下，海森矩阵编码了函数在各个方向上的曲率。

类似于方向导数的梯度公式（引理2.5），二阶方向导数与海森矩阵的关系如下。

引理 2.9：（海森矩阵与方向导数恒等式）若函数 f： $\mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$二阶可微，则函数 f 在 $\boldsymbol{\alpha} \in \mathbb{S}$处的二阶方向导数 $f''(\boldsymbol{\alpha}; \boldsymbol{d})$等于

$$f''(\boldsymbol{\alpha}; \boldsymbol{d}) = \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{\alpha}) \boldsymbol{d}, \tag{2.4}$$

对于任意单位范数向量 $\boldsymbol{d} \in \mathbb{R}^p$。

2.1.2 多元实函数

现在我们介绍几个依赖于导数及相关概念的经典逼近定理。

定理 2.10：（罗尔定理）设 $f(\alpha)$ 在闭区间 $[a,b]$ 上连续，且在 (a,b) 上可微。若 $f(a) = f(b)$，则至少存在一点 $\xi \in (a,b)$，使得 $f'(\xi) = 0$。

证明 [定理 2.10] 只需注意，要么 f 是常数，此时 $f'(\alpha) = 0$对所有 α 属于 (a,b) 的情形都成立；要么它在 $\xi \in (a,b)$ 处取得局部最大值或最小值，此时$f'(\xi) = 0$。 ■

罗尔定理主要用作证明普通中值定理（OMVT）的关键步骤。

定理 2.11：（普通中值定理（OMVT））设 $f(\alpha)$ 在闭区间 $[a,b]$ 上连续，且在 (a,b) 上可微。则存在一点 $\xi \in (a,b)$，使得

$$f(b) - f(a) = f'(\xi)(b-a).$$

证明 [定理 2.11]定义辅助函数 $g(\alpha) \triangleq f(\alpha) - \left(\frac{f(b)-f(a)}{b-a}\right)(\alpha - a)$。有 $g(a) = g(b)$。由罗尔定理，存在 $\xi \in (a,b)$，使得 $g'(\xi) = 0$，整理后即可得到所需的恒等式：

$$f(b) - f(a) = f'(\xi)(b-a).$$

证明至此完成。 ■

对于多变量函数，也有类似的结果，不过由于梯度的向量值性质，其形式略有不同。

定理 2.12：（中值定理（MVT））设 $f: \mathbb{S} \to \mathbb{R}$ 是在开集 $\mathbb{S} \subseteq \mathbb{R}^p$上连续可微的函数，给定两个点 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$。那么，存在一点 $\boldsymbol{\xi} \in [\boldsymbol{\alpha}, \boldsymbol{\beta}]$ 使得

$$f(\boldsymbol{\beta}) = f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\xi})^{\top}(\boldsymbol{\beta} - \boldsymbol{\alpha}).$$

证明 [定理 2.12] 不失一般性，假设 $\boldsymbol{\xi}(t) = \boldsymbol{\alpha} + t(\boldsymbol{\beta} - \boldsymbol{\alpha})$，且$t \in [0,1]$。然后定义 $g(t) \triangleq f(\boldsymbol{\xi}(t))$。我们可以对 $g(t)$应用普通中值定理，得到

$$g(1) = g(0) + g'(\gamma),$$

对于区间 $[0,1]$中的某个 γ。$g(t)$ 的导数为

$$g'(t) = \langle \nabla f(\boldsymbol{\xi}(t)), \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle,$$

其中 $\nabla f(\boldsymbol{\xi}(t)) = \left[\frac{\partial f}{\partial \alpha_1}(\boldsymbol{\xi}(t)) \ \frac{\partial f}{\partial \alpha_2}(\boldsymbol{\xi}(t)) \ldots, \frac{\partial f}{\partial \alpha_p}(\boldsymbol{\xi}(t))\right]^{\top}$. 因此，

$$g'(\gamma) = \langle \nabla f(\boldsymbol{\xi}(\gamma)), \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle.$$

由于 $\boldsymbol{\xi}(\gamma) = (1-\gamma)\boldsymbol{\alpha} + \gamma\boldsymbol{\beta}$，证明至此完成。 ■

定理 2.13：（线性近似定理）设 $f:\mathbb{S}\to\mathbb{R}$ 是定义在开集 $\mathbb{S}\subseteq\mathbb{R}^p$ 上的二次连续可微函数，且 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{S}$。那么，存在一点 $\boldsymbol{\xi}\in[\boldsymbol{\alpha},\boldsymbol{\beta}]$，使得

$$\begin{aligned} & f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\alpha})^{\top}\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{\beta}-\boldsymbol{\alpha}), \\ \text{or}\quad & f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})+o(\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2), \\ \text{or}\quad & f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})+\mathcal{O}(\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2). \end{aligned}$$

这个定理也称为一阶泰勒级数展开，它表明线性近似的误差与 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 之间距离的平方同阶。一阶泰勒级数展开 $f(\boldsymbol{\beta})\approx f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})$ 仅在点 $\boldsymbol{\alpha}$ 处的 $\boldsymbol{\beta}$ 邻域内是精确的，

定理 2.14：（二次近似定理）设 $f:\mathbb{S}\to\mathbb{R}$ 是开集 $\mathbb{S}\subseteq\mathbb{R}^p$ 上的二阶连续可微函数，且 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{S}$。则有

$$\begin{aligned} & f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\alpha})^{\top}\nabla^2 f(\boldsymbol{\alpha})(\boldsymbol{\beta}-\boldsymbol{\alpha})+o(\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2), \\ \text{or}\quad & f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^{\top}(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\alpha})^{\top}\nabla^2 f(\boldsymbol{\alpha})(\boldsymbol{\beta}-\boldsymbol{\alpha})+\mathcal{O}(\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^3). \end{aligned}$$

这个定理表明，二次逼近的误差与 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 之间距离的立方同阶，因此当 $\boldsymbol{\beta}$ 接近 $\boldsymbol{\alpha}$ 时，二次逼近更加准确。

2.1.3 微积分基本定理

现在，我们将在多变量情形下介绍微积分基本定理。这一结果在连接微分学与积分学方面起着关键作用，并为理解光滑函数的结构和行为提供了深刻见解。

定理 2.15：（微积分基本定理）设 $f:\mathbb{R}^p\to\mathbb{R}$ 是一个连续可微函数，并且 $\boldsymbol{\beta}$，$\boldsymbol{\alpha}\in\mathbb{R}^p$。那么函数值之差可以表示为

$$f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})=\int_0^1\langle\nabla f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha})),\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle\,d\mu. \tag{2.5a}$$

此外，若 f 是二阶连续可微的，则梯度之差满足

$$\nabla f(\boldsymbol{\beta})-\nabla f(\boldsymbol{\alpha})=\left(\int_0^1\nabla^2 f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha}))d\mu\right)\cdot(\boldsymbol{\beta}-\boldsymbol{\alpha}); \tag{2.5b}$$

$$\nabla f(\boldsymbol{\alpha}+\eta\boldsymbol{d})-\nabla f(\boldsymbol{\alpha})=\int_0^{\eta}\nabla^2 f(\boldsymbol{\alpha}+\mu\boldsymbol{d})\boldsymbol{d}d\mu. \tag{2.5c}$$

其中，$\nabla f(\boldsymbol{\beta})$表示在 $\boldsymbol{\beta}$处计算的 f梯度，$\nabla^2 f(\cdot)$表示海森矩阵，而$\langle\cdot,\ \cdot\rangle$ 表示标准欧几里得内积。此外，对于

二阶连续可微函数，我们还有

$$f(\boldsymbol{\beta}) = f(\boldsymbol{\alpha}) + \langle \nabla f(\boldsymbol{\alpha}), (\boldsymbol{\beta} - \boldsymbol{\alpha}) \rangle + \int_0^1 (1-\mu) \frac{\partial^2 f(\boldsymbol{\alpha} + \mu(\boldsymbol{\beta} - \boldsymbol{\alpha}))}{\partial \mu^2} d\mu. \tag{2.5d}$$

第一个等式（ 2 5 $g\ t\ \triangleq$.a）可以通过定义辅助标量函数 $f\ \boldsymbol{\alpha} + t\ \boldsymbol{\beta} - \boldsymbol{\alpha}$ 推导得出。然后，我们可以得到 $f\ \boldsymbol{\beta}\quad f\ \boldsymbol{\beta}\ = g\ 1\ = g\ 0\ + \int_0^1 g'\ t\ dt.$，即 。其余恒等式也可用类似方法证明。

2.2. 凸组合与凸集

本节将在 $\mathbb{R}^p$中向量集的背景下，介绍仿射组合、凸组合和锥组合及其相关包的概念。我们将为这些基本几何构造提供精确定义和数学表述。

内点与闭集

给定一个向量集 $\boldsymbol{\alpha}_1$， $\boldsymbol{\alpha}_2$、……、 $\boldsymbol{\alpha}_k \in \mathbb{R}^p$，这些向量的仿射组合是形如 $\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i$ 的向量，其中 $\sum_{i=1}^k \lambda_i = 1$，即这些点的线性组合，其系数（不一定非负）之和为 1。这些向量的凸组合是形如 $\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i$ 的向量，其中 $\lambda_i \geq 0$，对于满足 $\sum_{i=1}^k \lambda_i = 1$ 的 $i \in \{1, 2, \ldots, k\}$（即 $\{\lambda_i\}$ 属于 $\mathbb{R}^p$ 中的单位单纯形： $\boldsymbol{\lambda} \in \Delta_k$）；若只要求 $\lambda_i \geq 0$ 而不约束其和，则该组合称为这些向量的锥组合。

在 $\mathbb{R}^p$上固定一个范数后，我们在定义1.29中引入了开球与闭球的概念。这些概念自然引出了内点与相对内点的概念，它们是拓扑学和凸分析的基础。

定义 2.16（仿射包）。集合 $\mathbb{S}$的仿射包记作 aff($\mathbb{S}$)，它是包含 $\mathbb{S}$ 的最小仿射子空间（即向量子空间的一个平移）。形式上，给定集合 $\mathbb{S} \subseteq \mathbb{R}^p$， $\mathbb{S}$ 的仿射包定义为由 $\mathbb{S}$ 中的点构成的所有仿射组合的集合：

$$\text{aff}(\mathbb{S}) \triangleq \left\{ \sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_k \in \mathbb{S}, \lambda_i \in \mathbb{R}, \mathbf{1}^\top \boldsymbol{\lambda} = 1, k \in \mathbb{N} \right\}.$$

定义 2.17（内点与相对内点）。拓扑空间中集合 $\mathbb{S}$ 的内部是 $\mathbb{S}$ 中所有存在一个完全包含于 $\mathbb{S}$ 的邻域的点所构成的集合。形式上， $\mathbb{S}$ 的内部记作 int($\mathbb{S}$)，定义如下：

$$\text{int}(\mathbb{S}) = \{\boldsymbol{\alpha} \in \mathbb{S} \mid \exists \epsilon > 0 \text{ such that } \mathbb{B}(\boldsymbol{\alpha}, \epsilon) \subseteq \mathbb{S}\}.$$

其中， $\mathbb{B}(\boldsymbol{\alpha}, \epsilon)$ 表示以 $\boldsymbol{\alpha}$ 为中心、半径为 ϵ 的开球。

拓扑空间中集合 $\mathbb{S}$ 的相对内部是其相对于仿射包的内部。 $\mathbb{S}$ 的相对内部记作 relint($\mathbb{S}$)，定义如下：

$$\text{relint}(\mathbb{S}) = \{\boldsymbol{\alpha} \in \mathbb{S} \mid \exists \epsilon > 0 \text{ such that } \mathbb{B}_{\text{aff}(\mathbb{S})}(\boldsymbol{\alpha}, \epsilon) \subseteq \mathbb{S}\}.$$

其中， $\mathbb{B}_{\text{aff}(\mathbb{S})}(\boldsymbol{\alpha}, \epsilon)$ 是在 aff($\mathbb{S}$) 的子空间拓扑中，以 $\boldsymbol{\alpha}$ 为中心、半径为 ϵ 的开球。

换句话说，集合的内部 关注的是环境空间中的开球；相对内部 关注的是集合仿射包中的开球，也就是说，它只考虑实际包含 $\mathbb{S}$ 的低维仿射子空间中的邻域。这一点在高维情形和凸分析中尤为重要，因为集合可能位于一个低维子空间中。

例2.18（内点与相对内点）。 **$\mathbb{R}$中的区间。**设 $\mathbb{S} = [0,1]$。则 $\mathsf{int}(\mathbb{S}) = (0,\ 1)$，且 $\mathsf{relint}(\mathbb{S}) = (0,\ 1)$。这里，内部与相对内部相同，因为集合 $[0,1]$ 是 $\mathbb{R}$ 的子集，且 $\mathsf{aff}(\mathbb{S}) = \mathbb{R}$。

$\mathbb{R}^2$中的线段。设 $\mathbb{S} = \{(x,\ y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ y = 0\}$，其中 $\mathsf{aff}(\mathbb{S}) = \{(x,\ 0) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$。则 $\mathsf{int}(\mathbb{S}) = \varnothing$，且 $\mathsf{relint}(\mathbb{S}) = \{(x,\ 0) \mid 0 < x < 1\}$。内部为空，因为在 $\mathbb{R}^2$ 中不存在完全包含于 $\mathbb{S}$的开球。然而，相对内部是在线段（仿射包是直线 $y = 0$）背景下的开区间 $(0, 1)$。

$\mathbb{R}^2$中的三角形。设 $\mathbb{S}$ 是顶点分别位于 $(0,\ 0)$、$(1,\ 0)$ 和 $(0,\ 1)$ 的（填充）三角形。那么，内部由三角形内严格不在边界上的所有点组成；在此情况下，相对内部与内部相同，因为该三角形的仿射包是整个平面 $\mathbb{R}^2$。□

定义 2.19（开集、闭集、紧集与水平集）。如果集合 $\mathbb{S}_1$ 只由内点组成，则称其为开集。也就是说，对于每个 $\boldsymbol{\alpha} \in \mathbb{S}$，都存在一个标量$s > 0$，使得 $\mathbb{B}(\boldsymbol{\alpha},\ s) \subseteq \mathbb{S}_1$。

如果集合 $\mathbb{S}_2$ 的补集 $\mathbb{S}_2^c$ 是开集，则称集合 $\mathbb{S}_2$ 为闭集。或者，如果集合 $\mathbb{S}_2$ 包含其内点序列的所有极限点，则称其为闭集；也就是说，对于满足 $\boldsymbol{\alpha}_i \to \boldsymbol{\alpha}^\star$（当 $i \to \infty$ 时）的每个点序列 $\{\boldsymbol{\alpha}_i\}_{i \ge 1} \subseteq \mathbb{S}_2$，都有 $\boldsymbol{\alpha}^\star \in \mathbb{S}_2$。与此同时，集合 $\mathbb{S}$的闭包（记作 $\mathsf{cl}(\mathbb{S})$）定义为包含 $\mathbb{S}$的最小闭集：

$$\mathrm{cl}(\mathbb{S}) = \cap \{\mathbb{T} \mid \mathbb{S} \subseteq \mathbb{T}, \mathbb{T} \text{ is closed}\}.$$

若存在 $M > 0$，使得 $\mathbb{S} \subseteq \mathbb{B}(\mathbf{0},\ M)$，则称集合 $\mathbb{S} \subseteq \mathbb{R}^p$ 为有界，其中$\mathbb{B}(\mathbf{0},\ M)$是以原点为中心、半径有限的球 M。若集合 $\mathbb{S} \subseteq \mathbb{R}^p$ 是闭且有界的，则称其为紧。

给定函数 $f : \mathbb{S} \to \mathbb{R}$，其中 $\mathbb{S} \subseteq \mathbb{R}^p$，函数 f 在特定值 $c \in \mathbb{R}$ 处的水平集定义如下：

$$\mathrm{Lev}[f, c] = \{\boldsymbol{\alpha} \in \mathbb{S} \mid f(\boldsymbol{\alpha}) \le c\}.\ ^a$$

换言之，水平集 $\mathsf{Lev}[f, c]$ 由定义域 $\mathbb{S}$ 中所有满足函数 f 的取值小于或等于常数 c的点 $\boldsymbol{\alpha}$ 组成。

[a]。注意，我们使用方括号而非圆括号，以表示等号可以取到；这与定义 1.29 中闭球的记号一致。

练习 2.20（内部与闭包）。设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个内部非空的凸集。证明：

$$\mathrm{cl}(\mathrm{int}(\mathbb{S})) = \mathrm{cl}(\mathbb{S}) \qquad \text{and} \qquad \mathrm{int}(\mathrm{cl}(\mathbb{S})) = \mathrm{int}(\mathbb{S}).$$

凸集

对于向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$，任何形如 $\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}$ 且满足 $\lambda\in[0,1]$ 的点，都称为这两个向量的凸组合。对所有这类凸组合都封闭的集合称为凸集。标准定义如下：

定义 2.21（凸集）。如果集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 满足：对于任意 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}\in\mathbb{S}$和 $\lambda\in[0,1]$，点 $\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}$ 也属于 $\mathbb{S}$，则称其为凸集。

根据定义，很容易验证，集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 是凸集，当且仅当它包含其有限多个点的每一个凸组合。也就是说，对于任意 $k\in\mathbb{N}$，任意$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、……、$\boldsymbol{\alpha}_k\in\mathbb{S}$和 λ_1，λ_2、……、$\lambda_k\geq 0$，只要满足 $\sum_{i=1}^k\lambda_i=1$，凸组合$\sum_{i=1}^k\lambda_i\boldsymbol{\alpha}_i$ 也属于 $\mathbb{S}$。对于任意集合 $\mathbb{S}\subseteq\mathbb{R}^p$（可能是非凸集），其凸包定义为所有此类凸组合构成的集合：

定义2.22（凸包）。集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 的凸包记作 conv($\mathbb{S}$)，它是 $\mathbb{S}$ 中所有点的凸组合所构成的集合：

$$\operatorname{conv}(\mathbb{S})\triangleq\left\{\sum_{i=1}^k\lambda_i\boldsymbol{\alpha}_i\mid\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_k\in\mathbb{S},\boldsymbol{\lambda}\in\Delta_k,k\in\mathbb{N}\right\},$$

其中 $\Delta_k=\{(\lambda_1,\ \lambda_2,\ \ldots\ldots,\ \lambda_k)\mid\lambda_i\geq 0,\ \sum_i^k\lambda_i=1\}$ 是标准的（$k-1$）单位单纯形。

任意一族凸集的交集仍然是凸集。从几何上讲，凸集包含其任意两点之间的整条线段（见图2.3）。因此，这类集合没有“内凹”或凹区域。

例2.23。凸集的常见例子包括仿射子空间、线性子空间、半空间、多边形，以及形如 $\mathbb{B}_{\|\cdot\|}[\boldsymbol{c},r]$ 的范数球，其中 $\boldsymbol{c}$ 是中心，$r>0$ 是半径，且 $\|\cdot\|$ 是给定的范数（见定义1.29）。□

习题 2.24（凸集的闭包）。设 $\mathbb{S}\subseteq\mathbb{R}^p$ 是一个凸集。证明其闭包 cl($\mathbb{S}$) 也是凸集。

定义 2.25（锥与凸锥）。如果对于任意 $\boldsymbol{\alpha}\in\mathbb{S}$ 和 $\lambda\geq 0$，缩放后的向量 $\lambda\boldsymbol{\alpha}$ 也属于 $\mathbb{S}$，则称集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 为一个锥。如果集合 $\mathbb{S}$ 既是锥又是凸集，则称其为凸锥。

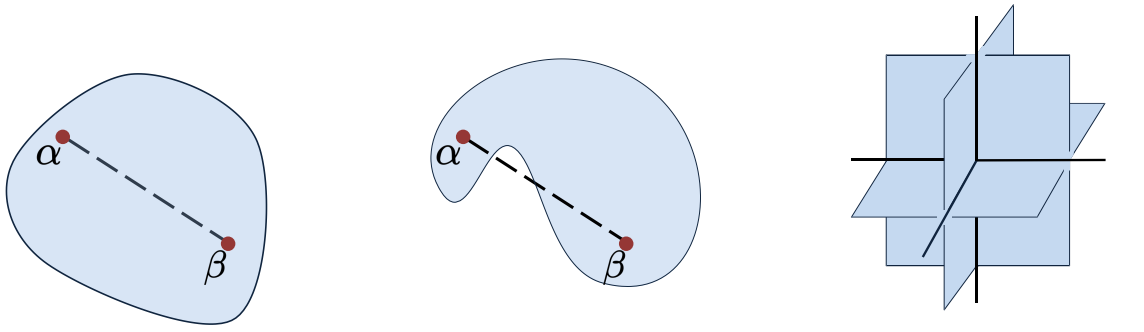
集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 的锥包是包含 $\mathbb{S}$ 的最小**凸锥**。

定义 2.26（锥包）。 $\mathbb{S}$ 的锥包记作 cone($\mathbb{S}$)，它由 $\mathbb{S}$中所有向量的锥组合构成：

$$\operatorname{cone}(\mathbb{S})\triangleq\left\{\sum_{i=1}^k\lambda_i\boldsymbol{\alpha}_i\mid\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_k\in\mathbb{S},\boldsymbol{\lambda}\in\mathbb{R}_+^k,k\in\mathbb{N}\right\}.$$

显然，零向量总是属于任意锥。若对于所有 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{S}$ 和 a，$b\geq 0$，锥组合 $a\boldsymbol{\alpha}+b\boldsymbol{\beta}\in\mathbb{S}$，则集合 $\mathbb{S}$ 是凸锥。

例 2.27（凸锥）。凸锥的常见例子包括：对所有 $i\in\{1,2,\ldots,p\}\}$，
$\mathbb{R}_+^p=\{\boldsymbol{\beta}\in\mathbb{R}^p\mid\beta_i\geq 0$的正正交象限，以及半正定矩阵



$\mathbb{S} \subseteq \mathbb{R}^p$ $\mathbb{S} \subseteq \mathbb{R}^p$ $\mathbb{B}_0(\mathbf{0}, 1) \subseteq \mathbb{R}^3$

凸集 **非凸集** **非凸集**

图 2.3： 如果一个集合包含其所有点的凸组合，那么该集合就是凸的。如果集合中哪怕只有一个点的凸组合落在集合之外，那么该集合就不是凸集。因此，凸集必须“没有鼓包”，也不能有向内的凹陷。值得注意的是，稀疏向量集合不满足这一性质，因此是非凸集。图示改编自Jain 等人(2017年)。

在 中的矩阵。凸锥的另一个重要例子是二阶锥（也称为洛伦兹锥）：

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^{p+1} \mid \sqrt{\textstyle\sum_{i=1}^{p} \beta_i^2} \leq \beta_{p+1}\right\}.$$ □

定义2.28（对偶锥和极锥）。设 $\mathbb{S} \subset \mathbb{R}^p$ 为一个锥。其对偶锥j，记作 $\mathbb{S}^*$，定义如下

$$\mathbb{S}^* \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle \geq 0 \text{ for all } \boldsymbol{\alpha} \in \mathbb{S}\}. \tag{2.6}$$

与对偶锥密切相关的是极锥，记作 $\mathbb{S}^\circ$，其定义如下：

$$\mathbb{S}^\circ \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle \leq 0 \text{ for all } \boldsymbol{\alpha} \in \mathbb{S}\} = -\mathbb{S}^*. \tag{2.7}$$

由于对偶锥定义为若干半空间的交集， $\mathbb{S}^*$ 总是闭且凸的，并且不难验证 $\mathbb{S}^*$ 本身是一个锥。如果 $\mathbb{S}$ 是闭锥，那么$\mathbb{S}^{**} = \mathbb{S}$。此外，如果 $\mathbb{T}$， $\mathbb{S} \subset \mathbb{R}^p$ 是满足 $\mathbb{T} \subset \mathbb{S}$的锥，那么 $\mathbb{S}^* \subset \mathbb{T}^*$。

例2.29（对偶锥）。正正交象限 $\mathbb{R}^p_+$ 的对偶锥就是其本身 $\mathbb{R}^p_+$ ；换言之， $\mathbb{R}^p_+$ 是自对偶的。□

现在我们来讨论极点的概念。

定义2.30（极点）。设 $\mathbb{S} \subset \mathbb{R}^p$ 为一个凸集。若点 $\boldsymbol{z} \in \mathbb{S}$ 满足：对于任意 $\boldsymbol{\alpha}$， $\boldsymbol{\beta} \in \mathbb{S}$ 和 $\lambda \in (0, 1)$，只要 $\boldsymbol{z} = \lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}$，就有 $\boldsymbol{z} = \boldsymbol{\alpha} = \boldsymbol{\beta}$，则称 $\boldsymbol{z} \in \mathbb{S}$ 为 $\mathbb{S}$ 的一个极点。

紧（即闭且有界）的凸集可以通过其极点得到完整刻画，具体如下所述。

定理 2.31：（通过极点表示的紧凸集）每个紧（即闭且有界）凸集都等于其极点的凸包。

证明 参见Rockafellar (1970年)，第 18 节。 ■

例如，如果 $\mathbb{S}$ 是 $\mathbb{R}^2$中的一个多边形，它的极点恰好就是它的顶点（零维面），该定理与多边形由其角点“构成”的直观理解相吻合。

2.3. 凸函数

一个密切相关的概念是凸函数，它在凸组合下表现出特定的不等式性质。在介绍凸函数之前，我们先说明，我们研究的是扩展实值函数 $f: \mathbb{R}^p \to \widetilde{\mathbb{R}} \triangleq (-\infty, \infty] = \mathbb{R} \cup \{+\infty\}$ [2]。算术运算以及$-\infty, \infty]$ 中的不等式均按自然方式解释；例如，$\alpha + \infty = \infty$ 对所有 $\alpha \in \mathbb{R}$，$\lambda \cdot \infty = \infty$ 当 $\lambda > 0$，$\alpha < \infty$ 对所有 $\alpha \in \mathbb{R}$，$\infty \leq \infty$。扩展实值函数的定义域 f 定义为

$$\operatorname{dom}(f) = \{\boldsymbol{\beta} \in \mathbb{R}^p \mid f(\boldsymbol{\beta}) \neq \infty\}.$$

定义在 $\operatorname{dom}(f) \neq \varnothing$ 上的函数称为适当：也就是说，(i) 函数不会取值 $-\infty$；(ii) 函数并非处处取值 $+\infty$。定义在子集 $\mathbb{S} \subset \mathbb{R}^p$ 上的函数$f: \mathbb{S} \to \mathbb{R}$，可以通过令 $f(\boldsymbol{\beta}) = \infty$（当 $\boldsymbol{\beta} \notin \mathbb{S}$ 时）扩展为扩展实值函数。该扩展满足 $\operatorname{dom}(f) = \mathbb{S}$，称为 f 的典范延拓。

现在回顾凸函数的定义。

定义 2.32（凸函数）。定义在凸集 $f: \mathbb{S} \to \mathbb{R}$ [a] 上的函数 $\mathbb{S} \subseteq \mathbb{R}^p$ 称为凸函数，如果

$$f(\lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}) \leq \lambda f(\boldsymbol{\alpha}) + (1-\lambda) f(\boldsymbol{\beta}), \text{ for any } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}, \lambda \in [0,1].$$

此外，f 称为严格凸函数，如果

$$f(\lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}) < \lambda f(\boldsymbol{\alpha}) + (1-\lambda) f(\boldsymbol{\beta}), \text{ for any } \boldsymbol{\alpha} \neq \boldsymbol{\beta} \in \mathbb{S}, \lambda \in (0,1).$$

a。或 $f: \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$，且 $\operatorname{dom}(f) \equiv \mathbb{S}$ 为凸集。

函数 f 称为凹（分别为严格凹），如果 $-f$ 是凸函数（分别为严格凸函数）。注意，任何凸函数的定义域必然是凸集，而定义在凸子集 $\mathbb{S} \subset \mathbb{R}^p$ 上的函数$f: \mathbb{S} \to \mathbb{R}$ 称为凸函数，当且仅当它在 $\mathbb{R}^p$ 上的典范延拓是凸函数。

一个等价且具有几何直观性的刻画是：函数 f 是凸函数，当且仅当它的上图集

$$\operatorname{epi}(f) \triangleq \{(\boldsymbol{\alpha}, r) \in \mathbb{S} \times \mathbb{R} \mid f(\boldsymbol{\alpha}) \leq r\} \subseteq \mathbb{R}^p \times \mathbb{R} \tag{2.8}$$

是 $\mathbb{R}^{p+1}$ 的一个凸集。

例 2.33（凸、严格凸）。令 $f(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \boldsymbol{X} \boldsymbol{\alpha} + \boldsymbol{y}^\top \boldsymbol{\alpha} + z$。则该函数为

- 当且仅当 $\boldsymbol{X} \succeq \mathbf{0}$ 时为凸；当且仅当 $\boldsymbol{X} \succ \mathbf{0}$ 时为严格凸。当且仅当
- $\boldsymbol{X} \preceq \mathbf{0}$ 时为凹；当且仅当 $\boldsymbol{X} \prec \mathbf{0}$ 时为严格凹。

2。虽然扩展实数轴通常定义为 $\widehat{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$，但为简化对适当函数的处理，本文采用约定 $\widetilde{\mathbb{R}} = (-\infty, \infty]$。

如果使用凸性的二阶刻画，证明就很容易（见定理2.47）。□

习题 2.34（方向导数的凸性）。设函数 f是连续可微凸函数。证明，方向导数（定义2.1） $g(\boldsymbol{d}) \triangleq f'(\boldsymbol{\alpha};\ \boldsymbol{d})$ 也是凸函数。

习题2.35（二次函数的凸性）。设 $f(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{y}^\top \boldsymbol{\alpha} + z$，其中 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 是对称的，$\boldsymbol{y} \in \mathbb{R}^p$，且 $z \in \mathbb{R}$。证明，$f(\boldsymbol{\alpha})$ 是凸的（分别地，严格凸的），当且仅当 $\boldsymbol{X} \succeq \boldsymbol{0}$（分别地，$\boldsymbol{X} \succ \boldsymbol{0}$）。

习题 2.36（示性函数的凸性）。证明，示性函数（也称为特征函数） $\delta_{\mathbb{S}}(\boldsymbol{\alpha}) : \mathbb{R}^p \to \mathbb{R}$，当 $\boldsymbol{\alpha}$ 属于集合$\mathbb{S}$ 时取值为 0，否则取值为 $+\infty$，是凸函数，当且仅当集合 $\mathbb{S}$ 是凸集。

习题 2.37（范数函数的凸性）。证明范数函数 $\|\boldsymbol{\alpha}\|_2^2$ 和 $\|\boldsymbol{\alpha}\|_1$ 是凸的。更一般地，证明

- ℓ_s-范数函数 $\|\cdot\|_s$ 在 $1 < s < \infty$时是严格凸的，而在 $s = 1$ 或 $s = \infty$时**不**是严格凸的。对于非递减凸函数 $f\colon\ \mathbb{R} \to (-\infty,\ \infty]$，以及定义在 $\mathbb{R}^p$上的范数 $\|\cdot\|$，函数 $g(\boldsymbol{\alpha}) = f(\|\boldsymbol{\alpha}\|)$是凸的。特别地，只要 $k \geq 1$，函数 $\boldsymbol{\alpha} \mapsto \|\boldsymbol{\alpha}\|^k$ 就是凸的。

上面的习题表明，标准范数函数是凸的。然而，所谓的“ℓ_0-范数”——定义为向量中非零项的数量——并非凸函数。尽管名称如此，它并非真正的范数，因为它不满足正齐次性：对于任意非零的 $\boldsymbol{\alpha} \in \mathbb{R}^p$，都有 $\|2\boldsymbol{\alpha}\|_0 = \|\boldsymbol{\alpha}\|_0 \neq 2\|\boldsymbol{\alpha}\|_0$；见定义1.24。

引理 2.38：（ℓ_0-范数不是凸函数）将 ℓ_0-范数函数定义为向量中非零项的数量，它不是凸函数。

证明 [引理2.38] 考虑一个简单的反例：$\mathbb{R}^2$。令 $\boldsymbol{\alpha} \triangleq \begin{bmatrix}1\\0\end{bmatrix}$ 且 $\boldsymbol{\beta} \triangleq \begin{bmatrix}0\\1\end{bmatrix}$。

那么，对于任意 $\lambda \in (0,\ 1)$，

$$\|\lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}\|_0 = 2 > 1 = \lambda\|\boldsymbol{\alpha}\|_0 + (1-\lambda)\|\boldsymbol{\beta}\|_0 .$$

这违反了凸性的定义。同样的思路可直接推广到 $\mathbb{R}^p$，从而完成证明。

■

矩阵的秩 作为其元素的函数来看，也不是凸函数。

引理 2.39：（秩不是凸的）秩函数 $\operatorname{rank}(\cdot)\colon\ \mathbb{R}^{p\times p} \to \mathbb{R}$ 在方阵空间上不是凸函数。

证明 [引理 2 的39] 我们给出一个反例，类似于引理证明中所用的反例 2.38。令

$$\boldsymbol{X} \triangleq \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \qquad \boldsymbol{Y} \triangleq \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}.$$

对于任意 $\lambda \in (0,\ 1)$,

$$\operatorname{rank}\left(\lambda \boldsymbol{X}+(1-\lambda) \boldsymbol{Y}\right)=2>1=\lambda \operatorname{rank}(\boldsymbol{X})+(1-\lambda) \operatorname{rank}(\boldsymbol{Y}).$$

将这个反例扩展到 $\mathbb{R}^{p \times p}$ 即可完成证明。 ■

2.3.1 凸函数的性质

下面给出一个由凸函数定义推导出的著名不等式，证明从略。

定理 2.40：（詹森不等式）设 $f: \mathbb{S} \rightarrow \mathbb{R}$ 是定义在凸子集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上的凸函数。对于任意有限点序列 $\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \ldots\ldots,\ \boldsymbol{\alpha}_k \in \mathbb{S}$ 以及任意非负权重序列 $\lambda_1,\ \lambda_2,\ \lambda_3,\ \ldots\ldots,\ \lambda_k$，只要满足 $\sum_{i=1}^k \lambda_i=1$，詹森不等式表明:

$$f\left(\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i\right) \leq \sum_{i=1}^k \lambda_i f(\boldsymbol{\alpha}_i).$$

如果 f 是凹函数，则不等号方向相反:

$$f\left(\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i\right) \geq \sum_{i=1}^k \lambda_i f(\boldsymbol{\alpha}_i).$$

在概率论中，如果$\mathbf{x}$ 是一个取值于 $\mathbb{S}$ 的随机向量，且 f 是一个凸函数，那么詹森不等式的形式为:

$$f(\mathbb{E}[\mathbf{x}]) \leq \mathbb{E}[f(\mathbf{x})],$$

其中， $\mathbb{E}[\cdot]$ 表示关于随机向量$\mathbf{x}$的期望算子。对于凹函数，不等式方向同样反转。

接下来，我们考虑凸集上的凸函数的最大值。

定理 2.41：（最大值的达到）设$\mathbb{S} \subset \mathbb{R}^p$ 是一个紧（闭且有界）凸集， $f: \mathbb{S} \rightarrow \mathbb{R}$ 是一个凸函数。那么， f 在 $\mathbb{S}$的某个极点处达到其最大值。

证明 [定理2.41]设 $\boldsymbol{\alpha} \in \mathbb{S}$使得 $f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\beta})$ 对所有 $\boldsymbol{\beta} \in \mathbb{S}$成立。由于 $\mathbb{S}$是其极点的凸包（定理2.31），我们可以写成 $\boldsymbol{\alpha}=\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i$其中 k，且 $\lambda_i>0$，$\sum_{i=1}^k \lambda_i=1$，并且每个 $\boldsymbol{\alpha}_i$ 都是 $\mathbb{S}$的极点。根据凸性以及 $\boldsymbol{\alpha}$的假设，我们有

$$f(\boldsymbol{\alpha})=f\left(\sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i\right) \leq \sum_{i=1}^k \lambda_i f(\boldsymbol{\alpha}_i) \leq \sum_{i=1}^k \lambda_i f(\boldsymbol{\alpha})=f(\boldsymbol{\alpha}).$$

因此，这两个不等式都是等式，这意味着对于所有 $i=1,\ 2,\ \ldots\ldots,\ k$， $f(\boldsymbol{\alpha}_i)=f(\boldsymbol{\alpha})$。因此，$f$ 的最大值在 $\mathbb{S}$ 的某个极点处取得。 ■

如果二元函数 $f(\boldsymbol{\alpha},\ \boldsymbol{\beta})$ 关于变量 $\boldsymbol{z}=(\boldsymbol{\alpha},\ \boldsymbol{\beta})$ 是凸函数，我们就称它关于两个自变量 $\boldsymbol{\alpha}\in\mathbb{R}^q$、$\boldsymbol{\beta}\in\mathbb{R}^p$ 联合凸。如下所述，对联合凸函数的一组变量进行部分最小化会保持其凸性。

定理 2.42：（部分下确界下的保持性）设 $f:\ \mathbb{R}^q\times\mathbb{R}^p\to(-\infty,\ \infty]$是一个联合凸函数。则函数 $g(\boldsymbol{\alpha})\triangleq\min_{\boldsymbol{\beta}\in\mathbb{R}^p}f(\boldsymbol{\alpha},\ \boldsymbol{\beta}),\quad \boldsymbol{\alpha}\in\mathbb{R}^q$，是凸函数。

证明 [定理 2.42] 为简单起见，假设对于每个 $\boldsymbol{\alpha}$<style id='12}，下确界都能取到。一般情形需要使用 ε-论证来处理。给定 $\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2\in\mathbb{R}^q$，存在$\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2\in\mathbb{R}^p$ 使得

$$f(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)=\min_{\boldsymbol{\beta}\in\mathbb{R}^p}f(\boldsymbol{\alpha}_1,\boldsymbol{\beta}),\quad f(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2)=\min_{\boldsymbol{\beta}\in\mathbb{R}^p}f(\boldsymbol{\alpha}_2,\boldsymbol{\beta}).$$

对于任意 $\lambda\in[0,1]$，对 g 的最小化以及 f 的联合凸性意味着

$$\begin{aligned}g(\lambda\boldsymbol{\alpha}_1+(1-\lambda)\boldsymbol{\alpha}_2)&\le f(\lambda\boldsymbol{\alpha}_1+(1-\lambda)\boldsymbol{\alpha}_2,\lambda\boldsymbol{\beta}_1+(1-\lambda)\boldsymbol{\beta}_2)\\&\le\lambda f(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)+(1-\lambda)f(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2)=\lambda g(\boldsymbol{\alpha}_1)+(1-\lambda)g(\boldsymbol{\alpha}_2).\end{aligned}$$

因此，g 是凸函数，证明完毕。 ■

显然，同样的推理表明，对联合凹函数进行部分最大化会得到凹函数。以下运算也保持凸性。这些结果留作习题。

练习 2.43（凸性的保持）。证明以下结论：

(i) 设 $\boldsymbol{X}\in\mathbb{R}^{p\times p}$ 和 $\boldsymbol{y}\in\mathbb{R}^p$，且设 $f:\ \mathbb{R}^p\to(-\infty,\infty]$ 是一个扩展实值凸函数。那么函数 $g(\boldsymbol{\alpha})=f(\boldsymbol{X}\boldsymbol{\alpha}+\boldsymbol{y})$ 是凸函数。(ii) 设 $f:\ \mathbb{R}\to\mathbb{R}$ 是凸且非递减的，且 $g:\ \mathbb{R}^p\to\mathbb{R}$ 是凸函数。那么函数 $h(\boldsymbol{\alpha})=f(g(\boldsymbol{\alpha}))$ 是凸函数。(iii) 设 $f_1,f_2,\ldots,f_k:\ \mathbb{R}^p\to(-\infty,\infty]$ 是扩展实值凸函数，且设 $\sigma_1,\sigma_2,\ldots,\sigma_k\in\mathbb{R}_+$。那么函数 $f=\sum_{i=1}^k\sigma_i f_i$ 是凸的。(iv) 设 $f_i:\ \mathbb{R}^p\to(-\infty,\infty],i\in\mathbb{I}$ 是扩展实值凸函数，其中 $\mathbb{I}$ 是给定的指标集。那么函数 $f(\boldsymbol{\alpha})=\max_{i\in\mathbb{I}}f_i(\boldsymbol{\alpha})$ 是凸函数。

提示：$\max_{i\in\mathbb{I}}f_i(\boldsymbol{\alpha})$ 的上图集是函数 f_i的上图集的交集。

定理 2.44：（凸函数中的局部即全局）设 $f:\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$（或$f:\ \mathbb{R}^p\to\mathbb{R}\cup\{\infty\}$）在凸集 $\mathbb{S}$上为凸函数。

（i） f 的局部最小值是全局最小值。如果 f 是严格凸函数，则最小值唯一。（ii） f 的极小化解集合是凸集。如果 f 是严格凸函数，则该集合是**单元素集合**。

^定理 2.44] (i) 的 [证明。设 $\boldsymbol{\theta}$ 是 f 在 $\mathbb{S}$上的一个局部最小值。存在一个标量 $\tau>0$，使得对于任意满足 $\boldsymbol{\theta}\in\mathbb{B}[\boldsymbol{\theta},\tau]$的 $\boldsymbol{\theta}\in\mathbb{S}$，都有 $f(\boldsymbol{\theta})\ge f(\boldsymbol{\theta})$。现在令 ^ ^$\gamma\in\mathbb{S}$满足 $\gamma\ne\boldsymbol{\theta}$。只需证明 $f(\gamma)\ge f(\boldsymbol{\theta})$。令 $\lambda\in(0,1]$满足

^ ^ ^ ^ ^ ^$\boldsymbol{\theta}+\lambda(\gamma-\boldsymbol{\theta})\in\mathbb{B}[\boldsymbol{\theta},\tau]$。这样的 λ的一个例子是 $\lambda=\frac{\tau}{\|\hat{\boldsymbol{\theta}}-\gamma\|_2}$。由于 $\boldsymbol{\theta}+\lambda(\gamma-\boldsymbol{\theta})\in\mathbb{B}[\boldsymbol{\theta},\tau]\cap\mathbb{S}$,

可得 $f(\boldsymbol{\theta}) \leq f(\boldsymbol{\theta}+\lambda(\boldsymbol{\gamma}-\boldsymbol{\theta}))$，因此根据詹森不等式（定理 2.40），$f(\boldsymbol{\theta}) \leq f(\boldsymbol{\theta}+\lambda(\boldsymbol{\gamma}-\boldsymbol{\theta})) \leq (1-\lambda)f(\boldsymbol{\theta}) + \lambda f(\boldsymbol{\gamma})$。因此，我们得到 $f(\boldsymbol{\theta}) \leq f(\boldsymbol{\gamma})$。

对上述论证稍作修改即可证明，在凸集上，严格凸函数的任何局部极小值实际上都是该函数在此集合上的严格全局极小值。

(ii). 令 $\Theta \triangleq \arg\min\{f(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{S}\}$。若 $\Theta = \varnothing$，结论显然成立。接下来假设 $\Theta \neq \varnothing$，并将最优值记为 f^*。令 $\boldsymbol{\theta}, \boldsymbol{\gamma} \in \Theta$ 为两个不同的极小化点，且 $\lambda \in [0,1]$。于是，根据詹森不等式，$f(\lambda\boldsymbol{\theta}+(1-\lambda)\boldsymbol{\gamma}) \leq \lambda f^* + (1-\lambda)f^* = f^*$，因此 $\lambda\boldsymbol{\theta} + (1-\lambda)\boldsymbol{\gamma}$ 也是一个极小化点，即它属于 Θ，从而证明了 Θ的凸性。现在假设 f 是严格凸的，且 Θ 非空；为证明 Θ是单元素集合，反设存在 $\boldsymbol{\theta}, \boldsymbol{\gamma} \in \Theta$ 使得 $\boldsymbol{\theta} \neq \boldsymbol{\gamma}$。那么$\frac{1}{2}\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\gamma} \in \mathbb{S}$，并且根据 f的严格凸性，我们有

$$f\left(\frac{1}{2}\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\gamma}\right) < \frac{1}{2}f(\boldsymbol{\theta}) + \frac{1}{2}f(\boldsymbol{\gamma}) = \frac{1}{2}f^* + \frac{1}{2}f^* = f^*,$$

这与 f^* 是最优值这一事实矛盾。因此，Θ 必须是单元素集合。

■

凸函数的每个局部最小值自动也是全局最小值，这正是凸优化问题能够采用高效求解方法的关键原因；更多详情请参见第2.9节。

2.3.2 凸函数的刻画

▶ **凸函数的一阶刻画。** 凸函数不一定可微。然而，当它们可微时，可以用梯度不等式来刻画。

定理 2.45：（梯度不等式） 设 $f: \mathbb{S} \to \mathbb{R}$ 是定义在凸集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上的连续可微函数。那么，f 在 $\mathbb{S}$ 上是凸函数，当且仅当

$$f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}) \leq f(\boldsymbol{\beta}), \text{ for any } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}. \tag{2.9a}$$

类似地，该函数在 $\mathbb{S}$ 上严格凸，当且仅当

$$f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}) < f(\boldsymbol{\beta}), \text{ for any } \boldsymbol{\alpha} \neq \boldsymbol{\beta} \in \mathbb{S}. \tag{2.9b}$$

从几何上看，这意味着凸函数的图像在任意点处都位于其切超平面之上。对于凹函数（或严格凹函数），不等号方向相反。

证明 [定理 2.45]为简洁起见，我们只证明（2.9a）；（2.9b）的证明类似。假设 f是凸的。设 $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \in \mathbb{S}$且 $\lambda \in [0,1]$。若 $\boldsymbol{\alpha} = \boldsymbol{\beta}$，则（2.9a）显然成立。因此，我们假设 $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$。由凸性

$$f(\lambda\boldsymbol{\beta} + (1-\lambda)\boldsymbol{\alpha}) \leq \lambda f(\boldsymbol{\beta}) + (1-\lambda)f(\boldsymbol{\alpha}) \implies \frac{f(\boldsymbol{\alpha} + \lambda(\boldsymbol{\beta}-\boldsymbol{\alpha})) - f(\boldsymbol{\alpha})}{\lambda} \leq f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}).$$

取 $\lambda \to 0^+$，左侧收敛于 f 在 $\boldsymbol{\alpha}$ 沿 $\boldsymbol{\beta}-\boldsymbol{\alpha}$方向的方向导数（定义 2.1），因此有 $f'(\boldsymbol{\alpha}; \boldsymbol{\beta}-\boldsymbol{\alpha}) \leq f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha})$。由于 f是

连续可微，因此由（ $f'(\boldsymbol{\alpha};\ \boldsymbol{\beta}-\boldsymbol{\alpha}) = \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})$，根据（2.1），可得（2.9a）。

反之，假设梯度不等式 (2.9a) 成立。令 $\boldsymbol{\alpha}$、$\boldsymbol{\beta} \in \mathbb{S}$，并令 $\lambda \in [0,1]$。我们将证明 $f(\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}) \leq \lambda f(\boldsymbol{\alpha}) + (1-\lambda) f(\boldsymbol{\beta})$。$\lambda = 0$或 $\lambda = 1$ 成立的情形是显然的。因此，我们假设 $\lambda \in (0,1)$。令 $\boldsymbol{z} \triangleq \lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta} \in \mathbb{S}$。则 $-\frac{\lambda}{1-\lambda}(\boldsymbol{\alpha}-\boldsymbol{z}) = (\boldsymbol{\beta}-\boldsymbol{z})$。结合梯度不等式可得

$$f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^\top(\boldsymbol{\alpha}-\boldsymbol{z}) \leq f(\boldsymbol{\alpha}) \qquad \text{and} \qquad f(\boldsymbol{z}) - \frac{\lambda}{1-\lambda}\nabla f(\boldsymbol{z})^\top(\boldsymbol{\alpha}-\boldsymbol{z}) \leq f(\boldsymbol{\beta}).$$

将第一个不等式乘以 $\frac{\lambda}{1-\lambda}$，并将其与第二个不等式相加，得到 $\frac{1}{1-\lambda} f(\boldsymbol{z}) \leq \frac{\lambda}{1-\lambda} f(\boldsymbol{\alpha}) + f(\boldsymbol{\beta})$，从而证明了 f 的凸性。证明完毕。 ■

定理 2.46：（凸函数梯度的单调性） 设 $f: \mathbb{S} \to \mathbb{R}$ 是定义在凸集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上的连续可微函数。那么，f 在 $\mathbb{S}$ 上为凸函数，当且仅当

$$\big\langle \nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta}), (\boldsymbol{\alpha}-\boldsymbol{\beta}) \big\rangle \geq 0, \quad \text{for all } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}. \tag{2.10a}$$

同理，f 在 $\mathbb{S}$ 上严格凸，当且仅当

$$\big\langle \nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta}), (\boldsymbol{\alpha}-\boldsymbol{\beta}) \big\rangle > 0, \quad \text{for all } \boldsymbol{\alpha} \neq \boldsymbol{\beta} \in \mathbb{S}. \tag{2.10b}$$

证明 [定理2.46] 我们只证明（2.10a）；严格凸的情形可类似证明。假设 f 在 $\mathbb{S}$上为凸函数。于是根据梯度不等式，对任意 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{S}$有

$$f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\beta}) + \nabla f(\boldsymbol{\beta})^\top(\boldsymbol{\alpha}-\boldsymbol{\beta}) \qquad \text{and} \qquad f(\boldsymbol{\beta}) \geq f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}).$$

将这两个不等式相加，函数值便消去，从而得到（2.10a）。

反过来，假设（2.10a）成立，并设 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{S}$。定义 $g(\mu) \triangleq f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha}))$ $\mu \in [0,1]$为一维函数。根据微积分基本定理（定理2.15），有

$$\begin{aligned} f(\boldsymbol{\beta}) = g(1) &= g(0) + \int_0^1 g'(\mu)\,d\mu = f(\boldsymbol{\alpha}) + \int_0^1 (\boldsymbol{\beta}-\boldsymbol{\alpha})^\top \nabla f\big(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha})\big)\,d\mu \\ &= f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}) + \int_0^1 (\boldsymbol{\beta}-\boldsymbol{\alpha})^\top \big(\nabla f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha})) - \nabla f(\boldsymbol{\alpha})\big)\,d\mu \\ &\geq f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}), \end{aligned}$$

其中，最后一个不等式由梯度的单调性得出。这表明 f根据定义2.32为凸函数。

■

▶ **凸函数的二阶刻画**。进一步假设函数是二阶连续可微的，则其凸性可通过其海森矩阵的半正定性（见定义1.17）来刻画

定理 2.47：（凸函数的半正定 Hessian）设 $f:\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$ 是定义在一个**开**凸集 $\mathbb{S}$上的二阶连续可微函数。则 f 是凸函数**当且仅当** $\nabla^2 f(\boldsymbol{\alpha})\succeq\mathbf{0}$ 对任意 $\boldsymbol{\alpha}\in\mathbb{S}$均成立。[a] 此外，**若** $\nabla^2 f(\boldsymbol{\alpha})\succ\mathbf{0}$ 对任意 $\boldsymbol{\alpha}\in\mathbb{S}$均成立，则该函数在 $\mathbb{S}$ [b]上严格凸。

[a]。注意到，可以放宽 $\mathbb{S}$ 的开性，以证明对于任意 $\boldsymbol{\alpha}\in\operatorname{int}(\mathbb{S})$，$\nabla^2 f(\boldsymbol{\alpha})\succeq\mathbf{0}$ 的凸性，其中

$$\operatorname{int}(\mathbb{S})=\{\boldsymbol{\alpha}\in\mathbb{S}\mid\exists\epsilon>0\text{ such that }\mathbb{B}(\boldsymbol{\alpha},\epsilon)\subseteq\mathbb{S}\},\tag{2.11}$$

且 $\mathbb{B}(\boldsymbol{\alpha},\epsilon)$ 表示以 $\boldsymbol{\alpha}$ 为中心、半径为 ϵ 的开球（定义 1.29）。[b]。注意，前一个条件既是充分条件也是必要条件，而后一个条件只是充分条件而非必要条件。例如，$f(x)=x^6$ 是严格凸的，但 $f''(x)=30x^4$ 在 $x=0$处等于零。

定理 [的证明 2.47] 假设对于所有 $\nabla^2 f(\boldsymbol{\alpha})\succeq\mathbf{0}$，都有 $\boldsymbol{\alpha}\in\mathbb{S}$，并令 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{S}$。根据线性近似定理（定理 2.13），存在 $\boldsymbol{\xi}\in[\boldsymbol{\alpha},\boldsymbol{\beta}]$（因此 $\boldsymbol{\xi}\in\mathbb{S}$），使得

$$f(\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\alpha})^\top\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{\beta}-\boldsymbol{\alpha}),\quad\boldsymbol{\xi}\in[\boldsymbol{\alpha},\boldsymbol{\beta}].\tag{2.12}$$

由于 $\nabla^2 f(\boldsymbol{\xi})\succeq\mathbf{0}$，因此有$(\boldsymbol{\beta}-\boldsymbol{\alpha})^\top\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{\beta}-\boldsymbol{\alpha})\geq 0$，由此根据(2.12)可得 $f(\boldsymbol{\beta})\geq f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})$，由定理 2.45可知该函数为凸函数。

反过来，假设 f 在 $\mathbb{S}$上为凸函数。令 $\boldsymbol{\alpha}\in\mathbb{S}$，并令 $\boldsymbol{\beta}\in\mathbb{R}^p$。由于 $\mathbb{S}$ 是开集，因此存在一个标量 $\varepsilon>0$，使得对于任意 $\boldsymbol{\alpha}+\lambda\boldsymbol{\beta}\in\mathbb{S}$，$\lambda\in(0,\varepsilon)$ 都成立。梯度不等式（定理 2.45）表明：

$$f(\boldsymbol{\alpha}+\lambda\boldsymbol{\beta})\geq f(\boldsymbol{\alpha})+\lambda\nabla f(\boldsymbol{\alpha})^\top\boldsymbol{\beta}.\tag{2.13}$$

此外，由二次近似定理（定理2.14），可得：

$$f(\boldsymbol{\alpha}+\lambda\boldsymbol{\beta})=f(\boldsymbol{\alpha})+\lambda\nabla f(\boldsymbol{\alpha})^\top\boldsymbol{\beta}+\frac{\lambda^2}{2}\boldsymbol{\beta}^\top\nabla^2 f(\boldsymbol{\alpha})\boldsymbol{\beta}+o(\lambda^2\|\boldsymbol{\beta}\|_2^2),$$

将其与(2.13)结合，可得不等式 $\frac{\lambda^2}{2}\boldsymbol{\beta}^\top\nabla^2 f(\boldsymbol{\alpha})\boldsymbol{\beta}+o(\lambda^2\|\boldsymbol{\beta}\|_2^2)\geq 0$ 对任意$\lambda\in(0,\varepsilon)$均成立。将后一个不等式除以 λ^2，可得 $\frac{1}{2}\boldsymbol{\beta}^\top\nabla^2 f(\boldsymbol{\alpha})\boldsymbol{\beta}+\frac{o(\lambda^2\|\boldsymbol{\beta}\|_2^2)}{\lambda^2}\geq 0$. 取 $\lambda\to 0^+$，可得 $\boldsymbol{\beta}^\top\nabla^2 f(\boldsymbol{\alpha})\boldsymbol{\beta}\geq 0$ 对任意 $\boldsymbol{\beta}\in\mathbb{R}^p$均成立，这意味着 $\nabla^2 f(\boldsymbol{\alpha})\succeq\mathbf{0}$对任意 $\boldsymbol{\alpha}\in\mathbb{S}$均成立。严格凸性的证明类似。 ■

2.4. 连续函数与闭函数，以及魏尔斯特拉斯定理

由于我们将在第2.8.3节中利用闭函数的概念来分析近端算子的性质——这一步依赖于魏尔斯特拉斯定理（定理2.56）中的闭性条件；而该定理又是研究近端梯度方法在第9.2.2节中收敛性的关键，因此我们在本节回顾相关概念。

连续函数是指其函数值不会发生突变的函数——也就是说，它没有不连续点。更正式地说，连续性意味着输入发生任意小的变化时，输出也会发生任意小的变化。连续函数是微积分和数学分析的基础，并具有若干重要性质。

定义 2.48（连续函数）。设 f： $\mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$。如果函数 f 在点 $\boldsymbol{\beta} \in \mathbb{S}$ 处满足：对于每个 $\epsilon > 0$，存在一个 $\delta > 0$，使得对于 $\mathbb{S}$中的所有 $\boldsymbol{\alpha}$，均有：

$$\forall \epsilon > 0, \exists \delta > 0: \quad \forall \boldsymbol{\alpha} \in \mathbb{S}, \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta \quad \Longrightarrow \quad |f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| < \epsilon. \tag{2.14}$$

换言之，只要选择足够接近 $\boldsymbol{\beta}$ 的 $\boldsymbol{\alpha}$，就能使 $f(\boldsymbol{\alpha})$ 的值任意接近 $f(\boldsymbol{\beta})$。

如果 f 在其定义域 $\mathbb{S}$ 的每一点都连续，那么称 f 在$\mathbb{S}$上连续，或简称为连续。

注 2.49（连续函数的性质）。连续函数具有若干重要性质：

1. 介值定理。如果 f 在闭区间 $[a, b]$ 上连续，且 y 介于 $f(a)$ 和 $f(b)$ 之间，那么存在某个 $c \in [a, b]$，使得 $f(c) = y$。这称为罗尔定理（定理 2.10）。 2. 极值定理。如果 f 在闭区间 $[a, b]$ 上连续，那么 f 在该区间上能取得最大值和最小值。关于这一点将在魏尔斯特拉斯定理（定理 2.56）中进一步讨论。 3. 连续函数的复合。如果 g 在 c 处连续，且 f 在 $g(c)$ 处连续，那么复合函数 $f \circ g$ 在 c 处连续。 4. 算术运算。如果 f 和 g 在 c 处连续，那么 $f + g$、 $f - g$、 fg 和 $\frac{f}{g}$ 也连续（条件是 $g(c) \neq 0$）。 5. 连续性与极限。函数 f 在 c 处连续，当且仅当 $\lim_{x \to c} f(x) = f(c)$。

–

例 2.50（连续函数）。以下是连续函数的示例：

- 多项式函数处处连续。有理函数在其定义域内处处连续（即分母不为零处）。正弦
- 和余弦等三角函数处处连续。指数函数和对数函数在各自的定义域内连续。绝对值函数 $|\alpha|$ 处处连续。
-
-
-

相比之下，常见的不连续点类型包括：

- 可去间断点。函数在某一点有一个空洞，但可以通过“填补”使其连续。跳跃间断点。左极限和右极限存在，但不相等。无穷间断点。函数在某一点趋于正无穷或负无穷。
- 振荡间断点。当输入趋近某一点时，函数不趋近于某个确定值。
-
-

连续函数通常更易于分析和处理，而不连续性可能导致复杂的行为，需要特殊处理。在应用中，识别不连续性的类型对于选择适当的解析或数值策略至关重要。□

定义 2.51（下半连续性（LSC））。函数 $f\colon \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ 称为在 $\boldsymbol{\alpha}^* \in \mathbb{R}^p$ 处下半连续（LSC），如果

$$f(\boldsymbol{\alpha}^*) \leq \liminf_{t\to\infty} f(\boldsymbol{\alpha}^{(t)})$$

对于任意满足当时的序列 $\{\boldsymbol{\alpha}^{(t)}\}_{t\geq 1} \subseteq \mathbb{R}^p$ $\boldsymbol{\alpha}^{(t)} \to \boldsymbol{\alpha}^*$ $t \to \infty$。换言之，当点趋近 $\boldsymbol{\alpha}^*$ 时，函数值的下极限必须大于或等于 $\boldsymbol{\alpha}^*$处的函数值。这意味着函数在 $\boldsymbol{\alpha}^*$处不会突然下降。若函数 $f\colon \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ 在 $\mathbb{R}^p$中的每一点都下半连续，则称其为下半连续。

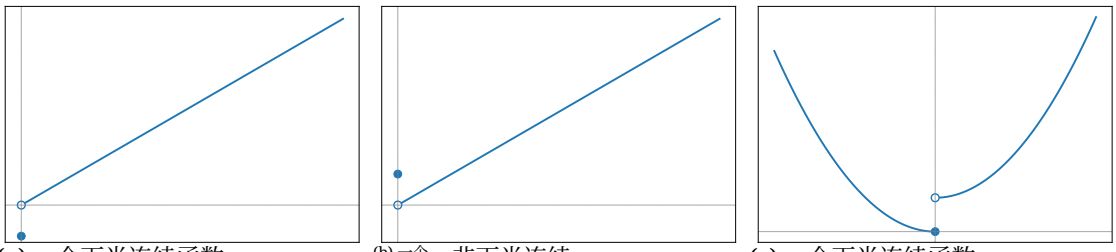
图2.4展示了下半连续函数和非下半连续函数的示例。请注意，在值域中允许出现扩展实值 $\{\infty\}$ 是定义下半连续性的标准做法。虽然并非对所有函数都严格必要，但这种扩展提供了更广泛的适用性，并且在优化和变分分析中至关重要。它使我们能够处理在上方无界的函数（例如约束条件的示性函数），同时仍可讨论其类似连续性的性质。

在优化中，下半连续性发挥着至关重要的作用，因为它有助于确保极小化解的存在。具体而言，如果一列可行点趋近于某个候选解，且目标函数是下半连续的，那么极限点处的函数值不会超过该序列函数值的下极限。在适当的紧性或强制性条件下，这就保证了全局最小值能够取到。

例如，在变分法和最优控制中，人们常通过证明某个泛函既下半连续又具有强制性（即$f(\boldsymbol{\alpha}) \to \infty$ 当 $\|\boldsymbol{\alpha}\|_2 \to \infty$时）来证明极小化解的存在。结合弱紧性论证，即可得到存在性结果（见定理(iv) 中的性质2.56）。

在数值优化和迭代算法中，下半连续性对于分析收敛性也很有价值。例如，如果某个算法为下半连续函数生成一个极小化序列，那么该序列的任一聚点都保证具有不大于函数值下极限的函数值——这通常意味着这些点是驻点或最优点。此外，与强凸性结合时，下半连续性（或闭性）可确保优化器的唯一性和存在性（定理2.82）。

基于这些原因，在本文的分析中，我们主要考虑下半连续（即 闭）函数。



(a) 一个下半连续函数——数。 (b) 一个非下半连续函数 (c) 一个下半连续函数——tion.

图 2.4： 下半连续函数与非下半连续函数。在数值方法和迭代算法中，下半连续性有助于证明收敛性质。例如，如果算法生成的序列正在最小化一个下半连续函数，则可以证明该序列的极限点是驻点。

定义 2.52（闭函数）。若函数 $f: \mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ 的上图集是 $\mathbb{S} \times \mathbb{R}$ 的积拓扑中的一个闭子集，则称其为闭函数。换言之，对于上图集 $\mathrm{epi}(f)$ 中任意一个收敛到极限 $(\boldsymbol{\alpha}^*, r^*)$ 的序列 $(\boldsymbol{\alpha}^{(t)}, r_t)$，都有 $f(\boldsymbol{\alpha}^*) \leq r^*$，这意味着极限点 $(\boldsymbol{\alpha}^*, r^*)$ 也属于该上图集。

习题 2.53（示性函数的闭性）。证明示性函数 $\delta_{\mathbb{S}}$（当 $\boldsymbol{\alpha}$属于集合 $\mathbb{S}$ 时取 0，$+\infty$ 否则取值），是闭函数，当且仅当 $\mathbb{S}$ 是闭集。

下面的习题说明，闭性在几种常见运算下保持不变：仿射变换、非负线性组合以及逐点最大化。

习题 2.54（闭性的保持）。证明以下结论：

（i）设 $\boldsymbol{X} \in \mathbb{R}^{p \times p}$，且设 $f: \mathbb{R}^p \to (-\infty, \infty]$ 是扩展实值闭函数。则函数 $g(\boldsymbol{\alpha}) = f(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{y})$ 是闭函数。（ii）设 $f_1, f_2, \ldots\ldots, f_k: \mathbb{R}^p \to (-\infty, \infty]$ 是扩展实值闭函数，且设 $\sigma_1, \sigma_2, \ldots\ldots, \sigma_k \in \mathbb{R}_+$。则函数 $f = \sum_{i=1}^k \sigma_i f_i$ 是闭函数。（iii）设 $f_i: \mathbb{R}^p \to (-\infty, \infty]$，$i \in \mathbb{I}$ 是扩展实值闭函数，其中 $\mathbb{I}$ 是给定的指标集。则函数 $f(\boldsymbol{\alpha}) = \max_{i \in \mathbb{I}} f_i(\boldsymbol{\alpha})$ 是闭函数。

提示：函数$\max_{i \in \mathbb{I}} f_i(\boldsymbol{\alpha})$ 的上图集是函数 f 的上图集的交集 $_i$.

闭函数的一个重要性质是三个基本概念之间的等价性：下半连续性、函数的闭性（即上图集为闭集），以及其所有次水平集的闭性。

定理 2.55：（闭性、下半连续性与水平集闭性的等价性）设 $f: \mathbb{R}^p \to \mathbb{R} \cup \{-\infty, \infty\}$。那么以下三个陈述等价：

(i) f 是下半连续的（下半连续性）。**(ii)** f 是闭的。**(iii)** 对任意 $\gamma \in \mathbb{R}$，水平集 $\mathrm{Lev}[f, \gamma] = \{\boldsymbol{\alpha} \in \mathbb{R}^p \mid f(\boldsymbol{\alpha}) \leq \gamma\}$ 是闭集。

证明 [定理 2.55](i) $\Longrightarrow$ **(ii)**。假设 f 是下半连续的。我们将证明 $\mathrm{epi}(f)$ 是闭集。设 $\{(\boldsymbol{\alpha}^{(t)}, y^{(t)})\}_{t \geq 1} \subseteq \mathrm{epi}(f)$ 是一个序列，满足 $(\boldsymbol{\alpha}^{(t)}, y^{(t)}) \to (\boldsymbol{\alpha}^*, y^*)$ 当 $t \to \infty$ 时。根据定义，对任意 $t \geq 1$,

$$f(\boldsymbol{\alpha}^{(t)}) \leq y^{(t)}.$$

由于 f 在 $\boldsymbol{\alpha}^*$处下半连续性，我们有

$$f(\boldsymbol{\alpha}^*) \leq \liminf_{t \to \infty} f(\boldsymbol{\alpha}^{(t)}) \leq \liminf_{t \to \infty} y^{(t)} = y^*,$$

这表明 $(\boldsymbol{\alpha}^*, y^*) \in \mathrm{epi}(f)$，从而证明 f是闭的。

(ii) $\Longrightarrow$ **(iii)**。假设 f 是闭的，即上图集 $\mathrm{epi}(f)$ 是闭集。对于任意 $\gamma \in \mathbb{R}$，我们旨在证明 $\mathrm{Lev}[f, \gamma]$ 是闭的。若 $\mathrm{Lev}[f, \gamma] = \varnothing$，则结论显然成立。否则，取序列 $\{\boldsymbol{\alpha}^{(t)}\}_{t \geq 1} \subseteq \mathrm{Lev}[f, \gamma]$，其收敛于 $\widehat{\boldsymbol{\alpha}}$。显然，对于任意 t，$(\boldsymbol{\alpha}^{(t)}, \gamma) \in \mathrm{epi}(f)$，并且当 $t \to \infty$时，$(\boldsymbol{\alpha}^{(t)}, \gamma) \to (\widehat{\boldsymbol{\alpha}}, \gamma)$。由 $\mathrm{epi}(f)$ 的闭性可知，$(\widehat{\boldsymbol{\alpha}}, \gamma) \in \mathrm{epi}(f)$，从而证明了 $\widehat{\boldsymbol{\alpha}} \in \mathrm{Lev}[f, \gamma]$ 是闭的这一事实。

(iii) $\Longrightarrow$ **(i)**。假设 f 的所有水平集都是闭的，我们证明它是下半连续的。反设 f 不是下半连续的，即存在 $\boldsymbol{\alpha}^* \in \mathbb{R}^p$ 和 $\{\boldsymbol{\alpha}^{(t)}\}_{t\geq 1} \subseteq \mathbb{R}^p$，使得 $\boldsymbol{\alpha}^{(t)} \to \boldsymbol{\alpha}^*$ 且 $\liminf_{t\to\infty} f(\boldsymbol{\alpha}^{(t)}) < f(\boldsymbol{\alpha}^*)$。取满足以下条件的 γ

$$\liminf_{t\to\infty} f(\boldsymbol{\alpha}^{(t)}) < \gamma < f(\boldsymbol{\alpha}^*). \tag{2.15}$$

于是存在一个子序列 $\{\boldsymbol{\alpha}_{n_t}\}_{t\geq 1}$ 使得 $f(\boldsymbol{\alpha}_{n_t}) \leq \gamma$ 对所有 $t \geq 1$ 均成立。根据水平集 Lev 的闭性 $[f,\gamma]$ 以及 $\boldsymbol{\alpha}_{n_t} \to \boldsymbol{\alpha}^*$ 当 $t \to \infty$ 时这一事实，可得$f(\boldsymbol{\alpha}^*) \leq \gamma$，这与 (2.15) 矛盾，从而证明 (iii) 蕴含 (i)。 ■

为保证全局极小点（也称全局最优点）的存在——即函数在给定集合上取得最小值的点——经典的魏尔斯特拉斯定理给出了充分条件。下面我们介绍魏尔斯特拉斯定理及其针对连续函数和真闭函数的若干推广。

定理 2.56：（魏尔斯特拉斯定理及其变体）我们将考察适用于不同类型函数和集合的魏尔斯特拉斯定理及其变体：

(i) 设 $f: \mathbb{S} \to (-\infty, \infty]$ 是定义在非空紧集（闭且有界） $\mathbb{S} \subseteq \mathbb{R}^p$上的**连续函数**。那么， f 在 $\mathbb{S}$上存在全局最小点和全局最大点。(ii) 设 $f: \mathbb{S} \to (-\infty, \infty]$ 是定义在非空闭集 $\mathbb{S} \subseteq \mathbb{R}^p$上的**连续函数**。假设所有水平集 $\text{Lev}[f,\gamma] = \{\boldsymbol{\alpha} \in \mathbb{S} \mid f(\boldsymbol{\alpha}) \leq \gamma\}$都有界。那么， f 在 $\mathbb{S}$上存在全局最小点。(iii) 设 $f: \mathbb{S} \to (-\infty, \infty]$ 是**连续强制函数**，且 $\mathbb{S} \subseteq \mathbb{R}^p$ 是非空闭集。那么， f 在 $\mathbb{S}$上存在全局最小点。（强制性保证函数在子集上有界。）（iv）设 $f: \mathbb{S} \to (-\infty, \infty]$ 是**正常闭函数**，并且满足以下条件之一：

(a) f 定义在一个非空紧集 $\mathbb{S}$上。(b) 存在一个非空有界水平集$\text{Lev}[f,\gamma] = \{\boldsymbol{\alpha} \in \mathbb{S} \mid f(\boldsymbol{\alpha}) \leq \gamma\}$。(c) f是强制的，即 $\lim_{\|\boldsymbol{\alpha}\|_2\to\infty} f(\boldsymbol{\alpha}) = \infty$。

那么， f 在 $\mathbb{S}$上存在一个全局最小点。$\min_{\boldsymbol{\alpha}\in\mathbb{S}} f(\boldsymbol{\alpha})$的极小化点集$\{\boldsymbol{\alpha} \in \mathbb{S} \mid f(\boldsymbol{\alpha}) \leq f(\boldsymbol{\beta}) \ \forall \boldsymbol{\beta} \in \mathbb{S}\}$ 非空且紧。

（iii）中的强制性（iii）确保函数在某个子集上有界。（iv）的三个条件（iv）。（a）~（iv）。（c）基本上确保函数 f（$\boldsymbol{\alpha}$）的最小值无法在无穷远处取到。（iv）中的结果（iv）在优化分析中有更广泛的应用。举一个反例，设 f =$\exp(-x)$定义在 $\mathbb{R}$上，它是一个正常闭函数。然而，函数的定义域及其某些水平集 f 并非有界，而且 f 不是强制的。因此，无法保证全局最优点能够取到。

2.5. 次梯度与共轭函数

凸函数的梯度不等式具体适用于连续可微函数。然而，利用次梯度这一概念，可以对其进行推广；次梯度在优化中发挥着关键作用，尤其适用于不可微或非光滑函数。

定义 2.57（次梯度与次微分）。设 $f\colon\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$。向量$\boldsymbol{g}\in\mathbb{R}^p$ 称为次梯度，它是 f 在 $\boldsymbol{\alpha}\in\mathbb{S}$ **处的次梯度，当且仅当**

$$f(\boldsymbol{\beta})\geq f(\boldsymbol{\alpha})+\langle\boldsymbol{g},\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle\,,\quad \text{for all } \boldsymbol{\beta}\in\mathbb{S}.$$

这个不等式称为次梯度不等式。 f在 $\boldsymbol{\alpha}$处的所有次梯度构成的集合，记作 $\partial f(\boldsymbol{\alpha})$，称为 f 在 $\boldsymbol{\alpha}$处的次微分：

$$\partial f(\boldsymbol{\alpha})\triangleq\{\boldsymbol{g}\in\mathbb{R}^p\mid f(\boldsymbol{\beta})\geq f(\boldsymbol{\alpha})+\langle\boldsymbol{g},\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle\,,\ \text{for all } \boldsymbol{\beta}\in\mathbb{S}\}\,.$$

f 在 $\boldsymbol{\alpha}$ 处的任意次梯度可记为

$$f'(\boldsymbol{\alpha})\in\partial f(\boldsymbol{\alpha}).$$

值得注意的是，次梯度的概念不仅适用于凸函数，也适用于非凸情形。然而，对于凸函数，可以保证其在定义域内部任意一点处的次微分非空（定理2.59）。如果 f 在 $\boldsymbol{\alpha}$处可微，则次微分只包含梯度：

$$\partial f(\boldsymbol{\alpha})=\{\nabla f(\boldsymbol{\alpha})\},$$

习题 2.58（范数的次微分）。设 $f(\boldsymbol{\alpha})=\|\boldsymbol{\alpha}\|_2$。证明

$$\partial f(\boldsymbol{\alpha})=\begin{cases}\left\{\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_2}\right\}, & \boldsymbol{\alpha}\neq\mathbf{0};\\ \mathbb{B}_2[\mathbf{0},1], & \boldsymbol{\alpha}=\mathbf{0},\end{cases}$$

其中 $\mathbb{B}_2[\mathbf{0},1]$ 表示 ℓ_2-范数下的闭单位球。此外，设 $g(\boldsymbol{\alpha})=\|\boldsymbol{\alpha}\|_1$。证明 $\partial g(\boldsymbol{\alpha})=\sum_{i=1}^p\partial g_i(\boldsymbol{\alpha})$，其中 $g_i(\boldsymbol{\alpha})=|\alpha_i|$ 且

$$\partial g_i(\boldsymbol{\alpha})=\begin{cases}\{\operatorname{sign}(\alpha_i)\boldsymbol{e}_i\}, & \alpha_i\neq 0;\\ [-\boldsymbol{e}_i,\boldsymbol{e}_i], & \alpha_i=0.\end{cases}$$

令 $\boldsymbol{u}\in\partial g(\boldsymbol{\alpha})$。我们有 $\|\boldsymbol{u}\|_\infty\leq 1$ 和 $\boldsymbol{u}^\top\boldsymbol{\alpha}=\|\boldsymbol{\alpha}\|_1$。这也表明

$$\operatorname{sign}(\boldsymbol{\alpha})\in\partial g(\boldsymbol{\alpha}),$$

其中 $\operatorname{sign}(\boldsymbol{\alpha})$ 返回每个分量的符号。由此也可得出绝对值函数的次梯度不等式。具体而言，对于任意 $a\neq 0$ 和任意 $b\in\mathbb{R}$，我们有

$$|a+b|\geq|a|+\operatorname{sign}(a)b,$$

当且仅当 $|a+b|-|a|$ 关于 b是线性的，即当 b 不会导致符号变化或穿过零点时，等号成立。

凸分析中的一个经典结果指出，凸集的相对内部总是非空的。此外，对于适当凸函数，在其有效定义域的相对内部中的每个点处，次微分都保证非空。下面的定理总结了这些关键事实，不作证明。

定理 2.59：（相对内部非空性（Rockafellar，1970年））设 $\mathbb{S}\subseteq\mathbb{R}^p$ 是一个非空凸集。那么，$\mathbb{S}$的相对内部，记作 $\operatorname{relint}(\mathbb{S})$，是

非空。此外，设 $f:\ \mathbb{R}^p \to (-\infty,\ \infty]$ 是一个适当凸函数，且 $\boldsymbol{\alpha} \in$relint(dom(f))。那么，f在 $\boldsymbol{\alpha}$ 处的次微分，记作 $\partial f(\boldsymbol{\alpha})$，非空。

费马定理，也称为驻点处的费马定理，是微积分和数学优化中的一项基本结果。它为可微函数在其定义域的内点处取得局部最优解（即局部最大值或局部最小值）提供了一个必要条件。在单变量情形中，费马定理给出了区间内部某个点的最优性条件——也就是说，适用于约束集为开区间的一维约束优化问题。

定理2.60：（费马定理：单变量函数的必要条件）设 $f:(a,b) \to \mathbb{R}$ 是定义在开区间 (a, b) 上的一维可微函数。若点 $\alpha^* \in (a,b)$（即 $\alpha^* \in \text{int}([a,b])$）是 f 的局部最大值或局部最小值，则 $f'(\alpha^*) = 0$。

换句话说，如果可微函数 f 在点α^*处取得局部最大值或局部最小值，那么该点处切线的斜率为零；函数在该点处的导数必为零。需要强调的是，这个条件对于 α^* 达到最优是必要的，但**不** 充分。导数为零但该点既非最大点也非最小点的情形确实存在，例如拐点。此外，费马定理**不** 适用于 f 定义域的边界点，也不适用于 f 不可微的点。

现在我们给出多变量优化中局部最小值的一阶必要条件。

推论 2.61：（最小点的一阶必要条件）设$\curvearrowright f:\ \mathbb{R}^p \to \mathbb{R}$ 是可微函数。如果 $\boldsymbol{\theta}$ 是 f的一个（局部）极小化解，那么

$$\nabla f(\widehat{\boldsymbol{\theta}}) = \mathbf{0}.$$

这样的点称为 f的驻点。

证明 [推论 2.61] 固定一个索引 $i \in \{1, 2, \ldots, p\}$，并定义一维函数 $g(\mu) = f(\widehat{\boldsymbol{\theta}} + \mu \boldsymbol{e}_i)$。注意到 g在 $\mu = 0$处可微，且 $g'(0) = \frac{\partial f}{\partial \theta_i}(\widehat{\boldsymbol{\theta}})$。$\widehat{}$由于 $\boldsymbol{\theta}$ 是 f的局部最小点，因此 $\mu = 0$ 是g的局部最小点，由此根据定理2.60立即可得 $g'(0) = 0$。这个等式恰好与 $\frac{\partial f}{\partial \theta_i}(\widehat{\boldsymbol{\theta}}) = 0$相同。由于这对任意 $i \in \{1, 2, \ldots, p\}$都成立，我们得到 $\nabla f(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$。 ■

大多数目标函数，尤其是具有多个局部极小值的目标函数，还包含局部极大值和其他满足费马定理必要条件的临界点。为了将真正的局部极小值与这些无关或非最优的临界点区分开来，我们需要额外的判据，例如二阶条件或凸性假设（Lu，2025年）。次微分对于刻画极小化点特别有用，尤其是在非光滑或不可微情形下。凸优化中的一个基本结果通过次微分给出了全局极小值的完整刻画。从本质上说，这是对可微点处费马最优性条件的推广：$\nabla f\ (\boldsymbol{\alpha}^*)\ = \mathbf{0}$。

定理 2.62：（无约束凸性的必要性/充分性）设 $f:\ \mathbb{R}^p \to (-\infty,\ \infty]$ 是一个适当凸函数。则，

$$\boldsymbol{\alpha}^* \in \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} f(\boldsymbol{\alpha})$$

当且仅当 $\mathbf{0} \in \partial f(\boldsymbol{\alpha}^*)$。

证明 [定理 2.62] 根据次梯度的定义（定义 2.57），有 $\boldsymbol{\alpha}^* \in$ 最小值点集合$_{\boldsymbol{\alpha}\in\mathbb{R}^p} f(\boldsymbol{\alpha})$ 当且仅当

$$f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\alpha}^*) + \langle \mathbf{0}, \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \rangle, \quad \text{for any } \boldsymbol{\alpha} \in \operatorname{dom}(f),$$

这等价于包含关系 $\mathbf{0} \in \partial f(\boldsymbol{\alpha}^*)$，该次微分的存在性由定理2.60给出。

■

现在我们介绍共轭函数的定义及其关键性质。

定义 2.63（共轭函数）。设 $f:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个扩展实值函数。函数 共轭函数 $f^*:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 定义为

$$f^*(\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha}\in\mathbb{R}^p}\{\langle \boldsymbol{\beta}, \boldsymbol{\alpha}\rangle - f(\boldsymbol{\alpha})\}, \quad \boldsymbol{\beta} \in \mathbb{R}^p.$$

注意，$f^*(\boldsymbol{\beta})$ 是关于 $\boldsymbol{\beta}$的仿射函数的逐点最大值，其中每个函数都是凸且闭的。因此，无论原函数 $f^*(\boldsymbol{\beta})$ 是否具有这些性质，其共轭函数 f 始终是凸且闭的（见习题2.43和2.54）。这引出了以下基本结果：

引理 2.64：（共轭函数的闭性和凸性）设 $f:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个扩展实值函数，则其共轭函数 f^* 是闭且凸的。

由定义可直接得到Fenchel不等式。

定理2.65：（Fenchel不等式）设 $f:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当的扩展实值函数。则对于任意 $\boldsymbol{\alpha},\ \boldsymbol{\beta} \in \mathbb{R}^p$,

$$f(\boldsymbol{\alpha}) + f^*(\boldsymbol{\beta}) \geq \langle \boldsymbol{\beta}, \boldsymbol{\alpha}\rangle.$$

双共轭 f^{**}（即 f^* 的共轭函数）在凸分析中发挥着核心作用。

习题2.66（双共轭的上界（Rockafellar，1970年））。设 $f:\ \mathbb{R}^p \to]-\infty, \infty]$是一个扩展实值函数。证明 $f(\boldsymbol{\alpha}) \geq f^{**}(\boldsymbol{\alpha})$ 对任意 $\boldsymbol{\alpha} \in \mathbb{R}^p$均成立。此外，设 $g:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当的**闭且凸**函数。证明 $g(\boldsymbol{\alpha}) = g^{**}(\boldsymbol{\alpha})$

由于这一性质，双共轭 f^{**} 常被称为 f 的凸包络（或凸闭包、凸松弛）——它是位于 f 下方的最大凸下半连续函数。

习题 2.67（共轭函数的正则性）。设 $f\colon \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当凸函数。证明 f^* 也是适当的。

练习 2.68（共轭函数运算）。设 $f\colon \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$。证明

(i) 设 $f_\tau(\boldsymbol{\alpha}) \triangleq f(\tau\boldsymbol{\alpha})$，其中 $\tau \neq 0$。则 $(f_\tau)^*(\boldsymbol{\beta}) = f^*(\boldsymbol{\beta}/\tau)$。(ii) 对于 $\tau > 0$，$(\tau f)^*(\boldsymbol{\beta}) = \tau f^*(\boldsymbol{\beta}/\tau)$。(iii) 设 $f_{(\boldsymbol{\zeta})} \triangleq f(\boldsymbol{\alpha}-\boldsymbol{\zeta})$，其中对于 $\boldsymbol{\zeta}\in\mathbb{R}^p$。则 $(f_{(\boldsymbol{\zeta})})^*(\boldsymbol{\beta}) = f^*(\boldsymbol{\beta}) + \langle \boldsymbol{\zeta}, \boldsymbol{\beta}\rangle$。

我们来计算几个例子的凸共轭。

例 2.69（自共轭）。设 $f(\boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2$ 对于 $\boldsymbol{\alpha}\in\mathbb{R}^p$。则 $f^*(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 = f(\boldsymbol{\beta})$，$\forall \boldsymbol{\beta}\in\mathbb{R}^p$。事实上，由于 $\langle \boldsymbol{\beta}, \boldsymbol{\alpha}\rangle \le \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2 + \frac{1}{2}\|\boldsymbol{\beta}\|_2^2$，我们有

$$f^*(\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha}\in\mathbb{R}^p}\{\langle \boldsymbol{\beta},\boldsymbol{\alpha}\rangle - f(\boldsymbol{\alpha})\} \le \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 .$$

对于反向不等式，只需在凸共轭的定义中令 $\boldsymbol{\alpha} = \boldsymbol{\beta}$，即可得到

$$f^*(\boldsymbol{\beta}) \ge \|\boldsymbol{\beta}\|_2^2 - \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 = \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 .$$

因此，等式成立。该函数是自共轭的：它是在 $\mathbb{R}^p$ 上满足 $f = f^*$ 的唯一函数（在某些情况下，除缩放和平移外）。□

联系次梯度与共轭函数的核心结果称为共轭次梯度定理，它通过凸函数及其芬切尔共轭的次微分，确立了二者之间优美的对偶关系。从本质上说，该定理刻画了一对点 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 互为对偶的条件——不仅意味着它们在 Fenchel不等式 $f(\boldsymbol{\alpha}) + f^*(\boldsymbol{\beta}) \ge \langle \boldsymbol{\alpha},\boldsymbol{\beta}\rangle$ 中取等，而且在几何意义上，一个点是函数在另一个点处的次梯度。这种联系是凸分析的核心，在优化、对偶理论和灵敏度分析中发挥着关键作用。此外，当函数是闭的（即下半连续）时，这种关系具有对称性：f 的次梯度与 f^* 的次梯度一一对应，体现了共轭运算的对合性质。

定理 2.70：（共轭次梯度定理）设 $f\colon \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当的**凸**函数。对于任意 $\boldsymbol{\alpha}, \boldsymbol{\beta}\in\mathbb{R}^p$，以下两个断言等价：

(i) $\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle = f(\boldsymbol{\alpha}) + f^*(\boldsymbol{\beta})$。（ii） $\boldsymbol{\beta}\in\partial f(\boldsymbol{\alpha})$】【。

此外，如果 f 是**闭**的，那么 (i) 和 (ii) 也等价于

(iii) $\boldsymbol{\alpha}\in\partial f^*(\boldsymbol{\beta})$.

或者，设 $f\colon \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当的**闭凸**函数。那么对于任意$\boldsymbol{\alpha}, \boldsymbol{\beta}\in\mathbb{R}^p$,

$$\partial f(\boldsymbol{\alpha}) = \arg\max_{\boldsymbol{u}\in\mathbb{R}^p}\{\langle \boldsymbol{\alpha},\boldsymbol{u}\rangle - f^*(\boldsymbol{u})\}; \tag{2.16a}$$

$$\partial f^*(\boldsymbol{\beta}) = \arg\max_{\boldsymbol{v}\in\mathbb{R}^p}\{\langle \boldsymbol{\beta},\boldsymbol{v}\rangle - f(\boldsymbol{v})\}. \tag{2.16b}$$

证明 [定理 2 的证明。70] 根据 $\boldsymbol{\beta}\in\partial f(\boldsymbol{\alpha})$ 的次梯度不等式，可得

$$f(\boldsymbol{u}) \ge f(\boldsymbol{\alpha}) + \langle \boldsymbol{\beta}, \boldsymbol{u}-\boldsymbol{\alpha}\rangle \quad\Longrightarrow\quad \langle \boldsymbol{\beta},\boldsymbol{\alpha}\rangle - f(\boldsymbol{\alpha}) \ge \langle \boldsymbol{\beta},\boldsymbol{u}\rangle - f(\boldsymbol{u}),\ \text{for all } \boldsymbol{u}\in\mathbb{R}^p.$$

对 $\boldsymbol{u}$ 取最大值，上述不等式等价于

$$\langle \boldsymbol{\beta}, \boldsymbol{\alpha} \rangle - f(\boldsymbol{\alpha}) \geq f^*(\boldsymbol{\beta}),$$

根据 Fenchel 不等式（定理 2.65），这等价于等式 $\langle \boldsymbol{\alpha},\ \boldsymbol{\beta} \rangle = f(\boldsymbol{\alpha}) + f^*(\boldsymbol{\beta})$，从而证明了 (i) 与 (ii) 等价。当 f 为真闭凸函数时，根据习题 2.66，$f^{**} = f$，这意味着 (i) 等价于 $\langle \boldsymbol{\alpha},\ \boldsymbol{\beta} \rangle = f^*(\boldsymbol{\beta}) + f^{**}(\boldsymbol{\alpha})$。根据上述论证，我们得出 (i) 等价于 $\boldsymbol{\alpha} \in \partial f^*(\boldsymbol{\beta}$。证明完毕。

■

2.6. 利普希茨函数

本节将进一步探讨 Lipschitz 连续性的性质，定义相关概念，并考察其对函数行为的影响，尤其是凸函数次微分集的有界性。

定义2.71（Lipschitz函数）。设 $L \geq 0$，且令 $f\colon \mathbb{S} \to \mathbb{R}$ 为定义在集合 $\mathbb{S} \subseteq \mathbb{R}^p$上的函数。那么，如果对于所有 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{S}$，都有

$$|f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| \leq L \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 .$$

如果 f 连续可微，则称其梯度在 $\mathbb{S}$ 上 *Lipschitz* 连续（或者等价地，称 f 是 *Lipschitz*连续可微），常数为 L，若对于所有 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$，均有

$$\|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_2 \leq L \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 .$$ [a]

梯度以常数 L 满足 Lipschitz 连续的函数类记为 ${C^1}_{,L}{}^1(\mathbb{S})$，或简记为 ${C^1}_{,L}{}^1$。

类似地，若对于所有 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$，均有，则称该函数为常数为 L 的二阶*Lipschitz*连续可微函数，记为 ${C^2}_{,L}{}^2(\mathbb{S})$。

$$\left\|\nabla^2 f(\boldsymbol{\alpha}) - \nabla^2 f(\boldsymbol{\beta})\right\|_2 \leq L \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 .$$

更一般地，设 $\mathbb{S} \subseteq \mathbb{R}^p$。我们用 ${C^k}_{,L}{}^s(\mathbb{S})$ 表示满足以下性质的函数类：

- 任意 $f \in {C^k}_{,L}{}^s(\mathbb{S})$ 在 $\mathbb{S}$ 上均为 k 次连续可微。其 s 阶导数在 $\mathbb{S}$ 上 Lipschitz 连续，
- 常数为 L：$\|\nabla^s f(\boldsymbol{\alpha}) - \nabla^s f(\boldsymbol{\beta})\|_2 \leq L \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$，对所有 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$ 均成立。

在本书中，我们主要考虑 $s = 0$、$s = 1$ 和 $s = 2$ 这几种情形。显然，我们总有 $s \leq k$。此外，若 $t \geq k$，则 ${C^t}_{,L}{}^s(\mathbb{S}) \subseteq {C^k}_{,L}{}^s(\mathbb{S})$。为简单起见，我们也将 s$C^k(\mathbb{S})$ 表示为 $\mathbb{S}$ 上的 k 次连续可微函数的集合。

[a] 梯度Lipschitz性可以用对偶范数定义：$\|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_* \leq L \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$，其中 $\|\cdot\|_*$ 是 $\|\cdot\|$ 的对偶范数。值得注意的是，向量 - 范数是自对偶的。在本例中，我们只考虑向量 ℓ_2- 范数。

例 2.72。考虑以下示例：

- 由于 $\left|\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1\right| \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_1 \leq \sqrt{p}\,\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$ 对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^p$ (参见习题 1.26)，函数 $f_1(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$ 因此是 $\sqrt{p}$-Lipschītz。由于
- $\left|\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_2 - \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2\right| \leq \|(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) - (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\|_2 \leq \|\boldsymbol{X}\|_2 \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$，函数 $f_2(\boldsymbol{\alpha}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_2$ 是 $\|\boldsymbol{X}\|_2$-Lipschitz，其中 $\|\boldsymbol{X}\|_2$ 表示 $\boldsymbol{X}$的谱范数。

函数 $f_3(\boldsymbol{\alpha}) \triangleq f_2^2(\boldsymbol{\alpha}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_2^2$不是 Lipschitz 连续的。不过，可以证明 $f_3(\boldsymbol{\alpha})$ 是梯度 Lipschitz 连续可微的，其常数为 $2\left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2$。 □

习题 2.73（二次函数的梯度Lipschitz性）。设 $f(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{y}^\top \boldsymbol{\alpha} + z$，其中 $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 是对称的，$\boldsymbol{y} \in \mathbb{R}^p$，且 $z \in \mathbb{R}$。证明 $f(\boldsymbol{\alpha})$ 的梯度是 $\|\boldsymbol{X}\|_2$-Lipschitz 的。这也说明，当 $\boldsymbol{X} = \boldsymbol{0}$ （一个仿射函数）时，梯度函数是 0-Lipschitz 的。

注 2.74（Lipschitz连续性蕴含连续性）。回顾一下，函数f 在点 $\boldsymbol{\beta}$ （定义 2.48）处连续，是指对于每个 $\epsilon > 0$，都存在一个 $\delta > 0$使得：

$$\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta \implies |f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| < \epsilon.$$

现在，假设 f是 Lipschitz 连续的，常数为 L，即 $|f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| \leq L\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$。给定 $\epsilon > 0$，取 $\delta = \frac{\epsilon}{L}$。那么，[a]如果 $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$，我们有：

$$|f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| \leq L\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < L \cdot \frac{\epsilon}{L} = \epsilon.$$

因此，f 在 $\boldsymbol{\beta}$处连续。由于 $\boldsymbol{\beta}$ 是任意的，f 在 $\mathbb{S}$上连续。[a] Lipschitz 条件限制了函数的“变化率”，防止出现可能破坏连续性的无限斜率或剧烈振荡。

[a]。此外，$\delta = \epsilon/L$ 的选取与 $\boldsymbol{\beta}$无关——它在整个定义域上均适用。这意味着 f 实际上是一致连续的。

习题 2.75. 令 $f_1 \in C^{k}_{,L_1}{}^{s}(\mathbb{S})$，$f_2 \in C^{k}_{,L_2}{}^{s}(\mathbb{S})$，并令 $a_1, a_2 \in \mathbb{R}$。定义 $L_3 = |a_1| \cdot L_1 + |a_2| \cdot L_2$。证明 $a_1 f_1 + a_2 f_2 \in C^{k}_{,L_3}{}^{s}(\mathbb{S})$。

习题 2.76. 证明函数 $f: \mathbb{R}^p \to \mathbb{R}$属于类 $C^{2}_{,L}{}^{1}(\mathbb{R}^p) \subset C^{1}_{,L}{}^{1}(\mathbb{R}^p)$ 当且仅当对所有 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 都有 $\left\|\nabla^2 f(\boldsymbol{\alpha})\right\|_2 \leq L$. [a] 提示：使用微积分基本定理（定理 2.15）。

[a]。由于 $\nabla^2 f(\boldsymbol{\alpha})$ 的谱范数是 $\nabla^2 f(\boldsymbol{\alpha})$ 的最大奇异值，其中 $\nabla^2 f(\boldsymbol{\alpha})$ 是对称的。这也表明 $-L\boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{\alpha}) \preceq L\boldsymbol{I}$。

现在我们来证明 Lipschitz连续性下次微分的有界性。

定理 2.77：（Lipschitz性与次微分集的有界性）设 $f: \mathbb{S} \to \mathbb{R}$ 为一个凸函数，并设 $\mathbb{X} \subseteq \operatorname{int}(\mathbb{S})$。考虑以下两个命题：

(i) $|f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta})| \leq L\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{X}$ 均成立。(ii) $\|\boldsymbol{g}\|_2 \leq L$ 对任意 $\boldsymbol{g} \in \partial f(\boldsymbol{\alpha})$ $\boldsymbol{\alpha} \in \mathbb{X}$ 均成立。于是，

(a) (ii) 推出 $\implies$ (i),

(b) 如果 $\mathbb{X}$ 是开集，则命题 (i) 成立当且仅当命题 (ii) 成立。

证明 [定理 2.77] (a). 假设 (ii) 成立，并令 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{X}$. 令 $\boldsymbol{g}_x \in \partial f(\boldsymbol{\alpha})$ 和 $\boldsymbol{g}_y \in \partial f(\boldsymbol{\beta})$。根据 $\boldsymbol{g}_x, \boldsymbol{g}_y$的定义以及柯西–施瓦茨不等式（定理 1.59），

$$f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta}) \leq \langle \boldsymbol{g}_x, \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle \leq \|\boldsymbol{g}_x\|_2 \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 \leq L \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 ;$$
$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq \langle \boldsymbol{g}_y, \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle \leq \|\boldsymbol{g}_y\|_2 \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 \leq L \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 ,$$

这就证明了 (i) 的有效性。

(b). 蕴含关系 (ii) $\implies$ (i) 已在 (a) 中证明。反过来，假设 (i) 成立。取 $\boldsymbol{\alpha} \in \mathbb{X}$和 $\boldsymbol{g} \in \partial f(\boldsymbol{\alpha})$。令 $\widetilde{\boldsymbol{g}} \triangleq \frac{\boldsymbol{g}}{\|\boldsymbol{g}\|_2}$ 满足 $\|\widetilde{\boldsymbol{g}}\|_2 = 1$和 $\langle \widetilde{\boldsymbol{g}}, \boldsymbol{g} \rangle = \|\boldsymbol{g}\|_2$。由于$\mathbb{X}$ 是开集，我们可以取足够小的 $\epsilon > 0$，使得 $\boldsymbol{\alpha} + \epsilon \widetilde{\boldsymbol{g}} \in \mathbb{X}$。由次梯度不等式可得， $f(\boldsymbol{\alpha} + \epsilon \widetilde{\boldsymbol{g}}) \geq f(\boldsymbol{\alpha}) + \epsilon \langle \boldsymbol{g}, \widetilde{\boldsymbol{g}} \rangle$，因此我们有

$$\epsilon \|\boldsymbol{g}\|_2 \leq f(\boldsymbol{\alpha} + \epsilon \widetilde{\boldsymbol{g}}) - f(\boldsymbol{\alpha}) \leq L \|\boldsymbol{\alpha} + \epsilon \widetilde{\boldsymbol{g}} - \boldsymbol{\alpha}\|_2 = L\epsilon \quad \implies \quad \|\boldsymbol{g}\|_2 \leq L.$$

证明至此完成。 ■

2.7. 强凸、强光滑函数

本节将对强凸性（*SC*）和强光滑性（*SS*，或简称光滑性）给出严格定义。虽然这两个概念都可以借助次梯度推广到不可微凸函数，但本书将讨论范围限制在可微函数。

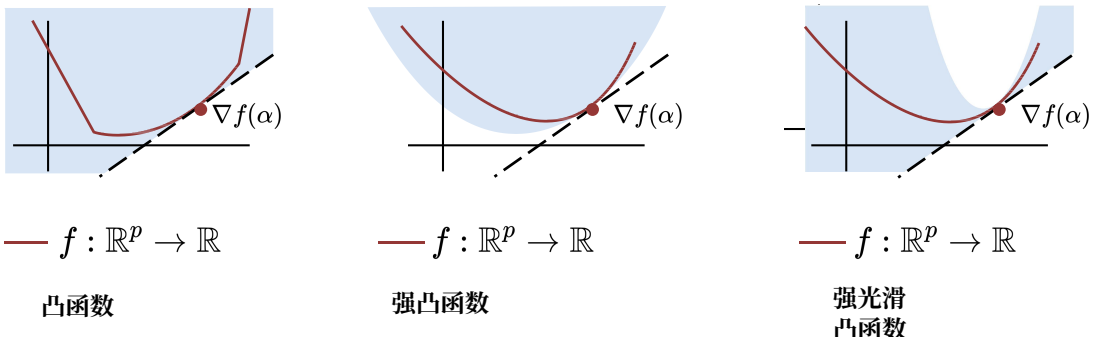


图 2.5： 凸函数始终位于其在任意点处的切线上方。相比之下，强凸函数不仅有切线作为下界，还有二次函数作为下界，这限制了它增长得过慢的程度。同样，光滑函数有二次模型作为上界，从而防止其增长得过快。强凸性与光滑性共同将函数的行为限制在两个二次包络之间。每幅图中的阴影区域表示函数图像允许所在的范围。此图改编自Jain 等人 (2017年)。

定义 2.78（强凸/光滑函数）。设 f： $\mathbb{S} \to \mathbb{R}$ 是在**凸**集合 $\mathbb{S} \subseteq \mathbb{R}^p$ [a]上定义的可微函数。则， f 称为 L_a-强凸（*SC*，或简称 L_a-凸）且 L_b-强光滑（*SS*，或简称 L_b-光滑），如果:

对每个 $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \in \mathbb{S}$，都有

$$\frac{L_a}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2 \leq f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), (\boldsymbol{\beta}-\boldsymbol{\alpha})\rangle \leq \frac{L_b}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2 . \ ^{b} \tag{2.17}$$

当函数二阶可微时，定理 2.47还表明

$$L_a \boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{\alpha}) \preceq L_b \boldsymbol{I}, \quad \text{for all } \boldsymbol{\alpha} \in \operatorname{int}(\mathbb{S}). \ ^{c} \tag{2.18}$$

$\nabla^2 f(\boldsymbol{\alpha})$的最小特征值至少为 L_a，而$\nabla^2 f(\boldsymbol{\alpha})$的最大特征值至多为 L_b:

$$L_a \leq \frac{\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{\alpha}) \boldsymbol{v}}{\|\boldsymbol{v}\|_2^2} \leq L_b, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{S}, \text{and nonzero } \boldsymbol{v} \in \mathbb{R}^p. \ ^{d} \tag{2.19}$$

显然，如果函数是 L_b'-强光滑的，那么对任意$L_b'' \geq L_b'$，它也是 L_b''-强光滑的。 L_a-强凸函数必然是凸函数（当$L_a = 0$时，它就退化为标准凸函数）。而如果函数是 L_a'-强凸的，那么对任意 $L_a'' \in$ ，它也是 L_a''-强凸的（0， L_a']。

---

a。请注意，强凸性和光滑性的概念主要适用于定义在凸集上的可微函数（至少大多适用于可微函数）。例如， $\mathbb{S}$ 可以设为$\mathbb{R}^p$。其原因将在下文中阐明。b。在许多文献中，第二个不等式称为 L_b-强光滑函数的 *descent lemma*，因为 $f(\boldsymbol{\beta}) \leq f(\boldsymbol{\alpha}) + \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}) + \frac{L_b}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2$，即函数值从 $f(\boldsymbol{\alpha})$更新到 $f(\boldsymbol{\beta})$时，其增量有上界。c。我们将定理 2.47 中的开集条件放宽为 $\boldsymbol{\alpha} \in \operatorname{int}(\mathbb{S})$。$d$。 $\frac{\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v}}{\|\boldsymbol{v}\|_2^2}$ 称为向量 $\boldsymbol{v}$ 相对于矩阵 $\boldsymbol{X}$的 *Rayleigh quotient*，其值介于 $\boldsymbol{X}$的最大和最小特征值之间；更多细节可参见 Lu（2021a，2025）。

强凸性和强光滑性是分析基于梯度的优化算法的基础（见第9章）。强凸性确保函数以足够快的速度向上弯曲，为其曲率提供正下界，并防止梯度在极小值点附近过快趋于零。反之，强光滑性对曲率施加上界，确保梯度不会变化得过于剧烈。这限制了梯度下降中允许的步长，并防止更新过于激进（见图2.5）。

定义 2.86 进一步将这些概念扩展到定义在非凸域上的函数。

练习 2.79（强凸函数与凸函数之和、强光滑函数之和）。证明以下两个命题:

- 设 f 是 L_a-强凸函数，且 g 是凸函数。证明 $f+g$ 是 L_a-强凸函数。设 f是 L_b'-
- 强光滑函数，且 g是 L_b''-强光滑函数。证明 $f+g$是 $(L_b' + L_b'')$-强光滑函数。

定理 2.80：（强光滑性质-O：梯度 Lipschitz 性与光滑性的等价性）设 $f: \mathbb{S} \to \mathbb{R}$ 是定义在一个**凸**集合 $\mathbb{S} \subseteq \mathbb{R}^p$上的可微函数。若 f 是梯度 Lipschitz 连续函数，常数为 L_b，即 $f \in C^{1,1}_{L_b}(\mathbb{S})$，则它也是 L_b-强光滑函数。

定理 2 的证明 [80]根据微积分基本定理（方程 (2.5a)），可得 $f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) = \langle \nabla f(\boldsymbol{\alpha}) \ \boldsymbol{\beta}-\boldsymbol{\alpha}\rangle + \int_0^1 \langle \nabla f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha})) - \nabla f(\boldsymbol{\alpha}) \ \boldsymbol{\beta}-\boldsymbol{\alpha}\rangle \, d\mu$.由于 $\mathbb{S}$是一个凸集，函数 $f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha}))$对 $\mu \in [0,1]$有定义。这

得到

$$|f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})-\langle\nabla f(\boldsymbol{\alpha}),\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle| = \left|\int_0^1 \langle\nabla f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha}))-\nabla f(\boldsymbol{\alpha}),\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle\, d\mu\right|$$
$$\overset{\dagger}{\leq} \int_0^1 \|\nabla f(\boldsymbol{\alpha}+\mu(\boldsymbol{\beta}-\boldsymbol{\alpha}))-\nabla f(\boldsymbol{\alpha})\|_2 \cdot \|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2\, d\mu \leq \int_0^1 L_b\mu\,\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2\, d\mu = \frac{L_b}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2,$$

其中，不等式（†）由柯西–施瓦茨不等式（定理 1.59）可得。因此，f 是 L_b-光滑的。

■

如果 f 是 L_b-光滑的并且存在全局极小点 $\boldsymbol{\alpha}^*$，那么一个重要结论是，可以利用光滑性所给出的二次上界，将任意点 $\boldsymbol{\alpha}$处的函数次优性 $f(\boldsymbol{\alpha})-f(\boldsymbol{\alpha}^*)$ 与梯度的范数联系起来。当与强凸性结合时，这一关系尤其有力，因为它可以得到紧的双侧界。

定理2.81：（SS&SC 性质-I：() () 的上界 $f\colon \mathbb{R}^p\to\mathbb{R}$）设 $f\colon \mathbb{R}^p\to\mathbb{R}$ 是定义在 $\mathbb{R}^p$ 上的可微函数，并且具有全局极小点 $\boldsymbol{\alpha}^*$。假设 $f(\boldsymbol{\alpha})$ 是 L_b-光滑且 L_a-强凸的。那么，

$$\frac{1}{2L_b}\|\nabla f(\boldsymbol{\alpha})\|_2^2 \leq f(\boldsymbol{\alpha})-f(\boldsymbol{\alpha}^*) \leq \frac{1}{2L_a}\|\nabla f(\boldsymbol{\alpha})\|_2^2, \qquad \text{for any } \boldsymbol{\alpha}\in\mathbb{R}^p.$$

定理 [的证明2.81] 由于 $\boldsymbol{\alpha}^*$ 是全局极小化解，对于任意 $\boldsymbol{\beta}\in\mathbb{R}^p$，应用光滑性给出的二次上界可得

$$f(\boldsymbol{\alpha}^*) \leq f(\boldsymbol{\beta}) \leq f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{L_b}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2.$$

固定 $\boldsymbol{\alpha}$，注意到上述不等式对任意 $\boldsymbol{\beta}$ 都成立，因此我们可以对不等式右侧取下确界：

$$f(\boldsymbol{\alpha}^*) \leq \min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{L_b}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2\right\} = f(\boldsymbol{\alpha})-\frac{1}{2L_b}\|\nabla f(\boldsymbol{\alpha})\|_2^2,$$

其中，最小值在 $\boldsymbol{\beta}=\boldsymbol{\alpha}-\frac{1}{L_b}\nabla f(\boldsymbol{\alpha})$ 处取得（因为定义域是 $\mathbb{R}^p$）.

对于第二部分，由于 f 是强凸的，我们有

$$\begin{aligned} f(\boldsymbol{\beta}) &\geq f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})+\frac{L_a}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2 \\ &\geq f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\alpha})+\frac{L_a}{2}\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\alpha}\right\|_2^2 = f(\boldsymbol{\alpha})-\frac{1}{2L_a}\|\nabla f(\boldsymbol{\alpha})\|_2^2, \end{aligned}$$

其中 $\widetilde{\boldsymbol{\beta}}\triangleq\boldsymbol{\alpha}-(1/L_a)\nabla f(\boldsymbol{\alpha})$ 使上述不等式右侧达到最小值。令 $\boldsymbol{\beta}\triangleq\boldsymbol{\alpha}^*$，证明完成。

■

上界的一个直接推论是以下实用判据：对于任意容差 $\varepsilon>0$,

$$\|\nabla f(\boldsymbol{\alpha})\|_2 \leq (2L_a\varepsilon)^{1/2} \quad\Longrightarrow\quad f(\boldsymbol{\alpha})-f(\boldsymbol{\alpha}^*)\leq\varepsilon, \tag{2.20}$$

因此，控制梯度范数即可为近似最优性提供证书。这推广了经典的凸函数最优性条件，该条件在定理2.62中陈述，其

指出 $\nabla f(\boldsymbol{\alpha}) = \mathbf{0}$ 刻画了可微情形下的精确极小化解。此外，该定理揭示了一项关键的增长性质：在极小化解附近，强凸且光滑的函数表现得像二次函数——既不会过于平坦，也不会过于陡峭。

下一个结果表明，仅有强凸性（即使不具备光滑性）也能保证极小化解的存在性与唯一性，并且在该解附近具有二次增长条件。

定理 2.82：（强凸性质 II：闭强凸函数极小化点的存在性与唯一性） 设 $f: \mathbb{R}^p \to (-\infty, \infty]$ 是一个可微 [a]、真闭且L_a-强凸函数（$L_a > 0$）。那么，

(i) f 有唯一极小化解。(ii) 对所有 $\boldsymbol{\alpha} \in \mathbb{R}^p$，$f(\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}^*) \geq \frac{L_a}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2^2$，其中 $\boldsymbol{\alpha}^*$是f的唯一极小化解。

[a] 注意，该定理对于不可微函数也成立，并且可以根据定理 2.59，在证明中使用次梯度。

Pro的 [定理 2.82] **(i).** 任取 $\boldsymbol{\alpha}_0 \in \mathbb{R}^p$。由强凸性可知

$$f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\alpha}_0) + \langle \nabla f(\boldsymbol{\alpha}_0), \boldsymbol{\alpha} - \boldsymbol{\alpha}_0 \rangle + \frac{L_a}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2$$

$$\implies f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\alpha}_0) - \frac{1}{2L_a} \|\nabla f(\boldsymbol{\alpha}_0)\|_2^2 + \frac{L_a}{2} \left\| \boldsymbol{\alpha} - \left(\boldsymbol{\alpha}_0 - \frac{1}{L_a} \nabla f(\boldsymbol{\alpha}_0) \right) \right\|_2^2 \quad \text{for any } \boldsymbol{\alpha} \in \mathbb{R}^p.$$

这表明

$$\mathrm{Lev}[f, f(\boldsymbol{\alpha}_0)] \subseteq \mathbb{B}_2 \left[\boldsymbol{\alpha}_0 - \frac{1}{L_a} \nabla f(\boldsymbol{\alpha}_0), \frac{1}{L_a} \|\nabla f(\boldsymbol{\alpha}_0)\|_2 \right].$$

由于 f 是闭的，上述水平集也是闭的（定理 2.55）；又因为它包含于一个球内，所以它也是有界的。因此，$\mathrm{Lev}[f, f(\boldsymbol{\alpha}_0)]$ 是紧集。由此可推知，在 $\mathrm{dom}(f)$ 上极小化 f 的最优解集，与在非空紧集 $\mathrm{Lev}[f, f(\boldsymbol{\alpha}_0)]$ 上极小化f的最优解集相同。根据适当闭函数的魏尔斯特拉斯定理（定理 2.56 中的 (iv).(a)），可知极小化解存在。为证明其唯一性，假设 $\widetilde{\boldsymbol{\alpha}}$和 $\boldsymbol{\alpha}^*$是 f的极小化解。那么 $f(\widetilde{\boldsymbol{\alpha}}) = f(\boldsymbol{\alpha}^*)$，其中 $f(\boldsymbol{\alpha}^*)$ 是 f的最小值。于是，根据定理 2.84 的性质 (iv)，

$$f(\boldsymbol{\alpha}^*) \leq f\left(\frac{1}{2}\widetilde{\boldsymbol{\alpha}} + \frac{1}{2}\boldsymbol{\alpha}^* \right) \leq \frac{1}{2} f(\widetilde{\boldsymbol{\alpha}}) + \frac{1}{2} f(\boldsymbol{\alpha}^*) - \frac{L_a}{8} \|\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2 = f(\boldsymbol{\alpha}^*) - \frac{L_a}{8} \|\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2,$$

这意味着 $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*$，从而证明了 f 的极小化解唯一。

（ii） 令 $\boldsymbol{\alpha}^*$ 为 f 的唯一极小化解。于是，根据凸函数的最优性条件（定理 2.62），$\nabla f(\boldsymbol{\alpha}^*) = \mathbf{0}$。因此，根据强凸性的定义，

$$f(\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}^*) \geq \langle \mathbf{0}, \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \rangle + \frac{L_a}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2^2 = \frac{L_a}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|_2^2, \quad \text{for any } \boldsymbol{\alpha} \in \mathbb{R}^p,$$

从而证明了断言（ii）。 ■

光滑性的定义并不要求函数是凸函数。不过，如果函数确实是凸的，则可以用以下方式对其进行等价刻画。

定理 2.83：（SS 与凸性的刻画定理（Beck，2017年）） 设 f： $\mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$ 是在凸集 $\mathbb{S}$上定义的可微凸 $\mathbb{S}$，并设$L_b > 0$。则以下命题等价:

(i) f 是 L_b-光滑的。(ii) 函数 $g(\boldsymbol{\alpha}) \triangleq \frac{L_b}{2}\boldsymbol{\alpha}^\top\boldsymbol{\alpha} - f(\boldsymbol{\alpha})$ 是凸函数。(iii) 对所有 $\boldsymbol{\alpha}$、$\boldsymbol{\beta} \in \mathbb{S}$， $f(\boldsymbol{\beta}) \leq f(\boldsymbol{\alpha}) + \langle \nabla f(\boldsymbol{\alpha})\ \boldsymbol{\beta} - \boldsymbol{\alpha}\rangle + \frac{L_b}{2}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$。(iv) 对所有 $\boldsymbol{\alpha}$、$\boldsymbol{\beta} \in \mathbb{S}$， $f(\boldsymbol{\beta}) \geq f(\boldsymbol{\alpha}) + \langle \nabla f(\boldsymbol{\alpha})\ \boldsymbol{\beta} - \boldsymbol{\alpha}\rangle + \frac{1}{2L_b}\|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_2^2$。(v) 对任意 $\boldsymbol{\alpha}$、$\boldsymbol{\beta} \in \mathbb{S}$ 和 $\lambda \in [0,1]$， $f(\lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}) \geq \lambda f(\boldsymbol{\alpha}) + (1-\lambda)f(\boldsymbol{\beta}) - \frac{L_b}{2}\lambda(1-\lambda)\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$。在 (iii) 和 (iv) 中反转 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$，我们还可以得到

$$\frac{1}{L_b}\|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_2^2 \leq \langle \nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta}\rangle \leq L_b\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2, \text{ for all } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}. \tag{2.21}$$

证明 [定理 2.83] **(i) $\Longrightarrow$ (ii).** 对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$，可得

$$\begin{aligned}\big(\nabla g(\boldsymbol{\alpha}) - \nabla g(\boldsymbol{\beta})\big)^\top(\boldsymbol{\alpha} - \boldsymbol{\beta}) &= L_b\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 - \big(\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\big)^\top(\boldsymbol{\alpha} - \boldsymbol{\beta}) \\ &\geq L_b\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2\|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_2 \geq 0,\end{aligned}$$

其中，最后一个不等式由定理2.80 及光滑性的定义可得。因此，$g(\boldsymbol{\alpha})$ 根据定理 2.46 是一个凸函数。

(ii) $\Longrightarrow$ (iii). 由于 $g(\boldsymbol{\alpha}) \triangleq \frac{L_b}{2}\|\boldsymbol{\alpha}\|_2^2 - f(\boldsymbol{\alpha})$ 是凸的，因此可得

$$g(\boldsymbol{\beta}) \geq g(\boldsymbol{\alpha}) + \nabla g(\boldsymbol{\alpha})^\top(\boldsymbol{\beta} - \boldsymbol{\alpha}), \quad \forall\, \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S},$$

其中， $\nabla g(\boldsymbol{\alpha}) = L_b\boldsymbol{\alpha} - \nabla f(\boldsymbol{\alpha})$。将梯度代入凸性条件，可得

$$\begin{aligned}&\frac{L_b}{2}\|\boldsymbol{\beta}\|_2^2 - f(\boldsymbol{\beta}) \geq \frac{L_b}{2}\|\boldsymbol{\alpha}\|_2^2 - f(\boldsymbol{\alpha}) + \big(L_b\boldsymbol{\alpha} - \nabla f(\boldsymbol{\alpha})\big)^\top(\boldsymbol{\beta} - \boldsymbol{\alpha}) \\ \Longrightarrow\quad & f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq \nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta} - \boldsymbol{\alpha}) + \frac{L_b}{2}\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2.\end{aligned} \tag{2.22}$$

(iii) $\Longrightarrow$ (iv). 我们可以假设 $\nabla f(\boldsymbol{\alpha}) \neq \nabla f(\boldsymbol{\beta})$，否则根据 f 的凸性，不等式 (iv) 显然成立。固定一个 $\boldsymbol{\alpha} \in \mathbb{S}$，考虑函数

$$g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{\beta} - \boldsymbol{\alpha}\rangle, \quad \boldsymbol{\beta} \in \mathbb{S},$$

其关于 $\boldsymbol{\beta}$ 是凸的。注意，对任意 $\boldsymbol{\beta} \in \mathbb{S}$， $\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \nabla f(\boldsymbol{\beta}) - \nabla f(\boldsymbol{\alpha})$。根据 (iii) 的条件，对任意 $\boldsymbol{\beta}$， $\boldsymbol{u} \in \mathbb{S}$，我们有

$$\begin{aligned}g_{\boldsymbol{\alpha}}(\boldsymbol{u}) &= f(\boldsymbol{u}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{u} - \boldsymbol{\alpha}\rangle \\ &\leq \big\{f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta}), \boldsymbol{u} - \boldsymbol{\beta}\rangle + \frac{L_b}{2}\|\boldsymbol{u} - \boldsymbol{\beta}\|_2^2\big\} - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{u} - \boldsymbol{\alpha}\rangle \\ &= f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{\beta} - \boldsymbol{\alpha}\rangle + \langle \nabla f(\boldsymbol{\beta}) - \nabla f(\boldsymbol{\alpha}), \boldsymbol{u} - \boldsymbol{\beta}\rangle + \frac{L_b}{2}\|\boldsymbol{u} - \boldsymbol{\beta}\|_2^2 \\ &= g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) + \langle \nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \boldsymbol{u} - \boldsymbol{\beta}\rangle + \frac{L_b}{2}\|\boldsymbol{u} - \boldsymbol{\beta}\|_2^2,\end{aligned} \tag{2.23}$$

由于 $\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = \mathbf{0}$，根据 $g_{\boldsymbol{\alpha}}$ 的凸性可知，$\boldsymbol{\alpha}$ 是 $g_{\boldsymbol{\alpha}}$ 的全局极小点，使得对所有 $\boldsymbol{u} \in \mathbb{S}.$，$0 = g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \leq g_{\boldsymbol{\alpha}}(\boldsymbol{u})$，。令 $\boldsymbol{\beta} \in \mathbb{S}$，并令 $\widetilde{\boldsymbol{g}} \triangleq \frac{\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})}{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2}$，使得$\|\widetilde{\boldsymbol{g}}\|_2 = 1$且 $\langle \nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \widetilde{\boldsymbol{g}} \rangle = \|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2$。定义 $\boldsymbol{u} \triangleq \boldsymbol{\beta} - \frac{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2}{L_b} \widetilde{\boldsymbol{g}}$，$g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ 的最优值表明

$$0 = g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \leq g_{\boldsymbol{\alpha}}\left(\boldsymbol{\beta} - \frac{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2}{L_b} \widetilde{\boldsymbol{g}}\right).$$

将前面的不等式与（2.23）结合，可得

$$\begin{aligned} 0 = g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) &\leq g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) - \frac{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2}{L_b} \langle \nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \widetilde{\boldsymbol{g}} \rangle + \frac{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2^2 \cdot \|\widetilde{\boldsymbol{g}}\|_2^2}{2L_b} = g_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) - \frac{\|\nabla g_{\boldsymbol{\alpha}}(\boldsymbol{\beta})\|_2^2}{2L_b} \\ &= f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) - \langle \nabla f(\boldsymbol{\alpha}), \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle - \frac{1}{2L_b} \|\nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\|_2^2, \end{aligned}$$

从而证明了（iv）。

(iii) $\Longrightarrow$ (v). 令 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$ 和 $\lambda \in [0,1]$，并记 $\boldsymbol{\xi} \triangleq \lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}$。由于 $\mathbb{S}$ 是凸函数，因此我们也有 $\boldsymbol{\xi} \in \mathbb{S}$。在 $\boldsymbol{\alpha}, \boldsymbol{\xi}$ 和 $\boldsymbol{\beta}, \boldsymbol{\xi}$ 的情形下应用（iii），并代入 $\boldsymbol{\xi}$ 的表达式，可得

$$\begin{aligned} f(\boldsymbol{\alpha}) &\leq f(\boldsymbol{\xi}) + (1-\lambda) \langle \nabla f(\boldsymbol{\xi}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle + \frac{L_b(1-\lambda)^2}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2, \\ f(\boldsymbol{\beta}) &\leq f(\boldsymbol{\xi}) + \lambda \langle \nabla f(\boldsymbol{\xi}), \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle + \frac{L_b\lambda^2}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2. \end{aligned}$$

将第一个不等式乘以 λ，第二个不等式乘以 $1-\lambda$，再将两者相加，得到不等式（v）。

(v) $\Longrightarrow$ (iii). 重新整理不等式（v）中的各项，可得它等价于

$$f(\boldsymbol{\beta}) \leq f(\boldsymbol{\alpha}) + \frac{f\big(\boldsymbol{\alpha} + (1-\lambda)(\boldsymbol{\beta} - \boldsymbol{\alpha})\big) - f(\boldsymbol{\alpha})}{1-\lambda} + \frac{L_b}{2}\lambda \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2.$$

取 $\lambda \to 1^-$，则前面的不等式变为 $f(\boldsymbol{\beta}) \leq f(\boldsymbol{\alpha}) + f'(\boldsymbol{\alpha};\ \boldsymbol{\beta} - \boldsymbol{\alpha}) + \frac{L_b}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$。由于 f 是可微函数，(2.1) 表明 $f'(\boldsymbol{\alpha};\ \boldsymbol{\beta} - \boldsymbol{\alpha}) = \langle \nabla f(\boldsymbol{\alpha})\ \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle$，由此得到所需结果。

■

定理 2.84：（强凸性的刻画定理（Beck，2017年）） 设 $f:\ \mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$ 是定义在凸集 $\mathbb{S}$ 上的凸且可微函数，并设 $L_a > 0$。则下列陈述等价：

(i) f 是 L_a-强凸的。(ii) $f(\boldsymbol{\beta}) \geq f(\boldsymbol{\alpha}) + \langle \nabla f(\boldsymbol{\alpha})\ \boldsymbol{\beta} - \boldsymbol{\alpha} \rangle + \frac{L_a}{2} \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2$ 对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$。(iii) $\langle \nabla f(\boldsymbol{\alpha}) - \nabla f(\boldsymbol{\beta})\, \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle \geq L_a \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ 对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$。(iv) $f(\lambda\boldsymbol{\alpha} + (1-\lambda)\boldsymbol{\beta}) \leq \lambda f(\boldsymbol{\alpha}) + (1-\lambda) f(\boldsymbol{\beta}) - \frac{L_a}{2}\lambda(1-\lambda) \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ 对任意 $\lambda \in [0,1]$ 和 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$。

证明与定理 2.83 的证明类似，留作习题。该定理表明，强凸函数是严格凸的。值得注意的是，结合定理 2.83 和

定理 2.84：如果 $f:\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$ 可微，L_a-强凸，且 L_b-光滑，定义在凸集 $\mathbb{S}$ 上，且 $0<L_a<L_b$，那么对于任意 $\boldsymbol{\alpha},\ \boldsymbol{\beta}\in\mathbb{S}$，

$$\frac{L_a}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2 \le f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})-\langle\nabla f(\boldsymbol{\alpha}),\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle \le\frac{L_b}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2; \quad (2.24a)$$

$$L_a\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2 \le\langle\nabla f(\boldsymbol{\alpha})-\nabla f(\boldsymbol{\beta}),\boldsymbol{\alpha}-\boldsymbol{\beta}\rangle \le L_b\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2; \quad (2.24b)$$

$$\frac{L_a\zeta}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2 \le\{\lambda f(\boldsymbol{\alpha})+(1-\lambda)f(\boldsymbol{\beta})\}-f\big(\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}\big) \le\frac{L_b\zeta}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2, \quad (2.24c)$$

其中 $\zeta\triangleq\lambda(1-\lambda)$ 对任意 $\lambda\in[0,1]$，且 (2.24a) 等价于定义 2.78。一些教材使用 (2.24c) 来定义强凸和强光滑，因为这种定义不要求函数可微。本书仍然重点讨论可微函数。当函数不可微时，根据定理 2.83 或定理 2.84，由于次微分非空，其中的梯度可以替换为次梯度。定理 2.59。此外，对于强凸和强光滑函数，我们还有以下刻画定理。

定理 2.85：（强凸性与强光滑性的刻画定理）设 $f:\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$ 是定义在凸集 $\mathbb{S}$ 上的可微、L_a-强凸且 L_b-光滑函数，其中 $L_a\le L_b$。那么，对任意 $\boldsymbol{\alpha},\ \boldsymbol{\beta}\in\mathbb{S}$，有

$$\langle\nabla f(\boldsymbol{\alpha})-\nabla f(\boldsymbol{\beta}),\boldsymbol{\alpha}-\boldsymbol{\beta}\rangle\ge\frac{L_aL_b}{L_a+L_b}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2+\frac{1}{L_a+L_b}\|\nabla f(\boldsymbol{\alpha})-\nabla f(\boldsymbol{\beta})\|_2^2. \quad (2.25)$$

当 $L_a=0$时，这就化简为定理 2.83 中的推论（2.21）。

证明 [定理 2.85]设 $g(\boldsymbol{\alpha})\triangleq f(\boldsymbol{\alpha})-\frac{1}{2}L_a\|\boldsymbol{\alpha}\|_2^2$。则 $\nabla g(\boldsymbol{\alpha})=\nabla f(\boldsymbol{\alpha})-L_a\boldsymbol{\alpha}$。利用定理 2.84 中的 (iii) 和定理 2.83 中的（2.21），可得 $g(\boldsymbol{\alpha})$ 是（L_b-L_a）-光滑的，并且

$$\frac{1}{L_b-L_a}\|\nabla g(\boldsymbol{\alpha})-\nabla g(\boldsymbol{\beta})\|_2^2\le\langle\nabla g(\boldsymbol{\alpha})-\nabla g(\boldsymbol{\beta}),\boldsymbol{\alpha}-\boldsymbol{\beta}\rangle,$$

这与该结果等价。 ■

在非凸优化中，我们通常关注的并非全局凸性或光滑性，而是这些性质的受限版本，它们仅在定义域的某些子集上成立——即使函数整体上是非凸的。

定义 2.86（受限强凸性/光滑性）。设 $f:\ \mathbb{S}\subseteq\mathbb{R}^p\to\mathbb{R}$是一个可微函数。若 f 在（可能是非凸的）集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 上满足受限凸性，则称其满足此性质；对于每个 $\boldsymbol{\alpha},\ \boldsymbol{\beta}\in\mathbb{S}$，均有

$$f(\boldsymbol{\beta})\ge f(\boldsymbol{\alpha})+\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha}). \quad (2.26)$$

此外，如果对于每个 $\boldsymbol{\alpha},\ \boldsymbol{\beta}\in\mathbb{S}$，都有以下结论，则称 f 在（可能是非凸的）集合 $\mathbb{S}\subseteq\mathbb{R}^p$ 上满足 L_a-受限强凸性（*RSC*）和 L_b-限制强光滑性（*RSS*）

$$\frac{L_a}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2\le f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})-\nabla f(\boldsymbol{\alpha})^\top(\boldsymbol{\beta}-\boldsymbol{\alpha})\le\frac{L_b}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2. \quad (2.27)$$

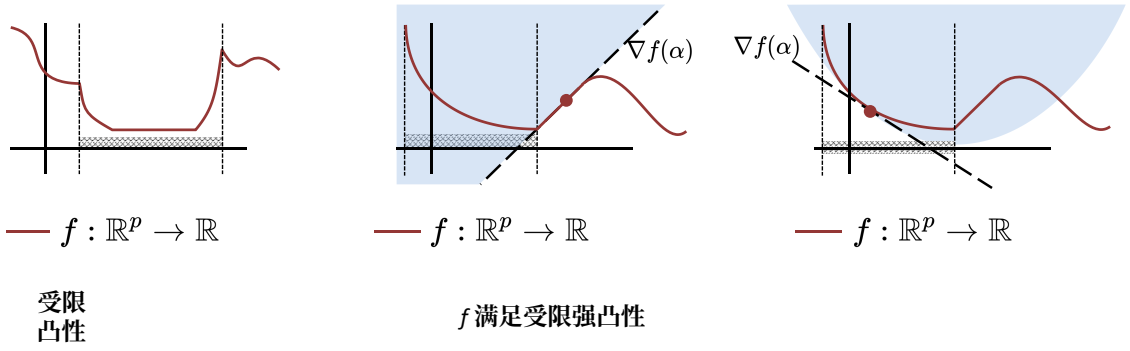


图 2.6： 受限凸性性质的示意图。第一个函数 f 在整个实数轴上表现为非凸，但在虚线垂直线所界定的阴影区域内呈现凸性特征。第二个和第三个函数整体上也都是非凸的，但表现出受限强凸性。在阴影区域之外，这些函数不满足凸性；然而，在所定义的区域内，它们具有强凸性。图改编自Jain 等人 (2017年)。

当函数二阶可微时，定理2.47还表明

$$L_a \boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{\alpha}) \preceq L_b \boldsymbol{I}, \tag{2.28}$$

对于任意 $\boldsymbol{\alpha} \in \text{int}(\mathbb{S}$，并且存在一个向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 使得 $\boldsymbol{\alpha} + \lambda\boldsymbol{\beta} \in \mathbb{S}$ 对于某个充分小的 λ （参见定理2.47的证明）。也就是说，

$$L_a \leq \frac{\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{\alpha}) \boldsymbol{v}}{\|\boldsymbol{v}\|_2^2} \leq L_b, \tag{2.29}$$

for 任意的 $\boldsymbol{\alpha} \in \text{int}(\mathbb{S})$ 和 $\boldsymbol{v} \in \mathbb{R}^p$，使得 $\boldsymbol{\alpha} + \lambda \boldsymbol{v} \in \mathbb{S}$ 对某个充分小的 λ .

图2.6 展示了这些受限性质。注意，当 $f(\boldsymbol{\alpha}) = \frac{1}{2n}\|\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2$，其中 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 且 $\boldsymbol{y} \in \mathbb{R}^n$，(2.29)变为 $L_a \leq \frac{\boldsymbol{v}^\top (\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}) \boldsymbol{v}}{\|\boldsymbol{v}\|_2^2} \leq L_b$。这给出了稀疏问题的 RSC 和 RSS 性质（定义 4.33）。

2.8. 投影、近端、布雷格曼近端与分离

本节介绍投影和近端算子的基本概念，它们是优化理论和凸分析中的重要工具。这些概念在第9章讨论的投影梯度下降法和近端梯度下降法中发挥核心作用，也广泛应用于各种理论结果的证明中。

我们首先定义一种基于凸函数的散度度量。

定义 2.87（Bregman 距离（Bregman，1965年））。Bregman距离（也称为布雷格曼散度）推广了平方欧几里得距离的概念，用于衡量空间中两个点之间的差异。对于凸且可微的函数 ϕ： $\mathbb{S} \to \mathbb{R}$，其中 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个凸集。相对于 ϕ，从点 $\boldsymbol{\beta}$ 到点 $\boldsymbol{\alpha}$ 的 Bregman 距离定义为：

$$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \phi(\boldsymbol{\alpha}) - \phi(\boldsymbol{\beta}) - \langle \nabla\phi(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle . \tag{2.30}$$

如果 ϕ 是一个非光滑凸函数，令 $\boldsymbol{g} \in \partial\phi(\boldsymbol{\beta})$ 为函数 ϕ 在点 $\boldsymbol{\beta}$处的一个次梯度向量。点 $\boldsymbol{\alpha}$和 $\boldsymbol{\beta}$之间的 Bregman 距离记作$D_\phi^{\boldsymbol{g}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$，定义为

$$D_\phi^{\boldsymbol{g}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \phi(\boldsymbol{\alpha}) - \phi(\boldsymbol{\beta}) - \langle \boldsymbol{g}, \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle . \tag{2.31}$$

为简洁起见，我们在下面的注释中总结 Bregman 距离的关键性质。

注释 2.88（Bregman 距离的性质）。Bregman 距离满足以下性质：

1. 非负性。 $\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0$，对所有的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}$. 均成立。当 ϕ 严格凸时，等号成立当且仅当 $\boldsymbol{\alpha} = \boldsymbol{\beta}$。2. 同一点的 Bregman 距离为零，即 $D_\phi(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 0$。3. 关于第一自变量的凸性。对于固定的 $\boldsymbol{\beta}$， $\mathcal{D}_\phi(\cdot, \boldsymbol{\beta})$ 关于其第一自变量是凸的。这一性质确保涉及在 $\boldsymbol{\alpha}$ 上对 $\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 进行最小化的问题适定——这是优化算法中经常利用的一项特征。4. 不对称性。一般而言， $\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \mathcal{D}_\phi(\boldsymbol{\beta}, \boldsymbol{\alpha})$。因此，Bregman 距离/散度不满足距离度量所要求的对称性。“距离”一词因历史原因而沿用。5. 闭性。Bregman 距离可以衡量两个点 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha}$ 的接近程度，因为对于连接 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha}$ 的直线上的任意点 $\boldsymbol{\xi}$， $D_\phi(\boldsymbol{\beta}, \boldsymbol{\alpha}) \geq D_\phi(\boldsymbol{\xi}, \boldsymbol{\alpha})$。6. 关于梯度的线性。如果 ϕ 是二次函数，则 Bregman 距离化为平方欧几里得距离的一个缩放形式。如果 ϕ 是线性的，则 Bregman 距离为零。7. 勾股性质。设 $\mathbb{T} \subseteq \mathbb{S}$ 是 $\mathbb{S}$ 的一个非空、闭且凸的子集，并令 $\mathcal{P}_{\mathbb{T}}^{\phi}(\boldsymbol{\beta})$ 表示相对于 $\mathcal{D}_\phi$ 将 $\boldsymbol{\beta}$ 投影到 $\mathbb{T}$ 上的投影（定义 2.90

）。那么对于任意 $\boldsymbol{\alpha} \in \mathbb{T}$,

$$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{D}_\phi(\boldsymbol{\alpha}, \mathcal{P}_{\mathbb{T}}^{\phi}(\boldsymbol{\beta})) + \mathcal{D}_\phi(\mathcal{P}_{\mathbb{T}}^{\phi}(\boldsymbol{\beta}), \boldsymbol{\beta})$$

这个恒等式与经典勾股定理相呼应，对于分析迭代算法的收敛至关重要。

8. 三点性质。 对于任意三个点 $\boldsymbol{\alpha}$， $\boldsymbol{\beta}$， $\boldsymbol{\xi} \in \mathbb{S}$，以下关系成立：

$$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \mathcal{D}_\phi(\boldsymbol{\beta}, \boldsymbol{\xi}) = \mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\xi}) + \langle \nabla\phi(\boldsymbol{\xi}) - \nabla\phi(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle \tag{2.32}$$

这个恒等式常用于推导界并确立算法保证。

9. 局部行为。 当 $\boldsymbol{\alpha}$ 接近 $\boldsymbol{\beta}$时， $\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 近似于 ϕ 在 $\boldsymbol{\beta}$ 处的二阶泰勒展开（根据定理 2.14中的二次近似定理）。具体而言，若 ϕ二阶连续可微，则对于足够小的 $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$,

$$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \approx \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\beta})^\top \nabla^2\phi(\boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta})$$

其中 $\nabla^2\phi(\boldsymbol{\beta})$ 表示 ϕ在 $\boldsymbol{\beta}$处的海森矩阵。

下面给出两个常用的强凸函数示例。

示例 2.89（欧几里得情形与负熵情形下的 Bregman）。注意，当$\phi(\boldsymbol{\alpha}) \triangleq \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2$：$\mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$，$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 变为

$$\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2,$$

这对应于平方欧几里得距离。令 $\phi(\boldsymbol{\alpha})$ 为负熵函数，定义为 $\phi(\boldsymbol{\alpha}) \triangleq \begin{cases} \sum_{i=1}^p \alpha_i \ln(\alpha_i) \\ \infty \end{cases}$

$\boldsymbol{\alpha} \in \mathbb{R}_+^p$；否则，布雷格曼距离 $\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 变为

$$\begin{aligned}\mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{i=1}^p \alpha_i \ln(\alpha_i) - \sum_{i=1}^p \beta_i \ln(\beta_i) - \sum_{i=1}^p (\ln(\beta_i)+1)(\alpha_i - \beta_i) \\ &= \sum_{i=1}^p \alpha_i \ln(\alpha_i/\beta_i) - \sum_{i=1}^p (\alpha_i - \beta_i).\end{aligned}$$

若 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 位于概率单纯形上（即 $\sum_{i=1}^p \alpha_i = \sum_{i=1}^p \beta_i = 1$），线性项便相互抵消，从而得到标准的*Kullback–Leibler*散度。□

布雷格曼距离在优化、机器学习、统计学和信息论等领域有广泛应用。在凸优化算法的设计中，它们尤为重要，因为它们提供了一种考虑几何结构的进展度量，能够反映凸函数所诱导的结构，衡量算法朝最优解推进的程度 ϕ。例如，布雷格曼距离是镜像下降方法的核心；镜像下降是梯度下降的一种强大推广，可在非欧几里得几何中运行；例如参见 Lu (2025年)。

接下来，我们介绍投影和近端算子的定义，以及一种称为布雷格曼近端算子(Lu，2025年）。

定义 2.90（投影、近端算子、布雷格曼近端算子）。设 $\mathbb{S} \subseteq \mathbb{R}^p$ 为非空**闭**集合，并设 $\boldsymbol{\beta} \in \mathbb{R}^p$。（正交）投影到 $\boldsymbol{\beta}$记为$\mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$：$\mathbb{R}^p \to \mathbb{S}$，定义为

$$\textbf{(Projection)}: \quad \widetilde{\boldsymbol{\beta}} \in \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) \triangleq \arg\min_{\boldsymbol{\alpha}\in\mathbb{S}} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 = \arg\min_{\boldsymbol{\alpha}\in\mathbb{S}} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2. \tag{2.33}$$

设 $f\colon \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 为一个**适当的**函数，并设 $\boldsymbol{\beta} \in \mathbb{R}^p$。$f$的近端映射记为 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$，定义为

$$\textbf{(Proximal)}: \quad \widehat{\boldsymbol{\beta}} \in \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) \triangleq \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \left(f(\boldsymbol{\alpha}) + \frac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 \right). \tag{2.34}$$

利用示性函数 $\delta_{\mathbb{S}}$（定义见习题2.36），并假设 $\mathbb{S}$非空，可以验证投影算子与近端算子之间满足关系

$$\mathcal{P}\mathrm{rox}_{\delta_{\mathbb{S}}}(\boldsymbol{\beta}) = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}), \quad \text{for all } \boldsymbol{\beta} \in \mathbb{S}. \tag{2.35}$$

将式（2.34）中的平方欧几里得距离替换为 Bregman 距离，可得布雷格曼近端映射，记作 $\mathcal{B}\mathrm{rox}_f(\boldsymbol{\beta})$，其形式为

$$\textbf{(Bregman)}: \quad \overline{\boldsymbol{\beta}} \in \mathcal{B}\mathrm{rox}_{f,\phi}(\boldsymbol{\beta}) \triangleq \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \big(f(\boldsymbol{\alpha}) + \mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \big), \tag{2.36}$$

其中 $\phi:\ \mathbb{R}^p \to \mathbb{R}\cup\{\infty\}$ 是一个适当的**凸**函数。根据例 2.89，

$$\mathcal{B}\mathrm{rox}_{f,\phi}(\boldsymbol{\beta}) \equiv \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) \quad \text{if } \phi(\boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2.$$

除非另有说明，我们假设投影算子作用于非空闭集，且近端算子应用于适当函数。以下例子说明了这些假设为何是必要的。

例2.91（非闭 $\mathbb{S}$集与非适当函数的反例）。考虑开集 $\mathbb{S} = \{\alpha \in \mathbb{R} \mid \alpha > 0\}$，并令 $\beta = 0$。于是投影问题变为：$\widehat{\beta} =$ 最小值点集合$_{\alpha>0}\,\alpha^2$.

- α^2 的下确界是 0，但该值无法取到，因为 $\alpha = 0 \notin \mathbb{S}$。因此，$\mathbb{S}$ 中没有点能够达到这个下确界。任何满足 $\alpha_t \to 0$ 的序列 $\{\alpha_t\}_{t>0} \subset \mathbb{S}$ 都能使 α^2最小化，但“解”
- 并不唯一，因为它取决于所选的序列。

因此，对于非闭集，投影没有良好定义。当 f 不是适当函数时，近端算子也会出现类似问题——例如，如果 $f \equiv +\infty$，那么该最小化问题没有可行解。□

习题2.92（近端演算）。设 $g:\ \mathbb{R}^p \to (-\infty,\ \infty]$ 是适当的，且 $\eta \neq 0$。证明：

(i) 设 $f(\boldsymbol{\alpha}) = \eta g(\boldsymbol{\alpha}/\eta)$。则

$$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \eta\mathcal{P}\mathrm{rox}_{g/\eta}(\boldsymbol{\beta}/\eta).$$

(ii) 设 $\eta \neq 0$ 和 $\boldsymbol{\xi} \in \mathbb{R}^p$。定义 $f(\boldsymbol{\alpha}) \triangleq g(\eta\boldsymbol{\alpha} + \boldsymbol{\xi})$。则

$$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \frac{1}{\eta}\left(\mathcal{P}\mathrm{rox}_{\eta^2 g}(\eta\boldsymbol{\beta} + \boldsymbol{\xi}) - \boldsymbol{\xi}\right).$$

在接下来的几个小节中，我们将讨论投影算子和近端算子的性质。Bregman近端算子的性质将在问题中讨论2.10~2.11。

2.8.1 投影算子的性质

下面介绍投影算子的一些关键性质。

引理 2.93：（投影性质-O）令 $\mathbb{S} \subseteq \mathbb{R}^p$ 为任意一个**闭集**，并令$\boldsymbol{\beta} \in \mathbb{R}^p$ 满足：$\widetilde{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$ 是 $\boldsymbol{\beta}$ 在集合 $\mathbb{S}$上的投影。那么，对于所有 $\boldsymbol{\alpha} \in \mathbb{S}$，有 $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2 \leq \left\|\boldsymbol{\alpha} - \boldsymbol{\beta}\right\|_2$。

证明 [引理2.93] 根据投影的定义可直接得到：$\mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$ =最小值点集合$_{\boldsymbol{\alpha}\in\mathbb{S}}\,\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$。■

注意，投影性质-O 对任意闭集都成立，无论该集合是否为凸集。然而，以下三个性质要求集合具有凸性。

引理 2.94：（投影性质 I）设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个闭**凸集**，且令$\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\widetilde{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$。那么，对所有 $\boldsymbol{\alpha} \in \mathbb{S}$，都有 $\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}},\ \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle \leq 0$，也就是说，两个向量之间的角大于或等于 90°。

证明 [引理 294]。反设 $\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}},\ \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle > 0$ 对某个 $\boldsymbol{\alpha} \in \mathbb{S}$成立。由于 $\mathbb{S}$ 是凸集，且 $\widetilde{\boldsymbol{\beta}}$, $\boldsymbol{\alpha} \in \mathbb{S}$，对于任意 $\lambda \in [0,1]$，我们有 $\boldsymbol{\xi}_\lambda \triangleq \lambda \cdot \boldsymbol{\alpha} + (1-\lambda) \cdot \widetilde{\boldsymbol{\beta}} \in \mathbb{S}$.现在我们将证明存在一个 $\lambda \in [0,1]$ 的取值，使得 $\|\boldsymbol{\beta} - \boldsymbol{\xi}_\lambda\|_2^2 < \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right\|_2^2$, ~，这与 $\boldsymbol{\beta}$ 是凸集中离 $\boldsymbol{\beta}$ 最近的点这一事实矛盾，从而证明该引理。为此，暂且令 $\lambda_1 \triangleq \frac{2\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}},\ \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle}{\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}\|_2^2}$。我们有

$$\|\boldsymbol{\beta} - \boldsymbol{\xi}_{\lambda_1}\|_2^2 = \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right\|_2^2 - 2\lambda_1 \left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle + \lambda_1^2 \left\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}\right\|_2^2 = \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right\|_2^2.$$

令 $g(\lambda) \triangleq \|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\|_2^2 - 2\lambda\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle + \lambda^2 \|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}\|_2^2$。可以证明，根据二次函数的性质，$\lambda = \lambda_1$无法取得 $g(\lambda)$ 的最小值。事实上，最小值在 $\lambda = \frac{\lambda_1}{2}$ 处取得。由于我们假设 $\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle > 0$，令 $0 < \lambda < \min\left\{1, \frac{2\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}},\ \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle}{\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}\|_2^2}\right\}$，则有 $\|\boldsymbol{\beta} - \boldsymbol{\xi}_\lambda\|_2^2 < \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right\|_2^2$。这与投影算子的性质矛盾，证毕。

■

引理 2.95：（投影性质 II）设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个闭**凸集**，并且$\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\widetilde{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$。那么，对于所有 $\boldsymbol{\alpha} \in \mathbb{S}$，我们有 $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2^2 \leq \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 - \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right\|_2^2$，这也意味着 $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2 \leq \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2$（前一个不等式可视为到凸集的投影的广义勾股定理。）

证明 [引理 2.95] 我们按如下方式展开平方范数：

$$\begin{aligned}\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 &= \left\|(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}) - (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\|_2^2 = \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2^2 + \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2^2 - 2\left\langle \widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\rangle \\ &\overset{\dagger}{\geq} \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2^2 + \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2^2 \geq \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2^2,\end{aligned}$$

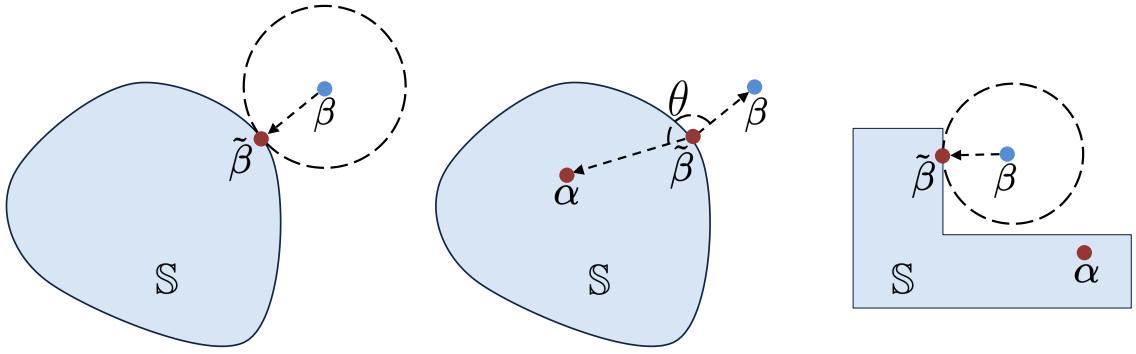
其中不等式 (†) 由投影性质 I 得出。 ■

引理 2.96：（投影性质 III）设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个闭**凸集**，且 $\boldsymbol{\beta} \in \mathbb{R}^p$。那么，$\mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) \triangleq$ 最小值点集合$_{\boldsymbol{\alpha} \in \mathbb{S}} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ 具有唯一的**最优解**。

证明 [引理 2.96] 由于 $\mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) \triangleq$ 最小值点集合$_{\boldsymbol{\alpha} \in \mathbb{S}} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ 中的目标函数是与正定矩阵相关联的二次函数，根据问题 2.4 可知，该目标函数是强制的，因此根据定理 (iv) 中的魏尔斯特拉斯定理 2.56，该问题至少存在一个最优解。此外，由于目标函数是严格凸的（同样，这是因为该目标函数是与正定矩阵相关联的二次函数），根据定理2.44 可知，只存在一个最优解。

■

凸性很重要。请注意，投影性质-I、II和III通常被称为投影的一阶性质，并且关键依赖于 $\mathbb{S}$ 的凸性。然而，投影性质 -O （通常称为零阶性质）对任意非空闭集都成立，无论其是否为凸集；见图 2.7。



投影性质-O **投影性质-I** **投影性质-II**

图2.7： 投影算子将点 $\boldsymbol{\beta}$ 映射到集合 $\mathbb{S}$中距离最近的点 $\tilde{\boldsymbol{\beta}}$ 。对于凸集，投影性质 I 保证向量 $\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}$（从投影点指向原始点）与任意向量 $\boldsymbol{\alpha}-\tilde{\boldsymbol{\beta}}$（从投影点指向集合中的另一点）之间的角 θ 至少为 $90°$。满足性质-I的集合自动满足性质-II，这保证将 $\boldsymbol{\beta}$ 投影到 $\mathbb{S}$上不会增加它与任何其他点$\boldsymbol{\alpha}\in\mathbb{S}$之间的距离。然而，非凸集可能不满足性质-II。

例 2.97（投影到 $\mathbb{R}^p$ 的子集上）。以下列出一些非空闭凸集及其正交投影的示例：

(i) 非负正交象限 $\mathbb{S}_1=\mathbb{R}^p_+$：$\mathcal{P}_{\mathbb{S}_1}(\boldsymbol{\beta})=\{\max\{\beta_i,0\}\}_{i=1}^p$。(ii) 盒集 $\mathbb{S}_2=\text{Box}[\boldsymbol{\ell},\boldsymbol{u}]\triangleq\{\boldsymbol{\alpha}\in\mathbb{R}^p\mid\boldsymbol{\ell}\le\boldsymbol{\alpha}\le\boldsymbol{u}\}$：$\mathcal{P}_{\mathbb{S}_2}(\boldsymbol{\beta})=\{\min(\max(\beta_i,\ell_i),u_i)\}_{i=1}^p$。(iii) 闭 ℓ_2-球 $\mathbb{S}_3=\mathbb{B}_2[\boldsymbol{a},\gamma]$：$\mathcal{P}_{\mathbb{S}_3}(\boldsymbol{\beta})=\boldsymbol{a}+\frac{\gamma\max\{\|\boldsymbol{\beta}-\boldsymbol{a}\|_2,\gamma\}}{(\boldsymbol{\beta}-\boldsymbol{a})}$。(iv) 半空间 $\mathbb{S}_4=\{\boldsymbol{\alpha}\mid\boldsymbol{c}^\top\boldsymbol{\alpha}\le\eta\}$：$\mathcal{P}_{\mathbb{S}_4}(\boldsymbol{\beta})=\boldsymbol{\beta}-\frac{\max\{\boldsymbol{c}^\top\boldsymbol{\beta}-\eta,0\}}{\|\boldsymbol{c}\|_2^2}\boldsymbol{c}$。(v) 仿射集 $\mathbb{S}_5=\{\boldsymbol{\alpha}\in\mathbb{R}^p\mid\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{y}\}$：$\mathcal{P}_{\mathbb{S}_5}(\boldsymbol{\beta})=\boldsymbol{\beta}-\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})$。

其中 $\boldsymbol{\ell}\in[-\infty,\infty)^p$，$\boldsymbol{u}\in(-\infty,\infty]^p$ 满足 $\boldsymbol{\ell}\le\boldsymbol{u}$, $\boldsymbol{a}\in\mathbb{R}^p$, $\gamma>0$, $\boldsymbol{c}\in\mathbb{R}^p\setminus\{\boldsymbol{0}\}$, $\eta\in\mathbb{R}$, $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 具有满行秩，使得 $\boldsymbol{X}\boldsymbol{X}^\top$ 是非奇异的，且 $\boldsymbol{y}\in\mathbb{R}^n$。(i)–(iv)中的结果显然成立。对于 (v)，这等价于求解以下优化问题：

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\quad\frac{1}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{y},$$

其中 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$，$\boldsymbol{y}\in\mathbb{R}^n$ 和 $\boldsymbol{\beta}\in\mathbb{R}^p$ 是已知的矩阵和向量。我们假设矩阵 $\boldsymbol{X}$ 满行秩（对于不满行秩的矩阵，可以删除线性相关的行，从而得到简化的约束条件）。对于等式约束，我们引入拉格朗日乘子 $\boldsymbol{\lambda}\in\mathbb{R}^n$ 并构造拉格朗日函数（参见第2.9节）：

$$L(\boldsymbol{\alpha},\boldsymbol{\lambda})=\frac{1}{2}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2^2+\boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{y}).$$

由于目标函数是凸的，且约束条件是仿射的，因此 Slater 条件显然成立。定理 2.115 表明，$\boldsymbol{\alpha}^*$ 是全局最优解，当且仅当存在 $\boldsymbol{\lambda}^* \in \mathbb{R}^n$ 使得 $\begin{cases} \boldsymbol{\alpha}^* - \boldsymbol{\beta} + \boldsymbol{X}^\top \boldsymbol{\lambda}^* = \mathbf{0} \\ \boldsymbol{X}\boldsymbol{\alpha}^* = \boldsymbol{y}. \end{cases}$，由此可得

$$\boldsymbol{X}\boldsymbol{\alpha}^* - \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\lambda}^* = \mathbf{0} \qquad \Longrightarrow \qquad \boldsymbol{\lambda}^* = (\boldsymbol{X}\boldsymbol{X}^\top)^{-1}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}).$$

代入 $\boldsymbol{\lambda}^*$ 后，即得所声称的公式。 □

2.8.2 分离定理

利用投影的性质，我们可以证明以下严格分离定理，它对于建立 Farkas引理（问题 2.12）至关重要。反过来，Farkas引理在推导 KKT 条件中发挥着基础作用；更多细节请参见例如 Lu(2025年。其他分离结果将在问题 2.8∼2.9 中讨论。

定理 2.98：（严格分离定理）设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个闭凸集合，且 $\boldsymbol{\beta} \notin \mathbb{S}$。那么存在 $\boldsymbol{w} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ 和 $\gamma \in \mathbb{R}$，使得

$$\boldsymbol{w}^\top \boldsymbol{\beta} > \gamma \qquad \text{and} \qquad \boldsymbol{w}^\top \boldsymbol{\alpha} \leq \gamma, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{S}.$$

定理 [的证明 2.98] 设 $\widetilde{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) \in \mathbb{S}$。根据投影性质 I（引理 2.94），

$$\left\langle \boldsymbol{\alpha} - \widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle \leq 0 \qquad \Longrightarrow \qquad (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top \boldsymbol{\alpha} \leq (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top \widetilde{\boldsymbol{\beta}}, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{S}.$$

记 $\boldsymbol{w} \triangleq \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \neq \mathbf{0}$（因为 $\boldsymbol{\beta} \notin \mathbb{S}$）和 $\gamma \triangleq (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top \widetilde{\boldsymbol{\beta}}$。那么，对于所有 $\boldsymbol{\alpha} \in \mathbb{S}$，我们有 $\boldsymbol{w}^\top \boldsymbol{\alpha} \leq \gamma$。另一方面，

$$\boldsymbol{w}^\top \boldsymbol{\beta} = (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top \boldsymbol{\beta} = (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^\top \widetilde{\boldsymbol{\beta}} = \left\| \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\|_2^2 + \gamma > \gamma,$$

从而得出所需结果。 ■

2.8.3 近端算子的性质

与投影算子类似，对于近端算子，若函数是闭的并满足一个温和的强制性条件（即当 $\|\boldsymbol{\alpha}\|_2 \to \infty$ 时，$f(\boldsymbol{\alpha}) \to \infty$），则可保证对于每个 $\boldsymbol{\beta} \in \mathbb{R}^p$，近端映射 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$ 非空。

引理 2.99：（近端性质 O：闭性和强制性条件下近端算子的非空性）设 $f: \mathbb{R}^p \to (-\infty, \infty]$ 是一个适当的 **闭**函数，并假设满足以下条件：

$$\text{the function } \boldsymbol{\alpha} \mapsto f(\boldsymbol{\alpha}) + \frac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 \text{ is coercive for any } \boldsymbol{\beta} \in \mathbb{R}^p.$$

那么，对于任意 $\boldsymbol{\beta} \in \mathbb{R}^p$，$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$ 都是**非空的**。

证明 [引理 2。99] 对于任意 $\boldsymbol{\alpha} \in \mathbb{R}$，适当函数 $g(\boldsymbol{\alpha}) \triangleq f(\boldsymbol{\alpha}) + \frac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ 作为两个闭函数之和是闭的。因此，根据适当、闭且强制函数的极小化解的标准存在性结果

（例如魏尔斯特拉斯定理 (iv) 中的定理 2.56），g的极小化点集合（恰好是 $\mathcal{P}\mathrm{rox}f(\beta)$）非空。 ■

引理 2.100：（近端性质 I） 设 $f:\mathbb{R}^p\to(-\infty,\infty]$ 是一个适当的**闭凸**函数。那么对于任意 $\beta,\beta\in\mathbb{R}^p$，以下三个陈述等价：

(i) $\widehat{\beta}=\mathcal{P}\mathrm{rox}f(\beta)$。 (ii) $\beta-\widehat{\beta}\in\partial f(\widehat{\beta})$。
(iii) 对任意 $\xi\in\mathbb{R}^p$，$\left\langle\beta-\widehat{\beta},\xi-\widehat{\beta}\right\rangle\le f(\xi)-f(\widehat{\beta})$。

证明 [引理2.100] 根据定义，$\widehat{\beta}=\mathcal{P}\mathrm{rox}f(\beta)$ 当且仅当 $\widehat{\beta}$ 是问题 min 的极小化解$_v\left\{f(v)+\frac{1}{2}\|v-\beta\|_2^2\right\}$，根据最优性条件（定理2.62）及次微分运算法则，这等价于关系 $0\in\partial f(\widehat{\beta})+\widehat{\beta}-\beta$.这就证明了 (i) 和 (ii) 等价。最后，根据次梯度的定义（定义2.57），命题 (ii) 中的隶属关系等价于 (iii)。 ■

当 $f=\delta_{\mathbb{S}}$，且 $\mathbb{S}$ 为非空闭凸集时，近端性质 I 中命题 (i) 与 (iii) 的等价性恰好归结为投影性质 I（引理 2.94）。

为与投影性质保持一致，我们也陈述近端性质二，尽管它在实践中较少使用。

引理2.101：（近端性质二） 设 $f:\mathbb{R}^p\to(-\infty,\infty]$ 为适当的**闭且凸**函数。那么对于任意 $\widehat{\beta}\triangleq\mathcal{P}\mathrm{rox}f(\beta)$ 和 $\xi\in\mathbb{R}^p$，可得 $\|\beta-\xi\|_2^2\ge\|\widehat{\beta}-\xi\|_2^2+\|\widehat{\beta}-\beta\|_2^2+2\left(f(\widehat{\beta})-f(\xi)\right)$.

引理 2.102：（近端性质-III） 设 $f:\mathbb{R}^p\to(-\infty,\infty]$ 是一个适当的 **闭凸**函数。那么 $\mathcal{P}\mathrm{rox}f(\beta)$ 对任意 **单元素集合** $\beta\in\mathbb{R}^p$。

引理 [的证明 2.102] 对任意 $\alpha\in\mathbb{R}^p$，

$$\mathcal{P}\mathrm{rox}_f(\beta)=\arg\min_{\alpha\in\mathbb{R}^p}g(\beta;\alpha),\tag{2.37}$$

其中 $g(\beta;\alpha)=f(\alpha)+\frac{1}{2}\|\alpha-\beta\|_2^2$。函数 $g(\beta;\alpha)$ 是闭强凸函数，因为它是闭强凸函数 $\frac{1}{2}\|\alpha-\beta\|_2^2$ 与闭凸函数 f 的和（练习 2.79 和练习 2.43）。$g(\beta;\alpha)$ 的适当性直接由 f 的适当性得出。因此，根据 SC 性质-II（定理 2.82；关于严格凸性下唯一性的结论也可参见定理 2.44），问题 (2.37) 存在唯一极小化解。 ■

近端性质-III 还证明了投影性质 III（引理 2.96），注意到 $\mathcal{P}\mathrm{rox}_{\delta_{\mathbb{S}}}(\beta)=\mathcal{P}_{\mathbb{S}}(\beta)$，对任意 $\beta\in\mathbb{S}$，若 $\mathbb{S}$是凸且闭的（示性函数是闭凸的，这由习题 2.36 和2.53可知）。

当 f 是真闭凸函数时，前述引理表明，对于任意 $\boldsymbol{\beta} \in \mathbb{R}^p$， $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$ 是一个单元素集合。在这些情形下——它们将构成本书讨论的大多数情形——我们将把 $\mathcal{P}\mathrm{rox}_f$ 视为从 $\mathbb{R}^p$ 到 $\mathbb{R}^p$ 的单值映射，也就是说，我们将直接写作 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \boldsymbol{\beta}$，而不写作 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \{\boldsymbol{\beta}\}$。

凸性与闭性。 请再次注意，近端性质 I、II 和 III 通常称为一阶性质；如果底层函数非凸，这些性质可能不成立。然而，投影性质-O 通常称为零阶性质，无论底层函数是否凸，始终成立。在所有情况下，我们都要求函数是闭的，以确保根据魏尔斯特拉斯定理（定理 2.56）能够取到极值。

定理 2.103：（投影与近端性质 IV：非扩张性） 设 $\mathbb{S} \subseteq \mathbb{R}^p$ 为闭的**凸集**，并设 $f : \mathbb{R}^p \to (-\infty, \infty]$ 为适当的**闭且凸**函数。则，

1. 坚定非扩张性。 对于任意 $\boldsymbol{\alpha}$， $\boldsymbol{\beta} \in \mathbb{R}^p$，可得

$$\|\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})\|_2^2 \leq \langle \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle ;$$
$$\|\mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) - \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})\|_2^2 \leq \langle \mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) - \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle .$$

2. 非扩张性。 对于任意 $\boldsymbol{\alpha}$， $\boldsymbol{\beta} \in \mathbb{R}^p$，可得

$$\|\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})\|_2 \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 ;$$
$$\|\mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) - \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})\|_2 \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 .$$

证明 [定理2.103] **投影算子。** 根据投影性质 I（引理 2.94），对于任意 $\boldsymbol{u} \in \mathbb{R}^p$ 和 $\boldsymbol{v} \in \mathbb{S}$，有 $\langle \boldsymbol{v} - \mathcal{P}_{\mathbb{S}}(\boldsymbol{u}) \ \ \boldsymbol{u} - \mathcal{P}_{\mathbb{S}}(\boldsymbol{u}) \rangle \leq 0$. 将此不等式应用于 $\boldsymbol{u} \triangleq \boldsymbol{\alpha}$、$\boldsymbol{v} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$ 和 $\boldsymbol{u} \triangleq \boldsymbol{\beta}$、$\boldsymbol{v} \triangleq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha})$：

$$\langle \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}), \boldsymbol{\alpha} - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) \rangle \leq 0 \qquad \text{and} \qquad \langle \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}), \boldsymbol{\beta} - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) \rangle \leq 0.$$

将两个不等式相加可得 $\langle \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) \ \ \boldsymbol{\alpha} - \boldsymbol{\beta} + \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) \rangle \leq 0$，这表明投影算子具有坚定非扩张性。

为证明非扩张性，注意到如果 $\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$，则不等式显然成立。因此，我们假设 $\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) \neq \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})$。根据柯西–施瓦茨不等式，有

$$\langle \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle \leq \|\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) - \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta})\|_2 \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 ,$$

将其与坚定非扩张性结合，即可得到所需结果。

近端算子。 记 $\boldsymbol{u} \triangleq \mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha})$， $\boldsymbol{v} \triangleq \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$，由近端性质 I（引理2.100）中 (i) 与 (ii) 的等价性可得

$$\boldsymbol{\alpha} - \boldsymbol{u} \in \partial f(\boldsymbol{u}) \qquad \text{and} \qquad \boldsymbol{\beta} - \boldsymbol{v} \in \partial f(\boldsymbol{v}).$$

根据次梯度不等式，我们有

$$f(\boldsymbol{v}) \geq f(\boldsymbol{u}) + \langle \boldsymbol{\alpha} - \boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u} \rangle \qquad \text{and} \qquad f(\boldsymbol{u}) \geq f(\boldsymbol{v}) + \langle \boldsymbol{\beta} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \rangle .$$

将这两个不等式相加，得到 $0 \geq \langle \boldsymbol{\beta} - \boldsymbol{\alpha} + \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \rangle$，这表明近端算子具有牢固的非扩张性。

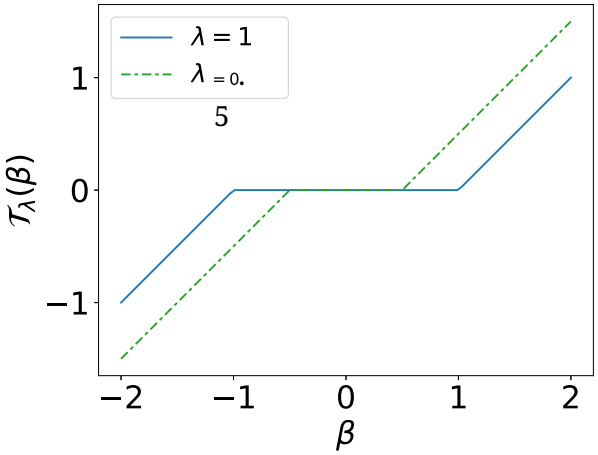


图 2.8： 不同 λ取值下软阈值函数 $\mathcal{T}_\lambda(\beta)$的示意图。

为证明非扩张性，若 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) = \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$，则该不等式显然成立。因此，我们假设 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) \neq \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})$。根据柯西 – 施瓦茨不等式，可得

$$\langle \mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) - \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle \leq \|\mathcal{P}\mathrm{rox}_f(\boldsymbol{\alpha}) - \mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta})\|_2 \cdot \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 ,$$

结合坚定非扩张性即可得到所需结果。 ■

最后请注意，投影算子的（坚定）非扩张性是近端算子结果的一个特例。事实上，令 $f \triangleq \delta_{\mathbb{S}}$ （$\mathbb{S}$的示性函数），则有$\mathcal{P}\mathrm{rox}_f = \mathcal{P}_{\mathbb{S}}$，且由于当 $\delta_{\mathbb{S}}$ 非空、闭且凸时， $\mathbb{S}$ 是适当、闭且凸的（习题 2.36和 2.53），近端非扩张性定理直接推出投影版本。

例 2.104（软阈值处理，ℓ_1-范数的近端算子）。令 $f(\boldsymbol{\alpha}) = \lambda \|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^p \psi(\alpha_i)$，其中 $\boldsymbol{\alpha} \in \mathbb{R}^p$、 $\lambda > 0$ 和 $\psi(u) = \lambda |u|$。则 $\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \{\mathcal{P}\mathrm{rox}_{\psi}(\beta_i)\}_{i=1}^p$，其中 $\mathcal{P}\mathrm{rox}_{\psi}(\beta) = \mathcal{T}_\lambda(\beta)$，而 $\mathcal{T}_\lambda$ 称为软阈值函数，定义为

$$\mathcal{T}_\lambda(\beta) \triangleq [|\beta| - \lambda]_+ \cdot \mathrm{sign}(\beta) = \begin{cases} \beta - \lambda, & \beta \geq \lambda; \\ 0, & |\beta| < \lambda; \\ \beta + \lambda, & \beta \leq -\lambda. \end{cases} \tag{2.38}$$

其中 $[u]_+ = \max\{u, 0\}$ 对任意 $u \in \mathbb{R}$，且符号函数定义为 $\mathrm{sign}(u) = u/|u|$当 $u \neq 0$时，这是通常的约定。有关该软阈值函数的示意图，请参见图 2.8 。更简洁地， f 的近端操作可表示为

$$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) = \mathcal{T}_\lambda(\boldsymbol{\beta}) = \{\mathcal{T}_\lambda(\beta_i)\}_{i=1}^p = [|\boldsymbol{\beta}| - \lambda \mathbf{1}]_+ \circ \mathrm{sign}(\boldsymbol{\beta}),$$

其中 $\circ$ 表示哈达玛积，$\mathbf{1}$是全1向量。 □

例2.105（块软阈值）。函数$f(\boldsymbol{\alpha}) = \lambda \|\boldsymbol{\alpha}\|_2$在 $\boldsymbol{\alpha} \in \mathbb{R}^p$下的缩放 ℓ_2-范数近端算子被称为块软阈值函数。它常用于组稀疏性（见9.4.2节）。该近端算子

具有闭式解：

$$\mathcal{P}\mathrm{rox}_{\lambda\|\cdot\|_2}(\boldsymbol{\beta}) = \left[1 - \frac{\lambda}{\|\boldsymbol{\beta}\|_2}\right]_+ \boldsymbol{\beta} = \begin{cases} \left(1 - \dfrac{\lambda}{\|\boldsymbol{\beta}\|_2}\right)\boldsymbol{\beta}, & \text{if } \|\boldsymbol{\beta}\|_2 > \lambda; \\ \mathbf{0}, & \text{if } \|\boldsymbol{\beta}\|_2 \le \lambda, \end{cases} \tag{2.39}$$

其中 $[u]_+ = \max(u, 0)$。为说明这一点，我们注意到目标函数具有径向对称性，因此极小化解 $\boldsymbol{\alpha}^*$ 必与 $\boldsymbol{\beta}$共线：$\boldsymbol{\alpha}^* = \gamma\boldsymbol{\beta}$，其中 $\gamma \ge 0$为某个标量。代入目标函数：

$$\lambda\|\gamma\boldsymbol{\beta}\|_2 + \frac{1}{2}\|\gamma\boldsymbol{\beta} - \boldsymbol{\beta}\|_2^2 = \lambda\gamma\|\boldsymbol{\beta}\|_2 + \frac{1}{2}(\gamma - 1)^2\|\boldsymbol{\beta}\|_2^2.$$

对 $\gamma \ge 0$求最小值，我们对 γ求导：

$$\lambda\|\boldsymbol{\beta}\|_2 + (\gamma - 1)\|\boldsymbol{\beta}\|_2^2 = 0 \quad \Longrightarrow \quad \gamma = 1 - \frac{\lambda}{\|\boldsymbol{\beta}\|_2}.$$

然而，由于 $\gamma \ge 0$，该解仅在 $\|\boldsymbol{\beta}\|_2 > \lambda$时有效。若 $\|\boldsymbol{\beta}\|_2 \le \lambda$，最小值在 $\gamma = 0$处取得，从而得到 $\boldsymbol{\alpha}^* = \mathbf{0}$。

作为一个特例，当维数 $p = 1$时，则 $\|\beta\|_2 = |\beta|$ 且

$$\mathcal{P}\mathrm{rox}_{\lambda|\cdot|}(\beta) = \mathrm{sign}(\beta)\max(|\beta| - \lambda, 0),$$

这就归结为示例 2.104中的标量软阈值算子。 □

2.8.4 莫罗分解

近端算子的一个关键性质是所谓的莫罗分解定理，它建立了适当闭凸函数的近端算子与其凸共轭（定义 2.63）的近端算子之间的关系。

定理 2.106：（莫罗分解）设 $f\colon \mathbb{R}^p \to (-\infty, \infty]$ 是一个真闭凸函数。那么，对于所有 $\boldsymbol{\beta} \in \mathbb{R}^p$，

$$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) + \mathcal{P}\mathrm{rox}_{f^*}(\boldsymbol{\beta}) = \boldsymbol{\beta}. \tag{2.40}$$

定理 [的证明 2.106] 设 $\widehat{\boldsymbol{\beta}} \triangleq \mathcal{P}$近端算子$_f(\boldsymbol{\beta})$。根据近端性质 I（引理 2.100）中断言 **(i)** $\widehat{\ }$ $\widehat{\ }$与 **(ii)** 的等价性，可得 $\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}} \in \partial f(\boldsymbol{\beta})$，根据共轭次梯度定理（定理 2.70）这等价于 $\boldsymbol{\beta} \in \partial f^*(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$。再次利用近端性质 I（引理 2.100），可得 $\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}} = \mathcal{P}$近端算子$_{f^*}(\boldsymbol{\beta})$。因此，

$$\mathcal{P}\mathrm{rox}_f(\boldsymbol{\beta}) + \mathcal{P}\mathrm{rox}_{f^*}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}.$$

证明至此完成。 ■

下一个结果给出了莫罗分解定理的一个有用推广。

定理 2.107：（扩展 Moreau 分解）令 $f\colon \mathbb{R}^p \to (-\infty, \infty]$ 为一个正常闭凸函数，并令 $\eta > 0$。那么，对于所有 $\boldsymbol{\beta} \in \mathbb{R}^p$，

$$\mathcal{P}\mathrm{rox}_{\eta f}(\boldsymbol{\beta}) + \eta\mathcal{P}\mathrm{rox}_{\eta^{-1}f^*}(\boldsymbol{\beta}/\eta) = \boldsymbol{\beta}. \tag{2.41}$$

证明 [定理 2.107] 利用莫罗分解，对于任意 $\boldsymbol{\beta} \in \mathbb{R}^p$，

$$\mathcal{P}\mathrm{rox}_{\eta f}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \mathcal{P}\mathrm{rox}_{(\eta f)^*}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \mathcal{P}\mathrm{rox}_{\eta f^*(\cdot/\eta)}(\boldsymbol{\beta}), \tag{2.42}$$

其中，第二个等式由共轭函数运算（习题 2.68）得出。根据近端演算（习题 2.92），

$$\mathcal{P}\mathrm{rox}_{\eta f^*(\cdot/\eta)}(\boldsymbol{\beta}) = \eta \mathcal{P}\mathrm{rox}_{\eta^{-1} f^*}(\boldsymbol{\beta}/\eta),$$

将其与（2.42）结合，即可得到所需的恒等式。 ■

2.9. 凸优化与最优性条件

我们在前几章中已经接触过优化问题的一般形式。更确切地说，凸优化问题（或简称凸问题）是指在凸集上最小化凸函数：

$$\textbf{(Convex optimization)}: \quad \begin{array}{ll} \min & f(\boldsymbol{\theta}) \quad \text{(convex function)} \\ \text{s.t.} & \boldsymbol{\theta} \in \mathbb{S} \quad \text{(convex set)}, \end{array} \tag{2.43}$$

其中，函数 $f: \mathbb{R}^p \to (-\infty, \infty]$ 称为目标函数（或损失函数），而 $\mathbb{S}$ 是约束集。满足约束条件的点 $\boldsymbol{\theta}$ 称为可行点。如果至少存在一个可行点，则称问题（2.43）是可行的。能够取得最小值的可行点$\boldsymbol{\theta}^*$ ——即对于所有可行的 $\boldsymbol{\theta}$，都有 $f(\boldsymbol{\theta}^*) \leq f(\boldsymbol{\theta})$——称为极小化解、最小点或最优点。值 $f(\boldsymbol{\theta}^*)$称为最优值。

例2.108（等式约束与不等式约束）。考虑以下问题：

$$\begin{array}{ll} \min & f(\boldsymbol{\theta}) \\ \text{s.t.} & g_i(\boldsymbol{\theta}) \leq 0, \quad i = \{1, 2, \ldots, m\}, \\ & h_j(\boldsymbol{\theta}) = 0, \quad j = \{1, 2, \ldots, n\}, \end{array} \tag{2.44}$$

其中 f，g_1，g_2，……，g_m：$\mathbb{R}^p \to \mathbb{R}$ 是凸函数，而 h_1，h_2，……，h_n：$\mathbb{R}^p \to \mathbb{R}$ 是仿射函数。由于目标函数是凸函数，且可行集是凸集，因此该问题是一个凸优化问题。可行集可以表示为

$$\mathbb{S} = \left(\bigcap_{i=1}^{m} \mathrm{Lev}[g_i, 0] \right) \cap \left(\bigcap_{j=1}^{n} \{\boldsymbol{\theta} \mid h_j(\boldsymbol{\theta}) = 0\} \right),$$

这是一个凸集，因为它是凸函数的水平集与超平面的交集，而这两者也都是凸集。□

例2.109（线性规划（LP））。线性规划问题涉及在满足线性等式约束和不等式约束的条件下，最小化线性目标函数：

$$\text{(LP)} \qquad \begin{array}{ll} \min & \boldsymbol{c}^\top \boldsymbol{\theta} + d \\ \text{s.t.} & \boldsymbol{G}\boldsymbol{\theta} \leq \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b}, \end{array}$$

其中 $\boldsymbol{G} \in \mathbb{R}^{n \times p}$、$\boldsymbol{h} \in \mathbb{R}^n$、$\boldsymbol{A} \in \mathbb{R}^{m \times p}$、$\boldsymbol{b} \in \mathbb{R}^m$ 和 $\boldsymbol{c} \in \mathbb{R}^p$。（为方便起见，如果对所有 i 都有 $a_i \leq 0$，我们记作 $\boldsymbol{a} \leq \boldsymbol{0}$。）这是一个凸优化问题，因为仿射函数

本质上是凸的。目标函数中的常数项 d 不影响极小化解的位置，在确定最优解时可以省略。一种常用的表述是线性规划的“标准形式”：

$$\text{(Standard LP)} \qquad \begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{\theta} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b}, \\ & \boldsymbol{\theta} \geq \mathbf{0}. \end{aligned}$$

任何线性规划都可以转换为这种标准形式。为此，我们引入非负松弛变量 $\boldsymbol{s}$，将不等式转化为等式约束： $\boldsymbol{G}\boldsymbol{\theta} + \boldsymbol{s} = \boldsymbol{h}$。此外，我们将$\boldsymbol{\theta}$ 表示为两个非负变量 $\boldsymbol{\theta} \triangleq \boldsymbol{\theta}^+ - \boldsymbol{\theta}^-$之差，其中 $\boldsymbol{\theta}^+$， $\boldsymbol{\theta}^- \geq \mathbf{0}$:

$$\text{(LP}') \qquad \begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{\theta}^+ - \boldsymbol{c}^\top \boldsymbol{\theta}^- + d \\ \text{s.t.} \quad & \boldsymbol{G}\boldsymbol{\theta}^+ - \boldsymbol{G}\boldsymbol{\theta}^- + \boldsymbol{s} = \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{\theta}^+ - \boldsymbol{A}\boldsymbol{\theta}^- = \boldsymbol{b}, \\ & \boldsymbol{s} \geq \mathbf{0}, \boldsymbol{\theta}^+ \geq \mathbf{0}, \boldsymbol{\theta}^- \geq \mathbf{0}, \end{aligned} \iff \begin{aligned} \min \quad & [\boldsymbol{c}^\top, -\boldsymbol{c}^\top, \mathbf{0}]\widetilde{\boldsymbol{\theta}} + d \\ \text{s.t.} \quad & [\boldsymbol{G}, -\boldsymbol{G}, \boldsymbol{I}]\widetilde{\boldsymbol{\theta}} = \boldsymbol{h}, \\ & [\boldsymbol{A}, -\boldsymbol{A}, \mathbf{0}]\widetilde{\boldsymbol{\theta}} = \boldsymbol{b}, \\ & \widetilde{\boldsymbol{\theta}} \geq \mathbf{0}, \end{aligned}$$

其中 $\widetilde{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \boldsymbol{\theta}^+ \\ \boldsymbol{\theta}^- \\ \boldsymbol{s} \end{bmatrix}$。这种重新表述采用了标准线性规划形式，其决策变量为 $\widetilde{\boldsymbol{\theta}}$。对于（LP）的每个可行解 $\boldsymbol{\theta}^*$，我们可以令 $\boldsymbol{s} \triangleq \boldsymbol{h} - \boldsymbol{G}\boldsymbol{\theta}^*$、 $\theta_i^+ \triangleq \max\{0, \theta_i\}$，并对所有 $i \in \{1, 2, \ldots, p\}$令$\theta_i^- \triangleq \max\{0, -\theta_i\}$，从而得到（LP′）的一个可行解。这表明，（LP′）的最优值小于或等于（LP）的最优值。反之，假设 $\boldsymbol{\theta}^+$， $\boldsymbol{\theta}^-$以及 $\boldsymbol{s}$ 是（LP′）的可行解。那么， $\boldsymbol{\theta} \triangleq \boldsymbol{\theta}^+ - \boldsymbol{\theta}^-$ 是（LP）的一个可行解，这表明（LP）的最优值小于或等于（LP′）的最优值。因此，优化问题（LP）和（LP′）是等价的。□

例 2.110（二次规划）。凸优化问题（2.44）称为二次规划（*QP*），当目标函数是半正定（二次凸）函数，且约束函数为仿射函数时。一般形式可写为:

$$\begin{aligned} \min \quad & \boldsymbol{\theta}^\top \boldsymbol{P}\boldsymbol{\theta} + 2\boldsymbol{q}^\top \boldsymbol{\theta} + r \\ \text{s.t.} \quad & \boldsymbol{G}\boldsymbol{\theta} \leq \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b}, \end{aligned}$$

其中 $\boldsymbol{P} \in \mathbb{R}^{p \times p}$ 是半正定的， $\boldsymbol{q} \in \mathbb{R}^p$、 $\boldsymbol{G} \in \mathbb{R}^{n \times p}$、 $\boldsymbol{h} \in \mathbb{R}^n$、 $\boldsymbol{A} \in \mathbb{R}^{m \times p}$和$\boldsymbol{b} \in \mathbb{R}^m$。如果目标函数和不等式约束都是凸二次函数:

$$\begin{aligned} \min \quad & \boldsymbol{\theta}^\top \boldsymbol{P}\boldsymbol{\theta} + 2\boldsymbol{q}^\top \boldsymbol{\theta} + r \\ \text{s.t.} \quad & \boldsymbol{\theta}^\top \boldsymbol{G}_i \boldsymbol{\theta} + 2\boldsymbol{h}_i^\top \boldsymbol{\theta} + c_i \leq 0, \quad i = \{1, 2, \ldots, k\}, \\ & \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b}, \end{aligned}$$

其中 $\boldsymbol{P}$， $\boldsymbol{G}_1$， $\boldsymbol{G}_2$， ……， $\boldsymbol{G}_k$均为半正定矩阵，这类问题称为二次约束二次规划（*QCQP*）。在 QCQP 中，我们在由椭球体（当每个 $\boldsymbol{G}_i \succ \mathbf{0}$ 时）与仿射子空间的交集所定义的可行域上最小化一个凸二次函数。□

2.9.1 凸集上的约束优化

我们首先将注意力限制在定义于一个**凸**集合$\mathbb{S}$上的优化问题。在这种情况下，如果目标函数 f 也是凸函数，那么该问题就成为前文讨论过的凸优化问题。在前面的推论2.61 或定理2.62中，我们为无约束问题引入了驻点——具体而言，即梯度为零的点 $\boldsymbol{\beta}^*$ $\nabla f(\boldsymbol{\beta}^*) = \mathbf{0}$，或者当函数不光滑时满足 $\mathbf{0} \in \partial f(\boldsymbol{\beta}^*)$ 的点。然而，对于约束问题，这一定义已不再充分，因此需要一种更一般的概念。

定义 2.111（凸集上约束问题的驻点）。考虑约束优化问题

$$(\mathrm{P}_C) \quad \min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{S}. \tag{2.45}$$

设 f 是定义在闭**凸** 集合 $\mathbb{S}$ 上的连续可微函数。那么，如果对于任意 $\boldsymbol{\beta} \in \mathbb{S}$， $\boldsymbol{\beta}^* \in \mathbb{S}$ 满足 $\nabla f(\boldsymbol{\beta}^*)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*) \geq 0$，则称其为 (P_C) 的驻点。

下面的结果表明，约束优化问题的局部最小点必定是驻点。

定理 2.112：（凸集上 (P_C) 的一阶必要条件）考虑约束优化问题 (P_C)：$\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$，满足约束 $\boldsymbol{\beta} \in \mathbb{S}$。设 f 是定义在闭**凸** 集合 $\mathbb{S}$ 上的连续可微函数，且 $\boldsymbol{\beta}^*$ 是 (P_C) 的局部最小值点。那么 $\boldsymbol{\beta}^*$ 是 (P_C) 的驻点。

定理 2.112] 的证明 [设 $\boldsymbol{\beta}^*$ 是 (P_C) 的一个局部最小值，并反设 $\boldsymbol{\beta}^*$ 不是 (P_C) 的驻点。那么存在某个 $\boldsymbol{\beta} \in \mathbb{S}$ 使得$\nabla f(\boldsymbol{\beta}^*)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*) < 0$。由引理 2.7 可知，存在 $\varepsilon \in (0, 1)$ 使得$f(\boldsymbol{\beta}^* + \eta \boldsymbol{d}) < f(\boldsymbol{\beta}^*)$ 对所有 $\eta \in (0, \varepsilon)$ 都成立，其中 $\boldsymbol{d} \triangleq \boldsymbol{\beta} - \boldsymbol{\beta}^*$。由于 $\mathbb{S}$是凸的，因此有 $\boldsymbol{\beta}^* + \eta \boldsymbol{d} = (1 - \eta)\boldsymbol{\beta}^* + \eta \boldsymbol{\beta} \in \mathbb{S}$，这与 $\boldsymbol{\beta}^*$ 是 (P_C) 的局部最小点这一假设矛盾。因此， $\boldsymbol{\beta}^*$ 必为驻点。 ■

当到凸集 $\mathbb{S}$的投影存在闭式表达式时，一阶条件可以等价地用投影算子表述。由此得到以下刻画。

推论 2.113：（在凸集上，关于投影的（$\mathbf{P}_C$）的必要性/充分性）考虑约束优化问题（P_C）：$\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$，满足 $\boldsymbol{\beta} \in \mathbb{S}$。设 f 是闭**凸**集 $\mathbb{S}$上的连续可微函数，且 $\eta > 0$ 是任意正标量。那么，当且仅当 $\boldsymbol{\beta}^*$ 是（P_C）的驻点时，

$$\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}^* - \eta \nabla f(\boldsymbol{\beta}^*)).$$

证明 [推论2.113] 根据投影性质 I（引理2.94），可得$\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}$（$\boldsymbol{\beta}^* - \eta \nabla f$（$\boldsymbol{\beta}^*$）） 当且仅当

$$\langle \boldsymbol{\beta}^* - \eta \nabla f(\boldsymbol{\beta}^*) - \boldsymbol{\beta}^*, \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle \leq 0 \quad \Longrightarrow \quad \nabla f(\boldsymbol{\beta}^*)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*) \geq 0, \text{ for any } \boldsymbol{\beta} \in \mathbb{S}.$$

这正是定义 2.111中对驻点的定义，证明完毕。 ■

2.9.2 等式/不等式约束

现在考虑例 2.108中提出的问题。此类优化问题通常表述为

$$\begin{aligned} &\min_{\boldsymbol{\beta}\in\mathbb{R}^p} f_0(\boldsymbol{\beta}) \\ \text{s.t.}\quad &\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}, \\ &f_i(\boldsymbol{\beta})\le z_i,\quad i\in\{1,2,\dots,q\}, \end{aligned} \tag{P}$$

其中，函数 $f_0:\mathbb{R}^p\to(-\infty,\infty]$ 再次称为目标函数或损失函数，函数 $f_1,f_2,\dots,f_q:\mathbb{R}^p\to(-\infty,\infty]$ 称为（不等式）约束函数，而 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$, $\boldsymbol{y}\in\mathbb{R}^n$ 定义了等式约束。注意，如果 $f_0,f_1,\dots,f_q$都是凸函数，那么问题（P）是一个**凸**优化问题。

等式约束可以等价地用不等式约束表示。具体而言，对于 $\boldsymbol{X}$ 的每一行 $\boldsymbol{\alpha}_i^\top$（即第 i 行），条件 $\boldsymbol{\alpha}_i^\top\boldsymbol{\beta}=y_i$ 等价于以下两个不等式

$$\boldsymbol{\alpha}_i^\top\boldsymbol{\beta}\le y_i \qquad\text{and}\qquad -\boldsymbol{\alpha}_i^\top\boldsymbol{\beta}\le -y_i.$$

由这些约束描述的可行点集合可以表示为

$$\mathbb{S}=\{\boldsymbol{\beta}\in\mathbb{R}^p\mid \boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}, f_i(\boldsymbol{\beta})\le z_i, i\in\{1,2,\dots,q\}\}. \tag{2.46}$$

因此，问题（P）等价于在可行集 f_0 上最小化 $\mathbb{S}$,

$$\min_{\boldsymbol{\beta}\in\mathbb{S}} f_0(\boldsymbol{\beta}). \tag{2.47}$$

此外，通过引入示性函数 $\delta_{\mathbb{S}}$ （参见例2.36），可以将该约束问题重新表述为一个包含复合目标函数的无约束优化问题

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta})\triangleq f_0(\boldsymbol{\beta})+\delta_{\mathbb{S}}(\boldsymbol{\beta}). \tag{2.48}$$

2.9.3 对偶问题

与形式为（$\boldsymbol{P}$）的优化问题相关联的拉格朗日函数（或简称拉格朗日函数）定义在 $\boldsymbol{\beta}\in\mathbb{R}^p$, $\boldsymbol{\lambda}\in\mathbb{R}^n$, $\boldsymbol{\nu}\in\mathbb{R}^q$, $\nu_i\ge 0$, $i\in\{1,2,\dots,q\}$ 上（为简洁起见，当 $\boldsymbol{\nu}\ge\mathbf{0}$ 的所有分量均非负时，我们记作 $\boldsymbol{\nu}$），定义如下:

$$L(\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\nu})\triangleq f_0(\boldsymbol{\beta})+\boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})+\sum_{i=1}^{q}\nu_i(f_i(\boldsymbol{\beta})-z_i)\triangleq f_0(\boldsymbol{\beta})+\boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})+\boldsymbol{\nu}^\top\boldsymbol{f}, \tag{2.49}$$

其中，向量 $\boldsymbol{\lambda}$ 和 $\boldsymbol{\nu}$ 称为拉格朗日乘子，且 $\boldsymbol{f}\triangleq\{f_i(\boldsymbol{\beta})-z_i\}_{i=1}^q\in\mathbb{R}^q$。拉格朗日对偶函数（或简称对偶函数）定义为

$$D(\boldsymbol{\lambda},\boldsymbol{\nu})\triangleq\min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\nu}),\quad \boldsymbol{\lambda}\in\mathbb{R}^n,\boldsymbol{\nu}\in\mathbb{R}^q,\boldsymbol{\nu}\ge\mathbf{0}. \tag{2.50}$$

如果函数 $\boldsymbol{\beta}\mapsto L(\boldsymbol{\beta},\ \boldsymbol{\lambda},\ \boldsymbol{\nu})$ 的下界不存在，则设 $D(\boldsymbol{\lambda},\ \boldsymbol{\nu})=-\infty$。重要的是，对偶函数始终是**凹的**，因为它是关于 $(\boldsymbol{\lambda},\ \boldsymbol{\nu})$ 的一族仿射函数的逐点下确界，即使原问题 (P) 不是凸的（参见习题 2.43，注意凹函数的相反数是凸函数）。

现在，假设 $\boldsymbol{\beta}$ 是问题 (P) 的一个可行点。那么 $\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}=\mathbf{0}$ 并且 $f_i(\boldsymbol{\beta}) \leq z_i$, i =1、2、.....、q。由于 $\boldsymbol{\nu} \geq \mathbf{0}$, 因此

$$\begin{aligned} &\boldsymbol{\lambda}^{\top}(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})+\boldsymbol{\nu}^{\top}\boldsymbol{f} \leq 0 \\ \Longrightarrow\ & L(\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\nu})=f_0(\boldsymbol{\beta})+\boldsymbol{\lambda}^{\top}(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})+\boldsymbol{\nu}^{\top}\boldsymbol{f} \leq f_0(\boldsymbol{\beta}). \end{aligned}$$

对左侧所有 $\boldsymbol{\beta} \in \mathbb{R}^p$ 取下确界（得到 $D(\boldsymbol{\lambda}, \boldsymbol{\nu})$），并对右侧所有可行的 $\boldsymbol{\beta}$ 取下确界（得到最优值 $f_0(\boldsymbol{\beta}^*)$），可得基本不等式

$$D(\boldsymbol{\lambda},\boldsymbol{\nu}) \leq f_0(\boldsymbol{\beta}^*), \quad \text{for all } \boldsymbol{\lambda} \in \mathbb{R}^n, \boldsymbol{\nu} \geq \mathbf{0}. \tag{2.51}$$

因此，对偶函数为原始问题的最优值提供了一个下界。为了获得尽可能紧的下界，我们将该对偶函数最大化，由此得到相对于对偶问题（P）的:

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n,\boldsymbol{\nu}\geq\mathbf{0}} D(\boldsymbol{\lambda},\boldsymbol{\nu}). \tag{D}$$

因此，原问题（P）通常被称为原始问题。

强对偶性

由于问题（D）涉及最大化一个凹函数，因此它等价于最小化凸函数 $-D$ 且满足正性约束 $\boldsymbol{\nu} \geq \mathbf{0}$ 的凸优化问题。满足 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 和 $\boldsymbol{\nu} \geq \mathbf{0}$ 的一对（$\boldsymbol{\lambda}$, $\boldsymbol{\nu}$）称为对偶可行。（对偶可行的）问题（D）的最大化解（$\boldsymbol{\lambda}^*$, $\boldsymbol{\nu}^*$）称为对偶最优解，或最优拉格朗日乘子。如果 $\boldsymbol{\beta}^*$ 是原始问题（P）的最优解，则三元组（$\boldsymbol{\beta}^*$, $\boldsymbol{\lambda}^*$, $\boldsymbol{\nu}^*$）称为原始-对偶最优解。不等式（2.51）意味着，对于对偶最优解（$\boldsymbol{\lambda}^*$, $\boldsymbol{\nu}^*$）和原始最优解 $\boldsymbol{\beta}^*$,

$$D(\boldsymbol{\lambda}^*,\boldsymbol{\nu}^*) \leq f_0(\boldsymbol{\beta}^*). \tag{2.52}$$

这种关系称为弱对偶性。对于大多数（但并非所有）凸优化问题，强对偶性 成立，即等号成立:

$$D(\boldsymbol{\lambda}^*,\boldsymbol{\nu}^*) = f_0(\boldsymbol{\beta}^*). \tag{2.53}$$

参见问题 2.12∼2.14 ，了解线性规划的强对偶性。

凸问题中强对偶性成立的一个著名充分条件是 Slater 条件，下面将以简化形式陈述。

定理 2.114：（凸问题的 Slater 条件资格）假设 f_0、f_1、f_2、.....、f_q 是凸函数，且 $\mathrm{dom}(f_0) = \mathbb{R}^p$。若存在 $\boldsymbol{\beta} \in \mathbb{R}^p$ 使得 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 且 $f_i(\boldsymbol{\beta}) < 0$ 对所有 $i \in \{1, 2, \ldots, q\}$都成立，则优化问题（P）具有强对偶性。

证明 例如，参见Bertsekas 等人 的第 6 章，或参见Lu (2025年）. ■

我们还回顾一些众所周知的最优性条件，这些条件在强对偶性成立时用拉格朗日函数表示

定理 2.115：（KKT 条件：强对偶性下的最优性）假设强对偶性成立；也就是说，问题（P）和（D）的最优值有限且$\widehat{\quad}$$\widehat{\quad}$相等。那么，$\boldsymbol{\beta}$ 和（$\boldsymbol{\lambda}$，$\widehat{\boldsymbol{\nu}}$分别是原始问题和对偶问题的最优解，当且仅当满足以下条件：

$\widehat{\quad}$$\widehat{\quad}$$\widehat{\quad}$(i) 原问题可行性。$\boldsymbol{\beta}$ 是 (P) 的可行解，即 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 且 $f_i(\boldsymbol{\beta}) \leq z_i$ 对所有 $i = 1, 2, \ldots, q$ 均成立。(ii) 对偶可行性。$\widehat{\boldsymbol{\nu}} \geq \mathbf{0}$。$\widehat{\quad}$(iii) 互补松弛。$\widehat{\nu}_i(f_i(\boldsymbol{\beta}) - z_i) = 0$ 对所有 $i = 1, 2, \ldots, q$ 均成立。$\widehat{\quad}$$\widehat{\quad}$$\widehat{\quad}$(iv) 驻点条件。$\boldsymbol{\beta} \in$ 最小值点集合$_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \widehat{\boldsymbol{\nu}})$。

$\widehat{\quad}$$\widehat{\quad}$**证明** [定理 2.115]**必要性**。令 f 和 d 分别表示原问题$\widehat{\quad}$$\widehat{\quad}$$\widehat{\quad}$和对偶问题的最优值。根据假设，强对偶性成立，因此 $f = d$。若 $\boldsymbol{\beta}$ 和 $\widehat{\quad}(\boldsymbol{\lambda}, \widehat{\boldsymbol{\nu}})$ 分别是问题 (P) 和 (D) 的最优解，则 (i) 和 (ii) 立即由可行性得出。此外，

$$\begin{aligned}\widehat{f} = \widehat{d} = D(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) &= \min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) \leq L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) \\ &= f(\widehat{\boldsymbol{\beta}}) + \left\langle \widehat{\boldsymbol{\lambda}}, \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y} \right\rangle + \sum_{i=1}^{q} \widehat{\nu}_i \big(f_i(\widehat{\boldsymbol{\beta}}) - z_i\big) \leq f(\widehat{\boldsymbol{\beta}}),\end{aligned}$$

$\widehat{\quad}$$\widehat{\quad}$$\widehat{\quad}$$\widehat{\quad}$其中，最后一个不等式由以下事实推出：$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$，且 $f_i(\boldsymbol{\beta}) \leq z_i$。由于 $f = f(\boldsymbol{\beta})$，$\widehat{\quad}$$\widehat{\quad}$上面的所有不等式都必须是等式。这意味着 $\boldsymbol{\beta} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \widehat{\boldsymbol{\nu}})$，即性质 (iv)；并且 $\sum_{i=1}^{q} \widehat{\nu}_i(f_i(\widehat{\boldsymbol{\beta}}) - z_i) = 0$。根据以下事实：$\widehat{\nu}_i(f_i(\widehat{\boldsymbol{\beta}}) - \widehat{\quad}\sim z_i) \leq 0$ 对任意 i 都成立，可推出 $\lambda_i(f_i(\boldsymbol{\beta}) - z_i) = 0$ 对任意 i 都成立，这表明性质 (iii)。

充分性。反过来，假设性质 (i)-(iv) 成立。那么，

$$\begin{aligned}D(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) &= \min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) = L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\nu}}) \\ &= f(\widehat{\boldsymbol{\beta}}) + \left\langle \widehat{\boldsymbol{\lambda}}, \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y} \right\rangle + \sum_{i=1}^{q} \widehat{\nu}_i (f_i(\widehat{\boldsymbol{\beta}}) - z_i) = f(\widehat{\boldsymbol{\beta}}).\end{aligned}$$

其中，第二个等式由性质 (iv) 得出，最后一个等式由$\widehat{\quad}$$\widehat{\quad}$性质 (i) 和 (iii) 得出。根据弱对偶定理，由于 $\boldsymbol{\beta}$ 和 ($\boldsymbol{\lambda}$，$\widehat{\boldsymbol{\nu}}$) 分别是原问题和对偶问题的可行解，且原问题与对偶问题的目标函数值相等，因此它们分别是对应问题的最优解。

■

备注 2.116（带等式约束的凸优化的 KKT 条件）。考虑仅包含等式约束的优化问题这一特殊情形：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} f_0(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y},$$

$\widehat{\quad}$其中 f_0 是**凸**且可微。因此，定理 2.114 和 2.115 表明 $\boldsymbol{\beta}$ 是$\widehat{\quad}$解，当且仅当 $\boldsymbol{\beta}$ 可行，且存在 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 使得：

$$\nabla_{\boldsymbol{\beta}} L(\widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda}) = \mathbf{0} = \nabla f_0(\widehat{\boldsymbol{\beta}}) - \boldsymbol{X}^\top \boldsymbol{\lambda},$$

其中，拉格朗日函数定义为 $L(\boldsymbol{\beta},\ \boldsymbol{\lambda}) = f_0(\boldsymbol{\beta}) + \langle \boldsymbol{\lambda},\ \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\rangle$，而向量$\boldsymbol{\lambda}$ 是与等式约束对应的拉格朗日乘子。几何解释是，$\nabla f_0(\boldsymbol{\beta}) \perp \mathcal{N}(\boldsymbol{X})$。

当 f_0 不可微时，驻点条件可推广为：$\widehat{\boldsymbol{\beta}}$ 是可行的，并且存在 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 使得

$$\boldsymbol{X}^\top \boldsymbol{\lambda} \text{ is a subgradient of } f_0 \text{ at } \widehat{\boldsymbol{\beta}}.$$

给定一个原始-对偶可行三元组（$\boldsymbol{\beta},\ \boldsymbol{\lambda},\ \boldsymbol{\nu}$）——也就是说，$\boldsymbol{\beta}$对于（P）是可行的，并且 $\boldsymbol{\lambda} \in \mathbb{R}^n$，$\boldsymbol{\nu} \geq \mathbf{0}$——原始-对偶间隙定义为

$$H(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = F(\boldsymbol{\beta}) - D(\boldsymbol{\lambda}, \boldsymbol{\nu}), \quad \text{with } F(\boldsymbol{\beta}) \triangleq f_0(\boldsymbol{\beta}) + \delta_{\mathbb{S}}(\boldsymbol{\beta}). \tag{2.54}$$

该量用于衡量 $\boldsymbol{\beta}$ 距离原始问题极小化解有多近，以及$\widehat{\ }\ \widehat{\ }(\boldsymbol{\lambda},\ \boldsymbol{\nu})$距离对偶问题极大化解有多近。如果（$\boldsymbol{\beta},\ \boldsymbol{\lambda},\ \widehat{\boldsymbol{\nu}}$）是原始-对偶最优解，并且强$\widehat{\ }\ \widehat{\ }$对偶性成立，则间隙消失：$H(\boldsymbol{\beta},\ \boldsymbol{\lambda},\ \widehat{\boldsymbol{\nu}}) = 0$。在实践中，原始-对偶间隙通常用作迭代优化算法中的停止准则。

为说明这些概念，我们现在推导 ℓ_∞最小化问题的对偶问题。（其对应的 ℓ_1最小化问题将在第3章中详细讨论。）

例 2.117（ℓ_∞-最小化问题）。考虑 ℓ_∞-最小化问题：

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_\infty \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{2.55}$$

与该问题相关的拉格朗日函数为

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \|\boldsymbol{\beta}\|_\infty + \langle \boldsymbol{\lambda}, (\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})\rangle .$$

The 相应的对偶函数通过对拉格朗日函数关于 $\boldsymbol{\beta}$取最小值获得：

$$D(\boldsymbol{\lambda}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \|\boldsymbol{\beta}\|_\infty + \left\langle \boldsymbol{X}^\top \boldsymbol{\lambda}, \boldsymbol{\beta} \right\rangle - \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle \right\}.$$

根据 Hölder 不等式，我们有 $\left\langle \boldsymbol{X}^\top \boldsymbol{\lambda},\ \boldsymbol{\beta} \right\rangle \leq \left\|\boldsymbol{X}^\top \boldsymbol{\lambda}\right\|_1 \|\boldsymbol{\beta}\|_\infty$。这意味着，如果 $\left\|\boldsymbol{X}^\top \boldsymbol{\lambda}\right\|_1 > 1$，那么存在 $\boldsymbol{\beta} \in \mathbb{R}^p$（通过适当选择其符号），使得 $\left\langle \boldsymbol{X}^\top \boldsymbol{\lambda},\ \boldsymbol{\beta} \right\rangle < -\|\boldsymbol{\beta}\|_\infty$。将 $\boldsymbol{\beta}$ 替换为 $\mu\boldsymbol{\beta}$，并令 $\mu \to \infty$，可知在这种情况下，$D(\boldsymbol{\lambda}) = -\infty$。另一方面，如果 $\left\|\boldsymbol{X}^\top \boldsymbol{\lambda}\right\|_1 \leq 1$，那么 $\|\boldsymbol{\beta}\|_\infty + \left\langle \boldsymbol{X}^\top \boldsymbol{\lambda},\ \boldsymbol{\beta} \right\rangle \geq 0$。取 $\boldsymbol{\beta} = \mathbf{0}$可得下确界，并且 $D(\boldsymbol{\lambda}) = -\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle$。因此，对偶问题等价于：

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} D(\boldsymbol{\lambda}) = \begin{cases} -\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle, & \text{if } \left\|\boldsymbol{X}^\top \boldsymbol{\lambda}\right\|_1 \leq 1; \\ -\infty, & \text{otherwise.} \end{cases}$$

由于只有在最大化 $D(\boldsymbol{\lambda})$为有限值时才有意义，对偶问题等价于

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} -\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top \boldsymbol{\lambda}\right\|_1 \leq 1. \tag{2.56}$$

等价地，可以将目标函数写为 $\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \langle \boldsymbol{\lambda},\ \boldsymbol{y}\rangle$ 将负号吸收到最大化运算中即可（但约束保持不变）。最后，根据 Slater 条件（定理2.114），只要原始问题（2.55）可行，这一原始-对偶对就满足强对偶性——因为目标函数 $\|\cdot\|_\infty$ 是凸的，且等式约束是仿射的。

□

鞍点性质

拉格朗日对偶性有一种自然的鞍点解释。为简便起见，我们考虑没有不等式约束的原始问题（P）；不过，同样的推理也可以直接推广到含有不等式约束的问题。

设 $(\boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)$ 是一个原始-对偶最优对。根据拉格朗日函数 L 的定义，我们有

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} L(\boldsymbol{\beta},\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} f_0(\boldsymbol{\beta}) + \boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}) = \begin{cases} f_0(\boldsymbol{\beta}), & \text{if } \boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}, \\ \infty, & \text{otherwise.} \end{cases} \tag{2.57}$$

换言之，只要 $\boldsymbol{\beta}$ 违反等式约束，该上确界就是无穷大。因此，原问题（P）的（可行）极小化解 $\boldsymbol{\beta}^*$ 满足

$$f_0(\boldsymbol{\beta}^*) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} L(\boldsymbol{\beta},\boldsymbol{\lambda}). \tag{2.58}$$

另一方面，根据拉格朗日对偶函数的定义，对偶最优向量$\boldsymbol{\lambda}^*$ 可取到

$$D(\boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta},\boldsymbol{\lambda}) \tag{2.59}$$

由弱对偶性可知，一般不等式

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta},\boldsymbol{\lambda}) \le \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} L(\boldsymbol{\beta},\boldsymbol{\lambda}). \tag{2.60}$$

当强对偶性成立时，等号成立：

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \min_{\boldsymbol{\beta}\in\mathbb{R}^p} L(\boldsymbol{\beta},\boldsymbol{\lambda}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} L(\boldsymbol{\beta},\boldsymbol{\lambda}). \tag{2.61}$$

在这种情况下，强对偶性成立时，最小化与最大化的顺序可以互换。这称为强极大-极小性质或鞍点性质。事实上，在强对偶性成立时，任意原始-对偶最优对 $(\boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)$ 都构成拉格朗日函数的一个鞍点：

$$L(\boldsymbol{\beta}^*,\boldsymbol{\lambda}) \le L(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*) \le L(\boldsymbol{\beta},\boldsymbol{\lambda}^*) \quad \text{for all } \boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\lambda}\in\mathbb{R}^n. \tag{2.62}$$

因此，只要强对偶性成立，联合求解原始问题和对偶问题就等价于寻找拉格朗日函数的一个鞍点。

2.9.4 复合函数

共轭函数自然地出现在对偶优化问题中，尤其是在著名的*Fenchel* 对偶定理中，下面我们将回顾该定理。考虑以下复合优化问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}) + G(\boldsymbol{\beta}).$$

本书讨论的所有相关优化问题都属于这一类；参见第3章。

例2.118。本书讨论的许多问题都采用这种形式。例如，令 $F(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2$和 $G(\boldsymbol{\beta}) = \lambda\|\boldsymbol{\beta}\|_1$，可得到拉格朗日 LASSO 问题（$\mathrm{L}_L$）；取 $G(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$和 $\widetilde{F}(\boldsymbol{z}) \triangleq F(\boldsymbol{X}\boldsymbol{\beta})$，其中 $\widetilde{F} = \delta_{\{\boldsymbol{y}\}}$（单元素集合 $\{\boldsymbol{y}\}$的示性函数），可得到 ℓ_1-最小化问题（P_1）；取 $G(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ 和

$\widetilde{F}(\boldsymbol{z}) \triangleq F(\boldsymbol{X}\boldsymbol{\beta})$，其中 $\widetilde{F}=\delta_{\mathbb{B}[\boldsymbol{y},\epsilon]}$（$\ell_2$-球，其中心为 $\boldsymbol{y}$，半径为 ϵ）会导出二次约束 ℓ_1-最小化问题（$\mathrm{P}_1,\ \epsilon$）。□

我们首先将原问题重新表述为约束优化问题：

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathbb{R}^p}\{F(\boldsymbol{\beta})+G(\boldsymbol{\alpha})\}\quad \text{s.t.}\quad \boldsymbol{\beta}=\boldsymbol{\alpha}.$$

于是，相应的拉格朗日函数为

$$L(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\lambda})=F(\boldsymbol{\beta})+G(\boldsymbol{\alpha})+\langle\boldsymbol{\lambda},\boldsymbol{\alpha}-\boldsymbol{\beta}\rangle=-[\langle\boldsymbol{\lambda},\boldsymbol{\beta}\rangle-F(\boldsymbol{\beta})]-[\langle-\boldsymbol{\lambda},\boldsymbol{\alpha}\rangle-G(\boldsymbol{\alpha})].$$

通过对原变量 $\boldsymbol{\beta},\boldsymbol{\alpha}$ 求拉格朗日函数的最小值得到对偶函数：

$$D(\boldsymbol{\lambda})=\min_{\boldsymbol{\beta},\boldsymbol{\alpha}}L(\boldsymbol{\beta},\boldsymbol{\alpha};\boldsymbol{\lambda})=-F^*(\boldsymbol{\lambda})-G^*(-\boldsymbol{\lambda}).$$

因此，相应的对偶问题——通常称为*Fenchel* 对偶——为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^p}\left\{-F^*(\boldsymbol{\lambda})-G^*(-\boldsymbol{\lambda})\right\}.$$

下文将陈述 Fenchel 对偶定理，该定理给出了原问题与其 Fenchel 对偶之间满足强对偶性的充分条件。

定理 2.119：（Fenchel 对偶定理；Rockafellar(1970年），定理 31.1）设 $F,\ G:\mathbb{R}^p\to(-\infty,\ \infty]$ 为适当凸函数。若 $\mathrm{relint}(\mathrm{dom}(F))\cap\mathrm{relint}(\mathrm{dom}(G))\neq\varnothing$，则

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{F(\boldsymbol{\beta})+G(\boldsymbol{\beta})\}=\max_{\boldsymbol{\lambda}\in\mathbb{R}^p}\left\{-F^*(\boldsymbol{\lambda})-G^*(-\boldsymbol{\lambda})\right\},$$

并且只要右侧的最大值有限，该最大值就能取到。

第2章习题

1. ℓ_0 **和** ℓ_1。对于任意标量 β，定义向量

$$\boldsymbol{u}_\beta\triangleq[\beta,\beta,\beta-1]^\top.$$

选择 β，使得 $\boldsymbol{u}_\beta$ 尽可能稀疏；也就是说，使 $\|\boldsymbol{u}_\beta\|_0$最小化。证明，使 $\|\boldsymbol{u}_\beta\|_0$ 最小化等价于使 $\|\boldsymbol{u}_\beta\|_1$ 最小化，具体而言，两个问题都在 β 的相同取值处达到全局最小值。

2. **秩与核范数**。对于任意标量 β，考虑矩阵

$$\boldsymbol{X}_\beta\triangleq\begin{bmatrix}0.5+\beta & 1 & 1\\ 0.5 & 0.5 & \beta\\ 0.5 & 1-\beta & 0.5\end{bmatrix}.$$

证明，使 β 的秩最小的 $\boldsymbol{X}_\beta$ 的取值也会使其核范数（即其奇异值之和）最小。

3.**近端无约束凸优化的必要性/充分性**。设 $f:\ \mathbb{R}^p\to\ (-\infty,\ \infty]$ 是一个正常闭凸函数。证明$\boldsymbol{\alpha}^*$ 是 f 的极小化解，当且仅当 $\boldsymbol{\alpha}^*=\mathcal{P}$近端算子$f(\boldsymbol{\alpha}^*)$。提示：使用定理 *2.62* 和近端性质 I（引理*2.100*）。

4. **二次函数的强制性**。设 $f(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{y}^\top\boldsymbol{\beta} + z$，其中$\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 是对称的，$\boldsymbol{y} \in \mathbb{R}^p$，且 $z \in \mathbb{R}$。证明，当 $\boldsymbol{X}$ 为正定时，$f(\boldsymbol{\beta})$是强制的（即，当 $\|\boldsymbol{\beta}\|_2 \to \infty$时，$f(\boldsymbol{\beta}) \to \infty$）。

5. 利用定理2.47的证明，建立（2.17）、（2.18）和（2.19）之间的等价性，将它们作为强凸性和强光滑性的刻画。

6. 证明备注2.88中所述的布雷格曼距离的性质。

7. 设 $f \in C^{2,\ {L}^{2}}(\mathbb{R}^p)$，即 f 是二阶连续可微的，且其海森矩阵是 L-Lipschitz 连续的。证明，对于所有 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$，以下不等式成立：

$$\left\|\nabla f(\boldsymbol{\beta}) - \nabla f(\boldsymbol{\alpha}) - \nabla^2 f(\boldsymbol{\alpha})(\boldsymbol{\beta}-\boldsymbol{\alpha})\right\|_2 \le \frac{L}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2;$$

$$\left|f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) - \langle\nabla f(\boldsymbol{\alpha}), \boldsymbol{\beta}-\boldsymbol{\alpha}\rangle - \frac{1}{2}\left\langle\nabla^2 f(\boldsymbol{\alpha})(\boldsymbol{\beta}-\boldsymbol{\alpha}), \boldsymbol{\beta}-\boldsymbol{\alpha}\right\rangle\right| \le \frac{L}{6}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^3.$$

此外，若 $\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2 \le M$，证明：

$$\nabla^2 f(\boldsymbol{\alpha}) - LM\boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{\beta}) \preceq \nabla^2 f(\boldsymbol{\alpha}) + LM\boldsymbol{I}.$$

提示：参见定理 *2.80* 和习题 *2.76*。

8. **一般分离定理**。设 $\mathbb{S}$，$\mathbb{T} \subseteq \mathbb{R}^p$ 是**凸**集合，且交集为空 $\mathbb{S}\cap\mathbb{T} = \varnothing$。证明存在一个向量 $\boldsymbol{z} \in \mathbb{R}^p$ 和一个标量 $\gamma \in \mathbb{R}$，使得

(i) 对所有 $\boldsymbol{\alpha} \in \mathbb{S}$，都有 $\langle\boldsymbol{z},\ \boldsymbol{\alpha}\rangle \ge \gamma$。

(ii) 对所有 $\boldsymbol{\alpha} \in \mathbb{T}$，都有 $\langle\boldsymbol{z},\ \boldsymbol{\alpha}\rangle \le \gamma$。

此外，如果 $\mathbb{S}$ 和 $\mathbb{T}$ 都是闭集合，且其中至少有一个是有界的，则不等式可以加强为严格不等式：

(i) 对所有 $\boldsymbol{\alpha} \in \mathbb{S}$，都有 $\langle\boldsymbol{z},\ \boldsymbol{\alpha}\rangle > \gamma$。

(ii) 对所有 $\boldsymbol{\alpha} \in \mathbb{T}$，都有 $\langle\boldsymbol{z},\ \boldsymbol{\alpha}\rangle < \gamma$。

9. **支撑超平面定理**。设 $\mathbb{S} \subseteq \mathbb{R}^p$ 是一个**凸**集合（不一定是闭集合），且 $\boldsymbol{\beta} \notin \mathbb{S}$。证明存在 $\boldsymbol{0} \ne \boldsymbol{w} \in \mathbb{R}^p$，使得

$$\boldsymbol{w}^\top\boldsymbol{\alpha} \le \boldsymbol{w}^\top\boldsymbol{\beta}, \quad \text{for any } \boldsymbol{\alpha} \in \mathbb{S}.$$

10. **布雷格曼近端性质**。假设满足以下条件：

- ϕ： $\mathbb{R}^p \to (-\infty,\ \infty]$ 是适当的**闭凸**函数，在 dom($\partial\phi$) 上可微。f： $\mathbb{R}^p \to ($
- $-\infty,\ \infty]$ 是适当的**闭凸**函数，且 dom(f) $\subseteq$dom(ϕ)。$\phi + \delta_{\mathrm{dom}(f)}$ 是 σ-强凸
- 的（$\sigma > 0$）。

证明该优化问题

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\{f(\boldsymbol{\alpha}) + \phi(\boldsymbol{\alpha})\}$$

存在唯一极小化解，且该解属于 dom(f) $\cap$ dom($\partial\phi$)。

11. **布雷格曼近端性质**。假设满足以下条件：

- ϕ： $\mathbb{R}^p \to (-\infty,\ \infty]$ 是适当的**闭凸**函数，在 dom($\partial\phi$) 上可微；f： $\mathbb{R}^p \to ($
- $-\infty,\ \infty]$ 是适当的**闭凸**函数，且 dom(f) $\subseteq$dom(ϕ)；$\phi + \delta_{\mathrm{dom}(f)}$ 是 σ-强凸
- 的（$\sigma > 0$）。

^证明对于任意 $\boldsymbol{\beta}, \boldsymbol{\beta} \in$ dom($\partial\phi$)，以下两个陈述等价：

(i) $\boldsymbol{\beta} = \mathcal{B}\mathrm{rox}_{f,\phi}(\boldsymbol{\beta}) \triangleq \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \{f(\boldsymbol{\alpha}) + \mathcal{D}_\phi(\boldsymbol{\alpha}, \boldsymbol{\beta})\}$. (ii) $\langle \nabla\phi(\boldsymbol{\beta}) - \nabla\phi(\boldsymbol{\beta}), \boldsymbol{\xi} - \boldsymbol{\beta}\rangle \leq f(\boldsymbol{\xi}) - f(\boldsymbol{\beta})$ 对任意 $\boldsymbol{\xi} \in \mathrm{dom}(f)$ 均成立。

12. **Farkas引理**。 设 $\boldsymbol{c} \in \mathbb{R}^n$ 和 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$。证明下列两个系统中**恰有一个**有解：

(i) $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}$, $\boldsymbol{c}^\top \boldsymbol{x} > 0$. (ii) $\boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}$, $\boldsymbol{y} \geq \boldsymbol{0}$. （即，$\boldsymbol{c}$ 是 $\boldsymbol{A}$ 各行的锥组合。）

等价地，这意味着以下两个命题等价：

(a) 蕴含关系 $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0} \implies \boldsymbol{c}^\top \boldsymbol{x} \leq 0$ 成立。(b) 存在 $\boldsymbol{y} \in \mathbb{R}^m_+$ 使得 $\boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}$。

13. **Farkas引理′**。 证明 Farkas引理（习题2.12）还蕴含以下广义结果。给定适当维数的矩阵 $\boldsymbol{D}$, $\boldsymbol{E}$, 以及 $\boldsymbol{F}$，**以下两个系统中恰有一个**有解：

(1) $\boldsymbol{D}\boldsymbol{x} < \boldsymbol{0}$, $\boldsymbol{E}\boldsymbol{x} \leq \boldsymbol{0}$, $\boldsymbol{F}\boldsymbol{x} = \boldsymbol{0}$, (2) $\boldsymbol{D}^\top \boldsymbol{u} + \boldsymbol{E}^\top \boldsymbol{w} + \boldsymbol{F}^\top \boldsymbol{v} = \boldsymbol{0}$, $\boldsymbol{u} \geq \boldsymbol{0}$, $\boldsymbol{w} \geq \boldsymbol{0}$, $\boldsymbol{1}^\top \boldsymbol{u} = 1$.

14. 使用Farkas引理证明例2.109中所述线性规划的强对偶定理：如果原线性规划可行且最优值有限，并且对偶问题也可行，那么两个问题都能取得相同的最优目标函数值。

第3章

稀疏优化问题与最优性条件

目录

3.1 稀疏优化问题.1233.1.1 稀疏回归问题.124 3.1.2 稀疏恢复问题.127 3.1.3 某些问题的等价性.132 3.1.4 ℓ_1-分析模型.133

3.2 最小二乘问题的最优性条件 134
3.2.1 凸优化中的最小二乘.136 3.2.2 加权最小二乘问题.137 3.2.3 回归及其他形式.138

3.3 ℓ_1-正则化问题的最优性条件 140
3.3.1 LASSO 对偶.142 3.3.2 对偶多面体投影与筛选规则.144 3.3.3 安全特征消除（SAFE）.147

3.4 ℓ_1-最小化问题的最优性条件 150
3.4.1 充要条件.151 3.4.2 对偶证书与唯一性.152 3.4.3 ℓ_1-含噪测量下的最小化.156

3.5 ℓ_1-分析模型的最优性条件*.160
3.5.1 无噪声情形下的唯一恢复.161 3.5.2 含噪信号中的唯一恢复.164 3.5.3 噪声信号中的稳定恢复.165 **第3章习题.168**



I在稀疏表示和压缩感知（也称为压缩感知、压缩采样或稀疏采样）领域，一个核心挑战是从相对少量的线性测量中恢复高维信号——

测量值。这一任务在数学上可表述为求解一个欠定线性系统

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y},$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ (其中 $n \ll p$) 是测量矩阵或感知矩阵， $\boldsymbol{y} \in \mathbb{R}^n$ 是观测到的数据向量，而 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是未知的系数向量，假设它是稀疏的，这意味着其元素中只有一个很小的子集非零。稀疏性假设至关重要：若没有这一假设，由于系统欠定，它会有无穷多个解。然而，当已知（或认为） $\boldsymbol{\beta}$ 是稀疏的——或近似稀疏的——在 $\boldsymbol{X}$ 满足适当条件时（例如，受限等距性质或非相干性；参见第 4章和第 5章），便可以从 $\boldsymbol{\beta}$ 中 $\boldsymbol{y}$唯一且稳定地重建。

本章着重探讨如何运用优化理论中的工具来求解这类稀疏逆问题。具体而言，我们将研究两类密切相关但概念上有所区别的稀疏优化问题：

(i) 稀疏回归（或稀疏约束）优化。 在这种设置中，稀疏性作为约束或正则化项，被纳入更广泛的回归或估计框架。典型的形式包括：在非零系数的数量（“ℓ_0-范数”约束）上施加约束以最小化损失函数（例如最小二乘），或者更易于处理地添加 ℓ_1-范数惩罚项（如$LASSO$）。这些方法广泛应用于统计学、机器学习和信号处理，用于特征选择和提升模型可解释性。

(ii) 稀疏恢复问题。 这里的主要目标是从不完整测量值中精确或稳定地重构稀疏信号，且测量值往往含有噪声。这种视角常见于应用数学和工程领域，尤其是压缩感知。通常通过求解原组合优化问题的凸松弛方法（例如基追踪）或非凸近似来实现恢复；在关于感知矩阵 $\boldsymbol{X}$ 的假设下，还可提供理论保证。

尽管这两种范式都旨在利用稀疏性，但侧重点有所不同：在数据丰富的情境中，稀疏回归通常优先考虑预测性能和模型简洁性，而稀疏恢复则强调利用最少的测量值忠实地重建信号。接下来，我们将阐述这两种方法的数学基础和理论见解，并说明优化如何成为稀疏建模的统一框架。

3.1. 稀疏优化问题

一方面，压缩感知——这一术语由 David Donoho、Emmanuel Cand`es 和 Terence Tao 首创（Donoho 等人，2001年；Donoho 和 Elad，2003；Candes 和 Tao，2005年）——已成为稀疏信号恢复领域的一项开创性范式，从根本上改变了我们处理欠定线性系统的方式。它利用了许多自然信号在某些基或框架中具有稀疏表示这一事实，从而能够仅凭远少于奈奎斯特速率所需的测量值实现精确重建。这直接挑战了传统观点，即此类逆问题本质上是病态的。压缩感知不仅在理论上优雅，也产生了深远的实际影响，改变了信号处理、信息论和统计学等领域，同时还

催生了新的研究方向——包括低秩矩阵补全和相位恢复（Baraniuk 等人，2011年；Foucart 和 Rauhut，2013；Jain 等人，2017年）。由此，它成为一个典型范例，展现了统一的数学洞见如何在当今数据密集时代重新定义数据采集与分析。

压缩感知的一项关键优势在于，它能够将采样与压缩合并为一个高效步骤。压缩感知不必先采集大量原始数据，再对其进行压缩，而是在采集信号时利用其稀疏性，从而大幅减少所需的测量值数量。这不仅降低了数据存储和传输成本，还加快了采集过程；因此，在磁共振成像（MRI）和无线通信等对时间要求严格的应用中尤其有价值，因为它能在不影响保真度的情况下缩短扫描时间或减少带宽使用（Han 等人，2013年）。此外，压缩感知对噪声具有固有的鲁棒性，也能抵御观测缺失，因此即使数据高度不完整，也能实现高质量的信号重构，从而显著提升速度、效率和整体系统性能。

尽管 S. Chen、D. Donoho 和 M. Saunders（Chen 等人，2001年）提出了用于稀疏恢复的基追踪表述，将其作为一种有原则的方法，通过 ℓ_1-范数最小化来求解欠定线性系统中最稀疏的解，但由 R. Tibshirani（Tibshirani，1996年）更早提出的 *LASSO*（最小绝对收缩与选择算子），已成为统计学和机器学习领域的主流框架（Hastie 等人，2015年）。这种流行度上的差异，很大程度上源于问题表述方式和学科优先事项的不同。在植根于应用数学、信号处理和信息论的压缩感知文献中，重点是从不完整测量值中精确或稳定地恢复真实的稀疏信号，且通常对感知矩阵有严格的假设（例如受限等距性质或互相干性；见第 4 章和第 5章）。基追踪及其考虑噪声的变体（如基追踪去噪（*BPDN*））与这一目标天然契合，因为它们直接在显式的数据保真约束下最小化 ℓ_1-范数。

相比之下，LASSO 起源于回归建模，其主要目标是在高维场景中提高预测准确性、增强模型可解释性并进行变量选择。LASSO 并不对残差误差施加硬约束，而是将稀疏性视为一种权衡：它最小化残差平方和与系数的 ℓ_1 惩罚项之和。这种拉格朗日形式在实践中更加灵活——它无需预先指定噪声水平或容差参数，并且能够无缝融入用于调整正则化强度的交叉验证流程。因此，LASSO 深受寻求稳健、可扩展的特征选择和模型正则化工具的统计学家与数据科学家青睐。随着时间推移，两种方法之间的区别逐渐模糊，因为在适当选择参数的情况下，可以证明它们彼此等价。尽管如此，它们各自的历史渊源和方法侧重点仍然影响着不同领域的研究者对稀疏优化的理解与应用。

现在，我们来探讨这两个经典问题的精确数学表述，以及与之密切相关的变体。

3.1.1 稀疏回归问题

在本小节中，我们首先探讨在稀疏性约束下最小化（一般）连续可微目标函数这一难题。具体来说，我们考虑

以下问题：

$$\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_0 \leq R, \tag{S_0}$$

其中，$f: \mathbb{R}^p \to \mathbb{R}$ 是连续可微的，且其梯度 Lipschitz 连续，常数为 L_b（在我们的情形中，若取 $f(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$，即最小二乘目标函数，则 $L_b = \left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2$；见例 2.72）；$R > 0$ 是满足 $R < p$的整数，而 $\|\boldsymbol{\beta}\|_0$ 是所谓的“ℓ_0-范数”，用于统计 $\boldsymbol{\beta}$中的非零分量个数。对于任意正数 $s > 0$，向量 $\boldsymbol{\beta}$的 ℓ_s-范数 [1] 定义为 $\|\boldsymbol{\beta}\|_s = (\sum_i |\beta_i|^s)^{1/s}$. 因此，$p \times 1$ 向量 $\boldsymbol{\beta}$的 ℓ_0-范数可定义为

$$\|\boldsymbol{\beta}\|_0 = \lim_{s\to 0} \|\boldsymbol{\beta}\|_s^s = \lim_{s\to 0} \sum_{i=1}^p |\beta_i|^s = \sum_{i=1}^p \mathbb{1}(\beta_i \neq 0) = |\{i \mid \beta_i \neq 0, \forall\, i \in \{1, 2, \ldots, p\}\}|\,.$$

因此，若 $\|\boldsymbol{\beta}\|_0 \ll p$，则称向量 $\boldsymbol{\beta}$ 为稀疏向量。需要注意的是，ℓ_0-范数实际上并不是范数，因为它不满足齐次性：对于任意非零标量 λ，都有 $\|\lambda\boldsymbol{\beta}\|_0 = \|\boldsymbol{\beta}\|_0$ （定义 1.24）。

问题（S_0）是一个约束优化问题，我们要寻找最佳模型——即使 f最小化的模型——同时满足硬稀疏约束。通过将解限制为至多含有 R 个非零系数，我们可以明确控制模型复杂度，提升可解释性并减轻过拟合。

LASSO

通过上述稀疏优化分析，我们发现，尽管在非常具体的条件下，可以保证诸如 迭代硬阈值（IHT；参见第9.1节）的算法收敛，但使用 ℓ_0-范数（或硬阈值化）来统计向量中非零元素的数量，会带来若干重大挑战和缺点：

- 非凸性。 ℓ_0范数本质上是非凸的。因此，涉及该范数的优化问题通常都是 NP 难的。求得全局最小值在计算上难以承受，尤其是在问题维数增大时。组合性质。最小化 ℓ_0 范数需要评估变量的所有可能子集，以确定满足约束条件的最稀疏解。这种组合搜索具有指数复杂度，使其在高维数据集上不可行。不稳定性。直接通过 ℓ_0 最小化得到的解对数据中的微小扰动非常敏感。这种缺乏鲁棒性的情况在噪声环境下，或处理含有错误或离群值的真实数据时，尤其成问题。缺乏连续性和可微性。 ℓ_0范数既不连续也不可微，因此无法与标准的基于梯度的优化方法兼容。由于缺乏光滑性，针对可微目标函数设计的迭代算法难以应用。难以纳入优化算法。由于其离散且不连续的性质， ℓ_0范数难以直接纳入大多数优化框架。因此，实际方法往往依赖松弛方法或替代函数——最常见的是凸 ℓ_1范数——使问题变得可处理。

1。为避免符号冲突，本文在此使用术语 ℓ_s范数，尽管文献中通常使用ℓ_p范数来表示同一概念。

相比之下， ℓ_1范数同样能够促进稀疏性：它鼓励得到许多系数恰好为零的解。这种特性源于其几何结构——在高维空间中， ℓ_1球（一个多面体）的角点位于坐标轴上，因此有利于得到稀疏解；参见图 3.2 和3.3。重要的是， ℓ_1范数是凸的。因此， ℓ_1正则化问题的任何局部最小值也是全局最小值。与非凸的 ℓ_0 情形相比，这一性质大大简化了优化。

具体而言，考虑用矩阵—向量形式表示的标准线性回归模型：

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}, \tag{3.1}$$

其中， $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是设计矩阵（或模型矩阵）， $\boldsymbol{\epsilon} \in \mathbb{R}^n$ 是噪声变量向量，$\boldsymbol{\beta}^* \in \mathbb{R}^p$ 是未知系数向量。标准线性回归或最小二乘问题对应于以下无约束优化问题：

$$(\textbf{Least squares}): \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} f(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 . \tag{L_2}$$

通过利用 ℓ_1-范数进行凸松弛，本章后续各节将构建*LASSO*（最小绝对值选择与收缩算子）框架，涵盖其约束形式和拉格朗日形式。第9.2 节和第5.4节将分别讨论实用算法和理论保证：

$$(\textbf{Constrained LASSO}): \qquad \min_{\|\boldsymbol{\beta}\|_1\le\Sigma} f(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 ; \tag{L_C}$$

$$(\textbf{Lagrangian LASSO}): \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}) \triangleq \left\{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\}. \tag{L_L}$$

约束 LASSO 使用与（L_2）中相同的最小二乘损失，但将解限制在 ℓ_1球 $\|\boldsymbol{\beta}\|_1 < \Sigma$内。拉格朗日 LASSO 也称为 ℓ_1-正则化最小二乘问题，它通过加性惩罚项引入稀疏性。根据拉格朗日对偶性，这两种形式在以下意义下等价：对于每个$\Sigma > 0$，都存在一个 $\lambda \ge 0$ ，使得（L_C）和（L_L）的解一致，其中λ可解释为与约束 $\|\boldsymbol{\beta}\|_1 \le \Sigma$相关联的拉格朗日乘子；参见推论 3.3。

当 λ足够大（或 Σ足够小）时，任何解 $\widehat{\boldsymbol{\beta}}$ 对于（L_C）或（L_L）而言，解通常是稀疏的——也就是说，它只有少数几个非零项，因此能够识别出对预测 $\boldsymbol{X}$有用的相关特征（ $\boldsymbol{y}$的列）。随着 λ逐渐减小，解向量 $\boldsymbol{\beta}$ 的稀疏性也逐渐降低。当 λ趋近于 0 时，解向量 $\boldsymbol{\beta}$ 变为使 $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$ 达到最小值的向量。也就是说， $\lambda > 0$ 通过最小化误差平方和目标函数 $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$ 与 ℓ_1-范数目标函数 $\|\boldsymbol{\beta}\|_1$，在两个目标之间取得平衡。更多详情请参见第 3.2.3 节。

注意，LASSO 的一个缩放版本也很常用，尤其是在统计学习中，按样本量 n 进行归一化可以提高可解释性：

$$(\textbf{约束 LASSO-II}): \qquad \min_{\|\boldsymbol{\beta}\|_1\le\Sigma} f(\boldsymbol{\beta}) \triangleq \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 ; \tag{L'_C}$$

$$(\textbf{Lagrangian LASSO-II}): \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}) \triangleq \left\{\frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\}. \tag{L'_L}$$

LASSO 有许多推广形式，以下是几个例子。在机器学习中，稀疏核回归有时被称为广义*LASSO* (Roth,

2004年）。组 *LASSO* 将稀疏性扩展到预先定义的变量组（Puig 等人，2011年）。分布式 *LASSO* 用于处理去中心化或联网环境下的稀疏线性回归（ Mateos 等人，2010 年）。有关这些扩展的更全面讨论，请参见第9 章。

3.1.2 稀疏恢复问题

LASSO 在促进稀疏性和实现特征选择方面的成功并非孤立现象；它深深植根于一个更广泛的数学框架，这一框架称为稀疏恢复，起源于信号处理和压缩感知。如上所述，LASSO 虽然是在统计背景下发展起来的——其主要目标是在存在噪声的情况下进行预测、估计并提高可解释性——但同样的 ℓ_1正则化原则也出现在确定性情形中，此时目标是从不完整的线性测量值中精确重建稀疏信号。这自然引出了基追踪及其抗噪扩展形式基追踪去噪（ *BPDN*）等表述，它们将稀疏恢复构造成约束优化问题，而非带惩罚项的优化问题。尽管二者的出发点不同——统计推断与精确或稳定的信号重建——但 ℓ_1范数的底层几何性质，以及保证成功恢复的条件（例如限制等距性或零空间性质；见第4章和第5章），为这些方法提供了统一的基础。本小节将介绍这一互补视角。

有多种方法可以从少量线性测量值中恢复稀疏信号 $\boldsymbol{\beta}$ 。我们先考虑解决稀疏恢复问题的一种自然方法。已知测量值 $\boldsymbol{y} \in \mathbb{R}^n$ （例如通过 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$ 生成，其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$），并且知道原始信号 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是稀疏的，因此很自然地可以尝试通过求解如下形式的优化问题来恢复信号：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}), \tag{PS$_0$}$$

其中 $\mathcal{B}(\boldsymbol{y})$ 用于保证 $\boldsymbol{\beta}$ 与观测到的测量值 $\boldsymbol{y}$一致。回顾一下，$\|\boldsymbol{\beta}\|_0 = |\mathrm{supp}(\boldsymbol{\beta})|$ 只是统计 $\boldsymbol{\beta}$中非零项的数量；因此，(PS$_0$) 寻找与数据一致的最稀疏信号。常见的 $\mathcal{B}(\boldsymbol{y})$ 选择包括：

- 如果测量值**精确且无噪声**，那么我们可以设 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \boldsymbol{X}\boldsymbol{z} = \boldsymbol{y}\}$。如果测量值受到少量**有界噪声**的污染，则可以改用 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{y}\|_2 \le \epsilon\}$。
-

在这两种情况下，(PS$_0$) 都指与观测值一致的最稀疏 $\boldsymbol{\beta}$ 。下面将进一步介绍这两种情形。

问题(S$_0$)（第 125 页）与(PS$_0$)之间的主要区别如下：

（i）目标函数。在（S$_0$）中，目标是在满足稀疏性约束的条件下，最小化数据保真项（例如最小二乘误差）。相比之下，（PS$_0$）直接最小化稀疏性（即 $\|\boldsymbol{\beta}\|_0$），同时满足数据一致性。（ii）约束条件。问题（S$_0$）对稀疏度施加硬约束（例如 $\|\boldsymbol{\beta}\|_0 \le R$），从而允许模型拟合数据的程度具有灵活性。相反，（PS$_0$）强制要求严格的数据保真度（精确或近似），并以稀疏性作为目标。（iii）应用。公式（S$_0$）通常用于数据含有噪声的统计场景，此时需要在拟合效果与稀疏性之间进行权衡（例如回归）。相比之下，（PS$_0$）源于信号处理和压缩感知，其重点是根据有限的测量值，对真正的稀疏信号进行精确或稳定重构。

（iv）计算复杂度。尽管这两个问题都涉及非凸的 ℓ_0-范数，（S_0）有时可以在附加假设下采用迭代启发式方法（例如迭代硬阈值）求解。然而，（PS_0）通常是 NP 难的，这是由于最小化 $\|\boldsymbol{\beta}\|_0$具有组合性质，因为它需要评估 $\boldsymbol{\beta}$ （Mallat 和 Zhang，1993 年；Natarajan，1995 年；Davis 等人，1997 年）。

▶ **数据变换**。 请注意，在 (PS_0) 中，我们隐含地假设信号 $\boldsymbol{\beta}$本身是稀疏的。然而，在许多实际情形中，稀疏性并非存在于标准基中，而是存在于某个变换域中，即 $\boldsymbol{\beta} = \boldsymbol{\Phi}\boldsymbol{\alpha}$，其中 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 是稀疏的，且 $\boldsymbol{\Phi} \in \mathbb{R}^{p\times p}$ 是一个已知字典（例如小波基或傅里叶基）。此时，我们可以将恢复问题重新表述为

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{y}), \tag{PS_0'}$$

对于无噪声测量，其中 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \boldsymbol{X}\boldsymbol{\Phi}\boldsymbol{z} = \boldsymbol{y}\}$；存在有界噪声时，则有 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \|\boldsymbol{X}\boldsymbol{\Phi}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \epsilon\}$。通过定义 $\boldsymbol{X} = \boldsymbol{X}\boldsymbol{\Phi}$，我们可以看到，($PS_0'$) 在结构上与 ($PS_0$) 相同，只是用 $\boldsymbol{X}$ $\boldsymbol{X}$替代了。 此外，在许多应用中，引入 $\boldsymbol{\Phi}$ 并不会显著增加设计满足理想性质（例如限制等距性）的 $\boldsymbol{X}$ $\boldsymbol{X}$ 的难度。因此，在本书其余大部分内容中，我们将注意力限制在情形 $\boldsymbol{\Phi} = \boldsymbol{I}$，即假设稀疏性直接存在于典范基中。不过需要指出的是，当 $\boldsymbol{\Phi}$ 是一般的（可能过完备的）字典而非标准正交基时，这种简化会带来局限性。具体而言，

$$\left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right\|_2 = \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \boldsymbol{\Phi}\widehat{\boldsymbol{\alpha}}\|_2 \neq \|\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}\|_2 ,$$

因此，关于 $\|\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}\|_2$ 的上界并不能直接转化为关于 $\left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right\|_2$的上界，而后者往往是应用中关注的量。关于这些影响的进一步讨论，见 Candes 等人(2011年)。

▶ **凸松弛**。尽管可以在对 $\boldsymbol{X}$作出适当假设的条件下分析（PS_0）的性能，但我们不采用这一路径，因为目标函数 $\|\cdot\|_0$ 是非凸的，因此该优化问题通常难以求解。事实上，对于任意测量矩阵 $\boldsymbol{X}$，即使是近似求得最优解也是 NP 难的。回想一下，向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 的非零项个数可以看作其 ℓ_s-拟范数在 $s \to 0$时的极限：

$$\sum_{i=1}^{p} |\beta_i|^s \quad \xrightarrow{s\to 0} \quad \sum_{i=1}^{p} \mathbb{1}(\{\beta_i \neq 0\}) = \|\boldsymbol{\beta}\|_0 .$$

这一观察启发我们用以下一系列问题替代 ℓ_0-最小化问题（PS_0）：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_s \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}), \quad 0 < s < 1. \tag{PS_s}$$

然而，这种方法存在明显缺陷。对于 $s > 1$， ℓ_s-范数无法促进稀疏性——即使是 1-稀疏向量也可能无法恢复（见问题 3.1）。对于 $0 < s < 1$，尽管它能更好地促进稀疏性，但所得问题仍然是非凸的，并且依然是 NP 难问题（见问题3.19）。一种更实用的替代方法是将 $\|\cdot\|_0$ 替换为其

在单位 ℓ_∞-球上的凸包络（详见引理 3.10 ）：即 ℓ_1-范数。这将得到凸规划

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}). \tag{PS$_1$}$$

只要 $\mathcal{B}(\boldsymbol{y})$ 是凸的，(PS$_1$) 在计算上就是可行的。在无噪声情形下$\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \boldsymbol{X}\boldsymbol{z} = \boldsymbol{y}\}$甚至可以将其转化为线性规划（Chen 和 Donoho，1994年；Chen 等人，2001年；Foucart 和 Rauhut，2013年）。

显然，将(PS$_0$)替换为(PS$_1$)，就能把一个难以处理的组合问题转化为一个易于处理的凸问题。然而，(PS$_1$)的解是否与(PS$_0$)的解相似，则不那么显而易见。尽管如此，我们仍有充分的直觉和几何理由认为，ℓ_1-最小化能够促进稀疏性。例如，正如稍后例3.1 和图3.1所示，对于任意 ℓ_1，ℓ_1-最小化问题的解恰好与 ℓ_s-最小化的解一致，而且确实是稀疏的。

此外，历史先例也支持这一看法。早在 1965年，Logan(1965年) 就指出，在一个较小区间内受到破坏的带限信号，可以通过最小化误差的 ℓ_1范数而完美重构。这种方法——在 ℓ_1中寻找与观测值最接近的带限信号——可以视为通过 ℓ_1最小化实现稳健恢复的早期实例，进一步印证了 ℓ_1范数非常适合处理稀疏误差这一观点。

基追踪及其其他形式

总之，稀疏表示的核心问题是 ℓ_0-范数最小化

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}, \tag{P$_0$}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，$\boldsymbol{\beta} \in \mathbb{R}^p$，且 $\boldsymbol{y} \in \mathbb{R}^n$。由于该问题是 NP 难，我们转而考虑其凸松弛，即所谓的基追踪问题，也称为无噪声信号恢复、ℓ_1-范数最小化，或简称为 ℓ_1-最小化问题）：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}. \tag{P$_1$}$$

这是一个凸优化问题，因为目标函数 $\|\boldsymbol{\beta}\|_1$ 是凸的，且约束 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$ 是仿射的。因此，这个基追踪问题可以看作是（P$_0$）的凸松弛。问题（P$_1$）等价于

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{u}} \sum_{i=1}^{p} u_i \quad & \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}; \\ & -u_i \le \beta_i \le u_i, \quad \forall\, i \in \{1,2,\ldots,p\}. \end{aligned} \tag{P$_1'$}$$

这是一种线性规划形式（参见例 2.109）。虽然（P$_1$）施加了精确的等式约束，但也可以将其重新表述为一个惩罚问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 + \sigma \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2, \tag{P$_{1,\sigma}$}$$

其中 $\sigma > 0$ 是惩罚参数。这种形式称为基追踪去噪（*BPDN*）问题（Chen 等人，2001年）。当 $\sigma \to \infty$时，（P$_{1,\ \sigma}$）的解收敛到（P$_1$）的解，因此这两个问题渐近等价。显然，通过设定 ℓ_1，惩罚P$_{1\text{问题}\ (\sigma}$）等价于拉格朗日 LASSO 问题（L$_L$）（第126页）。$\sigma \triangleq \frac{1}{2\lambda}$。

—

▶ **信号恢复与稀疏回归**。如前所述，尽管（$\mathrm{P}_{1,\ \sigma}$）与（L_L）在对应关系 $\sigma \triangleq \frac{1}{2\lambda}$ 下在数学上等价，但二者的解释及典型应用场景有显著差异。在拉格朗日 LASSO 问题（L_L）中，一主要目标是在满足 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2$正则化项 ℓ_1 的约束下，最小化最小二乘损失 $\lambda\|\boldsymbol{\beta}\|_1$。其根本目标是建立预测模型：在拟合数据的同时，鼓励估计系数具有稀疏性，以提高可解释性并防止过拟合。当问题出现在超定系统中时（$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 其中 $n \geq p$，且当 $p \geq n$时，LASSO 也适用 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$），由于噪声或模型不匹配，方程通常没有精确解，因此正则化项 $\lambda\|\boldsymbol{\beta}\|_1$ 可用作控制模型复杂度的机制（Lu，2021a）。

相比之下，惩罚型 ℓ_1-最小化问题 ($\mathrm{P}_{1,\sigma}$) 源于稀疏信号恢复，尤其是压缩感知。在这里，目标不是预测，而是从不完整且可能含噪的测量值中，精确或稳定地重建真正的稀疏信号。该系统通常是欠定的 ($p \gg n$)，这意味着未知量远多于观测值。稀疏性并非便利条件，而是一种使恢复成为可能的结构性假设。

这种区别在通信系统中具有实际意义。当 $p \gg n$时，传输完整信号 $\boldsymbol{\beta} \in \mathbb{R}^p$ 效率低下。发送方可以改为仅传输低维测量向量 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} \in \mathbb{R}^n$；由于 $n \gg p$，这所需的带宽要少得多。如果接收方同时知道 $\boldsymbol{X}$ 和 $\boldsymbol{y}$，就可以通过求解诸如 ($\mathrm{P}_{1,\sigma}$) 这样的问题，重建原始稀疏信号 $\boldsymbol{\beta}$。

例 3.1（使用 ℓ_0、ℓ_1 和 ℓ_2-范数最小化进行稀疏信号恢复）。我们比较使用 ℓ_0-范数和 ℓ_1-范数最小化，从欠定测量值中恢复稀疏信号的效果。我们还将其与 ℓ_2-范数最小化进行比较：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_2 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}. \tag{P_2}$$

该过程从一个大小为 60×120 的随机矩阵 $\boldsymbol{X}$ 和一个长度为 120 的稀疏向量 $\boldsymbol{\beta}^*$ 开始；稀疏性通过将稠密随机向量中的小值置零来实现。随后计算测量值 $\boldsymbol{y}$，即 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$，以模拟压缩感知场景：对于一个 120 维的稀疏信号，只有 60 个线性测量值可用。

图3.1 中的第一个子图展示了通过 ℓ_0 最小化得到的解。第二个子图展示了 ℓ_2 最小化的解。由于 ℓ_2范数倾向于将能量分散并使结果更加光滑，恢复信号是稠密的，并将误差分散到所有坐标上，因此相对于 $\boldsymbol{\beta}^*$而言，重构误差较大。第三个子图展示了 ℓ_1 最小化的解。得益于促进稀疏性的几何结构，ℓ_1范数得到的解与原始稀疏信号的支撑集和大小十分接近。正如预期，重构误差显著低于 ℓ_2方法，说明 ℓ_1最小化在稀疏恢复中的有效性。□

约束稀疏恢复

另一方面，观测信号通常会受到噪声污染。因此，理想化优化问题中的等式约束应放宽，以允许测量误差有界。这就得到了以下带有不

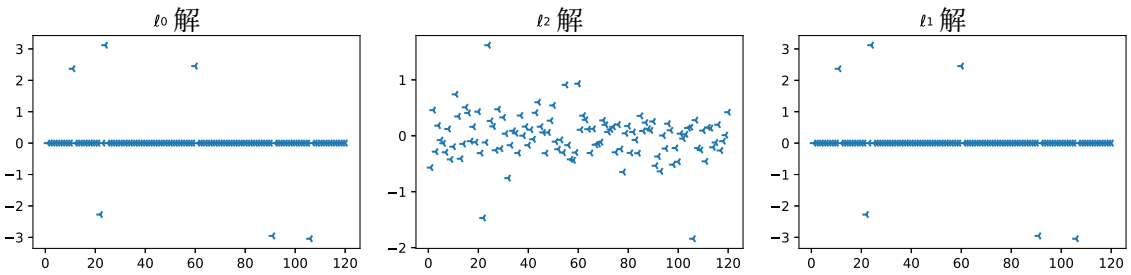


图 3.1： 稀疏信号恢复中 ℓ_0, ℓ_1 和 ℓ_2-范数最小化的示意图。

不等式约束：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2 \le \epsilon, \tag{$\mathrm{P}_{0,\epsilon}$}$$

其中， $\epsilon \ge 0$ 控制测量值 $\boldsymbol{y}$ 与模型预测 $\boldsymbol{X}\boldsymbol{\beta}$之间允许的偏差。

在实际情形中，也不可能以无限精度测量信号 $\boldsymbol{\beta} \in \mathbb{R}^p$ 。因此，观测到的测量向量 $\boldsymbol{y} \in \mathbb{R}^n$ 只是对真实线性测量 $\boldsymbol{X}\boldsymbol{\beta}$的近似，满足

$$\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\| \le \epsilon$$

对于某个 $\epsilon \ge 0$ 以及定义在 $\mathbb{R}^n$ 上的某个范数 $\|\cdot\|$ （最常见的是 ℓ_2-范数）。在这种情况下，良好的重建方法应给出一个估计值 $\boldsymbol{\beta} \in \mathbb{R}^p$ ，使其与真实信号 $\boldsymbol{\beta}$ 之间的距离受测量误差 ϵ控制。这种理想性质称为对测量噪声的鲁棒性 。

因此，无噪声基追踪问题（P_1）被以下凸优化问题所取代：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2 \le \epsilon, \tag{$\mathrm{P}_{1,\epsilon}$}$$

这通常称为二次约束基追踪 （或二次约束 ℓ_1-最小化、 ℓ_1-含噪测量最小化）问题，也称为含噪信号恢复问题。

另一种重要的变体是*Dantzig*选择器（*DS*）（Candes 和 Tao，2007年），其定义如下：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})\right\|_\infty \le D. \tag{$\mathrm{P}_{1,D}$}$$

这又是一个凸优化问题。该约束背后的直觉是，残差 $\boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}$ 与测量矩阵 $\boldsymbol{X}$的所有列 $\boldsymbol{x}_i$ 之间的相关性都应很小。事实上，

$$\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})\right\|_\infty = \max_{i\in\{1,2,\ldots,p\}} |\langle \boldsymbol{\epsilon}, \boldsymbol{x}_i\rangle| .$$

本章稍后针对问题 (P_1) 和 ($\mathrm{P}_{1,\epsilon}$) 建立的理论框架，也适用于 Dantzig选择器。不过，我们这里不深入探讨这些细节。更多见解请参阅问题3.17。

3.1.3 某些问题的等价性

基于二次约束基追踪问题 ($P_{1,\epsilon}$) 的解 $\boldsymbol{\beta}$ 与 (L_L) 和 (L_C)（第126页）中的拉格朗日函数问题和约束 LASSO 问题的解密切相关。为了建立这种联系，我们先证明一个更一般的结果。

定理 3.2：（约束问题与正则化问题的等价性） 设 $\|\cdot\|_a$，$\|\cdot\|_b$ 是 $\mathbb{R}^p$上的两个范数。给定 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，$\boldsymbol{y} \in \mathbb{R}^n$ 和 $\epsilon > 0$，考虑优化问题

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_a \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_b \le \epsilon. \tag{3.4}$$

假设（3.4）严格可行；也就是说，存在某个向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 使得 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_b < \epsilon$。设 $\boldsymbol{\beta}$ 是（3.4）的极小化解。那么存在一个参数 $\sigma \ge 0$，使得 $\boldsymbol{\beta}$ 也是正则化问题的极小化解

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_a + \sigma \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_b . \tag{3.5}$$

反之，对于任意 $\sigma > 0$，若 $\boldsymbol{\beta}$ 是（3.5）的极小化解，则存在 $\epsilon \ge 0$ 使得 $\boldsymbol{\beta}$ 是（3.4）的极小化解。 [a]

[a] 虽然我们只考虑实值向量，但通过识别 $\mathbb{C}^p = \mathbb{R}^{2p}$，该定理在复数情形下也成立。

证明 [定理 3.2] 问题 (3.4) 的拉格朗日函数为

$$L(\boldsymbol{\beta}, \lambda) = \|\boldsymbol{\beta}\|_a + \lambda(\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_b - \epsilon).$$

根据 Slater 条件（定理2.114），在所述严格 3.4可行性假设下，问题 (3.4) 满足强对偶性。因此，存在一个最优对偶乘子 $\lambda \ge 0$。鞍点性质 (2.62) 表明

$$L(\widehat{\boldsymbol{\beta}}, \widehat{\lambda}) \le L(\boldsymbol{\beta}, \widehat{\lambda})$$

$\boldsymbol{\beta} \mapsto L\ \boldsymbol{\beta}\ \lambda$ 对所有 $\boldsymbol{\beta} \in \mathbb{R}^p$ 均成立。因此，$\boldsymbol{\beta}$ 是函数 $-\lambda\epsilon$ (.) 的极小化解。由于常数项不会影响极小化解的位置，$\boldsymbol{\beta}$ 也是以下函数的极小化解

$$\|\boldsymbol{\beta}\|_a + \widehat{\lambda}\,\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_b ,$$

由此可得出第一个结论，其中 $\sigma = \lambda$。

反之，设 $\widehat{\boldsymbol{\beta}}$ 是 (3.5) 的极小化解，并令 $\lambda = \sigma$。取 $\epsilon = \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y}\right\|_b$。则拉格朗日对偶函数 D 满足

$$D(\lambda) = L(\widehat{\boldsymbol{\beta}}, \lambda) = \left\|\widehat{\boldsymbol{\beta}}\right\|_a .$$

根据弱对偶性，有 $D(\lambda) \le \|\boldsymbol{\beta}\|_a$ 对所有可行的 $\boldsymbol{\beta}$。由于 $\boldsymbol{\beta}$ 是可行的，并且根据 $\epsilon = \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y}\right\|_b$的选取，它达到此下界，因此必定是（3.4）的极小化解。 ■

这种一般等价性立即阐明了以下问题之间的关系式（$P_{1,\ \epsilon}$）、（$P_{1,\ \sigma}$）、（L_C）和（L_L）。

推论 3.3：（（P1，ϵ）、（P1，σ）、（L_C）和（L_L）的等价性）考虑问题（P1，ϵ）、（P1，σ）、（L_C）和（L_L）：

(i) 如上所述，通过设置 $\sigma \triangleq \frac{1}{2\lambda}$，惩罚型 ℓ_1 问题 (P1,σ) 与拉格朗日 LASSO 问题 (L_L) 等价。(ii) 如果 $\widehat{\boldsymbol{\beta}}$ 是拉格朗日 LASSO 问题 (L_L) 的极小化解，那么存在 $\epsilon = \epsilon_{\widehat{\boldsymbol{\beta}}}$，使得 $\widehat{\boldsymbol{\beta}}$ 是二次约束基追踪问题 (P1,ϵ) 的极小化解。(iii) 如果 $\widehat{\boldsymbol{\beta}}$ 是约束 LASSO 问题 (L_C) 的极小化解，那么存在 $\lambda = \lambda_{\widehat{\boldsymbol{\beta}}}$，使得 $\widehat{\boldsymbol{\beta}}$ 是拉格朗日 LASSO 问题 (L_L) 的极小化解。(iv) 如果 $\widehat{\boldsymbol{\beta}}$ 是二次约束基追踪 (P1,ϵ) 的极小化解，那么存在 $\sim\Sigma = \Sigma_{\widehat{\boldsymbol{\beta}}}$，使得 $\widehat{\boldsymbol{\beta}}$ 是约束 LASSO 问题 (L_C) 的极小化解。

需要注意的是，这些等价关系是隐含的：参数之间的映射（例如，λ 与 ϵ 之间，或 λ 与 Σ之间）取决于具体的极小化解 $\boldsymbol{\beta}$。因此，无法先验地进行变换——这需要知道解本身。正因如此，尽管理论上这些等价关系颇具启发性，但在实际应用中选择正则化参数时，它们的实用性有限。

3.1.4 ℓ_1-分析模型

注意，模型（P_1）、（$\mathrm{P}_{1,\sigma}$）和（$\mathrm{P}_{1,\epsilon}$）统称为 ℓ_1-合成模型。我们还考虑以下 ℓ_1-最小化形式：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1, \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}, \tag{Q_1}$$

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1 + \sigma\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2^2, \tag{$\mathrm{Q}_{1,\sigma}$}$$

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1, \quad \text{s.t.} \quad \left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2 \le \epsilon, \tag{$\mathrm{Q}_{1,\epsilon}$}$$

其中 ϵ，σ 为正参数。模型（Q_1）适用于无噪声情形，即测量误差满足 $\boldsymbol{\epsilon} = \mathbf{0}$。若 $\boldsymbol{\Psi} = \boldsymbol{I}$ 为单位矩阵，这些模型就完全化为 ℓ_1- 合成模型。若 $\boldsymbol{\Psi} \neq \boldsymbol{I}$，则称其为 ℓ_1- 分析模型 (Elad 等人，2007年；Candes 等人，2011年；Liu 等人，2012；Nam 等人，2013年；Zhang 等人，2016）。

需要将这些模型与（PS_0'）（第 128 页）中描述的数据变换区分开来：在分析框架中，变换作用于信号本身，而不是测量值。

在 ℓ_1-综合方法中，感兴趣信号表示为 $\boldsymbol{\beta}^* = \boldsymbol{D}\boldsymbol{b}$，其中 $\boldsymbol{D}$是某个字典，而 $\boldsymbol{b}$是一个稀疏系数向量。包括余稀疏分析模型（Nam 等人，2013年）和全变差模型（Rudin 等人，1992年）在内的 ℓ_1-分析模型有许多广为人知的实例，并受到广泛关注。人们期望底层信号与（过完备）字典 $\boldsymbol{\Psi} \in \mathbb{R}^{p\times q}$ 的各列（原子）之间具有 $q \ge p$稀疏相关性，即 $\boldsymbol{\Psi}^\top\boldsymbol{\beta}^*$ 是稀疏的。

通常，我们要求分析算子 $\boldsymbol{\Psi}$ 具有满行秩，使其成为有效的范数，适用于（问题 $\|\boldsymbol{\Psi}^\top\cdot\|_1$。）。当具有满行秩时，我们可以结合 $\|\cdot\|_a \triangleq \|\cdot\|_{\boldsymbol{\Psi}^\top}$ 应用定理 3.2，证明问题（Q_1，σ）与（Q_1，ϵ）等价。 $\boldsymbol{\beta}$ 3.3 $\|\cdot\|_{\boldsymbol{\Psi}^\top} \triangleq$

推论 3.4： （$(Q1,\ \sigma)$ **与** $(Q1,\ \epsilon)$**的等价性**）假设 $\boldsymbol{\Psi}$ 具有满行秩，并考虑问题（$Q1,\ \sigma$）和（$Q1,\ \epsilon$）：

(i) 若 $\widehat{\boldsymbol{\beta}}$ 是问题（$Q1,\sigma$）的极小化解，则存在 $\epsilon = \epsilon_{\widehat{\boldsymbol{\beta}}}$ 使得 $\widehat{\boldsymbol{\beta}}$ 是问题（$Q1,\epsilon$）的极小化解。

(ii) 若 $\widehat{\boldsymbol{\beta}}$ 是问题（$Q1,\epsilon$）的极小化解，则存在 $\sigma = \sigma_{\widehat{\boldsymbol{\beta}}}$ 使得 $\widehat{\boldsymbol{\beta}}$ 是问题（$Q1,\sigma$）的极小化解。

3.2. 最小二乘问题的最优性条件

为了引入稀疏优化问题的最优性条件，我们从基本的最小二乘问题开始：$\min_{\boldsymbol{\beta}} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$。对于这个问题，我们考虑以下问题：

- Q1：什么是最小二乘解？Q2：何时可以断定最小二乘解唯一？Q3：何时可以断定最
- 小范数最小二乘解唯一？
-

为回答这些问题，我们依据以下刻画最小二乘问题最优性条件的基本结果：

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}. \tag{3.7}$$

该条件成立，是因为目标函数 $f(\boldsymbol{\beta}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 是凸且可微的。因此，任何极小化解 $\boldsymbol{\beta}$ 都必须满足一阶最优性条件 $\nabla f(\widehat{\boldsymbol{\beta}}) = 2\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$ (推论 2.61)。

现在，我们通过以下定理对最小二乘问题给出统一的观点。随着讨论的深入，其基本思想将变得更加清晰。该定理回答了问题 Q1。

定理 3.5：（最小二乘问题的统一观点（Lu，2021b）） 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 且 $\boldsymbol{y} \in \mathbb{R}^n$。则最小二乘问题 $f(\boldsymbol{\beta}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 存在一个极小化解 $\boldsymbol{\beta} \in \mathbb{R}^n$ 当且仅当存在一个向量 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 使得

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^+\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}, \tag{3.8}$$

其中 $\boldsymbol{X}^+$ 表示 $\boldsymbol{X}$ 的伪逆（见1.2.5节）：

- 这表明，只有当**唯一的**最小二乘问题极小化解为 $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^+\boldsymbol{y}$时，$\boldsymbol{X}^+$ 才是 $\boldsymbol{X}$ 的左逆（即 $\boldsymbol{X}^+\boldsymbol{X} = \boldsymbol{I}$，并且 $\boldsymbol{X}$ 仅当 $\boldsymbol{X}$ 具有满列秩时才可左逆（Lu，2021a）。最优值为 $f(\widehat{\boldsymbol{\beta}}) = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^+)\boldsymbol{y}$。若 $\boldsymbol{\alpha} \neq \mathbf{0}$：$\|\boldsymbol{X}^+\boldsymbol{y}\|_2 \leq \|\boldsymbol{X}^+\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}\|_2$。

这意味着，任何使 $\boldsymbol{\beta}$ 达到最小值的向量 $f(\boldsymbol{\beta})$ 都必须具有（3.8）中给出的形式，其中：

- $\boldsymbol{X}^+\boldsymbol{y} \perp \mathcal{N}(\boldsymbol{X})$（由定理 1.56）是特解（最小范数解）。$(\boldsymbol{I} - \boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha} \in \mathcal{N}(\boldsymbol{X})$ 是齐次解，它体现了 $\boldsymbol{\beta}$ 中由 $\boldsymbol{X}$ 的零空间所产生的自由度。
-

定理 [的证明3。5]由（3.7），任意极小化解 $\boldsymbol{\beta}$ 都必须满足正规方程：$\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$。该方程意味着，向量 $\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}$ 与 $\boldsymbol{X}$ 的列空间正交，即误差 $\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}$ 位于 $\boldsymbol{X}^\top$ 的零空间中：

$$\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \perp \mathcal{C}(\boldsymbol{X}) \qquad \text{and} \qquad \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}^\top). \tag{3.9}$$

为求解 $\boldsymbol{\beta}$，注意到当 $\boldsymbol{y} \notin \mathcal{C}(\boldsymbol{X})$ 时，方程组 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 可能没有精确解。在这种情况下，我们寻求使 $\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2$ 最小的最小范数解。伪逆 $\boldsymbol{X}^+$ 给出了可能的最佳解，即唯一的最小范数最小二乘解（参见下文的论证）：

$$\boldsymbol{\beta}_{\text{particular}} = \boldsymbol{X}^+\boldsymbol{y}. \tag{3.10}$$

为说明 $\boldsymbol{\beta}_{\text{particular}}$ 满足（3.7），我们有 $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_{\text{particular}} = \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{y} = \boldsymbol{X}^\top\boldsymbol{y}$，这里用到了 $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^+ = \boldsymbol{X}^\top$（引理 1.55）。

然而，这只是一个解，可能还存在许多其他解。为了刻画完整的解集，我们考虑两种情况。

情况 1： $\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})$。在这种情况下，（3.8）中的解显然成立，因为矩阵的伪逆具有（$\boldsymbol{X}$（$\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X}$）$\boldsymbol{\alpha}=\mathbf{0}$）这些性质，且矩阵的伪逆具有唯一性（命题 1.54）。

情形 2： $\boldsymbol{y} \notin \mathcal{C}(\boldsymbol{X})$。定义残差 $\boldsymbol{e} \triangleq \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{\text{particular}} = \boldsymbol{y}-\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{y}$。我们要找出所有满足 $\left\|\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right\|_2^2 = \|\boldsymbol{e}\|_2^2$ 的向量 $\widetilde{\boldsymbol{\beta}}$。由此有两种代数可能性：

$$\boldsymbol{X}\widetilde{\boldsymbol{\beta}}-\boldsymbol{y} = \boldsymbol{e} \qquad \text{and} \qquad \boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}} = \boldsymbol{e}.$$

前一种情况意味着 $\boldsymbol{e} = \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{\text{particular}} = \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}$，由此得到 $\boldsymbol{y} = \boldsymbol{X}\frac{\boldsymbol{\beta}+\boldsymbol{\beta}_{\text{particular}}}{2}$，即 $\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})$，这与情形 2 的假设矛盾。因此，只有第二种情况成立。

由于 $\boldsymbol{\beta}$ 属于 $\mathbb{R}^p$，极小化点的完整集合由特解$\boldsymbol{\beta}_{\text{particular}} = \boldsymbol{X}^+\boldsymbol{y}$ 加上 $\boldsymbol{X}$ 的零空间中的任意向量组成（即任何满足$\boldsymbol{X}\boldsymbol{\beta}_{\text{null}} = \mathbf{0}$、$\boldsymbol{\beta}_{\text{null}} \in \mathcal{N}(\boldsymbol{X}) \triangleq \{\boldsymbol{z} \mid \boldsymbol{X}\boldsymbol{z} = \mathbf{0}\}$的 $\boldsymbol{\beta}_{\text{null}}$）。因此，$\boldsymbol{\beta}$ 的通解必定具有以下形式：

$$\boldsymbol{\beta} = \boldsymbol{X}^+\boldsymbol{y} + \boldsymbol{\beta}_{\text{null}}, \qquad \text{with} \quad \boldsymbol{\beta}_{\text{null}} \in \mathcal{N}(\boldsymbol{X}). \tag{3.11}$$

由于 $\boldsymbol{X}-\boldsymbol{X}\boldsymbol{X}^+\boldsymbol{X} = \mathbf{0}$ 由（1.34），到 $\boldsymbol{X}$ 的零空间的投影由以下式子给出 y:

$$\boldsymbol{\beta}_{\text{null}} = (\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}, \quad \text{for some } \boldsymbol{\alpha} \in \mathbb{R}^p. \tag{3.12}$$

这是因为（$\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X}$）是到 $\mathcal{N}$（$\boldsymbol{X}$）上的正交投影矩阵（参见第1.2.3 节和第 1.2.5节）。

另一方面，由于 $\boldsymbol{\beta}_{\text{null}} \in \mathcal{N}$（$\boldsymbol{X}$）和 $\mathcal{C}$（$\boldsymbol{X}^+$）$\equiv \mathcal{C}$（$\boldsymbol{X}^\top$）（见定理 1.56），可得

$$\left\|\boldsymbol{X}^+\boldsymbol{y} + (\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}\right\|_2^2 = \left\|\boldsymbol{X}^+\boldsymbol{y}\right\|_2^2 + \left\|(\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}\right\|_2^2.$$

这表明 帽子 $\|\boldsymbol{X}^+\boldsymbol{y}\|_2 \le \|\boldsymbol{X}^+\boldsymbol{y}+(\boldsymbol{I}-\boldsymbol{X}^+\boldsymbol{X})\boldsymbol{\alpha}\|_2$ 若 $\boldsymbol{\alpha} \neq \mathbf{0}$，并完成证明。 ■

3.2.1 凸优化中的最小二乘

关于凸函数的定理，为本节开头提出的问题 Q2 和 Q3 提供了严格的解答。定理 3.5 在大样本情形下，部分回答了问题 Q2——关于最小二乘解的唯一性。具体而言，若 $n > p =$ 秩$(\boldsymbol{X})$，则最小二乘解 $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$ 是唯一的。在这种情况下，矩阵

$$\boldsymbol{H} \triangleq \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

称为帽子矩阵，它将 $\boldsymbol{y}$ 正交投影到 $\boldsymbol{X}$的列空间上（参见第 1.2.3 节）。事实上，这是**唯一**能保证最小二乘解唯一的情形。

注意，Q2 和 Q3 都可以表述为凸优化问题:

$$\begin{aligned} &\text{(P2)}: \quad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} f(\boldsymbol{\beta}) = \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2 = \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{y}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{y}^\top\boldsymbol{y}; \\ &\text{(P3)}: \quad \min_{\boldsymbol{\beta}\in\mathbb{Q}_2} g(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2^2, \quad \mathbb{Q}_2 \triangleq \{\boldsymbol{\beta}\in\mathbb{R}^p \mid \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 = \min\}. \end{aligned} \tag{3.13}$$

这里，$f(\boldsymbol{\beta})$ 和 $g(\boldsymbol{\beta})$ 都是凸函数（习题 2.35），集合 $\mathbb{Q}_2$ 是凸集（参见定理 2.44）。因此，(P2) 和 (P3) 都是凸优化问题。定理 2.44还指出，如果函数是严格凸的，那么极小化解是唯一的。显然，$g(\boldsymbol{\beta})$ 是严格凸的（这也再次证实了 Q3）。而且，$f(\boldsymbol{\beta})$ 仅当 $\boldsymbol{X}^\top\boldsymbol{X}$ 是正定矩阵时才是严格凸的，也就是说，当 $\boldsymbol{X}$ 满列秩时（即 $n > p =$的秩为 $\boldsymbol{X}$）时。这为问题 Q2 提供了严格的答案。

结合这一分析与定理3.5，我们在以下推论中确立了最小范数最小二乘解的唯一性。

推论 3.6：（最小范数且唯一的最小二乘解）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和$\boldsymbol{y} \in \mathbb{R}^n$。

- 设 $\mathbb{Q}_2 \subset \mathbb{R}^p$ 表示 $\boldsymbol{\beta} \mapsto \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2$ 的所有极小化点构成的集合。优化问题

$$\min_{\boldsymbol{\beta}\in\mathbb{Q}_2} \|\boldsymbol{\beta}\|_2^2 \tag{3.14}$$

  有唯一解 $\boldsymbol{\beta}^+ = \boldsymbol{X}^+\boldsymbol{y}$。
- 若 $n \geq p$ 和 $\boldsymbol{X}$ 满列秩，则最小二乘问题

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 \tag{3.15}$$

  有唯一解 $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^+\boldsymbol{y}$。

注意，(3.15) 的极小化解对应于 $\boldsymbol{y}$ 在 $\boldsymbol{X}$的列空间上的正交投影。因此，矩阵 $\boldsymbol{X}\boldsymbol{X}^+$ 表示到 $\boldsymbol{X}$上的正交投影算子。当 $\boldsymbol{X}$ 满秩且 $n \geq p$时，矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$ 可逆。根据(1.35)，可得 $\boldsymbol{X}^+ =(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$。因此，$\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^+\boldsymbol{y}$ 满足正规方程

$$\boldsymbol{X}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{y}. \tag{3.16}$$

推论 3.7：（高维最小二乘解，ℓ_2-最小化）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 且 $p \geq n$，假设 $\boldsymbol{X}$ 满秩，并设 $\boldsymbol{y} \in \mathbb{R}^n$。那么约束优化

问题

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_2 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y} \tag{3.17}$$

有唯一解 $\widehat{\boldsymbol{\beta}}=\boldsymbol{X}^+\boldsymbol{y}$。

在此情形下，由（1.36），可得 $\boldsymbol{X}^+=\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}$ 若 $\boldsymbol{X}$ 满秩且 $p\geq n$；因此，方程组 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}$ 相容。因此，在上述推论的情形下，$\widehat{}$极小化解 $\boldsymbol{\beta}$ （3.17）满足第二类正规方程

$$\widehat{\boldsymbol{\beta}}=\boldsymbol{X}^\top\boldsymbol{\gamma}, \quad \text{where } \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\gamma}=\boldsymbol{y}. \tag{3.18}$$

最后，尽管正规方程（3.16）和（3.18）表明，最小二乘问题可以通过标准线性方程组求解器（例如高斯消元法）求解，但在数值计算中，使用专为最小二乘设计的专用算法更具优势。这些算法包括基于Cholesky分解、QR分解或奇异值分解（SVD）的方法，它们具有更高的稳定性和精度（Lu，2021b；Bj¨orck，2024）。

3.2.2 加权最小二乘问题

推论 3.7 给出了 ℓ_2-最小化问题的唯一解。我们还可以考虑加权 ℓ_2-最小化问题：

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \|\boldsymbol{\alpha}\|_{2,\boldsymbol{w}} \triangleq \left(\sum_{i=1}^p \alpha_i^2 w_i\right)^{1/2} \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{y}, \tag{$\mathrm{L}_{2,\boldsymbol{w}}$}$$

其中 $\boldsymbol{w}$ 是一个正权重向量（对所有 i，$w_i>0$ ）。引入对角权重矩阵 $\boldsymbol{D}_{\boldsymbol{w}}=\mathrm{diag}([w_1,w_2,\ldots,w_p])\in\mathbb{R}^{p\times p}$，并定义变量替换

$$\boldsymbol{\beta}\triangleq\boldsymbol{D}_{\boldsymbol{w}}^{1/2}\boldsymbol{\alpha}.$$

则 $\|\boldsymbol{\alpha}\|_{2,\boldsymbol{w}}^2=\|\boldsymbol{\beta}\|_2^2$，约束变为 $\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{w}}^{-1/2}\boldsymbol{\beta}=\boldsymbol{y}$。因此，$(\mathrm{L}_{2,\boldsymbol{w}})$ 的极小化解 $\widehat{\boldsymbol{\alpha}}$对应于标准最小范数问题的极小化解 $\boldsymbol{\beta}$：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_2 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{D}_{\boldsymbol{w}}^{-1/2}\boldsymbol{\beta}=\boldsymbol{y},$$

根据推论 3.6，解为

$$\widehat{\boldsymbol{\alpha}}=\boldsymbol{D}_{\boldsymbol{w}}^{-1/2}\widehat{\boldsymbol{\beta}}=\boldsymbol{D}_{\boldsymbol{w}}^{-1/2}(\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{w}}^{-1/2})^+\boldsymbol{y}. \tag{3.19}$$

特别地，如果 $p\geq n$ 且 $\boldsymbol{X}$ 满秩，那么推论 3.7 意味着伪逆可化简为

$$\widehat{\boldsymbol{\alpha}}=\boldsymbol{D}_{\boldsymbol{w}}^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{w}}^{-1}\boldsymbol{X}^\top)^{-1}\boldsymbol{y}. \tag{3.20}$$

定理 3.5 表明，标准最小二乘解 $\widehat{\boldsymbol{\beta}}=\boldsymbol{X}^+\boldsymbol{y}$ 与$\mathcal{N}(\boldsymbol{X})$正交。加权最小二乘情形也可以观察到类似的结果。

定理 3.8：（LS-III的正交关系）向量 $\widehat{\boldsymbol{\alpha}}\in\mathbb{R}^p$ 是（$\mathrm{L}_{2,\boldsymbol{w}}$）的极小化解，当且仅当

$$\langle\widehat{\boldsymbol{\alpha}},\boldsymbol{n}\rangle_{\boldsymbol{w}}=0, \quad \text{for all } \boldsymbol{n}\in\mathcal{N}(\boldsymbol{X}), \tag{3.21}$$

其中 $\langle\widehat{\boldsymbol{\alpha}},\ \boldsymbol{n}\rangle_{\boldsymbol{w}}\triangleq\sum_{i=1}^p\beta_i n_i w_i$。

证明 [定理 3。8] 令 $\widehat{\boldsymbol{\alpha}}$ 满足 $\boldsymbol{X}\widehat{\boldsymbol{\alpha}}=\boldsymbol{y}$。对于（$L_{2,\boldsymbol{w}}$），任意可行向量 $\boldsymbol{\alpha}\in\mathbb{R}^p$ 均可写为 $\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}+\boldsymbol{n}$，其中 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$。对于任意标量 $\eta\in\mathbb{R}$ 以及 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$，考虑扰动向量 $\widehat{\boldsymbol{\alpha}}+\eta\boldsymbol{n}$。其加权范数的平方为

$$\|\widehat{\boldsymbol{\alpha}}+\eta\boldsymbol{n}\|_{2,\boldsymbol{w}}^2=\|\widehat{\boldsymbol{\alpha}}\|_{2,\boldsymbol{w}}^2+\eta^2\|\boldsymbol{n}\|_{2,\boldsymbol{w}}^2+2\eta\langle\widehat{\boldsymbol{\alpha}},\boldsymbol{n}\rangle_{\boldsymbol{w}}. \tag{3.22}$$

因此，如果 $\langle\widehat{\boldsymbol{\alpha}},\boldsymbol{n}\rangle_{\boldsymbol{w}}=0$ 对所有 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$ 都成立，那么 $\eta=0$ 就是 $\eta\mapsto\|\widehat{\boldsymbol{\alpha}}+\eta\boldsymbol{n}\|_{2,\boldsymbol{w}}$ 的极小化解。因此，(3.21) 意味着 $\widehat{\boldsymbol{\alpha}}$ 是 ($L_{2,\boldsymbol{w}}$) 的极小化解。

反之，假设 $\widehat{\boldsymbol{\alpha}}$ 是 ($L_{2,\boldsymbol{w}}$) 的极小化解。那么，对于所有 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$，$\eta=0$ 都是 $\eta\mapsto\|\widehat{\boldsymbol{\alpha}}+\eta\boldsymbol{n}\|_{2,\boldsymbol{w}}$ 的极小化解。然而，如果 $\langle\widehat{\boldsymbol{\alpha}},\boldsymbol{n}\rangle_{\boldsymbol{w}}$ 不为零，那么根据 (3.22)，我们可以找到一个充分接近于 0 且与 $\langle\widehat{\boldsymbol{\alpha}},\boldsymbol{n}\rangle_{\boldsymbol{w}}$ 符号相反的非零 η，使得 $\|\widehat{\boldsymbol{\alpha}}+\eta\boldsymbol{n}\|_{2,\boldsymbol{w}}<\|\widehat{\boldsymbol{\alpha}}\|_{2,\boldsymbol{w}}$，这将导致与 $\widehat{\boldsymbol{\alpha}}$ 是极小化解相矛盾。■

3.2.3 回归及其他形式

由于最小二乘模型通常用于回归问题，因此一般假设样本数量大于特征数量（$n>p$）。在这种情况下，若设计矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 满列秩，则最小二乘解为

$$\boldsymbol{\beta}^{OLS}\triangleq(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y},$$

这通常称为普通最小二乘（*OLS*）解。

普通最小二乘解的一个常见问题是，当 $\boldsymbol{X}$ 接近奇异时会出现问题。设 $\boldsymbol{X}$ 的奇异值分解为 $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\in\mathbb{R}^{n\times p}$，其中 $\boldsymbol{U}\in\mathbb{R}^{n\times n}$ 和 $\boldsymbol{V}\in\mathbb{R}^{p\times p}$ 为正交矩阵，且 $\boldsymbol{\Sigma}\in\mathbb{R}^{n\times p}$ 的主对角线包含奇异值$\sigma_1\geq\sigma_2\geq\ldots\geq\sigma_p\geq 0$。因此，$\boldsymbol{X}^\top\boldsymbol{X}=\boldsymbol{V}(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})\boldsymbol{V}^\top\triangleq\boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top$，其中$\boldsymbol{S}\triangleq\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\ldots\ldots,\sigma_p^2)\in\mathbb{R}^{p\times p}$包含 $\boldsymbol{X}$的奇异值平方。当 $\boldsymbol{X}$接近奇异时，最小奇异值满足 $\sigma_p^2\approx 0$，这使得求逆（$\boldsymbol{X}^\top\boldsymbol{X}$）$^{-1}=\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}^\top$ 在数值上不稳定。因此，解$\boldsymbol{\beta}^{OLS}=(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ 可能对扰动极为敏感，甚至发散。为缓解这种不稳定性，通常会添加一个 ℓ_2-正则化项，从而得到以下正则化优化问题：

$$\boldsymbol{\beta}^{Tik}=\arg\min_{\boldsymbol{\beta}}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_2^2. \tag{3.23}$$

这种方法称为蒂霍诺夫正则化（或岭回归）（Tikhonov，1963年）。目标函数的梯度为 $2(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})\boldsymbol{\beta}-2\boldsymbol{X}^\top\boldsymbol{y}$，令其为零可得闭式解 $\boldsymbol{\beta}^{Tik}=(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$. 利用 $\boldsymbol{X}$ 的奇异值分解，逆矩阵变为 $(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}=\boldsymbol{V}(\boldsymbol{S}+\lambda\boldsymbol{I})^{-1}\boldsymbol{V}^\top$，其中 $\widetilde{\boldsymbol{S}}\triangleq(\boldsymbol{S}+\lambda\boldsymbol{I})=\mathrm{diag}(\sigma_1^2+\lambda,\sigma_2^2+\lambda,\ldots,\sigma_p^2+\lambda)$。因此，普通最小二乘解和蒂霍诺夫正则化解可表示为

$$\boldsymbol{\beta}^{OLS}=(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}=\boldsymbol{V}\left(\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\right)\boldsymbol{U}^\top\boldsymbol{y}; \tag{3.24a}$$

$$\boldsymbol{\beta}^{Tik}=(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{y}=\boldsymbol{V}\left((\boldsymbol{S}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\right)\boldsymbol{U}^\top\boldsymbol{y}, \tag{3.24b}$$

[2] 其中，$\left(\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\right)$ 的主对角线为 $\mathrm{diag}(\frac{1}{\sigma_1},\frac{1}{\sigma_2},\ldots\ldots,\frac{1}{\sigma_p})$；而$\left((\boldsymbol{S}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\right)$的主对角线为 $\mathrm{diag}(\frac{\sigma_1}{\sigma_1^2+\lambda},\frac{\sigma_2}{\sigma_2^2+\lambda},\ldots\ldots,\frac{\sigma_p}{\sigma_p^2+\lambda})$。后一表达式仍然表现良好

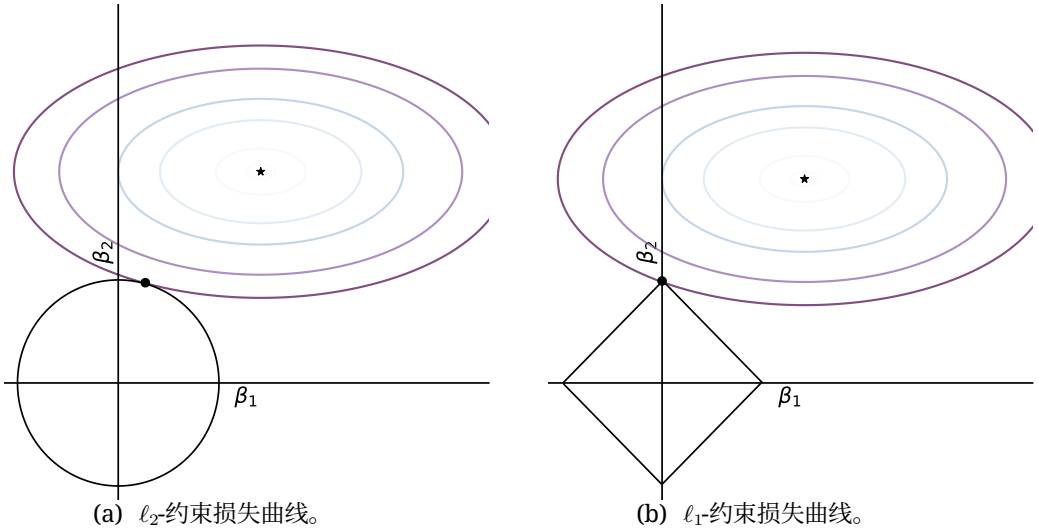
[2]。根据穿越恒等式（习题 3 4 $\boldsymbol{\beta}^{Tik}=$ 。），蒂霍诺夫解也可以写成 $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top+\lambda\boldsymbol{I}^{-1}\boldsymbol{y}$ ( )。

即使在 $\sigma_p \approx 0$时，只要 λ 选取得大于最小的非零奇异值平方即可。这种正则化还改善了系统的条件数（见注释1.38）：当σ_p 接近零时，

$$\kappa(\boldsymbol{X}^\top \boldsymbol{X}) = \sigma_1^2/\sigma_p^2 \qquad \to \qquad \kappa(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}) = (\lambda + \sigma_1^2)/(\lambda + \sigma_p^2).$$

总之，当 $\boldsymbol{\beta}^{OLS}$ $\boldsymbol{X}$ 病态或秩亏时，蒂霍诺夫正则化能够有效防止最小二乘估计发散。这种稳定作用提升了最小二乘算法及其变体（如交替最小二乘（Lu，2022年））的数值收敛性，使蒂霍诺夫正则化成为实践中广泛采用的技术。

练习 3.9. 利用奇异值分解证明，最优解 $\boldsymbol{\beta}^{Tik} = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$的范数随 λ 增大而非递增。



(a) ℓ_2-约束损失曲线。 (b) ℓ_1-约束损失曲线。

图 3.2： ℓ_2-约束与 ℓ_1-约束最小二乘问题的比较。彩色椭圆等高线表示目标函数关于参数 $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$ 的变化，假设不存在约束条件。

▶ **ℓ_1-正则化**。推论 3.3 中我们已经证明， ℓ_2-正则化模型与 ℓ_2-约束模型等价。图3.2(a) 展示了在 ℓ_2-约束模型下的几何解释，其中 $p = 2$。在这种情况下，估计系数会向零收缩，但永远不会恰好为零。图中，彩色椭圆形等高线描绘了无约束最小二乘目标函数关于 $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$的变化。该目标函数的全局最小值位于最内层椭圆的中心。然而，加入约束会限制 $\boldsymbol{\beta}$的可行域，使解无法达到这个无约束最小值。这一限制有助于缓解过拟合。

相比之下， ℓ_1正则化模型（L_L）（第 126页）会促进解的稀疏性；其约束形式见图 3.2(b) 。其中， ℓ_1正则化项将参数限制在菱形区域内，而 ℓ_2正则化项则将参数限制在圆形区域内。如图所示，最优

在 ℓ_1正则化下，解通常出现在菱形的角点（顶点）处——通常是某个系数（如 β_1）恰好为零的地方。相比之下， ℓ_2约束的圆形几何结构很少产生精确的零值，因此得到的解是稠密的（非稀疏的）。

ℓ_0-单位球促进稀疏性，但它不是凸集。而 ℓ_∞-单位球包含所有每个元素大小小于或等于 1 的向量。ℓ_0-单位球与 ℓ_∞-单位球的交集恰好是元素属于 $[-1,1]$的 1-稀疏向量集合。这一观察为使用 ℓ_1-范数来诱导稀疏性提供了另一种依据： ℓ_1-单位球是所有具有单位2.22）范数的 1-稀疏向量的凸包，也就是说，是 ℓ_∞-单位球与 ℓ_0-单位球交集的凸包。这一关系在图 ℓ_∞中以二维情形进行了说明。3.3。

引理 3.10：（ℓ_1-单位球）{The ℓ_1-unit ball is the convex hull of the intersection between the ℓ_0-unitballand the ℓ_∞-unitball.}

证明 [引理3 的10] $\mathbb{B}_1 \subseteq$ **凸包**$(\mathbb{B}_0 \cap \mathbb{B}_\infty)$。 设 $\boldsymbol{\beta} \in \mathbb{B}_1 \subset \mathbb{R}^p$ 使得 $\|\boldsymbol{\beta}\|_1 \leq 1$。如果令$\theta_i \triangleq |\beta_i|$ 且 $\theta_0 \triangleq 1 - \sum_{i=1}^p \theta_i$，则有 $\sum_{i=0}^p \theta_i = 1$ 根据构造， $\theta_i = |\beta_i| \geq 0$ 且

$$\theta_0 = 1 - \sum_{i=1}^{p} \theta_i = 1 - \|\boldsymbol{\beta}\|_1 \geq 0. \tag{3.25}$$

现在，我们可以将 $\boldsymbol{\beta}$ 表示为标准基向量与 $\boldsymbol{\beta}$ 各元素的符号函数的乘积的凸组合，即 sign$(\beta_1)\boldsymbol{e}_1$、sign$(\beta_2)\boldsymbol{e}_2$、.....、sign$(\beta_p)\boldsymbol{e}_p$，以及零向量$\mathbf{0}$:

$$\boldsymbol{\beta} = \sum_{i=1}^{p} \theta_i \operatorname{sign}(\beta_i)\boldsymbol{e}_i + \theta_0 \mathbf{0}, \tag{3.26}$$

其中这些向量都属于 $\mathbb{B}_0 \cap \mathbb{B}_\infty$。因此， $\mathbb{B}_1 \subseteq$ 凸包$(\mathbb{B}_0 \cap \mathbb{B}_\infty)$。

凸包$(\mathbb{B}_0 \cap \mathbb{B}_\infty) \subseteq \mathbb{B}_1$。 设 $\boldsymbol{\beta} \in$ 凸包$(\mathbb{B}_0 \cap \mathbb{B}_\infty) \subset \mathbb{R}^p$。根据凸包的定义，我们可以将 $\boldsymbol{\beta}$ 表示为

$$\boldsymbol{\beta} = \sum_{i=1}^{m} \theta_i \boldsymbol{\alpha}_i, \tag{3.27}$$

其中 $m > 0$, $\boldsymbol{\alpha}_1$、.....、 $\boldsymbol{\alpha}_m \in \mathbb{R}^p$ 各自只有一个元素的绝对值不超过 1， $\theta_i \geq 0$ 对所有 $1 \leq i \leq m$和 $\sum_{i=1}^m \theta_i = 1$均成立。根据三角不等式，我们有

$$\|\boldsymbol{\beta}\|_1 \leq \sum_{i=1}^{m} \theta_i \|\boldsymbol{\alpha}_i\|_1 = \sum_{i=1}^{m} \theta_i \|\boldsymbol{\alpha}_i\|_\infty \leq \sum_{i=1}^{m} \theta_i \leq 1, \tag{3.28}$$

其中等号成立是因为每个向量 $\boldsymbol{\alpha}_i$ 都只有一个非零元素。这立即推出 $\boldsymbol{\beta} \in \mathbb{B}_1$ ，并完成证明。

■

3.3. ℓ_1-正则化问题的最优性条件

我们此前已证明， ℓ_1正则化最小二乘估计量能够促进解的稀疏性：估计系数向量 $\boldsymbol{\beta}$ 中的许多分量会被精确地压至零，从而得到易于解释的紧凑模型。这一性质使 ℓ_1正则化在需要进行特征选择的高维场景中特别有价值。

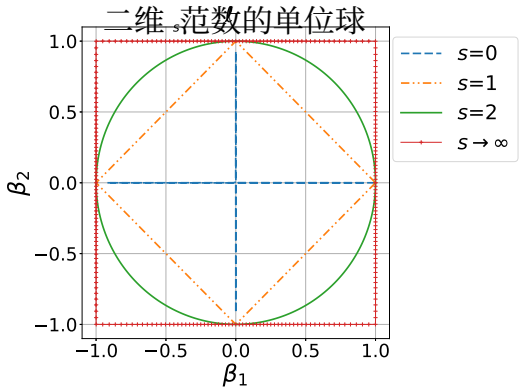


图3.3： 二维中 ℓ_0 球与 ℓ_∞球交集的示意图，以及它们的凸包所形成的 ℓ_1球。

为了更正式地分析这一行为，我们现在转而讨论 ℓ_1正则化最小二乘问题的最优性条件——该问题通常称为拉格朗日 LASSO（参见（L_L），第 126页）。其惩罚（拉格朗日）形式表述如下：

$$(\mathrm{L}_L) \quad \min_{\boldsymbol{\beta}} F(\lambda,\boldsymbol{\beta}) \triangleq \left\{ \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\}. \tag{3.29}$$

目标函数（3.29）关于 $\boldsymbol{\beta}$ 的次梯度为

$$\partial_{\boldsymbol{\beta}} F(\lambda,\boldsymbol{\beta}) = \frac{\partial F(\lambda,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + \lambda\nabla_{\boldsymbol{\beta}}\|\boldsymbol{\beta}\|_1 \triangleq -\boldsymbol{c} + \lambda\partial\|\boldsymbol{\beta}\|_1, \tag{3.30}$$

其中 $\boldsymbol{c} \triangleq \boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$ 是残差相关向量，向量 $\partial\|\boldsymbol{\beta}\|_1 = [\partial_{\beta_1}\|\boldsymbol{\beta}\|_1, \partial_{\beta_2}\|\boldsymbol{\beta}\|_1, \ldots, \partial_{\beta_p}\|\boldsymbol{\beta}\|_1]^\top$ 表示 ℓ_1范数 $\|\boldsymbol{\beta}\|_1$（练习2.58）的次微分，其中 i个元素

$$\partial|\beta_i| = \begin{cases} \{+1\}, & \beta_i > 0; \\ \{-1\}, & \beta_i < 0; \\ [-1,+1], & \beta_i = 0, \end{cases} \qquad \forall\, i = 1,2,\ldots,p. \tag{3.31}$$

由（3.30）可知，拉格朗日 LASSO 问题（L_L）的驻点必须满足 $\partial_{\boldsymbol{\beta}} F(\lambda,\boldsymbol{\beta}) = -\boldsymbol{c} + \lambda\partial\|\boldsymbol{\beta}\|_1 = \boldsymbol{0}$，即

$$\boldsymbol{c} = \lambda\partial\|\boldsymbol{\beta}\|_1. \tag{3.32}$$

等价地，对于每个 $i=1$，2，…，p，都有 $c_i = \lambda\partial|\beta_i|$。由于拉格朗日 LASSO 问题是凸的，该驻点条件对于最优性既是必要条件也是充分条件（定理2.62）。

最优性条件（3.32）可以用支撑集 $\mathbb{S} \triangleq \{i \mid \beta_i \neq 0\}$ 及其补集 $\mathbb{S}^c = \{1,2,\ldots,p\} \setminus \mathbb{S}$ 表示为

$$\boldsymbol{c}[\mathbb{S}] = \lambda \cdot \mathrm{sign}(\boldsymbol{\beta}[\mathbb{S}]) \qquad \text{and} \qquad |\boldsymbol{c}[\mathbb{S}^c]| \leq \lambda\boldsymbol{1}, \tag{3.33}$$

这表明，对于支撑集中的索引 $\mathbb{S}$，残差相关性的大小恰好等于 λ，并且其符号函数与 $\boldsymbol{\beta}$中相应系数的符号函数相同。等价地，条件（3.33）可以写成

$$|c_i| = \lambda, \quad \forall\, i \in \mathbb{S} \qquad \text{and} \qquad |c_i| \leq \lambda, \quad \forall\, i \in \mathbb{S}^c. \tag{3.34}$$

换言之，对于 λ活跃（非零）系数，残差相关性的绝对值等于 λ；对于非活跃（零）系数，该绝对值不超过 λ。因此，最大绝对残差相关性满足 $\|\boldsymbol{c}\|_\infty = \max\{c_i\} = \lambda$。

备注 3.11（约束 LASSO 的 KKT 条件）。式（3.34）中的条件与约束 LASSO 问题（L_C）（第 126 页）的 KKT 条件相似。这种等价性源于惩罚形式与约束形式 LASSO 之间共享的最优性结构。对于给定的 $\boldsymbol{\beta}$ 范数约束 ℓ_1，以下 KKT 条件（定理2.115）必须成立：

(i) 次梯度条件。 最小二乘损失的梯度（$-\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$）必须属于 $\nu\left\|\widehat{\boldsymbol{\beta}}\right\|_1$ 的次微分，其中 ν 是与该约束相关联的拉格朗日乘子。这可转化为：

- 对于任意满足 $\widehat{\beta}_i \neq 0$ 的 i： $\left|\boldsymbol{x}_i^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right| = \nu$（预测变量 i 与残差之间的绝对相关性恰好为 ν）。对于任意满足 $\widehat{\beta}_i = 0$ 的 i： $\left|\boldsymbol{x}_i^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right| \leq \nu$
- （绝对相关性不超过 ν）。

(ii) 互补松弛。$\nu(\left\|\widehat{\boldsymbol{\beta}}\right\|_1 - \Sigma) = 0$。如果 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1 < \Sigma$，则 $\nu = 0$（无正则化）。如果 $\nu > 0$，则 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1 = \Sigma$（约束处于激活状态）。

3.3.1 LASSO 对偶

对于拉格朗日 LASSO 问题（L_L），引入残差向量 $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$后，我们可以将原始问题等价地改写为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{r}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{r}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}. \tag{3.35}$$

令 $\boldsymbol{\nu} \in \mathbb{R}^n$ 表示与等式约束对应的拉格朗日乘子向量。该问题的拉格朗日函数为

$$L(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\nu}) = \frac{1}{2}\|\boldsymbol{r}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 - \boldsymbol{\nu}^\top(\boldsymbol{r} - \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta}). \tag{3.36}$$

通过对 $L(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\nu})$ 关于 $\boldsymbol{\beta}$ 和 $\boldsymbol{r}$ 求最小值，可得到拉格朗日对偶函数。为了求得拉格朗日对偶函数 $\min_{\boldsymbol{\beta},\boldsymbol{r}} L(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\nu})$，将包含 $\boldsymbol{\beta}$ 的项分离出来，并注意到根据 H $\left|\boldsymbol{\nu}^\top\boldsymbol{X}\boldsymbol{\beta}\right| \leq \|\boldsymbol{\beta}\|_1 \left\|\boldsymbol{X}^\top\boldsymbol{\nu}\right\|_\infty$ older 不等式（定理 1.60），我们得到

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \lambda\|\boldsymbol{\beta}\|_1 - \boldsymbol{\nu}^\top\boldsymbol{X}\boldsymbol{\beta} = \begin{cases} 0, & \text{if } \left\|\boldsymbol{X}^\top\boldsymbol{\nu}\right\|_\infty \leq \lambda; \\ -\infty, & \text{otherwise.} \end{cases} \tag{3.37}$$

接下来，我们将含有 $\boldsymbol{r}$ 的项分离出来，得到

$$\min_{\boldsymbol{r}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{r}\|_2^2 - \boldsymbol{\nu}^\top\boldsymbol{r} = -\frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{\nu}, \quad \text{with } \boldsymbol{r} = \boldsymbol{\nu}. \tag{3.38}$$

将关系式（3.37）和（3.38）代入拉格朗日表示（3.36），得到

$$\begin{aligned} \text{LASSO Dual:} \qquad \phi(\lambda) &\triangleq \max_{\boldsymbol{\nu}} \frac{1}{2}\left(\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{\nu}\|_2^2\right) \triangleq \max_{\boldsymbol{\nu}} D(\boldsymbol{\nu}) \\ &\text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\nu}\right\|_\infty \leq \lambda, \end{aligned} \tag{3.39}$$

其中，$D(\boldsymbol{\nu}) \triangleq \frac{1}{2}\left(\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{\nu}\|_2^2\right)$ 是对偶目标函数，$\phi(\lambda)$ 表示与惩罚参数 λ对应的最优对偶值。

从几何角度看，这种对偶形式相当于将响应向量 $\boldsymbol{y}$ 投影到可行集 $\mathbb{S}_\lambda \triangleq \{\boldsymbol{\nu} \in \mathbb{R}^n \mid \left\|\boldsymbol{X}^\top\boldsymbol{\nu}\right\|_\infty \leq \lambda\}$上。集合 $\mathbb{S}_\lambda$ 是 $2p$

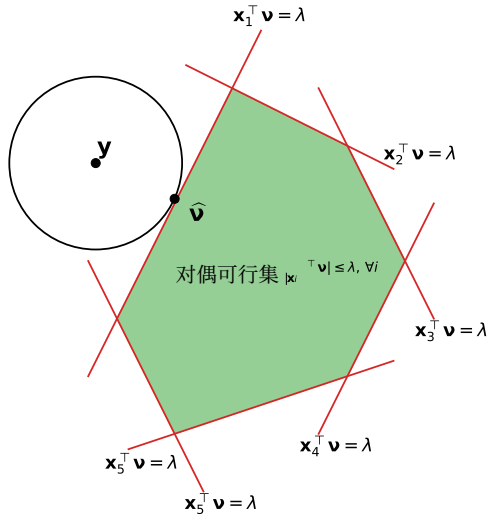


图 3.4： LASSO 的拉格朗日对偶，其中 $D(\boldsymbol{\nu}) = \frac{1}{2}(\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{\nu}\|_2^2)$。阴影区域为可行集 $\mathbb{S}_\lambda$。无约束的对偶问题解为 $\boldsymbol{\nu} = \boldsymbol{y}$，即零残差。对偶问题解 $\widehat{\boldsymbol{\nu}} = \mathcal{P}_{\mathbb{S}_\lambda}(\boldsymbol{y})$ 是将 $\boldsymbol{y}$ 投影到凸集 $\mathbb{S}_\lambda$ 上。

由 $\left|\boldsymbol{x}_i^\top \boldsymbol{\nu}\right| \le \lambda$ 定义的半空间中，每一个都是 $\mathbb{R}^n$ 中的一个凸多胞体。用投影和近端算子的语言来说，解由投影或近端映射 $\widehat{\boldsymbol{\nu}} = \mathcal{P}_{\mathbb{S}_\lambda}(\boldsymbol{y}) \equiv \mathcal{P}\mathrm{rox}^{\delta_{\mathbb{S}_\lambda}}(\boldsymbol{y})$ 给出。换句话说，$\widehat{\boldsymbol{\nu}}$ 是投影到多胞体 $\mathbb{S}_\lambda$ 上。图 3.4 展示了这一几何解释。

▶ **稀疏性的对偶可行性条件**。对于原始 LASSO 问题（L_L），强对偶性成立（定理 2.119）。那么，最优原始 $\boldsymbol{\beta}$ 解和对偶 $\widehat{\boldsymbol{\nu}}$ 解满足

$$\boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\nu}}. \tag{3.40}$$

对偶问题在识别原始解中的稀疏性方面发挥着关键作用：通过考察对偶变量 $\widehat{\boldsymbol{\nu}}$，我们可以确定 $\boldsymbol{\beta}$ 中哪些系数必须恰好为零。具体而言，最优性条件

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}\right| < \lambda \quad \Longrightarrow \quad \widehat{\beta}_i(\lambda) = 0, \quad i = 1, 2, \ldots, p, \tag{3.41}$$

直接由（3.34）和（3.38）得出。请注意，上述条件足以保证是指 β_i（λ）必须是非活跃预测变量，但这并非必要条件。事实上，即使$\widehat{\beta}_i$（λ）$= 0$，也有可能出现 $\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}\right| = \lambda$。在这种情况下，该预测变量位于活跃集的边界上，被视为边际非活跃。

▶ 非平凡解的最大值 λ。这种形式的 LASSO 对偶还揭示了使原始 LASSO 问题存在非平凡解 $\boldsymbol{\beta}_{\max} \neq \mathbf{0}$的惩罚参数最大值，记为 $\lambda = \lambda_{\max}$。如果 $\lambda = \lambda_{\max}$ 取值足够大，最优原始解将恒为零：$\boldsymbol{\beta} = \boldsymbol{\beta}_{\max} = \mathbf{0}$。为确定这样的 $\lambda_{\max}$，我们将 $\widehat{\boldsymbol{\beta}}_{\max} = \mathbf{0}$ 代入原始 LASSO 问题（L_L），得到 ϕ（$\lambda_{\max}$）$= \frac{1}{2}\|\boldsymbol{y}\|_2^2$，因为 $\boldsymbol{y} = \boldsymbol{\nu}$ 在（3.38）中成立。根据强对偶性，在 $\lambda = \lambda_{\max}$ 下，对偶问题在唯一解处达到值 ϕ（$\lambda_{\max}$）$= \frac{1}{2}\|\boldsymbol{y}\|_2^2$

$$\widehat{\boldsymbol{\nu}}_{\max} = \boldsymbol{y}. \tag{3.42}$$

点 $\widehat{\boldsymbol{\nu}}_{\max}$ 是一个对偶可行点，并满足约束条件$\left|\boldsymbol{x}_i^\top \boldsymbol{y}\right| \le \lambda_{\max}$，$i$ =1、2、.....、p。因此，这种 LASSO 对偶形式给出了使解 $\boldsymbol{\beta}_{\max}$ 非平凡的最大 $\lambda = \lambda_{\max}$，即

$$\lambda_{\max} = \max_{1 \le i \le p} \left|\boldsymbol{x}_i^\top \boldsymbol{y}\right| = \left\|\boldsymbol{X}^\top \boldsymbol{y}\right\|_\infty. \tag{3.43}$$

3.3.2 对偶多面体投影与筛选规则

(3.43) 中的量揭示了一个重要的直觉：与内积较大的协变量相比，与响应向量 $\boldsymbol{y}$ 内积较小的协变量，在 LASSO 解中获得非零系数的可能性较低。

这一观察启发了一种实用策略：在求解完整优化问题之前，我们或许可以消除无关的预测变量，从而降低计算成本。例如，在基因组学应用中，数据集往往涉及数百万个单核苷酸多态性（SNP），而我们通常预计最终模型中只会出现少数几个变量。

本小节将讨论利用这一思路的筛选规则。这些规则能够高效地识别并剔除非活跃预测变量（即在最优解中系数必定为零的特征），从而大幅加快 LASSO 的计算，同时不影响数值精度。由此得到的约化问题与原问题具有完全相同的解，但计算成本却低得多。

我们首先介绍一种基于向对偶可行集投影的几何筛选规则，称为对偶多面体投影（*DPP*）规则（Ghaoui 等人，2010年；Wang 等人，2013年）。

定理 3.12：（对偶多面体投影（DPP））假设给定一个$\widehat{}$解 $\widehat{\boldsymbol{\nu}}(\lambda')$，它是 LASSO 对偶（3.39）的解，以及在某个惩罚参数$0 < \lambda' \leq \lambda_{\max}$下的一个原始解 $\boldsymbol{\beta}(\lambda')$。令 λ 为一个非负且不同于 λ'的值。若对于给定的索引 i，以下不等式成立：

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}(\lambda')\right| < \lambda' - \|\boldsymbol{x}_i\|_2 \|\boldsymbol{y}\|_2 \left|1 - \frac{\lambda'}{\lambda}\right|, \tag{3.44}$$

$\widehat{}$那么，在 λ 处的 LASSO 解中，第 i 个系数为零：$\beta_i(\lambda) = 0$，其中 $\boldsymbol{x}_i$ 表示 $\boldsymbol{X}$的第 i 列。

定理 [的证明 3.12] 引入变量替换 $\boldsymbol{\mu} \triangleq \boldsymbol{\nu}/\lambda$。则 LASSO对偶问题（3.39）变为

$$\text{LASSO Dual:} \quad \max_{\boldsymbol{\mu}} \frac{1}{2}\left(\|\boldsymbol{y}\|_2^2 - \lambda^2 \|\boldsymbol{y}/\lambda - \boldsymbol{\mu}\|_2^2\right) \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top \boldsymbol{\mu}\right\|_\infty \leq 1. \tag{3.45}$$

因此，缩放对偶空间中的可行集是固定的凸多面体 $\widetilde{\mathbb{S}} \triangleq \{\boldsymbol{z} \in \mathbb{R}^n \left\|\boldsymbol{X}^\top \boldsymbol{z}\right\|_\infty \leq 1\}$，它与 λ 无关。最优缩放对偶变量 $\widehat{\boldsymbol{\mu}}(\lambda)$ 是将 $\boldsymbol{y}/\lambda$ 投影到可行集 $\mathbb{S}$ 上的$\widetilde{}$投影。

由（3.41），根据 LASSO 问题的驻点条件，我们有

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\mu}}(\lambda)\right| < 1 \quad \Longrightarrow \quad \widehat{\beta}_i(\lambda) = 0, \quad i = 1, 2, \ldots, n. \tag{3.46}$$

由于不同惩罚参数对应的多面体 λ 相同，根据投影算子的非扩张性（定理2.103），可得

$$\left\|\widehat{\boldsymbol{\mu}}(\lambda) - \widehat{\boldsymbol{\mu}}(\lambda')\right\|_2 = \left\|\mathcal{P}_{\widetilde{\mathbb{S}}}\left(\frac{\boldsymbol{y}}{\lambda}\right) - \mathcal{P}_{\widetilde{\mathbb{S}}}\left(\frac{\boldsymbol{y}}{\lambda'}\right)\right\|_2 \leq \left\|\frac{\boldsymbol{y}}{\lambda} - \frac{\boldsymbol{y}}{\lambda'}\right\|_2 = \|\boldsymbol{y}\|_2 \left|\frac{1}{\lambda} - \frac{1}{\lambda'}\right|. \tag{3.47}$$

利用三角不等式和柯西–施瓦茨不等式，由上述不等式及假设（3.44）可得

$$\begin{aligned}\left|\boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\mu}}(\lambda)\right| &\leq \left|\boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\mu}}(\lambda)-\boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\mu}}(\lambda')\right|+\left|\boldsymbol{x}_i^{\top}\widehat{\boldsymbol{\mu}}(\lambda')\right| \\ &< \|\boldsymbol{x}_i\|_2\left\|\widehat{\boldsymbol{\mu}}(\lambda)-\widehat{\boldsymbol{\mu}}(\lambda')\right\|_2+1-\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}\|_2\left|\frac{1}{\lambda'}-\frac{1}{\lambda}\right| \\ &\leq \|\boldsymbol{x}_i\|_2\|\boldsymbol{y}\|_2\left|\frac{1}{\lambda'}-\frac{1}{\lambda}\right|+1-\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}\|_2\left|\frac{1}{\lambda'}-\frac{1}{\lambda}\right|=1.\end{aligned}$$

使用（3.46）即可完成证明。 ■

上述定理表明，如果

$$\left|\boldsymbol{x}_i^{\top}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda')\right)\right|<\lambda'-\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}\|_2\left|1-\frac{\lambda'}{\lambda}\right|, \tag{3.48a}$$

那么变量 i 在惩罚水平 λ（即 $\beta_i(\lambda)=0$下不属于活跃集）。特别地，在 $\lambda'=\lambda_{\max}\equiv\left\|\boldsymbol{X}^{\top}\boldsymbol{y}\right\|_{\infty}$ 由（3.43）可知，上述不等式变为

$$\left|\boldsymbol{x}_i^{\top}\boldsymbol{y}\right|<\lambda_{\max}-\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}\|_2\left(\frac{\lambda_{\max}}{\lambda}-1\right). \tag{3.48b}$$

该对偶多胞体投影（*DPP*）规则随后建议，只要（i3.48b）成立，就丢弃第个预测变量（Wang 等人，2013年）。

▶ DPP规则的变体。另一方面，假设 $\frac{\boldsymbol{y}}{\lambda'}\notin\widetilde{\mathbb{S}}\triangleq\{\boldsymbol{z}\in\mathbb{R}^n\mid\left\|\boldsymbol{X}^{\top}\boldsymbol{z}\right\|_{\infty}\leq 1\}$，即 $\lambda'\in(0,\lambda_{\max})$。在这种情况下，缩放对偶解满足 $\frac{\boldsymbol{y}}{\lambda'}\neq\mathcal{P}_{\widetilde{\mathbb{S}}}(\frac{\boldsymbol{y}}{\lambda'})\equiv\widehat{\boldsymbol{\mu}}(\lambda')$（回顾一下，$\widehat{\boldsymbol{\mu}}(\lambda')=\widehat{\boldsymbol{\nu}}(\lambda')/\lambda'$ 是在惩罚参数 $\lambda=\lambda'$ 下的缩放对偶解）。令$\boldsymbol{\mu}(\eta)\triangleq\widehat{\boldsymbol{\mu}}(\lambda')+\eta\left(\frac{\boldsymbol{y}}{\lambda'}-\widehat{\boldsymbol{\mu}}(\lambda')\right)$ 对于 $\eta\geq 0$，即 $\boldsymbol{\mu}(\eta)$ 是位于从$\widehat{\boldsymbol{\mu}}(\lambda')$出发、指向与 $\frac{\boldsymbol{y}}{\lambda'}-\widehat{\boldsymbol{\mu}}(\lambda')$ 相同方向的射线上的点。我们可以观察到，$\mathcal{P}_{\widetilde{\mathbb{S}}}(\boldsymbol{\mu}(\eta))=\widehat{\boldsymbol{\mu}}(\lambda')$，即将 $\boldsymbol{\mu}(\eta)$ 投影到集合 $\mathbb{S}$ 上所得的结果也是 $\widehat{\boldsymbol{\mu}}(\lambda')$。根据投影算子的非扩张性（定理 2.103），我们有

$$\begin{aligned}\left\|\widehat{\boldsymbol{\mu}}(\lambda)-\widehat{\boldsymbol{\mu}}(\lambda')\right\|_2&=\left\|\mathcal{P}_{\widetilde{\mathbb{S}}}\left(\frac{\boldsymbol{y}}{\lambda}\right)-\mathcal{P}_{\widetilde{\mathbb{S}}}(\boldsymbol{\mu}(\eta))\right\|_2 \\ &\leq\left\|\frac{\boldsymbol{y}}{\lambda}-\boldsymbol{\mu}(\eta)\right\|_2=\left\|\eta\left(\frac{\boldsymbol{y}}{\lambda'}-\widehat{\boldsymbol{\mu}}(\lambda')\right)-\left(\frac{\boldsymbol{y}}{\lambda}-\widehat{\boldsymbol{\mu}}(\lambda')\right)\right\|_2.\end{aligned} \tag{3.49}$$

显然，当 $\eta=1$时，不等式（3.49）可化为（3.47）中的不等式。由于不等式（3.49）对所有 $\eta\geq 0$ 都成立，我们可以通过以下方式得到更紧的上界：

$$\left\|\widehat{\boldsymbol{\mu}}(\lambda)-\widehat{\boldsymbol{\mu}}(\lambda')\right\|_2\leq\min_{\eta\geq 0}\|\eta\boldsymbol{u}_1-\boldsymbol{u}_2\|_2, \tag{3.50}$$

其中 $\boldsymbol{u}_1\triangleq\frac{\boldsymbol{y}}{\lambda'}-\widehat{\boldsymbol{\mu}}(\lambda')$ 且 $\boldsymbol{u}_2\triangleq\frac{\boldsymbol{y}}{\lambda}-\widehat{\boldsymbol{\mu}}(\lambda')$。当 $\lambda'=\lambda_{\max}$时，我们可以令 $\boldsymbol{u}_1=\operatorname{sign}(\boldsymbol{x}_*^{\top}\boldsymbol{y})\boldsymbol{x}_*$，其中 $\boldsymbol{x}_*\triangleq\operatorname{argmax}_{\boldsymbol{x}_i}\left|\boldsymbol{x}_i^{\top}\boldsymbol{y}\right|$。不等式（3.50）右侧的最小化问题具有闭式解：

$$\min_{\eta\geq 0}\|\eta\boldsymbol{u}_1-\boldsymbol{u}_2\|_2=\varphi(\lambda',\lambda)\triangleq\begin{cases}\|\boldsymbol{u}_2\|_2, & \text{if } \langle\boldsymbol{u}_1,\boldsymbol{u}_2\rangle<0;\\ \left\|\frac{\langle\boldsymbol{u}_1,\boldsymbol{u}_2\rangle}{\|\boldsymbol{u}_1\|_2^2}\boldsymbol{u}_1-\boldsymbol{u}_2\right\|_2, & \text{otherwise.}\end{cases} \tag{3.51}$$

这个改进后的上界引出了增强型 DPP 筛选规则，如下所述。

推论 3.13：（增强型 DPP 规则）设我们给定一个解 $\widehat{\boldsymbol{\nu}}(\lambda')$ 和一个LASSO 对偶（3.39）的原始解 $\boldsymbol{\beta}(\lambda')$，其对应于某个惩罚参数 $0 < \lambda' \leq \lambda_{\max}$。设 λ 为一个非负且不同于 λ' 的值。若对于给定的索引 i，以下不等式成立：

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}(\lambda')\right| < \lambda' - \lambda' \left\|\boldsymbol{x}_i\right\|_2 \varphi(\lambda', \lambda), \tag{3.52}$$

则 $\beta_i(\lambda) = 0$，其中 $\boldsymbol{x}_i$ 是 $\boldsymbol{X}$的第 i列。

上述推论表明，如果

$$\left|\boldsymbol{x}_i^\top \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda')\right)\right| < \lambda' - \lambda' \left\|\boldsymbol{x}_i\right\|_2 \varphi(\lambda', \lambda), \tag{3.53a}$$

则变量 i 在惩罚水平 λ下不属于活跃集。在 $\lambda' = \lambda_{\max} \equiv \left\|\boldsymbol{X}^\top \boldsymbol{y}\right\|_\infty$<style id='10'}根据（3.43），该条件变为

$$\left|\boldsymbol{x}_i^\top \boldsymbol{y}\right| < \lambda_{\max} - \lambda_{\max} \left\|\boldsymbol{x}_i\right\|_2 \varphi(\lambda_{\max}, \lambda). \tag{3.53b}$$

同样，增强型 DPP 规则建议：如果满足（3.53b），则舍弃第 i 个变量。

▶ 筛选规则。方程（3.48a）或（3.53a）为高效求解拉格朗日 LASSO 问题（L_L）提供了实用的筛选规则。假设我们已知某个惩罚参数 λ' 下的最优对偶解 $\widehat{\boldsymbol{\nu}}(\lambda')$（例如，广为人知的情形 $\widehat{\boldsymbol{\nu}}(\lambda_{\max}) = \boldsymbol{y}$）。利用这些信息，我们可以界定在邻近参数 λ 下对偶问题解 $\widehat{\boldsymbol{\nu}}(\lambda)$ 可能所在的位置。通过界定 $\widehat{\boldsymbol{\nu}}(\lambda)$ 的这一可能位置，我们可以检查 $\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}(\lambda)\right|$是否必然小于 λ。如果是，则相应的预测变量 $\boldsymbol{x}_i$ 在 λ 下处于非活跃状态（即 $\beta_i(\lambda) = 0$），可以安全地从优化问题中移除，而不会影响最终解。

序贯 DDP

受对偶多面体投影定理（定理3.12）及上述筛选原则的启发，我们可以构造一种序贯 DPP 策略。具体而言，考虑一个递减的惩罚参数序列：$\lambda_0 = \lambda_{\max} > \lambda_1 > \lambda_2 > \ldots > \lambda_m$。我们首先在 $\lambda_0 = \lambda_{\max}$ 处应用 DPP 筛选规则，剔除那些可确定为非活跃的预测变量。由于 $\widehat{\boldsymbol{\nu}}(\lambda_{\max}) = \boldsymbol{y}$，这一初始筛选尤其简单有效。

剔除非活跃变量后，我们使用第 λ_1 节中的任意算法求解9.3处的约化 LASSO 问题，从而得到精确的原始解 $\boldsymbol{\beta}(\lambda_1)$。根据原始解与对偶解的关系式（3.40），我们进而恢复对偶问题解：$\widehat{\boldsymbol{\nu}}(\lambda_1) = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}(\lambda_1)$。这一新计算得到的对偶问题解使我们能够再次应用 DPP 规则——此时界更紧——为下一个参数 λ_2筛除非活跃预测变量。迭代重复这一过程，即可形成一个序贯筛选框架，在每一步逐步缩减问题规模，在保持精确性的同时显著加快计算。

推论 3.14：（序贯 DPP）对于拉格朗日 LASSO 问题（L_L），假设给定参数值序列 $\lambda_{\max} = \lambda_0 > \lambda_1 > \ldots > \lambda_m$。那么对于任意整数 $0 \leq t < m$，如果已知 $\beta_i(\lambda_{t+1})$，则有 $= 0$，并且以下结论成立：

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}(\lambda_t)\right| \equiv \left|\boldsymbol{x}_i^\top \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda_t)\right)\right| < \lambda_t - \left\|\boldsymbol{x}_i\right\|_2 \left\|\boldsymbol{y}\right\|_2 \left(\frac{\lambda_t}{\lambda_{t+1}} - 1\right), \quad i = 1, 2, \ldots, p.$$

or
$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}(\lambda_t)\right| \equiv \left|\boldsymbol{x}_i^\top \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda_t)\right)\right| < \lambda_t - \lambda_t \left\|\boldsymbol{x}_i\right\|_2 \varphi(\lambda_t, \lambda_{t+1}), \quad i = 1, 2, \ldots, p.$$

3.3.3 安全特征消除（SAFE）

SAFE（安全特征消除）规则是稀疏优化问题中使用的另一种强大技术，尤其适用于LASSO型回归，可高效识别并消除那些能够确定在最优解中系数为零的特征（设计矩阵的列），而无需求解完整问题。这能显著加快计算速度，尤其是在处理包含许多无关特征的高维数据时。

SAFE背后的直觉源自式（3.41）中的驻点条件，现回顾如下：

$$\left|\boldsymbol{x}_i^\top \widehat{\boldsymbol{\nu}}\right| < \lambda \quad \Longrightarrow \quad \widehat{\beta}_i(\lambda) = 0, \quad i = 1, 2, \ldots, p, \tag{3.54}$$

SAFE规则构造了一个已知包含对偶最优解$\widehat{\boldsymbol{\nu}}$的约束集 Θ 。随后，它用整个集合 Θ替换不等式测试中的 $\widehat{\boldsymbol{\nu}}$ 。具体而言，若

$$\left|\boldsymbol{x}_i^\top \boldsymbol{\nu}\right| < \lambda, \quad \text{for all } \boldsymbol{\nu} \in \Theta, \tag{3.55}$$

那么，式（3.54）中的条件对 Θ 中的所有元素均成立。因此， i-个分量 $\widehat{}$的原始最优解必定为零： $\beta_i = 0$。

构造对偶约束集

我们首先构造一个集合 Θ 并保证该集合包含$\widehat{\boldsymbol{\nu}}$ LASSO 对偶问题（3.39）的对偶最优解：

$$\begin{aligned} \text{LASSO Dual:} \qquad \phi(\lambda) \triangleq \max_{\boldsymbol{\nu}} \left\{ D(\boldsymbol{\nu}) \triangleq \frac{1}{2}\left(\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{\nu}\|_2^2\right) = \boldsymbol{y}^\top \boldsymbol{\nu} - \frac{1}{2}\|\boldsymbol{\nu}\|_2^2 \right\} \\ \text{s.t.} \quad \left\|\boldsymbol{X}^\top \boldsymbol{\nu}\right\|_\infty \le \lambda, \end{aligned} \tag{3.56}$$

在惩罚参数 λ下。为此，我们将 Θ定义为两个集合 Θ_1和 Θ_2的交集，它们分别由不同的最优性条件导出。

▶ 构造 Θ_1。第一个集合 Θ_1基于以下事实：对偶目标函数 $D(\boldsymbol{\nu})$在 $\widehat{\boldsymbol{\nu}}$处取得最大值。也就是说，对于任意对偶可行点 $\boldsymbol{\nu}$，都有 $D(\widehat{\boldsymbol{\nu}}) \ge D(\boldsymbol{\nu})$。设$\boldsymbol{\nu}_\eta$ 是问题（3.56）的一个特定对偶可行点，并定义 $\gamma \triangleq D(\boldsymbol{\nu}_\eta)$。于是必有 $D(\widehat{\boldsymbol{\nu}}) \ge \gamma$，因此集合 $\Theta_1 \triangleq \{\boldsymbol{\nu} \mid D(\boldsymbol{\nu}) \ge \gamma\}$ 包含对偶最优解 $\widehat{\boldsymbol{\nu}}$，即$\widehat{\boldsymbol{\nu}} \in \Theta_1$。

获得此类可行点的一种实用方法是$\boldsymbol{\nu}_\eta$ 采用对偶缩放。具体而言，设 $\widehat{\boldsymbol{\nu}}_0$ 是已知的、对应于较大惩罚参数 $\lambda_0 > \lambda$的对偶可行点（例如式（3.42）所示）。然后定义 $\boldsymbol{\nu}_\eta \triangleq \eta\widehat{\boldsymbol{\nu}}_0$ 其中标量 $\eta \in \mathbb{R}$ 的选取应使$\boldsymbol{\nu}_\eta$对于惩罚参数为 λ的对偶问题仍然可行；也就是说，$\left\|\boldsymbol{X}^\top \boldsymbol{\nu}_\eta\right\|_\infty \le \lambda$ 或者 $|\eta| \le \lambda/\lambda_0$。现在，我们选择 η 使该区间上的对偶目标函数最大：

$$\gamma = \max_\eta \left\{ D(\eta\widehat{\boldsymbol{\nu}}_0) \mid |\eta| \le \frac{\lambda}{\lambda_0} \right\} = \max_\eta \left\{ -\frac{1}{2}a\eta^2 + b\eta \mid |\eta| \le \frac{\lambda}{\lambda_0} \right\},$$

其中 $a \triangleq \|\widehat{\boldsymbol{\nu}}_0\|_2^2 > 0$和 $b \triangleq \left|\boldsymbol{y}^\top \widehat{\boldsymbol{\nu}}_0\right|$。由 $\eta = \lambda/\lambda_0$可得 $\gamma = -\frac{a}{2}\left(\frac{\lambda}{\lambda_0}\right)^2 + b\frac{\lambda}{\lambda_0}$，使得

$$\widehat{\boldsymbol{\nu}} \in \Theta_1 \triangleq \{\boldsymbol{\nu} \mid D(\boldsymbol{\nu}) \ge \gamma\}. \tag{3.57}$$

▶ 构造 Θ_2。第二个集合 Θ_2源自对偶问题 (3.56) 的一阶最优性条件。回想一下，对于定义在凸可行集上的凸最大化问题，点 $\widehat{\boldsymbol{\nu}}_0$ 是最优的，当且仅当对于每个在惩罚参数 λ_0下对偶问题 (3.56) 可行的对偶点 $\boldsymbol{\nu}$，都有 $\boldsymbol{g}^\top(\boldsymbol{\nu}-\widehat{\boldsymbol{\nu}}_0)\geq 0$，其中 $\boldsymbol{g}\triangleq\nabla D(\widehat{\boldsymbol{\nu}}_0)=\boldsymbol{y}-\widehat{\boldsymbol{\nu}}_0$（关于凸约束优化中的一般最优性条件，参见定理 2.112）。现在考虑惩罚参数 $\lambda\leq\lambda_0$。在惩罚参数 λ 下对偶问题 (3.56) 可行的任意对偶点 $\boldsymbol{\nu}$都满足 $\left\|\boldsymbol{X}^\top\boldsymbol{\nu}\right\|_\infty\leq\lambda\leq\lambda_0$，因此对于参数 λ_0下的对偶问题 (3.56) 也同样对偶可行。由于 $\widehat{\boldsymbol{\nu}}$ （惩罚参数为 λ时的对偶最优解）在 λ下可行，因此它在 λ_0下也可行。因此，我们定义

$$\widehat{\boldsymbol{\nu}}\in\Theta_2\triangleq\left\{\boldsymbol{\nu}\mid\boldsymbol{g}^\top(\boldsymbol{\nu}-\widehat{\boldsymbol{\nu}}_0)\geq 0\right\}. \tag{3.58}$$

综合这两个条件，对偶最优解必须位于这两个集合的交集中：

$$\Theta=\Theta_1\cap\Theta_2.$$

该集合 Θ 提供了一个经过证明且足够紧的区域，其中包含 $\widehat{\boldsymbol{\nu}}$，从而可以根据前文所述规则安全地消除特征。

安全特征消除

^识别原始解的第 i 个分量是否为零的任务——若为零，便可安全地从拉格朗日 LASSO 问题（ $\boldsymbol{\beta}$ ）中移除第 i 个特征（即设计矩阵 i 的第 $\boldsymbol{x}_i$ 列 $\boldsymbol{X}$）——归结为验证以下条件：L_L—

$$\lambda>\left|\boldsymbol{\nu}^\top\boldsymbol{x}_i\right|=\max(\boldsymbol{\nu}^\top\boldsymbol{x}_i,-\boldsymbol{\nu}^\top\boldsymbol{x}_i)\quad\text{s.t.}\quad\boldsymbol{\nu}\in\Theta. \tag{3.59}$$

等价地，这可以表示为：

$$\lambda>\max(S(\gamma,\boldsymbol{x}_i),S(\gamma,-\boldsymbol{x}_i)),$$

其中，$S(\gamma,\boldsymbol{x}_i)$ 表示以下带有约束条件 $\boldsymbol{\nu}\in\Theta_1$ 和 $\boldsymbol{\nu}\in\Theta_2$的凸优化问题的最优值：

$$S(\gamma,\boldsymbol{x}_i)\triangleq\max_{\boldsymbol{\nu}}\left\{\boldsymbol{x}_i^\top\boldsymbol{\nu}\quad\text{s.t.}\quad D(\boldsymbol{\nu})\geq\gamma,\ \boldsymbol{g}^\top(\boldsymbol{\nu}-\widehat{\boldsymbol{\nu}}_0)\geq 0\right\}. \tag{3.60}$$

值得注意的是，这个问题有闭式解：

$$S(\gamma,\boldsymbol{x}_i)=\begin{cases}\widehat{\boldsymbol{\nu}}_0{}^\top\boldsymbol{x}_i+A_iB, & \text{if }\ \|\boldsymbol{g}\|_2^2\|\boldsymbol{x}_i\|_2<C\boldsymbol{x}_i^\top\boldsymbol{g};\\ \boldsymbol{y}^\top\boldsymbol{x}_i-\|\boldsymbol{x}_i\|_2C, & \text{if }\ \|\boldsymbol{g}\|_2^2\|\boldsymbol{x}_i\|_2\geq C\boldsymbol{x}_i^\top\boldsymbol{g},\end{cases} \tag{3.61}$$

其中 $A_i\triangleq(\|\boldsymbol{x}_i\|_2^2-(\boldsymbol{x}_i^\top\boldsymbol{g})^2/\|\boldsymbol{g}\|_2^2)^{1/2}$，$B\triangleq(C^2-\|\boldsymbol{g}\|_2^2)^{1/2}$，以及 $C\triangleq(\|\boldsymbol{y}\|_2^2-2\gamma)^{1/2}$。**证明** [（3.61）] 我们通过构造问题（3.60）的拉格朗日对偶来推导该结果。为两个约束条件引入非负对偶变量 μ_1 和 μ_2，得到：

$$\begin{aligned}S(\gamma,\boldsymbol{x}_i)&=\max_{\mu_1,\mu_2\geq0}L(\mu_1,\mu_2)\triangleq\max_{\mu_1,\mu_2\geq0}\min_{\boldsymbol{\nu}}\boldsymbol{x}_i^\top\boldsymbol{\nu}-\mu_1(D(\boldsymbol{\nu})-\gamma)-\mu_2\boldsymbol{g}^\top(\boldsymbol{\nu}-\widehat{\boldsymbol{\nu}}_0)\\&=\max_{\mu_1,\mu_2\geq0}\left\{\{\mu_1\gamma+\mu_2\boldsymbol{g}^\top\widehat{\boldsymbol{\nu}}_0\}+\min_{\boldsymbol{\nu}}\{\boldsymbol{x}_i^\top\boldsymbol{\nu}-\mu_1D(\boldsymbol{\nu})-\mu_2\boldsymbol{g}^\top\boldsymbol{\nu}\}\right\}\\&=\max_{\mu_1,\mu_2\geq0}\left\{\{\mu_1\gamma+\mu_2\boldsymbol{g}^\top\widehat{\boldsymbol{\nu}}_0\}+\mu_1\min_{\boldsymbol{\nu}}\left(\frac{\boldsymbol{x}_i^\top-\mu_1\boldsymbol{y}^\top-\mu_2\boldsymbol{g}^\top}{\mu_1}\boldsymbol{\nu}+\frac{1}{2}\|\boldsymbol{\nu}\|_2^2\right)\right\}.\end{aligned}$$

因此，

$$S(\gamma, \boldsymbol{x}_i) = \max_{\mu_1, \mu_2 \geq 0} L(\mu_1, \mu_2),$$

其中，根据 $\boldsymbol{g} = \boldsymbol{y} - \widehat{\boldsymbol{\nu}}_0$，我们有

$$L(\mu_1, \mu_2) = \boldsymbol{x}_i^\top \boldsymbol{y} - \frac{\mu_1}{2} C^2 - \frac{1}{2\mu_1} \|\boldsymbol{x}_i\|_2^2 - \frac{\mu_2^2}{2\mu_1} \|\boldsymbol{g}\|_2^2 + \frac{\mu_2}{\mu_1} \boldsymbol{x}_i^\top \boldsymbol{g} - \mu_2 \|\boldsymbol{g}\|_2^2, \tag{3.62}$$

且 $C \triangleq (\|\boldsymbol{y}\|_2^2 - 2\gamma)^{1/2}$。为使 $L(\mu_1, \mu_2)$ 最大化，我们首先对 (3.62) 关于 μ_2 求导，并令其等于零：

$$-\mu_2 \|\boldsymbol{g}\|_2^2 + \boldsymbol{x}_i^\top \boldsymbol{g} - \mu_1 \|\boldsymbol{g}\|_2^2 = 0.$$

由于 $\mu_2 \geq 0$，我们得到阈值条件：$\mu_2 = \max\{0, \boldsymbol{x}_i^\top \boldsymbol{g}/\|\boldsymbol{g}\|_2^2 - \mu_1\}$。这引出两种情况。当 $\mu_1 \geq \boldsymbol{x}_i^\top \boldsymbol{g}/\|\boldsymbol{g}\|_2^2$ 时，我们有 $\mu_2 = 0$, $\mu_1 = \|\boldsymbol{x}_i\|_2/C$，因此，$S(\gamma, \boldsymbol{x}_i)$ 的取值为：

$$S(\gamma, \boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{y} - \|\boldsymbol{x}_i\|_2 C.$$

另一方面，当 $\mu_1 \leq \boldsymbol{x}_i^\top \boldsymbol{g}/\|\boldsymbol{g}\|_2^2$ 时，我们有 $\mu_2 = \boldsymbol{x}_i^\top \boldsymbol{g}/\|\boldsymbol{g}\|_2^2 - \mu_1$。将其代入 $L(\mu_1, \mu_2)$，并对 μ_1 求导，可得最优性条件：

$$B^2 \mu_1^2 = A_i^2,$$

其中 $A_i \triangleq (\|\boldsymbol{x}_i\|_2^2 - (\boldsymbol{x}_i^\top \boldsymbol{g})^2/\|\boldsymbol{g}\|_2^2)^{1/2}$ 且 $B \triangleq (C^2 - \|\boldsymbol{g}\|_2^2)^{1/2}$。将 μ_1 和 μ_2代入（3.62），$S(\gamma, \boldsymbol{x}_i)$ 的取值为：$S(\gamma, \boldsymbol{x}_i) = \widehat{\boldsymbol{\nu}}_0^\top \boldsymbol{x}_i + A_i B$. ■

上述分析直接导出了 LASSO 的 SAFE规则。

定理3.15：（SAFE–LASSO（Ghaoui 等人，2010年））考虑带惩罚参数 λ的 LASSO 问题（L_λ）。设 $\lambda_0 \geq \lambda$ 是一个正则化参数，对应的最优原始解 $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ 已知。记 $\boldsymbol{x}_i$ 为设计矩阵 $\boldsymbol{X}$的第 i列。定义指标集

$$\mathbb{I} = \{i \mid \lambda > \max(S(\gamma, \boldsymbol{x}_i), S(\gamma, -\boldsymbol{x}_i))\},$$

其中，函数 $S(\gamma, \boldsymbol{x}_i)$ 定义为

$$S(\gamma, \boldsymbol{x}_i) = \begin{cases} \widehat{\boldsymbol{\nu}}_0{}^\top \boldsymbol{x}_i + A_i B(\gamma), & \text{if } \|\boldsymbol{g}\|_2^2 \|\boldsymbol{x}_i\|_2 < C(\gamma) \boldsymbol{x}_i^\top \boldsymbol{g}; \\ \boldsymbol{y}^\top \boldsymbol{x}_i - \|\boldsymbol{x}_i\|_2 C(\gamma), & \text{if } \|\boldsymbol{g}\|_2^2 \|\boldsymbol{x}_i\|_2 \geq C(\gamma) \boldsymbol{x}_i^\top \boldsymbol{g}, \end{cases}$$

with

$$\widehat{\boldsymbol{\nu}}_0 = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0, \quad \boldsymbol{g} \triangleq \boldsymbol{y} - \widehat{\boldsymbol{\nu}}_0, \quad a_0 \triangleq \|\widehat{\boldsymbol{\nu}}_0\|_2^2,$$

$$b_0 \triangleq \left|\boldsymbol{y}^\top \widehat{\boldsymbol{\nu}}_0\right|, \quad \gamma \triangleq -\frac{a_0}{2}\left(\frac{\lambda}{\lambda_0}\right)^2 + b_0 \frac{\lambda}{\lambda_0},$$

$$C(\gamma) = \left(\|\boldsymbol{y}\|_2^2 - 2\gamma\right)^{1/2}, \quad B(\gamma) = \left(C(\gamma)^2 - \|\boldsymbol{g}\|_2^2\right)^{1/2}, \quad A_i \triangleq \left(\|\boldsymbol{x}_i\|_2^2 - \frac{(\boldsymbol{x}_i^\top \boldsymbol{g})^2}{\|\boldsymbol{g}\|_2^2}\right)^{1/2}.$$

那么，对于每个索引 $i \in \mathbb{I}$，其对应的元素 $\boldsymbol{\beta}$ 均为零：$\beta_i = 0$。因此，可以安全地从 $\boldsymbol{x}_i$ 中移除特征 $\boldsymbol{X}$，然后再求解完整LASSO问题（L_λ）。

在实践中，如果没有可用的先前解 $\widehat{\boldsymbol{\beta}}_0$，我们可以使用平凡解 $\widehat{\boldsymbol{\beta}}_0 = \mathbf{0}$进行初始化，该解对应于最大的正则化参数 $\lambda_0 = \lambda_{\max} \triangleq \left\|\boldsymbol{X}^\top \boldsymbol{y}\right\|_\infty$。在这种情况下，SAFE 检验会大幅简化。具体而言，条件$\lambda > \max(S(\gamma, \boldsymbol{x}_i)\ S(\gamma, -\boldsymbol{x}_i))$ 可化简为 $\lambda > \rho_i \lambda_{\max}$，其中

$$\rho_i \triangleq \frac{\|\boldsymbol{y}\|_2 \|\boldsymbol{x}_i\|_2 + \left|\boldsymbol{y}^\top \boldsymbol{x}_i\right|}{\|\boldsymbol{y}\|_2 \|\boldsymbol{x}_i\|_2 + \lambda_{\max}}. \tag{3.63}$$

对于缩放数据集，其中响应和所有特征均经过归一化，使得 $\|\boldsymbol{y}\|_2 =1$ 且对每个 i 都成立，量 ρ_i 具有清晰的几何解释：

$$\rho_i = \frac{1 + |\cos \alpha_i|}{1 + \max\limits_{1 \le j \le p} |\cos \alpha_j|}, \tag{3.64}$$

其中 α_i 表示 i第 $\boldsymbol{x}_i$ 个特征与响应向量 $\boldsymbol{y}$之间的角度。因此，SAFE 检验根据特征与响应的相对对齐程度来剔除特征——具体而言，是比较它们与对齐程度最高的特征有多接近。在这种情况下，SAFE规则与标准的基于相关性的特征选择方法十分相似（Ghaoui 等人，2010年）。

3.4. ℓ_1最小化问题的最优性条件

回顾一下，给定测量值 $\boldsymbol{y}$，并假设原始信号 $\boldsymbol{\beta}$ 是稀疏或可压缩的，恢复 $\boldsymbol{\beta}$ 的一种自然方法是求解如下形式的优化问题：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}), \tag{3.65}$$

其中 $\mathcal{B}(\boldsymbol{y})$ 确保重建信号 $\boldsymbol{\beta}$ 与测量值$\boldsymbol{y}$一致。例如，如果测量值精确且无噪声，我们可以定义 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{\beta} \,|\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}\}$。存在小幅有界噪声时，我们也可以改用 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{\beta} \mid \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \le \epsilon\}$。在这两种情况下，(3.65) 都旨在求得满足测量约束条件的最稀疏解 $\boldsymbol{\beta}$。

为了使问题更易处理，一种常见策略是用凸替代 ℓ_1-范数替换 ℓ_0-范数。这会得到如下放宽后的优化问题：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}). \tag{3.66}$$

只要 $\mathcal{B}(\boldsymbol{y})$ 是凸的，3.66) 就成为一个凸优化问题，并且可以高效求解。

利用 ℓ_1-最小化来促进稀疏性由来已久。至少可以追溯到 Beurling 关于从部分观测进行傅里叶变换外推的研究（Beurling，1938）。在不同的背景下，Logan(1965年) 于 1965 年证明，即使带限信号在一个小区间内受到破坏，仍然可以完美恢复（后来 Donoho 和 Logan (1992年) 对此进行了扩展）。他的方法是寻找与观测数据最接近的带限信号，采用的度量是 ℓ_1-范数——这进一步印证了 ℓ_1 非常适合处理稀疏误差这一观点。

因此，有多种充分理由可以预期 ℓ_1-最小化能够为稀疏信号恢复提供精确解。此外，它还为难以求解的 ℓ_0-最小化问题提供了一种计算高效的替代方案。本节将分析 ℓ_1-最小化的最优性，并在第10章随后讨论求解该问题的算法。

3.4.1 充要条件

具体而言，现在考虑系统 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 以及 ℓ_1-最小化问题（P_1）（第129页）：

$$(\mathrm{P}_1) \quad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{3.67}$$

向量 $\widehat{\boldsymbol{\beta}}$ 是（P_1）的解，当且仅当：

$$\|\widehat{\boldsymbol{\beta}} + \boldsymbol{n}\|_1 \geq \|\widehat{\boldsymbol{\beta}}\|_1, \quad \forall\, \boldsymbol{n} \in \mathbb{R}^p \quad \text{s.t.} \quad \boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}) \quad \text{and} \quad \boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}},$$

其中，$\{i \mid \widehat{\beta}_i \neq 0\}$ $|\widehat{\beta}_i + n_i| \geq |\widehat{\beta}_i| +$ $\widehat{\beta}_i\ n_i$ 属于 $\boldsymbol{X}$的零空间，即 $\boldsymbol{X}\boldsymbol{n} = \boldsymbol{0}$。鉴于 $\widehat{\boldsymbol{\beta}}$支撑于 $\mathbb{S} \triangleq \mathrm{supp}(\widehat{\boldsymbol{\beta}})$，且不等式符号为 sign( )，我们有

$$\left\|\widehat{\boldsymbol{\beta}} + \boldsymbol{n}\right\|_1 = \sum_{i\in\mathbb{S}} \left|\widehat{\beta}_i + n_i\right| + \sum_{i\in\mathbb{S}^c} |n_i| \geq \sum_{i\in\mathbb{S}} \left|\widehat{\beta}_i\right| + \sum_{i\in\mathbb{S}} \mathrm{sign}(\widehat{\beta}_i) n_i + \sum_{i\in\mathbb{S}^c} |n_i|\,.$$

该分析表明了（P_1）的充要条件。

定理 3.16：（（P_1）**的充要条件**）考虑优化问题（P_1）。那么，$\boldsymbol{\beta}$ 是（P_1）的解，当且仅当对于所有 $\boldsymbol{n} \in \mathcal{N}$（$\boldsymbol{X}$），都有

$$\sum_{i\in\mathbb{S}} \mathrm{sign}(\widehat{\beta}_i) n_i \leq \sum_{i\in\mathbb{S}^c} |n_i|\,,$$

其中 $\mathbb{S} \triangleq$ 支撑集$(\widehat{\boldsymbol{\beta}}) = \{i \mid \widehat{\beta}_i \neq 0, i \in \{1, 2, \ldots, p\}\}$。

证明 [定理 3。16] **必要性**。设 $\boldsymbol{\beta}$ 是（P_1 $\gamma > 0$）的一个解。选取充分小，使得每个非零分量的符号函数 β_i 在 $\beta_i - \gamma n_i$中保持不变。于是对于 $i \in \{1, 2, \ldots, p\}$，

$$\left|\widehat{\beta}_i - \gamma n_i\right| = \begin{cases} \widehat{\beta}_i - \gamma n_i = \widehat{\beta}_i - \gamma \mathrm{sign}(\widehat{\beta}_i) n_i, & \text{if } \widehat{\beta}_i > 0; \\ -(\widehat{\beta}_i - \gamma n_i) = -\widehat{\beta}_i - \gamma \mathrm{sign}(\widehat{\beta}_i) n_i, & \text{if } \widehat{\beta}_i < 0; \\ \gamma\,|n_i|\,, & \text{otherwise,} \end{cases} \qquad \forall\, i \in \{1, 2, \ldots, p\}.$$

现在反设存在 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，使得 $\sum_{i\in\mathbb{S}}$ 符号函数$(\beta_i) n_i > \sum_{i\in\mathbb{S}^c} |n_i|$。这意味着

$$\left\|\widehat{\boldsymbol{\beta}} - \gamma\boldsymbol{n}\right\|_1 = \|\widehat{\boldsymbol{\beta}}\|_1 - \gamma \sum_{i\in\mathbb{S}} \mathrm{sign}(\widehat{\beta}_i) n_i + \gamma \sum_{i\in\mathbb{S}^c} |n_i| < \|\widehat{\boldsymbol{\beta}}\|_1,$$

从而导致矛盾。因此，所述不等式必然成立。

充分性。设存在另一个可行点 $\boldsymbol{\beta}'$，使得 $\boldsymbol{X}\boldsymbol{\beta}' = \boldsymbol{y}$ 且 $\|\boldsymbol{\beta}'\|_1 < \|\widehat{\boldsymbol{\beta}}\|_1$。定义 $\boldsymbol{n} \triangleq \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}'$，由于 $\boldsymbol{X}\boldsymbol{\beta}' = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$，其满足 $\boldsymbol{X}\boldsymbol{n} = \boldsymbol{0}$。这意味着$\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$。利用单位矩阵 $\|\boldsymbol{\beta}'\|_1 = \|\widehat{\boldsymbol{\beta}} - \boldsymbol{n}\|_1$，我们进行如下分析：

$$\begin{aligned} \|\boldsymbol{\beta}'\|_1 = \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{n}\right\|_1 &= \sum_{i\in\mathbb{S}} \left|\widehat{\beta}_i - n_i\right| + \sum_{i\in\mathbb{S}^c} |n_i| \\ &\geq \sum_{i\in\mathbb{S}} |\widehat{\beta}_i| - \sum_{i\in\mathbb{S}} \mathrm{sign}(\widehat{\beta}_i) n_i + \sum_{i\in\mathbb{S}^c} |n_i| \geq \sum_{i\in\mathbb{S}} \left|\widehat{\beta}_i\right| = \|\widehat{\boldsymbol{\beta}}\|_1, \end{aligned}$$

其中，第一个不等式由以下事实得出：$\left|\widehat{\beta}_i - n_i\right| \geq \left|\widehat{\beta}_i\right| -$符号函数$(\widehat{\beta}_i) n_i$；而第二个不等式则由给定条件 $\sum_{i\in\mathbb{S}}$符号函数$(\beta_i) n_i \leq \sum_{i\in\mathbb{S}^c} |n_i|$得出。由于假设为 $\|\boldsymbol{\beta}'\|_1 < \|\widehat{\boldsymbol{\beta}}\|_1$，我们得到了矛盾。证明完毕。 ■

3.4.2 对偶证书与唯一性

我们在例 ℓ_∞-最小化问题中引入了2.117的对偶问题。令 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。该优化问题的对偶问题为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_\infty \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \tag{3.68a}$$

is

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle \boldsymbol{y}, \boldsymbol{\lambda}\rangle \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_1 \le 1. \tag{3.68b}$$

ℓ_1 最小化问题的对偶问题可以用类似的方法推导。

命题 3.17：（ℓ_1-最小化问题的对偶）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。优化问题（$\mathrm{P}_1$）的对偶问题为

$$(\mathrm{P}_1) \quad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \tag{3.69a}$$

is

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} D(\boldsymbol{\lambda}) = \begin{cases} \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle, & \text{if } \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \le 1; \\ -\infty, & \text{otherwise.} \end{cases} \tag{3.69b}$$

显然，只需在满足 $\boldsymbol{\lambda}$ 的点 $D(\boldsymbol{\lambda}) > -\infty$上取最大值。将该约束明确写出，问题（3.69a）的对偶问题为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \le 1. \tag{3.69c}$$

命题 [的证明：3.17] 与（P_1）相关的拉格朗日函数为

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \|\boldsymbol{\beta}\|_1 + \langle \boldsymbol{\lambda}, \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\rangle .$$

因此，拉格朗日对偶函数为

$$D(\boldsymbol{\lambda}) \triangleq \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 - \left\langle \boldsymbol{X}^\top\boldsymbol{\lambda}, \boldsymbol{\beta}\right\rangle + \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle .$$

如果（$\boldsymbol{X}^\top\boldsymbol{\lambda}$）$_i > 1$，那么可以将 β_i 设为任意大，使得 D（$\boldsymbol{\lambda}$）$\to -\infty$。如果（$\boldsymbol{X}^\top\boldsymbol{\lambda}$）$_i < -1$，同样如此。如果 $\left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \le 1$，根据 H¨older 不等式（定理 1.60）

$$\left|(\boldsymbol{X}^\top\boldsymbol{\lambda})^\top\boldsymbol{\beta}\right| \le \|\boldsymbol{\beta}\|_1 \cdot \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \le \|\boldsymbol{\beta}\|_1 ,$$

因此，将 $\boldsymbol{\beta}$ 设为零并令 $D(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle$，即可使拉格朗日函数达到最小值。证明完毕。 ■

根据定理 2.114，对于这对原问题和对偶优化问题，只要原问题（3.69a）是可行的，强对偶性就成立。

有趣的是，ℓ_1-最小化问题的对偶问题的解通常能为原始最优解的支撑集提供有价值的信息。设$\widehat{\ }$原始最优解 $\boldsymbol{\beta}$ 支撑于

$$\mathbb{S} \triangleq \operatorname{supp}(\widehat{\boldsymbol{\beta}}) = \{i \mid \widehat{\beta}_i \neq 0\}.$$

注意，$\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}$，其中 $\widehat{\boldsymbol{\beta}}_{\mathbb{S}} \triangleq \widehat{\boldsymbol{\beta}}[\mathbb{S}] \in \mathbb{R}^{|\mathbb{S}|}$ 表示由 $\mathbb{S}$索引的 $\widehat{\boldsymbol{\beta}}$的子向量，而$\boldsymbol{X}_{\mathbb{S}} \triangleq \boldsymbol{X}[:, \mathbb{S}] \in \mathbb{R}^{n\times|\mathbb{S}|}$ 表示由 $\mathbb{S}$索引的列组成的 $\boldsymbol{X}$ 的子矩阵。

引理3.18：（使用对偶最优解确定原始最优解的支撑集）假设原问题（3.69a）是可行的。设 $\widehat{\boldsymbol{\lambda}}$ 是对偶问题（3.69c）的一个最优解。那么，对于每个原始最优解 $\widehat{\boldsymbol{\beta}}$，

$$(\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}})_i = \operatorname{sign}(\widehat{\beta}_i), \quad \text{for all } \widehat{\beta}_i \neq 0 \text{ (i.e., } i \in \mathbb{S}). \tag{3.70}$$

证明 [引理 3.18] 原问题的可行性意味着强对偶性（由定理 2.114）。因此，

$$\|\widehat{\boldsymbol{\beta}}\|_1 = \boldsymbol{y}^\top\widehat{\boldsymbol{\lambda}} = (\boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}}) = \sum_{i=1}^{p}(\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}})_i\widehat{\beta}_i.$$

另一方面，根据 Hölder 不等式，我们有

$$\sum_{i=1}^{p}(\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}})_i\widehat{\beta}_i \leq \|\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}}\|_\infty\|\widehat{\boldsymbol{\beta}}\|_1 \leq \|\widehat{\boldsymbol{\beta}}\|_1,$$

当且仅当对所有 $\widehat{\beta}_i \neq 0.$，$(\boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}})_i =$符号函数$(\widehat{\beta}_i)$时，等号成立。这就完成了证明。 ■

在稀疏恢复或 ℓ_1-正则化优化的研究中，一个核心问题是 ℓ_1-最小化问题（P_1）的最优解是否 **唯**一。由于凸性，在温和条件下可以保证解的存在；然而，解的唯一性往往取决于额外的结构性假设，尤其是与数据矩阵的几何性质 $\boldsymbol{X}$ 以及解的支撑集有关的假设。前文的分析建立了必要最优性条件，揭示了最优解 $\boldsymbol{\beta}$ 满足涉及拉格朗日乘子的某些符号和对偶可行性条件。特别是，向量$\boldsymbol{X}^\top\boldsymbol{\lambda}$ 在刻画 $\widehat{\boldsymbol{\beta}}$的支撑集和稀疏模式方面起着关键作用。基于这些认识，我们现在转而探讨一个重要问题：在什么条件下，这个最优解能够被唯一确定？由此引出以下关键结果：该结果基于 ℓ_1-最小化问题（P_1）的 对偶证书 的性质，给出了其解唯一性的充分条件，并阐明了设计矩阵的 受限等距性质（*RIP*）（第4章和第5章）。

考虑一个稀疏向量 $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^p$ 其支撑集为 $\mathbb{S}$，满足 $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}$。如果存在一个向量 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 满足 $\boldsymbol{X}^\top\boldsymbol{\lambda}$ 等于 $\widehat{\boldsymbol{\beta}}$ 在 $\mathbb{S}$ 上的符号函数，并且在其他位置的大小小于 1，那么 $\boldsymbol{\lambda}$ 对于对偶问题（3.69b）是可行的，因此根据弱对偶性，对于任意 $\|\boldsymbol{\beta}\|_1 \geq \boldsymbol{y}^\top\boldsymbol{\lambda}$ 满足原始问题（3.69a）可行的 $\boldsymbol{\beta} \in \mathbb{R}^p$，由引理 3.18，我们有

$$\|\boldsymbol{\beta}\|_1 \geq \boldsymbol{y}^\top\boldsymbol{\lambda} = (\boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top\boldsymbol{\lambda} = (\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{X}^\top\boldsymbol{\lambda}) = \sum_{i=1}^{p}\widehat{\beta}_i\operatorname{sign}(\widehat{\beta}_i) = \|\widehat{\boldsymbol{\beta}}\|_1. \tag{3.71}$$

由于对于所有 $\widehat{\beta}_i \neq 0$，$(\boldsymbol{X}^\top\boldsymbol{\lambda})_i = \operatorname{sign}(\widehat{\beta}_i)$，从几何上看，$\boldsymbol{X}^\top\boldsymbol{\lambda}$ 是 ℓ_1-范数在 $\boldsymbol{\beta}$处的次梯度。该次梯度与由 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 给出的可行性超平面正交（超平面内的任意向量 $\boldsymbol{n}$ 都是两个可行向量之差，因此满足 $\boldsymbol{X}\boldsymbol{n} = \boldsymbol{0}$）。因此，对于任意其他可行向量 $\boldsymbol{\beta}$

$$\|\boldsymbol{\beta}\|_1 \geq \|\widehat{\boldsymbol{\beta}}\|_1 + (\boldsymbol{X}^\top\boldsymbol{\lambda})^\top(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) = \|\widehat{\boldsymbol{\beta}}\|_1 + \boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \|\widehat{\boldsymbol{\beta}}\|_1, \tag{3.72}$$

其中第一个不等式由次梯度不等式（定义 2.57）推出。

或者，我们在备注2.116中讨论了带等式约束的凸优化问题的 KKT 条件；这是凸优化问题的 Slater 条件以及强对偶性下的 KKT 条件（定理 2.114 和 2.115）的结果。由于 $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ 是凸的，备注 2.116 和方程（3.31）表明，当 $\boldsymbol{\beta}$ 是最优解时，条件是 $\widehat{\boldsymbol{\beta}}$ 可行且存在 $\boldsymbol{\alpha} = \boldsymbol{X}^\top \boldsymbol{\lambda}$，其中 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 表示拉格朗日乘子（即 $\boldsymbol{\alpha}$ 属于 $\boldsymbol{X}$的行空间，并且 $\boldsymbol{\alpha} \perp \mathcal{N}(\boldsymbol{X})$）满足

$$\begin{cases} \alpha_i = \operatorname{sign}(\widehat{\beta}_i), & \widehat{\beta}_i \neq 0 \text{ (i.e., } i \in \mathbb{S}); \\ |\alpha_i| \leq 1, & \widehat{\beta}_i = 0 \text{ (i.e., } i \in \mathbb{S}^c), \end{cases} \tag{3.73}$$

其中，$\widehat{\boldsymbol{\beta}}$ 支撑于 $\mathbb{S} \triangleq \operatorname{supp}(\widehat{\boldsymbol{\beta}}) = \{i \mid \widehat{\beta}_i \neq 0\}$。

这两个论证表明，可以利用某个对偶向量的存在性来验证给定的原始可行向量是否为最优解，但它们不能保证唯一性。事实证明，只要要求 $\boldsymbol{X}^\top \boldsymbol{\lambda}$ 在 $\mathbb{S}^c$ 上的大小严格小于 1，就足以保证唯一性（前提是 $\boldsymbol{X}_{\mathbb{S}}$ 满秩）。在这种情况下，我们称对偶变量 $\boldsymbol{\lambda}$ 为 ℓ_1 最小化问题的对偶证书 。

定理 3.19：（(P_1)的对偶证书）考虑 ℓ_1最小化问题（P_1）。设 $\widehat{\boldsymbol{\beta}}$ 是（P_1）的一个最优解，且 $\boldsymbol{\alpha} = \boldsymbol{X}^\top \boldsymbol{\lambda}$ 满足以下条件：进一步假设

(i) $\alpha_i = (\boldsymbol{X}^\top \boldsymbol{\lambda})_i = \text{符号函数}(\widehat{\beta}_i)$，当 $\widehat{\beta}_i \neq 0$ 对所有 $i \in \{1, 2, \ldots, p\}$ 均成立。
(ii) $|\alpha_i| = |(\boldsymbol{X}^\top \boldsymbol{\lambda})_i| < 1$，当 $\widehat{\beta}_i = 0$ 对所有 $i \in \{1, 2, \ldots, p\}$均成立。
(iii) $\boldsymbol{X}_{\mathbb{S}}$ 具有满列秩 [a]，其中 $\mathbb{S} \triangleq$ 支撑集$(\widehat{\boldsymbol{\beta}}) = \{i \mid \widehat{\beta}_i \neq 0\}$。

那么 $\widehat{\boldsymbol{\beta}}$ 就是**方程**（P_1）的唯一解。任何满足（i）–（iii）的此类 $\boldsymbol{\alpha}$ 或 $\boldsymbol{\lambda}$ 均称为对偶证书。

[a]。这也可由 RIP 条件推出；参见第 4 章和引理 5.12。

证明 [定理319] 任取 $\boldsymbol{\beta} \in \mathbb{R}^p$ 为一个可行点，并定义 $\boldsymbol{n} \triangleq \boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}$。由于$\boldsymbol{X}_{\mathbb{S}}$ 满列秩，因此 $\boldsymbol{n}_{\mathbb{S}^c} \neq \boldsymbol{0}$ 除非 $\boldsymbol{n} = \boldsymbol{0}$，否则 $\boldsymbol{n}_{\mathbb{S}}$ 将是 $\boldsymbol{X}_{\mathbb{S}}$零空间中的一个非零向量。条件（ii）蕴含

$$\|\boldsymbol{n}_{\mathbb{S}^c}\|_1 > \boldsymbol{\alpha}^\top \boldsymbol{n}(\mathbb{S}^c),$$

其中 $\boldsymbol{n}(\mathbb{S}^c) \in \mathbb{R}^p$是一个投影，它将向量中除索引属于 $\mathbb{S}^c$的元素之外的所有元素都置为零（见定义1.1）。利用 $|\widehat{\beta}_i + n_i| \geq |\widehat{\beta}_i| +$符号函数$(\widehat{\beta}_i) n_i$ 以及 $\boldsymbol{\beta}$ 支撑于 $\mathbb{S}$这一事实，上述不等式蕴含

$$\|\boldsymbol{\beta}\|_1 = \left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} + \boldsymbol{n}_{\mathbb{S}}\right\|_1 + \|\boldsymbol{n}_{\mathbb{S}^c}\|_1 > \left\|\widehat{\boldsymbol{\beta}}\right\|_1 + \boldsymbol{\alpha}^\top \boldsymbol{n}(\mathbb{S}) + \boldsymbol{\alpha}^\top \boldsymbol{n}(\mathbb{S}^c) = \left\|\widehat{\boldsymbol{\beta}}\right\|_1 + \boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{n} = \left\|\widehat{\boldsymbol{\beta}}\right\|_1.$$

另一方面，如果 $\boldsymbol{n}_{\mathbb{S}^c} = \boldsymbol{0}$，那么由于 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩，

$$\boldsymbol{X}\boldsymbol{n} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{n}_{\mathbb{S}} = \boldsymbol{0} \quad \Longrightarrow \quad \boldsymbol{n}_{\mathbb{S}} = \boldsymbol{0} \quad \Longrightarrow \quad \boldsymbol{n} = \boldsymbol{0}.$$

因此，对于任意的 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，除非 $\boldsymbol{n} = \boldsymbol{0}$，否则 $\left\|\widehat{\boldsymbol{\beta}} + \boldsymbol{n}\right\|_1 > \left\|\widehat{\boldsymbol{\beta}}\right\|_1$。证明完毕。
■

请注意，理想情况下，我们希望分析满足这个欠定系统的向量 $\boldsymbol{\lambda}$，该系统包含 $|\mathbb{S}|$ 个方程，并且 $\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda}$ 的 ℓ_∞-范数尽可能小。遗憾的是，这个优化问题的解

$$\min_{\boldsymbol{\lambda}} \left\|\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda}\right\|_\infty \quad \text{s.t.} \quad \boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda} = \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) \tag{3.74}$$

不存在闭式解。不过，相关的最小二乘问题

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_2 \quad \text{s.t.} \quad \boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda} = \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) \tag{3.75}$$

确实有一个显式解，因此我们可以改为分析它。注意，由于 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩（$\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}$ 可逆），解为

$$\boldsymbol{\lambda} = \boldsymbol{X}_{\mathbb{S}} (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1} \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) \qquad \text{and} \qquad \boldsymbol{\alpha} \triangleq \boldsymbol{X}^\top \boldsymbol{\lambda}, \tag{3.76}$$

这是一个高维最小二乘问题（推论 3.7），形式为 $\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda} =$ 符号函数$(\widehat{\boldsymbol{\beta}}_{\mathbb{S}})$。因此，

- 如果对所有 $i \in \mathbb{S}^c$ 都有 $|\alpha_i| \le 1$，那么 $\boldsymbol{\beta}$ 是一个解；
- 如果对所有 $i \in \mathbb{S}^c$ 都有 $|\alpha_i| < 1$，那么 $\boldsymbol{\beta}$ 是唯一解。

对于给定的系数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，定理 3.19 中的条件可改述如下：指标集 $\mathbb{S} \triangleq$ 支撑集$(\widehat{\boldsymbol{\beta}})$ 和 $\mathbb{T} \triangleq \mathbb{S}^c$ 满足：

(i) $\mathcal{N}(\boldsymbol{X}(\mathbb{S})) = \{\boldsymbol{0}\}$：这保证线性方程组 $\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}} = \boldsymbol{y}$ 有唯一解。(ii) 存在一个向量 $\boldsymbol{\alpha} \in \mathbb{R}^p$，使得 $\boldsymbol{\alpha} \in \mathcal{C}(\boldsymbol{X}^\top)$、$\boldsymbol{\alpha}_{\mathbb{S}} =$ 符号函数$(\widehat{\boldsymbol{\beta}}_{\mathbb{S}})$，且 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty < 1$：这确保任何最优解都必须满足 $\boldsymbol{\beta}_{\mathbb{T}} = \boldsymbol{0}$。

事实上，受限等距性质（*RIP*）保证对于所有足够小的支撑集，$\mathbb{S}$ 以及任意符号模式，这些条件都成立；详见定义4.30 和第5.3 节，了解更多详情。

因此，定理3.19中的论证可以自然地分为两部分。

► $\mathcal{N}(\boldsymbol{X}(\mathbb{S})) = \{\boldsymbol{0}\}$ 条件 (i) $\boldsymbol{0} \neq \boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}(\mathbb{S}))$[3] $\boldsymbol{\beta}_\eta \triangleq \widehat{\boldsymbol{\beta}} + \eta\widetilde{\boldsymbol{n}}$ η $\widetilde{\boldsymbol{n}}_{\mathbb{S}} = \boldsymbol{n}_{\mathbb{S}}$ (()) **是唯一性的必要条件**。若 (( )) ，则对任意标量 ，所有 都是最优的，其中 且 $\widetilde{\boldsymbol{n}}_{\mathbb{S}^c} = \boldsymbol{0}$。为此，我们注意到 $\boldsymbol{\beta}_\eta$ 是可行的，因为 $\boldsymbol{X}\boldsymbol{\beta}_\eta = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$。由于$\widehat{\boldsymbol{\beta}}$ 是唯一的极小化解，因此 $\|\boldsymbol{\beta}_\eta\|_1 > \|\widehat{\boldsymbol{\beta}}\|_1$。但对于 0 附近足够小的 η，即$|\eta| < \min_{i\in\mathbb{S}} |\widehat{\beta}_i| / \|\boldsymbol{n}\|_\infty$，可知 sign$(\widehat{\boldsymbol{\beta}} + \eta\widetilde{\boldsymbol{n}}) =$sign$(\widehat{\boldsymbol{\beta}})$，由此可得

$$\|\boldsymbol{\beta}_\eta\|_1 = \left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} + \eta\boldsymbol{n}_{\mathbb{S}}\right\|_1 = \left\langle \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}), (\widehat{\boldsymbol{\beta}}_{\mathbb{S}} + \eta\boldsymbol{n}_{\mathbb{S}}) \right\rangle = \|\widehat{\boldsymbol{\beta}}\|_1 + \left\langle \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}), \eta\boldsymbol{n}_{\mathbb{S}} \right\rangle.$$

这就产生了矛盾，因为我们总能选取一个足够小的 $\eta \neq 0$，使得 $\langle$sign$(\widehat{\boldsymbol{\beta}}_{\mathbb{S}})\ \eta\boldsymbol{n}_{\mathbb{S}}\rangle <$0。因此，$\boldsymbol{X}_{\mathbb{S}}$ 必须具有满列秩。

► **条件（ii）保证** $\boldsymbol{\beta}_{\mathbb{T}} = \boldsymbol{0}$。为说明这一点，我们引入以下优化问题：

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty \quad \text{s.t.} \quad \boldsymbol{\alpha} \in \mathcal{C}(\boldsymbol{X}^\top),\ \boldsymbol{\alpha}_{\mathbb{S}} = \operatorname{sign}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}). \tag{3.77}$$

如果最优目标函数值严格小于 1，那么存在满足条件（ii）的 $\boldsymbol{\alpha}$，因此条件（ii）也是必要的。

设$\boldsymbol{\beta}_{\mathbb{S}}$ 且$\mathbb{R}^p$$\boldsymbol{v}_{\mathbb{T}}$满足$\boldsymbol{0}$,$\boldsymbol{v}_{\mathbb{S}}$并设$\boldsymbol{Q}$sign$(\in \mathbb{R}^{p\times r}$ 包含 $\mathcal{N}(\boldsymbol{X})$ 的相互正交归一基。若 $\boldsymbol{v} \in \mathcal{C}(\boldsymbol{X}^\top)$，则令 $\boldsymbol{\alpha} \triangleq \boldsymbol{v}$，证明即告完成。否则，对某个 $\boldsymbol{u} \in \mathbb{R}^p$，令 $\boldsymbol{\alpha} \triangleq \boldsymbol{v} + \boldsymbol{u}$。那么，以下问题等价

$$\boldsymbol{\alpha} \in \mathcal{C}(\boldsymbol{X}^\top) \iff \boldsymbol{Q}^\top\boldsymbol{\alpha} = \boldsymbol{0} \iff \boldsymbol{Q}^\top\boldsymbol{u} = -\boldsymbol{Q}^\top\boldsymbol{v},$$

3。再次注意，$\boldsymbol{X}(\mathbb{S}) \in \mathbb{R}^{n\times p}$ 和 $\boldsymbol{X}_{\mathbb{S}} = \boldsymbol{X}[:\ \mathbb{S}] \in \mathbb{R}^{n\times|\mathbb{S}|}$；参见定义 1.1。

^其中 $\boldsymbol{u}_{\mathbb{S}} = \mathbf{0}$ 因为 $\boldsymbol{\alpha}_{\mathbb{S}} =$符号函数$(\boldsymbol{\beta}_{\mathbb{S}}) = \boldsymbol{v}_{\mathbb{S}}$。由于 $\boldsymbol{v}_{\mathbb{T}} = \mathbf{0}$，我们还有 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty = \|\boldsymbol{u}_{\mathbb{T}}\|_\infty$。因此，我们可以转换（3.77）并将 $\boldsymbol{\alpha} \in \mathcal{C}(\boldsymbol{X}^\top)$ 重写为以下等价问题：

$$\min_{\boldsymbol{u}\in\mathbb{R}^p} \|\boldsymbol{u}_{\mathbb{T}}\|_\infty \quad \text{s.t.} \quad \boldsymbol{Q}^\top \boldsymbol{u} = -\boldsymbol{Q}^\top \boldsymbol{v},\ \boldsymbol{u}_{\mathbb{S}} = \mathbf{0}. \tag{3.78}$$

若最优目标值严格小于 1，则条件（ii）是必要的。问题 (3.78) 可行，且目标函数值有限。(3.78) 的对偶问题为

$$\max_{\boldsymbol{\gamma}\in\mathbb{R}^p} (\boldsymbol{Q}^\top \boldsymbol{v})^\top \boldsymbol{\gamma} \quad \text{s.t.} \quad \|(\boldsymbol{Q}\boldsymbol{\gamma})_{\mathbb{T}}\|_1 \le 1. \tag{3.79}$$

根据假设，问题 (3.78) 满足强对偶性。因此，若 (3.79) 的最优目标值严格小于 1，则条件（ii）是必要的。

引理 3.20 假设 $\widehat{\boldsymbol{\beta}}$ 是支撑集为 $\mathbb{S} \triangleq \operatorname{supp}(\widehat{\boldsymbol{\beta}})$ 的唯一解，并令 $\mathbb{T} \triangleq \mathbb{S}^c$。则以下原始-对偶问题的最优目标函数值严格小于 1：

$$\begin{aligned} &\min_{\boldsymbol{u}} \|\boldsymbol{u}_{\mathbb{T}}\|_\infty \quad \text{s.t.} \quad \boldsymbol{Q}^\top \boldsymbol{u} = -\boldsymbol{Q}^\top \boldsymbol{v},\ \boldsymbol{u}_{\mathbb{S}} = \mathbf{0}; \\ &\max_{\boldsymbol{\gamma}} (\boldsymbol{Q}^\top \boldsymbol{v})^\top \boldsymbol{\gamma} \quad \text{s.t.} \quad \|(\boldsymbol{Q}\boldsymbol{\gamma})_{\mathbb{T}}\|_1 \le 1. \end{aligned}$$

^**引理** [的证明 3.20] 由于 $\boldsymbol{\beta}$ 是 ℓ_1-最小化问题 (P_1) 的唯一解，

$$\|\widehat{\boldsymbol{\beta}}\|_1 < \left\|\widehat{\boldsymbol{\beta}} + \boldsymbol{n}\right\|_1 = \left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} + \boldsymbol{n}_{\mathbb{S}}\right\|_1 + \|\boldsymbol{n}_{\mathbb{T}}\|_1, \quad \forall\, \boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}.$$

类似于定理 3.16 证明中的论证，这表明 $\langle \boldsymbol{v}_{\mathbb{S}},\ \boldsymbol{n}_{\mathbb{S}} \rangle < \|\boldsymbol{n}_{\mathbb{T}}\|_1$。因此，

- 如果 $\widehat{\boldsymbol{\gamma}} = \mathbf{0}$，则 $\|\widehat{\boldsymbol{u}}_{\mathbb{T}}\|_\infty =(\boldsymbol{Q}^\top \boldsymbol{v})^\top \widehat{\boldsymbol{\gamma}} = \mathbf{0}$。如
- 果 $\widehat{\boldsymbol{\gamma}} \neq \mathbf{0}$，则 $\boldsymbol{n} \triangleq \boldsymbol{Q}\widehat{\boldsymbol{\gamma}} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$ 满足

$$\|\widehat{\boldsymbol{u}}_{\mathbb{T}}\|_\infty = (\boldsymbol{Q}^\top \boldsymbol{v})^\top \widehat{\boldsymbol{\gamma}} = \langle \boldsymbol{v}_{\mathbb{S}}, \boldsymbol{n}_{\mathbb{S}} \rangle < \|\boldsymbol{n}_{\mathbb{T}}\|_1 = \|(\boldsymbol{Q}\widehat{\boldsymbol{\gamma}})_{\mathbb{T}}\|_1 \le 1.$$

在这两种情况下，最优目标值都严格小于 1。 ■

3.4.3 ℓ_1-含噪测量下的最小化

现在考虑观测值受到噪声污染的情况；也就是说，$\boldsymbol{y} =\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$。在这种情况下，我们采用以下两种鲁棒变体来处理 ℓ_1-最小化问题，即 ($\mathrm{P}_{1,\sigma}$) 和 ($\mathrm{P}_{1,\epsilon}$)（第 129 页和第 131 页）：

$$(\mathrm{P}_{1,\sigma}) \qquad \min \|\boldsymbol{\beta}\|_1 + \sigma \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2; \tag{3.80a}$$

$$(\mathrm{P}_{1,\epsilon}) \qquad \min \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \le \epsilon. \tag{3.80b}$$

如推论 3.3 所示，问题（$\mathrm{P}_{1,\ \sigma}$）和（$\mathrm{P}_{1,\ \epsilon}$）在以下意义下是等价的：对于适当选择的 σ 和 ϵ，它们具有相同的解集。下面的引理确立了一个更强的性质：对于每个问题的整个解集，残差范数和 ℓ_1-范数都是常数。

引理 3.21：（关于 (P1, σ) 和 (P1, ϵ)） 设 $\sigma > 0$。假设函数 $\|\boldsymbol{\beta}\|_1 + \sigma\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 在凸集 $\mathbb{C}$ 上为常数。那么，$\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}$和 $\|\boldsymbol{\beta}\|_1$ 在 $\mathbb{C}$上也都是常数。

令 $\mathbb{Q}_\sigma$ 表示问题（P1, σ）的解集。由于 $\|\boldsymbol{\beta}\|_1 + \sigma\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 在 $\mathbb{Q}_\sigma$上为常数，上述结果表明，在问题（P1, σ）中，$\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}$ 和 $\|\boldsymbol{\beta}\|_1$ 在 $\mathbb{Q}_\sigma$ 上为常数。

此外，令 $\mathbb{Q}_\epsilon$ 表示问题（P1, ϵ）的解集。于是，$\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2$和 $\|\boldsymbol{\beta}\|_1$ 在 $\mathbb{Q}_\epsilon$上均为常数

证明 [引理 3。21] 只需考虑凸集包含多个点的情形。设 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2 \in \mathbb{C}$ 其中 $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$。考虑连接 $\boldsymbol{\beta}_1$与 $\boldsymbol{\beta}_2$ $\mathbb{L}$ 的线段。由 $\mathbb{C}$的凸性可知 $\mathbb{L} \subset \mathbb{C}$。因此，$c \triangleq \|\boldsymbol{\beta}\|_1 + \sigma\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 在 $\mathbb{L}$ 上为常数。假设 $\boldsymbol{X}\boldsymbol{\beta}_1 - \boldsymbol{y} \neq \boldsymbol{X}\boldsymbol{\beta}_2 - \boldsymbol{y}$。那么，对于任意 $0 < \lambda < 1$，有

$$
\begin{aligned}
&\sigma\|\boldsymbol{X}(\lambda\boldsymbol{\beta}_1 + (1-\lambda)\boldsymbol{\beta}_2) - \boldsymbol{y}\|_2^2 + \big\|(\lambda\boldsymbol{\beta}_1 + (1-\lambda)\boldsymbol{\beta}_2)\big\|_1 \\
&= \sigma\|\lambda(\boldsymbol{X}\boldsymbol{\beta}_1 - \boldsymbol{y}) + (1-\lambda)(\boldsymbol{X}\boldsymbol{\beta}_2 - \boldsymbol{y})\|_2^2 + \big\|\lambda(\boldsymbol{\beta}_1) + (1-\lambda)(\boldsymbol{\beta}_2)\big\|_1 \\
&< \lambda\left(\sigma\|\boldsymbol{X}\boldsymbol{\beta}_1 - \boldsymbol{y}\|_2^2 + \big\|\boldsymbol{\beta}_1\big\|_1\right) + (1-\lambda)\left(\sigma\|\boldsymbol{X}\boldsymbol{\beta}_2 - \boldsymbol{y}\|_2^2 + \big\|\boldsymbol{\beta}_2\big\|_1\right) \\
&= \lambda c + (1-\lambda)c = c,
\end{aligned}
$$

其中，严格不等式来自 $\sigma\|\cdot\|_2^2$ 的严格凸性和$\|\boldsymbol{\beta}\|_1$的凸性。这意味着，$\mathbb{L}$ 上的点 $\lambda\boldsymbol{\beta}_1 +($ $1-\lambda)\boldsymbol{\beta}_2$ 取得的值低于 c，由此导致矛盾。因此，我们有 $\boldsymbol{X}\boldsymbol{\beta}_1 - \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_2 - \boldsymbol{y}$，进而推出 $\|\boldsymbol{\beta}_1\|_1 = \|\boldsymbol{\beta}_2\|_1$。

对于问题（P1,ϵ），若 $\mathbf{0} \in \mathbb{Q}_\epsilon$，则有 $\mathbb{Q}_\epsilon = \{\mathbf{0}\}$，结论显然成立。否则，假设 $\mathbf{0} \notin \mathbb{Q}_\epsilon$。由于对所有 $\boldsymbol{\beta} \in \mathbb{Q}_\epsilon$而言，最优目标函数值 $\|\boldsymbol{\beta}\|_1$ 都是常数，因此只需证明对所有 $\boldsymbol{\beta} \in \mathbb{Q}_\epsilon$都有 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 = \epsilon$。反设存在一个非零的 $\widetilde{\boldsymbol{\beta}} \in \mathbb{Q}_\epsilon$ 使得 $\big\|\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{y}\big\|_2 < \epsilon$，则可找到一个以 $\boldsymbol{\beta}$ 为中心、半径 $r > 0$ 足够小的非空球 $\mathbb{T}$，使得对所有 $\boldsymbol{\alpha} \in \mathbb{T}$都有 $\|\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{y}\|_2 < \epsilon$。令$\gamma \triangleq \min\left(\frac{r}{2\|\widetilde{\boldsymbol{\beta}}\|_2}, \frac{1}{2}\right)$。我们有 $(1-\gamma)\,\widetilde{\boldsymbol{\beta}} \in \mathbb{T}$，使得 $(1-\gamma)\,\widetilde{\boldsymbol{\beta}}$满足问题（P1,ϵ）中的约束，这意味着 $\big\|(1-\gamma)\widetilde{\boldsymbol{\beta}}\big\|_1 < \big\|\widetilde{\boldsymbol{\beta}}\big\|_1$，从而导致矛盾，并得证。

■

利用引理 3.21，我们可以在与无噪声问题（P_1）的定理 3.19 中相同的条件下，为这两个鲁棒模型建立对偶证书。

定理 3.22：（对偶证书：适用于 (P1,σ) 和 (P1,ϵ)）假设 $\boldsymbol{\beta}$ 是以下问题的解：（P1,σ）或（P1,ϵ）。那么，当且仅当定理 $\boldsymbol{\beta}$ 中所述的相同条件对于3.19 成立时，$\boldsymbol{\beta}$才是唯一解。

定理 [的证明：定理 3.22] 由于引理3.21 的结果对于问题（P1,σ）和（P1,ϵ）完全相同，因此我们仅给出问题（P1,σ）的证明。

根据假设，$\mathbb{Q}_\sigma$（P1, σ）的解集非空，因此我们选取 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$。令 $\widetilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$，根据引理 $\boldsymbol{\beta}$，它与3.21中对的选择无关。现在考虑以下无噪声 ℓ_1-最小化问题：

$$
\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \widetilde{\boldsymbol{y}}, \tag{3.81}
$$

并令 $\mathbb{Q}_q$ 表示其解集。

现在，我们证明 $\mathbb{Q}_\sigma = \mathbb{Q}_q$。由于 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ 且 $\|\boldsymbol{\beta}\|_1 = \|\widetilde{\boldsymbol{\beta}}\|_1$ 对所有 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$ 都成立，并且反过来，任何满足 $\boldsymbol{\beta}$ 和 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ 的 $\|\boldsymbol{\beta}\|_1 = \|\widetilde{\boldsymbol{\beta}}\|_1$ 都必须属于 $\mathbb{Q}_\sigma$，因此只需证明 $\|\boldsymbol{\beta}\|_1 = \|\widetilde{\boldsymbol{\beta}}\|_1$ 对任意 $\boldsymbol{\beta} \in \mathbb{Q}_q$成立。假设这不成立，那么由于问题(3.81) 存在可行解 $\boldsymbol{\beta}$，且目标函数有限，因此 $\mathbb{Q}_q$ 非空，并且存在 $\boldsymbol{z} \in \mathbb{Q}_q$满足 $\|\boldsymbol{z}\|_1 < \|\widetilde{\boldsymbol{\beta}}\|_1$。然而，$\|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{y}\|_2 = \|\widetilde{\boldsymbol{y}} - \boldsymbol{y}\|_2 = \|\boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{y}\|_2$和 $\|\boldsymbol{z}\|_1 < \|\widetilde{\boldsymbol{\beta}}\|_1$ 意味着 $\boldsymbol{z}$ 是问题（$\mathrm{P}_{1,\ \sigma}$）的一个严格优于 $\widetilde{\boldsymbol{\beta}}$的解，这与假设 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$矛盾。因此，$\mathbb{Q}_\sigma = \mathbb{Q}_q$。

由于 $\mathbb{Q}_\sigma = \mathbb{Q}_q$，$\boldsymbol{\beta}$ 是问题（$\mathrm{P}_{1,\ \sigma}$）的唯一解，当且仅当它是问题（3.81）的唯一解。由于问题（3.81）与问题（P_1）具有相同的形式，应用定理 3.19 中针对问题（P_1）的部分，我们得出，$\boldsymbol{\beta}$ 是问题（$\mathrm{P}_{1,\ \sigma}$）的唯一解，当且仅当与定理3.19 相同的条件成立。 ■

定理 3.23：（（$\mathrm{P}_{1,\ \epsilon}$）在测量噪声下的误差界）设 $\boldsymbol{\beta}^*$ 为满足 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 且 $\|\boldsymbol{\epsilon}\|_2 \le \epsilon$的原始信号。再设 $\boldsymbol{\beta}$ 为问题（$\mathrm{P}_{1,\ \epsilon}$）的一个解：

$$(\mathrm{P}_{1,\epsilon}) \qquad \min \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \le \epsilon.$$

将真实信号的支撑集定义为 $\mathbb{S} \triangleq \mathrm{supp}(\boldsymbol{\beta}^*)$，将其补集定义为 $\mathbb{T} \triangleq \mathbb{S}^c$。则重构误差满足

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \le C_3\epsilon + C_4\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1,$$

其中 $C_3 \triangleq 2\sqrt{|\mathbb{S}|}\overline{\cdot\mathcal{U}}(\mathbb{S})$、$C_4 \triangleq \|\boldsymbol{X}\|_2\sqrt{|\mathbb{S}|}\overline{\cdot\mathcal{U}}(\mathbb{S}) + 1$，并且

$$\mathcal{U}(\mathbb{S}) \triangleq \max_{\substack{\mathrm{supp}(\boldsymbol{u})=\mathbb{S},\\ \boldsymbol{u}\neq\boldsymbol{0}}} \frac{\|\boldsymbol{u}\|_2}{\|\boldsymbol{X}\boldsymbol{u}\|_2}.$$ [a]

[a] $\mathcal{U}(\mathbb{S})$ 与 RIP 界的一侧有关；参见定义4.30。

证明 [定理 3。23] 注意 $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 = \|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\|_1 + \|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1$，其中第一项满足

$$\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1 \le \sqrt{|\mathbb{S}|}\cdot\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_2 \le \sqrt{|\mathbb{S}|}\cdot\mathcal{U}(\mathbb{S})\cdot\left\|\boldsymbol{X}_{\mathbb{S}}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}})\right\|_2,$$

其中，第一个不等式由向量范数的标准界（习题 1.26）得出，最后一个不等式则由 $\mathcal{U}(\mathbb{S})$ 的定义得出。令 $\boldsymbol{d} \in \mathbb{R}^p$ 使得 $\boldsymbol{d}_{\mathbb{S}} = \boldsymbol{\beta}_{\mathbb{S}}$且 $\boldsymbol{d}_{\mathbb{T}} = \boldsymbol{0}$，即 $\mathrm{supp}(\boldsymbol{d}) = \mathbb{S}$。因此，根据三角不等式，

$$\left\|\boldsymbol{X}_{\mathbb{S}}(\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}})\right\|_2 = \|\boldsymbol{X}(\boldsymbol{d} - \boldsymbol{\beta}^*)\|_2 \le \left\|\boldsymbol{X}(\boldsymbol{d} - \widehat{\boldsymbol{\beta}})\right\|_2 + \|\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2.$$

由于 $\widehat{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta}^*$ 均为可行解，因此有 $\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 \le 2\epsilon$。谱范数的定义以及向量范数的标准界（习题1.26）也表明

$$\left\|\boldsymbol{X}(\boldsymbol{d} - \widehat{\boldsymbol{\beta}})\right\|_2 \le \|\boldsymbol{X}\|_2\left\|\boldsymbol{d} - \widehat{\boldsymbol{\beta}}\right\|_2 \le \|\boldsymbol{X}\|_2\left\|\boldsymbol{d} - \widehat{\boldsymbol{\beta}}\right\|_1 = \|\boldsymbol{X}\|_2\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1$$

结合这些结果可得

$$\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1 \le \sqrt{|\mathbb{S}|}\cdot\mathcal{U}(\mathbb{S})\cdot(\|\boldsymbol{X}\|_2\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1 + 2\epsilon).$$

重新排列后，结果得证。 ■

假设 $\boldsymbol{\beta}^*$ 和 $\boldsymbol{\alpha}$ 满足与定理3.19相同的条件。则向量 $\boldsymbol{\alpha}=\boldsymbol{X}^\top\boldsymbol{\lambda}$ 满足

$$\boldsymbol{\alpha}_{\mathbb{S}}=\operatorname{sign}(\boldsymbol{\beta}^*_{\mathbb{S}}) \qquad \text{and} \qquad \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty<1,$$

其中 $\mathbb{S}=\operatorname{supp}(\boldsymbol{\beta}^*)$ 且 $\mathbb{T}=\mathbb{S}^c$。令 $\boldsymbol{\beta}$ 为（P_1）的一个解。则

$$\begin{aligned}&\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}}\|_1\geq\left\langle\boldsymbol{\alpha}_{\mathbb{S}},\widehat{\boldsymbol{\beta}}_{\mathbb{S}}\right\rangle;\\&\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1\leq(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}\left(\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1-\left\langle\boldsymbol{\alpha}_{\mathbb{T}},\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\right\rangle\right),\end{aligned}$$

其中第二个不等式由三角不等式 $\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1-\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty\leq(\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1-\langle\boldsymbol{\alpha}_{\mathbb{T}},\ \widehat{\boldsymbol{\beta}}_{\mathbb{T}}\rangle)$ 得出。因此，

$$\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1\leq(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}\left(\|\widehat{\boldsymbol{\beta}}\|_1-\left\langle\boldsymbol{\alpha},\widehat{\boldsymbol{\beta}}\right\rangle\right)\triangleq(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}D_1^{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}^*),$$

其中 $D_1^{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\beta}},\ \boldsymbol{\beta}^*)\triangleq\|\widehat{\boldsymbol{\beta}}\|_1-\|\boldsymbol{\beta}^*\|_1-\langle\boldsymbol{\alpha},\ \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\rangle$ 是与$\|\cdot\|_1$ 相关的 Bregman 距离（参见定义2.87），且 $\boldsymbol{\alpha}\in\partial\|\boldsymbol{\beta}^*\|_1$。

由于 $\boldsymbol{\alpha}\in\mathcal{C}(\boldsymbol{X}^\top)$，因此对于某个向量 $\boldsymbol{\lambda}$（即对偶向量），有 $\boldsymbol{\alpha}=\boldsymbol{X}^\top\boldsymbol{\lambda}$。利用 $\boldsymbol{\beta}$ 的最优性条件，可得

$$\begin{aligned}D_1^{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}^*)&\triangleq\|\widehat{\boldsymbol{\beta}}\|_1-\|\boldsymbol{\beta}^*\|_1-\left\langle\boldsymbol{\alpha},\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\rangle\\&\leq 0+\left\langle\boldsymbol{X}^\top\boldsymbol{\lambda},\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\rangle\leq\|\boldsymbol{\lambda}\|_2\left\|\boldsymbol{X}(\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}})\right\|_2\leq 2\|\boldsymbol{\lambda}\|_2\epsilon.\end{aligned}$$

因此，我们可以将补集中的最优解上界估计为

$$\|\widehat{\boldsymbol{\beta}}_{\mathbb{T}}\|_1\leq 2(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}\|\boldsymbol{\lambda}\|_2\epsilon. \tag{3.82}$$

上述分析给出了定理 3.23的以下推论

推论 3.24：（含测量噪声时($\mathrm{P}_{1,\epsilon}$) 的误差界）假设 $\boldsymbol{\beta}^*$和 $\boldsymbol{\alpha}=\boldsymbol{X}^\top\boldsymbol{\lambda}$ 满足与定理3.19相同的条件，并设 $\boldsymbol{\beta}^*$ 为满足 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$ 且 $\|\boldsymbol{\epsilon}\|_2\leq\epsilon$ 的原始信号。设 $\boldsymbol{\beta}$ 为问题（$\mathrm{P}_{1,\epsilon}$）的一个解：

$$(\mathrm{P}_{1,\epsilon})\qquad \min\|\boldsymbol{\beta}\|_1\quad\text{s.t.}\quad\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2\leq\epsilon.$$

定义 $\mathbb{S}\triangleq\operatorname{supp}(\boldsymbol{\beta}^*)$ 和 $\mathbb{T}\triangleq\mathbb{S}^c$。那么，

$$\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1\leq C\epsilon,$$

其中

$$C\triangleq 2\sqrt{|\mathbb{S}|}\cdot\mathcal{U}(\mathbb{S})+\frac{2\|\boldsymbol{\lambda}\|_2(\|\boldsymbol{X}\|_2\sqrt{|\mathbb{S}|}\cdot\mathcal{U}(\mathbb{S})+1)}{1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty}.$$

and

$$\mathcal{U}(\mathbb{S})\triangleq\max_{\substack{\operatorname{supp}(\boldsymbol{u})=\mathbb{S},\\ \boldsymbol{u}\neq\boldsymbol{0}}}\frac{\|\boldsymbol{u}\|_2}{\|\boldsymbol{X}\boldsymbol{u}\|_2}.$$

请注意，条件 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty < 1$ 至关重要：当 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty \to 1$ 时，常数会发散，这反映出当对偶证书在支撑集补集上的作用较弱时，恢复可能存在不稳定性。

对于惩罚形式，也可以推导出类似的误差界（$\mathrm{P}_{1,\sigma}$）：

$$\min \|\boldsymbol{\beta}\|_1 + \sigma \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$$

前提是 $\sigma > 0$ 选取适当。这是由推论3.3中建立的题目（$\mathrm{P}_{1,\epsilon}$）与（$\mathrm{P}_{1,\sigma}$）之间的等价性得出的；此处省略细节。

3.5. ℓ_1-分析模型的最优性条件*

现在考虑以下 ℓ_1-分析形式：

$$(Q_1) \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1, \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}; \tag{3.83a}$$

$$(Q_{1,\sigma}) \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1 + \sigma \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2; \tag{3.83b}$$

$$(Q_{1,\epsilon}) \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1, \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \le \epsilon, \tag{3.83c}$$

其中 ϵ、σ 为正参数，$\boldsymbol{X} \in \mathbb{R}^{n\times p}$、$\boldsymbol{\beta} \in \mathbb{R}^p$ 和 $\boldsymbol{\Psi} \in \mathbb{R}^{p\times q}$。当观测值无噪声，即 $\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} = \boldsymbol{\epsilon} = \boldsymbol{0}$ 时，问题（3.83a）是合适的。若 $\boldsymbol{\Psi} = \boldsymbol{I}$ 为单位矩阵，这些模型便化为标准的 ℓ_1-合成模型（即第129页的（P_1）、第129页的（$\mathrm{P1}_{,\sigma}$）以及第131页的（$\mathrm{P1}_{,\epsilon}$））。我们着重讨论高维情形 $n \le p$（其合理性见引理 3.33 或定理 3.34）。

对于这些模型的证书，我们考虑以下条件。

条件 3.25. 设 $\boldsymbol{\Psi} \in \mathbb{R}^{q\times p}$ 和 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。给定 $\boldsymbol{\beta} \in \mathbb{R}^p$，指标集 $\mathbb{S} = \mathrm{supp}(\boldsymbol{\Psi}^\top\boldsymbol{\beta}) \subset \{1, 2, \dots, q\}$ 和 $\mathbb{T} \triangleq \mathbb{S}^c$ 满足

(i) $\mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top) \cap \mathcal{N}(\boldsymbol{X}) = \{\boldsymbol{0}\}$;

(ii) There exists $\boldsymbol{\alpha} \in \mathbb{R}^q$ such that $\boldsymbol{\Psi}\boldsymbol{\alpha} \in \mathcal{C}(\boldsymbol{X}^\top)$, $\boldsymbol{\alpha}_{\mathbb{S}} = \mathrm{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\overline{\boldsymbol{\beta}})$, and $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty < 1$.

该条件是定理 3.29 中经典 ℓ_1-最小化问题 (P_1) 的对偶证书条件的自然推广。再次注意，$\boldsymbol{\Psi}(\mathbb{T}) \in \mathbb{R}^{q\times p}$ 和 $\boldsymbol{\Psi}_{\mathbb{T}} = \boldsymbol{\Psi}[:\mathbb{T}] \in \mathbb{R}^{n\times|\mathbb{T}|}$；参见定义 1.1。当 $\boldsymbol{\Psi} = \boldsymbol{I}_p$ 时，对于所有 $i \in \mathbb{T}\}$，$\mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top) = \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \beta_i = 0$。于是，条件 3.25 (i) 简化为要求 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩，与定理 3.29 完全相同。

更一般地，条件 3.25 (i) 保证不存在非零扰动 $\Delta\boldsymbol{\beta}$同时满足 $\boldsymbol{\Psi}(\mathbb{T})^\top\boldsymbol{\beta} = \boldsymbol{\Psi}(\mathbb{T})^\top(\boldsymbol{\beta}+\Delta\boldsymbol{\beta})$ 和 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}(\boldsymbol{\beta}+\Delta\boldsymbol{\beta})$。否则，对于某个充分小的 $\mathbb{M} = [\boldsymbol{\beta} - \gamma\Delta\boldsymbol{\beta}, \boldsymbol{\beta} + \gamma\Delta\boldsymbol{\beta}]$，存在一个非空区间 $\gamma > 0$，使得$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}$ 且 $\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1$ 在 $\boldsymbol{\beta} \in \mathbb{M}$上为常数；因此，$\boldsymbol{\beta}$ 不可能是唯一极小化解。

条件 3.25（ii）断言严格互补对偶证书 $\boldsymbol{\alpha}$ 的存在性。为此，回顾问题 (Q_1) 的最优性条件：$\boldsymbol{0} \in \boldsymbol{\Psi}\partial\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1 - \boldsymbol{X}^\top\boldsymbol{\lambda}$，其中 $\boldsymbol{\lambda}$是拉格朗日乘子向量（见备注 2.116）。在综合情形（$\boldsymbol{\Psi} = \boldsymbol{I}$）下，$\boldsymbol{X}^\top\boldsymbol{\lambda}$ 是 $\boldsymbol{\beta}$ 处 ℓ_1 范数的一个次梯度（见定理 3.19）；然而，由于 $\boldsymbol{\Psi}$ 的变换，在此 ℓ_1-分析情形下，$\boldsymbol{X}^\top\boldsymbol{\lambda}$不再是次梯度。我们可以将该条件改写为 $\boldsymbol{0} = \boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{X}^\top\boldsymbol{\lambda}$，其中 $\boldsymbol{\alpha} \in \partial\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1$，这由 ℓ_1范数的次微分（习题 2.58）转化为$\boldsymbol{\alpha}_{\mathbb{S}} = \mathrm{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{\beta})$ 和 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty \le 1$。

因此，向量 $\boldsymbol{\alpha}$ 证明了 $\boldsymbol{\beta}=\boldsymbol{\beta}$ 的最优性。关于唯一性和鲁棒性，我们稍后将证明严格不等式 $\|\boldsymbol{\alpha}_{\mathbb{S}}\|_{\infty}<1$ 至关重要。

综合这些观测值可以确认，条件 3.25 确实是定理 3.19 中的对偶证书条件向 ℓ_1-分析框架的自然推广。我们还引入一个稍微更一般的版本：

条件 3.26。 设 $\boldsymbol{\Psi}\in\mathbb{R}^{q\times p}$ 和 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$。给定 $\boldsymbol{\beta}\in\mathbb{R}^{p}$，令 $\mathbb{S}=\operatorname{supp}(\boldsymbol{\Psi}^{\top}\boldsymbol{\beta})$ 。存在一个非空指标集 $\mathbb{T}\subseteq\mathbb{S}^{c}$，使得指标集 $\mathbb{S}$、 $\mathbb{T}$ 和 $\mathbb{I}\triangleq(\mathbb{S}\cup\mathbb{T})^{c}$满足

(a) $\mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^{\top})\cap\mathcal{N}(\boldsymbol{X})=\{\mathbf{0}\}$; (b) 存在 $\boldsymbol{\alpha}\in\mathbb{R}^{q}$，使得 $\boldsymbol{\Psi}\boldsymbol{\alpha}\in\underline{\mathcal{C}}(\boldsymbol{X}^{\top})$、 $\boldsymbol{\alpha}_{\mathbb{S}}=\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{\beta})$、 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_{\infty}<1$，以及$\|\boldsymbol{\alpha}_{\mathbb{I}}\|_{\infty}\leq 1$。

在条件 3.26中，选择较小的 $\mathbb{T}$ 会放宽要求 (b)，但会扩大零空间$\mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^{\top})$，从而收紧要求 (a)。尽管这种表述在选择 $\mathbb{T}$时提供了更大的灵活性，但我们将证明条件 3.25 和 3.26 实际上是等价的。

我们还采用以下标准假设：

假设3.27。 考虑方程 (3.83) 中的问题，我们假设

1. 矩阵 $\boldsymbol{X}$ 满行秩。 2. 矩阵 $\boldsymbol{\Psi}$ 满行秩。
3. 最大奇异值 $\sigma_{\max}(\boldsymbol{\Psi})=1$。

假设 1 和假设 2 可避免冗余方程或变换系数，并且通常用于压缩感知和逆问题中。假设 3 并非必要：任意矩阵 $\boldsymbol{\Psi}$ 都可以通过除以 $\sigma_{\max}(\boldsymbol{\Psi})$ 进行归一化，这只会重新缩放 ℓ_1-惩罚项，而不会改变解集。

3.5.1 无噪声情形下的唯一恢复

我们首先建立问题 (Q_1) 的解

定理 3.28： （(Q_1) **的充要条件**）考虑优化问题 (Q_1)。向量 $\boldsymbol{\beta}$ 是 (Q_1 的解，当且仅当对所有 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$,

$$\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\right\rangle\leq\left\|\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\boldsymbol{n}\right\|_{1},$$

其中 $\mathbb{S}\triangleq\operatorname{supp}(\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}})$。

证明 [定理 3 的28] **必要性。** 假设 $\boldsymbol{\beta}$ 是（Q_1）的一个解。由于函数 $f(\eta)=\boldsymbol{X}(\boldsymbol{\beta}+\eta\boldsymbol{n})$ 连续可微，我们可以找到足够小的 η邻域，使得在 0 附近 $\operatorname{sign}(\boldsymbol{\Psi}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n}))=\operatorname{sign}(\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}})$. 当 η足够小且非零时，我们有

$$\begin{aligned}
\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_{1}&\leq\|\boldsymbol{\Psi}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n})\|_{1}=\|\boldsymbol{\Psi}_{\mathbb{S}}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n})\|_{1}+\|\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n})\|_{1}\\
&\leq\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n}),\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n}))\right\rangle+\|\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\widehat{\boldsymbol{\beta}}\|_{1}+\|\eta\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\boldsymbol{n}\|_{1}\\
&=\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\right\rangle+\|\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\widehat{\boldsymbol{\beta}}\|_{1}+\|\eta\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\boldsymbol{n}\|_{1}\\
&=\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}},\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\right\rangle+\eta\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\right\rangle+\|\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\widehat{\boldsymbol{\beta}}\|_{1}+\|\eta\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\boldsymbol{n}\|_{1}\\
&=\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_{1}+\eta\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\right\rangle+\|\eta\boldsymbol{\Psi}_{\mathbb{S}^{c}}^{\top}\boldsymbol{n}\|_{1}.
\end{aligned}$$

因此，对于任意 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，我们有 $\langle \boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\ \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\rangle \leq \|\boldsymbol{\Psi}_{\mathbb{S}^c}^{\top}\boldsymbol{n}\|_1$。

充分性。假设所述不等式对所有 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$ 都成立。反设存在一个可行点 $\boldsymbol{\beta}'$，使得 $\boldsymbol{X}\boldsymbol{\beta}' = \boldsymbol{y}$ 且 $\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}'\|_1 < \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1$。定义 $\boldsymbol{n} \triangleq \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}'$，由于 $\boldsymbol{X}\boldsymbol{\beta}' = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$，它满足 $\boldsymbol{X}\boldsymbol{n} = \boldsymbol{0}$。这意味着 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$。利用单位矩阵 $\|\boldsymbol{\beta}'\|_1 = \|\widehat{\boldsymbol{\beta}} - \boldsymbol{n}\|_1$，我们考虑目标函数的分解：

$$\begin{aligned}\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}'\|_1 &= \left\|\boldsymbol{\Psi}^{\top}(\widehat{\boldsymbol{\beta}} - \boldsymbol{n})\right\|_1 = \sum_{i\in\mathbb{S}}\left|\boldsymbol{\Psi}_i^{\top}(\widehat{\beta}_i - n_i)\right| + \sum_{i\in\mathbb{S}^c}\left|\boldsymbol{\Psi}_i^{\top}n_i\right| \\ &\geq \sum_{i\in\mathbb{S}}\left|\boldsymbol{\Psi}_i^{\top}\widehat{\beta}_i\right| - \sum_{i\in\mathbb{S}}\operatorname{sign}(\boldsymbol{\Psi}_i^{\top}\widehat{\beta}_i)\boldsymbol{\Psi}_i^{\top}n_i + \sum_{i\in\mathbb{S}^c}\left|\boldsymbol{\Psi}_i^{\top}n_i\right| \geq \sum_{i\in\mathbb{S}}\left|\boldsymbol{\Psi}_i^{\top}\widehat{\beta}_i\right| = \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1,\end{aligned}$$

其中，第一个不等式由 $|b-n| \geq |b| - \operatorname{sign}(b)n$ 这一事实得出，第二个不等式则由给定条件 $\langle \boldsymbol{\Psi}_{\mathbb{S}}^{\top}\boldsymbol{n},\ \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})\rangle \leq \|\boldsymbol{\Psi}_{\mathbb{S}^c}^{\top}\boldsymbol{n}\|_1$ 得出。由于假设为 $\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}'\|_1 < \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1$，因此得到矛盾。故不存在这样的 $\boldsymbol{\beta}'$，而且 $\widehat{\boldsymbol{\beta}}$ 确实是最优解。证明完毕。

■

下面的定理指出，条件 3.25 和 3.26 等价，并且它们是问题（Q_1）的唯一解 $\widehat{\boldsymbol{\beta}}$ 的必要且充分条件。

定理 3.29：（Q_1）（**Zhang 等人，2016**）在假设 1 下，设 $\widehat{\boldsymbol{\beta}}$ 是问题（Q_1）的一个解。以下各项等价：

(i) $\widehat{\boldsymbol{\beta}}$ 是唯一的。(ii) 条件 3.25 对 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ 成立。(iii) 条件 3.26 对 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ 成立。—

定理 [3 的证明 3.29] 这个证明改编自 Grasmair 等人 (2011年)；Zhang 等人 (2015年、2016)。我们通过证明以下蕴含关系来建立等价性：(iii) $\Rightarrow$ (i) $\Rightarrow$(ii) $\Rightarrow$ (iii)。

(iii) $\Rightarrow$ (i)。考虑任意扰动 $\boldsymbol{\beta} + \boldsymbol{n}$，其中 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\boldsymbol{0}\}$。由于根据条件 3.26，$\boldsymbol{\alpha}_{\mathbb{S}} = \text{符号函数}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})$，$\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty < 1$，且 $\|\boldsymbol{\alpha}_{\mathbb{I}}\|_\infty \leq 1$，我们可以取一个次梯度 $\boldsymbol{g} \in \partial\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1$，使得 $\boldsymbol{g}_{\mathbb{S}} = \boldsymbol{\alpha}_{\mathbb{S}} \equiv \text{符号函数}(\boldsymbol{\Psi}_{\mathbb{S}}^{\top}\widehat{\boldsymbol{\beta}})$，$\boldsymbol{g}_{\mathbb{I}} = \boldsymbol{\alpha}_{\mathbb{I}}$ ($\|\boldsymbol{g}_{\mathbb{I}}\|_\infty = \|\boldsymbol{\alpha}_{\mathbb{I}}\|_\infty \leq 1$)，并且$\|\boldsymbol{g}_{\mathbb{T}}\|_\infty \leq 1$ 满足 $\langle \boldsymbol{g}_{\mathbb{T}},\ \boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\rangle = \|\boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\|_1$（这与 $\boldsymbol{\alpha}$ 类似，不过这里不要求严格不等式；实际上，$\|\boldsymbol{g}_{\mathbb{T}}\|_\infty = 1$ 需要满足后一个等式）。接下来，次梯度不等式（定义 2.57）表明

$$\begin{aligned}\|\boldsymbol{\Psi}^{\top}(\widehat{\boldsymbol{\beta}} + \boldsymbol{n})\|_1 &\geq \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1 + \langle \boldsymbol{\Psi}\boldsymbol{g}, \boldsymbol{n}\rangle \overset{\dagger}{=} \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1 + \langle \boldsymbol{\Psi}\boldsymbol{g} - \boldsymbol{\Psi}\boldsymbol{\alpha}, \boldsymbol{n}\rangle \\ &= \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1 + \left\langle \boldsymbol{g} - \boldsymbol{\alpha}, \boldsymbol{\Psi}^{\top}\boldsymbol{n}\right\rangle \overset{\ddagger}{=} \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1 + \left\langle \boldsymbol{g}_{\mathbb{T}} - \boldsymbol{\alpha}_{\mathbb{T}}, \boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\right\rangle \\ &\geq \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1 + \|\boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\|_1(1 - \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty),\end{aligned}$$

其中，等式 (†) 由 $\boldsymbol{\Psi}\boldsymbol{\alpha} \in \mathcal{C}\left(\boldsymbol{X}^{\top}\right) = \mathcal{N}\left(\boldsymbol{X}\right)^{\perp}$ 和 $\boldsymbol{n} \in \mathcal{N}\left(\boldsymbol{X}\right)$ 得出；等式 (‡) 由 $\boldsymbol{g}$ 的定义得出；最后一个不等式则是对 H¨older 不等式 $\langle \boldsymbol{\alpha}_{\mathbb{T}}、\ \boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\rangle \leq \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty \|\boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\|_1$ 和 $\langle \boldsymbol{g}_{\mathbb{T}}、\ \boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\rangle = \|\boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\|_1$的应用。由于 $\boldsymbol{n} \in \mathcal{N}\left(\boldsymbol{X}\right) \setminus \{\boldsymbol{0}\}$，且 $\mathcal{N}\left(\boldsymbol{\Psi}\left(\mathbb{T}\right)^{\top}\right) \cap \mathcal{N}\left(\boldsymbol{X}\right) = \{\boldsymbol{0}\}$，因此有 $\|\boldsymbol{\Psi}_{\mathbb{T}}^{\top}\boldsymbol{n}\|_1 > 0$。结合条件$\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty < 1$，对于每个 $\boldsymbol{n} \in \mathcal{N}\left(\boldsymbol{X}\right) \setminus \{\boldsymbol{0}\}$，都有 $\|\boldsymbol{\Psi}^{\top}\left(\widehat{\boldsymbol{\beta}} + \boldsymbol{n}\right)\|_1 > \|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\|_1$。这就证明了 $\widehat{\boldsymbol{\beta}}$ 是问题（Q_1）的唯一极小化解。

$\widehat{\quad}$ $\widehat{\quad}$ (i) $\Rightarrow$ (ii)。对于每个 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$，都有 $\boldsymbol{X}(\boldsymbol{\beta}+\eta\boldsymbol{n}) = \boldsymbol{X}\boldsymbol{\beta}$。由于函数 $\backsim f(\eta) = \boldsymbol{X}(\boldsymbol{\beta}+\eta\boldsymbol{n})$ 连续可微，我们可以在 0 附近取足够小的 η，使得 sign$(\boldsymbol{\Psi}^\top(\widehat{\boldsymbol{\beta}}+\eta\boldsymbol{n})) =$sign$(\boldsymbol{\Psi}^\top\widehat{\boldsymbol{\beta}})$。由于 $\widehat{\boldsymbol{\beta}}$是唯一解，对于足够小且非零的$\eta$，定理 3.28 的证明给出了以下严格不等式

$$\left\|\boldsymbol{\Psi}^\top\widehat{\boldsymbol{\beta}}\right\|_1 < \left\|\boldsymbol{\Psi}^\top\widehat{\boldsymbol{\beta}}\right\|_1 + \eta\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{n}, \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\right\rangle + \left\|\eta\boldsymbol{\Psi}_{\mathbb{S}^c}^\top\boldsymbol{n}\right\|_1.$$

因此，对于任意 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$，我们有

$$\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{n}, \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\right\rangle < \left\|\boldsymbol{\Psi}_{\mathbb{S}^c}^\top\boldsymbol{n}\right\|_1. \tag{3.84}$$

如果条件 $\mathcal{N}(\boldsymbol{\Psi}(\mathbb{S}^c)^\top) \cap \mathcal{N}(\boldsymbol{X}) = \{\mathbf{0}\}$ 不成立，我们可以选取一个非零向量$\boldsymbol{n} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{S}^c)^\top) \cap \mathcal{N}(\boldsymbol{X})$。我们还有 $-\boldsymbol{n} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{S}^c)^\top) \cap \mathcal{N}(\boldsymbol{X})$。由 (3.84)，我们有$\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{n}$、符号函数$(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\rangle < 0$和 $-\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{n}$、符号函数$(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\rangle < 0$，这导致矛盾，并表明 $\mathcal{N}(\boldsymbol{\Psi}(\mathbb{S}^c)^\top) \cap \mathcal{N}(\boldsymbol{X}) = \{\mathbf{0}\}$，即条件 3.25 (i)。

接下来，我们构造 $\boldsymbol{\alpha}$ 使其满足条件 3.25 (ii)。定义 $\widetilde{\boldsymbol{\alpha}} \in \mathbb{R}^q$ 为 $\widetilde{\boldsymbol{\alpha}}_{\mathbb{S}} =$ 符号函数$(\boldsymbol{\Psi}_{\mathbb{S}}^\top \widehat{\boldsymbol{\beta}})$ 和 $\widetilde{\boldsymbol{\alpha}}_{\mathbb{S}^c} = \mathbf{0}$。

- 如果这样的 $\widetilde{\boldsymbol{\alpha}}$ 满足 $\boldsymbol{\Psi}\widetilde{\boldsymbol{\alpha}} \in \mathcal{C}(\boldsymbol{X}^\top)$，则问题解决。如果 $\boldsymbol{\Psi}\widetilde{\boldsymbol{\alpha}} \notin \mathcal{C}(\boldsymbol{X}^\top) \equiv \mathcal{N}(\boldsymbol{X})^\perp$，
- 那么我们将构造一个新向量，使其满足条件 3.25 (ii)。

设 $\boldsymbol{Q} \in \mathbb{R}^{p\times r}$ 为一个矩阵，其列构成 $\mathcal{N}(\boldsymbol{X})$ 的一组基。由此可知，$\boldsymbol{\gamma} \triangleq \boldsymbol{Q}^\top\boldsymbol{\Psi}\widetilde{\boldsymbol{\alpha}} \in \mathbb{R}^r$ 必为非零向量。考虑如下优化问题

$$\min_{\boldsymbol{u}\in\mathbb{R}^q} \|\boldsymbol{u}\|_\infty \quad \text{s.t.} \quad \boldsymbol{Q}^\top\boldsymbol{\Psi}\boldsymbol{u} = -\boldsymbol{\gamma} \text{ and } \boldsymbol{u}_{\mathbb{S}} = \mathbf{0}. \tag{3.85}$$

对于问题 (3.85) 的任意极小化解 $\widehat{\boldsymbol{u}}$，都有 $\boldsymbol{Q}^\top\boldsymbol{\Psi}(\widetilde{\boldsymbol{\alpha}}+\widehat{\boldsymbol{u}}) = \mathbf{0}$，使得 $\boldsymbol{\Psi}(\widetilde{\boldsymbol{\alpha}}+\widehat{\boldsymbol{u}})$ 与零空间 $\mathcal{N}(\boldsymbol{X})$ 正交，由此可得 $\boldsymbol{\Psi}(\widetilde{\boldsymbol{\alpha}}+\widehat{\boldsymbol{u}}) \in \mathcal{N}(\boldsymbol{X})^\perp \equiv \mathcal{C}(\boldsymbol{X}^\top)$ 且 $(\widetilde{\boldsymbol{\alpha}}+\widehat{\boldsymbol{u}})_{\mathbb{S}} = \widetilde{\boldsymbol{\alpha}}_{\mathbb{S}} =$符号函数$(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})$。因此，我们将利用强对偶性证明问题 (3.85) 的目标函数严格小于 1，从而完成条件 3.25 (ii) 部分的证明。为此，我们将问题 (3.85) 改写为如下等价形式：

$$\min_{\boldsymbol{u}\in\mathbb{R}^q} \|\boldsymbol{u}_{\mathbb{S}^c}\|_\infty \quad \text{s.t.} \quad \boldsymbol{Q}^\top\boldsymbol{\Psi}_{\mathbb{S}^c}\boldsymbol{u}_{\mathbb{S}^c} = -\boldsymbol{\gamma}, \tag{3.86}$$

其拉格朗日对偶问题（即一个 ℓ_∞最小化问题的对偶问题；参见例2.117）为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^r} \langle\boldsymbol{\lambda}, \boldsymbol{\gamma}\rangle \quad \text{s.t.} \quad \left\|\boldsymbol{\Psi}_{\mathbb{S}^c}^\top\boldsymbol{Q}\boldsymbol{\lambda}\right\|_1 \le 1. \tag{3.87}$$

注意到 $\boldsymbol{Q}\boldsymbol{\lambda} \in \mathcal{N}(\boldsymbol{X})$，且 $|\langle\boldsymbol{\lambda}, \boldsymbol{\gamma}\rangle| = \left|\left\langle\boldsymbol{\lambda}, \boldsymbol{Q}^\top\boldsymbol{\Psi}\widetilde{\boldsymbol{\alpha}}\right\rangle\right| = \left|\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{Q}\boldsymbol{\lambda}, \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\right\rangle\right|$。利用（3.84），对于任意 $\boldsymbol{\lambda}$，有

$$|\langle\boldsymbol{\lambda}, \boldsymbol{\gamma}\rangle| = \begin{cases} \left|\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{Q}\boldsymbol{\lambda}, \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\right\rangle\right| = 0, & \text{if } \boldsymbol{Q}\boldsymbol{\lambda} = \mathbf{0}; \\ \left|\left\langle\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{Q}\boldsymbol{\lambda}, \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})\right\rangle\right| < \left\|\boldsymbol{\Psi}_{\mathbb{S}^c}^\top\boldsymbol{Q}\boldsymbol{\lambda}\right\|_1 \le 1, & \text{otherwise.} \end{cases}$$

因此，问题（3.87）是可行的，且其目标函数值严格小于 1。根据强对偶性（由于原问题可行，且目标函数是凸且连续的，该性质成立），问题（3.85）和（3.86）达到其最小值，且最优值也严格小于 1。令 $\widehat{\boldsymbol{u}}$ 为一个最优解。定义 $\boldsymbol{\alpha} = \widetilde{\boldsymbol{\alpha}} + \widehat{\boldsymbol{u}}$。则：

- $\boldsymbol{\alpha}_{\mathbb{S}} = \widetilde{\boldsymbol{\alpha}}_{\mathbb{S}} = \operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\widehat{\boldsymbol{\beta}})$,
- $\|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_\infty = \|\widehat{\boldsymbol{u}}_{\mathbb{S}^c}\|_\infty < 1$,
- $\boldsymbol{Q}^\top\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{Q}^\top\boldsymbol{\Psi}(\widetilde{\boldsymbol{\alpha}}+\widehat{\boldsymbol{u}}) = \boldsymbol{\gamma} - \boldsymbol{\gamma} = \mathbf{0}$, so $\boldsymbol{\Psi}\boldsymbol{\alpha} \in \mathcal{N}(\boldsymbol{X})^\perp \equiv \mathcal{C}(\boldsymbol{X}^\top)$.

因此，$\boldsymbol{\alpha}$ 满足条件3.25 (ii)。

(ii) $\Rightarrow$ (iii)。令 $\mathbb{T} \triangleq \mathbb{S}^c$ 和 $\mathbb{I} \triangleq \varnothing$。那么条件3.25 直接推出条件3.26，从而完成了蕴含关系的循环。 ■

3.5.2 含噪信号中的唯一恢复

与定理 3.29类似，在存在噪声的情况下，条件 3.25 和 3.26 仍然等价，并给出了 $\boldsymbol{\beta}$ 问题（$Q_{1,\ \sigma}$）或问题（$Q_{1,\ \epsilon}$）的解唯一性的充要条件。为证明这一点，我们需要以下引理。其证明与引理 3.21 的证明类似，故略去。

引理 3.30：（解集上的常数，适用于 $(Q_{1,\sigma})$ 和 $(Q_{1,\epsilon})$） 设 $\sigma > 0$。若函数 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 + \sigma\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2^2$ 在凸集 $\mathbb{C}$上为常数，则 $\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}$和 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1$ 在 $\mathbb{C}$上均为常数。

令 $\mathbb{Q}_\sigma$ 表示问题（$Q_{1,\ \sigma}$）的解集。由于 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 + \sigma\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2^2$ 在 $\mathbb{Q}_\sigma$ 上为常数，上述结果表明，在问题（$Q_{1,\ \sigma}$）中，$\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}$ 和 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1$ 在 $\mathbb{Q}_\sigma$ 上为常数。

此外，假设 $\boldsymbol{\Psi}$ 在问题（$Q_{1,\ \epsilon}$）中满行秩。那么 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2$和 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1$ 在 $\mathbb{Q}_\epsilon$ 上均为常数，其中 $\mathbb{Q}_\epsilon$ 表示问题（$Q_{1,\ \epsilon}$）的解集

定理 3.31：（（$Q_{1,\ \sigma}$）的对偶证书）以及（$Q_{1,\ \epsilon}$）（Zhang 等人，2016））在满足假设 1 的条件下，设 $\boldsymbol{\beta}$ 是问题（$Q_{1,\ \sigma}$）的一个解；或者在满足假设 1 和假设 2 的条件下，设 $\widehat{\boldsymbol{\beta}}$ 是问题（$Q_{1,\ \epsilon}$）的一个解。则以下陈述等价：

(i) 解 $\widehat{\boldsymbol{\beta}}$ 是唯一的。(ii) 条件 3.25 对 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ 成立。(iii) 条件 3.26 对 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$成立。

证明 [定理 3。31]证明过程与定理 3.22 的证明非常相似，并且依赖于引理 3.30。由于引理 3.30 为两个问题（$Q_{1,\sigma}$）和（$Q_{1,\epsilon}$）的解集给出了相同的结构性质，我们仅对问题（$Q_{1,\sigma}$）给出论证；（$Q_{1,\epsilon}$）的情形可类似推出。

设 $\mathbb{Q}_\sigma$ 表示问题（$Q_{1,\sigma}$）的（非空）解集，并任取 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$。令 $\widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$，根据引理 3.30，它与 $\boldsymbol{\beta}$ 的选取无关。我们引入以下约束优化问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \widehat{\boldsymbol{y}}, \tag{3.88}$$

并令 $\mathbb{Q}_q$ 表示其解集。

我们断言 $\mathbb{Q}_\sigma = \mathbb{Q}_q$。由于 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$且对所有 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$，$\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 = \left\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\right\|_1$；反过来，任何满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$和 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 = \left\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\right\|_1$的 $\boldsymbol{\beta}$都属于 $\mathbb{Q}_\sigma$，因此只需证明对于任意 $\boldsymbol{\beta} \in \mathbb{Q}_q$，$\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}\right\|_1 = \left\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\right\|_1$。反设此结论不成立，那么由于问题（3.88）以 $\boldsymbol{\beta}$ 为可行解且目标函数有限，因此 $\mathbb{Q}_q$ 非空，并且存在满足 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{z}\right\|_1 < \left\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\right\|_1$的 $\boldsymbol{z} \in \mathbb{Q}_q$。但是，$\|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{y}\|_2 = \|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|_2 = \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y}\right\|_2$和 $\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{z}\right\|_1 < \left\|\boldsymbol{\Psi}^{\top}\widehat{\boldsymbol{\beta}}\right\|_1$意味着，相较于 $\boldsymbol{\beta}$，$\boldsymbol{z}$是问题（$Q_{1,\sigma}$）的一个严格更优解，这与假设 $\boldsymbol{\beta} \in \mathbb{Q}_\sigma$矛盾。因此，范数必相等，且 $\mathbb{Q}_\sigma = \mathbb{Q}_q$。

由于 $\mathbb{Q}_\sigma = \mathbb{Q}_q$， $\boldsymbol{\beta}$ 是问题（$Q_{1,\sigma}$）的唯一解，当且仅当它是问题（3.88）的唯一解。但问题（3.88）与问题（Q_1）形式相同。应用定理 3.29，可得唯一性成立，当且仅当条件 3.25 。或条件 3.26 满足（其中 $\boldsymbol{\beta} = \boldsymbol{\beta}$）。证明完毕。 ■

3.5.3 噪声信号中的稳定恢复

为陈述鲁棒性结果，定义

$$\mathcal{U}(\mathbb{T}) \triangleq \max_{\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top) \setminus \{\mathbf{0}\}} \frac{\|\boldsymbol{\beta}\|_2}{\|\boldsymbol{X}\boldsymbol{\beta}\|_2}. \tag{3.89}$$

条件 3.25 (i) 保证 $0 < \mathcal{U}(\mathbb{T}) < +\infty$。若 $\boldsymbol{\Psi} = \boldsymbol{I}$，则任何非零 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top) \setminus \{\mathbf{0}\}$ 都是支撑于 $\mathbb{T}^c$的稀疏（非零）向量；因此， $\mathcal{U}(\mathbb{T})$ 等于子矩阵最小奇异值的倒数 $\boldsymbol{X}_{\mathbb{T}^c}$ （见（1.29））。

引理 3.32 假设向量 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha}$ 满足条件 3.25。设 $\mathbb{S} \triangleq \operatorname{supp}(\boldsymbol{\Psi}^\top \boldsymbol{\beta})$ 且 $\mathbb{T} \triangleq \mathbb{S}^c$。那么，

$$\left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta} - \boldsymbol{\Psi}^\top \overline{\boldsymbol{\beta}}\right\|_1 \leq C_3 \left\|\boldsymbol{X}(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}})\right\|_2 + C_4 D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \overline{\boldsymbol{\beta}}), \quad \forall \boldsymbol{\beta}, \tag{3.90}$$

其中 $D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\beta}) \triangleq \left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}\right\|_1 - \left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}\right\|_1 - \langle \boldsymbol{\Psi}\boldsymbol{\alpha}, \boldsymbol{\beta} - \boldsymbol{\beta} \rangle$ 是与函数 $\left\|\boldsymbol{\Psi}^\top \cdot\right\|_1$相关联的 Bregman 距离，其中 $\boldsymbol{\alpha} \in \partial \left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}\right\|_1$ （见定义 2.87），绝对常数C_3， C_4 由下式给出

$$C_3 = \mathcal{U}(\mathbb{T})\sqrt{|\mathbb{S}|}, \quad \text{and} \quad C_4 = \frac{1 + \operatorname{cond}(\boldsymbol{\Psi}) \|\boldsymbol{X}\|_2 C_3}{1 - \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty},$$

其中，$\operatorname{cond}(\boldsymbol{\Psi}) = \sigma_{\max}(\boldsymbol{\Psi})/\sigma_{\min}(\boldsymbol{\Psi})$表示 $\boldsymbol{\Psi}$的条件数。

引理 [的证明3。32] 此证明改编自Zhang 等人(2016)，分两步进行。第一部分证明，对于任意 $\boldsymbol{u} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top)$，

$$\left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta} - \boldsymbol{\Psi}^\top \overline{\boldsymbol{\beta}}\right\|_1 \leq \left(1 + \frac{C_3 \|\boldsymbol{X}\|_2}{\sigma_{\min}(\boldsymbol{\Psi})}\right) \left\|\boldsymbol{\Psi}^\top (\boldsymbol{\beta} - \boldsymbol{u})\right\|_1 + C_3 \left\|\boldsymbol{X}(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}})\right\|_2. \tag{3.91a}$$

第二部分证明

$$f(\boldsymbol{\beta}) \triangleq \min \left\{ \left\|\boldsymbol{\Psi}^\top (\boldsymbol{\beta} - \boldsymbol{u})\right\|_1 \mid \boldsymbol{u} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top) \right\} \leq (1 - \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1} D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \overline{\boldsymbol{\beta}}). \tag{3.91b}$$

结合这两部分，并利用 C_4 的定义，即可得到所需结果。

((3.91a)的证明。 设 $\boldsymbol{u} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top)$，这也意味着 $\boldsymbol{\Psi}_{\mathbb{T}}^\top \boldsymbol{u} = \mathbf{0}$ 且 $\operatorname{supp}(\boldsymbol{\Psi}^\top \boldsymbol{u}) = \mathbb{S}$。根据三角不等式，我们得到

$$\left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta} - \boldsymbol{\Psi}^\top \overline{\boldsymbol{\beta}}\right\|_1 \leq \left\|\boldsymbol{\Psi}^\top (\boldsymbol{\beta} - \boldsymbol{u})\right\|_1 + \left\|\boldsymbol{\Psi}^\top (\boldsymbol{u} - \overline{\boldsymbol{\beta}})\right\|_1. \tag{3.92}$$

由于支撑集$(\boldsymbol{\Psi}^\top \overline{\boldsymbol{\beta}}) = \mathbb{S}$，这意味着 $\overline{\boldsymbol{\beta}} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top)$，因此我们有 $\overline{\boldsymbol{u}} - \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top)$，从而$\left\|\overline{\boldsymbol{u}} - \boldsymbol{\beta}\right\|_2 \leq \mathcal{U}(\mathbb{T}) \left\|\boldsymbol{X}(\overline{\boldsymbol{u}} - \boldsymbol{\beta})\right\|_2$，其中 $\mathcal{U}(\mathbb{T}) < +\infty$由条件 3.25(i) 得出。可以

可以证明，支撑集$(\boldsymbol{\Psi}^\top(\boldsymbol{u}-\boldsymbol{\beta}))=\mathbb{S}$，由此可得

$$\begin{aligned}\left\|\boldsymbol{\Psi}^\top(\boldsymbol{u}-\overline{\boldsymbol{\beta}})\right\|_1 &\le \sqrt{|\mathbb{S}|}\left\|\boldsymbol{\Psi}^\top(\boldsymbol{u}-\overline{\boldsymbol{\beta}})\right\|_2 \le \sqrt{|\mathbb{S}|}\left\|\boldsymbol{u}-\overline{\boldsymbol{\beta}}\right\|_2 \\ &\le \sqrt{|\mathbb{S}|}\mathcal{U}(\mathbb{T})\left\|\boldsymbol{X}(\boldsymbol{u}-\overline{\boldsymbol{\beta}})\right\|_2 = C_3\left\|\boldsymbol{X}(\boldsymbol{u}-\overline{\boldsymbol{\beta}})\right\|_2,\end{aligned} \tag{3.93}$$

其中，第一个不等式由向量范数的标准界（习题1.26）得到；我们使用了假设 $\sigma_{\max}(\boldsymbol{\Psi})=1$ （假设 3）以及定义 $C_3=\mathcal{U}(\mathbb{T})\sqrt{|\mathbb{S}|}$。此外，利用奇异值界（1.29）以及再次使用向量范数的标准界（习题 1.26），可得

$$\begin{aligned}&\left\|\boldsymbol{X}(\boldsymbol{u}-\overline{\boldsymbol{\beta}})\right\|_2 \le \left\|\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{u})\right\|_2 + \left\|\boldsymbol{X}(\boldsymbol{\beta}-\overline{\boldsymbol{\beta}})\right\|_2 \le \|\boldsymbol{X}\|_2\|\boldsymbol{\beta}-\boldsymbol{u}\|_2 + \left\|\boldsymbol{X}(\boldsymbol{\beta}-\overline{\boldsymbol{\beta}})\right\|_2 \\ &\le \frac{\|\boldsymbol{X}\|_2\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\|_2}{\sigma_{\min}(\boldsymbol{\Psi})} + \left\|\boldsymbol{X}(\boldsymbol{\beta}-\overline{\boldsymbol{\beta}})\right\|_2 \le \frac{\|\boldsymbol{X}\|_2\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\|_1}{\sigma_{\min}(\boldsymbol{\Psi})} + \left\|\boldsymbol{X}(\boldsymbol{\beta}-\overline{\boldsymbol{\beta}})\right\|_2.\end{aligned} \tag{3.94}$$

因此，结合（3.92）、（3.93）和（3.94）后，我们得到（3.91a）。

（3.91b）的证明。由于 $\boldsymbol{\alpha}_{\mathbb{S}}=\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{\beta})$和 $\mathbb{S}=\operatorname{supp}(\boldsymbol{\Psi}^\top\boldsymbol{\beta})$，有 $\langle\boldsymbol{\Psi}\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1$，因此 Bregman 距离可简化为 $D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta},\boldsymbol{\beta})=\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1-\langle\boldsymbol{\Psi}\boldsymbol{\alpha},\boldsymbol{\beta}\rangle$。因此，只需证明

$$f(\boldsymbol{\beta})\le(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}(\left\|\boldsymbol{\Psi}^\top\boldsymbol{\beta}\right\|_1-\langle\boldsymbol{\Psi}\boldsymbol{\alpha},\boldsymbol{\beta}\rangle).$$

由于 $\boldsymbol{u}\in\mathcal{N}(\boldsymbol{\Psi}(\mathbb{T})^\top)$ 等价于 $\boldsymbol{\Psi}_{\mathbb{T}}^\top\boldsymbol{u}=\boldsymbol{0}$，因此（3.91b）中最小化问题的拉格朗日函数为

$$L(\boldsymbol{u},\boldsymbol{\nu})=\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\|_1+\left\langle\boldsymbol{\nu},\boldsymbol{\Psi}_{\mathbb{T}}^\top\boldsymbol{u}\right\rangle=\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\|_1+\langle\boldsymbol{\Psi}_{\mathbb{T}}\boldsymbol{\nu},\boldsymbol{u}-\boldsymbol{\beta}\rangle+\langle\boldsymbol{\Psi}_{\mathbb{T}}\boldsymbol{\nu},\boldsymbol{\beta}\rangle.$$

于是，$f(\boldsymbol{\beta})=\min_{\boldsymbol{u}}\max_{\boldsymbol{\nu}}L(\boldsymbol{u},\boldsymbol{\nu})$。根据极小极大定理，我们推导出

$$\begin{aligned}f(\boldsymbol{\beta}) &= \max_{\boldsymbol{\nu}}\min_{\boldsymbol{u}}L(\boldsymbol{u},\boldsymbol{\nu}) = \max_{\boldsymbol{\nu}}\min_{\boldsymbol{u}}\left\{\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\|_1+\langle\boldsymbol{\Psi}_{\mathbb{T}}\boldsymbol{\nu},\boldsymbol{u}\rangle\right\} \\ &\overset{\dagger}{=} \max_{\boldsymbol{\nu}}\min_{\boldsymbol{u}}\max_{\boldsymbol{w}\in\mathbb{B}_\infty}\left\{\left\langle\boldsymbol{w},\boldsymbol{\Psi}^\top(\boldsymbol{\beta}-\boldsymbol{u})\right\rangle+\langle\boldsymbol{\Psi}_{\mathbb{T}}\boldsymbol{\nu},\boldsymbol{u}\rangle\right\} \\ &= \max_{\boldsymbol{z}\in\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})}\min_{\boldsymbol{u}}\max_{\boldsymbol{w}\in\mathbb{B}_\infty}\{\langle\boldsymbol{\Psi}\boldsymbol{w},\boldsymbol{\beta}\rangle+\langle\boldsymbol{u},\boldsymbol{z}-\boldsymbol{\Psi}\boldsymbol{w}\rangle\} \\ &\overset{\ddagger}{=} \max_{\boldsymbol{z}}\{\langle\boldsymbol{z},\boldsymbol{w}\rangle\mid\boldsymbol{z}\in\boldsymbol{\Psi}\cdot\mathbb{B}_\infty\cap\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})\} \\ &= \max_{\boldsymbol{z}}\{\langle c\boldsymbol{\Psi}\boldsymbol{\alpha}+\boldsymbol{z},\boldsymbol{\beta}\rangle\mid\boldsymbol{z}\in\boldsymbol{\Psi}\cdot\mathbb{B}_\infty\cap\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})\}-\langle c\boldsymbol{\Psi}\boldsymbol{\alpha},\boldsymbol{\beta}\rangle,\ \forall c>0 \\ &= c\max_{\boldsymbol{z}}\left\{\langle\boldsymbol{\Psi}\boldsymbol{\alpha}+\boldsymbol{z},\boldsymbol{\beta}\rangle\mid\boldsymbol{z}\in c^{-1}\boldsymbol{\Psi}\cdot\mathbb{B}_\infty\cap\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})\right\}-c\langle\boldsymbol{\Psi}\boldsymbol{\alpha},\boldsymbol{\beta}\rangle,\ \forall c>0,\end{aligned}$$

其中，等式（†）由对偶范数的定义（见（1.3））得出；等式（‡）是令 $\boldsymbol{z}-\boldsymbol{\Psi}\boldsymbol{w}=\boldsymbol{0}$ 得出的（否则，对 $\boldsymbol{u}$ 的最小值为 $-\infty$）；$\mathbb{B}_\infty\triangleq\{\boldsymbol{w}\mid\|\boldsymbol{w}\|_\infty\le1\}$ 表示单位 ℓ_∞-球，$\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})$ 表示 $\boldsymbol{\Psi}_{\mathbb{T}}$ 的列空间。令 $c\triangleq(1-\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty)^{-1}$ 和 $\mathbb{B}_{\mathbb{T}}\triangleq\{\boldsymbol{b}\in\mathbb{R}^q\mid\boldsymbol{b}_{\mathbb{S}}=\boldsymbol{0}\}$。由于 $\boldsymbol{\alpha}_{\mathbb{S}}=\operatorname{sign}(\boldsymbol{\Psi}_{\mathbb{S}}^\top\boldsymbol{\beta})$，且根据条件 3.25（ii）有 $\|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty<1$，因此我们有 $c<+\infty$，并得到

$$(\boldsymbol{\alpha}+c^{-1}\mathbb{B}_\infty\cap\mathbb{B}_{\mathbb{T}})\subset\mathbb{B}_\infty,$$

由此我们得出

$$(\boldsymbol{\Psi}\boldsymbol{\alpha}+c^{-1}\boldsymbol{\Psi}\cdot\mathbb{B}_\infty\cap\mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}}))\subset\boldsymbol{\Psi}\cdot\mathbb{B}_\infty.$$

因此，对于任意 $\boldsymbol{z} \in c^{-1}\boldsymbol{\Psi} \cdot \mathbb{B}_\infty \cap \mathcal{C}(\boldsymbol{\Psi}_{\mathbb{T}})$，均有 $\boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{z} \in \boldsymbol{\Psi} \cdot \mathbb{B}_\infty$，由 H¨older 不等式可推出 $\langle \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{z},\ \boldsymbol{\beta} \rangle \leq \left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}\right\|_1$。因此，$f(\boldsymbol{\beta}) \leq c(\left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}\right\|_1 - \langle \boldsymbol{\Psi}\boldsymbol{\alpha},\ \boldsymbol{\beta} \rangle)$。证明完毕。 ■

引理 3.33 设 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 为满足 $\operatorname{supp}(\boldsymbol{\Psi}^\top \boldsymbol{\beta}^*) = \mathbb{S}$的固定向量。假设存在一个次梯度 $\boldsymbol{\alpha} \in \partial \left\|\boldsymbol{\Psi}^\top \boldsymbol{\beta}^*\right\|_1$ 满足 $\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{X}^\top \boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{X}^\top)$，其中由$\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{X}^\top \boldsymbol{\lambda}$ 和 $\boldsymbol{X}$ 的满秩性（其中 $n \leq p$），可得 $\boldsymbol{\lambda} = (\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{\alpha}$。那么，对于每个 $\epsilon > 0$ 以及满足 $\|\boldsymbol{X}\boldsymbol{\beta}^* - \boldsymbol{y}\|_2 \leq \epsilon$的数据向量 $\boldsymbol{y}$，以下两个陈述成立：

1. 问题（$\mathrm{Q}_{1,\sigma}$）的每个极小化解 $\boldsymbol{\beta}_\sigma$ 都满足

$$D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta}_\sigma, \boldsymbol{\beta}^*) \leq \frac{(\epsilon + \sigma \|\boldsymbol{\lambda}\|_2 /2)^2}{\sigma} \quad \text{and} \quad \|\boldsymbol{X}\boldsymbol{\beta}_\sigma - \boldsymbol{y}\|_2 \leq \epsilon + \sigma \|\boldsymbol{\lambda}\|_2 ;$$

2. 问题（$\mathrm{Q}_{1,\epsilon}$）的每个极小化解 $\boldsymbol{\beta}_\epsilon$ 都满足 $D_1^{\boldsymbol{\alpha}}(\boldsymbol{\beta}_\epsilon, \boldsymbol{\beta}^*) \leq 2\epsilon \|\boldsymbol{\lambda}\|_2$。

证明 参见 Burger 和 Osher (2004年) 的定理 3 以及 Grasmair 等人(2011年) 的引理 3.5。 ■

根据上述引理，我们现在证明，条件 3.25 保证问题($\mathrm{Q}_{1,\sigma}$) 和($\mathrm{Q}_{1,\epsilon}$)对 $\boldsymbol{y}$ 中任意测量噪声具有鲁棒性。

定理 3.34：（测量噪声下 ($\mathrm{Q}_{1,\sigma}$) 和 ($\mathrm{Q}_{1,\epsilon}$)的误差界（Zhang 等人，2016））在假设 1∼3 下，令 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 为给定的信号，且令$\mathbb{S} \triangleq \operatorname{supp}(\boldsymbol{\Psi}^\top \boldsymbol{\beta}^*)$ 和 $\mathbb{T} \triangleq \mathbb{S}^c$。对于任意噪声 $\boldsymbol{\epsilon}$，令 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 和 $\epsilon = \|\boldsymbol{\epsilon}\|_2$。若条件 3.25 在 $\boldsymbol{\beta} = \boldsymbol{\beta}^*$下成立，则

(i) 对于任意 $C_0 > 0$，存在常数 $C_1 > 0$，使得问题 ($\mathrm{Q}_{1,\sigma}$) 的每个极小化解 $\boldsymbol{\beta}_\sigma$ 在参数 $\sigma = C_0\epsilon$ 下均满足

$$\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}_\sigma - \boldsymbol{\beta}^*)\right\|_1 \leq C_1\epsilon;$$

（ii）问题（$\mathrm{Q}_{1,\ \epsilon}$）的每个极小化解 $\boldsymbol{\beta}_\epsilon$ 均满足

$$\left\|\boldsymbol{\Psi}^\top(\boldsymbol{\beta}_\epsilon - \boldsymbol{\beta}^*)\right\|_1 \leq C_2\epsilon.$$

常数 C_1 和 C_2 的具体形式如下。定义

$$\boldsymbol{\lambda} = (\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{\alpha}, \quad C_3 = \mathcal{U}(\mathbb{T})\sqrt{|\mathbb{S}|}, \quad \text{and} \quad C_4 = \frac{1 + \operatorname{cond}(\boldsymbol{\Psi})\|\boldsymbol{X}\|_2 C_3}{1 - \|\boldsymbol{\alpha}_{\mathbb{T}}\|_\infty},$$

由此我们有

$$C_1 = 2C_3 + C_3C_0\|\boldsymbol{\lambda}\|_2 + \frac{(1 + C_0\|\boldsymbol{\lambda}\|_2/2)^2 C_4}{C_0},$$
$$C_2 = 2C_3 + 2C_4\|\boldsymbol{\lambda}\|_2 .$$

证明 [定理 3。34]首先，我们推导出

$$\begin{aligned}\left\|\boldsymbol{\Psi}^{\top}\boldsymbol{\beta}_{\sigma}-\boldsymbol{\beta}^{*}\right\|_{1} &\leq C_{3}\left\|\boldsymbol{X}(\boldsymbol{\beta}_{\sigma}-\boldsymbol{\beta}^{*})\right\|_{2}+C_{4}D_{1}^{\boldsymbol{\alpha}}(\boldsymbol{\beta}_{\sigma},\boldsymbol{\beta}^{*})\\ &\leq C_{3}\left\|\boldsymbol{X}\boldsymbol{\beta}_{\sigma}-\boldsymbol{y}\right\|_{2}+C_{3}\left\|\boldsymbol{X}\boldsymbol{\beta}^{*}-\boldsymbol{y}\right\|_{2}+C_{4}D_{1}^{\boldsymbol{\alpha}}(\boldsymbol{\beta}_{\sigma},\boldsymbol{\beta}^{*})\\ &\leq C_{3}(\epsilon+\sigma\left\|\boldsymbol{\lambda}\right\|_{2})+C_{3}\epsilon+C_{4}\frac{(\epsilon+\sigma\left\|\boldsymbol{\lambda}\right\|_{2}/2)^{2}}{\sigma},\end{aligned}$$

其中，第一个和第三个不等式分别由引理3.32 和3.33推出。代入 $\sigma=C_{0}\epsilon$ 并化简，即可得到第 (i) 部分的上界。对于约束形式，应用引理3.33 中相应的上界，可类似地推出第 (ii) 部分。

■

注意，条件 $\left\|\boldsymbol{\alpha}_{\mathbb{T}}\right\|_{\infty}<1$ 起着至关重要的作用：常数 C_{4} 与 $1-\left\|\boldsymbol{\alpha}_{\mathbb{T}}\right\|_{\infty}$ 成反比，因此这一间隙直接决定了稳定性。此外， C_{4} 的分子也取决于 $\boldsymbol{\Psi}$ 和 $\boldsymbol{X}$ 的全局性质，而 C_{3} 仅取决于支撑集 $\mathbb{S}$ 的受限几何。

第3章习题

1. **关于 的不可恢复性** $s>1$。设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 满足 $n<p$。证明存在一个 1-稀疏向量，它不是

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{p}}\left\|\boldsymbol{\beta}\right\|_{s}\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y},$$

的极小化解，其中 $s>1$。

2. **在唯一性条件下的** (P_{1}) **的线性无关性**。设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 为测量 ^矩阵。假设优化问题的极小化解 $\boldsymbol{\beta}$ 满足

$$(\mathrm{P}_{1})\quad\min_{\boldsymbol{\beta}\in\mathbb{R}^{p}}\left\|\boldsymbol{\beta}\right\|_{1}\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y},$$

^是唯一的。则列集合 $\{\boldsymbol{x}_{i},i\in\operatorname{supp}(\boldsymbol{\beta})\}$ 线性无关。特别地，

$$\left\|\widehat{\boldsymbol{\beta}}\right\|_{0}=\left|\operatorname{supp}(\widehat{\boldsymbol{\beta}})\right|\leq n.$$

3. ***S*-范数**。设 $\|\cdot\|$ 是 $\mathbb{R}^{p}$上的一个有效向量范数。再设 $\boldsymbol{S}\in\mathbb{R}^{n\times p}$ 具有满列秩。证明 $\|\cdot\|_{\boldsymbol{S}}\triangleq\|\boldsymbol{S}\cdot\|$ 也是一个有效范数，并满足定义 1.24中的三条范数公理。

4. **穿越恒等式**。设 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是两个 $n\times p$ 矩阵。证明以下结果：

$$\boldsymbol{X}^{\top}(\boldsymbol{I}_{n}+\boldsymbol{Y}\boldsymbol{X}^{\top})^{-1}=(\boldsymbol{I}_{p}+\boldsymbol{X}^{\top}\boldsymbol{Y})^{-1}\boldsymbol{X}^{\top}.\tag{3.95}$$

利用上述结果，证明对于任意 $n\times p$ 矩阵 $\boldsymbol{Z}$ 和标量 $\lambda>0$，有：

$$\boldsymbol{Z}^{\top}(\lambda\boldsymbol{I}_{n}+\boldsymbol{Z}\boldsymbol{Z}^{\top})^{-1}=(\lambda\boldsymbol{I}_{p}+\boldsymbol{Z}^{\top}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\top}.\tag{3.96}$$

穿越恒等式得名于这样的事实：我们从左侧推入一个矩阵，它会从右侧出来。提示：用适当的矩阵对上述恒等式进行前乘和后乘。

5. **交替最小二乘（ALS）**（Lu，2021a）。给定一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，使用最小二乘（带或不带正则化）模型求一个矩阵分解$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{Z}$。

6. **数据最小二乘**。普通最小二乘法考虑了响应变量 $\boldsymbol{y}$ 中的误差，而数据最小二乘法（*DLS*）则考虑预测变量中的误差：

$$\boldsymbol{\beta}^{DLS} = \arg\min_{\boldsymbol{\beta},\widetilde{\boldsymbol{X}}} \left\|\widetilde{\boldsymbol{X}}\right\|_F^2, \quad \text{s.t.} \quad \boldsymbol{y} \in \mathcal{C}(\boldsymbol{X}+\widetilde{\boldsymbol{X}}), \tag{3.97}$$

其中，$\widetilde{\boldsymbol{X}}$ 表示对 $\boldsymbol{X}$ 的扰动（即预测变量中的噪声）。也就是说，$(\boldsymbol{X}+\widetilde{\boldsymbol{X}})\boldsymbol{\beta}^{DLS} = \boldsymbol{y}$，假设测得的响应 $\boldsymbol{y}$ 无噪声。拉格朗日函数及其关于 $\boldsymbol{\beta}$ 的梯度分别为

$$\begin{aligned} L(\boldsymbol{\beta},\widetilde{\boldsymbol{X}},\boldsymbol{\lambda}) &= \operatorname{tr}(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^\top) + \boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}\boldsymbol{\beta}-\boldsymbol{y}); \\ \nabla_{\widetilde{\boldsymbol{X}}} L(\boldsymbol{\beta},\widetilde{\boldsymbol{X}},\boldsymbol{\lambda}) &= \widetilde{\boldsymbol{X}} + \boldsymbol{\lambda}\boldsymbol{\beta}^\top = \boldsymbol{0} \quad \Longrightarrow \quad \widetilde{\boldsymbol{X}} = -\boldsymbol{\lambda}\boldsymbol{\beta}^\top, \end{aligned}$$

其中，$\boldsymbol{\lambda} \in \mathbb{R}^n$ 是拉格朗日乘子向量。证明原问题等价于

$$\arg\min_{\boldsymbol{\beta}} \frac{(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})}{\boldsymbol{\beta}^\top\boldsymbol{\beta}}.$$

7. **总体最小二乘**。与数据最小二乘类似，总体最小二乘（*TLS*）考虑预测变量和响应变量中的误差。TLS 问题可表述为：

$$\boldsymbol{\beta}^{TLS} = \arg\min_{\boldsymbol{\beta},\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{y}}} \left\|[\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{y}}]\right\|_F^2, \quad \text{s.t.} \quad (\boldsymbol{y}+\widetilde{\boldsymbol{y}}) \in \mathcal{C}(\boldsymbol{X}+\widetilde{\boldsymbol{X}}), \tag{3.98}$$

其中，$\boldsymbol{X}$ 和 $\widetilde{\boldsymbol{y}}$ 分别是预测变量和响应变量中的扰动。证明该问题等价地可表述为

$$\boldsymbol{\beta}^{TLS} = \arg\min_{\boldsymbol{z},\boldsymbol{D}} \|\boldsymbol{D}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{D}\boldsymbol{z} = -\boldsymbol{C}\boldsymbol{z}, \tag{3.99}$$

其中 $\boldsymbol{C} \triangleq [\boldsymbol{X},\boldsymbol{y}] \in \mathbb{R}^{n\times(p+1)}$、$\boldsymbol{D} \triangleq [\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{y}}] \in \mathbb{R}^{n\times(p+1)}$ 和 $\boldsymbol{z} \triangleq \begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$。

8. 讨论 Θ_1、Θ_2 和 Θ 在 SAFE 规则的对偶空间中的几何关系。画出图形。你观察到了什么?

9. 严格证明 (3.63) 和 (3.64) 中的结果。

10. **带截距的LASSO的SAFE**。证明 SAFE–LASSO定理（定理 3.15）可应用于带截距的 LASSO 问题：

$$\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\gamma}-\boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

其中 $\boldsymbol{\gamma} \in \mathbb{R}^n$ 表示截距项。提示：使用变换。

11. **弹性网的SAFE**。证明 SAFE–LASSO 定理（定理3.15）可应用于弹性网问题：

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\sigma\|\boldsymbol{\beta}\|_2^2.$$

提示：使用一种变换。

12. 类似于推论3.14中的序贯DPP规则，设计一种利用 SAFE 规则来降低求解 LASSO问题时内存使用量的算法。

13. **拟合值的分段线性更新**。 考虑约束 LASSO 问题（L_C）（第 126页）：

$$\min_{\|\boldsymbol{\beta}\|_1 \leq \Sigma} f(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2.$$

假设两个约束半径 Σ_1 和 Σ_2位于解路径的同一区域内，在该区域中，活跃集 $\mathbb{S}=\{i:\beta_i\neq 0\}$ 和符号向量 $\boldsymbol{s}_{\mathbb{S}}$ 保持不变。这里，完整符号向量 $\boldsymbol{s}\in\mathbb{R}^p$ 按分量定义为

$$s_i=\begin{cases}\operatorname{sign}(\beta_i), & \beta_i\neq 0;\\ \in[-1,1], & \beta_i=0,\end{cases}$$

证明相应的解 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 满足

$$\boldsymbol{\beta}_2-\boldsymbol{\beta}_1=(\Sigma_2-\Sigma_1)\frac{(\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}{\boldsymbol{s}_{\mathbb{S}}^\top(\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}.$$

也就是说，解 $\boldsymbol{\beta}$ 沿固定方向随 Σ 线性变化

$$\boldsymbol{v}=\frac{(\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}{\boldsymbol{s}_{\mathbb{S}}^\top(\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}.$$

14. **对偶问题与原始最优解**。 引理3.18的逆命题成立吗?

15. 证明 (3.78) 的对偶问题为 (3.79).

16. 推导 $\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\|\boldsymbol{\beta}\|_2$ subject to $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}$ 的对偶问题

17. **Dantzig选择器问题的对偶证书**。 考虑以下 Dantzig 选择器（DS）问题($\mathrm{P}_{1,D}$):

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\quad\|\boldsymbol{\beta}\|_1\quad\text{s.t.}\quad\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})\right\|_\infty\leq D. \tag{3.100}$$

(a). 推导问题 (3.100) 的对偶问题。
(b). 提 供问题 (3.100) 的唯一最优解的对偶证书分析。

提示：参见定理*3.19*关于 (P_1)。

18. **组基追踪（GBP）**。考虑以下组基追踪（GBP）问题:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{pT}}\quad\sum_{t=1}^{T}\|\boldsymbol{\beta}_t\|_2\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}, \tag{3.101}$$

其中 $\boldsymbol{\beta}=[\boldsymbol{\beta}_1^\top,\boldsymbol{\beta}_2^\top,\ldots,\boldsymbol{\beta}_T^\top]^\top$，且 $\forall t\in\{1,2,\ldots,T\},\boldsymbol{\beta}_t\in\mathbb{R}^p$。
假设：矩阵 $\boldsymbol{X}$ 满行秩，且解集 $\mathbb{B}$（3.101）非空。

（a）推导问题（3.101）的对偶问题。（b）为问题（3.101）的唯一最优解提供对偶证书分析。

提示：参见定理*3.19*关于 (P_1)。

19. **关于 ℓ_s-最小化问题的 NP难性**（$0<s<1$（**Foucart 和 Rauhut，2013年**）。给定一个矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 和一个向量 $\boldsymbol{y}\in\mathbb{R}^n$，$\ell_s$-最小化问题旨在寻找一个向量 $\boldsymbol{\beta}\in\mathbb{R}^p$，使其 ℓ_s-拟范数在满足线性约束的条件下达到最小

$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$。假设划分问题的 NP完全性：给定整数 x_1、x_2、.....、x_k，判断指标集 $\{1,2,\ldots,k\}$ 是否可以拆分为两个不相交子集 $\mathbb{S}$和 $\mathbb{T} \subset \{1,2,\ldots,k\}$，使得 $\sum_{i\in\mathbb{S}} x_i = \sum_{j\in\mathbb{T}} x_j$；证明 ℓ_s-最小化问题是 NP 难。为建立这一归约，考虑如下构造。定义矩阵 $\boldsymbol{X}$ 和向量 $\boldsymbol{y}$ 如下：

$$\boldsymbol{X} \triangleq \begin{bmatrix} x_1 & x_2 & \ldots & x_k & x_1 & x_2 & \ldots & x_k \\ 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = [0,1,1,\ldots,1]^\top.$$

第二部分

设计矩阵

第4章

设计矩阵与性质

目录

4.1 零空间、受限特征值及其他设计性质 177

4.1.1 零空间性质-II.1784.1.2 稳定零空间性质.1804.1.3 鲁棒零空间性质.182

4.2 SPARK、互相干性与稀疏恢复的唯一性 184

4.2.1 Spark.1844.2.2 互相干性.1864.2.3 ℓ_1-相干性.1904.2.4 互相干性和 ℓ_1-相干性的界.193

4.3 受限等距性质（RIP）.196

4.3.1 受限正交性性质（ROP）.1994.3.2 RIC、ROC 与其他常数之间的关系 . . . . . 2014.3.3 受限等距性质的上界 .2044.3.4 限制等距性质与稳定性 .2074.3.5 限制等距性质、零空间性质及其变体.209**第4章习题.217**

In 稀疏表示和压缩感知（也称为压缩感知、压缩采样或稀疏采样；参见第3.1节），必须求解一个欠定线性系统

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y},$$

其中，$\boldsymbol{\beta}$ 是一个仅含少数非零项的稀疏向量，$\boldsymbol{y}$ 是测量向量（或响应），$\boldsymbol{X}$ 是设计矩阵（或感知矩阵）。长期以来，稀疏性一直应用于信号处理和逼近理论，例如用于图像压缩（Pennebaker 和 Mitchell，1992年；Taubman 等，2002年；Rabbani 和 Joshi，2002年；Foucart 和 Rauhut，2013年）和去噪（Donoho，2002年；Joy 等，2013年）；它也应用于统计学和学习理论，以防止过拟合（Vapnik，2013年；Lu, 2021b）。此外，稀疏性在统计估计和模型选择（Tibshirani，1996年；Lu, 2021b）、人类视觉系统建模（Olshausen 和 Field，1996年）以及图像处理中都发挥着核心作用；在图像处理中，多尺度小波变换能够为自然图像生成近似稀疏表示。本章将讨论设计矩阵在求解稀疏线性系统中的作用。

将信号变换到新的基或框架中——即将其表示为 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$，从而$\boldsymbol{\beta}$ 表示该信号在该基中的表示——可以得到更简洁的表示。这种压缩对于降低数据存储和传输成本很有价值，而在许多应用中这些成本可能相当可观。因此，与其传输原始的高维信号，不如只传输基或框架展开中（通常为数不多的）非零分析系数。当非零系数的数量较少时，我们称该信号具有稀疏表示。稀疏信号模型能够实现很高的压缩速率；在压缩感知中，关于信号在某个已知基或框架中是稀疏的先验知识，使我们能够用远少于其环境维数所暗示的测量值来重构原始信号。对于稀疏数据，通常只需存储或传输非零系数；其余系数都可以安全地假定为零。

从数学上讲，若信号 $\boldsymbol{\beta}$ 至多包含 k 个非零项，即 $\|\boldsymbol{\beta}\|_0 \leq k$，则称其为 k-稀疏。我们将所有 k-稀疏信号 的集合记为

$$\mathbb{B}_0[k] = \{\boldsymbol{\beta} : \|\boldsymbol{\beta}\|_0 \leq k\}, \tag{4.1}$$

[1] 这里使用方括号强调等号可以成立。采用这种记法，问题（S_0）（第125页）可改写为：

$$\min\ f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{B}_0[R].$$

在实践中，许多信号在标准基下并不稀疏，但在某个其他基 $\boldsymbol{\Phi}$ 下可以有稀疏表示。在这种情况下，我们仍称 $\boldsymbol{\beta}$ 为 k-稀疏，意即对于某个满足 $\|\boldsymbol{\alpha}\|_0 \leq k$ 的系数向量 $\boldsymbol{\alpha}$，有 $\boldsymbol{\beta} = \boldsymbol{\Phi}\boldsymbol{\alpha}$。

在许多应用中，稀疏优化涉及最小化目标函数

$$f(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2,$$

其中，$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是设计矩阵，$\boldsymbol{y} \in \mathbb{R}^n$ 是观测响应向量。值得注意的是，当设计矩阵为单位矩阵时，它保留了信号的几何结构，

1。回顾一下，$\mathbb{B}_0[k] = \mathbb{B}_0[\mathbf{0}, k]$ 表示所有 k-稀疏向量的集合（定义1.29）。尽管有时被非正式地称为“ℓ_0-球”，“ℓ_0-范数”并不是真正的范数。

从而能够精确恢复稀疏向量。我们将把这一思想形式化，并建立关于 $\boldsymbol{X}$ 的精确条件，以保证普遍恢复——也就是说，能够根据其测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$唯一重构每个 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 。

然而，能够识别某些稀疏向量的设计矩阵，并不一定能确保普遍恢复。例如，考虑一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和两个不同的k-稀疏向量 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2 \in \mathbb{B}_0[k]$ ，它们满足 $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$，但却有 $\boldsymbol{X}\boldsymbol{\beta}_1 = \boldsymbol{X}\boldsymbol{\beta}_2 = \boldsymbol{y}$。在这种情况下，测量值无法区分 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$，因此不可能实现唯一恢复。因此，这样的设计矩阵无法支持普遍恢复。此外，这种失败并不限于单对向量：它通常意味着整个子空间中的稀疏向量，或无穷多对稀疏向量，都无法根据其测量值被唯一识别。

因此，设计矩阵在将稀疏向量从高维空间（维数为 p）投影到低维空间（维数为 n）时，必须保留其几何结构，尤其是在 $n \ll p$的情况下。本章中的零空间性质及其他性质对这一要求进行了形式化描述。

基本记号

现在介绍一些基本记号。对于任意指标集 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$，定义以下集合：

$$\mathbb{C}[\mathbb{S}] \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 \leq \|\boldsymbol{\beta}_{\mathbb{S}}\|_1\}; \tag{4.2a}$$

$$\mathbb{C}[\mathbb{S};\gamma] \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 \leq \gamma \|\boldsymbol{\beta}_{\mathbb{S}}\|_1\}; \tag{4.2b}$$

$$\mathbb{C}[k] \triangleq \bigcup_{\mathbb{S}:|\mathbb{S}|=k} \mathbb{C}[\mathbb{S}], \tag{4.2c}$$

其中，$\mathbb{S}^c$ 表示 $\mathbb{S}$ 的补集。这里使用“方括号”（而非圆括号），表示定义中的不等式允许取等号。因此，$\mathbb{C}[\mathbb{S}] = \mathbb{C}[\mathbb{S};1]$ 是由满足以下条件的向量构成的凸锥：其在 $\mathbb{S}$ 之外的 ℓ_1-质量不超过在 $\mathbb{S}$ 之内的质量；也就是说，这些向量的大部分权重集中在 $\mathbb{S}$ 中的坐标上。相比之下，$\mathbb{C}[k]$ 是所有这类锥在大小为 k 的指标集上的并（非凸）；它刻画了这样一类向量：其大部分 ℓ_1-范数集中在 k 的某个子集上。注意，

$$\mathbb{B}_0[k] \subset \mathbb{C}[k],$$

因为任意 k-稀疏向量的所有非零项——因而其全部 ℓ_1-质量——都位于至多包含 k 个坐标的集合内。此外，对于固定的 $\mathbb{S}$，有

$$\mathbb{C}[\mathbb{S};\gamma_1] \subseteq \mathbb{C}[\mathbb{S};\gamma_2], \quad \text{whenever } \gamma_1 \leq \gamma_2.$$

在某些情形下，需要使用严格不等式。在这种情况下，可以用“圆括号”排除等号，从而重新表述上述定义：

$$\mathbb{C}(\mathbb{S}) \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 < \|\boldsymbol{\beta}_{\mathbb{S}}\|_1\}; \tag{4.3a}$$

$$\mathbb{C}(\mathbb{S};\gamma) \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 < \gamma \|\boldsymbol{\beta}_{\mathbb{S}}\|_1\}; \tag{4.3b}$$

$$\mathbb{C}(k) \triangleq \bigcup_{\mathbb{S}:|\mathbb{S}|=k} \mathbb{C}(\mathbb{S}), \tag{4.3c}$$

在进一步讨论这些性质之前，我们再引入一个定义。最佳 ℓ_s-误差 k-稀疏近似（也称为到 ℓ_s-稀疏向量 k-集合的距离）定义如下：

$$\sigma_k(\boldsymbol{\beta})_s \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]} \left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right\|_s. \tag{4.4}$$

4.1. 零空间、受限特征值及其他设计性质

现在，我们介绍文献中广泛研究的设计矩阵的若干理想性质（例如参见Candes 和 Tao (2005年)；Cohen 等人(2009年)；Raskutti 等人(2010年)；Baraniuk 等人(2011年)；Jain 等人(2014年)；Hastie 等人(2015年)；Jain 等人(2017年)）。

我们从零空间性质开始。

定义 4.1（零空间性质（NSP））。矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 若在集合零空间性质（*NSP*）或受限零空间性质（*RNP*）上满足 $\mathbb{S} \subseteq\{1,2,\ldots,p\}$，记作 NSP($\mathbb{S}$)，则称其满足

$$\mathcal{N}(\boldsymbol{X}) \cap \mathbb{C}[\mathbb{S}] = \{\mathbf{0}\}, \tag{4.5a}$$

其中 $\mathcal{N}(\boldsymbol{X}) \triangleq \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \boldsymbol{X}\boldsymbol{\beta} = \mathbf{0}\}$ 表示 $\boldsymbol{X}$ 的零空间（定义1.8）。等价地，该条件可表示为：

$$\text{or} \quad \|\boldsymbol{\beta}_{\mathbb{S}}\|_1 < \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1, \qquad \text{for all } \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}; \tag{4.5b}$$

$$\text{or} \quad 2\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 < \|\boldsymbol{\beta}\|_1, \qquad \text{for all } \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}. \tag{4.5c}$$

另一方面，$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 若满足阶数为 k的零空间性质（NSP），记作 NSP(k)，则 $\mathcal{N}(\boldsymbol{X}) \cap \mathbb{C}[k] = \{\mathbf{0}\}$。注意，对于给定的 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$，只要某个指标集 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 < \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1$ 满足 $\mathbb{S} \subset \{1,2,\ldots,p\}$，那么对任意满足 $|\mathbb{S}| \le k$ 的指标集 k，其中包含 $\boldsymbol{\beta}$绝对值最大的 k，阶数为4.5b或（4.5c）便可等价地表述为

$$\|\boldsymbol{\beta}\|_1 < 2\sigma_k(\boldsymbol{\beta})_1, \qquad \text{for all } \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}, \tag{4.5d}$$

其中，$\sigma_k(\boldsymbol{\beta})_s \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]} \left\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right\|_s$ 表示到 ℓ_s稀疏向量的 k距离。

阶为 k 的 NSP（即 NSP(k)）是相对于并集 $\mathbb{C}[k] = \bigcup_{\mathbb{S}:\ |\mathbb{S}|=k} \mathbb{C}[\mathbb{S}]$定义的，而 $\mathbb{S}$上的 NSP（即 NSP($\mathbb{S}$)）则适用于固定的指标集 $\mathbb{S} \subseteq \{1,2,\ldots,p\}$。因此，与阶为 k的 NSP 相比，$\mathbb{S}$上的 NSP 是一个较弱的条件，如果 $|\mathbb{S}| = k$。换言之，当锥 $\mathbb{C}[\mathbb{S}]$ 中唯一位于 $\boldsymbol{X}$ 的零空间内的向量是零向量时，NSP($\mathbb{S}$)性质成立。

回顾一下，阶为 k 的 NSP 表述如下：对于每个非零向量 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$ 以及每个满足 $|\mathbb{S}| \le k$的指标集 $\mathbb{S} \subset \{1,2,\ldots,p\}$，都有 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 < \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1$。换言之，$\boldsymbol{X}$ 的零空间中的任意非零向量，在任意至多包含 k 个坐标的子集之外，都必须比在该子集之内更加“分散”。这意味着，$\mathcal{N}(\boldsymbol{X})$中的非零向量不可能将其大部分 ℓ_1-质量集中在任意 k 个坐标上。直接 consecuencia 是，零空间不能包含任何非零 k稀疏向量。

练习 4.2。设 $\boldsymbol{X}$ 是满足 $2k$ 阶零空间性质的设计矩阵。证明，对于任意两个不同的 k-稀疏向量 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2 \in \mathbb{B}_0[k]$，若满足 $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$，则有 $\boldsymbol{X}\boldsymbol{\beta}_1 \neq \boldsymbol{X}\boldsymbol{\beta}_2$。

▶ **NSP 的缩放、重排和扩充性质**。考虑将原始测量矩阵 $\boldsymbol{X}$ 替换为新的测量矩阵 $\boldsymbol{X}$ 或 $\boldsymbol{X}$，定义如下：

如下:

$$\widehat{\boldsymbol{X}} \triangleq \boldsymbol{P}\boldsymbol{X}, \quad \text{where } \boldsymbol{P} \text{ is some invertible } n \times n \text{ matrix};$$
$$\widetilde{\boldsymbol{X}} \triangleq \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}, \quad \text{where } \boldsymbol{Y} \text{ is some } n' \times p \text{ matrix}.$$

当 $\boldsymbol{P}$ 为对角矩阵时, $\boldsymbol{P}\boldsymbol{X}$ 对应于 $\boldsymbol{X}$的缩放版本; 当 $\boldsymbol{P}$ 为置换矩阵（定义1.14）时, $\boldsymbol{P}\boldsymbol{X}$ 表示 $\boldsymbol{X}$的重新排列（即行置换）版本。我们注意到

$$\mathcal{N}(\widehat{\boldsymbol{X}}) = \mathcal{N}(\boldsymbol{X}) \qquad \text{and} \qquad \mathcal{N}(\widetilde{\boldsymbol{X}}) \subseteq \mathcal{N}(\boldsymbol{X}). \tag{4.6}$$

因此, 如果 $\boldsymbol{X}$满足 NSP 条件, 那么 $\boldsymbol{X}$ 和 $\boldsymbol{X}$也都满足该条件。

如果一个设计矩阵满足阶为 $2k$的 NSP, 那么不存在两个不同的 k-稀疏向量被映射到同一个测量向量（见习题4.2）。这种在 k-稀疏向量集合上的单射性, 对于保证稀疏恢复中的全局唯一性至关重要。比 NSP 更强的条件是受限特征值（RE）性质; 该性质不仅排除了零空间中的稀疏向量, 还确保稀疏向量在测量矩阵的作用下仍保留其范数的相当大一部分。

定义 4.3（受限特征值（RE）性质）。若矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足常数为 $\mu > 0$ 、作用于集合 $\mathbb{S}$上的受限特征值（RE）性质, 则对所有 $\boldsymbol{\beta} \in \mathbb{S}$, 有

$$\mu \le \frac{\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|_2^2} \qquad \Longleftrightarrow \qquad \mu \cdot \|\boldsymbol{\beta}\|_2^2 \le \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 ,$$

其中, 第一个不等式右侧表示向量 $\boldsymbol{\beta}$ 相对于矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$的瑞利商 [a] , 而该矩阵也作为最小二乘损失函数 $f(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 的海森矩阵。

[a] 更多详情请参见, 例如, Lu (2021a)。

这一性质意味着, 稀疏向量不仅不会落在 $\boldsymbol{X}$的零空间中, 而且在由 $\boldsymbol{X}$定义的线性变换下也具有良态: 它们的欧几里得范数至多相差一个因子 μ。特别地, 如果 RE 性质对所有 $2k$-稀疏向量组成的集合（即 $\mathbb{S} = \mathbb{B}_0[2k]$）成立, 那么任意两个 k-稀疏向量 $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2 \in \mathbb{B}[k]$之差 $\boldsymbol{\beta}_1\boldsymbol{\beta}_2$ 都是 $2k$-稀疏的, 并且有

$$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{B}_0[k] \qquad \Longrightarrow \qquad \|\boldsymbol{X}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\|_2^2 \ge \mu \cdot \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_2^2 . \tag{4.7}$$

因此, 映射 $\boldsymbol{\beta} \mapsto \boldsymbol{X}\boldsymbol{\beta}$ 对 k-稀疏向量之间的距离的保持误差不超过常数 μ。这种稳定性对于高维统计学和压缩感知中的可辨识性与稳健恢复都至关重要。

4.1.1 零空间性质-II

NSP 条件并非唯一能够确保零空间中的向量不集中的条件。在实践中, 无论是通过基追踪还是其他重建方法, 我们试图恢复的向量都很少是精确稀疏的。精确稀疏性通常只是一种理想化假设; 更现实地说, 信号只是近似稀疏的, 也就是说, 它们在某种范数下接近一个 k-稀疏向量。在这种情况下, 我们希望重建

误差能够由真实向量与其最佳 k-稀疏近似之间的差距来控制。这种理想性质称为相对于稀疏性缺陷的稳定性 。

为了使这一想法形式化，我们引入一种称为零空间性质-II（NSP′）的替代表述。

定义 4.4（零空间性质-II（NSP′））。矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 在集合 $\mathbb{S} \subset \{1,2,\ldots,p\}$ 上满足零空间性质′，其常数为 C，记作 NSP′($\mathbb{S}$, C)，若 $|\mathbb{S}| = k$ 且存在常数 $C > 0$ 使得

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \leq C\frac{\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1}{\sqrt{k}}, \quad \text{for all } \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}). \tag{4.8}$$

称其满足零空间性质′ 的阶为 k 、常数为 C，记作 NSP′(k, C)；若对于任意满足 $|\mathbb{S}| \leq k$ 的集合 $\mathbb{S} \subseteq \{1,2,\ldots,p\}$，它都满足常数为 C 的 NSP′。

请注意，在零空间性质-II′的定义中可以取等号，而标准 NSP 的定义中则不能取等号。

与标准 NSP 一样，NSP′ 条件将这样一种想法形式化：零空间 $\boldsymbol{X}$ 中没有非零向量会过度集中在一个小的坐标子集上。例如，假设 $\boldsymbol{\beta}$ 恰好是 k-稀疏的。那么存在一个集合 $\mathbb{S}$，满足 $|\mathbb{S}| \leq k$，使得$\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 = \mathbf{0}$。将其代入式（4.8）可得 $\boldsymbol{\beta}_{\mathbb{S}} = \mathbf{0}$，这意味着 $\boldsymbol{\beta} = \mathbf{0}$。因此，如果 $\boldsymbol{X}$ 满足 NSP′，那么 k中的唯一 $\mathcal{N}(\boldsymbol{X}$-稀疏向量是零向量 $\boldsymbol{\beta} = \mathbf{0}$。也就是说，$\mathcal{N}(\boldsymbol{X}) \cap \mathbb{B}_0[k] = \{\mathbf{0}\}$，这与标准 NSP 条件类似。

为了全面理解 NSP′ 在稀疏恢复中的意义，我们现在考虑当真实信号 $\boldsymbol{\beta}$ 不一定稀疏时，如何评估重建算法的性能。令 $\Delta\colon \mathbb{R}^n \to \mathbb{R}^p$ 表示一个特定的恢复映射（例如，解码器）。我们尤其关注如下形式的误差保证：

$$\|\Delta(\boldsymbol{X}\boldsymbol{\beta}) - \boldsymbol{\beta}\|_2 \leq C\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}, \quad \text{for all } \boldsymbol{\beta}, \tag{4.9}$$

[2] 其中 $\sigma_k(\boldsymbol{\beta})_1$ 表示最佳 ℓ_1-距离 k项近似：$\sigma_k(\boldsymbol{\beta})_1 \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]} \|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|_1$，也就是说，它表示向量 $\boldsymbol{\beta}$ 与 ℓ_1-范数下的球 $\mathbb{B}_0[k]$之间的距离。这个上界保证了两个重要性质：

- **精确恢复**所有可能的 k-稀疏解（即，如果 $\boldsymbol{\beta}' \in \mathbb{B}_0[k]$，则 $\boldsymbol{\beta}' = \Delta(\boldsymbol{X}\boldsymbol{\beta}')$ 。）**对**非
- 稀疏信号的鲁棒性，此时重构误差与 $\boldsymbol{\beta}$ 可由一个 k-稀疏向量近似到何种程度成正比。

这类保证称为实例最优(Cohen 等人，2009年；Foucart 和 Rauhut，2013年)，因为误差界会针对每个具体信号 $\boldsymbol{\beta}$进行调整，而不仅仅适用于某个受限类别（例如严格稀疏的向量）。这类保证也称为均匀保证，因为它们对所有可能的 $\boldsymbol{\beta}$都成立。

(4.9) 中范数的选择并非唯一。也可以使用其他 ℓ_s-范数来衡量重构误差。不过，范数的选择会影响可实现保证的强度以及相应的零空间条件。例如，若将 (4.9) 中的 ℓ_2-ℓ_1 结构替换为 ℓ_2-ℓ_2 上界，通常就需要显著增加测量值；事实上，若样本数量不大到难以承受，这种保证是不可能实现的；参见定理 5.19 或 Cohen 等人 (2009年)。因此，(4.9) 的形式代表了通过线性测量实现稳定恢复的最强可行均匀保证。此外，第 ℓ_1-ℓ_1 上界将在第 4.1.2 节中讨论。

2. 这也与实例最优性有关；定义 5.18。

以下结果改编自 Cohen 等人(2009年)，它表明，任何满足 (4.9) 的恢复算法，都必然要求测量矩阵 $\boldsymbol{X}$ 满足′ 阶为 $2k$ 的 NSP。

引理 4.5：（NSP 下的稀疏恢复′ （Cohen 等人，2009年）） 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$是一个测量矩阵，且 $\Delta:\ \mathbb{R}^n \to \mathbb{R}^p$ 是任意恢复映射。如果这一对（$\boldsymbol{X}$，Δ）满足误差界（4.9），那么 $\boldsymbol{X}$ 满足′ 阶零空间性质 $2k$。

引理 4.5] 的证明：令 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，并令 $\mathbb{S}$ 为与 $\boldsymbol{n}$ 中大小最大的 $2k$ 个元素对应的指标集。将 $\mathbb{S}$ 划分为两个不相交子集 $\mathbb{S}_0$ 和 $\mathbb{S}_1$，每个子集的大小均为 k。定义 $\boldsymbol{\beta} \triangleq \boldsymbol{n}(\mathbb{S}_1) + \boldsymbol{n}(\mathbb{S}^c)$ 和 $\boldsymbol{\beta}' \triangleq -\boldsymbol{n}(\mathbb{S}_0)$，使得 $\boldsymbol{n} = \boldsymbol{\beta} - \boldsymbol{\beta}'$。 [3] 根据构造，$\boldsymbol{\beta}' \in \mathbb{B}_0[k]$。因此，实例最优保证（4.9）意味着精确恢复：$\boldsymbol{\beta}' = \Delta(\boldsymbol{X}\boldsymbol{\beta}')$。此外，由于 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，我们有

$$\boldsymbol{X}\boldsymbol{n} = \boldsymbol{X}\left(\boldsymbol{\beta} - \boldsymbol{\beta}'\right) = \boldsymbol{0} \qquad \Longrightarrow \qquad \boldsymbol{X}\boldsymbol{\beta}' = \boldsymbol{X}\boldsymbol{\beta}. \tag{4.10}$$

由此可得 $\boldsymbol{\beta}' = \Delta(\boldsymbol{X}\boldsymbol{\beta})$。现在考虑重构误差：

$$\|\boldsymbol{n}_{\mathbb{S}}\|_2 \le \|\boldsymbol{n}\|_2 = \left\|\boldsymbol{\beta} - \boldsymbol{\beta}'\right\|_2 = \|\boldsymbol{\beta} - \Delta(\boldsymbol{X}\boldsymbol{\beta})\|_2 \le C\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}} = \sqrt{2}C\frac{\|\boldsymbol{n}_{\mathbb{S}^c}\|_1}{\sqrt{2k}}, \tag{4.11}$$

其中，最后一个不等式由（4.9）得出。证明完毕。 ■

4.1.2 稳定零空间性质

我们已经介绍了 NSP′ 性质，它为零空间中的向量提供了一个 ℓ_1-ℓ_2 上界。现在我们转而讨论一个基于 ℓ_1-ℓ_1 上界的类似条件。具体而言，我们将证明，基追踪在零空间性质的一个略微加强的版本下是稳定的，该版本称为稳定零空间性质（见定理 5.7）。

定义 4.6（稳定零空间性质）。 若矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 在集合 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 上满足常数为$0 < \rho < 1$的稳定零空间性质（稳定 *NSP*），记作 SNSP($\mathbb{S}, \rho$)，则

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \rho\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1, \quad \text{for all } \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}).$$

若对每个满足 $|\mathbb{S}| \le k$ 的子集$\mathbb{S} \subset \{1, 2, \ldots, p\}$，它都满足常数为 $0 < \rho < 1$ 的稳定 NSP，则称其满足阶数为 k 、常数为 $0 < \rho < 1$的稳定 NSP，记作 SNSP(k, ρ)。

再次注意，在稳定 NSP 的定义中可以取到等号，而在标准 NSP 的定义中则取不到等号。

以下结果表明，相对于集合 $\mathbb{S}$的 ℓ_1稳定 NSP，向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 与另一个 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 可产生相同测量值（$\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{X}\boldsymbol{\beta}$）的向量之间的 ℓ_1-范数之差，可以控制它们之间的距离。这一事实对于在稳定 NSP 下建立基追踪的误差界起着关键作用（见定理5.7）。

3。给定 $\boldsymbol{\beta} \in \mathbb{R}^p$，注意 $\boldsymbol{\beta}(\mathbb{S})$ 表示这是一个 $\mathbb{R}^p$ 向量，其对应于 $\mathbb{S}^c$ 的元素全为零，而对应于 $\mathbb{S}$ 的元素与 $\boldsymbol{\beta}$ 相同；参见定义 1.1。

定理4.7：（稳定 NSP 的标准推论）一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足关于 $\mathbb{S}$ 的稳定 NSP，其常数为 $0<\rho<1$，当且仅当

$$\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 \le \frac{1+\rho}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1) \tag{4.12}$$

对于所有向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{R}^p$，其中 $\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}$。

注意，如果 $\mathbb{S}$ 由 $\boldsymbol{\beta}$中大小最大的 k 个元素所对应的索引组成，那么$\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 = \sigma_k(\boldsymbol{\beta})_1$, 最佳 k 项 ℓ_1 逼近误差。因此，如果 $\boldsymbol{X}$ 满足阶为 k 、常数为 $0<\rho<1$ 的稳定 NSP，则有

$$\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 \le \frac{1+\rho}{1-\rho}\big(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\sigma_k(\boldsymbol{\beta})_1\big), \quad \text{for all } \boldsymbol{\alpha},\boldsymbol{\beta} \text{ with } \boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}. \tag{4.13}$$

在证明定理 4.7之前，我们先建立一个有用的辅助不等式。

引理 4.8 对于任意集合 $\mathbb{S}\subset\{1,2,\dots,p\}$ 和任意向量 $\boldsymbol{\beta}$，$\boldsymbol{\alpha}\in\mathbb{R}^p$，以下不等式成立：

$$\|(\boldsymbol{\beta}-\boldsymbol{\alpha})_{\mathbb{S}^c}\|_1 \le \|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + \|(\boldsymbol{\beta}-\boldsymbol{\alpha})_{\mathbb{S}}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 .$$

证明 [引理4.8]根据三角不等式，有

$$\|\boldsymbol{\beta}\|_1 = \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \|(\boldsymbol{\beta}-\boldsymbol{\alpha})_{\mathbb{S}}\|_1 + \|\boldsymbol{\alpha}_{\mathbb{S}}\|_1 ;$$
$$\|(\boldsymbol{\beta}-\boldsymbol{\alpha})_{\mathbb{S}^c}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_1 .$$

将两个不等式相加，即得所需结果。 ■

定理 4.7] $\Leftarrow$ 的证明 [。假设矩阵 $\boldsymbol{X}$ 对所有满足 $\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}$ 的向量$\boldsymbol{\alpha}$、$\boldsymbol{\beta}\in\mathbb{R}^p$ 都满足 (4.12)。给定一个向量 $\boldsymbol{u}\in\mathcal{N}(\boldsymbol{X})$，由于 $\boldsymbol{X}\boldsymbol{u}(\mathbb{S}^c) = -\boldsymbol{X}\boldsymbol{u}(\mathbb{S})$，我们可以对 $\boldsymbol{\beta}\triangleq -\boldsymbol{u}(\mathbb{S})$ 和 $\boldsymbol{\alpha}\triangleq \boldsymbol{u}(\mathbb{S}^c)$ 应用 (4.12)。由于 $\boldsymbol{\beta}(\mathbb{S}^c) = \boldsymbol{0}$，我们有

$$\|\boldsymbol{u}\|_1 \le \frac{1+\rho}{1-\rho}(\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 - \|\boldsymbol{u}_{\mathbb{S}}\|_1) \qquad \Longrightarrow \qquad \|\boldsymbol{u}_{\mathbb{S}}\|_1 \le \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 .$$

这证明了在 $\mathbb{S}$ 上具有常数 $0<\rho<1$ 的稳定 NSP。

$\Rightarrow$。反过来，假设矩阵 $\boldsymbol{X}$ 满足常数为 $0<\rho<$1、定义在 $\mathbb{S}$上的稳定 NSP。对于 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}\in\mathbb{R}^p$ 满足 $\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}$，由于 $\boldsymbol{u}\triangleq\boldsymbol{\alpha}-\boldsymbol{\beta}\in\mathcal{N}(\boldsymbol{X})$，稳定 NSP 给出$\|\boldsymbol{u}_{\mathbb{S}}\|_1 \le \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1$。将引理4.8 应用于 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha}=\boldsymbol{\beta}+\boldsymbol{u}$，可得

$$\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \le \|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + \|\boldsymbol{u}_{\mathbb{S}}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 .$$

将 $\|\boldsymbol{u}_{\mathbb{S}}\|_1 \le \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1$ 代入此不等式可得

$$\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \le \|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 .$$

由于 $\rho<1$，我们可以重新整理得到

$$\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \le \frac{1}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1).$$

再次使用$\|\boldsymbol{u}_{\mathbb{S}}\|_1 \le \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1$，我们得出

$$\|\boldsymbol{u}\|_1 = \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \|\boldsymbol{u}_{\mathbb{S}}\|_1 \le (1+\rho)\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \le \frac{1+\rho}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1).$$

证明至此完成。 ■

4.1.3 鲁棒零空间性质

在现实情境中，不可能以无限精度测量信号 $\boldsymbol{\beta} \in \mathbb{R}^p$ 。这意味着观测到的测量向量 $\boldsymbol{y} \in \mathbb{R}^n$ 只是对理想测量值 $\boldsymbol{X}\boldsymbol{\beta} \in \mathbb{R}^n$的近似，满足

$$\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\| \leq \epsilon$$

对某个 $\epsilon \geq 0$ 以及定义在 $\|\cdot\|$ 上的某个范数 $\mathbb{R}^n$ （通常是 ℓ_2-范数）。在这种情况下，重建算法应给出一个估计值 $\widehat{\ }$，使其与原向量的距离 $\boldsymbol{\beta} \in \mathbb{R}^p$ 受测量误差 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 控制。这种理想行为通常称为对测量误差的鲁棒性。我们将证明，若将标准基追踪问题（P_1）替换为凸优化程序（$\mathrm{P}_{1,\ \epsilon}$（第 131页））：

$$(\mathrm{P}_{1,\epsilon}) \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\| \leq \epsilon, \tag{4.14}$$

则在进一步强化零空间性质——即鲁棒零空间性质——的条件下，可以保证稳健恢复；更多结果参见定理5.8。

定义 4.9（ℓ_s-鲁棒零空间性质）。令 $s \geq 1$。若矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 ℓ_s-鲁棒零空间性质（ℓ_s-鲁棒 *NSP*），则称其具有阶数为 k （相对于 $\mathbb{R}^n$上的范数 $\|\cdot\|$），常数为 $0 < \rho < 1$和 $\tau > 0$，并且在集合 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 上成立，记为 RNSP(ℓ_s, $\mathbb{S}$, ρ, τ)，如果

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_s \leq \frac{\rho}{k^{(1-\frac{1}{s})}} \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \tau \|\boldsymbol{X}\boldsymbol{\beta}\|, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p. \tag{4.15a}$$

当 $s = 2$、$\rho = C$和 $\tau = 0$时，这就化简为 NSP′ （定义 4.4）。在特殊情形 $s = 1$ 下，矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 称为满足 稳健 *NSP*（相对于 $\|\cdot\|$），其常数为 $0 < \rho < 1$ 和 $\tau > 0$，定义在集合 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$上，记作 RNSP($\mathbb{S}$, ρ, τ)，如果

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \leq \rho \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \tau \|\boldsymbol{X}\boldsymbol{\beta}\|, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p. \tag{4.15b}$$

如果它满足 ℓ_s-稳健 *NSP* 或阶为 k、常数为 $0 < \rho < 1$和 $\tau > 0$ 的稳健 NSP，则称其满足相应条件；也就是说，它分别满足 ℓ_s-稳健 NSP，或相对于**任意**满足 $|\mathbb{S}| \leq k$ 的集合 $\mathbb{S} \subset \{1, 2, \ldots, p\}$、常数为 ρ, τ 的稳健 NSP。

请注意，稳健 NSP 推广了稳定 NSP（定义 4.6）：当 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$时，项 $\tau\|\boldsymbol{X}\boldsymbol{\beta}\|$ 消失，稳健 NSP 就恰好化简为稳定 NSP。

注 4.10。对于 $1 \leq t \leq s$，范数不等式 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_t \leq k^{(\frac{1}{t}-\frac{1}{s})}\|\boldsymbol{\beta}_{\mathbb{S}}\|_s$(参见问题 4.9) 表明，$\ell_s$-稳健 NSP 的常数为 $0 < \rho < 1$ 和 $\tau > 0$ 蕴含：对于任意满足 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 的 $|\mathbb{S}| \leq k$,

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_t \leq \frac{\rho}{k^{(1-\frac{1}{t})}} \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \tau k^{(\frac{1}{t}-\frac{1}{s})} \|\boldsymbol{X}\boldsymbol{\beta}\|, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p. \tag{4.16}$$

因此，对于 $1 \leq t \leq s$，ℓ_s-稳健 NSP 蕴含具有相同常数的 ℓ_t-稳健 NSP，只是在依赖测量的项上相差一个因子 $k^{(\frac{1}{t}-\frac{1}{s})}$。特别地，$\ell_s$-稳健 NSP 比 ℓ_1-稳健 NSP 是更强的条件，因为它以更强的范数（例如 ℓ_2 或 ℓ_∞）控制 $\mathbb{S}$ 上的信号，同时仍然只依赖于尾部的 ℓ_1-范数和测量残差。

注意，上述定义并不要求 $\boldsymbol{\beta}$ 属于 $\mathcal{N}(\boldsymbol{X}$ 的零空间。事实上，如果 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$，那么项 $\tau\|\boldsymbol{X}\boldsymbol{\beta}\|$ 就会消失。在这种情况下，稳健 NSP 恰好简化为定义4.6中给出的稳定 NSP。因此，稳健 NSP 确实推广了稳定 NSP。

与定理4.7类似，我们现在证明一个更强的等价性，它对任意固定指标集 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$均成立。

定理 4.11：（稳健 NSP 的标准推论）矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$相对于集合 $\mathbb{S}$ 满足常数为 $0 < \rho < 1$ 和 $\tau > 0$ 的稳健 NSP，当且仅当

$$\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_1 \leq \frac{1+\rho}{1-\rho}\big(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\big) + \frac{2\tau}{1-\rho}\|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\| \tag{4.17}$$

对所有向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$均成立。

证明 [定理 4.11] $\Leftarrow$。 证明过程与定理 4.7的证明类似。假设不等式（4.17）对所有 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$均成立。对于任意 $\boldsymbol{u} \in \mathbb{R}^p$，将$\boldsymbol{\beta} \triangleq -\boldsymbol{u}$（$\mathbb{S}$）和 $\boldsymbol{\alpha} \triangleq \boldsymbol{u}$（$\mathbb{S}^c$）代入（4.17）可得

$$\|\boldsymbol{u}\|_1 \leq \frac{1+\rho}{1-\rho}(\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 - \|\boldsymbol{u}_{\mathbb{S}}\|_1) + \frac{2\tau}{1-\rho}\|\boldsymbol{X}\boldsymbol{u}\| \qquad \Longrightarrow \qquad \|\boldsymbol{u}_{\mathbb{S}}\|_1 \leq \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \tau\|\boldsymbol{X}\boldsymbol{u}\|.$$

这证明了在 $\mathbb{S}$上，常数为 $0 < \rho < 1$ 和 $\tau > 0$ 的稳健 NSP。

$\Rightarrow$。 反之，假设矩阵 $\boldsymbol{X}$ 满足稳健 NSP，常数为 $0 < \rho < 1$ 且 $\tau > 0$ 定义在 $\mathbb{S}$上。令 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$ 为任意元素，并设 $\boldsymbol{u} \triangleq \boldsymbol{\alpha} - \boldsymbol{\beta}$。则 $\boldsymbol{u} \in \mathbb{R}^p$，由稳健 NSP 和引理 4.8 可得

$$\begin{aligned}\|\boldsymbol{u}_{\mathbb{S}}\|_1 &\leq \rho\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \tau\|\boldsymbol{X}\boldsymbol{u}\|;\\ \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 &\leq \|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + \|\boldsymbol{u}_{\mathbb{S}}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1.\end{aligned}$$

将第二个不等式中的项 $\|\boldsymbol{u}_{\mathbb{S}}\|_1$ 用第一个不等式代入，可得

$$\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \leq \frac{1}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 + \tau\|\boldsymbol{X}\boldsymbol{u}\|).$$

再次利用稳健 NSP 不等式，我们推导出

$$\begin{aligned}\|\boldsymbol{u}\|_1 = \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \|\boldsymbol{u}_{\mathbb{S}}\|_1 &\leq (1+\rho)\|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \tau\|\boldsymbol{X}\boldsymbol{u}\|\\ &\leq \frac{1+\rho}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1) + \frac{2\tau}{1-\rho}\|\boldsymbol{X}\boldsymbol{u}\|.\end{aligned}$$

证明至此完成。 ■

类似地，对于 ℓ_s-稳健 NSP，我们得到以下标准推论。该结果通过将 $s \geq t \geq 1$ 的 ℓ_1 误差估计替换为 ℓ_t误差估计，改进了先前的误差界。

定理 4.12：（ℓ_s-稳健 NSP 的标准推论）设 $1 \leq t \leq s$。假设矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 k、常数为 $0 < \rho < 1$和 $\tau > 0$的 ℓ_s-稳健 NSP。那么，对于所有 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$，

$$\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_t \leq \frac{C}{k^{(1-\frac{1}{t})}}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\sigma_k(\boldsymbol{\beta})_1) + Dk^{(\frac{1}{t}-\frac{1}{s})}\|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\|, \tag{4.18}$$

其中 $C \triangleq (1+\rho)^2/(1-\rho)$，且 $D \triangleq \tau(3+\rho)/(1-\rho)$。

证明 [定理 412]由（4.16），ℓ_s-稳健 NSP 蕴含 ℓ_1-稳健和 ℓ_t-稳健 NSP（$t \leq s$），其形式为：

$$\|\boldsymbol{u}_{\mathbb{S}}\|_1 \leq \rho \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \tau k^{(1-\frac{1}{s})} \|\boldsymbol{X}\boldsymbol{u}\|, \tag{4.19}$$

$$\|\boldsymbol{u}_{\mathbb{S}}\|_t \leq \frac{\rho}{k^{(1-\frac{1}{t})}} \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 + \tau k^{(\frac{1}{t}-\frac{1}{s})} \|\boldsymbol{X}\boldsymbol{u}\|, \tag{4.20}$$

对所有 $\boldsymbol{u} \in \mathbb{R}^p$ 以及所有 $\mathbb{S} \subset \{1,2,\ldots,p\}$ 满足 $|\mathbb{S}| \leq k$的情形。因此，根据（4.19），应用定理 4.11，并将 $\mathbb{S}$选为 k 中模最大的元素所构成的指标集，$\boldsymbol{\alpha}-\boldsymbol{\beta}$可得

$$\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 \leq \frac{1+\rho}{1-\rho}(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\sigma_k(\boldsymbol{\beta})_1) + \frac{2\tau}{1-\rho} k^{(1-\frac{1}{s})} \|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\|. \tag{4.21}$$

接着，将 $\mathbb{S}$ 选为 k 中模最大的元素所构成的指标集，$\boldsymbol{\alpha}-\boldsymbol{\beta}$，问题 4.2表明

$$\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_t = \|(\boldsymbol{\alpha}-\boldsymbol{\beta})_{\mathbb{S}^c}\|_t + \|(\boldsymbol{\alpha}-\boldsymbol{\beta})_{\mathbb{S}}\|_t \leq \frac{1}{k^{(1-\frac{1}{t})}} \|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 + \|(\boldsymbol{\alpha}-\boldsymbol{\beta})_{\mathbb{S}}\|_t.$$

现在将（4.20）中 $\boldsymbol{u} = \boldsymbol{\alpha}-\boldsymbol{\beta}$ 对应的界以及（4.21）中的界代入此不等式：

$$\begin{aligned}
\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_t &\overset{(4.20)}{\leq} \frac{1}{k^{(1-\frac{1}{t})}} \|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 + \frac{\rho}{k^{(1-\frac{1}{t})}} \|(\boldsymbol{\alpha}-\boldsymbol{\beta})_{\mathbb{S}^c}\|_1 + \tau k^{(\frac{1}{t}-\frac{1}{s})} \|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\| \\
&\leq \frac{1+\rho}{k^{(1-\frac{1}{t})}} \|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_1 + \tau k^{(\frac{1}{t}-\frac{1}{s})} \|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\| \\
&\overset{(4.21)}{\leq} C(\|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\beta}\|_1 + 2\sigma_k(\boldsymbol{\beta})_1) + D\|\boldsymbol{X}(\boldsymbol{\alpha}-\boldsymbol{\beta})\|.
\end{aligned}$$

证明至此完成。 ■

4.2. 火花、互相干性与稀疏恢复的唯一性

满行秩的欠定线性系统

$$\boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} = \boldsymbol{y}_{n\times 1}, \quad \text{with} \quad n < p,$$

具有无穷多个解。假设我们要寻找最稀疏解——也就是非零项最少的解。这种解是否可能是唯一的？我们又该如何找到最稀疏解？本节将简要讨论第一个问题（另见第 5 章以进一步了解）；第二个问题将在第8章和第10章中讨论。

4.2.1 SPARK

如果我们希望能够从测量值 $\boldsymbol{X}\boldsymbol{\beta}$ 中恢复所有稀疏信号 $\boldsymbol{\beta}$，那么对于任意两个不同的向量 $\boldsymbol{\beta}$，$\boldsymbol{\beta}' \in \mathbb{B}_0[k] = \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta}\|_0 \leq k\}$，都必须有 $\boldsymbol{X}\boldsymbol{\beta} \neq \boldsymbol{X}\boldsymbol{\beta}'$。否则，测量值 $\boldsymbol{y}$ 将无法让我们区分 $\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}'$。更形式化地说，如果$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}'$，那么 $\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}') = \boldsymbol{0}$，且其差满足 $\boldsymbol{\beta}-\boldsymbol{\beta}' \in \mathbb{B}_0[2k]$。因此，当且仅当其零空间 $\mathcal{N}(\boldsymbol{X})$ 不包含来自 $\mathbb{B}_0[2k]$的任何（非零）向量时，$\boldsymbol{X}$才唯一表示所有 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 。

这个条件可以通过几种等价的方式来刻画。其中最常见的一种是利用矩阵的 spark $\boldsymbol{X}$，这是由 Donoho 和 Elad (2003) 提出的一个关键概念，在分析稀疏解的唯一性时发挥着核心作用。

定义4.13（spark（Gorodnitsky 和 Rao，1997 年；Donoho 和 Elad，2003））。 给定一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，其 *spark*，记作 $\sigma = \text{spark}(\boldsymbol{X})$，定义为使得存在一个子集的最小可能数量，其中该子集包含 σ 个来自 $\boldsymbol{X}$ 的线性相关列。也就是说，

$$\sigma = \min\{s：\text{存在一个由 } \boldsymbol{X} \text{ 中 } s \text{ 列组成的子集，这些列线性相关}\}.$$

始终有

$$\sigma = \text{spark}(\boldsymbol{X}) \in [2, \min\{n, p\} + 1].$$

spark 与 *Kruskal*秩 密切相关，记作 $\text{rank}_k(\boldsymbol{X})$ 或 $k_{\boldsymbol{X}}$，其定义为满足以下条件的最大整数 r：每个由 r 列组成的集合 $\boldsymbol{X}$ 都是线性无关的（例如，参见 Lu (2021a)，了解更多详情）。回顾一下，矩阵的（标准）秩是 $\boldsymbol{X}$ 中线性无关列的最大数量（定义1.9）。虽然秩和 Kruskal 秩（或 spark）的定义看起来相似，但计算 spark 要困难得多：它需要对 $\boldsymbol{X}$ 的所有可能列子集进行组合搜索。根据定义，任何满足 $\boldsymbol{\beta}$ 的非零向量 $\boldsymbol{X\beta} = \boldsymbol{0}$ 至少有 $\text{spark}(\boldsymbol{X})$ 个非零项。换言之，

$$\|\boldsymbol{\beta}\|_0 \geq \text{spark}(\boldsymbol{X}) \qquad \text{if} \qquad \boldsymbol{X\beta} = \boldsymbol{0}, \text{ and } \boldsymbol{\beta} \neq \boldsymbol{0}, \tag{4.22}$$

因为少于 $\text{spark}(\boldsymbol{X})$ 列不可能形成非平凡的线性组合并使其等于零。

这一性质为欠定系统中稀疏解的唯一性提供了一个简单而有力的判据。

定理 4.14：（基于 spark 的唯一性（Gorodnitsky 和 Rao，1997 年；Donoho 和 Elad，2003；Bruckstein 等人，2009年）） 若线性方程组 $\boldsymbol{X\beta} = \boldsymbol{y}$ 存在一个解 $\boldsymbol{\beta}$ 满足 $\|\boldsymbol{\beta}\|_0 < \text{spark}(\boldsymbol{X})/2$，那么该解必然是最稀疏解（且是唯一的）。

证明 [定理4.14] 设 $\boldsymbol{\beta}$ 是满足 $\boldsymbol{X\beta} = \boldsymbol{y}$ 的任意另一个解。那么 $\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}) = \boldsymbol{0}$，因此 $\boldsymbol{\beta} - \boldsymbol{\beta}$ 属于 $\boldsymbol{X}$ 的零空间。由（4.22），

$$\|\boldsymbol{\beta}\|_0 + \|\widetilde{\boldsymbol{\beta}}\|_0 \geq \|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\|_0 \geq \text{spark}(\boldsymbol{X}). \tag{4.23}$$

由于我们有一个满足 $\|\boldsymbol{\beta}\|_0 < \text{spark}(\boldsymbol{X})/2$ 的解，因此可以得出，任何其他解 $\boldsymbol{\beta}$ 的非零元素个数必然大于 $\text{spark}(\boldsymbol{X})/2$，这证明了 $\boldsymbol{\beta}$ 是唯一的最稀疏解。

■

该定理还可以用另一种方式表述如下。

推论 4.15 对于任意向量 $\boldsymbol{y} \in \mathbb{R}^n$，至多存在一个解 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$，当且仅当 $\text{spark}(\boldsymbol{X}) > 2k$。

证明 [推论4。15] 我们首先假设，对于每个 $\boldsymbol{y} \in \mathbb{R}^n$，至多存在一个满足 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 的解 $\boldsymbol{y} = \boldsymbol{X\beta}$。反设 $\text{spark}(\boldsymbol{X}) \leq 2k$。根据 spark 的定义，这意味着存在一个至多包含 $2k$ 列的集合 $\boldsymbol{X}$，其中的列线性相关。因此，存在一个非零向量 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，使得 $\boldsymbol{n} \in \mathbb{B}_0[2k]$。此时，由于 $\boldsymbol{n} \in \mathbb{B}_0[2k]$，我们可以写成 $\boldsymbol{n} = \boldsymbol{\beta} - \boldsymbol{\beta}'$，其中 $\boldsymbol{\beta}$，$\boldsymbol{\beta}' \in \mathbb{B}_0[k]$，

由此可得 $\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}')=\mathbf{0}$，因此 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{X}\boldsymbol{\beta}'$。但这与“至多存在一个解 $\boldsymbol{\beta}\in\mathbb{B}_0[k]$ 使得 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}$”的假设矛盾。因此，我们必有 $\operatorname{spark}(\boldsymbol{X})>2k$。

反之，假设 $\operatorname{spark}(\boldsymbol{X})>2k$。设对于某个 $\boldsymbol{y}\in\mathbb{R}^n$，存在两个向量 $\boldsymbol{\beta},\boldsymbol{\beta}'\in\mathbb{B}_0[k]$ 使得 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{X}\boldsymbol{\beta}'$。那么 $\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}')=\mathbf{0}$，因此 $\boldsymbol{n}\triangleq\boldsymbol{\beta}-\boldsymbol{\beta}'\in\mathcal{N}(\boldsymbol{x})$。此外，由于 $\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}'$ 的非零元素至多有 k 个，它们的差满足 $\|\boldsymbol{n}\|_0\leq 2k$。然而，由于 $\operatorname{spark}(\boldsymbol{X})>2k$，$\boldsymbol{X}$ 的零空间不包含范数为 $\leq 2k$ 的非零向量。因此，$\boldsymbol{n}=\mathbf{0}$，这意味着 $\boldsymbol{\beta}=\boldsymbol{\beta}'$。这证明了解的唯一性，从而完成证明。 ■

显然，矩阵的 spark 信息量很大：$\operatorname{spark}(\boldsymbol{X})$ 的值越大，就越能保证较稀疏解的唯一性。spark 最大能有多大？根据定义，$2\leq\operatorname{spark}(\boldsymbol{X})\leq\min\{n,p\}+1$。例如，如果 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 的元素独立地取自某个连续分布（例如高斯分布），那么以概率 1，任意由 $\min\{n,p\}$ 列组成的子集都线性无关。特别地，当 $n<p$时，几乎必然有 $\operatorname{spark}(\boldsymbol{X})=n+1$，因为在 $\mathbb{R}^n$中，线性无关的列至多有 n 个。因此，对于非零元素少于 $(n+1)/2$ 个的任意解，都能保证其唯一性。

4.2.2 互相干性

虽然 spark、标准 NSP 及其鲁棒变体都能为稀疏信号的恢复提供理论保证（详见第 5.1 节），但验证一个一般矩阵 $\boldsymbol{X}$ 是否满足这些性质中的任意一种，在计算上都是不可行的。具体而言，每个条件都需要检查数量级为 $\binom{p}{k}$ 的子矩阵，从而导致组合的复杂度。例如，定理 4.14表明，对于线性方程组 $\boldsymbol{\beta}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}$的任意最稀疏解 $\|\boldsymbol{\beta}\|_0\leq\operatorname{spark}(\boldsymbol{X})/2$都成立。然而，计算矩阵的 spark 至少与求解（ ℓ_0-最小化问题（P_0）一样困难（第 129页）。因此，在许多实际场景中，最好使用易于计算的 $\boldsymbol{X}$性质，同时仍能提供有意义的恢复保证。矩阵的互相干性就是这样一种性质。

互相干性提供了一种简单而有效的度量，用于衡量测量矩阵是否适合稀疏恢复。一般来说，互相干性越小，恢复算法的性能越好。

定义 4.16（互相干性（Mallat 和 Zhang，1993 年；Donoho 等人，2001年；Bruckstein 等人，2009年））。给定矩阵的互相干性（也称为成对相干性或简称相干性）定义为 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 任意两列不同列之间归一化内积绝对值的最大值。记 $\boldsymbol{X}$中的第 i 列为 $\boldsymbol{X}$，$\boldsymbol{x}_i$则互相干性表示为

$$\mu(\boldsymbol{X})=\max_{1\leq i,j\leq p,i\neq j}\frac{|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|}{\|\boldsymbol{x}_i\|_2\|\boldsymbol{x}_j\|_2}\in[0,1]. \tag{4.24}$$

$\boldsymbol{X}$ 的互相干性刻画了其各列之间线性相关的程度。它量化了任意一对不同列之间的最大相似度：相干性越低，各列就越接近正交。这一点很有利，因为这意味着每一列都能提供独特的信息——这是实现可靠稀疏表示和恢复的关键要求。

对于正交矩阵（定义1.39），其列两两正交归一，因此互相干性恰好为零。相反，对于 $p>n$

（即列数多于行数的情况），$\mu(\boldsymbol{X})$ 必须严格为正。在这种情况下，我们的目标是尽可能减小 $\mu(\boldsymbol{X})$，以尽可能复现正交矩阵的优良性质。

为说明两种极端情况，现在考虑一个满足 $n \geq p$ 的矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。在这种情况下，$\mu(\boldsymbol{X}) = 0$ 当且仅当 $\boldsymbol{X}$ 的各列正交归一。特别地，当 $\boldsymbol{X}$ 为方阵（$n = p$）时，这恰好意味着 $\boldsymbol{X}$ 是一个正交矩阵。

然而，在压缩感知中，我们主要关注满足 $n < p$ 的欠定情形。在此情形下，所有列不可能彼此完全正交，而且相干性能够达到多小也存在基本限制。粗略而言，较小的相干性意味着由适量列构成的子矩阵具有良好的条件数，这对稳定稀疏恢复至关重要。可以证明，任何满足 $p > n$ 的矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，其互相干性总满足

$$\mu(\boldsymbol{X}) \in \left[\sqrt{\frac{p-n}{n(p-1)}}, 1\right].$$

这个下界称为*Welch*界（Rosenfeld，1997 年；Strohmer和Heath Jr，2003；Welch，2003）；定理4.27给出了证明。注意，当$p \gg n$时，这个上界近似简化为 $\mu(\boldsymbol{X}) \geq 1/\sqrt{n}$。

换言之，互相干性从逐元素的层面反映了格拉姆矩阵 $\boldsymbol{G} = \boldsymbol{X}^\top \boldsymbol{X}$ 与 p 维单位矩阵 $\boldsymbol{I}$ 的接近程度。

引理 4.17：（互相干性相对于谱范数的解释）设$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个列范数均为 1 的矩阵。则

$$\left\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\right\|_2 \leq \max_{i\in\{1,2,\ldots,p\}} \sum_{j\in\{1,2,\ldots,p\}\setminus\{i\}} \left|\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle\right|, \tag{4.25}$$

其中，$\|\cdot\|_2$ 表示矩阵的谱范数。

引理 [4. 的证明。17] 设 $\boldsymbol{H} \triangleq \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}$。根据格尔什戈林圆盘定理（定理 1.3），每个特征值 λ of $\boldsymbol{H}$ 都满足 $|\lambda - h_{ii}| \leq \sum_{j\neq i} |h_{ij}|$. 由于 $\boldsymbol{H}$ 的对角元素为零（$h_{ii} = 0$），因此可简化为

$$|\lambda| \leq \sum_{j\neq i} \left|\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle\right|.$$

对所有 $i \in \{1, 2, \ldots, p\}$ 取最大值，可得

$$\|\boldsymbol{H}\|_2 = \max_{\lambda\in\Lambda(\boldsymbol{H})} |\lambda| \leq \max_{i\in\{1,2,\ldots,p\}} \sum_{j\neq i} \left|\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle\right|.$$

其中，$\Lambda(\boldsymbol{H})$ 表示 $\boldsymbol{H}$ 的谱。证明完毕。 ■

比互相干性更强的条件是互不相干条件。此条件是针对任意子集 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 定义的。

定义 4.18（互非相干性）。矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的互非相干性（也称为不可表示条件）对于子集 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 成立，当且仅当

存在某个 $\gamma > 0$，使得

$$\max_{i \in \mathbb{S}^c} \left\| (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1} \boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{x}_i \right\|_1 \leq 1 - \gamma. \tag{4.26}$$

为了解释这一条件，请注意，子矩阵 $\boldsymbol{X}_{\mathbb{S}} \in \mathbb{R}^{n \times k}$（其中 $|\mathbb{S}| = k$）由 $\boldsymbol{X}$ 中以支撑集 $\mathbb{S}$ 为索引的列组成。对于补集 $\mathbb{S}^c$ 中的每个索引 i，向量 $(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1} \boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{x}_i$ 包含将 $\boldsymbol{x}_i$ 回归到 $\boldsymbol{X}_{\mathbb{S}}$ 的各列上所得到的系数。该向量的 ℓ_1 范数度量了 $\boldsymbol{x}_i$ 能够在多大程度上表示为 $\boldsymbol{X}_{\mathbb{S}}$ 中各列的线性组合。在理想情况下，每个满足 $i \in \mathbb{S}^c$ 的列 $\boldsymbol{x}_i$ 都应与 $\boldsymbol{X}_{\mathbb{S}}$ 中的所有列正交。在这种情况下，回归系数将全部为零，因此有 $\gamma = 1$。当然，在满足 $p \gg n$ 的高维场景中，这种完美的正交性是不可能实现的。然而，我们仍然可以期待某种形式的“近似正交”，而当 γ 接近 1 时，这正是互不相干条件所刻画的性质。

值得注意的是，“互不相干性”有时也被非正式地用来指列之间的两两相关性较小——也就是说，互相干性较小 $\mu(\boldsymbol{X})$。然而，定义 4.18 中的正式互不相干条件要严格得多，因为它涉及列组的整体行为，而不只是两两交互。

互相干性有时可以与其他基本性质联系起来，例如 spark、零空间性质以及受限等距性质（RIP；作为后续参考，参见定义 4.30）。例如，spark 与互相干性通过格尔什戈林圆盘定理（定理 1.3）相关联，如下面的命题所示。

命题 4.19：（spark & 互相干性）对于任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，有

$$\operatorname{spark}(\boldsymbol{X}) \geq 1 + \frac{1}{\mu(\boldsymbol{X})}. \tag{4.27}$$

命题 [的证明 4.19] 首先，将 $\boldsymbol{X}$ 的各列归一化为单位 ℓ_2范数，得到一个新矩阵 $\widetilde{\boldsymbol{X}}$。此操作保留 spark 和互相干性。令$\boldsymbol{G} = \widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}}$ 为对应的格拉姆矩阵。其元素满足：

$$\{g_{kk} = 1,\ \forall 1 \leq k \leq p\} \quad \text{and} \quad \{|g_{ij}| \leq \mu(\boldsymbol{X}),\ \forall 1 \leq i, j \leq p,\ i \neq j\}.$$

现在考虑任意一个 $l \times l$ 主子矩阵，其对应于 $\boldsymbol{G}$的一个子集 l 列，这些列取自 $\widetilde{\boldsymbol{X}}$（并计算它们的子格拉姆矩阵）。根据格尔什戈林圆盘定理（定理 1.3 或定理 1.20），如果该子式是对角占优的（即，如果 $\sum_{j \neq i} |g_{ij}| < |g_{ii}|$对每个 i都成立），那么 $\boldsymbol{G}$ 的这个子矩阵是正定的，因此这些 l 列取自 $\widetilde{\boldsymbol{X}}$是线性无关的，并满足：

- $g_{ii} = 1,\ 1 \leq i \leq l$;
- $|g_{ij}| \leq \mu(\boldsymbol{X}),\ 1 \leq i, j \leq l,\ i \neq j$.

由于 $l < \operatorname{spark}(\boldsymbol{X})$，spark 条件和正定性意味着 $(l-1)\mu(\boldsymbol{X}) < 1$或者，等价地，$l < 1 + 1/\mu(\boldsymbol{X})$。因此，我们可以选择 l 使得 $l < 1 + 1/\mu(\boldsymbol{X}) \leq l + 1 \leq \operatorname{spark}(\boldsymbol{X})$，从而得到 $\operatorname{spark}(\boldsymbol{X}) \geq 1 + 1/\mu(\boldsymbol{X})$。 ■

通过结合定理4.14 或推论 4.15 与命题 4.19，我们得到以下关于 $\boldsymbol{X}$ 的条件，该条件可保证稀疏解的唯一性。

定理 4.20：（互相干性下的唯一性（Donoho 等人，2001年；Donoho 和 Elad，2003））考虑线性方程组 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$。如果该方程组存在一个解 $\boldsymbol{\beta}$ 满足

$$\|\boldsymbol{\beta}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{X})}\right),$$

那么解 $\boldsymbol{\beta}$ 必然是（唯一的）最稀疏解。等价地，如果

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{X})}\right),$$

那么对于每个测量向量 $\boldsymbol{y} \in \mathbb{R}^n$，至多存在一个信号 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 满足 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$。

这个定理再次部分回答了以下基本问题：

- 何时能够保证最稀疏解的唯一性？能否验证一个候选解对于（P_0）是
- 全局最优的？

我们已经看到，任何足够稀疏的解——具体而言，比上述界更稀疏的解——都能保证在所有稀疏程度相近的解中具有唯一性。因此，这样的解也必然是（P_0）的全局极小点。这表明，寻求稀疏解并非只是启发式方法；在适当条件下，它能形成一个适定的逆问题，并具有强有力的理论保证。

比较定理 4.14 和 4.20，我们发现它们在结构上相似，但依赖于不同的假设。定理 4.14基于 spark，具有锐性，通常远强于定理 4.20；后者仅依赖于互相干性，而互相干性为 spark 提供了一个下界。

事实上，根据 Welch 界（即 $1/\sqrt{n}$（由此可知，$\mu(\boldsymbol{X}) \in \left[\sqrt{\frac{p-n}{n\,p-1}}(),\ 1\right])^{-}$，互相干性不会小于约 $1/\sqrt{n}$(as implied bythe Welch bound, i.e., $\mu(\boldsymbol{X}) \in \left[\sqrt{\frac{p-n}{n\,p-1}}(),\ 1\right]$，因此定理 4.20所保证的稀疏度大致不会超过 $\sqrt{n}/2$。也就是说，定理 4.20结合 Welch 界，为稀疏度 k 给出了一个上界，确保利用相干性可保证唯一性：$k = \mathcal{O}(\sqrt{n})$。相比之下，spark 最大可达 n（假设 $p > n$），此时定理 4.14允许稀疏度达到 $n/2$——这是一个宽松得多的上界。

现在我们证明，互相干性也蕴含 NSP 条件。

命题 4.21：（相干性 $\Longrightarrow$ NSP）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。假设对于某个整数 $k \in \{1, 2, \ldots, p\}$，互相干性满足上界 $\mu(\boldsymbol{X}) < \frac{1}{3k}$。那么 $\boldsymbol{X}$满足 k 阶 NSP。[a]

[a]。因此，根据推论 5.3，ℓ_1-最小化（P_1）能够精确恢复每个 k-稀疏向量 .

命题的证明 [4.任意性不失一般性，我们可以假设 $\boldsymbol{X}$ 的所有列均已归一化为单位 ℓ_2-范数，即 $\|\boldsymbol{x}_i\|_2 = 1$ 对所有 i（因为互相干性和 NSP 在列缩放下保持不变）。为简化分析，我们更一般地假设 $\mu(\boldsymbol{X}) < \frac{\delta}{k}$ 对于某个 $\delta > 0$，并确定能够保证 NSP 的最小 δ。我们将看到，$\delta = 1/3$ 足以满足要求。

设 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 为任意子集，且 $|\mathbb{S}| = k$，并设 $\boldsymbol{\beta} \in \mathbb{C}[\mathbb{S}] \setminus \{\boldsymbol{0}\}$。为验证 NSP $\mathcal{N}(\boldsymbol{X}) \cap \mathbb{C}[\mathbb{S}] = \{\boldsymbol{0}\}$，只需证明 $\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 > 0$，因此我们从下界开始

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 = \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}} + \boldsymbol{X}_{\mathbb{S}^c}\boldsymbol{\beta}_{\mathbb{S}^c}\|_2^2 \geq \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 + 2\boldsymbol{\beta}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}^c}\boldsymbol{\beta}_{\mathbb{S}^c}.$$

现在我们对交叉项取上界。利用互相干性的定义和三角不等式，

$$2\left|\boldsymbol{\beta}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}^c}\boldsymbol{\beta}_{\mathbb{S}^c}\right| \leq 2\left|\sum_{i\in\mathbb{S}}\sum_{j\in\mathbb{S}^c}|\beta_i|\,|\beta_j|\,\langle \boldsymbol{x}_i,\boldsymbol{x}_j\rangle\right|$$
$$\overset{(i)}{\leq} 2\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_1\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,\mu(\boldsymbol{X}) \overset{(ii)}{\leq} \frac{2\delta\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_1^2}{k} \overset{(iii)}{\leq} 2\delta\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2,$$

其中，不等式 (i) 根据互相干性的定义得出；不等式 (ii) 利用了关于 $\mu(\boldsymbol{X})$ 的假设上界以及以下事实： $\boldsymbol{\beta} \in \mathbb{C}[\mathbb{S}]$；不等式 (iii) 则根据以下事实得出：$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \leq \sqrt{k}\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2$ （习题 1.26），因为 $\mathbb{S}$ 的基数至多为 k。结合这两个不等式，可得

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \geq \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 - 2\delta\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2. \tag{4.28}$$

下面需要对 $\|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$ 给出下界。根据引理 4.17，有

$$\left\|\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \leq \max_{i\in\mathbb{S}}\sum_{j\in\mathbb{S}\setminus\{i\}}|\langle \boldsymbol{x}_i,\boldsymbol{x}_j\rangle| \leq k\frac{\delta}{k} = \delta.$$

这表明 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}-\boldsymbol{I}$的所有特征值都位于区间 $[-\delta,\delta]$ 和 $-\delta\boldsymbol{I} \preceq \boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}-\boldsymbol{I} \preceq \delta\boldsymbol{I}$内。因此，$\|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \geq (1-\delta)\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$，结合上界（4.28），我们可以得出$\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 > (1-3\delta)\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$，从而 $\delta = 1/3$保证严格正性，这就证明了 $\boldsymbol{X}$满足 k阶的 NSP。 ■

此外，格尔什戈林圆盘定理（定理 1.3）的另一项直接应用，将互相干性与限制等距性质联系起来（参见定义 4.30 ）。

命题 4.22：（相干性 $\Longrightarrow$限制等距性质）若 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的列均为单位范数，且相干性为 $\mu = \mu(\boldsymbol{X})$，则 $\boldsymbol{X}$满足阶数为 k、 $\delta = (k-1)\mu$的限制等距性质，对所有 $k < 1/\mu$均成立。

证明沿用命题 4.19中的相同推理，利用子格拉姆矩阵的对角占优性。

这些结果凸显了较小的互相干性在压缩感知测量矩阵设计中的重要性。确定性构造和随机构造的相干性界都已得到广泛研究。例如，存在尺寸为$n \times n^2$ 的显式确定性矩阵，可达到韦尔奇下界 $\mu(\boldsymbol{X}) = 1/\sqrt{n}$，例如由 Alltop 序列生成的加博尔框架（Herman 和 Strohmer，2009年）以及等角紧框架（Tropp，2008年）。使用这类矩阵，可以通过 k 稀疏信号实现稳定恢复，所需测量值数量为$n = \mathcal{O}(k^2 \ln p)$。

相比之下，对于元素独立同分布、均值为零且方差有限的随机矩阵，在渐近情形（n, $p \to\infty$）下，相干性集中于 $\mu(\boldsymbol{X}) = \sqrt{(2\ln p)/n}$ 附近（Cai and Jiang, 2011; Cand, 2008; Donoho, 2006）。将其代入基于相干性的稀疏性上界，可得 $k = \mathcal{O}(\sqrt{n}/(\ln p))$，重新整理后又可得到$n = \mathcal{O}(k^2 \ln p)$。不过，本书不再深入探讨这些细节，而与 RIP 相关的结果将在第 6 章和第 7 章介绍。

4.2.3 ℓ_1-相干性

比互相干性更一般的概念是 ℓ_1-相干性，也称为 Babel 函数。

定义 4.23（ℓ_1-相干性（Foucart 和 Rauhut，2013年））。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。矩阵 ℓ_1-相干性函数（或*Babel* 函数） μ_1 定义为矩阵 $\boldsymbol{X}$ 在 $1 \le k \le p-1$ 上的

$$\mu_1(\boldsymbol{X}, k) = \max_{\substack{\mathbb{S}\subset\{1,\dots,p\} \\ |\mathbb{S}|=k}} \max_{j\notin\mathbb{S}} \left\{ \sum_{i\in\mathbb{S}} \frac{|\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle|}{\|\boldsymbol{x}_i\|_2 \|\boldsymbol{x}_j\|_2} \right\}.$$

不难验证，对于 $1 \le k \le p-1$,

$$\mu(\boldsymbol{X}) \le \mu_1(\boldsymbol{X}, k) \le k \cdot \mu(\boldsymbol{X}), \tag{4.29a}$$

更一般地，对于 $1 \le k,\ t \le p-1$ 且 $k+t \le p-1$,

$$\max\{\mu_1(\boldsymbol{X}, k), \mu_1(\boldsymbol{X}, t)\} \le \mu_1(\boldsymbol{X}, k+t) \le \mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, t). \tag{4.29b}$$

互相干性刻画了 $\boldsymbol{X}$ 任意两个不同列之间的最坏情况两两相关性，仅衡量最大的绝对值归一化内积。相比之下，ℓ_1-相干性量化了任意一组 k 列对另一列所能产生的总干扰。具体而言，它衡量固定列 $\boldsymbol{x}_j$ 与任意 k 个其他列之间绝对值归一化内积之和的最大值。

注意，根据（4.29a），我们有 $\mu_1(\boldsymbol{X}, 1) = \mu_1(\boldsymbol{X})$，因此互相干性就是 ℓ_1-相干性在 $k=1$ 时的特殊情形。

互相干性和 ℓ_1-相干性函数在正交矩阵 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ 左乘下都保持不变。事实上，$\boldsymbol{QX}$ 的列为 $\boldsymbol{Qx}_1$，$\boldsymbol{Qx}_2$，……，$\boldsymbol{Qx}_p$，由于正交变换保持内积，

$$\langle \boldsymbol{Qx}_i, \boldsymbol{Qx}_j\rangle = \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle .$$

此外，根据柯西–施瓦茨不等式 $|\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle| \le \|\boldsymbol{x}_i\|_2 \|\boldsymbol{x}_j\|_2$，归一化内积的大小始终不超过 1。因此，互相干性满足 $\mu(\boldsymbol{X}) \le 1$。

如前所述，$\mu(\boldsymbol{X}) = 0$ 当且仅当 $\boldsymbol{X}$ 的列构成一个标准正交系。特别地，当 $\boldsymbol{X}$ 是方阵（$n = p$）时，这恰好意味着 $\boldsymbol{X}$ 是一个正交矩阵。然而，在压缩感知中，我们主要关注欠定情形，此时 $n < p$。在这种情况下，所有列不可能完全正交，而且互相干性能够小到什么程度也存在根本性的限制。我们将在第 4.2.4 节探讨这些限制。目前，我们只需指出，相干性较小意味着由适量列构成的子矩阵是良态的，这对稳定稀疏恢复至关重要。

我们已经在命题 4.22 中证明，在温和条件下，互相干性蕴含限制等距性质（定义 4.30）。对于 ℓ_1-相干性，也有类似的结果。

定理 4.24：（ℓ_1-相干性与限制等距性质）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个矩阵，其列为 ℓ_2-归一化列（即每一列的范数均为单位范数），且 $1 \le k \le p$。那么，对于每个满足 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$（即 $\|\boldsymbol{\beta}\|_0 \le k$）的向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，

$$\big(1 - \mu_1(\boldsymbol{X}, k-1)\big) \|\boldsymbol{\beta}\|_2^2 \le \|\boldsymbol{X\beta}\|_2^2 \le \big(1 + \mu_1(\boldsymbol{X}, k-1)\big) \|\boldsymbol{\beta}\|_2^2 .$$

等价地，对于每个满足 $|\mathbb{S}| \le k$ 的指标集 $\mathbb{S} \subseteq \{1,2,\dots,p\}$，格拉姆矩阵 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 的所有特征值都落在区间内

$$\lambda(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}) \in [1-\mu_1(\boldsymbol{X},k-1), 1+\mu_1(\boldsymbol{X},k-1)]\,.$$

特别地，如果 $\mu_1(\boldsymbol{X},\ k-1) < 1$，那么 $\boldsymbol{X}$ 满足阶数为 k、常数为$\delta = \mu_1(\boldsymbol{X},\ k-1)$ 的 RIP（见定义 4.30），并且 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 对所有这样的 $\mathbb{S}$ 都是非奇异的。

证明 [定理 4.24] 设 $\mathbb{S} \subseteq \{1,2,\dots,p\}$ 且满足 $|\mathbb{S}| \le k$。由于 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$是对称半正定的，因此它存在一组由特征向量构成的正交归一基，且特征值均为实数且非负（定理 1.18）。分别用 $\lambda_{\min}$ 和 $\lambda_{\max}$表示其最小和最大特征值。对于任意支撑于 $\mathbb{S}$ 上的 $\boldsymbol{\beta} \in \mathbb{R}^p$，有 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}$，因此，$\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 = \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$ 在集合 $\{\boldsymbol{\beta} \in \mathbb{R}^p, \operatorname{supp}(\boldsymbol{\beta}) \subseteq \mathbb{S},\ \|\boldsymbol{\beta}\|_2 = 1\}$ 上的最大值为 $\lambda_{\max}$，最小值为 $\lambda_{\min}$。 [4] 因此，关于 $\|\boldsymbol{X}\boldsymbol{\beta}\|_2$ 的上述界等价于定理中所述的特征值包含关系。

现在，由于 $\boldsymbol{X}$ 的列均已归一化，$\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 的对角元素都等于 1。应用格尔什戈林圆盘定理（定理 1.3），每个特征值都至少位于一个以 1 为圆心、半径为

$$r_i = \sum_{j\in\mathbb{S}, j\neq i} \left|(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})_{ij}\right| = \sum_{j\in\mathbb{S}, j\neq i} |\langle \boldsymbol{x}_j, \boldsymbol{x}_i\rangle| \le \mu_1(\boldsymbol{X}, k-1), \quad i \in \mathbb{S},$$

其中，不等式由 $\mu_1(\boldsymbol{X},\ k-1)$ 的定义得出（因为 $|\mathbb{S}\setminus\{i\}| \le k-1$）。由于所有特征值都是实数，它们必定位于区间 $[1-\mu_1(\boldsymbol{X},\ k-1)\ 1+\mu_1(\boldsymbol{X},\ k-1)]$,内，正如所声称的那样。 ■

上述结果表明，若 $\mu_1(\boldsymbol{X},\ k-1) < 1$，则所有此类$\mathbb{S}$对应的 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 均为非奇异矩阵。此处展开书写.....

上述结果表明，若 $\mu_1(\boldsymbol{X},\ k-1) < 1$，则对于满足 $|\mathbb{S}| \le k$ 的每个指标集 $\mathbb{S}$，格拉姆矩阵 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 均为非奇异矩阵。这保证了 $\boldsymbol{X}$ 的任意 k 列都是线性无关的——这是稀疏恢复的一项关键性质。下面的推论进一步加强了这一结论，给出了保证任意 $2k$ 列仍然线性无关的条件；这通常是分析正交匹配追踪等算法或确立 k-稀疏解唯一性所必需的。

推论 4.25 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的列经过 ℓ_2 归一化，并设 $k \ge 1$ 为整数。若

$$\mu_1(\boldsymbol{X},k) + \mu_1(\boldsymbol{X},k-1) < 1,$$

则对于任意满足 $|\mathbb{S}| \le 2k$ 的指标集 $\mathbb{S} \subseteq \{1,2,\dots,p\}$，矩阵 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 可逆。因此，$\boldsymbol{X}_{\mathbb{S}}$ 具有满列秩。特别地，当

$$\mu(\boldsymbol{X}) < \frac{1}{2k-1}.$$

‘

证明 [推论 4.25] 由不等式（4.29b），可得 $\mu_1(\boldsymbol{X},2k-1) \le \mu_1(\boldsymbol{X},k) + \mu_1(\boldsymbol{X},k-1) < 1$。现在考虑任意一个指标集 $\mathbb{S} \subseteq \{1,2,\dots,p\}$ 满足 $|\mathbb{S}| \le 2k$。应用定理 4.24，取稀疏度 $2k$，我们得到其最小特征值

4。因为最大值$_{\boldsymbol{\beta}\neq 0}\ \frac{\boldsymbol{\beta}^{\top}\boldsymbol{A}\boldsymbol{\beta}}{\boldsymbol{\beta}^{\top}\boldsymbol{\beta}} = \lambda_{\max}$ 和最小值$_{\boldsymbol{\beta}\neq 0}\ \frac{\boldsymbol{\beta}^{\top}\boldsymbol{A}\boldsymbol{\beta}}{\boldsymbol{\beta}^{\top}\boldsymbol{\beta}} = \lambda_{\min}$。

$\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 满足 $\lambda_{\min} \geq 1-\mu_1(\boldsymbol{X}, 2k-1) > 0$。因此，$\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 是正定的，因而可逆。为证明 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩，假设 $\boldsymbol{X}_{\mathbb{S}}\boldsymbol{z} = \boldsymbol{0}$对某个向量 $\boldsymbol{z} \in \mathbb{R}^{|\mathbb{S}|}$成立，这意味着 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{z} = \boldsymbol{0}$。由于 $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 可逆，可得$\boldsymbol{z} = \boldsymbol{0}$。这证明了第一个结论。最后，利用不等式 (4.29a)，我们得到上界$\mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, k-1) \leq (2k-1)\mu(\boldsymbol{X}) < 1$ 当且仅当 $\mu(\boldsymbol{X}) < 1/(2k-1)$。 ■

4.2.4 互相干性和 ℓ_1-相干性的界

本小节中，我们为矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ with $n < p$的互相干性和 ℓ_1-相干性函数分别建立下界。

达到这些下界的矩阵具有一种独特的结构：其列构成等角紧框架，定义如下。

定义4.26（等角系统与紧框架系统）。具有 ℓ_2-归一化列的矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 称为等角的，如果存在 $C \geq 0$ 使得

$$|\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle| = C \quad \text{for all } i \neq j,\ i, j \in \{1, 2, \ldots, p\}.$$

如果存在常数 $\lambda > 0$，使得以下任一等价条件成立，则称矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为紧框架：

(i) $\|\boldsymbol{\alpha}\|_2^2 = \lambda \sum_{j=1}^{p} \langle \boldsymbol{\alpha}, \boldsymbol{x}_j \rangle^2$ 对所有 $\boldsymbol{\alpha} \in \mathbb{R}^n$ 均成立，(ii) $\boldsymbol{\alpha} = \lambda \sum_{j=1}^{p} \langle \boldsymbol{\alpha}, \boldsymbol{x}_j \rangle \boldsymbol{x}_j$ 对所有 $\boldsymbol{\alpha} \in \mathbb{R}^n$均成立，(iii) $\boldsymbol{X}\boldsymbol{X}^{\top} = \frac{1}{\lambda}\boldsymbol{I}_n$。

顾名思义，若一组 ℓ_2-归一化向量满足上述两个性质，则称其为等角紧框架。

恰好是这类系统能够达到下文所述下界，该下界称为*Welch*界。

定理 4.27：（Welch界：互相干性的下界（Rosenfeld，1997 年；Strohmer和 Heath Jr，2003；Welch，2003）） 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 且 $p > n$。则其互相干性满足

$$\mu(\boldsymbol{X}) \geq \sqrt{\frac{p-n}{n(p-1)}}. \tag{4.30}$$

当且仅当 $\boldsymbol{X}$ 的列构成等角紧框架时，等号成立。注意，当 $p \gg n$ 时，下界约为 $\mu(\boldsymbol{X}) \geq 1/\sqrt{n}$。

证明 [定理4。27] 不失一般性，假设 $\boldsymbol{X}$ 具有 ℓ_2-归一化列。定义格拉姆矩阵 $\boldsymbol{G} \triangleq \boldsymbol{X}^{\top}\boldsymbol{X} \in \mathbb{R}^{p\times p}$ 以及矩阵 $\boldsymbol{H} \triangleq \boldsymbol{X}\boldsymbol{X}^{\top} \in \mathbb{R}^{n\times n}$。由于 $\boldsymbol{X}$ 的列是 ℓ_2-归一化的，因此有 $\operatorname{tr}(\boldsymbol{G}) = \sum_{i=1}^{p} \|\boldsymbol{x}_i\|_2^2 = p$。回顾弗罗贝尼乌斯内积 $\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F \triangleq \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}^{\top}) = \sum_{i,j=1}^{n} a_{ij}b_{ij}$ 诱导出弗罗贝尼乌斯范数 $\|\cdot\|_F$ (定义 1.30)。根据柯西–施瓦茨不等式，

$$\operatorname{tr}(\boldsymbol{H}) \leq \|\boldsymbol{H}\|_F \|\boldsymbol{I}_n\|_F = \sqrt{n}\sqrt{\operatorname{tr}(\boldsymbol{H}\boldsymbol{H}^{\top})}. \tag{4.31}$$

利用迹的循环不变性，我们计算得

$$\operatorname{tr}(\boldsymbol{H}\boldsymbol{H}^\top)=\operatorname{tr}\big((\boldsymbol{X}\boldsymbol{X}^\top)(\boldsymbol{X}\boldsymbol{X}^\top)\big)=\operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X})=\operatorname{tr}(\boldsymbol{G}\boldsymbol{G}^\top)=\sum_{i,j=1}^{p,p}|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|^2$$
$$=\sum_{i=1}^{p}\|\boldsymbol{x}_i\|_2^2+\sum_{i,j=1,i\neq j}^{p,p}|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|^2=p+\sum_{i,j=1,i\neq j}^{p,p}|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|^2.$$

再次利用迹的循环不变性，可得 $\operatorname{tr}(\boldsymbol{G})=\operatorname{tr}(\boldsymbol{H})$，因此我们有

$$p^2\leq n\left(p+\sum_{i,j=1,i\neq j}^{p,p}|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|^2\right). \tag{4.32}$$

根据互相干性的定义，

$$|\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle|\leq\mu(\boldsymbol{X}),\quad \text{for all } 1\leq i\neq j\leq p, \tag{4.33}$$

我们得到 $p^2\leq n(p+(p^2-p)\mu(\boldsymbol{X})^2)$，整理后即为所述上界（4.30）。最后，当且仅当（4.31）和（4.33）同时取等号时，等号成立。（4.31）取等号意味着 $\boldsymbol{H}=\lambda\boldsymbol{I}_n$ 对某个 $\lambda>0$成立，即 $\boldsymbol{X}$ 是一个紧框架。（4.33）取等号意味着所有非对角内积的大小都相同，即 $\boldsymbol{X}$ 是等角的。因此，当且仅当 $\boldsymbol{X}$ 构成一个等角紧框架时，等号成立。 ■

对于参数值较小的情形，Welch界可以推广到 ℓ_1-相干性函数。

定理 4.28：（Welch界：ℓ_1-相干性的下界） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 且$p>n$。则其 ℓ_1-相干性函数满足

$$\mu_1(\boldsymbol{X},k)\geq k\sqrt{\frac{p-n}{n(p-1)}},\quad \text{for}\quad k<\sqrt{p-1}. \tag{4.34}$$

当且仅当 $\boldsymbol{X}$ 的列构成一个等角紧框架时，等号成立。

证明依赖于以下辅助结果。

引理 4.29 设 $s<\sqrt{m}$，并令 $(\alpha_1,\alpha_2,\ldots,\alpha_m)$ 为一个非负、非递增序列，满足

$$\alpha_1\geq\alpha_2\geq\ldots\geq\alpha_m\geq 0\quad\text{and}\quad\alpha_1^2+\alpha_2^2+\ldots+\alpha_m^2\geq\frac{m}{s^2}.$$

Then

$$\alpha_1+\alpha_2+\ldots+\alpha_s\geq 1,$$

当且仅当 $\alpha_1=\alpha_2=\ldots=\alpha_m=1/s$ 时，等号成立。

证明 [引理4。29] 只需证明其逆否命题：

$$\left\{\begin{array}{l}\alpha_1\geq\alpha_2\geq\ldots\geq\alpha_m\geq 0\\ \alpha_1+\alpha_2+\ldots+\alpha_s\leq 1\end{array}\right\}\Longrightarrow\alpha_1^2+\alpha_2^2+\ldots+\alpha_m^2\leq\frac{m}{s^2},$$

且仅当所有 $\alpha_i = 1/s$ 时取等号。这等价于最大化凸函数

$$f(\alpha_1, \alpha_2, \ldots, \alpha_m) \triangleq \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_m^2$$

在凸集上

$$\mathbb{S} \triangleq \{(\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m \mid \alpha_1 \geq \ldots \geq \alpha_m \geq 0 \text{ and } \alpha_1 + \ldots + \alpha_s \leq 1\}.$$

由于 f 是凸的，而 $\mathbb{S}$ 是紧凸多面体，因此 f 在 $\mathbb{S}$ 上的最大值在 $\mathbb{S}$ 的某个极点（顶点）处取得。$\mathbb{S}$ 的顶点出现在足够多的不等式约束成为等式约束时。除对称性和排列顺序外，相关情形如下：

- 若 $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$，则 $f(\alpha_1, \alpha_2, \ldots\ldots, \alpha_m) = 0$；若对 $1 \leq \ell \leq s$ 有
- $\alpha_1 + \ldots + \alpha_s = 1$ 且 $\alpha_1 = \ldots = \alpha_\ell > \alpha_{\ell+1} = \ldots = \alpha_m = 0$，则 $\alpha_1 = \ldots = \alpha_\ell = 1/\ell$，因此 $f(\alpha_1, \alpha_2, \ldots\ldots, \alpha_m) = 1/\ell$；若对 $s < \ell \leq m$ 有 $\alpha_1 + \ldots + \alpha_s = 1$ 且
- $\alpha_1 = \ldots = \alpha_\ell > \alpha_{\ell+1} = \ldots = \alpha_m = 0$，则 $\alpha_1 = \ldots = \alpha_\ell = 1/s$，因此 $f(\alpha_1, \alpha_2, \ldots\ldots, \alpha_m) = \ell/s^2$。

由于 $s < \sqrt{m}$，我们有

$$\max_{(\alpha_1,\ldots,\alpha_m)\in\mathbb{S}} f(\alpha_1, \ldots, \alpha_m) = \max\left\{\max_{1\leq\ell\leq k} \frac{1}{\ell}, \max_{s<\ell\leq m} \frac{\ell}{s^2}\right\} = \max\left\{1, \frac{m}{s^2}\right\} = \frac{m}{s^2},$$

等号仅在 $\ell = m$ 和 $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 1/s$ 时成立。 ■

证明 [定理4.28] 不失一般性，我们假设 $\boldsymbol{X}$ 具有 ℓ_2-归一化列。由不等式 (4.32)，我们有 $\sum_{i,j=1,i\neq j}^{p} |\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle|^2 \geq p^2 - p = p(p-n)$。这意味着

$$\max_{i\in\{1,2,\ldots,p\}} \sum_{j=1,j\neq i}^{p} |\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle|^2 \geq \frac{1}{p} \sum_{i,j=1,i\neq j}^{p,p} |\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle|^2 \geq \frac{p-n}{n}.$$

对于使该最大值达到的索引 $i^\dagger \in \{1, 2, \ldots, p\}$，考虑绝对内积构成的（$p-1$）维向量：$\{|\langle \boldsymbol{x}_{i^\dagger}, \boldsymbol{x}_j \rangle|\}_{j=1,j\neq i^\dagger}^{p}$。将这些值按非递增顺序重新排列为 $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{p-1} \geq 0$，使得

$$\beta_1^2 + \beta_2^2 + \ldots + \beta_{p-1}^2 \geq \frac{p-n}{n}.$$

应用引理 4.29并结合 $m = p-1$和 $s = k$，并定义 $\alpha_\ell \triangleq \left(\frac{1}{k}\sqrt{n\,(p-1)\,/\,(p-n)}\right)\beta_\ell$ 可得 $\alpha_1 + \alpha_2 + \ldots + \alpha_k \geq 1$。根据 ℓ_1-相干性的定义，可知

$$\mu_1(\boldsymbol{X}, k) \geq \beta_1 + \beta_2 + \ldots + \beta_k \geq k\sqrt{\frac{p-n}{n(p-1)}},$$

从而得到所需的上界（4.34）。

最后，假设（4.34）中的等式成立。于是所有中间不等式都必须取等。特别地：

- （4.32）中的等号成立意味着 $\boldsymbol{X}$ 是一个紧框架（如定理 4.27的证明所示）。引理
- 4.29中的等号成立意味着，涉及列$i^\dagger$的所有非对角内积具有相同的大小：$|\langle \boldsymbol{x}_{i^\dagger}, \boldsymbol{x}_j \rangle| = \sqrt{(p-n)\,/\,(n\,(p-1))}$ 对所有 $j \in \{1, 2, \ldots, p\}$，$j \neq i^\dagger$均成立。

这保证了 $i^\dagger$ 的选取在 $\{1,2,\ldots,p\}$中必然是任意的，因此不同列之间所有成对内积的大小都相同。因此， $\boldsymbol{X}$ 是等角的。反过来，证明等角紧框架使（4.34）中的等号成立，可由上述讨论轻松得出。 ■

4.3. 受限等距性质（RIP）

NSP 条件对于确立如下形式的误差保证，既是必要条件，也是充分条件

$$\|\Delta(\boldsymbol{X}\boldsymbol{\beta})-\boldsymbol{\beta}\|_2 \le C\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}, \quad \text{with} \quad \sigma_k(\boldsymbol{\beta})_s \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]} \left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|_s. \tag{4.35}$$

然而，这些保证假设测量值无噪声。在实际情况下，测量值会受到噪声污染（例如，由传感器误差或量化导致），因此需要更强的条件来确保稳定且鲁棒的恢复。为此，Candes 和 Tao (2005年) 提出了受限等距性质（RIP）。这是压缩感知中的一项基本条件，我们在讨论互相干性时曾简要提及。

定义 4.30（受限等距性质（RIP））。若矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 满足阶数为 k $(k<p)$ 且常数为$\delta_k\in[0,1)$ 的受限等距性质（*RIP*），记作 RIP(k,δ_k)，则对于 k-稀疏向量 $\boldsymbol{\beta}\in\mathbb{B}_0[k]\triangleq\{\boldsymbol{\alpha}\mid\|\boldsymbol{\alpha}\|_0\le k\}$，有

$$(1-\delta_k)\cdot\|\boldsymbol{\beta}\|_2^2 \le \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \le (1+\delta_k)\cdot\|\boldsymbol{\beta}\|_2^2. \tag{4.36a}$$

利用谱范数单位矩阵恒等式 $\left\|\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I}\right\|_2=$最大值$_{\boldsymbol{u}\in\mathbb{R}^p:\ \|\boldsymbol{u}\|_2=1}\left|\boldsymbol{u}^\top(\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I})\boldsymbol{u}\right|=$最大值$_{\boldsymbol{u}\in\mathbb{R}^p:\ \|\boldsymbol{u}\|_2=1}\left|\|\boldsymbol{X}\boldsymbol{u}\|_2^2-1\right|$ 根据（1.9），条件（4.36a）可以等价地表述为

$$\left\|\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}}-\boldsymbol{I}_{|\mathbb{S}|}\right\|_2\le\delta_k, \quad \text{for all } \mathbb{S}\subseteq\{1,2,\ldots,p\} \text{ with } |\mathbb{S}|\le k., \tag{4.36b}$$

其中， $\boldsymbol{X}_{\mathbb{S}}\triangleq\boldsymbol{X}[:\mathbb{S}]\in\mathbb{R}^{n\times|\mathbb{S}|}$ 表示来自指标集 $\mathbb{S}$的 $\boldsymbol{X}$ 的列。

使 RIP$(k,\ \delta_k)$ 成立的最小常数 δ_k 称为阶为 k的限制等距常数（*RIC*）。其定义为

$$\delta_k=\min\left\{\delta\mid(1-\delta)\|\boldsymbol{v}\|_2^2\le\|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{v}\|_2^2\le(1+\delta)\|\boldsymbol{v}\|_2^2,\ \forall\ |\mathbb{S}|\le k,\forall\,\boldsymbol{v}\in\mathbb{R}^{|\mathbb{S}|}\right\} \tag{4.37a}$$

$$=\max_{\mathbb{S}\subseteq\{1,2,\ldots,p\},|\mathbb{S}|\le k}\left\|\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}}-\boldsymbol{I}_{|\mathbb{S}|}\right\|_2. \tag{4.37b}$$

请注意，(4.37a) 或 (4.37b) 中的任一表达式都意味着 $\boldsymbol{X}_{\mathbb{S}}$的所有奇异值都位于该区间

$$\sigma(\boldsymbol{X}_{\mathbb{S}})\in[\sqrt{1-\delta_k},\sqrt{1+\delta_k}] \qquad \text{with} \qquad \delta_k\le 1.$$

（见定理 1.52）。等价地， $\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}}$ 的特征值位于

$$\lambda(\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}})\in[1-\delta_k,1+\delta_k] \qquad \text{with} \qquad \delta_k\le 1.$$

[5] 严格不等式 $\delta_k<1$ 进一步表明， $\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}_{\mathbb{S}}$ 的最小特征值为正，这意味着对于任意满足 $|\mathbb{S}|\le k$的指标集 $\mathbb{S}$， $\boldsymbol{X}_{\mathbb{S}}$ 都具有满列秩，

5。量 $\frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|_2^2}$ 称为矩阵 *Rayleigh quotient* 的 $\boldsymbol{X}^\top\boldsymbol{X}$，当 $\boldsymbol{\beta}$ 变化时，它始终介于其最小和最大特征值之间。

根据正定矩阵的特征值刻画定理（定理 1.18）。因此， $\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}$ 是非奇异的。

特征值（或奇异值）与 RIP 常数之间的这种关系，对于推导后文要用到的关键不等式至关重要。

注4.31（IHT 不等式）。由（4.36b），可得

$$\left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{\beta}\right\|_2=\left\|\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}-\boldsymbol{\beta}_{\mathbb{S}}\right\|_2\leq\left\|\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}-\boldsymbol{I}_{|\mathbb{S}|}\right\|_2\|\boldsymbol{\beta}_{\mathbb{S}}\|_2\leq\delta_k\|\boldsymbol{\beta}\|_2\,,\tag{4.38}$$

其中，$\mathrm{supp}(\boldsymbol{\beta})=\mathbb{S}$。这是迭代算法（例如迭代硬阈值法（*IHT*））中常用的上界。或者，应用三角不等式可得：

$$\begin{aligned}\left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}\right\|_2&\leq\|\boldsymbol{\beta}\|_2+\left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{\beta}\right\|_2\leq(1+\delta_k)\|\boldsymbol{\beta}\|_2\,;\\ \left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}\right\|_2&\geq\|\boldsymbol{\beta}\|_2-\left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{\beta}\right\|_2\geq(1-\delta_k)\|\boldsymbol{\beta}\|_2\,.\end{aligned}$$

因此，我们得到关键不等式：

$$(1-\delta_k)\|\boldsymbol{\beta}\|_2\leq\left\|\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}\right\|_2\leq(1+\delta_k)\|\boldsymbol{\beta}\|_2\,.\tag{4.39}$$

非正式地说，若矩阵 $\boldsymbol{X}$ 满足限制等距性质，则δ_k 对于足够大的稀疏度 k而言很小。注意，对于支撑集为 $\mathbb{S}$的向量 $\boldsymbol{\beta}$，有 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}$。因此，限制等距性质本质上要求 $\boldsymbol{X}$ 的每个子矩阵（通过选取至多 k 列得到）都是良态的——也就是说，它在稀疏向量上的作用近似于等距变换。事实上，对于稀疏优化问题，阶数为 $2k$ 的限制等距性质尤为相关（例如，参见定理4.50）。确实，条件（4.36a）意味着 $\|\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\|_2>0$ 对于任意两个不同的 k稀疏度为 $\boldsymbol{\beta}_1$的向量 $\boldsymbol{\beta}_2\in\mathbb{B}_0[k]$。因此，它们的测量向量也彼此不同： $\boldsymbol{X}\boldsymbol{\beta}_1\neq\boldsymbol{X}\boldsymbol{\beta}_2$。

以稀疏恢复为例，假设 $\boldsymbol{\beta}^*$ 是 k-稀疏的。那么：

- 只要 $\boldsymbol{X}$ 的任意 $2k$ 列都是线性无关的，我们就能根据无噪声测量值 $\boldsymbol{X}\boldsymbol{\beta}^*=\boldsymbol{y}$ 唯一恢复 $\boldsymbol{\beta}^*\in\mathbb{B}_0[k]$。这是因为，如果对于两个 k-稀疏向量 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$，有 $\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)=\boldsymbol{0}$，那么 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$，从而确保解的唯一性。反之，如果 $\boldsymbol{X}$ 包含一组 $2k$ 个线性相关列，则无法保证 $\boldsymbol{\beta}^*$ 的唯一恢复。在这种情况下，存在一个非零向量 $\boldsymbol{n}=\boldsymbol{\beta}_1-\boldsymbol{\beta}_2$（其中$\|\boldsymbol{n}\|_0\leq 2k$），使得 $\boldsymbol{X}\boldsymbol{n}=\boldsymbol{0}$，这意味着存在两个不同的 k-稀疏向量 $\boldsymbol{X}\boldsymbol{\beta}_1=\boldsymbol{X}\boldsymbol{\beta}_2$。

这个例子说明了为什么限制等距性质（RIP）的 $2k$ 阶需要一个下界。如果该下界不成立——即如果 $(1-\delta_{2k})$ 太小或为零——那么就可能找到两个不同的 k-稀疏向量 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$，它们具有不相交支撑集，并满足 $\boldsymbol{X}\boldsymbol{\beta}_1=\boldsymbol{X}\boldsymbol{\beta}_2$。RIP 的下界确保不同的稀疏信号会被映射为彼此充分分离的测量向量，从而避免恢复歧义。

然而，RIP 可能具有一定的限制性，因为它要求向量范数的畸变以 $(1\pm\delta_k)$ 的形式对称地有界，其中 $\delta_k\in[0,1)$。根据定义，如果矩阵 $\boldsymbol{X}$ 是正交的（即 $\boldsymbol{X}\boldsymbol{X}^{\top}=\boldsymbol{X}^{\top}\boldsymbol{X}=\boldsymbol{I}$），那么 $\delta_k=0$，因为对于所有 $\boldsymbol{\beta}$，$\|\boldsymbol{X}\boldsymbol{\beta}\|_2=\|\boldsymbol{\beta}\|_2$。因此，非零 RIC δ_k 可以被解释为衡量 $\boldsymbol{X}$ 偏离正交程度的指标，但仅限于考虑至多涉及 k 列的稀疏线性组合时。

回顾一下，表达式 $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}/\|\boldsymbol{v}\|_2^2$ 被称为非零向量 $\boldsymbol{v}$ 相对于矩阵 $\boldsymbol{A} \in \mathbb{R}^{p\times p}$的瑞利商。该量始终介于 $\boldsymbol{A}$ 的最大特征值和最小特征值之间。最大特征值与最小特征值之比（称为条件数）刻画了矩阵的良态程度（Lu，2021a，2025年）；条件数越小，说明矩阵的条件越好。在此背景下，RIP 条件意味着两个重要性质：

- 由 δ_k $1-\delta_k\ \|\boldsymbol{\beta}\|_2^2 \le \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \le 1+\delta_k\ \|\boldsymbol{\beta}\|_2^2$ $\boldsymbol{X}$ 控制的范数保持。不等式 ( ) ( ) 保证，对于每个 k-稀疏向量 $\boldsymbol{\beta}$，其欧几里得范数均可由近
- 似保持，偏差由 δ_k 控制。良态子矩阵。当 δ_k 不太大时，从 $\boldsymbol{X}$ 中任意选取 k 列构成的每个子矩阵都是良态的。换言之，这些子矩阵接近正交归一。此性质对于确保稀疏恢复算法的稳定性和准确性至关重要。

▶ **RIP 经缩放、重排和增广后的性质？** 回顾 (4.6) 可知，NSP 在测量矩阵的某些变换下保持不变——具体而言，当测量值被缩放、重排或增广时。（因此，通过 k- 稀疏恢复的 ℓ_1-最小化不受此类操作影响；参见第 5 章。）然而，这些相同的操作可能会使受限等距常数（RIC）恶化。重排测量值对应于将测量矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 替换为 $\boldsymbol{P}\boldsymbol{X}$，其中 $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ 是一个置换矩阵（正交矩阵的一种特殊情况）。由于正交变换保持奇异值不变，即对于任意正交矩阵 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$，(σ_i)$\boldsymbol{Q}\boldsymbol{X}$ ($=\sigma_i$)$\boldsymbol{X}$，因此此操作不会改变受限等距常数（RIC）。

重新缩放测量值——即用 $\boldsymbol{D}\boldsymbol{X}$ 替换 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，其中 $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ 是一个对角矩阵——也可能增大受限等距常数。这甚至适用于标量缩放：对某个 $\eta \in \mathbb{R}$，用 $\eta\boldsymbol{X}$ 替换 $\boldsymbol{X}$。例如，若 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 $\delta_k(\boldsymbol{X}) < 1/3$，则缩放后的矩阵 $2\boldsymbol{X}$ 的受限等距常数为 $\delta_k(2\boldsymbol{X}) = 5/6$。

另一方面，增加一个测量值相当于在 $\boldsymbol{X}$ 中追加一行，这也可能增大受限等距常数。例如，设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 $\delta_k(\boldsymbol{X}) < 1$，$\widetilde{\ }$，并任取 $\delta > \delta_k(\boldsymbol{X})$。通过追加行向量 $[0,\dots,0,\sqrt{1+\delta}]$ 构造新矩阵 $\underline{\boldsymbol{X}}$。考虑 1-稀疏向量 $\boldsymbol{\beta} \triangleq [0,\dots,0,1]^\top$。那么，由 $\|\widetilde{\boldsymbol{X}}\boldsymbol{\beta}\|_2^2 \ge 1+\delta$，
$\widetilde{\ }$ $\widetilde{\ }$ 可得 $\delta_1(\boldsymbol{X}) \ge \delta$，进而有 $\delta_k(\boldsymbol{X}) > \delta_k(\boldsymbol{X})$（根据习题 4.34）。

习题 4.32。 给定 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，令 $\mu > 0$ 和 $\nu > 0$ 分别为满足以下条件的最大和最小常数

$$\mu \cdot \|\boldsymbol{\beta}\|_2^2 \le \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \le \nu \cdot \|\boldsymbol{\beta}\|_2^2, \quad \text{for all } k\text{-sparse vectors } \boldsymbol{\beta} \in \mathbb{R}^p.$$

求出 使 $\delta_k(\eta)$ 最小的缩放因子 $\eta > 0$。提示：考虑 $(\nu-\mu)/(\nu+\mu)$。 .

还需注意，在我们对 RIP 的定义中，假设界在 1 两侧对称（即形式为 $1\pm\delta_k$）。不过，采用这种对称形式仅仅是为了方便符号表示。在实践中，可以考虑非对称界，从而得到 RIP 的一种实用推广——这在设计矩阵未受到严格控制的情形下尤为相关，例如基因表达分析。这种推广由受限强凸性（RSC）和受限光滑性（RSS）的概念所刻画。

定义 4.33（受限强凸/光滑性性质）。 矩阵$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 称为满足 μ-阶受限强凸性（*RSC*）和 ν-阶受限光滑性（*RSS*）性质，阶数为 k 当且仅当对所有 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$,

$$\mu \cdot \|\boldsymbol{\beta}\|_2^2 \leq \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \leq \nu \cdot \|\boldsymbol{\beta}\|_2^2 . \tag{4.40}$$

RIP 与 RSC/RSS 的关键区别在于它们的常数。RIP 要求下界和上界以 1 为中心对称，即 $1 \pm \delta_k$，而 RSC/RSS 则没有这样的对称性。读者可能会注意到，定义 4.33与定义 2.86之间有密切联系；定义 2.86 中为一般函数定义了受限强凸性和光滑性。事实上，定义 2.86 可以看作是将定义 4.33推广到任意（可能是非二次函数的）函数（Jain 等人，2014年；Lu，2025年）。对于二阶可微函数，这两个定义本质上都约束海森矩阵的受限特征值。

重要的是，对于任意满足式（4.40）中非对称界的矩阵 $\boldsymbol{X}$，我们总能对其进行缩放，使其满足式（4.36a）中的对称受限等距条件。具体而言，定义

$$\widetilde{\boldsymbol{X}} \triangleq \sqrt{2/(\nu+\mu)}\boldsymbol{X}.$$

那么 $\widetilde{\boldsymbol{X}}$ 满足阶数为 k、常数为 $\delta_k = (\nu-\mu)/(\nu+\mu)$ 的 RIP。虽然我们在此未明确证明这一点，但验证起来很直接。此外，本书中所有假设 RIP 成立的定理，只要对 $\boldsymbol{X}$ 进行某种缩放后满足 RIP，实际上仍然成立。因此，不失一般性，我们可以将注意力限制在式（4.36a）中的对称形式上。

RIP 的另一个有用性质是其关于稀疏度的单调性：如果 $\boldsymbol{X}$ 满足阶为 k、常数为 δ_k 的 RIP，那么对于任意 $k' < k$，它都会自动满足阶为 k'、常数为 $\delta_{k'} \leq \delta_k$的 RIP。

习题 4.34（受限等距常数的非递减性）。 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 RIP 条件。证明：

$$\delta_1 \leq \delta_2 \leq \ldots \leq \delta_p. \tag{4.41}$$

此外，如果 $\boldsymbol{X}$ 满足阶为 k 且常数充分小的 RIP，那么对于某些 γ，它也会自动满足阶为 γk 的 RIP，不过此时的常数会稍大一些。

练习 4.35（限制等距性质的变体（Needell 和 Tropp，2009年））。 假设 $\boldsymbol{X}$ 满足阶为 k、常数为 δ_k的 RIP。设 γ 为一个正整数。证明 $\boldsymbol{X}$满足阶为 $k' = \gamma\lfloor \frac{k}{2} \rfloor$、常数为 $\delta_{k'} < \gamma \cdot \delta_k$ 的 RIP，其中 $\lfloor\cdot\rfloor$ 表示向下取整函数。

对于 $\gamma = 1,2$，这一结果显而易见；但对于 $\gamma \geq 3$（以及 $k \geq 4$），它使我们能够将 RIP 的受限等距保证从阶 k 扩展到显著更高的稀疏度级别。不过请注意，要使所得上界有意义（例如 $\delta_{k'} < 1$），原始常数 δ_k 必须足够小。

4.3.1 受限正交性性质（ROP）

请注意，不等式（4.36a）意味着，对于任意满足 $\mathbb{S} \subseteq \{1,2,\ldots,p\}$ 的指标集 $|\mathbb{S}| < k$，格拉姆矩阵 $\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}$ 的特征值都位于区间 $[1-\delta_k, 1+\delta_k]$中。特别是，当 $k = 1$时，常数 δ_1 反映了 $\boldsymbol{X}$ 的列范数偏离 1 的程度：$\delta_1 = 0$ 当且仅当 $\boldsymbol{X}$ 的每一列都具有单位欧几里得范数时成立。一个相关概念

用于刻画不同列（或互不相交的列子集）之间非正交性的，是受限正交性性质（*ROP*）。

为了引出这一概念，我们先从一个简单但重要的观察开始，它自然地引出了受限正交常数（*ROC*）的定义。

命题 4.36：（RIP 的标准推论）设 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$ 为向量，且满足 $\|\boldsymbol{\alpha}\|_0 \leq s$ 和 $\|\boldsymbol{\beta}\|_0 \leq t$。若 $\operatorname{supp}(\boldsymbol{\alpha}) \cap \operatorname{supp}(\boldsymbol{\beta}) = \varnothing$，则

$$|\langle \boldsymbol{X}\boldsymbol{\alpha}, \boldsymbol{X}\boldsymbol{\beta} \rangle| \leq \delta_{s+t} \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 .$$

证明 [命题 4.36] 设 $\mathbb{S} \triangleq \operatorname{supp}(\boldsymbol{\alpha}) \cup \operatorname{supp}(\boldsymbol{\beta})$，并将 $\boldsymbol{\alpha}_{\mathbb{S}}$，$\boldsymbol{\beta}_{\mathbb{S}} \in \mathbb{R}^{|\mathbb{S}|}$ 分别记为 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$ 在坐标 $\mathbb{S}$上的限制。由于 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的支撑集互不相交，$\langle \boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\beta}_{\mathbb{S}} \rangle = 0$。于是我们计算得：

$$\begin{aligned} |\langle \boldsymbol{X}\boldsymbol{\alpha}, \boldsymbol{X}\boldsymbol{\beta} \rangle| &= |\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}} \rangle - \langle \boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\beta}_{\mathbb{S}} \rangle| = \left| \left\langle (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}} - \boldsymbol{I})\boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\beta}_{\mathbb{S}} \right\rangle \right| \\ &\leq \left\| \boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}} - \boldsymbol{I} \right\|_2 \|\boldsymbol{\alpha}_{\mathbb{S}}\|_2 \|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \leq \delta_{s+t} \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 , \end{aligned}$$

其中，最后一个不等式由（4.37b）以及 $\|\boldsymbol{\alpha}_{\mathbb{S}}\|_2 = \|\boldsymbol{\alpha}\|_2$ 和$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 = \|\boldsymbol{\beta}\|_2$的事实可得。 ■

这一结果促使人们提出了一个与 RIP 互补的条件，即受限正交性性质。

定义 4.37（受限正交性性质（ROP））。如果矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 对阶为（k，k'）且常数为 $\theta_{k,\,k'}$（$k + k' \leq p$）的受限正交性性质（*ROP*），则称其满足 ROP(k，k'，$\theta_{k,\,k'}$)；也就是说，对于任意**不相交的**向量 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 和 $\boldsymbol{\beta}' \in \mathbb{B}_0[k']$（即 $\langle \boldsymbol{\beta}, \boldsymbol{\beta}' \rangle = 0$），均有

$$\left| \left\langle \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{X}\boldsymbol{\beta}' \right\rangle \right| \leq \theta_{k,k'} \cdot \|\boldsymbol{\beta}\|_2 \left\|\boldsymbol{\beta}'\right\|_2 . \tag{4.42a}$$

满足条件的最小常数 $\theta_{k,\,k'}$ 称为限制正交常数（*ROC*）。它可以定义为使 ROP(k，k'，$\theta_{k,\,k'}$)成立的所有参数 θ 的下确界，即

$$\begin{aligned} &\theta_{k,k'} = \min \mathbb{S}; \\ &\mathbb{S} \triangleq \left\{ \theta \mid |\langle \boldsymbol{X}_{\mathbb{I}}\boldsymbol{v}_1, \boldsymbol{X}_{\mathbb{I}'}\boldsymbol{v}_2 \rangle| \leq \theta \cdot \|\boldsymbol{v}_1\|_2 \|\boldsymbol{v}_2\|_2 , \forall\, |\mathbb{I}| \leq k, |\mathbb{I}'| \leq k', \forall \boldsymbol{v}_1 \in \mathbb{R}^{|\mathbb{I}|}, \boldsymbol{v}_2 \in \mathbb{R}^{|\mathbb{I}'|}, \mathbb{I} \cap \mathbb{I}' = \varnothing \right\}. \end{aligned} \tag{4.42b}$$

等价地，ROC 具有如下显式刻画

$$\theta_{k,k'} = \max \left\{ \left\| \boldsymbol{X}_{\mathbb{I}}^\top \boldsymbol{X}_{\mathbb{I}'} \right\|_2, \mathbb{I} \cap \mathbb{I}' = \varnothing, |\mathbb{I}| \leq k, \left|\mathbb{I}'\right| \leq k' \right\}. \tag{4.42c}$$

注意，由于对称性，$\theta_{k,\,k'} \equiv \theta_{k',\,k}$。

RIP 和 ROP 常数密切相关。为说明这一点，考虑定义 4.37 中的情形，并且不失一般性地假设 $\|\boldsymbol{\beta}\|_2 = \left\|\boldsymbol{\beta}'\right\|_2 = 1$（因为这两种性质都是尺度不变的）。由于 $\mathbb{I}$和 $\mathbb{I}'$互不相交，因此存在 $\boldsymbol{\alpha} \triangleq \boldsymbol{\beta} + \boldsymbol{\beta}' \in \mathbb{B}_0[k + k']$和 $\boldsymbol{\alpha}' \triangleq \boldsymbol{\beta} - \boldsymbol{\beta}' \in \mathbb{B}_0[k + k']$——$(k + k')$-稀疏向量，使得 $\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\beta}'$ 且$\boldsymbol{X}\boldsymbol{\alpha}' = \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}'$——满足

$$\begin{aligned} (1 - \delta_{k+k'}) \cdot \left\|\boldsymbol{\beta} + \boldsymbol{\beta}'\right\|_2^2 \leq \left\|\boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\beta}')\right\|_2^2 \leq (1 + \delta_{k+k'}) \cdot \left\|\boldsymbol{\beta} + \boldsymbol{\beta}'\right\|_2^2 ; \\ (1 - \delta_{k+k'}) \cdot \left\|\boldsymbol{\beta} - \boldsymbol{\beta}'\right\|_2^2 \leq \left\|\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}')\right\|_2^2 \leq (1 + \delta_{k+k'}) \cdot \left\|\boldsymbol{\beta} - \boldsymbol{\beta}'\right\|_2^2 . \end{aligned}$$

由于 $\|\boldsymbol{\beta}-\boldsymbol{\beta}'\|_2^2=\|\boldsymbol{\beta}+\boldsymbol{\beta}'\|_2^2=2$，这些不等式可简化为：

$$\begin{aligned}2(1-\delta_{k+k'})\le\|\boldsymbol{X}(\boldsymbol{\beta}+\boldsymbol{\beta}')\|_2^2\le 2(1+\delta_{k+k'});\\ 2(1-\delta_{k+k'})\le\|\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}')\|_2^2\le 2(1+\delta_{k+k'}).\end{aligned}\tag{4.43}$$

将两个平方范数相减，并利用单位矩阵

$$\|\boldsymbol{X}(\boldsymbol{\beta}+\boldsymbol{\beta}')\|_2^2-\|\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}')\|_2^2=4\left\langle\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{X}\boldsymbol{\beta}'\right\rangle,$$

还可以进一步推导出$\theta_{k,\,k'}$ 与 $\delta_{k+k'}$之间的定量联系，说明较小的 RIP 常数意味着较小的受限正交常数；见命题4.40。

练习4.38（受限正交常数的非递减性）。证明，式（4.42b）中定义的受限正交常数满足以下单调性：

$$\begin{aligned}\theta_{k,k'}\le\theta_{s,k'},\quad\text{if}\quad k\le s;\\ \theta_{k,k'}\le\theta_{k,s},\quad\text{if}\quad k'\le s.\end{aligned}\tag{4.44}$$

4.3.2 受限等距常数、受限正交常数及其他常数之间的关系

本小节将探讨与设计矩阵 $\boldsymbol{X}$ 相关的各种常数之间的联系，包括受限等距常数、互相干性和受限正交常数。

命题 4.39：（受限等距常数与相干性）设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 的列经过 ℓ_2 归一化，即对所有 $i\in\{1,2,\dots,p\}$，都有 $\|\boldsymbol{x}_i\|_2=1$。则受限等距常数（RIC）满足

$$\delta_1=0,\quad\delta_2=\mu(\boldsymbol{X}),\quad\delta_k\le\mu_1(\boldsymbol{X},k-1)\le(k-1)\mu(\boldsymbol{X}),\quad k\ge 2.$$

证明 [4.39] ℓ_2 $\|\boldsymbol{X}\boldsymbol{e}_i\|_2^2=\|\boldsymbol{e}_i\|_2^2$ 1 命题。列的 -归一化意味着，对每个标准基向量 ，因此，RIP 条件对 -稀疏向量恰好成立，从而得到 $\delta_1=0$。对于 δ_2，任取一对不同的索引 $i\neq j$。相应子矩阵的格拉姆矩阵满足

$$\delta_2=\max_{1\le i\neq j\le p}\left\|\boldsymbol{X}[:,\{i,j\}]^\top\boldsymbol{X}[:,\{i,j\}]-\boldsymbol{I}_2\right\|_2.$$

该对称矩阵的特征值 $\boldsymbol{X}[:\ \{i,j\}]^\top\boldsymbol{X}[:\ \{i,j\}]-\boldsymbol{I}=\begin{bmatrix}0 & \langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle\\ \langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle & 0\end{bmatrix}$为 $\pm|\langle\boldsymbol{x}_i,\ \boldsymbol{x}_j\rangle|$，因此其谱范数等于 $|\langle\boldsymbol{x}_i,\ \boldsymbol{x}_j\rangle|$。对$1\le i\neq j\le p$ 取最大值可得 $\delta_2=\mu$（$\boldsymbol{X}$），即互相干性。不等式 $\delta_k\le\mu_1$（$\boldsymbol{X}$，$k-1$）$\le$（$k-1$）μ（$\boldsymbol{X}$）由定理 4.24 和（4.29a）得出。

■

现在，我们来比较受限等距常数与受限正交常数。该结果能够控制由 $\boldsymbol{X}$ 中大小分别为 s和 t 的不相交列子集所张成的空间之间的主角。这样的控制很有价值，因为它使我们能够仅依据涉及受限等距常数（RIC）的界来保证精确稀疏恢复；更多详情请参见第5.3 节。

命题 4.40：（受限等距常数与受限正交常数（Candes 和 Tao，2005年；Foucart 和 Rauhut，2013年））对于任意正整数 s，t，都有

$$\theta_{s,t} \leq \delta_{s+t} \leq \begin{cases} \theta_{s,t} + \max\{\delta_s, \delta_t\}; \\ \frac{1}{s+t}(s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t}). \end{cases}$$

在特殊情形 $t = s$ 下，可得

$$\theta_{s,s} \leq \delta_{2s} \leq \delta_s + \theta_{s,s}.$$

证明 [命题 4.40] **不等式** $\theta_{s,\,t} \leq \delta_{s+t}$。不等式 $\theta_{s,\,t} \leq \delta_{s+t}$ 可直接由定义得出。或者，利用齐次性，只需证明

$$\left| \left\langle \sum_{i\in\mathbb{I}} \beta_i \boldsymbol{x}_i, \sum_{j\in\mathbb{J}} \beta_j \boldsymbol{x}_j \right\rangle \right| \leq \delta_{s+t} \tag{4.45}$$

当 $\mathbb{I}$，$\mathbb{J} \subseteq \{1, 2, \ldots, p\}$ 互不相交（$\mathbb{I} \cap \mathbb{J} = \varnothing$）时，$|\mathbb{I}| \leq s$，$|\mathbb{J}| \leq t$，且 $\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\|\boldsymbol{\beta}_{\mathbb{I}}\|_2 = \|\boldsymbol{\beta}_{\mathbb{J}}\|_2 = 1$，即 $\sum_{i\in\mathbb{I}} |\beta_i|^2 = \sum_{j\in\mathbb{J}} |\beta_j|^2 = 1$。由此可知，（4.43）表明

$$2(1-\delta_{s+t}) \leq \left\| \sum_{i\in\mathbb{I}} \beta_i \boldsymbol{x}_i + \sum_{j\in\mathbb{J}} \beta_j \boldsymbol{x}_j \right\|_2^2 \leq 2(1+\delta_{s+t});$$

$$2(1-\delta_{s+t}) \leq \left\| \sum_{i\in\mathbb{I}} \beta_i \boldsymbol{x}_i - \sum_{j\in\mathbb{J}} \beta_j \boldsymbol{x}_j \right\|_2^2 \leq 2(1+\delta_{s+t}).$$

不等式（4.45）随后由平行四边形恒等式 $4\langle \boldsymbol{a},\ \boldsymbol{b}\rangle = \|\boldsymbol{a}+\boldsymbol{b}\|_2^2 - \|\boldsymbol{a}-\boldsymbol{b}\|_2^2$ 对任意向量 $\boldsymbol{a}$，$\boldsymbol{b}$均成立。

不等式 $\delta_{s+t} \leq \theta_{s,\,t} + \max\{\delta_s, \delta_t\}$。对于反向不等式，再次利用齐次性，考虑任意一个 $(s+t)$-稀疏向量 $\boldsymbol{\beta}$，满足 $\|\boldsymbol{\beta}\|_2 = 1$。将其支撑集划分为 $\mathbb{S} = \mathbb{I} \cup \mathbb{J}$，其中 $|\mathbb{I}| \leq s$、$|\mathbb{J}| \leq t$，并定义 $\nu \triangleq \sum_{i\in\mathbb{I}} |\beta_i|^2$，使得 $1 - \nu = \sum_{j\in\mathbb{J}} |\beta_j|^2$。于是

$$\begin{aligned} \left| \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 - 1 \right| &= \left| \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 - \|\boldsymbol{\beta}\|_2^2 \right| \\ &\leq \left| \|\boldsymbol{X}_{\mathbb{I}}\boldsymbol{\beta}_{\mathbb{I}}\|_2^2 - \nu \right| + \left| \|\boldsymbol{X}_{\mathbb{J}}\boldsymbol{\beta}_{\mathbb{J}}\|_2^2 - (1-\nu) \right| + 2\left| \langle \boldsymbol{X}_{\mathbb{I}}\boldsymbol{\beta}_{\mathbb{I}}, \boldsymbol{X}_{\mathbb{J}}\boldsymbol{\beta}_{\mathbb{J}} \rangle \right| \\ &\leq \delta_s\nu + \delta_t(1-\nu) + 2\theta_{s,t}\sqrt{\nu(1-\nu)} \leq \max\{\delta_s, \delta_t\}(\nu + 1 - \nu) + \theta_{s,t} \\ &= \max\{\delta_s, \delta_t\} + \theta_{s,t}, \end{aligned} \tag{4.46}$$

其中，第一个不等式来自三角不等式，第二个不等式由 (4.36a) 和 (4.42a) 得出，它们是受限等距常数（RIC）和受限正交常数的定义。

不等式 $\delta_{s+t} \leq \frac{1}{s+t}(s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,\,t})$。这个不等式给出了受限等距常数（RIC）与受限正交常数之间更精确的关系。再次令 $\boldsymbol{\beta} \in \mathbb{R}^p$ 为一个 $(s+t)$-稀疏向量，满足$\|\boldsymbol{\beta}\|_2 = 1$。根据 (4.46)，可得

$$\left| \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2 \right| \leq \delta_s\nu + \delta_t(1-\nu) + 2\theta_{s,t}\sqrt{\nu(1-\nu)} \triangleq f(\nu), \tag{4.47}$$

其中 $f(\nu)$：$[0,1] \to \mathbb{R}$。可以证明，存在一个标量 $\nu^* \in [0,1]$ 使得该函数在 $[0, \nu^*]$ 上非递减，随后在 $[\nu^*, 1]$上非递增；参见习题 4.10。

根据这个 ν^* 相对于 $s/(s+t)$ 的位置，函数 f 要么在 $[0, s/(s+t)]$ 上非递减，要么在 $[s/(s+t), 1]$ 上非递增。通过适当选择向量 $\boldsymbol{\beta}_{\mathbb{I}}$，我们总可以假设 $\|\boldsymbol{\beta}_{\mathbb{I}}\|_2^2$ 位于其中一个区间内。事实上，如果 $\boldsymbol{\beta}_{\mathbb{I}}$ 由 $\boldsymbol{\beta}$ 中模值最小的 s 个分量构成，而 $\boldsymbol{\beta}_{\mathbb{J}}$ 由 $\boldsymbol{\beta}$中模值最大的 t个分量构成，那么

$$\frac{\|\boldsymbol{\beta}_{\mathbb{I}}\|_2^2}{s} \leq \frac{\|\boldsymbol{\beta}_{\mathbb{J}}\|_2^2}{t} = \frac{1-\|\boldsymbol{\beta}_{\mathbb{I}}\|_2^2}{t}, \quad \text{so that } \|\boldsymbol{\beta}_{\mathbb{I}}\|_2^2 \leq \frac{s}{s+t}.$$

否则，如果 $\boldsymbol{\beta}_{\mathbb{I}}$ 由 $\boldsymbol{\beta}$中模最大的 s 个分量组成，则有 $\|\boldsymbol{\beta}_{\mathbb{I}}\|_2^2 \geq s/(s+t)$。因此，我们得到

$$\left|\|\boldsymbol{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| \leq f\left(\frac{s}{s+t}\right) = \delta_s \frac{s}{s+t} + \delta_t \frac{t}{s+t} + 2\theta_{s,t}\frac{\sqrt{st}}{s+t}.$$

证明完毕。 ■

本小节最后说明，高阶的受限等距常数（RIC）和受限正交常数可以用低阶的相应常数给出上界。

命题 4.41：（不同阶的受限等距常数和受限正交常数） 令 k、s和$t \geq 1$为整数，且 $t \geq s$。则

$$\theta_{t,k} \leq \sqrt{\frac{t}{s}}\theta_{s,k},$$
$$\delta_t \leq \frac{t-d}{s}\delta_{2s} + \frac{d}{s}\delta_s, \quad \text{where } d \triangleq \gcd(s,t).$$

在特殊情况下，$t = cs$ 对于某个整数 $c \geq 1$，这可简化为

$$\delta_{cs} \leq c \cdot \delta_{2s}.$$

证明 [命题 4 的41] 设 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 是一个 t-稀疏向量，且 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是一个 k-稀疏向量，二者的支撑集不相交。我们旨在证明以下两个不等式：

$$|\langle \boldsymbol{X}\boldsymbol{\alpha}, \boldsymbol{X}\boldsymbol{\beta}\rangle| \leq \sqrt{\frac{t}{s}}\theta_{s,k}\|\boldsymbol{\alpha}\|_2\|\boldsymbol{\beta}\|_2, \tag{4.48}$$

$$\left|\|\boldsymbol{X}\boldsymbol{\alpha}\|_2^2 - \|\boldsymbol{\alpha}\|_2^2\right| \leq \left(\frac{t-d}{s}\delta_{2s} + \frac{d}{s}\delta_s\right)\|\boldsymbol{\alpha}\|_2^2. \tag{4.49}$$

设 d 是 s 和 t的公因数。我们定义整数 μ，ν，使得

$$s = \mu d, \qquad t = \nu d.$$

设 $\mathbb{T} = \{j_1, j_2, \ldots, j_t\}$表示 $\boldsymbol{\alpha}$的支撑集。我们考虑由以下方式定义的、大小为 s的 ν个子集 $\mathbb{S}_1$, $\mathbb{S}_2$, ..., $\mathbb{S}_\nu \subseteq \mathbb{T}$

$$\mathbb{S}_i = \{j_{(i-1)d+1}, j_{(i-1)d+2}, \ldots, j_{(i-1)d+s}\},$$

其中，索引按模 t计算。通过这种构造，每个索引 $j \in \mathbb{T}$ 恰好出现在子集 $\mathbb{S}_i$中的 $s/d = \mu$ 个子集中。因此，我们可以将 $\boldsymbol{\alpha}$ 分解为

$$\boldsymbol{\alpha} = \frac{1}{\mu}\sum_{i=1}^{\nu}\boldsymbol{\alpha}(\mathbb{S}_i), \qquad \|\boldsymbol{\alpha}\|_2^2 = \frac{1}{\mu}\sum_{i=1}^{\nu}\|\boldsymbol{\alpha}(\mathbb{S}_i)\|_2^2.$$

再次注意，$\boldsymbol{\alpha}(\mathbb{S}_i) \in \mathbb{R}^p$ (当 $\boldsymbol{\alpha}_{\mathbb{S}_i} \in \mathbb{R}^s$；参见定义 1.1）。

（证明4.48）。 利用上面的分解以及受限正交常数的定义，我们得到

$$\langle \boldsymbol{X}\boldsymbol{\alpha}, \boldsymbol{X}\boldsymbol{\beta}\rangle \le \frac{1}{\mu}\sum_{i=1}^{\nu}\langle \boldsymbol{X}\boldsymbol{\alpha}(\mathbb{S}_i), \boldsymbol{X}\boldsymbol{\beta}\rangle \le \frac{1}{\mu}\sum_{i=1}^{\nu}\theta_{s,k}\,\|\boldsymbol{\alpha}(\mathbb{S}_i)\|_2\,\|\boldsymbol{\beta}\|_2$$
$$\le \theta_{s,k}\frac{\sqrt{\nu}}{\mu}\sqrt{\left(\sum_{i=1}^{\nu}\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2^2\right)}\,\|\boldsymbol{\beta}\|_2 = \theta_{s,k}\sqrt{\frac{\nu}{\mu}}\,\|\boldsymbol{\alpha}\|_2\,\|\boldsymbol{\beta}\|_2,$$

其中，最后一个不等式由柯西–施瓦茨不等式得出 $\boldsymbol{a}^\top\boldsymbol{b} \le \|\boldsymbol{a}\|_2\,\|\boldsymbol{b}\|_2$，其方法是代入 $\boldsymbol{a}=\mathbf{1}$ 和 $b_i = \|\boldsymbol{\alpha}(\mathbb{S}_i)\|_2$，对所有 i均成立。

（4.49）的证明。我们先写出

$$\|\boldsymbol{X}\boldsymbol{\alpha}\|_2^2 - \|\boldsymbol{\alpha}\|_2^2 = \left|\left\langle(\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I})\boldsymbol{\alpha},\boldsymbol{\alpha}\right\rangle\right| \le \frac{1}{\mu^2}\sum_{1\le i\le\nu}\sum_{1\le j\le\nu}\left|\left\langle(\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I})\boldsymbol{\alpha}(\mathbb{S}_i),\boldsymbol{\alpha}(\mathbb{S}_j)\right\rangle\right|$$
$$= \frac{1}{\mu^2}\left(\sum_{1\le i\ne j\le\nu}\left|\left\langle(\boldsymbol{X}^\top\boldsymbol{X}-\boldsymbol{I})\boldsymbol{\alpha}(\mathbb{S}_i),\boldsymbol{\alpha}(\mathbb{S}_j)\right\rangle\right| + \sum_{1\le i\le\nu}\left|\left\langle(\boldsymbol{X}_{\mathbb{S}_i}^\top\boldsymbol{X}_{\mathbb{S}_i}-\boldsymbol{I})\boldsymbol{\alpha}_{\mathbb{S}_i},\boldsymbol{\alpha}_{\mathbb{S}_i}\right\rangle\right|\right)$$
$$\overset{\dagger}{\le} \frac{1}{\mu^2}\left(\sum_{1\le i\ne j\le\nu}\delta_{2s}\,\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2\,\|\boldsymbol{\alpha}_{\mathbb{S}_j}\|_2 + \sum_{1\le i\le\nu}\delta_s\,\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2^2\right) = \frac{\delta_{2s}}{\mu^2}\left(\sum_{1\le i\le\nu}\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2\right)^2 - \frac{\delta_{2s}-\delta_s}{\mu^2}\sum_{1\le i\le\nu}\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2^2$$
$$\overset{\ddagger}{\le}\left(\frac{\delta_{2s}\nu}{\mu^2}-\frac{\delta_{2s}-\delta_s}{\mu^2}\right)\sum_{1\le i\le\nu}\|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2^2 = \left(\frac{\nu}{\mu}\delta_{2s}-\frac{\delta_{2s}-\delta_s}{\mu}\right)\|\boldsymbol{\alpha}\|_2^2 = \left(\frac{t}{s}\delta_{2s}-\frac{1}{\mu}(\delta_{2s}-\delta_s)\right)\|\boldsymbol{\alpha}\|_2^2,$$

其中，不等式（†）由命题4.36得出，而不等式（‡）再次由柯西–施瓦茨不等式得出 $\boldsymbol{a}^\top\boldsymbol{b} \le \|\boldsymbol{a}\|_2\,\|\boldsymbol{b}\|_2$，其方法是代入 $\boldsymbol{a}=\mathbf{1}$ 和$b_i = \|\boldsymbol{\alpha}_{\mathbb{S}_i}\|_2$，对所有 i均成立。为使后一个上界尽可能紧，我们取 μ 尽可能小，即选择 d 作为 s和 t的最大公约数。证明完毕。

■

4.3.3 受限等距性质的界

与互相干性或 ℓ_1-相干性（见第4.2.4节）一样，理解矩阵 k的第 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 个受限等距常数（RIC）可以有多小非常重要。在与压缩感知相关的情形下——即 $p \ge Cn$时——下面的定理4.42 表明，RIC必须满足$\delta_k \ge c\sqrt{k/n}$。当 $k=2$时，由命题 $\mu(\boldsymbol{X}) \ge c'/\sqrt{n}$可知，互相干性满足4.22，这与定理4.27中的 Welch 界相似。

定理 4.42：（受限等距常数（Krahmer 和 Ward，2011年；Foucart 和 Rauhut，2013年）） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 和 $2\le k\le p$。假设 $p\ge Cn$ 且 $\delta_k\le\delta_*$，其中常数 c，C和 δ_* 仅彼此相关。则

$$n \ge c\frac{k}{\delta_k^2}. \tag{4.50}$$

例如，$c=1/162$、$C=30$和 $\delta_*=2/3$ 这些选取都是有效的。

定理 [的证明 4.42] 我们首先注意到，当 $k=1$时，该结论不成立，$\delta_1=0$ 因为只要 $\boldsymbol{X}$ 的所有列都归一化为单位 ℓ_2-范数（见命题 4.39）。令 $t\triangleq\lfloor k/2\rfloor\ge 1$，并将 $\boldsymbol{X}$ 划分为大小为 $n\times t$ 的块（最后一块的列数可能较少）：

$$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_\nu], \quad p \le \nu t.$$

根据定义4.30和 4.37、习题 4.34以及命题 4.40，对于任意互不相同的$i,\ j \in \{1,2,\ldots,\nu\},\ i \neq j$,

$$\left\|\boldsymbol{X}_i^\top \boldsymbol{X}_i - \boldsymbol{I}\right\|_2 \leq \delta_t \leq \delta_k, \qquad \left\|\boldsymbol{X}_i^\top \boldsymbol{X}_j\right\|_2 \leq \theta_{t,t} \leq \delta_{2t} \leq \delta_k.$$

因此，$\boldsymbol{X}_i^\top \boldsymbol{X}_i$ 的所有特征值以及 $\boldsymbol{X}_i^\top \boldsymbol{X}_j$ 的所有奇异值均满足 y

$$1-\delta_k \leq \lambda_h(\boldsymbol{X}_i^\top \boldsymbol{X}_i) \leq 1+\delta_k, \qquad \sigma_h(\boldsymbol{X}_i^\top \boldsymbol{X}_j) \leq \delta_k, \qquad h=1,2,\ldots,t$$

定义格拉姆矩阵及其对偶问题：

$$\boldsymbol{G} \triangleq \boldsymbol{X}^\top \boldsymbol{X} = [\boldsymbol{X}_i^\top \boldsymbol{X}_j]_{1\leq i,j\leq \nu} \in \mathbb{R}^{p\times p}, \qquad \boldsymbol{H} \triangleq \boldsymbol{X}\boldsymbol{X}^\top \in \mathbb{R}^{n\times n}.$$

下界：$\mathrm{tr}(\boldsymbol{H})$。利用迹的循环不变性以及每个$\boldsymbol{X}_i^\top \boldsymbol{X}_i$ 都是对称矩阵这一事实，我们得到

$$\mathrm{tr}(\boldsymbol{H}) = \mathrm{tr}(\boldsymbol{G}) = \sum_{i=1}^{\nu} \mathrm{tr}(\boldsymbol{X}_i^\top \boldsymbol{X}_i) = \sum_{i=1}^{\nu}\sum_{h=1}^{t} \lambda_h(\boldsymbol{X}_i^\top \boldsymbol{X}_i) \geq \nu t(1-\delta_k). \tag{4.51}$$

利用弗罗贝尼乌斯范数得到上界。回顾一下，对于任意矩阵 $\boldsymbol{A}_1,\ \boldsymbol{A}_2$，弗罗贝尼乌斯内积满足 $\langle \boldsymbol{A}_1,\ \boldsymbol{A}_2\rangle_F = \mathrm{tr}(\boldsymbol{A}_2^\top \boldsymbol{A}_1)$，根据柯西–施瓦茨不等式，

$$\mathrm{tr}(\boldsymbol{H})^2 = \langle \boldsymbol{I}_n, \boldsymbol{H}\rangle_F^2 \leq \|\boldsymbol{I}_n\|_F^2 \, \|\boldsymbol{H}\|_F^2 = n\mathrm{tr}(\boldsymbol{H}^\top \boldsymbol{H}).$$

现在，再次利用循环不变性以及 $\boldsymbol{X}\boldsymbol{X}^\top = \sum_i \boldsymbol{X}_i \boldsymbol{X}_i^\top$这一事实，我们有

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{H}^\top \boldsymbol{H}) &= \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top) = \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}) = \mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^\top) \\
&= \sum_{i=1}^{\nu} \mathrm{tr}\left(\sum_{j=1}^{\nu} \boldsymbol{X}_i^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{X}_i\right) \overset{\dagger}{=} \sum_{1\leq i\neq j\leq \nu}\sum_{h=1}^{t} \sigma_h(\boldsymbol{X}_i^\top \boldsymbol{X}_j)^2 + \sum_{i=1}^{\nu}\sum_{h=1}^{t} \lambda_h(\boldsymbol{X}_i^\top \boldsymbol{X}_i)^2 \\
&\leq \nu(\nu-1)t\delta_k^2 + \nu t(1+\delta_k)^2.
\end{aligned}$$

其中，等式（†）由对称矩阵的迹等于其特征值之和这一事实，以及对于任意矩阵$\boldsymbol{A}$，$\sigma(\boldsymbol{A})^2 = \lambda(\boldsymbol{A}^\top \boldsymbol{A}) = \lambda(\boldsymbol{A}\boldsymbol{A}^\top)$这一事实得出。这意味着上界

$$\mathrm{tr}(\boldsymbol{H})^2 \leq n\mathrm{tr}(\boldsymbol{H}^\top \boldsymbol{H}) \leq n\nu t\big((\nu-1)\delta_k^2 + (1+\delta_k)^2\big). \tag{4.52}$$

结合界（4.51）和（4.52），得到样本数量的下界：

$$n \geq \frac{\nu t(1-\delta_k)^2}{(\nu-1)\delta_k^2 + (1+\delta_k)^2}. \tag{4.53}$$

若$(\nu-1)\,\delta_k^2 < (1+\delta_k)^2/5$，利用 $\delta_k \leq 2/3$，可得

$$n > \frac{\nu t(1-\delta_k)^2}{6(1+\delta_k)^2/5} \geq \frac{5(1-\delta_k)^2}{6(1+\delta_k)^2} p \geq \frac{1}{30} p,$$

这与我们的假设矛盾。因此，我们有 $(\nu-1)\delta_k^2 \geq (1+\delta_k)^2/5$，再次利用 $\delta_k \leq 2/3$ 和 $k \leq 3t$ 可得，

$$n \geq \frac{\nu t(1-\delta_k)^2}{6(\nu-1)\delta_k^2} \geq \frac{1}{54\delta_k^2} t \geq \frac{1}{162\delta_k^2} k.$$

证明至此完成。 ■

我们将在第 7 章（例如，推论 7.26）中看到，某些随机矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 以高概率满足 $\delta_k \le \delta$，对任意固定的 $\delta \in (0,\ 1)$，只要

$$n \ge C\delta^{-2} k \ln(ep/k).$$

因此，下界（4.50）除对数因子外是最优的。

定理 4.42考察了 RIP 常数 δ 固定时的测量值界。我们还可以提出一个互补的问题：如果不考虑 δ 的具体取值，而只关注问题维数 n，p和k，满足 RIP 至少需要多少测量值？在这种情况下，我们可以推导出 n的一个简单下界。为此，我们首先证明一个关键引理，它将在主要结果的证明中用到。

引理 4.43 设 k 和 p 为满足 $k < p/2$的整数。那么存在一个集合$\mathbb{S} \subset \mathbb{B}_0[k]$，使得

(i) $\|\boldsymbol{\beta}\|_2 \le \sqrt{k}$ 对所有 $\boldsymbol{\beta} \in \mathbb{S}$均成立。(ii) $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 \ge \sqrt{k/2}$ 对所有互不相同的 $\boldsymbol{\alpha},\ \boldsymbol{\beta} \in \mathbb{S}$ 且满足 $\boldsymbol{\beta} \ne \boldsymbol{\alpha}$ 的情形均成立。(iii) $\ln|\mathbb{S}| \ge \frac{k}{2}\ln\binom{p}{k}$。

证明 [引理4.43] 考虑集合

$$\mathbb{T} \triangleq \{\boldsymbol{\beta} \in \{0,+1,-1\}^p \mid \|\boldsymbol{\beta}\|_0 = k\}.$$

根据构造，每个 $\boldsymbol{\beta} \in \mathbb{T}$ 都满足 $\|\boldsymbol{\beta}\|_2 = \sqrt{k}$。因此，$\mathbb{T}$ 的任意子集都自动满足条件 (i)。

注意，$\mathbb{T}$ 的基数为 $|\mathbb{T}| = \binom{p}{k}2^k$。此外，由于 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_0 \le \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2$（当 $\boldsymbol{\alpha},\ \boldsymbol{\beta} \in \mathbb{T}$ 时），可知若 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 \le k/2$，则 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_0 \le k/2$。因此，对于任意固定的 $\boldsymbol{\beta} \in \mathbb{T}$，平方 ℓ_2-距离不超过 $k/2$ 的 $\boldsymbol{\alpha} \in \mathbb{T}$ 数量满足

$$\left|\{\boldsymbol{\alpha} \in \mathbb{T} : \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 \le k/2\}\right| \le \left|\{\boldsymbol{\alpha} \in \mathbb{T} \mid \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_0 \le k/2\}\right| \le \binom{p}{k/2}3^{k/2}.$$

现在我们贪心地构造 $\mathbb{S}$：从 $\mathbb{T}$ 开始，迭代地选择在 ℓ_2-范数下彼此至少相距 $\sqrt{k/2}$ 的向量。选取 j 个点后，$\mathbb{T}$ 中至多有 $j\binom{p}{k/2}3^{k/2}$ 个点会被排除在后续选择之外。因此，只要

$$|\mathbb{S}|\binom{p}{k/2}3^{k/2} \le \binom{p}{k}2^k \equiv |\mathbb{T}|. \tag{4.54}$$

存在这样一个大小为 $|\mathbb{S}|$ 的集合 $\mathbb{S}$。为验证这个不等式，注意到

$$\frac{\binom{p}{k}}{\binom{p}{k/2}} = \frac{(k/2)!(p-k/2)!}{k!(p-k)!} = \prod_{i=1}^{k/2}\frac{p-k+i}{k/2+i} \ge \left(\frac{p}{k} - \frac{1}{2}\right)^{k/2},$$

其中不等式成立，是因为函数 $f(i) \triangleq (p-k+i)/(k/2+i)$ 在 i [6]上递减。现在令 $|\mathbb{S}| = (p/k)^{k/2}$。于是

$$|\mathbb{S}|\left(\frac{3}{4}\right)^{k/2} = \left(\frac{3p}{4k}\right)^{k/2} = \left(\frac{p}{k} - \frac{p}{4k}\right)^{k/2} \overset{\dagger}{\le} \left(\frac{p}{k} - \frac{1}{2}\right)^{k/2} \le \frac{\binom{p}{k}}{\binom{p}{k/2}}, \tag{4.55}$$

6。由于在假设 ~~$k < p/2$~~ 成立时，$f'(i) = \frac{3k/2-p}{(k/2+i)^2}$ 为负。

其中不等式 (†) 由 $k < p/2$ 的假设可得。结合 (4.54) 和 (4.55)，可知 (4.54) 对 $|\mathbb{S}| = (p/k)^{k/2}$成立，从而证明了该引理。 ■

利用这一打包引理，我们现在推导出 RIP 所需测量数量的一个基本下界。

定理 4.44：（RIP 的测量数量下界（Garnaev 和 Gluskin，1984年；Baraniuk 等人，2011年）） 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 $2k$ 、常数为$\delta \in (0, \frac{1}{2}]$ [a]）的 RIP。则，

$$n \geq Ck \ln\left(\frac{p}{k}\right), \quad \text{with } C \triangleq \frac{1}{2\ln(\sqrt{24}+1)} \approx 0.28. \tag{4.56}$$

[a] 注意，限制条件 $\delta \leq \frac{1}{2}$ 是为方便起见而作出的；稍微调整常数后，同样的论证适用于任意 $\delta_{\max} < 1$。

证明 [定理4.44] 由于 $\boldsymbol{X}$ 满足阶为 $2k$的 RIP，因此，对任意 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{S}$ （其中 $\mathbb{S}$来自引理4.43），我们有 $\boldsymbol{\beta} - \boldsymbol{\alpha} \in \mathbb{B}_0[2k]$，因此

$$\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\alpha}\|_2 \geq \sqrt{1-\delta}\,\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 \geq \sqrt{k/4}, \quad \forall\, \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{S}. \tag{4.57}$$

同理，利用 RIP 的上界，

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_2 \leq \sqrt{1+\delta}\,\|\boldsymbol{\beta}\|_2 \leq \sqrt{3k/2}, \quad \forall\, \boldsymbol{\beta} \in \mathbb{S}. \tag{4.58}$$

下界告诉我们，对于任意一对点 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{S}$，若分别以 $\boldsymbol{X}\boldsymbol{\beta}$ 和 $\boldsymbol{X}\boldsymbol{\alpha}$ 为球心、以 $\sqrt{k/4}/2 = \sqrt{k/16}$ 为半径作球，则这些球互不相交。反过来，上界告诉我们，所有这些球构成的集合都包含在一个半径为$\sqrt{3k/2} + \sqrt{k/16}$的大球内。令 $\mathbb{B}_2[r] = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq r\}$ 表示 ℓ_2-球。这意味着

$$\begin{aligned}
&\operatorname{Vol}\left(\mathbb{B}_2\left[\sqrt{3k/2} + \sqrt{k/16}\right]\right) \geq |\mathbb{S}| \cdot \operatorname{Vol}\left(\mathbb{B}_2\left[\sqrt{k/16}\right]\right) \\
\Longrightarrow\quad &\left(\sqrt{3k/2} + \sqrt{k/16}\right)^n \geq |\mathbb{S}| \cdot \left(\sqrt{k/16}\right)^n \\
\Longrightarrow\quad &(\sqrt{24}+1)^n \geq |\mathbb{S}| \\
\Longrightarrow\quad &n \geq \frac{\ln|\mathbb{S}|}{\ln(\sqrt{24}+1)}.
\end{aligned} \tag{4.59}$$

根据引理 $|\mathbb{S}|$ 中的上界4.43，证明得证。 ■

定理 4.42 给出了 n 的一个下界，该下界显式依赖于限制等距性质常数δ_k 和稀疏度 k，但不依赖于 p （环境维数），条件 $p \geq Cn$除外。定理 4.44 给出了 n 的一个下界，该下界依赖于 k和对数比值 $\ln(p/k)$，但假设 δ 的取值范围固定（例如，$\delta \leq 1/2$），且其上界中未显式涉及 δ。虽然上述证明略显间接，但也可以通过 p -球的 Gelfand 宽度 k 推导出 ℓ_1 和 Garnaev 和 Gluskin，1984 年）下类似的缩放关系。

4.3.4 限制等距性质与稳定性

我们将在本书后文看到，如果矩阵 $\boldsymbol{X}$ 满足 RIP，那么这一条件足以保证一大类算法能够成功恢复稀疏

信号的含噪测量值；参见第5.3节。这里，我们考察 RIP 是否也是必要条件。显然，如果我们希望从无噪声测量值中精确恢复所有 k-稀疏信号 $\boldsymbol{\beta}$ ，那么 RIP 中的下界是必要的 $\boldsymbol{X}\boldsymbol{\beta}$，理由与 NSP 是必要条件的原因相同（参见习题4.2）。事实上，通过引入以下稳定性概念，我们可以进一步说明 RIP 的必要性。

定义 4.45（C-稳定（Baraniuk 等人，2011年））。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个感知矩阵，且 $\Delta : \mathbb{R}^n \to \mathbb{R}^p$ 是一个恢复算法。我们称这对（$\boldsymbol{X}$， Δ）是 C-稳定的，如果对于任意 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 以及任意 $\boldsymbol{\epsilon} \in \mathbb{R}^n$，

$$\|\Delta(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) - \boldsymbol{\beta}\|_2 \leq C \|\boldsymbol{\epsilon}\|_2 . \tag{4.60}$$

这个定义只是说明，在测量值中加入少量噪声，不应导致恢复信号产生任意大的误差。下面的引理表明，任何能够从含噪测量中稳定恢复信号的解码算法——即使该算法在计算上可能不切实际——的存在，都意味着 $\boldsymbol{X}$ 必须满足（4.36a）中的下界，其常数由 C决定。

引理4.46：（C-稳定（Baraniuk 等人，2011年））如果这对（$\boldsymbol{X}$， Δ）是 C-稳定的，那么

$$\frac{1}{C}\|\boldsymbol{\beta}\|_2 \leq \|\boldsymbol{X}\boldsymbol{\beta}\|_2 \tag{4.61}$$

对所有 $\boldsymbol{\beta} \in \mathbb{B}_0[2k]$。

证明 [引理 4.46] 设 $\boldsymbol{\alpha}$， $\boldsymbol{\gamma} \in \mathbb{B}_0[k]$。定义

$$\boldsymbol{\epsilon}_\beta \triangleq \frac{\boldsymbol{X}(\boldsymbol{\alpha} - \boldsymbol{\gamma})}{2} \qquad \text{and} \qquad \boldsymbol{\epsilon}_\alpha = \frac{\boldsymbol{X}(\boldsymbol{\gamma} - \boldsymbol{\alpha})}{2},$$

并注意到

$$\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}_\beta = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_\alpha = \frac{\boldsymbol{X}(\boldsymbol{\gamma} + \boldsymbol{\alpha})}{2}.$$

设 $\widehat{\boldsymbol{\beta}} \triangleq \Delta(\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}_\beta) = \Delta(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_\alpha)$。根据三角不等式和$C$-稳定性的定义，

$$\begin{aligned}\|\boldsymbol{\gamma} - \boldsymbol{\alpha}\|_2 = \left\|\boldsymbol{\gamma} - \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2 &\leq \left\|\boldsymbol{\gamma} - \widehat{\boldsymbol{\beta}}\right\|_2 + \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\alpha}\right\|_2 \\ &\leq C\|\boldsymbol{\epsilon}_\beta\|_2 + C\|\boldsymbol{\epsilon}_\alpha\|_2 \leq C\|\boldsymbol{X}\boldsymbol{\gamma} - \boldsymbol{X}\boldsymbol{\alpha}\|_2 .\end{aligned}$$

由于这对任意 $\boldsymbol{\alpha}$， $\boldsymbol{\gamma} \in \mathbb{B}_0[k]$ 都成立，因此对任意 $\boldsymbol{\beta} \in \mathbb{B}_0[2k]$ （令$\boldsymbol{\beta} = \boldsymbol{\gamma} - \boldsymbol{\alpha}$），都有 $\|\boldsymbol{\beta}\|_2 \leq C\|\boldsymbol{X}\boldsymbol{\beta}\|_2$，证明完毕。 ■

注意，由于 $C \to 1$，矩阵 $\boldsymbol{X}$ 必须满足（4.36a）中的下界，其中 $\delta_k \triangleq 1 - 1/C^2 \to 0$。因此，为了减小噪声对重建信号的影响，我们必须选择 $\boldsymbol{X}$ 使其满足具有更严格（即更小）常数的 RIP 下界。

由于只需要限制等距性质的下界——而不需要上界——我们可以通过简单地对其进行缩放，避免重新设计 $\boldsymbol{X}$ 。事实上，如果 $\boldsymbol{X}$ 满足带有$\delta_{2k} < 1$的限制等距性质，那么缩放后的矩阵 $\eta\boldsymbol{X}$ 将满足不等式（4.61）以及任意所需的常数C，前提是适当选择 η 。在噪声大小与 $\boldsymbol{X}$相互独立的情况下，这一推理是成立的：缩放 $\boldsymbol{X}$ 会增大测量值中的“信号”成分，而不会影响噪声，从而提高信噪比。原则上，这可以使噪声相对于信号变得可忽略。

然而，在实践中，我们无法任意缩放 $\boldsymbol{X}$，因为在许多现实场景中，噪声确实取决于 $\boldsymbol{X}$。例如，假设噪声向量 $\boldsymbol{\epsilon}$源自使用具有 B 比特的有限动态范围量化器进行的量化。如果原始测量值位于$[-R, R]$，且量化器针对该值域进行了校准，那么将 $\boldsymbol{X}$ 缩放 η倍会使测量值覆盖 $[-\eta R, \eta R]$。为此，量化器的动态范围也必须按 η进行缩放，从而产生量化误差 $\eta\boldsymbol{\epsilon}$。因此，重构误差不会减小——它会随 η成比例缩放，无法带来任何净改进。

4.3.5 RIP、NSP 及其变体

本小节将说明，若一个矩阵满足 RIP，则它也满足 NSP 的多种形式。因此，RIP 严格强于 NSP。

我们首先说明 RIP 如何推出标准 NSP。下述观察——这是对向量范数标准界的一种推广（习题 1.26）——在我们的论证中经常用到，因此我们将其从证明中单独列出。

引理 4.47：（有序元素范数不等式（Foucart 和 Rauhut，2013年））给定 $s > t > 0$，若 $\boldsymbol{u} \in \mathbb{R}^p$ 和 $\boldsymbol{v} \in \mathbb{R}^m$ 满足

$$\max_{i \in \{1,2,\ldots,p\}} |u_i| \leq \min_{j \in \{1,2,\ldots,m\}} |v_j| , \tag{4.62}$$

then

$$\|\boldsymbol{u}\|_s \leq \frac{p^{1/s}}{m^{1/t}} \|\boldsymbol{v}\|_t .$$

当 $t = 1$、$s = 2$ 和 $p = m$ 时，有一个特殊情形，此时得到

$$\|\boldsymbol{u}\|_2 \leq \frac{1}{\sqrt{p}} \|\boldsymbol{v}\|_1 .$$

证明 [引理4.47] 为证明这一点，注意：

$$\frac{\|\boldsymbol{u}\|_s}{p^{1/s}} = \left[\frac{1}{p}\sum_{i=1}^{p} |u_i|^s\right]^{1/s} \leq \max_{1 \leq i \leq p} |u_i| ,$$

$$\frac{\|\boldsymbol{v}\|_t}{m^{1/t}} = \left[\frac{1}{m}\sum_{j=1}^{m} |v_j|^t\right]^{1/t} \geq \min_{1 \leq j \leq m} |v_j| .$$

证明完毕。 ■

下面的引理建立了满足 RIP 的矩阵的零空间中任意向量的最大元素与最小元素的 ℓ_1-范数之间的关键关系，表明在适当的 RIP 条件下，重要分量的能量严格受不重要分量能量的上界约束。

引理 4.48：（限制等距性质 ℓ_1-ℓ_1 上界引理）设 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 满足阶为 $2k$、常数为 δ_{2k} 的限制等距性质。对于任意非零的 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$ 以及由 k个

个 $\boldsymbol{\beta}$坐标中（按大小）最大的坐标组成的指标集，我们有

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \leq \rho \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 \quad \text{with} \quad \rho = \frac{\delta_{2k}}{1-2\delta_{2k}}.$$

特别地，如果 $\delta_{2k} < \frac{1}{3}$，那么 $\rho < 1$。

证明 [引理 4.48] 考虑 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$，并令 $\mathbb{S} \triangleq \mathbb{S}_0$ 为对应于 k 中绝对值最大的 $\boldsymbol{\beta}$ 个元素的索引集合。将 $\mathbb{S}^c$ 划分为不相交块$\mathbb{S}_1$，$\mathbb{S}_2$，..... ，每块至多包含 k 个元素，并使各块内元素的幅值按块的顺序非递增。由于 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$，我们有如下分解

$$\boldsymbol{\beta} = \boldsymbol{\beta}(\mathbb{S}) + \sum_{i\geq 1} \boldsymbol{\beta}(\mathbb{S}_i) \qquad \Longrightarrow \qquad \mathbf{0} = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}) + \sum_{i\geq 1} \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i),$$

where $\boldsymbol{\beta}(\mathbb{S}) \in \mathbb{R}^p$（见定义 1.1）。与 $\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})$ 取内积，得到

$$\mathbf{0} = \langle \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}), \boldsymbol{X}\boldsymbol{\beta} \rangle = \|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})\|_2^2 + \sum_{i\geq 1} \langle \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}), \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i) \rangle .$$

重新整理并取绝对值得到

$$\|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})\|_2^2 \leq \sum_{i\geq 1} |\langle \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}), \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i) \rangle| .$$

根据 RIP，对于任意两个支撑集不相交的向量 $\boldsymbol{u}$，$\boldsymbol{v}$ 满足 $\|\boldsymbol{u}\|_0 + \|\boldsymbol{v}\|_0 \leq 2k$时，命题4.36 表明

$$|\langle \boldsymbol{X}\boldsymbol{u}, \boldsymbol{X}\boldsymbol{v} \rangle| \leq \delta_{2k} \|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2 .$$

此外，RIP 下界给出 $\|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})\|_2^2 = \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \geq (1-\delta_{2k}) \|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$。应用这些结果，我们得到：

$$(1-\delta_{2k}) \|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \leq \delta_{2k} \|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \sum_{i\geq 1} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \quad \Longrightarrow \quad (1-\delta_{2k}) \|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \leq \delta_{2k} \sum_{i\geq 1} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 . \tag{4.63}$$

由于各块中的元素按非递增顺序排列，对于每个 $i \geq 1$，我们有：

$$\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_\infty \leq \frac{1}{k} \sum_{v\in\mathbb{S}_{i-1}} |\beta_v| = \frac{1}{k} \|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\|_1, \quad \forall\, i \geq 1.$$

根据向量范数的标准界（习题1.26），ℓ_2-范数满足：$\mathbb{S}_i$ 块

$$\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \leq \sqrt{k} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_\infty \leq \sqrt{k} \cdot \frac{1}{k} \|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\|_1 = \frac{1}{\sqrt{k}} \|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\|_1, \quad \forall\, i \geq 1.$$

（注意，该不等式可直接由引理4.47 得到，使得 $\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \leq \frac{1}{\sqrt{k}} \|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\|_1$。）将这些不等式对 $i \geq 1$ 求和可得：

$$\sum_{i\geq 1} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \leq \frac{1}{\sqrt{k}} \sum_{i\geq 1} \|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\|_1 = \frac{1}{\sqrt{k}} \sum_{i\geq 0} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_1 = \frac{1}{\sqrt{k}} \big(\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 + \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 \big). \tag{4.64}$$

[7] 使用（4.64）对（4.63）右侧进行上界估计，并使用 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \sqrt{k}\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2$ （练习1.26）替换 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1$：

$$\begin{aligned}(1-\delta_{2k})\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 &\le \delta_{2k}\cdot\frac{1}{\sqrt{k}}\big(\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_1+\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,\big)\\ &\le \delta_{2k}\cdot\frac{1}{\sqrt{k}}\big(\sqrt{k}\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2+\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,\big)=\delta_{2k}\big(\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2+\tfrac{1}{\sqrt{k}}\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,\big).\end{aligned}$$

重新整理，将 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_2$ 项归集起来，并使用 $\frac{1}{\sqrt{k}}\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{S}}\|_2$，得到：

$$\frac{1}{\sqrt{k}}(1-2\delta_{2k})\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le (1-2\delta_{2k})\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \le \frac{\delta_{2k}}{\sqrt{k}}\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,.$$

因此，有

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \frac{\delta_{2k}}{1-2\delta_{2k}}\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,.$$

因此，我们可以取 $\rho \triangleq \dfrac{\delta_{2k}}{1-2\delta_{2k}}$。特别地，若 $\delta_{2k} \le \frac{1}{3}$，则 $1-2\delta_{2k} > 1/3$ 和 $\rho < 1$。 ■

注释 4.49。在 $\|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})\|_2^2 = \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \ge (1-\delta_{2k})\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2$ 中，向量 $\boldsymbol{\beta}_{\mathbb{S}}$ 实际上是k-稀疏的，但却被视为 $2k$-稀疏的。因此，本可以采用更紧的上界 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \le \|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})\|_2^2/(1-\delta_k)$（见习题 4.34）。也可以使用受限正交常数 $\theta_{k,\,k} < \delta_{2k}$ 而不是 δ_{2k}（见命题 4.40）。采用这些改进可得到更好的充分条件 $\delta_k + 2\theta_{k,\,k} < 1$ 而非较为保守的要求 $\delta_{2k} < 1/3$。

利用上面的引理，我们现在证明 RIP 是 NSP（定义 4.1）成立的充分条件。

定理 4.50：（RIP $\Longrightarrow$ NSP）如果矩阵 $\boldsymbol{X}$ 满足 RIP$(2k,\ \delta_{2k})$，且 $\delta_{2k} < 1/3$，那么它满足 k阶的 NSP。

引理 [的证明 4.50] 根据技术引理 4.48：如果 $\boldsymbol{X}$ 满足 RIP$(2k,\ \delta_{2k})$，那么对于任意非零的 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$，以及由 $\boldsymbol{\beta}$中绝对值最大的 k 个元素构成的集合 $\mathbb{T}$，

$$\|\boldsymbol{\beta}_{\mathbb{T}}\|_1 \le \rho\,\|\boldsymbol{\beta}_{\mathbb{T}^c}\|_1\,,\quad \rho = \frac{\delta_{2k}}{1-2\delta_{2k}}.$$

如果 $\delta_{2k} \le \frac{1}{3}$，那么 $\rho < 1$。

任取满足 $|\mathbb{S}| = k$的 $\mathbb{S} \subset \{1,2,\ldots,p\}$，并令 $\mathbb{T}$为由 $\boldsymbol{\beta}$中绝对值最大的 k个元素构成的集合。根据 $\mathbb{T}$的构造：

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{T}}\|_1\,,\quad \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1 \ge \|\boldsymbol{\beta}_{\mathbb{T}^c}\|_1\,.$$

因此，

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{T}}\|_1 \le \rho\,\|\boldsymbol{\beta}_{\mathbb{T}^c}\|_1 \le \rho\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1\,,\quad \text{for all } \mathbb{S} \text{ with } |\mathbb{S}| = k.$$

这就证明了 $\boldsymbol{X}$的 k 阶 NSP。 ■

[7]。请注意，如果你转而尝试直接证明形式看起来更简单的上界 $\sum_i \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \le \frac{1}{\sqrt{k}}\,\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1$，一般来说会失败——我们需要额外的 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1$ 项，如式（4.64）所示。

值得注意的是，RIP$(2k,\ \delta_{2k})$ 条件对大量子矩阵施加了约束，具体而言，涉及其中的 $\binom{p}{2k}$ 个。不过，与成对非相干条件不同，RIP 常数 δ_{2k} 本身并不必然依赖于 k （除该性质的阶数外）。

随机矩阵理论的结果表明，在测量数量 n 按 $n \gtrsim k\ln\frac{ep}{k}$ 的速度增长时，各种随机系综以高概率满足 RIP。例如，标准高斯随机矩阵 $\boldsymbol{X}$ 的元素独立同分布，服从 $\mathcal{N}(0, \frac{1}{n})$；详情见推论7.26。因此，与成对非相干性相比，基于 RIP 的框架只需显著更少的测量值即可实现精确稀疏恢复；成对非相干性通常要求 $n \gtrsim k^2 \ln p$ (Foucart 和 Rauhut，2013年)。另一方面，RIP 的一项主要实际局限是，与成对非相干性不同，它极难直接验证，因为必须检查的子矩阵数量是组合级的 $\binom{p}{2k}$ 。

接下来我们证明，RIP 也蕴含 NSP′。我们先介绍一个关键引理，它将在定理4.54的证明中用到。该结果对任意向量 $\boldsymbol{\beta} \in \mathbb{R}^p$都成立，而不仅仅适用于 $\boldsymbol{X}$的零空间中的向量。（注意，如果 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X})$，论证可以大大简化。）

引理 4.51 设 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 是正交向量。则

$$\|\boldsymbol{\alpha}\|_2 + \|\boldsymbol{\beta}\|_2 \le \sqrt{2}\,\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_2\,.$$

证明 [引理 4.51]定义 2×1向量 $\boldsymbol{u} = [\|\boldsymbol{\alpha}\|_2\,, \|\boldsymbol{\beta}\|_2]^\top$。根据向量范数的标准界（习题 1.26），我们有 $\|\boldsymbol{u}\|_1 \le \sqrt{2}\,\|\boldsymbol{u}\|_2$。由此得到

$$\|\boldsymbol{\alpha}\|_2 + \|\boldsymbol{\beta}\|_2 \le \sqrt{2}\sqrt{\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2}.$$

由于 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 正交，$\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2 = \|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_2^2$，证毕。 ■

下一个引理给出了向量中连续各块 k 个最大元素（按大小排列）的 ℓ_2-范数之和的上界，该上界用其尾部的 ℓ_1-范数表示。它量化了向量的能量如何随着分量的重要性逐渐降低而衰减。

引理 4.52：（分块有序集不等式） 设 $\mathbb{S}_0 \subseteq \{1,2,\dots,p\}$ 是任意指标集，且满足 $|\mathbb{S}_0| \le k$。对于任意向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，定义 $\mathbb{S}_1$ 为 $\boldsymbol{\beta}_{\mathbb{S}_0^c}$中 k 个绝对值最大的元素的索引，$\mathbb{S}_2$ 为接下来 k 个绝对值最大的元素的索引，依此类推。则

$$\sum_{i\ge 2} \|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \le \frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}.$$

引理 [4 的证明。52]根据构造，$\boldsymbol{\beta}_{\mathbb{S}_i}$ 中各元素的幅值按非递增顺序排列，且 i。特别地，当 $i \ge 2$ 时，

$$\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_\infty \le \frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\right\|_1}{k}, \quad \text{for } i \ge 2,$$

因为 $\mathbb{S}_i$ 中的每个元素的绝对值都不大于 $\mathbb{S}_{i-1}$中各元素幅值的平均值。利用这一结果并应用向量范数的标准界（习题 1.26），可得

$$\sum_{i\geq 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 \leq \sqrt{k}\sum_{i\geq 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_\infty \leq \frac{1}{\sqrt{k}}\sum_{i\geq 1}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_1 = \frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}.$$

证明到此完成。 ■

以下结果由 Emmanuel Candes 提出，为支撑集大小至多为 $2k$ 的向量范数提供了一个关键估计值。它在建立 RIP 与稀疏恢复保证之间的联系中发挥着核心作用。

引理 4.53：（Candes RIP ℓ_2-ℓ_1 上界引理（Candes，2008年） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$满足阶 $2k$ 的 RIP（其中 $2k\leq p$），并设 $\boldsymbol{\beta}\in\mathbb{R}^p$，$\boldsymbol{\beta}\neq\mathbf{0}$为任意向量。设$\mathbb{S}_0\subseteq\{1,2,\ldots,p\}$ 满足 $|\mathbb{S}_0|\leq k$。定义 $\mathbb{S}_1$ 为 k 中大小（按绝对值）最大的 $\boldsymbol{\beta}_{\mathbb{S}_0^c}$个元素对应的索引，并令 $\mathbb{S}\triangleq\mathbb{S}_0\cup\mathbb{S}_1$。则

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \leq \mu\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}} + \nu\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \boldsymbol{X}\boldsymbol{\beta}\rangle|}{\|\boldsymbol{\beta}_{\mathbb{S}}\|_2}, \tag{4.65}$$

其中 $\mu=\sqrt{2}\delta_{2k}/(1-\delta_{2k})$ 且 $\nu=1/(1-\delta_{2k})$。

证明 [引理 4.53] 由于 $\boldsymbol{\beta}_{\mathbb{S}}\in\mathbb{B}_0[2k]$，限制等距性质的下界给出

$$(1-\delta_{2k})\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \leq \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 = \langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \boldsymbol{X}\boldsymbol{\beta}\rangle - \left\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \sum_{i\geq 2}\boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle, \tag{4.66}$$

其中，集合 $\mathbb{S}_i$ 对于 $i\geq 2$的定义与引理 4.52中的相同，即 $\mathbb{S}_1$ 是对应于 k 绝对值最大的元素的指标集，定义 $\mathbb{S}_2$ 为对应于接下来 k 个绝对值最大元素的指标集，依此类推。上述等式成立是因为$\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}=\boldsymbol{X}\boldsymbol{\beta}-\sum_{i\geq 2}\boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}$。现在我们对第二项进行上界估计。利用内积的线性性，

$$\begin{aligned}
\left|\left\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \sum_{i\geq 2}\boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle\right| &= \left|\sum_{i\geq 2}\left\langle \boldsymbol{X}_{\mathbb{S}_0}\boldsymbol{\beta}_{\mathbb{S}_0}, \boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle + \sum_{i\geq 2}\left\langle \boldsymbol{X}_{\mathbb{S}_1}\boldsymbol{\beta}_{\mathbb{S}_1}, \boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle\right| \\
&\leq \sum_{i\geq 2}\left|\left\langle \boldsymbol{X}_{\mathbb{S}_0}\boldsymbol{\beta}_{\mathbb{S}_0}, \boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle\right| + \sum_{i\geq 2}\left|\left\langle \boldsymbol{X}_{\mathbb{S}_1}\boldsymbol{\beta}_{\mathbb{S}_1}, \boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle\right| \\
&\overset{\dagger}{\leq} \delta_{2k}\sum_{i\geq 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_2\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 + \delta_{2k}\sum_{i\geq 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_2\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 \\
&\overset{\ddagger}{\leq} \sqrt{2}\delta_{2k}\|\boldsymbol{\beta}_{\mathbb{S}}\|_2\sum_{i\geq 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 \leq \sqrt{2}\delta_{2k}\|\boldsymbol{\beta}_{\mathbb{S}}\|_2\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}.
\end{aligned}$$

其中，不等式（†）由 RIP 的标准推论（命题 4.36）得到，不等式（‡）由以下事实得到：$\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_2+\left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_2\leq\sqrt{2}\|\boldsymbol{\beta}_{\mathbb{S}}\|_2$ 根据引理 4.51，

最后一个不等式由引理 4.52得到。将该上界代回可得

$$(1-\delta_{2k})\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2^2 \le |\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \boldsymbol{X}\boldsymbol{\beta}\rangle| + \left|\left\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \sum_{i\ge 2}\boldsymbol{X}_{\mathbb{S}_i}\boldsymbol{\beta}_{\mathbb{S}_i}\right\rangle\right|$$
$$\le |\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}, \boldsymbol{X}\boldsymbol{\beta}\rangle| + \sqrt{2}\delta_{2k}\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2\,\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}.$$

整理后即可得到所需结果。 ■

本引理的一个改进版本——专门用于给 $\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|_2$ 设定 $\|\boldsymbol{\alpha}\|_1 \le \|\boldsymbol{\beta}\|_1$上界——见引理5.16 ，并将用于建立无噪声信号恢复保证。

现在我们证明，若一个矩阵满足 RIP，那么它也满足零空间性质′ （定义 4.4）。

定理 4.54：（RIP $\Longrightarrow$零空间性质′） 设 $\boldsymbol{X}$ 满足阶为 $2k$ 的 RIP，且$\delta_{2k} < \sqrt{2}-1$。则 $\boldsymbol{X}$ 满足阶为 $2k$ 、常数为′ 的零空间性质

$$C = \frac{\sqrt{2}\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}.$$

定理 [的证明 4.54] 回顾一下，引理 4.53 对任意向量 $\boldsymbol{\beta}\in\mathbb{R}^p$均成立。为证明该定理，我们将此引理应用于如下特殊情形： $\boldsymbol{\beta}\in\mathcal{N}(\boldsymbol{X})$ ，即 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{0}$。令 $\mathbb{S}\subseteq\{1,2,\dots,p\}$ 表示对应于 $2k$ 中最大（按绝对值计）的若干个元素的指标集 $\boldsymbol{\beta}$。只需证明

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \le C\frac{\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1}{\sqrt{k}}. \tag{4.67}$$

为此，令 $\mathbb{S}_0$ 为 $\mathbb{S}$ 中包含 k 个最大元素的子集 $\boldsymbol{\beta}$，并定义$\mathbb{S}_1=\mathbb{S}\setminus\mathbb{S}_0$ （使得 $|\mathbb{S}_1|\le k$）。应用引理4.53 并注意到 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{0}$，引理中的第二项为零，因此得到

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \le \mu\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}, \tag{4.68}$$

其中$\mu=\sqrt{2\delta_{2k}}/(1-\delta_{2k})$。利用向量范数的标准界（习题1.26），可得

$$\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1 = \left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_1 + \left\|\boldsymbol{\beta}_{\mathbb{S}^c}\right\|_1 \le \sqrt{k}\left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_2 + \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1. \tag{4.69}$$

结合（4.68）和（4.69）可得

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \le \mu\left(\left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_2 + \frac{\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1}{\sqrt{k}}\right) \quad\Longrightarrow\quad (1-\mu)\,\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \le \mu\frac{\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1}{\sqrt{k}},$$

其中，该推论成立是因为 $\left\|\boldsymbol{\beta}_{\mathbb{S}_1}\right\|_2 \le \|\boldsymbol{\beta}_{\mathbb{S}}\|_2$。假设 $\delta_{2k} < \sqrt{2}-1$ 确保 $\mu<1$，因此我们可以除以 $1-\mu$ 而不改变不等式的方向，从而得到常数为 $C \triangleq \frac{\mu}{1-\mu} = \frac{\sqrt{2\delta_{2k}}}{1-\sqrt{2}\delta_{2k}}$ 的（4.67）。证明完毕。 ■

接下来的论证将证明 RIP 蕴含稳健 NSP（见定义 4.9）。证明依赖于以下辅助结果，即所谓的平方根提升不等式。它可以视为标准向量范数不等式的一个精细化对应结果 $\|\boldsymbol{\beta}\|_1 \le \sqrt{p}\,\|\boldsymbol{\beta}\|_2$ （见习题 1.26）。

引理 4.55：（平方根提升不等式）设 $\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_p \geq 0$。则

$$\sqrt{\beta_1^2 + \beta_2^2 + \ldots + \beta_p^2} \leq \frac{\beta_1 + \beta_2 + \ldots + \beta_p}{\sqrt{p}} + \frac{\sqrt{p}}{4}(\beta_1 - \beta_p).$$

也就是说，$\|\boldsymbol{\beta}\|_2 \leq \frac{\|\boldsymbol{\beta}\|_1}{\sqrt{p}} + \frac{\sqrt{p}}{4}(\beta_1 - \beta_p)$。

证明 [引理455] 只需证明以下等价的蕴含关系：

$$\left\{ \begin{array}{l} \beta_1 \geq \beta_2 \geq \ldots \geq \beta_p \geq 0 \\ \frac{\beta_1+\beta_2+\ldots+\beta_p}{\sqrt{p}} + \frac{\sqrt{p}}{4}\beta_1 \leq 1 \end{array} \right\} \Longrightarrow \sqrt{\beta_1^2 + \beta_2^2 + \ldots + \beta_p^2} + \frac{\sqrt{p}}{4}\beta_p \leq 1.$$

定义凸函数

$$f(\beta_1, \beta_2, \ldots, \beta_p) \triangleq \sqrt{\beta_1^2 + \beta_2^2 + \ldots + \beta_p^2} + \frac{\sqrt{p}}{4}\beta_p$$

并考虑凸多胞体

$$\mathbb{T} \triangleq \left\{ [\beta_1, \ldots, \beta_p] \in \mathbb{R}^p \mid \beta_1 \geq \ldots \geq \beta_p \geq 0 \text{ and } \frac{\beta_1 + \beta_2 + \ldots + \beta_p}{\sqrt{p}} + \frac{\sqrt{p}}{4}\beta_1 \leq 1 \right\}.$$

由于 f 是凸的，且 $\mathbb{T}$ 是紧凸集，因此 f 在 $\mathbb{T}$ 上的最大值在 $\mathbb{T}$的某个极点（顶点）处取得。顶点由将定义 $p+1$ 的 p 个不等式设为等式而得到。我们考察相关情形：

- 若 $\beta_1 = \ldots = \beta_p = 0$，则 $f(\beta_1, \beta_2, \ldots\ldots, \beta_p) = 0$。若 $(\beta_1 + \beta_2 + \ldots + \beta_p)/\sqrt{p} + \sqrt{p}\beta_1/4 = 1$ 且对于某个 $1 \leq k \leq n-1$有 $\beta_1 = \ldots = \beta_k > \beta_{k+1} = \ldots = \beta_p = 0$，则有 $\beta_1 = \ldots = \beta_k = 4\sqrt{p}/(4k+p)$，因此$f(\beta_1, \ldots\ldots, \beta_p) = 4\sqrt{p}/(4k+p) \leq 1$。若 $(\beta_1 + \beta_2 + \ldots + \beta_p)/\sqrt{p} + \sqrt{p}\beta_1/4 = 1$ 且 $\beta_1 = \ldots = \beta_p > 0$，则有 $\beta_1 = \ldots = \beta_p = 4/(5\sqrt{p})$，因此 $f(\beta_1, \ldots\ldots, \beta_p) = 4/5 + 1/5 = 1$。

因此，$\max_{[\beta_1,\ldots,\beta_p]\in\mathbb{T}} f(\beta_1, \beta_2, \ldots\ldots, \beta_p) = 1$，引理得证。 ■

定理 4.56：（RIP $\Longrightarrow$ 稳健 NSP）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 $2k$ 阶的 RIP 条件，其 RIC

$$\delta_{2k} < \frac{77 - \sqrt{1337}}{82} \approx 0.4931. \tag{4.70}$$

那么，$\boldsymbol{X}$ 满足 ℓ_2-稳健 NSP，阶为 k（相对于 $\|\cdot\| = \|\cdot\|_2$），其常数 $0 < \rho < 1$和 $\tau > 0$ 仅依赖于 δ_{2k}。

定理 [的证明 4.56] 我们需要找到常数 $0 < \rho < 1$ 和 $\tau > 0$，使得对于任意$\boldsymbol{\beta} \in \mathbb{R}^p$ 以及任意 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 满足 $|\mathbb{S}| = k$,

$$\|\boldsymbol{\beta}_{\mathbb{S}}\|_2 \leq \frac{\rho}{\sqrt{k}} \|\boldsymbol{\beta}\|_1 + \tau \|\boldsymbol{X}\boldsymbol{\beta}\|_2 . \tag{4.71}$$

给定 $\boldsymbol{\beta} \in \mathbb{R}^p$，只需考虑由 $\boldsymbol{\beta}$ 中模值最大的 k 个元素组成的指标集 $\mathbb{S} \triangleq \mathbb{S}_0$。与之前一样，我们将 $\mathbb{S}^c$ 划分为互不相交的块 $\mathbb{S}_1$，$\mathbb{S}_2$，..... ，每个块至多包含

k 个元素，使得各块内元素的大小按块的顺序非递增。我们先对向量 $\boldsymbol{\beta}(\mathbb{S}_0)+\boldsymbol{\beta}(\mathbb{S}_1)$ 应用限制等距性质，该向量是 $2k$-稀疏的：

$$(1-\delta_{2k})(\|\boldsymbol{\beta}(\mathbb{S}_0)+\boldsymbol{\beta}(\mathbb{S}_1)\|_2^2) \le \|\boldsymbol{X}(\boldsymbol{\beta}(\mathbb{S}_0)+\boldsymbol{\beta}(\mathbb{S}_1))\|_2^2 = \Big\|\boldsymbol{X}\boldsymbol{\beta}-\sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\Big\|_2^2$$
$$= \Big\|\sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\Big\|_2^2 - 2\left\langle \boldsymbol{X}\boldsymbol{\beta}, \sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\right\rangle + \|\boldsymbol{X}\boldsymbol{\beta}\|_2^2. \quad (4.72)$$

用 μ 表示最后两项之和。对于第一项，展开平方范数：

$$\begin{aligned}\left\|\sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\right\|_2^2 &= \left\langle \sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i), \sum_{j\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_j)\right\rangle \\ &= \sum_{i\ge 2}\|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\|_2^2 + \sum_{i,j\ge 2, i\ne j}\langle \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i), \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_j)\rangle \\ &\le (1+\delta_{2k})\sum_{i\ge 2}\|\boldsymbol{\beta}(\mathbb{S}_i)\|_2^2 + \sum_{i,j\ge 2, i\ne j}\delta_{2k}\|\boldsymbol{\beta}(\mathbb{S}_i)\|_2\|\boldsymbol{\beta}(\mathbb{S}_j)\|_2 \\ &= \sum_{i\ge 2}\|\boldsymbol{\beta}(\mathbb{S}_i)\|_2^2 + \delta_{2k}\left(\sum_{i\ge 2}\|\boldsymbol{\beta}(\mathbb{S}_i)\|_2\right)^2, \end{aligned} \quad (4.73)$$

其中，不等式是由 RIP 条件以及 RIP 的标准推论（命题 4.36）得出的。

对于每个块 $\mathbb{S}_i$，令 β_i^- 和 β_i^+ 分别表示 $\mathbb{S}_i$ 上 $\boldsymbol{\beta}$ 的元素的最小绝对值和最大绝对值。定义 $\eta \triangleq \sum_{i\ge 2}\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_1 = \|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1 - \|\boldsymbol{\beta}_{\mathbb{S}_1}\|_1$。那么我们可以得到上界：

$$\sum_{i\ge 2}\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2^2 = \sum_{i\ge 2}\sum_{j\in\mathbb{S}_i}|\beta_j|^2 \le \sum_{i\ge 2}\beta_2^+\sum_{j\in\mathbb{S}_i}|\beta_j| = \beta_2^+\eta.$$

其次，根据引理4.55以及块的单调性（即 $\beta_i^- \ge \beta_{i+1}^+$），可得

$$\sum_{i\ge 2}\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_2 \le \sum_{i\ge 2}\left(\frac{\|\boldsymbol{\beta}_{\mathbb{S}_i}\|_1}{\sqrt{k}} + \frac{\sqrt{k}}{4}(\beta_i^+ - \beta_i^-)\right) \le \frac{\eta}{\sqrt{k}} + \frac{\sqrt{k}}{4}\beta_2^+.$$

将上述两个不等式代入（4.73），得到

$$\left\|\sum_{i\ge 2}\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\right\|_2^2 \le \beta_2^+\eta + \delta_{2k}\left(\frac{\eta}{\sqrt{k}} + \frac{\sqrt{k}}{4}\beta_2^+\right)^2.$$

将上述不等式代入（4.72），考虑到 $\|\boldsymbol{\beta}_{\mathbb{S}_0}+\boldsymbol{\beta}_{\mathbb{S}_1}\|_2^2 = \|\boldsymbol{\beta}_{\mathbb{S}_0}\|_2^2 + \|\boldsymbol{\beta}_{\mathbb{S}_1}\|_2^2 \ge \|\boldsymbol{\beta}_{\mathbb{S}_0}\|_2^2 + k\beta_2^{+2}$，并利用不等式 $\eta = \|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1 - \|\boldsymbol{\beta}_{\mathbb{S}_1}\|_1 \le \|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1 - k\beta_2^+ = (1-\nu)\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1$，其中 $\nu \triangleq k\beta_2^+/\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1$ 可得

$$\begin{aligned}(1-\delta_{2k})\|\boldsymbol{\beta}_{\mathbb{S}_0}\|_2^2 &\le -(1-\delta_{2k})k\beta_2^{+2} + \beta_2^+\eta + \delta_{2k}\left(\frac{\eta}{\sqrt{k}} + \frac{\sqrt{k}}{4}\beta_2^+\right)^2 + \mu \\ &\le -(1-\delta_{2k})\frac{\nu^2}{k}\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1^2 + \frac{\nu(1-\nu)}{k}\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1^2 + \delta_{2k}\left(\frac{1-\nu}{\sqrt{k}} + \frac{\nu}{4\sqrt{k}}\right)^2\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1^2 + \mu \quad (4.74) \\ &= \left[-(1-\delta_{2k})\nu^2 + \nu(1-\nu) + \delta_{2k}\left(1-\frac{3\nu}{4}\right)^2\right]\frac{\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\|_1^2}{k} + \mu. \end{aligned}$$

我们需要证明方括号中的量除以（$1-\delta_{2k}$）小于 1，或者等价地，证明该量减去（$1-\delta_{2k}$）后为负。这个差是关于 ν 的二次表达式，表示为

$$f(\nu) \triangleq -\left(2-\frac{25}{16}\delta_{2k}\right)\nu^2+\left(1-\frac{3}{2}\delta_{2k}\right)\nu-1+2\delta_{2k} \triangleq -a\nu^2+b\nu-c.$$

该表达式在 $\nu^* \triangleq b/(2a)$ 处取得最大值，因此

$$f(\nu) \le f(\nu^*) = \frac{b^2}{4a}-c = \frac{(1-3\delta_{2k}/2)^2}{4(2-25\delta_{2k}/16)}-1+2\delta_{2k}.$$

容易证明 $f(\nu^*)<0$ 当且仅当 $41\delta_{2k}^2-77\delta_{2k}+28>0$。因此 $f(\nu)<0$ 只要 δ_{2k} 小于 $41x^2-77x+28$ 的最小根，就成立。这正是条件（4.70）。在此条件下，（4.74）中方括号内的量至多为 $\rho^2(1-\delta_{2k})$，其中常数 $0<\rho<1$ 仅取决于 δ_{2k}。为证明所需的不等式（4.71），由（4.74），可得

$$\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_2^2 \le \frac{\rho^2\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1^2}{k}+\frac{\mu}{1-\delta_{2k}}. \tag{4.75}$$

对于 μ 的估计，即（4.72）右侧最后两项之和，利用柯西–施瓦茨不等式和三角不等式，可得

$$\mu \le 2\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2\sum_{i\ge 2}\left\|\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S}_i)\right\|_2+\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2^2 \le 2\sqrt{1+\delta_{2k}}\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2\sum_{i\ge 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2+\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2^2.$$

由引理 4.47，可得 $\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 \le \left\|\boldsymbol{\beta}_{\mathbb{S}_{i-1}}\right\|_1/\sqrt{k}$，因此有

$$\sum_{i\ge 2}\left\|\boldsymbol{\beta}_{\mathbb{S}_i}\right\|_2 \le \frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}.$$

将上述两个不等式代入（4.75），可得

$$\begin{aligned}\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_2^2 &\le \frac{\rho^2\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1^2}{k}+2\frac{\sqrt{1+\delta_{2k}}}{1-\delta_{2k}}\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}+\frac{\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2^2}{1-\delta_{2k}}\\ &\le \left(\frac{\rho^2\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1^2}{k}+2\frac{\sqrt{1+\delta_{2k}}}{1-\delta_{2k}}\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2\frac{\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}+\frac{1+\delta_{2k}}{\rho^2(1-\delta_{2k})^2}\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2^2\right)\\ &= \left(\frac{\rho\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}}+\frac{\sqrt{1+\delta_{2k}}}{\rho(1-\delta_{2k})}\left\|\boldsymbol{X}\boldsymbol{\beta}\right\|_2\right)^2.\end{aligned}$$

令 $\tau \triangleq \sqrt{1+\delta_{2k}}/(\rho(1-\delta_{2k}))$，即可完成证明。 ■

第4章习题

1. 证明集合 $\mathbb{C}[k] \triangleq \bigcup_{\mathbb{S}:\ |\mathbb{S}|=k} \mathbb{C}[\mathbb{S}]$ 是非凸的。

2. ℓ_s-d **稀疏恢复的距离**。对于任意 $s > t > 0$ 以及任意 $\boldsymbol{\beta} \in \mathbb{R}^p$，证明

$$\sigma_k(\boldsymbol{\beta})_s \leq \frac{c_{t,s}}{k^{(\frac{1}{t}-\frac{1}{s})}} \|\boldsymbol{\beta}\|_t ,$$

其中 $c_{t,\ s} \triangleq \left[\binom{t}{s}^{\frac{t}{s}} \left(1 - \frac{t}{s}\right)^{(1-\frac{t}{s})} \right]^{\frac{1}{t}} \leq 1.$ 在 $t = 1$ 和 $s = 2$ 的特殊情况下，不等式变为

$$\sigma_k(\boldsymbol{\beta})_2 \leq \frac{1}{2\sqrt{k}} \|\boldsymbol{\beta}\|_1 .$$

3. **正交对的互相干性**。定义两个正交基 $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p] \in \mathbb{R}^p$ 和 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p] \in \mathbb{R}^p$ 之间的互相干性为

$$\mu(\boldsymbol{U}, \boldsymbol{V}) \triangleq \max_{1 \leq i,j \leq p} |\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle| .$$

证明

$$1/\sqrt{p} \leq \mu(\boldsymbol{U}, \boldsymbol{V}) \leq 1.$$

4. 验证定义4.26中紧框架系统的三个陈述之间的等价性：**(i)** $\Longrightarrow$(ii) $\Longrightarrow$ (iii) $\Longrightarrow$ **(i)**。

5. **等角系统的上界**。证明，ℓ_2归一化向量 $\mathbb{R}^n$ 中的等角系统 $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$ 的基数 p满足

$$p \leq \frac{n(n+1)}{2}.$$

此外，若等号成立，证明系统 $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$ 构成 $\mathbb{R}^n$ 的紧框架。

6. 证明 (4.37a) 与 (4.37b) 等价。
7. 证明 (4.42b) 与 (4.42c) 等价。
8. **RIP 的标准推论**。对于命题4.36的一个具体情形，设 $\boldsymbol{X}$满足阶 $2k$ 、常数为 δ_{2k}的 RIP 条件，并设 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 是支撑集互不相交且支撑集大小为 $\leq k$ 的向量（因此 RIP 适用于$\boldsymbol{\alpha} \pm \boldsymbol{\beta}$）。直接证明

$$|\langle \boldsymbol{X}\boldsymbol{\alpha}, \boldsymbol{X}\boldsymbol{\beta} \rangle| \leq \delta_{2k} \|\boldsymbol{\alpha}\|_2 \|\boldsymbol{\beta}\|_2 .$$

提示：不要使用命题 *4.36*中的结果。展开 $\left\| \boldsymbol{X}(\widetilde{\boldsymbol{\alpha}} \pm \widetilde{\boldsymbol{\beta}}) \right\|_2^2$ 并相减，~其中 $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_2$ 且 $\boldsymbol{\beta} = \boldsymbol{\beta}/\|\boldsymbol{\beta}\|_2$；然后缩放回去。

9. **向量范数**。设 $\boldsymbol{\beta} \in \mathbb{R}^p$。证明

$$\|\boldsymbol{\beta}\|_t \leq p^{(\frac{1}{t}-\frac{1}{s})} \|\boldsymbol{\beta}\|_s , \quad \text{for } 1 \leq t \leq s.$$

10. 证明存在一个标量 $\nu^* \in [0,1]$ ，使得定义于（4.47）中的函数在 $[0, \nu^*]$ 上是非递减的，并且在 $[\nu^*, 1]$上是非递增的。
11. **RIP 的性质**。设 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 满足阶为 k 、常数为 δ_k的限制等距性质。证明

$$\|\boldsymbol{X}\boldsymbol{\beta}\|_2 \leq \sqrt{1+\delta_k} \left(\|\boldsymbol{\beta}\|_2 + \frac{\|\boldsymbol{\beta}\|_1}{\sqrt{k}} \right), \quad \text{for any } \boldsymbol{\beta} \in \mathbb{R}^p.$$

12. **RIP 的性质**。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 k 、常数为 $\delta_k < 1$的限制等距性质，且设 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 满足 $|\mathbb{S}| \leq k$。证明

$$\left\|(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})^{-1}\right\|_2 \leq \frac{1}{1-\delta_k}.$$

还要证明

$$\left\|\boldsymbol{X}_{\mathbb{S}}(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})^{-1}\right\|_2 \leq \sqrt{1+\delta_k}/(1-\delta_k).$$

提示：半正定矩阵或正定矩阵的谱范数等于该矩阵的最大特征值。接着考察矩阵逆矩阵的特征值。利用这一事实： $\boldsymbol{X}$ 的奇异值平方是 $\boldsymbol{X}^{\top}\boldsymbol{X}$的特征值。

13. **子集强凸性（SSC）**。如果对所有满足 $|\mathbb{S}| \leq k$的指标集$\mathbb{S} \subset \{1, 2, \ldots, n\}$，下列不等式对所有 $\boldsymbol{\beta} \in \mathbb{R}^p$均成立，则称矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 k阶的 μ_k-子集强凸性（SSC）性质和 ν_k-子集光滑性（SSS）性质：

$$\mu_k \|\boldsymbol{\beta}\|_2^2 \leq \left\|\boldsymbol{X}^{\mathbb{S}}\boldsymbol{\beta}\right\|_2^2 \leq \nu_k \|\boldsymbol{\beta}\|_2^2,$$

其中， $\boldsymbol{X}^{\mathbb{S}} \in \mathbb{R}^{|\mathbb{S}|\times p}$ 表示通过选取由 $\mathbb{S}$索引的行构成的 $\boldsymbol{X}$ 子矩阵。证明 SSC 性质意味着不存在满足 $|\mathbb{S}| \leq k$ 的子集 $\mathbb{S}$ ，也不存在满足 $\boldsymbol{X}^{\mathbb{S}}\boldsymbol{\beta}_1 = \boldsymbol{X}^{\mathbb{S}}\boldsymbol{\beta}_2$的两个不同向量 $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2 \in \mathbb{R}^p$ 。

第5章

稀疏优化分析

目录

5.1 NSP 下的稀疏恢复.222 5.1.1 无噪声信号恢复.222 5.1.2 含噪信号恢复.224 **5.2 相干性下的稀疏恢复.**226 **5.3 RIP 下的稀疏恢复.**227 5.3.1 对偶证书：再探.227 5.3.2 无噪声信号恢复.232 5.3.3 实例最优保证.234 5.3.4 含噪信号恢复.235 5.3.5 使用 Dantzig 选择器的信号恢复.238

5.4 稀疏线性回归.239

5.4.1 LASSO与稳健线性回归.241 5.4.2 LASSO误差界.244 5.4.3 高斯噪声与预言机不等式.247 **第5章习题.**249



I在许多现实应用中——从信号处理、成像到机器学习和基因组学——我们遇到的数据往往具有稀疏性：也就是说，只有少数分量承载着有意义的信息。利用这一结构性质使我们能够从有限或含噪的观测中高精度地恢复信号或估计模型，同时具备较高的计算效率。本章深入探讨稀疏优化的理论基础，重点关注能够从线性测量中准确恢复稀疏信号的条件。

我们从第 5.1 节开始，考察基于零空间性质（NSP）的恢复保证。零空间性质是一项基本条件，可确保通过凸优化精确恢复稀疏信号。我们分析无噪声和有噪声两种环境，并探讨稳定性与鲁棒性的理论见解。

第 5.2 节研究基于相干性的恢复方法。相干性用于衡量测量矩阵各列之间的相似程度。这种方法有助于直观理解，并为设计具有良好恢复性质的感知矩阵提供实用工具。接下来，第 5.3节聚焦于受限等距性质（RIP），该性质为稀疏恢复提供了更强大、适用范围更广的保证。我们介绍无噪声和有噪声情形下的结果，包括实例最优误差界，并分析 Dantzig 选择器——一种替代 ℓ_1-最小化的方法。

稀疏优化的理论基础 ℓ_1-最小化自然引出了第5.4节对 LASSO（最小绝对收缩与选择算子）的讨论。LASSO 是统计学和机器学习中广泛用于带稀疏诱导正则化的回归问题的技术。我们将 LASSO 与稳健线性回归方法进行比较，并考察其在不同条件下的行为。

为了说明设计矩阵的性质与稀疏恢复之间的关联，考虑以下结果：

定理 5.1：（RIP 条件下 ℓ_0 的精确恢复）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个满足 $2k$ 阶 RIP 条件的设计矩阵，且 $k \geq 1$ 并具有受限等距常数$\delta_{2k} < 1$。设 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 满足 $|\mathbb{S}| \leq k$。令 $\boldsymbol{y} \triangleq \boldsymbol{X}\boldsymbol{\beta}^*$，其中 $\boldsymbol{\beta}^*$ 是任意一个支撑于 $\mathbb{S}$ 的向量（即 $\mathbb{S} =\text{supp}(\boldsymbol{\beta}^*)$）。那么 $\boldsymbol{\beta}^*$ 是问题（$\mathbf{P}_0$）（第129页）的唯一极小化解：

$$(\mathrm{P}_0) \quad \min \|\boldsymbol{\beta}\|_0 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}.$$

定理 5.1] 的证明。 $\boldsymbol{y}$ 存在两个不同的 k-稀疏表示： $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* = \boldsymbol{X}\boldsymbol{\beta}'$，其中 $\boldsymbol{\beta}^*$ 和 $\boldsymbol{\beta}'$ 分别支撑于集合 $\mathbb{S}$ 和 $\mathbb{S}'$，且 $|\mathbb{S}|$、$|\mathbb{S}'| \leq k$。那么 $\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}') = \mathbf{0}$. 根据构造，差值 $\boldsymbol{\beta}^* - \boldsymbol{\beta}'$ 支撑于 $\mathbb{S} \cup \mathbb{S}'$，其基数至多为 $2k$。应用 RIP 条件（见方程 (4.36a)）并利用 $\delta_{2k} < 1$，得到 $\left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}'\right\|_2^2 = 0$，这与两个表示互不相同的假设矛盾。

■

请注意，这个定理是一个抽象存在性结果：它确立了理论上的可能性，但没有提供一种高效算法来恢复 $\mathbb{S}$ 或系数 β_i^*，使其由 $\boldsymbol{y}$和 $\{\boldsymbol{x}_i\}_{i\in\{1,2,\ldots,p\}}$得到，穷举（蛮力）搜索除外——如第 3章所述。相比之下，正如我们接下来将展示的，施加稍强的条件——例如对限制等距常数 δ_{2k} 施加更严格的上界，或要求满足其他结构性质——即可确保凸 ℓ_1最小化程序（P_1）能够精确恢复 $\boldsymbol{\beta}^*$。

例如，Candes 和 Tao (2005年) 证明，若 $\boldsymbol{X} \in \text{RIP}(2k,\ 0。33)$（定义 4.30），则非凸 ℓ_0-范数最小化问题（P_0）与凸

ℓ_1-最小化问题（P_1）（第 129 页）对所有 k-稀疏向量具有相同的解。此外，在存在噪声（$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$）时，$\ell_1$-最小化能够稳定地逼近最稀疏的近似解，且稳定性常数受到良好控制（见第 5.3 节）。随后，Candes (2008年) 给出了更精确的充分条件：$\delta_{2k}<0.4142$。在此上界下，ℓ_1-最小化能够在无噪声情形下精确恢复任意 k-稀疏信号，并在存在噪声时提供稳健的估计值。本章将讨论与此条件相关的更多结果，并将分析进一步扩展到 LASSO 类问题。

5.1. NSP 下的稀疏恢复

在习题 4.2中，我们证明了，阶数为 $2k$ 的 NSP 足以唯一恢复不同的 k-稀疏信号；当恢复算法满足引理 4.5 中式（4.9）所述的特定上界时，NSP′ 则是必要条件。本节将探讨在无噪声情形下，NSP 成为稀疏信号唯一恢复的充分且必要条件时所需满足的条件。我们还将证明，在噪声环境中，相应的误差界仍然保持稳定。

5.1.1 无噪声信号恢复

我们首先阐明 NSP（定义 4.1）与通过基追踪精确恢复稀疏向量（即 ℓ_1-最小化）之间的联系。

定理 5.2：（NSP 下 ℓ_1 的精确恢复（$\mathbb{S}$）） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$。那么，每个 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 支撑于集合 $\mathbb{S}$ 的向量都是（**唯一** 解，当且仅当P_1）满足 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$，当且仅当 $\boldsymbol{X}$ 满足相对于 $\mathbb{S}$的 NSP 条件，即 NSP（$\mathbb{S}$）。

定理 5.2] 的证明 [固定一个指标集 $\mathbb{S}$，并假设每个支撑于 $\mathbb{S}$ 的向量 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 都是在约束 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\alpha}$ 下的 $\|\boldsymbol{\alpha}\|_1$ 的唯一极小化解，其中 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*=\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}^*_{\mathbb{S}}$。考虑任意非零 $\boldsymbol{\beta}\in\mathcal{N}(\boldsymbol{X})\setminus\{\boldsymbol{0}\}$。则 $\boldsymbol{\beta}(\mathbb{S})$ 满足 $\boldsymbol{X}(-\boldsymbol{\beta}(\mathbb{S}^c))=\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})$，[1] 由于 $\boldsymbol{\beta}\neq\boldsymbol{0}$，我们有 $-\boldsymbol{\beta}(\mathbb{S}^c)\neq\boldsymbol{\beta}(\mathbb{S})$。根据所假设的 ℓ_1 极小化解的唯一性，可得 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1<\|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1$，这正是 NSP($\mathbb{S}$) 条件。

反之，假设 NSP($\mathbb{S}$) 条件成立。设 $\boldsymbol{\beta}\in\mathbb{R}^p$ 支撑于 $\mathbb{S}$，且令 $\boldsymbol{\alpha}\in\mathbb{R}^p$、$\boldsymbol{\alpha}\neq\boldsymbol{\beta}$ 满足 $\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}$。定义 $\boldsymbol{u}\triangleq\boldsymbol{\beta}-\boldsymbol{\alpha}\in\mathcal{N}(\boldsymbol{X})\setminus\{\boldsymbol{0}\}$。于是，利用三角不等式和 NSP($\mathbb{S}$) 条件，

$$\|\boldsymbol{\beta}\|_1\le\|\boldsymbol{\beta}_{\mathbb{S}}-\boldsymbol{\alpha}_{\mathbb{S}}\|_1+\|\boldsymbol{\alpha}_{\mathbb{S}}\|_1=\|\boldsymbol{u}_{\mathbb{S}}\|_1+\|\boldsymbol{\alpha}_{\mathbb{S}}\|_1<\|\boldsymbol{u}_{\mathbb{S}^c}\|_1+\|\boldsymbol{\alpha}_{\mathbb{S}}\|_1=\|-\boldsymbol{\alpha}_{\mathbb{S}^c}\|_1+\|\boldsymbol{\alpha}_{\mathbb{S}}\|_1=\|\boldsymbol{\alpha}\|_1.$$

因此，$\boldsymbol{\beta}$ 在所有可行解中唯一地使 ℓ_1-范数达到最小，证明完成。
■

令支撑集 $\mathbb{S}$ 遍历所有大小至多为 k 的子集，我们立即得到以下推论。

推论 5.3：（在 ℓ_1 满足阶为 k 的零空间性质时的精确恢复） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$。那么，每个 k-稀疏向量 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 都是 （$\mathbf{P}_1$）在 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$ 下的唯一解，当且仅当 $\boldsymbol{X}$ 满足阶为 k的零空间性质。

该结果表明，只要满足 k 阶零空间性质（NSP），ℓ_1-最小化问题（P_1）就能对所有 ℓ_0-稀疏信号精确求解组合的P_0）最小化问题。k。

1. 再次注意，$\boldsymbol{\beta}$（$\mathbb{S}$）$\in\mathbb{R}^p$ 以及 $\boldsymbol{\beta}_{\mathbb{S}}\in\mathbb{R}^{|\mathbb{S}|}$；参见定义 1.1。

事实上，假设 $\boldsymbol{\beta}^*$ 是 k-稀疏的，并通过 ℓ_1-最小化从 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$中唯一恢复。设 $\boldsymbol{\alpha}$ 是问题（P_0）的一个解；那么 $\|\boldsymbol{\alpha}\|_0 \le \|\boldsymbol{\beta}^*\|_0 \le k$，因此 $\boldsymbol{\alpha}$ 也是 k-稀疏的。由于对所有有 ℓ_1 -稀疏向量， k解都是唯一的，因此必有 $\boldsymbol{\alpha}=\boldsymbol{\beta}^*$。下面的定理将形式化地阐述 NSP 条件下 ℓ_0 与 ℓ_1 问题之间的等价性。

定理5.4：（ℓ_0 以及 ℓ_1 NSP 下的等价性）设 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 是**唯**一的解，满足 ℓ_0 问题（P_0），且支撑集为 $\mathbb{S}$。那么，基追踪问题（P_1）有一个**唯**一的解，且该解等于 $\boldsymbol{\beta}^*$ 当且仅当 $\boldsymbol{X}$ 满足相对于 $\mathbb{S}$的 NSP 条件，即 NSP($\mathbb{S}$）。

[定理 5.4] 的证明：假设 $\boldsymbol{X}$ 满足 NSP($\mathbb{S}$)。令 $\boldsymbol{\beta} \in \mathbb{R}^p$ 为问题（P_1）的唯一解，并定义误差向量 $\Delta \triangleq \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$。我们的目标是证明 $\Delta=\mathbf{0}$，为此，只需证明 $\Delta \in \mathcal{N}(\boldsymbol{X}) \cap \mathbb{C}(\mathbb{S})$。由于 $\boldsymbol{\beta}^*$ 和 $\boldsymbol{\beta}$ 分别是 ℓ_0 和 ℓ_1 问题的最优（因而可行）解，它们给出相同的测量值 $\boldsymbol{X}\boldsymbol{\beta}^*=\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}$，这表明 $\boldsymbol{X}\Delta=\mathbf{0}$，因此 $\Delta \in \mathcal{N}(\boldsymbol{X})$。此外，由于 $\boldsymbol{\beta}^*$ 是 ℓ_1-最小化问题（P_1）的可行解，$\boldsymbol{\beta}$ 的唯一最优性意味着 $\|\widehat{\boldsymbol{\beta}}\|_1 < \|\boldsymbol{\beta}^*\|_1 = \|\boldsymbol{\beta}^*_{\mathbb{S}}\|_1 + \|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1$，由此可得

$$\|\boldsymbol{\beta}^*_{\mathbb{S}}\|_1 > \|\widehat{\boldsymbol{\beta}}\|_1 \ge \|\widehat{\boldsymbol{\beta}}_{\mathbb{S}}\|_1 = \|\boldsymbol{\beta}^*_{\mathbb{S}}+\Delta_{\mathbb{S}}\|_1 + \|\Delta_{\mathbb{S}^c}\|_1 \ge \|\boldsymbol{\beta}^*_{\mathbb{S}}\|_1 - \|\Delta_{\mathbb{S}}\|_1 + \|\Delta_{\mathbb{S}^c}\|_1,$$

等号成立是因为 $\boldsymbol{\beta}=\boldsymbol{\beta}^*+\Delta$ 和 $\boldsymbol{\beta}^*_{\mathbb{S}^c}=\mathbf{0}$，最后一个不等式则由三角不等式得出。重新整理各项可得 $\Delta \in \mathbb{C}(\mathbb{S})$。然而，由于 $\backsim\Delta\mathcal{N}(\boldsymbol{X}) \cap \mathbb{C}(\mathbb{S}) \neq \{\mathbf{0}\}$ 会违反 NSP($\mathbb{S}$)，因此必有 $\Delta=\mathbf{0}$。于是 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$。

反过来，假设问题 (P_1) 有唯一解，且该解等于 $\boldsymbol{\beta}^* \in \mathbb{R}^p$，其支撑集为 $\mathbb{S}$。对于任意非零的 $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$，向量 $\boldsymbol{\beta}(\mathbb{S})$ 是满足 $\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})$ 的 $\|\boldsymbol{\alpha}\|_1$ 的唯一极小化解。由于 $\boldsymbol{X}\boldsymbol{\beta}=\mathbf{0}$ 和 $\boldsymbol{\beta} \neq \mathbf{0}$，我们有 $\boldsymbol{X}(-\boldsymbol{\beta}(\mathbb{S}^c)) = \boldsymbol{X}\boldsymbol{\beta}(\mathbb{S})$ 且 $-\boldsymbol{\beta}(\mathbb{S}^c) \neq \boldsymbol{\beta}(\mathbb{S})$。因此可得 $\|\boldsymbol{\beta}_{\mathbb{S}}\|_1 < \|\boldsymbol{\beta}_{\mathbb{S}^c}\|_1$，从而确立了 NSP($\mathbb{S}$) 条件。证明完毕。 ■

注意，任何重建方法最好也能对测量系统的某些变换保持稳健——例如对 $\boldsymbol{X}$的行进行重新缩放、重新排序或增补。基追踪具有这类不变性，如（4.6）所示。

最后，我们将本节的分析推广到更一般的 ℓ_s-最小化问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \qquad \text{s.t.} \quad \boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}, \tag{P_s}$$

其中 $0<s\le 1$。尽管由于凸性， $s=1$ 的情形在实践中最为重要，但 $0<s<1$ 的非凸问题具有理论意义，因为它们能更好地逼近 ℓ_0 目标函数。直观地说， s 的值越小，所得解就应越接近真实的 ℓ_0极小化解。为使这一点形式化，我们引入 NSP 的 ℓ_s-类比形式。

下一个结果的证明与定理 5.2 的证明相仿，并留作习题。该证明依赖于函数 $\boldsymbol{\beta} \mapsto \|\boldsymbol{\beta}\|_s^s$ 对 $0<s\le 1$满足三角不等式这一事实；参见问题 5.2。

定理 5.5：（ℓ_s的精确恢复）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 且 $0<s\le 1$。那么，每个k-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 都是（P_s）在 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$ 下的唯一解，当且仅当对每个指标集 $\mathbb{S} \subseteq \{1,2,\dots,p\}$ 满足 $|\mathbb{S}| \le k$，

$$\|\boldsymbol{\alpha}_{\mathbb{S}}\|_s < \|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_s, \quad \text{for all } \boldsymbol{\alpha} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}. \tag{5.1}$$

（5.1）中的条件实际上是标准 NSP 条件（4.5b）在 ℓ_s-情形下的对应形式，该标准 NSP 条件见定义 4.1。

现在我们证明，只要 $0 < t < s \leq 1$，通过 ℓ_s-最小化成功进行稀疏恢复，就意味着通过 ℓ_t-最小化也能成功恢复——也就是说，使用较小的指数至少能提供同样强的恢复保证。

定理 5.6：（$\ell_t < \ell_s$的精确恢复）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 且 $0 < t < s \leq 1$。如果每个 k-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 都是方程（$\mathbf{P}_s$）在 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$下的唯一解，那么每个$k$-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 也都是方程（P_t）在 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$下的唯一解。

定理 [5.的证明6] 根据定理 5.5，只需证明以下结论：对于任意非零$\boldsymbol{\alpha} \in \mathcal{N}(\boldsymbol{X})$，$\setminus \{\mathbf{0}\}$如果 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 表示一个由 k 中绝对值最大的元素对应的索引组成的集合 $\boldsymbol{\alpha}$，那么

$$\sum_{j\in\mathbb{S}} |\alpha_j|^t < \sum_{i\in\mathbb{S}^c} |\alpha_i|^t, \tag{5.2}$$

前提是将 t 换成 s 后，同样的不等式仍成立。

注意，由于 $\boldsymbol{\alpha} \neq \mathbf{0}$ 和 $\mathbb{S}$ 包含绝对值最大的 k 个元素，补集 $\boldsymbol{\alpha}_{\mathbb{S}^c}$ 不可能恒为零，除非 $\boldsymbol{\alpha}$ 是 k-稀疏的——但即使在这种情况下，NSP 条件也要求 $\boldsymbol{\alpha} = \mathbf{0}$，这就产生了矛盾。因此，$\boldsymbol{\alpha}_{\mathbb{S}^c} \neq \mathbf{0}$，且下方的分母为正。

将 (5.2) 两边除以 $\sum_{i\in\mathbb{S}^c} |\alpha_i|^t > 0$，所需的不等式等价于

$$\sum_{j\in\mathbb{S}} \frac{|\alpha_j|^t}{\sum_{i\in\mathbb{S}^c} |\alpha_i|^t} = \sum_{j\in\mathbb{S}} \frac{1}{\sum_{i\in\mathbb{S}^c} (|\alpha_i| / |\alpha_j|)^t} < 1. \tag{5.3}$$

现在注意，对每个 $i \in \mathbb{S}^c$和 $j \in \mathbb{S}$，比值 $|\alpha_i| / |\alpha_j| \leq 1$由 $\mathbb{S}$的构造可知。这使得 (5.3) 的左侧成为关于 $0 < t \leq 1$的非递减函数。因此，它在 $t < s$ 处的值不超过其在 s处的值。但根据假设和定理 5.5，当 s时，不等式严格成立，即该表达式在 < 1 时为 $t = s$。因此，当 < 1 时，它也为 $t < s$，从而得证 (5.2)。证明完毕。 ■

5.1.2 含噪信号恢复

能够从无噪声测量中完美重构稀疏信号，是一项前景可观的理论成果。然而，在大多数现实应用中，测量值不可避免地会受到某种形式的噪声污染。例如，在数字计算机上处理数据时，我们必须用有限数量的比特表示信号，这会引入量化误差。此外，物理测量系统还会受到各种噪声的影响，例如热噪声、电子噪声或传感器噪声，具体取决于实现方式和环境。

或许有些出人意料的是，我们可以对标准的 ℓ_1-最小化程序进行修改

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \qquad \text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{y}).$$

从而在各种常见的噪声模型下稳定地恢复稀疏（或近似稀疏的）信号，例如模型（$\mathrm{P}_{1,\ \epsilon}$）（第 131 页）使用 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{z} \mid \|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \epsilon\}$。

前面的结果在各种 NSP 条件下建立了基追踪问题的精确恢复。接下来，我们首先利用稳定 NSP（定义 4.6）或稳健 NSP（定义 4.9）证明，任意解都位于真实向量的某个邻域内，而该邻域的大小由最佳 k项逼近误差控制。

定理 5.7：（ℓ_1 **在稳定 NSP 下的稳定性**）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶数为 k、常数为 $0<\rho<1$ 的稳定 $\boldsymbol{\beta}^* \in \mathbb{R}^p$。则，对于任意 $\boldsymbol{\beta}$，方程（P_1）的每个解 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$ 在测量值 ℓ_1下都满足

$$\left\|\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{(1-\rho)}\sigma_k(\boldsymbol{\beta}^*)_1, \tag{5.4}$$

其中，$\sigma_k(\boldsymbol{\beta}^*)_1 \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]}\left\|\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\|_1$ 表示最佳 ℓ_1 项近似的 k 误差。

定理 [的证明 5.7] 令 $\mathbb{S}\subseteq\{1,2,\ldots,p\}$ 为由 k 个对应于大小最大的元素的索引组成的集合 $\boldsymbol{\beta}^*$。根据定义，$\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1=\sigma_k\left(\boldsymbol{\beta}^*\right)_1$。由于 $\boldsymbol{\beta}$ 是（P_1）的极小化解，因此满足 $\|\widehat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}^*\|_1$ 以及 $\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{X}\boldsymbol{\beta}^*$。应用定理 4.7 并令$\boldsymbol{\alpha}=\boldsymbol{\beta}$ 以及 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$，可得

$$\left\|\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{(1-\rho)}\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1 = \frac{2(1+\rho)}{(1-\rho)}\sigma_k(\boldsymbol{\beta}^*)_1,$$

这正是式（5.4）中的上界。 ■

与定理 5.2 和推论 5.3不同的是，仅凭稳健 NSP，我们无法再保证 ℓ_1-极小化解的唯一性——不过，只有在退化情形下才会出现非唯一性。尽管如此，即使存在多个极小化点，定理仍保证每个解 $\boldsymbol{\beta}$ 满足（P_1）且 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$ 的解都满足误差界（5.4）。

下面的定理推广了定理 5.7，并将其作为二次约束 $\epsilon=0$ 最小化问题（$\mathrm{P}_{1,\ \epsilon}$）在特殊情形下的结果恢复出来。

定理 5.8：（（$\mathrm{P}_{1,\ \epsilon}$）在稳健 NSP 下的稳定性）设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 满足阶为 k、常数为 $0<\rho<1$、$\tau>0$ 和 $\|\cdot\|=\|\cdot\|_2$的稳健 NSP。则对于任意$\boldsymbol{\beta}^*\in\mathbb{R}^p$，每个解 $\boldsymbol{\beta}$ 满足（$\mathrm{P}_{1,\ \epsilon}$），且测量值为 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$、噪声满足$\|\boldsymbol{\epsilon}\|_2\leq\epsilon$，都满足 ℓ_1-误差界：

$$\left\|\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{(1-\rho)}\sigma_k(\boldsymbol{\beta}^*)_1+\frac{4\tau}{1-\rho}\epsilon. \tag{5.5}$$

证明 [定理58] 令$\mathbb{S}\subseteq\{1,2,\ldots,p\}$ 为一组 k 对应于中大小最大的 $\boldsymbol{\beta}^*$索引。根据定义，$\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1=\sigma_k\left(\boldsymbol{\beta}^*\right)_1$。由于 $\boldsymbol{\beta}$ 是（$\mathrm{P}_{1,\ \epsilon}$）的极小化解，因此满足 $\|\widehat{\boldsymbol{\beta}}\|_1\leq\|\boldsymbol{\beta}^*\|_1$和 $\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}^*\right\|\leq 2\epsilon$。应用定理4.11（稳健NSP 的标准推论），令 $\boldsymbol{\alpha}=\boldsymbol{\beta}$ 且 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$，可得

$$\left\|\boldsymbol{\beta}^*-\widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{(1-\rho)}\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1+\frac{2\tau}{1-\rho}\left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\right\|_2 \leq \frac{2(1+\rho)}{(1-\rho)}\sigma_k(\boldsymbol{\beta}^*)_1+\frac{4\tau}{1-\rho}\epsilon.$$

这正是式（5.5）中的上界。 ■

注意，当 $\boldsymbol{\beta}^*$ 的支撑集 $= \mathbb{S}$with $|\mathbb{S}| = k$，也就是说， $\boldsymbol{\beta}^*$ 恰好是 k-稀疏的时，上界变为

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{4\tau}{1-\rho}\epsilon.$$

此外，当 $\epsilon = 0$时，该结果可化为推论 5.3。

下面的推论通过不仅在 ℓ_1 中，而且对所有 $1 \leq t \leq 2$ 在 ℓ_t 中提供误差界，进一步强化了上述稳健性保证。

推论 5.9：（(P_1, ϵ) **在** ℓ_2**-稳健 NSP 下的稳定性**）令 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 k 、常数为 $0 < \rho < 1$、 $\tau > 0$ 和 $\|\cdot\| = \|\cdot\|_2$ 的ℓ_2-稳健 NSP。那么，对于任意$\boldsymbol{\beta}^* \in \mathbb{R}^p$，具有测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 和噪声 $\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$ 的（ P_1, ϵ) 的每个解 $\boldsymbol{\beta}$ 都满足 ℓ_t- 误差界：

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_t \leq \frac{C}{k^{(1-\frac{1}{t})}}\sigma_k(\boldsymbol{\beta}^*)_1 + Dk^{(\frac{1}{t}-\frac{1}{2})}\epsilon, \quad 1 \leq t \leq 2, \tag{5.6}$$

其中，常数 C、 $D > 0$ 仅取决于 ρ和 τ（具体而言， $C \triangleq 2(1+\rho)^2/(1-\rho)$ 以及 $D \triangleq 2\tau($ $3+\rho)/(1-\rho)$）。特别地，极端情形下的界为：

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq C\sigma_k(\boldsymbol{\beta}^*)_1 + D\sqrt{k}\epsilon, \qquad t = 1;$$
$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_2 \leq \frac{C}{\sqrt{k}}\sigma_k(\boldsymbol{\beta}^*)_1 + D\epsilon, \qquad t = 2.$$

推论 5.9] 的证明 [误差界 (5.6) 可由定理 4.12（ ℓ_s-稳健 NSP 的标准推论）如下得到：令 $\mathbb{S}$ 为与 $\boldsymbol{\beta}^*$中（按大小）最大的$\widehat{}$个元素对应的 k 索引集合。于是 $\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_1 = \sigma_k(\boldsymbol{\beta}^*)_1$。 (P1,$_\epsilon$) 的解 $\boldsymbol{\beta}$ 满足 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1 \leq \|\boldsymbol{\beta}^*\|_1$ 和 $\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^*\right\|_2 \leq 2\epsilon$。因此，利用定理 4.12，并取 $\boldsymbol{\alpha} = \widehat{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ ，即可证明所需的上界 (5.6)。 ■

我们注意到， $\sigma_k(\boldsymbol{\beta}^*)_1$ 的系数在 $t = 1$ 时为常数，在 $t = 2$时按 $1/\sqrt{k}$的规律衰减；而 ϵ 的系数在 $t = 1$ 时按 $\sqrt{k}$ 的规律增长，但在$t = 2$时保持不变。这反映了不同范数下稀疏性驱动的近似误差与噪声敏感性之间众所周知的权衡。

5.2. 相干性下的稀疏恢复

在本节中，我们将证明，较小的 ℓ_1-相干性（见定义 4.23）也能保证 ℓ_1-最小化的成功。

定理 5.10 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为一个具有 ℓ_2-归一化列的矩阵。若

$$\mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, k-1) < 1, \tag{5.7}$$

则每个 k-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 都可以通过求解 ℓ_1-最小化问题（P_1），从测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$中精确恢复。

证明 [定理 5 的10] 根据推论5.3，只需验证 $\boldsymbol{X}$ 满足阶为 k 的 NSP 条件，即验证

$$\|\boldsymbol{\alpha}_{\mathbb{S}}\|_1 < \|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_1 \tag{5.8}$$

对每个非零向量 $\boldsymbol{\alpha} \in \mathcal{N}(\boldsymbol{X}) \setminus \{\mathbf{0}\}$ 以及每个满足 $|\mathbb{S}| = k$ 的指标集 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$，均有此性质。由于 $\boldsymbol{X}\boldsymbol{\alpha} = \mathbf{0}$，可得 $\sum_{j=1}^{p} \alpha_j \boldsymbol{x}_j = \mathbf{0}$。将两边与某个列 $\boldsymbol{x}_i$（其中 $i \in \mathbb{S}$）取内积，并利用 $\|\boldsymbol{x}_i\|_2 = 1$，可得

$$\alpha_i = \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle = -\sum_{j=1, j\neq i}^{p} \alpha_j \langle \boldsymbol{x}_j, \boldsymbol{x}_i \rangle = -\sum_{h \in \mathbb{S}^c} \alpha_h \langle \boldsymbol{x}_h, \boldsymbol{x}_i \rangle - \sum_{j \in \mathbb{S}, j\neq i} \alpha_j \langle \boldsymbol{x}_j, \boldsymbol{x}_i \rangle ,$$

由此可得

$$|\alpha_i| \leq \sum_{h \in \mathbb{S}^c} |\alpha_h| \, |\langle \boldsymbol{x}_h, \boldsymbol{x}_i \rangle| + \sum_{j \in \mathbb{S}, j\neq i} |\alpha_j| \, |\langle \boldsymbol{x}_j, \boldsymbol{x}_i \rangle| .$$

对所有 $i \in \mathbb{S}$ 求和并交换求和顺序，可得

$$\begin{aligned} \|\boldsymbol{\alpha}_{\mathbb{S}}\|_1 = \sum_{i \in \mathbb{S}} |\alpha_i| &\leq \sum_{h \in \mathbb{S}^c} |\alpha_h| \left(\sum_{i \in \mathbb{S}} |\langle \boldsymbol{x}_h, \boldsymbol{x}_i \rangle| \right) + \sum_{j \in \mathbb{S}} |\alpha_j| \left(\sum_{i \in \mathbb{S}, i\neq j} |\langle \boldsymbol{x}_j, \boldsymbol{x}_i \rangle| \right) \\ &\leq \sum_{h \in \mathbb{S}^c} |\alpha_h| \cdot \mu_1(\boldsymbol{X}, k) + \sum_{j \in \mathbb{S}} |\alpha_j| \cdot \mu_1(\boldsymbol{X}, k-1) \\ &= \mu_1(\boldsymbol{X}, k) \, \|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_1 + \mu_1(\boldsymbol{X}, k-1) \, \|\boldsymbol{\alpha}_{\mathbb{S}}\|_1 . \end{aligned}$$

重新整理各项，得到 $\big(1 - \mu_1(\boldsymbol{X}, k-1)\big) \, \|\boldsymbol{\alpha}_{\mathbb{S}}\|_1 \leq \mu_1(\boldsymbol{X}, k) \, \|\boldsymbol{\alpha}_{\mathbb{S}^c}\|_1$。在假设（5.7）下，有 $1 - \mu_1(\boldsymbol{X}, k-1) > \mu_1(\boldsymbol{X}, k)$，这意味着（5.8）。因此，阶数为 k 的零空间性质成立，由推论 ℓ_1可知，通过 ℓ_1-最小化可实现精确恢复。5.3。 ■

5.3. RIP 下的稀疏恢复

本节将证明，对于受限等距常数较小的测量矩阵，通过基追踪可以成功实现稀疏恢复。我们将给出这一结果的几种证明。第一种证明使用条件 $\delta_k + \theta_{k,k} + \theta_{k,2k} < 1$，该条件由对偶证书定理（定理3.19）推得。第二种证明更简单，也更直观。它表明，条件 $\delta_{2k} < 1/3$ 足以保证所有 k-稀疏向量都能通过 ℓ_1-最小化实现精确恢复。其余证明则更为复杂。我们将证明，较弱的条件 $\delta_{2k} < 0.4142$ 或 $\delta_{2k} < 0.4931$ 实际上足以确保所有 k-稀疏向量都能通过 ℓ_1-最小化实现稳定且鲁棒的恢复。

5.3.1 对偶证书：再探

如定理 3.19所述，向量 $\boldsymbol{\beta}^*$支撑于 $\mathbb{S} = \mathrm{supp}(\boldsymbol{\beta}^*)$，是（$\mathrm{P}_1$）的唯一极小化解，当且仅当子矩阵 $\boldsymbol{X}_{\mathbb{S}}$ 满秩，且存在向量 $\boldsymbol{\lambda}$ 满足以下两个性质：

(i) $\langle \boldsymbol{\lambda}, \ \boldsymbol{x}_i \rangle =$ 对所有 $i \in \mathbb{S}$ 均有 $\mathrm{sign}(\beta_i^*)$;
(ii) $|\langle \boldsymbol{\lambda}, \ \boldsymbol{x}_i \rangle| < 1$，对所有 $i \in \mathbb{S}^c$ 均成立。

这个向量 $\boldsymbol{\lambda}$ 称为问题 的对偶证书 (P_1)。接下来的两个引理表明，在受限等距性质（RIP）下，这种对偶证书是存在的。引理 5.11 证明内积 $\langle \boldsymbol{\lambda}, \boldsymbol{x}_i \rangle$ 对于 $i \in \mathbb{S}^c$ 都有适当的上界，但可能有一个很小的“例外集”不满足这一点。引理 5.12进一步强化了这一结果，彻底消除了例外集。

引理 5.11：（对偶重建性质，ℓ_2 版本）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个满足阶为 $k \geq 1$ 、常数为 $\delta_k < 1$的 RIP 条件的设计矩阵，并设 $\boldsymbol{\beta}^*$ 是一个支撑于 $\mathbb{S} \subset \{1,2,\ldots,p\}$ 且满足 $|\mathbb{S}| \leq k$ 的实向量。那么存在一个向量 $\boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{X})$，使得 $\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle =$ 对于所有 $i \in \mathbb{S}$均有 $\operatorname{sign}(\beta_i^*)$。此外，存在一个与 $\mathbb{S}$不相交的“例外集” $\mathbb{T} \subset \{1,2,\ldots,p\}$，其基数至多为$|\mathbb{T}| \leq k'$ $(k' \geq 1)$，使得

$$|\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle| \leq \frac{\theta_{k,k'}}{(1-\delta_k)\sqrt{k'}} \cdot \sqrt{k}, \quad \text{for all } i \in (\mathbb{S}\cup\mathbb{T})^c$$

and

$$\|\boldsymbol{X}_{\mathbb{T}}\boldsymbol{\lambda}\|_2 = \left(\sum_{i\in\mathbb{T}} |\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle|^2\right)^{1/2} \leq \frac{\theta_{k,k'}}{1-\delta_k} \cdot \sqrt{k}.$$

此外，$\|\boldsymbol{\lambda}\|_2 \leq C\sqrt{k}$ 对于某个仅取决于 δ_k 的常数 $C > 0$ 。

证明 [引理 511] 证明略有不同，但基于 Candes 和 Tao (2005年)。RIP 条件（4.36a）意味着特征值满足 $\lambda(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}) \in [1-\delta_k, 1+\delta_k]$。由于我们假设 $\delta_k < 1$， $\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}$ 可逆，且

$$\left\|(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\right\|_2 \leq \frac{1}{1-\delta_k} \quad \text{and} \quad \left\|\boldsymbol{X}_{\mathbb{S}}(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\right\|_2 \leq \frac{\sqrt{1+\delta_k}}{1-\delta_k} \tag{5.9}$$

根据习题 4.12。定义 $\boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{X})$ 为 $\boldsymbol{\lambda} \triangleq \boldsymbol{X}_{\mathbb{S}}(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)$。显然，$\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{\lambda} = \operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)$，即 $\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle = \operatorname{sign}(\beta_i^*)$ 对所有 $i \in \mathbb{S}$ 均成立。此外，根据 (5.9) 和 (1.29)， $\|\boldsymbol{\lambda}\|_2 \leq C \cdot \|\operatorname{sign}(\beta_i^*)\|_2$，其中 $C \triangleq \sqrt{1+\delta_k/(1-\delta_k)}$。现在令 $\mathbb{S}' \subseteq \{1,2,\ldots,p\}\setminus\mathbb{S}$，即 $\mathbb{S}'$ 是 $\{1,2,\ldots,p\}$ 中任意一个与 $\mathbb{S}$ 不相交且满足 $|\mathbb{S}'| \leq k'$ 的集合，并令 $\boldsymbol{d}_{\mathbb{S}'} \triangleq \{d_i\}_{i\in\mathbb{S}'} \in \mathbb{R}^{|\mathbb{S}'|}$ 任意。于是，ROP 条件 (4.42a) 和 (5.9) 给出

$$\begin{aligned}\left|\left\langle \boldsymbol{X}_{\mathbb{S}'}^\top \boldsymbol{\lambda}, \boldsymbol{d}_{\mathbb{S}'}\right\rangle\right| &= |\langle \boldsymbol{\lambda}, \boldsymbol{X}_{\mathbb{S}'}\boldsymbol{d}_{\mathbb{S}'}\rangle| = \left|\left\langle \sum_{i\in\mathbb{S}} \left((\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)\right)_i \boldsymbol{x}_i, \sum_{i\in\mathbb{S}'} d_i \boldsymbol{x}_i\right\rangle\right| \\ &\leq \theta_{k,k'} \cdot \left\|(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)\right\|_2 \cdot \|\boldsymbol{d}_{\mathbb{S}'}\|_2 \leq \frac{\theta_{k,k'}}{1-\delta_k}\|\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)\|_2 \cdot \|\boldsymbol{d}_{\mathbb{S}'}\|_2 .\end{aligned}$$

由于 $\boldsymbol{d}_{\mathbb{S}'}$ 是任意的，根据 Hölder 不等式，令 $\boldsymbol{d}_{\mathbb{S}'} \triangleq \boldsymbol{X}_{\mathbb{S}'}^\top \boldsymbol{\lambda} / \left\|\boldsymbol{X}_{\mathbb{S}'}^\top \boldsymbol{\lambda}\right\|_2$ 可得

$$\left\|\boldsymbol{X}_{\mathbb{S}'}^\top \boldsymbol{\lambda}\right\|_2 = \left(\sum_{i\in\mathbb{S}'} |\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle|^2\right)^{1/2} \leq \frac{\theta_{k,k'}}{1-\delta_k}\|\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)\|_2 \tag{5.10}$$

只要 $\mathbb{S}' \subseteq \{1,2,\ldots,p\}\setminus\mathbb{S}$ 且 $|\mathbb{S}'| \leq k'$。最后，定义例外集

$$\mathbb{T} \triangleq \left\{ i \in \{1,2,\ldots,p\}\setminus\mathbb{S} \mid |\langle \boldsymbol{\lambda}, \boldsymbol{x}_i\rangle| > \frac{\theta_{k,k'}}{(1-\delta_k)\sqrt{k'}} \cdot \|\operatorname{sign}(\boldsymbol{\beta}_{\mathbb{S}}^*)\|_2 \right\},$$

则 $|\mathbb{T}|$ 必须满足 $|\mathbb{T}| \leq k'$，否则我们可以通过取 5.10 的一个子集$\mathbb{S}'$，该子集属于 $\mathbb{T}$ 且基数为 k'，从而得到矛盾。 ■

现在，我们通过迭代消除例外集，构造一个对 $|\langle \boldsymbol{\lambda}$进行更好控制的$\boldsymbol{x}_i\rangle|$ 对偶向量，使其对支撑集 $\mathbb{S}$之外的索引 i 成立 $\mathbb{T}$，同时保持$\mathbb{S}$ 上的取值不变。以下结果及其证明改编自Candes 和 Tao (2005年)。

引理 5.12：（对偶重建性质，ℓ_∞ 版本）假设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个满足 RIP 和受限正交性条件的设计矩阵，其 $k \geq 1$ 和常数满足 $\delta_k + \theta_{k,\ k} + \theta_{k,\ 2k} < 1$，并设 $\boldsymbol{\beta}^*$ 是一个支撑于 $\mathbb{S} \subset \{1, 2, \ldots, p\}$ 且满足$|\mathbb{S}| \leq k$的实向量。那么存在一个向量 $\boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{X})$，使得 $\langle \boldsymbol{\lambda}, \boldsymbol{x}_i \rangle =$ 对所有 $i \in \mathbb{S}$都有 $\operatorname{sign}(\beta_i^*)$。此外，$\boldsymbol{\lambda}$满足

$$|\langle \boldsymbol{\lambda}, \boldsymbol{x}_i \rangle| \leq \frac{\theta_{k,k}}{(1 - \delta_k - \theta_{k,2k})} < 1, \quad \text{for all } i \in \mathbb{S}^c. \tag{5.11}$$

证明 [引理5.12] 集合 $\mathbb{S}_0 \triangleq \mathbb{S}$。根据引理 5.11，存在一个向量 $\boldsymbol{\lambda}^{(1)} \triangleq \boldsymbol{X}_{\mathbb{S}_0}(\boldsymbol{X}_{\mathbb{S}_0}^\top \boldsymbol{X}_{\mathbb{S}_0})^{-1}$ 符号函数$(\boldsymbol{\beta}^*_{\mathbb{S}_0}) \in \mathcal{C}(\boldsymbol{X})$ 和一个集合 $\mathbb{S}_1 \subseteq \{1, 2, \ldots, p\}$ 使得

$$\begin{aligned}
&\mathbb{S}_0 \cap \mathbb{S}_1 = \varnothing \ \text{ with } \ |\mathbb{S}_1| \leq k;\\
&\left\langle \boldsymbol{\lambda}^{(1)}, \boldsymbol{x}_i \right\rangle = \operatorname{sign}(\beta_i^*), \qquad \text{for all } i \in \mathbb{S}_0;\\
&\left|\left\langle \boldsymbol{\lambda}^{(1)}, \boldsymbol{x}_i \right\rangle\right| \leq \frac{\theta_{k,k}}{(1-\delta_k)}, \qquad \text{for all } i \in (\mathbb{S}_0 \cup \mathbb{S}_1)^c;\\
&\left\|\boldsymbol{X}_{\mathbb{S}_1}^\top \boldsymbol{\lambda}^{(t+1)}\right\|_2 \leq \frac{\theta_{k,k}}{1-\delta_k}\sqrt{k};\\
&\left\|\boldsymbol{\lambda}^{(1)}\right\|_2 \leq C\sqrt{k}.
\end{aligned}$$

由于假设 $\delta_k + \theta_{k,\ k} + \theta_{k,\ 2k} < 1$ 意味着 $\delta_{2k} \leq \delta_k + \theta_{k,\ k} < 1$ 根据命题 4.40，使得 $\boldsymbol{X}$ 满足阶数为 $2k$的 RIP 条件。然后，我们迭代构造一个集合 $\mathbb{S}_t$，其中 $|\mathbb{S}_t| \leq k$ 对于 $t \geq 2$成立。令 $\mathbb{I}_t \triangleq \mathbb{S}_t \cup \mathbb{S}_0$，并对 $i \in \mathbb{I}_t$ 定义 $\boldsymbol{b}^{(t)} \in \mathbb{R}^{|\mathbb{I}_t|}$，使得

$$b_i^{(t)} = \begin{cases} \left\langle \boldsymbol{\lambda}^{(t)}, \boldsymbol{x}_i \right\rangle, & i \in \mathbb{S}_t; \\ 0, & i \in \mathbb{S}_0. \end{cases}$$

进一步定义 $\boldsymbol{\lambda}^{(t+1)} \triangleq \boldsymbol{X}_{\mathbb{I}_t}(\boldsymbol{X}_{\mathbb{I}_t}^\top \boldsymbol{X}_{\mathbb{I}_t})^{-1}\boldsymbol{b}^{(t)}$，其中 $(\boldsymbol{X}_{\mathbb{I}_t}^\top \boldsymbol{X}_{\mathbb{I}_t})$ 可逆，因为 $\boldsymbol{X}$ 满足阶数为 $2k$的 RIP 条件。迭代应用引理 5.11 可得一个向量序列$\boldsymbol{\lambda}^{(t+1)} \triangleq \boldsymbol{X}_{\mathbb{I}_t}(\boldsymbol{X}_{\mathbb{I}_t}^\top \boldsymbol{X}_{\mathbb{I}_t})^{-1}\boldsymbol{b}^{(t)} \in \mathcal{C}(\boldsymbol{X})$ 和集合 $\mathbb{S}_{t+1} \subseteq \{1, 2, \ldots, p\}$，对所有 $t \geq 1$ 均满足

$$\mathbb{S}_t \cap \ (\mathbb{S}_0 \cup \mathbb{S}_{t+1}) \ = \varnothing \text{ 与} |\mathbb{S}_{t+1}| \leq k; \tag{5.12a}$$

$$\left\langle \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{x}_i \right\rangle = \left\langle \boldsymbol{\lambda}^{(t)}, \boldsymbol{x}_i \right\rangle, \qquad \text{for all } i \in \mathbb{S}_t; \tag{5.12b}$$

$$\left\langle \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{x}_i \right\rangle = 0, \qquad \text{for all } i \in \mathbb{S}_0; \tag{5.12c}$$

$$\left|\left\langle \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{x}_i \right\rangle\right| \leq \frac{\theta_{k,k}}{1-\delta_k}\left(\frac{\theta_{k,2k}}{1-\delta_{2k}}\right)^t \leq \frac{\theta_{k,k}}{1-\delta_{2k}}\left(\frac{\theta_{k,2k}}{1-\delta_{2k}}\right)^t, \quad \forall\, i \in (\mathbb{S}_0 \cup \mathbb{S}_t \cup \mathbb{S}_{t+1})^c; \tag{5.12d}$$

$$\left\|\boldsymbol{X}_{\mathbb{S}_{t+1}}^\top \boldsymbol{\lambda}^{(t+1)}\right\|_2 \leq \frac{\theta_{k,k}}{1-\delta_k}\left(\frac{\theta_{k,2k}}{1-\delta_{2k}}\right)^t \sqrt{k} \leq \frac{\theta_{k,k}}{1-\delta_{2k}}\left(\frac{\theta_{k,2k}}{1-\delta_{2k}}\right)^t \sqrt{k}; \tag{5.12e}$$

$$\left|\boldsymbol{\lambda}^{(t+1)}\right| \leq \left(\frac{\theta_{k,\ 2k}}{1-\delta_{2k}}\right)^{t-1} C, \tag{5.12f}$$

其中我们使用了以下事实：$\delta_k \leq \delta_{2k}$（习题 4.34）。不等式（5.12d）成立是因为 $|\mathbb{I}_t| \leq 2k$（受限正交常数满足 $\theta_{k,\ 2k} \equiv \theta_{2k,\ k}$，因此参数 $\theta_{k,\ 2k}$ 而不是 $\theta_{k,\ k}$），以及 $\|\boldsymbol{b}^{(t)}\|_2 = \|\boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{\lambda}^{(t)}\|_2$ 的迭代论证：

$$\left\|\boldsymbol{X}_{\mathbb{S}_{t+1}}^\top \boldsymbol{\lambda}^{(t+1)}\right\|_2 \leq \frac{\theta_{k,2k}}{1-\delta_{2k}}\left\|\boldsymbol{b}^{(t)}\right\|_2 = \frac{\theta_{k,2k}}{1-\delta_{2k}}\left\|\boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{\lambda}^{(t)}\right\|_2.$$

不等式（5.12e）因此相应地成立。

根据假设，我们有 $\theta_{k,\ 2k}+\delta_{2k}\le\theta_{k,\ 2k}+\delta_k+\theta_{k,\ k}<1$ （命题 4.40），使得$\theta_{k,\ 2k}\frac{2k}{1-\delta_{2k}}<1$。因此，如果我们设 $\boldsymbol{\lambda}\triangleq\sum_{t=1}^{\infty}(-1)^{t-1}\boldsymbol{\lambda}^{(t)}$，那么该级数绝对收敛，从而，$\boldsymbol{\lambda}$ 是 $\mathcal{C}(\boldsymbol{X})$中的一个定义良好的向量。现在我们考察其系数

$$\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle=\sum_{t=1}^{\infty}(-1)^{t-1}\left\langle\boldsymbol{\lambda}^{(t)},\boldsymbol{x}_i\right\rangle,\quad\text{for } i\in\{1,2,\dots,p\}.\tag{5.13}$$

一方面，对于 $i\in\mathbb{S}_0$，由构造可知 $\langle\boldsymbol{\lambda}^{(1)},\ \boldsymbol{x}_i\rangle=\operatorname{sign}(\beta_i^*)$ 且$\langle\boldsymbol{\lambda}^{(t)},\ \boldsymbol{x}_i\rangle=0$ 对所有 $t\ge2$ 均成立，因此，

$$\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle=\operatorname{sign}(\beta_i^*),\quad\text{for all } i\in\mathbb{S}_0.$$

另一方面，我们考察带有 $i\in\mathbb{S}_0^c$ 的 i。例如，令 $i\in\mathbb{S}_2$，则有

$$\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle=\sum_{t=1}^{\infty}(-1)^{t-1}\left\langle\boldsymbol{\lambda}^{(t)},\boldsymbol{x}_i\right\rangle=\sum_{t\notin\{2,3\}}^{\infty}(-1)^{t-1}\left\langle\boldsymbol{\lambda}^{(t)},\boldsymbol{x}_i\right\rangle,$$

其中，项 $-\langle\boldsymbol{\lambda}^{(2)},\ \boldsymbol{x}_i\rangle+\langle\boldsymbol{\lambda}^{(3)},\ \boldsymbol{x}_i\rangle$ 根据（5.12b）相互抵消。其他项满足（5.12d）。因此，$\langle\boldsymbol{\lambda},\ \boldsymbol{x}_i\rangle$ 可由几何级数得到：

$$\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle\le\frac{\theta_{k,k}}{1-\delta_{2k}-\theta_{k,2k}},\quad\text{for all } i\in\mathbb{S}^c.$$

证明至此完成。 ■

引理5.12 有效地解决了对偶重构问题：它保证存在一个向量 $\boldsymbol{\lambda}\in\mathcal{C}(\boldsymbol{X})$，该向量满足本节开头所述的性质（i）和（ii）。将此引理与定理 3.19 中的对偶证书结果结合起来，即可立即确立 ℓ_1-最小化问题（P_1）的解的唯一性。或者，也可以仅利用 RIP 直接得到相同的结论。这就引出了以下恢复保证。

定理 5.13：（在 RIP 和 ROP 下精确恢复 ℓ_1 （Candes 和 Tao，2005年）） 设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 是满足 RIP 和 ROP 条件的设计矩阵，其中$k\ge1$ 且

$$\delta_k+\theta_{k,k}+\theta_{k,2k}<1,\tag{5.14}$$

并令 $\mathbb{S}\subset\{1,2,\dots,p\}$ 为任意指标集，满足 $|\mathbb{S}|\le k$。令 $\boldsymbol{y}\triangleq\boldsymbol{X}\boldsymbol{\beta}^*$，其中 $\boldsymbol{\beta}^*$ 是一个支撑于 $\mathbb{S}$ 的任意向量。那么，$\boldsymbol{\beta}^*$ 是该优化问题的**唯一**极小化解

$$(\mathrm{P}_1)\qquad\min\|\boldsymbol{\beta}\|_1\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}.$$

证明 [定理 5 的13] 凸分析中的标准结果保证至少存在一个极小化解 $\widehat{\boldsymbol{\beta}}$（$\mathrm{P}_1$）。我们需要证明 $\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}^*$。由于 $\boldsymbol{\beta}^*$ 是（P_1）的可行点，并且 $\widehat{\boldsymbol{\beta}}$ 是（P_1）的极小化解，因此可得

$$\left\|\widehat{\boldsymbol{\beta}}\right\|_1\le\|\boldsymbol{\beta}^*\|_1=\sum_{i\in\mathbb{S}}|\beta_i^*|.\tag{5.15}$$

根据引理5.12，存在一个向量 $\boldsymbol{\lambda}$ 满足本节开头所述的性质（i）和（ii）。于是可得

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}\|_1 &= \sum_{i\in\mathbb{S}}\left|\beta_i^* + (\widehat{\beta}_i - \beta_i^*)\right| + \sum_{i\in\mathbb{S}^c}\left|\widehat{\beta}_i\right| \geq \sum_{i\in\mathbb{S}}\operatorname{sign}(\beta_i^*)\left(\beta_i^* + (\widehat{\beta}_i - \beta_i^*)\right) + \sum_{i\in\mathbb{S}^c}\widehat{\beta}_i|\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle| \\ &= \|\boldsymbol{\beta}^*\|_1 + \sum_{i\in\mathbb{S}}(\widehat{\beta}_i - \beta_i^*)|\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle| + \sum_{i\in\mathbb{S}^c}\widehat{\beta}_i|\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle| = \|\boldsymbol{\beta}^*\|_1 + \left\langle\boldsymbol{\lambda}, \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^*\right\rangle \\ &= \|\boldsymbol{\beta}^*\|_1 + \langle\boldsymbol{\lambda}, \boldsymbol{y} - \boldsymbol{y}\rangle = \|\boldsymbol{\beta}^*\|_1 .\end{aligned} \tag{5.16}$$

结合（5.15）和（5.16）可得 $\|\widehat{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}^*\|_1$。由于 $|\langle\boldsymbol{\lambda},\ \boldsymbol{x}_i\rangle|$ 对于所有 $i\in\mathbb{S}^c$都严格小于 1，因此这特别地迫使 $\beta_i = 0$ 对所有 $i\in\mathbb{S}^c$都成立。为说明这一点，若 $\operatorname{supp}(\boldsymbol{\beta}) \neq \mathbb{S}$，则（5.16）还表明

$$0 = \sum_{i\in\mathbb{S}}(\beta_i^* - \widehat{\beta}_i)\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle - \sum_{i\in\mathbb{S}^c}\widehat{\beta}_i\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle = \|\boldsymbol{\beta}^*\|_1 - \underbrace{\sum_{i\in\mathbb{S}}\operatorname{sign}(\beta_i^*)\widehat{\beta}_i}_{\leq\sum_{i\in\mathbb{S}}|\widehat{\beta}_i|} - \underbrace{\sum_{i\in\mathbb{S}^c}\widehat{\beta}_i\langle\boldsymbol{\lambda},\boldsymbol{x}_i\rangle}_{<\sum_{i\in\mathbb{S}^c}|\widehat{\beta}_i|} > 0,$$

这与 $\|\widehat{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}^*\|_1$ 这一事实矛盾。因此，$\operatorname{supp}(\widehat{\boldsymbol{\beta}}) = \mathbb{S}$ 且 $\boldsymbol{y} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}^*$，

应用（4.36a）中的 RIP 条件以及 $\delta_k < 1$ 这一事实，我们得出 $\boldsymbol{\beta}_{\mathbb{S}} = \boldsymbol{\beta}_{\mathbb{S}}^*$。这证明了 $\boldsymbol{\beta}^*$ 的唯一性。 ■

注意，根据命题 4.40 ，（5.14）的假设意味着 $\delta_{2k} < 1$，而 $\delta_k + \delta_{2k} + \delta_{3k} < 1$ 又蕴含这一假设。因此，在 5.14）中，条件大致比在 5.1 定理的 ℓ_0-最小化模型（$\delta_{2k} < 1$）中所需的条件“严格三倍”。根据前面的讨论，我们得到以下简单结果，它要求 $\delta_{2k} < 1/3$ ，并且其条件大致也比 5.1 定理所需的条件“严格三倍”。

定理 5.14：（ℓ_1 在 RIP 条件下的精确恢复）设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 满足阶数为 $2k$ 、受限等距常数为

$$\delta_{2k} < \frac{1}{3}. \tag{5.17}$$

那么，每个 k-稀疏向量 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 都是以下问题的**唯一**解

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}^*.$$

证明 [定理 定理5.14] 证明直接由推论5.3 和定理4.50得出。 ■

探讨精确恢复之外的问题，并建立稳定性和鲁棒性，是很有启发性的——也就是说，当测量值含有噪声或信号仅近似稀疏时，相关保证仍然成立。下面的习题将引导我们进行这样的推广。

练习 5.15。 细化定理 5.14 的证明，以确立在条件 k下，通过基追踪实现的 $\delta_{2k} < 1/3$-稀疏恢复的稳定性和鲁棒性。具体而言，证明重构误差可由测量噪声水平以及真实信号的最佳 k项逼近误差所界定。 提示：参见定理 *5.7*，或参见第*5.3.4* 节了解更多细节。

5.3.2 无噪声信号恢复

练习 5.15 讨论了在阶数为 $2k$、常数为 $\delta_{2k}<1/3$ 的限制等距性质下的稳定恢复。现在我们将这一要求放宽为较弱的条件 $\delta_{2k}<\sqrt{2}-1$。为此，我们采用以下一般性结果，它细化了 Candes RIP 界引理（引理4.53）。

引理 5.16：（Candes RIP 界细化引理（Candes，2008年））假设 $\boldsymbol{X}$ 满足阶数为 $2k$、常数为 $\delta_{2k}<\sqrt{2}-1$的 RIP 条件。设 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\in\mathbb{R}^p$已知，并定义 $\boldsymbol{d}\triangleq\boldsymbol{\alpha}-\boldsymbol{\beta}$。令 $\mathbb{S}_0$ 表示 k 中绝对值最大的元素所对应的指标集，并令 $\boldsymbol{\beta}$，令 $\mathbb{S}_1$ 表示 k 中绝对值最大的元素所对应的指标集 $\boldsymbol{d}_{\mathbb{S}_0^c}$。设$\mathbb{S}\triangleq\mathbb{S}_0\cup\mathbb{S}_1$。若 $\|\boldsymbol{\alpha}\|_1\le\|\boldsymbol{\beta}\|_1$，则

$$\|\boldsymbol{d}\|_2\le C_0\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}+C_1\frac{|\langle\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2},\tag{5.18}$$

其中

$$C_0=2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}\qquad\text{and}\qquad C_1=\frac{2}{1-(1+\sqrt{2})\delta_{2k}}.$$

引理 [的证明 5.16] 由于 $\boldsymbol{d}=\boldsymbol{d}\left(\mathbb{S}\right)+\boldsymbol{d}\left(\mathbb{S}^c\right)$，由三角不等式可得

$$\|\boldsymbol{d}\|_2\le\|\boldsymbol{d}_{\mathbb{S}}\|_2+\|\boldsymbol{d}_{\mathbb{S}^c}\|_2\,.\tag{5.19}$$

我们首先证明上界 $\|\boldsymbol{d}_{\mathbb{S}^c}\|_2$。根据分块有序集合不等式（引理 4.52），我们有 e

$$\|\boldsymbol{d}_{\mathbb{S}^c}\|_2=\left\|\sum_{i\ge2}\boldsymbol{d}_{\mathbb{S}_i}\right\|_2\le\sum_{i\ge2}\|\boldsymbol{d}_{\mathbb{S}_i}\|_2\le\frac{\left\|\boldsymbol{d}_{\mathbb{S}_0^c}\right\|_1}{\sqrt{k}},\tag{5.20}$$

其中，集合 $\mathbb{S}_i$ 的定义与前面相同，即 $\mathbb{S}_1$ 是对应于 $\boldsymbol{d}_{\mathbb{S}_0^c}$ 中绝对值最大的 k 个元素的指标集，$\mathbb{S}_2$ 是对应于接下来绝对值最大的 k 个元素的指标集，依此类推。

接下来，我们给出 $\left\|\boldsymbol{d}_{\mathbb{S}_0^c}\right\|_1$的上界。由于 $\|\boldsymbol{\beta}+\boldsymbol{d}\|_1=\|\boldsymbol{\alpha}\|_1\le\|\boldsymbol{\beta}\|_1$，根据三角不等式可得

$$\begin{aligned}\|\boldsymbol{\beta}\|_1\ge\|\boldsymbol{\beta}+\boldsymbol{d}\|_1&=\left\|\boldsymbol{\beta}_{\mathbb{S}_0}+\boldsymbol{d}_{\mathbb{S}_0}\right\|_1+\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}+\boldsymbol{d}_{\mathbb{S}_0^c}\right\|_1\\&\ge\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_1-\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+\left\|\boldsymbol{d}_{\mathbb{S}_0^c}\right\|_1-\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1.\end{aligned}$$

重新排列各项，并回顾 $\sigma_k(\boldsymbol{\beta})_1=\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1=\|\boldsymbol{\beta}-\boldsymbol{\beta}(\mathbb{S}_0)\|_1$，我们得到

$$\begin{aligned}\left\|\boldsymbol{d}_{\mathbb{S}_0^c}\right\|_1&\le\|\boldsymbol{\beta}\|_1-\left\|\boldsymbol{\beta}_{\mathbb{S}_0}\right\|_1+\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1\\&\le\|\boldsymbol{\beta}-\boldsymbol{\beta}(\mathbb{S}_0)\|_1+\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+\left\|\boldsymbol{\beta}_{\mathbb{S}_0^c}\right\|_1=\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+2\sigma_k(\boldsymbol{\beta})_1.\end{aligned}\tag{5.21}$$

将（5.21）代入（5.20）得到

$$\|\boldsymbol{d}_{\mathbb{S}^c}\|_2\le\frac{\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+2\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}\le\|\boldsymbol{d}_{\mathbb{S}_0}\|_2+\frac{2\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}},\tag{5.22}$$

其中，最后一个不等式来自向量范数的标准界（习题1.26）。注意到 $\|\boldsymbol{d}_{\mathbb{S}_0}\|_2\le\|\boldsymbol{d}_{\mathbb{S}}\|_2$，将上述不等式代入式（5.19）可得

$$\|\boldsymbol{d}\|_2\le2\|\boldsymbol{d}_{\mathbb{S}}\|_2+\frac{2\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}.\tag{5.23}$$

现在对 $\|\boldsymbol{d}_{\mathbb{S}}\|_2$进行上界估计。结合 Candes RIP 界引理（引理4.53）和式（5.21），并再次应用向量范数的标准界（习题1.26），得到

$$\begin{aligned}\|\boldsymbol{d}_{\mathbb{S}}\|_2 &\leq \mu\frac{\|\boldsymbol{d}_{\mathbb{S}_0^c}\|_1}{\sqrt{k}}+\nu\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2}\leq \mu\frac{\|\boldsymbol{d}_{\mathbb{S}_0}\|_1+2\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}+\nu\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2}\\ &\leq \mu\|\boldsymbol{d}_{\mathbb{S}_0}\|_2+2\mu\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}+\nu\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2},\end{aligned}\tag{5.24}$$

其中 μ，ν 在引理 4.53中定义。由于 $\|\boldsymbol{d}_{\mathbb{S}_0}\|_2\leq\|\boldsymbol{d}_{\mathbb{S}}\|_2$，上述不等式变为

$$(1-\mu)\|\boldsymbol{d}_{\mathbb{S}}\|_2\leq 2\mu\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}+\nu\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2}.\tag{5.25}$$

假设 $\delta_{2k}<\sqrt{2}-1$ 保证 $\mu=\frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}}<1$。将两边除以（$1-\mu$），并将所得上界代入（5.23），可得

$$\|\boldsymbol{d}\|_2\leq\left(\frac{4\mu}{1-\mu}+2\right)\frac{\sigma_k(\boldsymbol{\beta})_1}{\sqrt{k}}+\frac{2\nu}{1-\mu}\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2}.\tag{5.26}$$

利用 μ 和 ν 的定义可得所需的常数。 ■

引理 5.16 为一类 ℓ_1-最小化算法建立了误差界，并假设测量矩阵 $\boldsymbol{X}$ 满足 RIP。为了推导具体的恢复保证，必须分析约束 $\mathcal{B}(\boldsymbol{y})$ 如何影响项 $|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\ \boldsymbol{X}\boldsymbol{d}\rangle|$。例如，考虑测量值精确的无噪声情形：$\mathcal{B}(\boldsymbol{y})=\{\boldsymbol{\alpha}\mid\boldsymbol{X}\boldsymbol{\alpha}=\boldsymbol{y}\}$。此时，我们得到以下结果。

定理 5.17：（(P_1) 在限制等距性质（Candes，2008年）下的误差界）假设$\boldsymbol{X}$ 满足 $2k$ 阶的限制等距性质，且 $\delta_{2k}<\sqrt{2}-1$，并令 $\boldsymbol{y}=\boldsymbol{X}\bar{\boldsymbol{\beta}}^*$ 其中 $\boldsymbol{\beta}^*\in\mathbb{R}^p$。那么，解 $\widehat{\boldsymbol{\beta}}$ 满足（P_1）的

$$\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_2\leq C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}},\tag{5.27}$$

其中 C_0 是引理 5.16 中定义的常数。

证明 [定理 5.17] 应用引理 5.16 取 $\boldsymbol{\alpha}\triangleq\widehat{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta}\triangleq\boldsymbol{\beta}^*$，可得对于 $\boldsymbol{d}=\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$，

$$\|\boldsymbol{d}\|_2\leq C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}}+C_1\frac{|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|}{\|\boldsymbol{d}_{\mathbb{S}}\|_2}.\tag{5.28}$$

此外，由于 $\widehat{\boldsymbol{\beta}}$，$\boldsymbol{\beta}^*$ 是可行的（即 $\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{X}\boldsymbol{\beta}^*$），因此我们还有 $\boldsymbol{X}\boldsymbol{d}=\boldsymbol{0}$。于是，上界中的第二项消失，从而得到所声称的不等式。 ■

定理 5.17 相当引人注目。特别是，若 $\boldsymbol{\beta}\in\mathbb{B}_0[k]\triangleq\{\boldsymbol{\alpha}\mid\|\boldsymbol{\alpha}\|_0\leq k\}$，那么只要 $\boldsymbol{X}$ 满足限制等距性质（RIP），我们就能精确恢复任意 k-稀疏向量 $\boldsymbol{\beta}^*$。乍看之下，这种精确恢复似乎令人惊讶，也暗示该方法可能对噪声高度敏感。然而，正如我们接下来将看到的，引理 5.16 也可用于证明该方法实际上是稳定的（见定理 5.20）。

最后请注意，定理5.17 假设 RIP 对 $\boldsymbol{X}$成立。我们可以轻松地调整论证，改为假设 NSP，依据的是已知结论：RIP 蕴含 NSP（参见第4.3.5节）。

5.3.3 实例最优保证

现在我们简要回到无噪声情形，更仔细地考察恢复非稀疏信号的实例最优保证。首先，我们来介绍实例最优性的概念。

定义 5.18（实例最优性）。给定 $s \geq 1$，如果一对测量矩阵$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和重构映射 $\Delta: \mathbb{R}^n \to \mathbb{R}^p$ 满足以下条件，则称其为 ℓ_s阶实例最优， k常数为 $C > 0$：

$$\|\boldsymbol{\beta} - \Delta(\boldsymbol{X}\boldsymbol{\beta})\|_s \leq C\sigma_k(\boldsymbol{\beta})_s, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p,$$

其中 $\sigma_k(\boldsymbol{\beta})_s \triangleq \min_{\widehat{\boldsymbol{\beta}}\in\mathbb{B}_0[k]} \|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|_s$。

在定理 5.17中，我们给出了问题（ ℓ_2范数重构误差的上界P_1）：

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C_0 \frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}}. \tag{5.29}$$

这一结果可以推广到其他范数：对于任意 $s \in [1,2]$，都可以给出重构误差在 ℓ_s范数下的上界。例如，只需对相同的论证稍作修改，同样可以证明：

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \leq C\sigma_k(\boldsymbol{\beta}^*)_1;$$

详见定理 5.21 和 Candes （2008年）以了解详细信息。这自然引出了一个问题：我们能否将右侧替换为 σ_k（$\boldsymbol{\beta}^*$）$_2$,，即得到如下形式的上界

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C\sigma_k(\boldsymbol{\beta}^*)_2?$$

遗憾的是，这样的 ℓ_2-实例最优保证需要付出测量数量方面的高昂代价，具体如下述定理所示。

定理 5.19：（ℓ_2-实例最优性（Cohen 等人，2009年））设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满秩 n （$p \geq n$），并设 Δ： $\mathbb{R}^n \to \mathbb{R}^p$ 是一个满足下式的恢复算法

$$\|\boldsymbol{\beta} - \Delta(\boldsymbol{X}\boldsymbol{\beta})\|_2 \leq C\sigma_k(\boldsymbol{\beta})_2, \qquad \text{for some } k \geq 1, \tag{5.30}$$

（即， ℓ_2-阶实例最优）。那么测量数量必须满足$n > \left(1 - \sqrt{1 - 1/C^2}\right) p$。

定理 5 的证明 [19] 令 $\boldsymbol{d} \in \mathbb{R}^p$ 为 $\boldsymbol{X}$零空间中的任意向量，即 $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{X})$。将 $\boldsymbol{d}$分解为 $\boldsymbol{d} = \boldsymbol{d}(\mathbb{S}) + \boldsymbol{d}(\mathbb{S}^c)$，其中 $\mathbb{S} \subseteq \{1,2,\ldots,p\}$是任意指标集，满足$|\mathbb{S}| \leq k$。令 $\boldsymbol{\beta} = \boldsymbol{d}(\mathbb{S}^c)$。由于 $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{X})$，我们有 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}_{\mathbb{S}^c}\boldsymbol{d}_{\mathbb{S}^c} = \boldsymbol{X}\boldsymbol{d} - \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}} = -\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}}$。另一方面，由于 $\boldsymbol{d}(\mathbb{S}) \in \mathbb{B}_0[k]$， $\sigma_k(\boldsymbol{d}(\mathbb{S}))_2 = 0$；由 (5.30) 可知， $\Delta(\boldsymbol{X}\boldsymbol{\beta}) \equiv \Delta(-\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}}) = -\boldsymbol{d}_{\mathbb{S}}$。因此， $\|\boldsymbol{\beta} - \Delta(\boldsymbol{X}\boldsymbol{\beta})\|_2 = \|\boldsymbol{d}_{\mathbb{S}^c} - (-\boldsymbol{d}_{\mathbb{S}})\|_2 = \|\boldsymbol{d}\|_2 \leq C\sigma_k(\boldsymbol{\beta})$。此外，根据 $\sigma_k(\boldsymbol{\beta})_2$,的定义，我们有 $\sigma_k(\boldsymbol{\beta})_2 \leq \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2$，对所有 $\boldsymbol{\alpha} \in \mathbb{B}_0[k]$都成立，包括 $\boldsymbol{\alpha} = \boldsymbol{0}$。因此，令 $\boldsymbol{\alpha} = \boldsymbol{0}$，即可得到$\|\boldsymbol{d}\|_2 \leq C\|\boldsymbol{d}_{\mathbb{S}^c}\|_2$。利用$\|\boldsymbol{d}\|_2^2 = \|\boldsymbol{d}_{\mathbb{S}}\|_2^2 + \|\boldsymbol{d}_{\mathbb{S}^c}\|_2^2$这一事实，可得

$$\|\boldsymbol{d}_{\mathbb{S}}\|_2^2 = \|\boldsymbol{d}\|_2^2 - \|\boldsymbol{d}_{\mathbb{S}^c}\|_2^2 \leq \|\boldsymbol{d}\|_2^2 - \frac{1}{C^2}\|\boldsymbol{d}\|_2^2 = \left(1 - \frac{1}{C^2}\right)\|\boldsymbol{d}\|_2^2.$$

只要 $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{X})$ 且 $|\mathbb{S}| \leq k$， $\boldsymbol{d}$ 和 $\mathbb{S}$ 的取法就任意；特别地，令 $\{\boldsymbol{u}_i\}_{i=1}^{p-n}$ 为 $\mathcal{N}(\boldsymbol{X})$的一组标准正交基，并按如下方式定义向量 $\{\boldsymbol{d}_j\}_{j=1}^{p}$：

$$\boldsymbol{d}_j = \sum_{i=1}^{p-n} u_{ij}\boldsymbol{u}_i \equiv \sum_{i=1}^{p-n} \langle \boldsymbol{e}_j, \boldsymbol{u}_i \rangle \boldsymbol{u}_i = \sum_{i=1}^{p-n} (\boldsymbol{u}_i \boldsymbol{u}_i^\top)\boldsymbol{e}_j \triangleq \boldsymbol{P}_{\mathcal{N}}\boldsymbol{e}_j,$$

其中 $\boldsymbol{e}_j$ 表示 j个标准基向量，u_{ij} 表示 j的第 $\boldsymbol{u}_i$个元素，并且$\boldsymbol{P}_{\mathcal{N}} \triangleq \sum_{i=1}^{p-n} (\boldsymbol{u}_i \boldsymbol{u}_i^\top)$表示到 $\mathcal{N}(\boldsymbol{X})$上的正交投影（参见定义1.42）。由于 $\|\boldsymbol{P}_{\mathcal{N}}\boldsymbol{e}_j\|_2^2 + \left\|\boldsymbol{P}_{\mathcal{N}}^{\perp}\boldsymbol{e}_j\right\|_2^2 = \|\boldsymbol{e}_j\|_2^2 = 1$，因此有 $\|\boldsymbol{d}_j\|_2 \leq 1$。于是，令 $\mathbb{S} = \{j\}$对于 $\boldsymbol{d}_j$，可得

$$\left|\sum_{i=1}^{p-n}(u_{ij})^2\right|^2 = (d_{jj})^2 \leq \left(1 - \frac{1}{C^2}\right)\|\boldsymbol{d}_j\|_2^2 \leq 1 - \frac{1}{C^2}.$$

对 $j = 1$，2，……，p 求和，得到

$$p\sqrt{1 - 1/C^2} \geq \sum_{j=1}^{p}\sum_{i=1}^{p-n}(u_{ij})^2 = \sum_{i=1}^{p-n}\sum_{j=1}^{p}(u_{ij})^2 = \sum_{i=1}^{p-n}\|\boldsymbol{u}_i\|_2^2 = p - n,$$

重新整理后，即完成证明。 ■

因此，若我们希望得到形如（5.30）的上界，且该上界对所有信号 $\boldsymbol{\beta}$ 都成立，且常数为$C \approx 1$，那么无论采用何种重构算法，我们都必须进行几乎与环境维数一样多的测量值，即 $n \approx p$。

5.3.4 含噪信号恢复

在第 5.1.2 节中，我们讨论了稳健 NSP 或稳健 NSP 下的恢复保证，它们确保了噪声环境中的稳健信号重构。RIP 同样支持此类稳定性和稳健性保证，并在压缩感知理论中发挥着核心作用。文献中投入了大量工作来确定 RIC 的最大允许值δ_{2k}。具体而言，Candes（2008年）证明，$\delta_{2k} < \sqrt{2} - 1 \approx 0.4142$ 足以确保任意 k-稀疏信号都能精确恢复。Foucart 和 Lai（2009年）随后将该上界改进为 $\delta_{2k} < 2\left(3 - \sqrt{2}\right)/7 \approx 0.4531$，Foucart（2010年）之后又将其改进为 $\bar{\delta}_{2k} < 0.4652$。后续改进包括 Cai 等人（2009年）的结果，他们证明当 k 是 4 的倍数或 k 非常大时，$\delta_{2k} < 0.4721$便足够。同一批作者后来（在 Foucart（2010年）中）针对较大的 k，将这一结果进一步推进到 $\delta_{2k} < 0.4734$ 。最值得注意的是，Mo and Li（2011年）推导出了目前已知最优的一般上界，证明对于任意 k，只要 $\delta_{2k} < \left(77 - \sqrt{1337}\right)/82 \approx 0.\overline{4931}$，就能保证精确恢复。

现在我们转向分析一致有界噪声下的重构误差。下面的结果给出了问题（P_1，ϵ）的最坏情况误差界，并通过将充分条件从 $\delta_{2k} < 1/3$ 放宽为限制更少的$\delta_{2k} < \sqrt{2} - 1 \approx 0.4142$，改进了定理5.14。

定理 5.20：（$(\mathrm{P}_{1,\epsilon})$在限制等距性质（Candes，2008年）下的误差界）假设测量矩阵 $\boldsymbol{X}$ 满足 $2k$ 阶限制等距性质，且常数为$\delta_{2k} < \sqrt{2} - 1 \approx 0.4142$，并令 $\boldsymbol{y} = \bar{\boldsymbol{X}}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 其中 $\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$。那么，解 $\boldsymbol{\beta}$ 满足 ($\mathrm{P}_{1,\epsilon}$的误差界为

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 \leq C_0 \frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}} + C_2\epsilon, \tag{5.31}$$

其中

$$C_0 = 2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}} \qquad \text{and} \qquad C_2 = 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}.$$

证明 [定理 5.20] 设 $\boldsymbol{d} \triangleq \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$。我们的目标是求 $\|\boldsymbol{d}\|_2 \triangleq \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2$ 的上界。由于$\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$, $\boldsymbol{\beta}^* \in \mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{\alpha} \mid \|\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \leq \epsilon\}$ 是可行的，并且 $\boldsymbol{\beta}$ 是极小化解，因此有$\left\|\widehat{\boldsymbol{\beta}}\right\|_1 \leq \|\boldsymbol{\beta}^*\|_1$。于是我们可以利用引理 5.16，取 $\boldsymbol{\alpha} \triangleq \widehat{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta} \triangleq \boldsymbol{\beta}^*$，接下来只需给出 $|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}}、\ \boldsymbol{X}\boldsymbol{d}\rangle|$的上界，其中 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 是满足 $|\mathbb{S}| \leq k$的任意对象。为此，利用三角不等式，有

$$\|\boldsymbol{X}\boldsymbol{d}\|_2 = \left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\|_2 = \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\right\|_2 \leq \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y}\right\|_2 + \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\|_2 \leq 2\epsilon,$$

其中，最后一个不等式成立是因为 $\boldsymbol{\beta}, \boldsymbol{\beta}^* \in \mathcal{B}(\boldsymbol{y})$。结合 RIP 和柯西–施瓦茨不等式，可得

$$|\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}}, \boldsymbol{X}\boldsymbol{d}\rangle| \leq \|\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}}\|_2 \|\boldsymbol{X}\boldsymbol{d}\|_2 \leq 2\epsilon\sqrt{1+\delta_{2k}}\,\|\boldsymbol{d}_{\mathbb{S}}\|_2\,.$$

因此，

$$\|\boldsymbol{d}\|_2 \leq C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}} + C_1 2\epsilon\sqrt{1+\delta_{2k}} = C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}} + C_2\epsilon.$$

证明完毕。 ■

假设我们有一个预言机估计量，它能给出稀疏向量 $\boldsymbol{\beta}^*$ 的精确支撑集；也就说，它能给出对应于其非零项的 k 索引，记作 $\mathbb{S}_0$。在这种情况下，一种自然的重建策略是使用限制在已知支撑集上的伪逆：

$$\widehat{\boldsymbol{\beta}}_{\mathbb{S}_0} = \boldsymbol{X}_{\mathbb{S}_0}^{+}\boldsymbol{y} = (\boldsymbol{X}_{\mathbb{S}_0}^{\top}\boldsymbol{X}_{\mathbb{S}_0})^{-1}\boldsymbol{X}_{\mathbb{S}_0}^{\top}\boldsymbol{y}, \qquad \widehat{\boldsymbol{\beta}}_{\mathbb{S}_0^c} = \boldsymbol{0},$$

其中假设 $\boldsymbol{X}_{\mathbb{S}_0}$ 满列秩，因此方程 $\boldsymbol{y} = \boldsymbol{X}_{\mathbb{S}_0}\boldsymbol{\beta}_{\mathbb{S}_0}$ 有唯一解。在这一假设下，给定含噪测量 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，重构误差为

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 = \left\|(\boldsymbol{X}_{\mathbb{S}_0}^{\top}\boldsymbol{X}_{\mathbb{S}_0})^{-1}\boldsymbol{X}_{\mathbb{S}_0}^{\top}(\boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}) - \boldsymbol{\beta}^*\right\|_2 = \left\|(\boldsymbol{X}_{\mathbb{S}_0}^{\top}\boldsymbol{X}_{\mathbb{S}_0})^{-1}\boldsymbol{X}_{\mathbb{S}_0}^{\top}\boldsymbol{\epsilon}\right\|_2.$$

注意，$\boldsymbol{X}_{\mathbb{S}_0}$ 的奇异值位于范围 $[1/\sqrt{1+\delta_{2k}}, 1/\sqrt{1-\delta_{2k}}]$ 内（参见定义 4.30后面的讨论）。因此，如果我们考虑所有满足 $\boldsymbol{\epsilon}$ 的噪声向量 $\left\|\boldsymbol{X}^{\top}\boldsymbol{\epsilon}\right\|_{\infty} \leq \epsilon$上的最坏情况恢复误差，那么重构误差满足

$$\frac{\epsilon}{\sqrt{1+\delta_{2k}}} \leq \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 \leq \frac{\epsilon}{\sqrt{1-\delta_{2k}}}. \tag{5.32}$$

因此，即使完美掌握真实支撑集（即处于预言机设定下），最佳可能的恢复保证也无法将定理 5.20 中的上界改进超过一个常数因子，这凸显了 ℓ_1-最小化在限制等距性质下的近似最优性。

我们可以利用 RIP 与 ℓ_2-稳健 NSP 之间的联系，进一步改进恢复保证。具体而言，利用推论 5.9，我们得到如下改进结果。

定理 5.21：（($P_{1,\epsilon}$)**在 RIP 下的误差界（MoandLi，2011年）**）设测量矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶数为 $2k$ 、常数为

$$\delta_{2k} < \frac{77-\sqrt{1337}}{82} \approx 0.4931, \tag{5.33}$$

并令观测值为 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，其中 $\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$。那么，问题 ($P_{1,\epsilon}$) 的任意解 $\widehat{\boldsymbol{\beta}}$都满足以下误差界

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq C\sigma_k(\boldsymbol{\beta}^*)_1 + D\sqrt{k}\epsilon,$$

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_2 \leq \frac{C}{\sqrt{k}}\sigma_k(\boldsymbol{\beta}^*)_1 + D\epsilon,$$

其中常数 $C, D > 0$ 仅取决于 δ_{2k}。

证明 [定理 5.21]RIP 条件 (5.33) 意味着 $\boldsymbol{X}$ 满足 ℓ_2-阶稳健 NSP k (相对于 $\|\cdot\| = \|\cdot\|_2$)，根据定理 4.56。于是，所述误差界可直接由推论 5.9 得出。

■

事实上，通过插值论证，可立即得到相应的 ℓ_s误差估计值，适用于所有 $1 \leq s \leq 2$。

▶ 高斯噪声。我们还可以分析这些方法在存在高斯噪声时的性能（Haupt and Nowak, 2006年）。为简化讨论，我们只考虑真实信号 $\boldsymbol{\beta}^*$ 恰好是 k-稀疏的情形，即 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k] \triangleq \{\boldsymbol{\alpha} \mid \|\boldsymbol{\alpha}\|_0 \leq k\}$。在此情形下，最佳 k 项逼近误差为零： $\sigma_k(\boldsymbol{\beta}^*)_1 = 0$，因此定理 5.20 中的误差界仅取决于噪声向量 $\boldsymbol{\epsilon}$。假设 $\boldsymbol{\epsilon} \in \mathbb{R}^n$ 的元素独立同分布，服从均值为零、方差为σ^2的高斯分布，即 $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$。由于高斯分布 $\mathcal{N}(0, \sigma^2)$ 本身是次高斯分布（事后看来，更多细节见第7章），我们可以应用定理7.22等集中结果。具体而言，存在一个常数 $\kappa_0 > 0$ ，使得对于任意 $e > 0$,

$$\Pr\left(\|\boldsymbol{\epsilon}\|_2^2 \geq (1+e)n\sigma^2\right) \leq \exp\left(-\kappa_0 n e^2\right). \tag{5.34}$$

令 $e = 3$，可得

$$\Pr\left(\|\boldsymbol{\epsilon}\|_2 \geq 2\sqrt{n}\sigma\right) \leq \exp\left(-9\kappa_0 n\right). \tag{5.35}$$

由此可得以下关于高斯噪声这一特殊情形的推论。

推论5.22：（($P_{1,\epsilon}$) **在限制等距性质与高斯噪声下的误差界**）假设测量矩阵 $\boldsymbol{X}$ 满足阶数为 $2k$ 且常数为 $\delta_{2k} < \sqrt{2}-1$的限制等距性质。进一步假设 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k]$ ，且观测值由 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$给出，其中 $\boldsymbol{\epsilon}$的元素独立同分布于 $\mathcal{N}(0, \sigma^2)$。那么当 $\mathcal{B}(\boldsymbol{y})$ $= \{\boldsymbol{\alpha} \mid \|\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \leq 2\sqrt{n}\sigma\}$时，满足 $\epsilon = 2\sqrt{n}\sigma$ 的($P_{1,\epsilon}$) 的解$\widehat{\boldsymbol{\beta}}$ 满足

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 \leq 8\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}\sqrt{n}\sigma \tag{5.36}$$

其成立的概率至少为 $1-\exp(-9\kappa_0 n)$。

通过注意到 5.20 ，可直接由定理 $\sigma_k(\boldsymbol{\beta}^*)_1=0$ 当 $\boldsymbol{\beta}^*\in\mathbb{B}_0[k]$ 时得出证明。类似的结果也可用于定理 5.21，从而在高斯噪声下以高概率得到类似的 ℓ_1-误差界和 ℓ_2-误差界。

5.3.5 使用 Dantzig 选择器进行信号恢复

现在我们考虑一种略有不同的噪声模型。定理 5.20 和 5.21 假设噪声的 ℓ_2-范数 $\|\boldsymbol{\epsilon}\|_2$ 很小，而下面的定理则分析一种称为 *Dantzig* 选择器 的不同恢复算法的性能，其假设是$\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty$很小（Candes 和 Tao，2007年）；参见习题 3.17。正如我们将看到的，这一假设使得分析该算法在高斯噪声下的性能尤为简单。

定理 5.23：（RIP 条件下 Dantzig选择器的误差界）假设测量矩阵 $\boldsymbol{X}$ 满足 $2k$ 阶的 RIP，且 $\delta_{2k}<\sqrt{2}-1$，并且我们观测到如下形式的测量值 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$，其中 $\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty\le\epsilon$。那么，当$\mathcal{B}(\boldsymbol{y})=\{\boldsymbol{\alpha}\mid\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{y})\right\|_\infty\le\epsilon\}$时，解 $\widehat{\boldsymbol{\beta}}$ 满足（PS_1）（第 129页）

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\|_2\le C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}}+C_3\sqrt{k}\epsilon, \tag{5.37}$$

其中

$$C_0=2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}\qquad\text{and}\qquad C_3=\frac{4\sqrt{2}}{1-(1+\sqrt{2})\delta_{2k}}. \tag{5.38}$$

定理 [的证明 5.23] 证明与定理 5.20的证明非常相似。由于 $\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty\le\epsilon$，我们同样有 $\boldsymbol{\beta}^*\in\mathcal{B}(\boldsymbol{y})$，所以 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1\le\|\boldsymbol{\beta}^*\|_1$。因此，引理 5.16 适用。令$\boldsymbol{d}=\boldsymbol{\beta}-\boldsymbol{\beta}^*$。与前面的证明一样，我们希望给 $|\langle\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\ \boldsymbol{X}\boldsymbol{d}\rangle|$一个上界。首先，根据三角不等式，

$$\left\|\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\right\|_\infty\le\left\|\boldsymbol{X}^\top(\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{y})\right\|_\infty+\left\|\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*)\right\|_\infty\le 2\epsilon,$$

其中，最后一个不等式成立，是因为 $\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}^*$ 都属于 $\mathcal{B}(\boldsymbol{y})$。应用柯西–施瓦茨不等式可得

$$|\langle\boldsymbol{X}_{\mathbb{S}}\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}\boldsymbol{d}\rangle|=\left|\left\langle\boldsymbol{d}_{\mathbb{S}},\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}\boldsymbol{d}\right\rangle\right|\le\|\boldsymbol{d}_{\mathbb{S}}\|_2\left\|\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}\boldsymbol{d}\right\|_2.$$

最后，由于 $\left\|\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\right\|_\infty\le 2\epsilon$， $\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}$ 的每个元素的绝对值都不超过2ϵ。由于 $|\mathbb{S}|=|\mathbb{S}_0|+|\mathbb{S}_1|=2k$ 根据引理 5.16，我们得到$\left\|\boldsymbol{X}_{\mathbb{S}}^\top\boldsymbol{X}\boldsymbol{d}\right\|_2\le\sqrt{2k}(2\epsilon)$。结合这些估计可得

$$\|\boldsymbol{d}\|_2\le C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}}+C_1 2\sqrt{2k}\epsilon=C_0\frac{\sigma_k(\boldsymbol{\beta}^*)_1}{\sqrt{k}}+C_3\sqrt{k}\epsilon,$$

证明完成。 ■

▶ **Dantzig选择器的高斯噪声**。我们同样可以将定理5.23 应用于高斯噪声的情形。假设 $\boldsymbol{X}$ 的列均已归一化为单位 ℓ_2范数。那么， $\boldsymbol{X}^\top \boldsymbol{\epsilon}$ 的每个分量都是均值为零、方差为σ^2的高斯随机变量。利用标准高斯尾界（见引理1.84），可得

$$\Pr\left(\left|(\boldsymbol{X}^\top \boldsymbol{\epsilon})_i\right| \geq t\sigma\right) \leq \exp\left(-\frac{t^2}{2}\right), \quad \text{for } i = 1, 2, \ldots, p. \tag{5.39}$$

对所有坐标应用并集界（定理 p 坐标（定理 1.61），并令 $t = 2\sqrt{\ln p}$，可得

$$\Pr\left(\left\|\boldsymbol{X}^\top \boldsymbol{\epsilon}\right\|_\infty \geq 2\sqrt{\ln p}\sigma\right) \leq p\exp(-2\ln p) = \frac{1}{p}. \tag{5.40}$$

将此代入定理 5.23，并注意到当 $\sigma_k(\boldsymbol{\beta}^*)_1 = 0$ 时 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k]$，即可得到以下推论。

推论5.24：（限制等距性质与高斯噪声下 Dantzig选择器的误差界（Candes 和 Tao, 2007年））设测量矩阵 $\boldsymbol{X}$ 的列范数为单位范数，且满足阶数为 $2k$、参数为 $\delta_{2k} < \sqrt{2} - 1$的限制等距性质。此外，假设$\boldsymbol{\beta}^* \in \mathbb{B}_0[k]$，测量值由 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$给出，其中 $\boldsymbol{\epsilon}$的元素独立同分布于 $\mathcal{N}(0, \sigma^2)$。那么，当 $\mathcal{B}(\boldsymbol{y}) = \{\boldsymbol{\alpha} \mid \left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{y})\right\|_\infty \leq 2\sqrt{\ln p}\sigma\}$时，$(\mathrm{PS}_1)$的解$\widehat{\boldsymbol{\beta}}$ 满足

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 \leq \frac{8\sqrt{2}}{1 - (1+\sqrt{2})\delta_{2k}}\sqrt{k\ln p}\sigma, \tag{5.41}$$

以至少 $1 - \frac{1}{p}$ 的概率。

忽略精确的常数和确切的成功概率（我们尚未对其进行优化），我们注意到，如果 $n = \mathcal{O}(k\ln p)$，那么推论 5.22 和 5.24 给出的误差界基本上具有可比性。然而，这里存在一个微妙但重要的区别：对于固定的n 和 p，推论5.24中的上界会根据稀疏度 k进行调整，当 k 较小时提供更紧的保证。相比之下，推论5.22 中的上界不会随着k 的减小而改善。因此，尽管两种方法的总体性能相近，但在真实稀疏度 k 显著小于环境维数的场景中，Dantzig 选择器可能更为理想。有关这些方法相对优点的更详细比较，请参见 Candes 和 Tao （2007年）。

5.4 稀疏线性回归

在许多现代数据问题中——从基因组学到金融——我们会遇到这样的数据集：特征数量 p 远远超过观测值数量 n。在这种高维场景下，经典线性回归会失效：它变得不稳定，容易过拟合，并且通常会产生难以解释甚至无法解释的模型。稀疏线性回归通过寻求仅有少量系数非零的解来应对这些挑战，在拟合模型的同时有效地执行特征选择。

从本质上说，稀疏线性回归建立在我们熟悉的线性模型之上

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

其中 $\boldsymbol{y} \in \mathbb{R}^n$ 是响应向量， $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 包含预测变量， $\boldsymbol{\beta} \in \mathbb{R}^p$ 是我们希望估计的系数向量，而 $\boldsymbol{\epsilon}$ 表示噪声。在

稀疏情形下，关键假设是 $\boldsymbol{\beta}$ 的大多数元素都恰好为零——也就是说，真正影响结果的预测变量只有少数几个。这一假设不仅反映了许多领域中的科学现实（例如，可能只有少数几个基因会影响某种疾病），还能够带来泛化能力更强、也更容易解释的模型。

实现稀疏性最广泛使用的方法是 LASSO（最小绝对值收缩与选择算子），由 Robert Tibshirani 于 1996年提出（Tibshirani，1996）。LASSO估计量求解以下优化问题

$$\widehat{\boldsymbol{\beta}}_{\text{lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\},$$

其中，$\lambda \geq 0$是控制稀疏程度的调节参数。ℓ_1 惩罚$\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$ 促使许多系数恰好为零——这是岭回归所用的 ℓ_2 惩罚不具备的性质。随着 λ 增大，更多系数被收缩为零，从而得到越来越稀疏的模型。在实践中，通常通过交叉验证来选择 λ，以平衡预测准确度与模型简洁性。

尽管 LASSO 计算效率高，且在许多情形下表现良好，但它也有局限性。例如，当预测变量高度相关时，LASSO 往往只从一组变量中选出一个，可能会遗漏其他相关变量。它还可能因过度收缩较大的系数而引入偏差。弹性网等替代方法——结合了 ℓ_1 和 ℓ_2 惩罚——有助于缓解这些问题，尤其适用于特征成组或高度相关的情况。另一种极端方法是最佳子集选择，它直接在非零系数数量（即 $\|\boldsymbol{\beta}\|_0 \leq k$）受约束的条件下最小化预测误差。尽管这种方法在统计学上颇具吸引力，但其组合复杂度高，对于大型 p通常不可行；不过，优化领域近期的进展重新引起了人们对它的兴趣。

稀疏回归的优势不止体现在计算方面。在全基因组关联研究等应用中，研究人员可能会测量数万个基因的表达水平，但预计其中只有几十个与生物学相关。稀疏模型不仅能提高预测性能，还能找出候选基因，以便进一步通过实验验证。同样，在信号处理中，稀疏表示是压缩感知等技术的基础；只要信号在某个基下是稀疏的，就能用远少于传统方法所需的测量值准确重构信号。

从理论角度来看，即使在 $p \gg n$ 的情况下，只要设计矩阵 $\boldsymbol{X}$ 满足某些条件，也有可能一致地恢复真实支撑集（即非零系数的集合）。其中值得注意的是受限特征值条件，它的基本要求是相关特征与无关特征之间的相关性不能太强。在这些假设下，随着样本量增大，LASSO 等方法能够以高概率证明可恢复正确的稀疏结构。

当然，稀疏性也伴随着权衡。施加稀疏结构会给系数估计引入偏差，但通常可通过显著降低方差来弥补，从而降低整体预测误差。此外，虽然 LASSO 等凸形式能够很好地扩展到大规模问题，但非凸惩罚（例如 SCAD 或 MCP (Kim 等人，2008年；Zhang，2010年）能够提供更优的统计性质，代价是计算复杂度更高，并且存在收敛到局部极小值的风险。

总而言之，稀疏线性回归秉持简约原则——最简单的解释往往是最好的——将一个病态的高维问题转化为可处理的问题。本节主要聚焦于 LASSO 的理论基础，尤其是它能够一致地恢复真实稀疏模型所需满足的条件。值得注意的是，LASSO 与压缩感知和信号处理中的稀疏恢复这一更广泛的框架密切相关。因此，LASSO 可以视为

基于 ℓ_1 的稀疏恢复在统计学中的具体实现，适用于数据含噪的情形；其主要目标是预测和一致的支撑集估计，而非精确重建。通过自动识别并仅保留信息量最大的特征，它构建出的模型不仅准确，而且易于解释——在大数据时代，这种组合既难得又宝贵。无论是在生物学、经济学、工程还是机器学习领域，稀疏回归始终是将纷繁复杂转化为可付诸行动的洞见的基础工具。

5.4.1 LASSO 与稳健线性回归

回归分析是许多领域中的基础问题，包括统计学、监督式机器学习和优化。最小化平方误差通常会得到对数据中的微小变化高度敏感的解。为减轻这种敏感性，人们提出了各种正则化方法。其中最知名且被广泛引用的方法包括蒂霍诺夫正则化（ℓ_2 惩罚项）（Tikhonov，1963年）和 LASSO（Tibshirani，1996年）。

为了降低直接求解优化问题（P_1的计算复杂度，考虑线性回归情形：给定一个观测响应向量 $\boldsymbol{y} \in \mathbb{R}^n$ 和一个设计矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，目标是找到一个系数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 使得

$$\widehat{y}_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \quad i = 1, 2, \ldots, n, \tag{5.42}$$

或者等价地，$\widehat{\boldsymbol{y}} = \sum_{i=1}^p \beta_i \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{\beta}$。

Tibshirani(1996年) 为线性回归提出了 LASSO（最小绝对收缩与选择算子）。LASSO 的核心思想是在系数向量的 Σ 范数 ℓ_1上界约束下，最小化预测误差平方和，即模型（L_C）（第 126 页）：

$$(\mathrm{L}_C) \qquad \min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \qquad \text{s.t.} \quad \|\boldsymbol{\beta}\|_1 = \sum_{i=1}^p |\beta_i| \le \Sigma. \tag{5.43}$$

上界 Σ 是一个调节参数。当 Σ 足够大时，约束不起作用，解就退化为普通最小二乘回归。然而，当 Σ较小时，某些系数会恰好变为零，从而得到 $\boldsymbol{\beta}$的稀疏估计。因此，选择适当的 Σ即可实现自动变量选择。

LASSO 的一种等价表述采用其拉格朗日形式，即模型（L_L）（第 126 页）：

$$(\mathrm{L}_L) \qquad \min_{\boldsymbol{\beta}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{5.44}$$

其中 $\lambda \ge 0$ 控制正则化的强度，并且与 Σ 有关（参见推论 3.3）。

LASSO 将 ℓ_1 范数正则化与线性回归结合起来，并具有两个关键性质：

（i）收缩效应。它将估计系数向零收缩，从而降低模型复杂度。（ii）选择效应。它会自动将某些系数精确设为零，从而执行特征选择并得到稀疏解。

因此，LASSO 既是一种收缩方法，也是一种变量选择方法。它与小波系数的软阈值处理和前向逐步回归密切相关；参见第 9.2.4 节。

在实际应用中，观测到的设计矩阵 $\boldsymbol{X}$ 可能受到噪声甚至对抗性扰动的影响。为确保在面对这类不确定性时的鲁棒性，鲁棒线性回归 通过求解以下极小极大问题来考虑最坏情况：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\max_{\Delta\boldsymbol{X}\in\mathbb{U}}\|\boldsymbol{y}-(\boldsymbol{X}+\Delta\boldsymbol{X})\boldsymbol{\beta}\|_2\,, \tag{5.45}$$

其中，$\Delta\boldsymbol{X}=[\boldsymbol{\delta}_1,\boldsymbol{\delta}_2,\ldots,\boldsymbol{\delta}_p]$ 表示施加于 $\boldsymbol{X}$的每一列（特征）的扰动，并且

$$\mathbb{U}\triangleq\left\{[\boldsymbol{\delta}_1,\boldsymbol{\delta}_2,\ldots,\boldsymbol{\delta}_p]\mid\|\boldsymbol{\delta}_i\|_2\le c_i,\ i=1,2,\ldots,p\right\} \tag{5.46}$$

称为不确定集，也就是矩阵 $\boldsymbol{X}$的允许扰动集合。该不确定集在特征间非耦合，这意味着不同特征上的扰动分别受到独立约束——不同于在各特征之间施加联合约束的耦合不确定集。

值得注意的是，当不确定集在每个特征上都是非耦合且有界的 ℓ_2范数集合时，鲁棒回归问题可以转化为一个易于处理的等价形式，如下述定理所述。

定理 5.25：（鲁棒回归（Xu 等人，2008年））具有不确定集 $\mathbb{U}$（5.46）所给出的鲁棒回归问题（5.45）等价于以下 ℓ_1正则化回归问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2+\sum_{i=1}^{p}c_i\,|\beta_i|\right\}. \tag{5.47}$$

定理 [的证明 5.25] 固定 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$。我们证明

$$\max_{\Delta\boldsymbol{X}\in\mathbb{U}}\|\boldsymbol{y}-(\boldsymbol{X}+\Delta\boldsymbol{X})\boldsymbol{\beta}^*\|_2=\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2+\sum_{i=1}^{p}c_i\,|\beta_i^*|\,.$$

根据三角不等式，左边可写为

$$\begin{aligned}
&\max_{\Delta\boldsymbol{X}\in\mathbb{U}}\left\|\boldsymbol{y}-(\boldsymbol{X}+\Delta\boldsymbol{X})\boldsymbol{\beta}^*\right\|_2\\
&=\max_{[\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_p]:\|\boldsymbol{\delta}_i\|_2\le c_i}\left\|\boldsymbol{y}-\left(\boldsymbol{X}+[\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_p]\right)\boldsymbol{\beta}^*\right\|_2=\max_{[\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_p]:\|\boldsymbol{\delta}_i\|_2\le c_i}\left\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*-\sum_{i=1}^{p}\beta_i^*\boldsymbol{\delta}_i\right\|_2\\
&\le\max_{[\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_p]:\|\boldsymbol{\delta}_i\|_2\le c_i}\left\{\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2+\sum_{i=1}^{p}\|\beta_i^*\boldsymbol{\delta}_i\|_2\right\}\le\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2+\sum_{i=1}^{p}|\beta_i^*|\,c_i.
\end{aligned} \tag{5.48}$$

为证明等式成立，定义单位向量 $\boldsymbol{v}$ 如下

$$\boldsymbol{v}\triangleq\begin{cases}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*)/\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2, & \text{if } \boldsymbol{X}\boldsymbol{\beta}^*\neq\boldsymbol{y};\\ \text{any unit norm vector}, & \text{otherwise},\end{cases}\qquad\text{and}\qquad\boldsymbol{\delta}_i^*\triangleq-c_i\text{sign}(\beta_i^*)\boldsymbol{v},\ \forall\, i.$$

由于 $\|\boldsymbol{\delta}_i^*\|_2\le c_i$，可知 $\Delta\boldsymbol{X}^*\triangleq[\boldsymbol{\delta}_1^*,\boldsymbol{\delta}_2^*,\ldots,\boldsymbol{\delta}_p^*]\in\mathbb{U}$，由此得到

$$\begin{aligned}
&\max_{\Delta\boldsymbol{X}\in\mathbb{U}}\|\boldsymbol{y}-(\boldsymbol{X}+\Delta\boldsymbol{X})\boldsymbol{\beta}^*\|_2\ge\|\boldsymbol{y}-(\boldsymbol{X}+\Delta\boldsymbol{X}^*)\boldsymbol{\beta}^*\|_2\\
&=\left\|\boldsymbol{y}-\left(\boldsymbol{X}+[\boldsymbol{\delta}_1^*,\boldsymbol{\delta}_2^*,\ldots,\boldsymbol{\delta}_p^*]\right)\boldsymbol{\beta}^*\right\|_2=\left\|\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\right)-\sum_{i=1}^{p}\left(-\beta_i^*c_i\text{sign}(\beta_i^*)\boldsymbol{v}\right)\right\|_2\\
&=\left\|\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\right)+\left(\sum_{i=1}^{p}c_i\,|\beta_i^*|\right)\boldsymbol{v}\right\|_2=\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2+\sum_{i=1}^{p}c_i\,|\beta_i^*|\,.
\end{aligned} \tag{5.49}$$

结合（5.48）和（5.49），对于任意固定的 $\boldsymbol{\beta}^*$，即可得到所需的等式

$$\max_{\Delta \boldsymbol{X} \in \mathbb{U}} \|\boldsymbol{y} - (\boldsymbol{X} + \Delta \boldsymbol{X})\boldsymbol{\beta}^*\|_2 = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\|_2 + \sum_{i=1}^{p} c_i \, |\beta_i^*| \, .$$

对两边关于 $\boldsymbol{\beta}$ 取最小值，即完成证明。 ■

当将 ℓ_1范数替换为任意范数 $\|\cdot\|_s$ 时，前述结果可推广如下。

推论 5.26：（稳健回归）设

$$\mathbb{U}_a \triangleq \left\{ [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_p] \mid \|\boldsymbol{\delta}_i\|_s \leq c_i, \; i = 1, 2, \ldots, p \right\}. \tag{5.50}$$

那么，以下稳健回归问题

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \max_{\Delta \boldsymbol{X} \in \mathbb{U}_a} \|\boldsymbol{y} - (\boldsymbol{X} + \Delta \boldsymbol{X})\boldsymbol{\beta}\|_s \, ,$$

带有不确定集 $\mathbb{U}_a$，等价于以下任意范数 $\|\cdot\|_s$ 正则化回归问题：

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_s + \sum_{i=1}^{p} c_i \, |\beta_i| \right\}. \tag{5.51}$$

因此，稳健回归得到的目标函数将 ℓ_2（或更一般地说，ℓ_s）损失与加权的 ℓ_1 惩罚相结合。下面的结果表明，对于 ℓ_2 损失，这种形式基本等价于 LASSO 中使用的标准正则化最小二乘模型。

定理 5.27：（等价形式（Xu 等人，2008年））以下优化问题

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \lambda_1 \|\boldsymbol{\beta}\|_1\} \qquad \text{and} \qquad \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_1\} \tag{5.52}$$

等价，即对于每个 $\lambda_1 > 0$，都存在一个相应的$\lambda_2 > 0$（反之亦然），使得两个问题具有相同的极小化点集合。

证明 参见 Xu 等人（2008年）的附录。 ■

现在，假设我们令 $c_i = \lambda$ 对所有 i 均成立，并假设 $\boldsymbol{x}_i$ 的各列 $\boldsymbol{X}$ 正交归一。在这些条件下，鲁棒回归问题（5.47）可简化为

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \lambda \|\boldsymbol{\beta}\|_1\} \, . \tag{5.53}$$

根据定理 5.27，该问题等价于拉格朗日 LASSO 形式（5.44）。换句话说，具有逐特征 ℓ_2-有界扰动且采用正交归一设计的稳健线性回归问题，等价于 LASSO 问题。

将 LASSO 解视为鲁棒回归问题的解具有两个重要含义：

(i) 正则化的可解释性。鲁棒性为 ℓ_1 正则化项提供了具体的解释——它对应于对抗最坏情况下的数据扰动，将数学惩罚项与一种切实的物理属性（抗噪性）联系起来。

(ii) 解析性杠杆。更重要的是，鲁棒性是一种强大的结构性质，可用于分析解的行为。例如，LASSO 估计量的稀疏性可以理解为其鲁棒性的结果：在对抗性扰动下，最优策略是只依赖最稳定（即非零）的特征。此外，鲁棒解按其构造在最坏情况场景下是最优的，这为不确定或含噪环境提供了有力的保证。

5.4.2 LASSO误差界

我们通过介绍 LASSO 问题的若干理论结果来结束本节。$\widehat{}$给定一个 LASSO 估计值 $\boldsymbol{\beta} \in \mathbb{R}^p$，其质量可根据应用目标，通过多种方式进行评估。在某些情形下，主要关注的是 $\boldsymbol{\beta}$的预测$\widehat{}$性能。在这种情况下，我们使用（样本内）预测损失来评估它：

$$\mathcal{L}_{\text{pred}}(\widehat{\boldsymbol{\beta}}; \boldsymbol{\beta}^*) = \frac{1}{n} \left\| \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^* \right\|_2^2, \tag{5.54a}$$

它衡量的是观测设计矩阵 $\boldsymbol{X}$上的均方预测误差。在其他应用中——例如医学成像、遥感和压缩感知——人们主要关注准确恢复未知参数向量 $\boldsymbol{\beta}^*$ 本身。在这种情况下，更适合考虑参数估计损失，其定义为估计量与真实参数之间的平方ℓ_2距离：

$$\mathcal{L}_2(\widehat{\boldsymbol{\beta}}; \boldsymbol{\beta}^*) = \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \right\|_2^2. \tag{5.54b}$$

最后，当目标是变量选择或支撑集恢复时，可以使用一种离散损失，用来检验估计向量与真实向量的带符号支撑集是否相同：

$$\mathcal{L}_{\text{vs}}(\widehat{\boldsymbol{\beta}}; \boldsymbol{\beta}^*) = \begin{cases} 0, & \text{if } \operatorname{sign}(\widehat{\beta}_i) = \operatorname{sign}(\beta_i^*) \text{ for all } i \in \{1, 2, \ldots, p\}; \\ 1, & \text{otherwise.} \end{cases} \tag{5.54c}$$

$\widehat{}$只有当 $\boldsymbol{\beta}$ 正确识别出 $\boldsymbol{\beta}^*$中非零项的位置和符号时，这种损失才等于零。

为简洁起见，本小节着重讨论前两种损失函数——预测损失和参数估计损失的理论保证。为建立这些结果，我们需要以下技术引理。

引理 5.28 设 $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\|\boldsymbol{\beta}\|_1 \le \|\boldsymbol{\alpha}\|_1$，其中 $\operatorname{supp}(\boldsymbol{\alpha}) \triangleq \mathbb{S} \subseteq \{1, 2, \ldots, p\}$。记 $\boldsymbol{e} \triangleq \boldsymbol{\beta} - \boldsymbol{\alpha}$。则 $\|\boldsymbol{e}_{\mathbb{S}^c}\|_1 \le \|\boldsymbol{e}_{\mathbb{S}}\|_1$，即 $\boldsymbol{e} \in \mathbb{C}[\mathbb{S};\ 1]$（定义见（4.2））， 其中$\mathbb{S}^c$ 表示 $\mathbb{S}$的补集。

证明 [引理 5。28] 令 $\mathbb{T} \triangleq$ 支撑集$(\boldsymbol{\beta})$。假设 $\|\boldsymbol{\beta}\|_1 \le \|\boldsymbol{\alpha}\|_1$ 表明

$$\sum_{i \in \mathbb{T}} |\beta_i| \le \sum_{i \in \mathbb{S}} |\alpha_i| \quad \Longrightarrow \quad \sum_{i \in \mathbb{T} \setminus \mathbb{S}} |\beta_i| + \sum_{i \in \mathbb{S} \cap \mathbb{T}} |\beta_i| \le \sum_{i \in \mathbb{S} \setminus \mathbb{T}} |\alpha_i| + \sum_{i \in \mathbb{S} \cap \mathbb{T}} |\alpha_i|$$

$$\Longrightarrow \quad \sum_{i \in \mathbb{T} \setminus \mathbb{S}} |\beta_i| \le \sum_{i \in \mathbb{S} \setminus \mathbb{T}} |\alpha_i| + \sum_{i \in \mathbb{S} \cap \mathbb{T}} |\alpha_i - \beta_i|,$$

从而得到所需结果。 ■

有了这一认识，我们现在给出一个结果，该结果为线性观测模型 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$ 下的参数估计误差 $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\|_2$ 提供了一个上界，其中 $\boldsymbol{\beta}^*$ 是 k-稀疏的，且支撑于子集 $\mathbb{S}\subseteq\{1,2,\ldots,p\}$（即 $|\mathbb{S}|=k$）。

定理 5.29：（LASSO 优化器误差的界）假设观测值服从线性模型 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$，其中 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$，且 $\boldsymbol{\beta}^*$ 是 k-稀疏的，支撑集为$\mathbb{S}\subseteq\{1,2,\ldots,p\}$（即 $|\mathbb{S}|=k$）。假设设计矩阵 $\boldsymbol{X}$ 满足受限特征值（RE）性质（定义4.3），其常数为 $n\mu>0$，定义于 $\mathbb{C}[\mathbb{S};3]$。[a] 则以下误差界成立：

(i) 约束 *LASSO*。任何基于 $\boldsymbol{\beta}$ 约束 LASSO（L_C）的估计值$\Sigma=\|\boldsymbol{\beta}^*\|_1$ 都满足上界

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\|_2\leq\frac{4\sqrt{k}}{\mu n}\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty. \tag{5.55}$$

(ii) 拉格朗日 *LASSO*。给定正则化参数 $\lambda\geq2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty>0$ [b]，任意估计值 $\boldsymbol{\beta}$ 由正则化 LASSO（L_L）得到的估计值都满足上界

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\|_2\leq\frac{3}{\mu}\sqrt{k}\frac{\lambda}{n}. \tag{5.56}$$

[a]。注意，第一个结果只要求受限特征值性质在子集 $\mathbb{C}[\mathbb{S}$上成立；1]。[b]。如果在（L′L 126, $\lambda\geq$）中使用缩放拉格朗日 LASSO 损失，$\frac{2}{n}\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\|_\infty>0$ 则要求变为。

证明 [定理 5。29]**(i) 约束 LASSO**。由于 $\boldsymbol{\beta}^*$ 是可行的，且 $\boldsymbol{\beta}$ 是最优的，因此有不等式 $\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|_2^2\leq\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^*\|_2^2$。定义误差向量 $\widehat{\boldsymbol{e}}\triangleq\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$，代入关系式 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$ 可得

$$(\text{Constrained LASSO}):\qquad\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2\leq2\boldsymbol{\epsilon}^\top\boldsymbol{X}\widehat{\boldsymbol{e}}. \tag{5.57}$$

应用霍尔德不等式（定理 1.60）可得右侧的上界 $2\left|\boldsymbol{\epsilon}^\top\boldsymbol{X}\widehat{\boldsymbol{e}}\right|\leq2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty\|\widehat{\boldsymbol{e}}\|_1$。不等式 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1\leq\Sigma=\|\boldsymbol{\beta}^*\|_1$ 意味着 $\widehat{\boldsymbol{e}}\in\mathbb{C}[\mathbb{S};1]$（引理 5.28）：

$$\left\|\widehat{\boldsymbol{\beta}}\right\|_1\leq\|\boldsymbol{\beta}^*\|_1\quad\Longrightarrow\quad\|\widehat{\boldsymbol{e}}\|_1=\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1+\|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1\leq2\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1\overset{\dagger}{\leq}2\sqrt{k}\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_2\leq2\sqrt{k}\|\widehat{\boldsymbol{e}}\|_2, \tag{5.58}$$

其中，不等式（†）由柯西–施瓦茨不等式（方程（1.38））可得：

$$\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1=|\widehat{\boldsymbol{e}}_{\mathbb{S}}|^\top\mathbf{1}\leq\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_2\|\mathbf{1}\|_2=\sqrt{k}\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_2\leq\sqrt{k}\|\widehat{\boldsymbol{e}}\|_2. \tag{5.59}$$

注意，（5.58）在条件 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1\leq\|\boldsymbol{\beta}^*\|_1$下成立，而（5.59）对任意 $\mathbb{S}$和 $\widehat{\boldsymbol{e}}$均成立。另一方面，将受限特征值性质应用于不等式（5.57）的左侧可得 $n\mu\|\widehat{\boldsymbol{e}}\|_2^2\leq\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2$。结合这些不等式即可得到所需结果。

（ii）拉格朗日 LASSO。定义函数

$$G(\boldsymbol{e})\triangleq\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}(\boldsymbol{\beta}^*+\boldsymbol{e})\|_2^2+\lambda\|\boldsymbol{\beta}^*+\boldsymbol{e}\|_1. \tag{5.60}$$

注意到，根据构造，$\widehat{\boldsymbol{e}} \triangleq \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ 使 G 最小，因此我们有 $G(\widehat{\boldsymbol{e}}) \leq G(\mathbf{0})$。代入关系 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 得

$$\text{(Lagrangian LASSO)}: \qquad \|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \leq 2\boldsymbol{\epsilon}^\top \boldsymbol{X}\widehat{\boldsymbol{e}} + 2\lambda\left\{\|\boldsymbol{\beta}^*\|_1 - \|\boldsymbol{\beta}^* + \widehat{\boldsymbol{e}}\|_1\right\}. \tag{5.61}$$

由于 $\boldsymbol{\beta}^*_{\mathbb{S}^c} = \mathbf{0}$，我们有 $\|\boldsymbol{\beta}^*\|_1 = \|\boldsymbol{\beta}^*_{\mathbb{S}}\|_1$，并且

$$\|\boldsymbol{\beta}^* + \widehat{\boldsymbol{e}}\|_1 = \|\boldsymbol{\beta}^*_{\mathbb{S}} + \widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 + \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1 \geq \|\boldsymbol{\beta}^*_{\mathbb{S}}\|_1 - \|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 + \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1.$$

将这些关系式代入不等式（5.61）并应用 H¨older 不等式，可得

$$\begin{aligned}\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 &\leq 2\boldsymbol{\epsilon}^\top \boldsymbol{X}\widehat{\boldsymbol{e}} + 2\lambda\left\{\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 - \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1\right\} \\ &\leq 2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty \|\widehat{\boldsymbol{e}}\|_1 + 2\lambda\left\{\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 - \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1\right\}.\end{aligned} \tag{5.62}$$

由于 $2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty \leq \lambda$ 根据假设，利用 $\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 \leq \sqrt{k}\,\|\widehat{\boldsymbol{e}}\|_2$ 由（5.59）这一事实，我们有：

$$0 \leq \|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \leq \lambda\left\{\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 + \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1\right\} + 2\lambda\left\{\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 - \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1\right\} \leq 3\sqrt{k}\lambda\,\|\widehat{\boldsymbol{e}}\|_2. \tag{5.63}$$

不等式（5.63）还意味着存在 $\|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1 \leq 3\,\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1$ 使得 $\widehat{\boldsymbol{e}} \in \mathbb{C}[\mathbb{S};\ 3]$（见（4.2）中的定义）。应用 μ-RE 条件于 $\widehat{\boldsymbol{e}}$ 可确保 $\mu\,\|\widehat{\boldsymbol{e}}\|_2^2 \leq \frac{1}{n}\,\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2$。将这个下界与不等式（5.63）结合可得 $\mu\,\|\widehat{\boldsymbol{e}}\|_2^2 \leq 3\sqrt{k}\frac{\lambda}{n}\,\|\widehat{\boldsymbol{e}}\|_2$，这就意味着所需结果。 ■

定理 5.30：(LASSO 预测误差的界) 假设观测值服从线性模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。考虑正则化参数为 $\lambda \geq 2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty$ [a] 的拉格朗日 LASSO（L_L）。

(i) 若 $\|\boldsymbol{\beta}^*\|_1 < R_1$，则任意最优解 $\widehat{\boldsymbol{\beta}}$ 均满足

$$\left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\|_2^2 \leq 12\,R_1\,\lambda.\ ^{b} \tag{5.64}$$

(ii) 进一步假设 $\boldsymbol{\beta}^*$ 是 k-稀疏的，且支撑于子集 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$（即 $|\mathbb{S}| = k$）。若模型矩阵 $\boldsymbol{X}$ 在 $\mathbb{C}[\mathbb{S}$ 上满足受限特征值性质，其参数为 $n\mu > 0$；$3]$，则任意最优解 $\widehat{\boldsymbol{\beta}}$ 均满足

$$\left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\|_2^2 \leq 9\frac{k\lambda^2}{n\mu}. \tag{5.65}$$

[a] 同样，若在（L_L'）中使用缩放拉格朗日 LASSO 损失，则要求变为 $\lambda \geq \frac{2}{n}\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty > 0$。[b] 更精细的上界见问题 5.12。

证明 [定理 5.30] 该证明稍作改编，基于 Hastie 等人 (2015年); Lu(2025年).

(i). 再次，我们定义误差向量 $\widehat{\boldsymbol{e}} \triangleq \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$。根据不等式（5.61），可得

$$0 \leq \|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \leq 2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty \|\widehat{\boldsymbol{e}}\|_1 + 2\lambda\left\{\,\|\boldsymbol{\beta}^*\|_1 - \|\boldsymbol{\beta}^* + \widehat{\boldsymbol{e}}\|_1\,\right\} \tag{5.66}$$

$$\overset{\dagger}{\leq} 2\left\{\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty - \lambda\right\}\|\widehat{\boldsymbol{e}}\|_1 + 4\lambda\,\|\boldsymbol{\beta}^*\|_1 \overset{\ddagger}{\leq} \lambda\Big\{-\|\widehat{\boldsymbol{e}}\|_1 + 4\,\|\boldsymbol{\beta}^*\|_1\Big\}, \tag{5.67}$$

其中，不等式（†）由三角不等式 $-(\|\boldsymbol{\beta}^*\|_1-\|\widehat{\boldsymbol{e}}\|_1) \le \|\boldsymbol{\beta}^* - (-\widehat{\boldsymbol{e}})\|_1$ [2]得出，而不等式（‡）则由假设 $\lambda \ge 2\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty$得出。这表明$\|\widehat{\boldsymbol{e}}\|_1 \le 4\|\boldsymbol{\beta}^*\|_1 \le 4R_1$。结合这些不等式，并在（$\|\boldsymbol{\beta}^*\|_1 - \|\boldsymbol{\beta}^*+\widehat{\boldsymbol{e}}\|_1 \le \|\boldsymbol{\beta}^* - (\boldsymbol{\beta}^*+\widehat{\boldsymbol{e}})\|_1 = \|\widehat{\boldsymbol{e}}\|_1$）中应用三角不等式5.66，可得

$$\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \le 2\left\{\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty + \lambda\right\}\|\widehat{\boldsymbol{e}}\|_1 \le 12\lambda R_1, \tag{5.68}$$

由此得到了所需结果。

(ii). 根据式（5.63），我们有$\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \le 3\sqrt{k}\lambda\|\widehat{\boldsymbol{e}}\|_2$. [3] 根据定理 5.29的证明，误差向量 $\widehat{\boldsymbol{e}}$ 属于锥 $\mathbb{C}[\mathbb{S};\ 3]$, 因此， μ-RE 条件保证$\|\widehat{\boldsymbol{e}}\|_2^2 \le \frac{1}{n\mu}\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2$。结合这些不等式即可得到所需结果。 —

备注 5.31（兼容性条件，ℓ_1 估计上界）。另一个相关条件是设计矩阵 $\boldsymbol{X}$ 上的兼容性条件。该条件相对于真实支撑集 $\mathbb{S}=\text{supp}(\boldsymbol{\beta}^*)$ 定义，并要求存在某个兼容性常数$\mu>0$, 使得

$$\frac{1}{n}\|\boldsymbol{X}\boldsymbol{u}\|_2^2 \ge \frac{\mu}{k}\|\boldsymbol{u}_{\mathbb{S}}\|_1^2, \quad \text{for all } \boldsymbol{u}\in\mathbb{C}[\mathbb{S};3],\ \text{i.e.,}\ \|\boldsymbol{u}_{\mathbb{S}^c}\|_1 \le 3\|\boldsymbol{u}_{\mathbb{S}}\|_1, \tag{5.69}$$

其中 $k \triangleq \|\boldsymbol{\beta}^*\|_0$。直观而言，该条件确保真正活跃的预测变量之间的相关性不会过高。注意，不等式（5.63）还意味着

$$\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^*\right\|_2^2 \le 3\lambda\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1. \tag{5.70}$$

由于 $\widehat{\boldsymbol{e}} = \boldsymbol{\beta} - \boldsymbol{\beta}^* \in \mathbb{C}[\mathbb{S};\ 3]$ 根据（5.63），我们可以将其与兼容性条件结合起来，以界定参数估计误差。具体而言，将（5.69）中的下界应用于（5.70）的左侧可得

$$\frac{n\mu}{k}\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1^2 \le \left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^*\right\|_2^2 \le 3\lambda\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1.$$

将两边除以 $\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1$，并利用 $\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}^c}\right\|_1 \le 3\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1$ 这一事实，可知 $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 \le 4\left\|\widehat{\boldsymbol{\beta}}_{\mathbb{S}} - \boldsymbol{\beta}^*_{\mathbb{S}}\right\|_1$，于是得到

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 \le \frac{k\lambda}{n\mu}. \tag{5.71}$$

这为参数估计损失提供了一个 ℓ_1上界。与（ℓ_1）中的5.56上界相比，现在稀疏度对 k 的依赖是线性的，而不是 $\sqrt{k}$。

5.4.3 高斯噪声与预言机不等式

我们考虑线性模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，其中噪声向量 $\boldsymbol{\epsilon}\in\mathbb{R}^n$ 的元素是均值为零、方差为 σ^2 的独立同分布高斯分布。假设设计矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 固定，且每一列满足 $\max_{i=1,\ 2,\ldots,p}\|\boldsymbol{x}_i\|_2 \le \sqrt{n}$。（注意，这种归一化总是可以通过

2. $|\|\boldsymbol{\beta}\| - \|\boldsymbol{\alpha}\|| \le \|\boldsymbol{\beta}-\boldsymbol{\alpha}\|$ 适用于任意向量范数。3. 注意，Hastie 等人 (2015年) 使用（5.68）以及这一事实 $\|\widehat{\boldsymbol{e}}\|_1 = \|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 + \|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1 \le 4\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1 \le 4\sqrt{k}\|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_2 \le 4\sqrt{k}\|\widehat{\boldsymbol{e}}\|_2$ 来得到 $\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \le 12\sqrt{k}\|\widehat{\boldsymbol{e}}\|_2$，由此得到一个较松的上界。

缩放实现。）在这些假设下，随机向量 $\boldsymbol{X}^\top\boldsymbol{\epsilon}$ 的元素相互独立，服从均值为零、方差至多为 $\max_{i=1,\ 2,\ldots,p}\|\boldsymbol{x}_i\|_2^2\sigma^2 \le n\sigma^2$ 的高斯分布。利用标准高斯尾界（见引理 1.84），我们有

$$\Pr\left(\left|(\boldsymbol{X}^\top\boldsymbol{\epsilon})_i\right| \ge t\sqrt{n}\sigma\right) \le \exp\left(-\frac{t^2}{2}\right), \quad \text{for } i=1,2,\ldots,p.$$

对所有坐标应用并集界（定理 p 坐标（定理 1.61），并令参数$t=\sqrt{2(\ln(2p)+\xi)}$，可得

$$\Pr\left(\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty \ge \sigma\sqrt{2n(\ln(2p)+\xi)}\right) \le \frac{1}{2}\exp(-\xi), \quad \forall\xi>0. \tag{5.72}$$

因此，取 $\lambda = 2\sigma\sqrt{2n(\ln(2p)+\xi)} = \mathcal{O}\big(\sqrt{n\ln(p)}\big)$，则定理 5.30 以至少 $1-\exp(-\xi)/2$的概率成立。因此，预测损失的界可写为：

$$\frac{\left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\right\|_2^2}{n} \le \mathcal{O}\big(R_1\sqrt{\frac{\ln(p)}{n}}\big); \tag{5.73a}$$

$$\frac{\left\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\right\|_2^2}{n} \le \mathcal{O}\big(\frac{|\mathbb{S}|\ \ln(p)}{n}\big). \tag{5.73b}$$

第一个界（5.73a）仅取决于 ℓ_1-范数半径 R_1，称为“慢速率”，因为均方预测误差以 $1/\sqrt{n}$ 的速率衰减。相比之下，第二个界（5.73b）称为“快速率”，其衰减 $1/n$更快。

请注意，快速率依赖于更强的假设：具体而言，$\boldsymbol{\beta}^*$ 恰好是k-稀疏的（即具有硬稀疏性），而且——更重要的是——设计矩阵 $\boldsymbol{X}$ 满足 RE 条件。原则上，良好的预测性能不应要求对 $\boldsymbol{X}$ 施加此类结构条件，这表明 RE 假设可能只是当前证明技术造成的假象，而非根本要求。

在分析 LASSO 的约束形式（问题（L_C））时，其中 $\Sigma\ge 0$ 作为调节参数。令 $\Sigma=\|\boldsymbol{\beta}^*\|_1$ 可确保真实系数向量 $\boldsymbol{\beta}^*$ 是（L_C）的可行解；在此选择下，（5.73）中的相同界以至少$1-\exp(-\xi)/2$的概率成立。

即使不假设线性模型 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}$，我们仍然可以推导出一个有意义的预测损失上界，用以量化其相对于最佳线性预测器的超额风险。相反，假设

$$\boldsymbol{y} = f_0(\boldsymbol{X})+\boldsymbol{\epsilon},$$

对于某个未知函数 $f_0\colon\mathbb{R}^p\to\mathbb{R}$，其中 $f_0(\boldsymbol{X})=[f_0(\boldsymbol{x}^1)\ldots,f_0(\boldsymbol{x}^n)]^\top\in\mathbb{R}^n$ 表示 $\widehat{}$将 f_0 逐行应用于 $\boldsymbol{X}$的各行。于是，对于约束LASSO 问题（L_C）的任意解 $\boldsymbol{\beta}$ 以及任意满足 $\boldsymbol{\beta}$ 的 $\|\boldsymbol{\beta}\|_1\le\Sigma$，我们有

$$\left\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|_2^2 \le \left\|\boldsymbol{y}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\|_2^2.$$

重新排列各项可得

$$\frac{1}{2}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\|_2^2 \le \left\langle\boldsymbol{y}-\boldsymbol{X}\overline{\boldsymbol{\beta}},\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\rangle = \left\langle f_0(\boldsymbol{X})-\boldsymbol{X}\overline{\boldsymbol{\beta}},\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\rangle + \left\langle\boldsymbol{\epsilon},\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\rangle.$$

对右侧的第一个内积应用极化恒等式 $\|\boldsymbol{a}\|_2^2+\|\boldsymbol{b}\|_2^2-\|\boldsymbol{a}-\boldsymbol{b}\|_2^2=2\langle\boldsymbol{a},\ \boldsymbol{b}\rangle$ 可得

$$\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\|_2^2\leq\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2+\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\overline{\boldsymbol{\beta}}\right\|_2^2-\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2+2\left\langle\boldsymbol{X}^\top\boldsymbol{\epsilon},\widehat{\boldsymbol{\beta}}-\overline{\boldsymbol{\beta}}\right\rangle.$$

重新整理后，我们得到

$$\begin{aligned}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2&\leq\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2+2\left\langle\boldsymbol{X}^\top\boldsymbol{\epsilon},\widehat{\boldsymbol{\beta}}-\overline{\boldsymbol{\beta}}\right\rangle\leq\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2+4\Sigma\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty\\&\leq\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2+4\Sigma\sigma\sqrt{2n(\ln(2p)+\xi)},\end{aligned}$$

其中，最后一个不等式以至少 $1-\exp(-\xi)/2$ 的概率成立。由于该结论对所有满足 $\|\boldsymbol{\beta}\|_1\leq\Sigma$ 的 $\boldsymbol{\beta}$ 同时成立，且概率至少为 $1-\exp(-\xi)/2$，我们可以对右侧这类 $\boldsymbol{\beta}$ 取下确界，从而得到预言机不等式：

$$\frac{1}{n}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-f_0(\boldsymbol{X})\right\|_2^2\leq\min_{\|\boldsymbol{\beta}\|_1\leq\Sigma}\left(\frac{1}{n}\left\|\boldsymbol{X}\boldsymbol{\beta}-f_0(\boldsymbol{X})\right\|_2^2\right)+4\Sigma\sigma\sqrt{\frac{2(\ln(2p)+\xi)}{n}},\tag{5.74}$$

以至少 $1-\exp(-\xi)/2$ 的概率成立。这个不等式表明，就预测损失而言，约束 LASSO 的表现几乎不逊于 f_0 的最佳 ℓ_1-约束线性预测器。

令 $\boldsymbol{\beta}^*$ 表示 $\frac{1}{n}\left\|\boldsymbol{X}\boldsymbol{\beta}-f_0(\boldsymbol{X})\right\|_2^2$ 在 $\|\boldsymbol{\beta}\|_1\leq\Sigma$ 上的一个极小化解。[4] 那么，根据三角不等式和（5.74），我们进一步得到

$$\frac{1}{n}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}^*\right\|_2^2\leq4\Sigma\sigma\sqrt{\frac{2(\ln(2p)+\xi)}{n}},\tag{5.75}$$

以至少 $1-\exp(-\xi)/2$ 的概率成立。该结果比预言机不等式更强：它表明，LASSO 的预测与最佳 ℓ_1-约束线性预测器的预测十分接近——即使该预测器本身与真实回归函数 f_0 相差甚远。

第5章习题

1. (Candes 和 Tao，2005年) 设 $\boldsymbol{X}$ 满足 $\boldsymbol{X}\boldsymbol{Z}=\boldsymbol{0}$，并设 $k\geq1$ 是满足定理 5.13假设的一个数。令 $\boldsymbol{y}=\boldsymbol{Z}\boldsymbol{\beta}+\boldsymbol{z}$，其中 $\boldsymbol{z}$ 是一个支撑集大小至多为 k的实向量。证明 $\boldsymbol{\beta}$ 是下式的唯一极小化解

$$(\mathrm{P}_1')\qquad\min_{\boldsymbol{b}\in\mathbb{R}^p}\|\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{b}\|_1.$$

2. **ℓ_s-拟范数的三角不等式**。对于 $0<s<1$，证明 ℓ_s-拟范数的 s 次幂满足三角不等式

$$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_s^s\leq\|\boldsymbol{\alpha}\|_s^s+\|\boldsymbol{\beta}\|_s^s,\quad\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{R}^p.$$

还要证明

$$\|\boldsymbol{\alpha}_1+\ldots+\boldsymbol{\alpha}_k\|_s\leq k^{\max(1/s-1,0)}(\|\boldsymbol{\alpha}_1\|_s+\ldots+\|\boldsymbol{\alpha}_k\|_s),\quad\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_k\in\mathbb{R}^p.$$

4。根据魏尔斯特拉斯定理（定理 2.56），由于我们是在紧集上对连续函数进行最小化，因此下确界可以取到。

3. **在 ℓ_1-相干性下的恢复（Foucart 和 Rauhut，2013年）**。设 $\nu < 1/2$，并假设矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足

$$\mu_1(\boldsymbol{X}, k) \leq \nu.$$

证明：对于任意满足 $\precsim\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$ 的 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 以及来自线性模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 的观测值，任何在约束 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \leq \epsilon$ 下的 $\|\boldsymbol{\beta}\|_1$ 的极小化解 $\boldsymbol{\beta}$ 都满足 ℓ_1-误差界

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq D_1\sigma_k(\boldsymbol{\beta}^*)_1 + D_2 k\epsilon,$$

其中 D_1 和 D_2 是仅依赖于 ν 的某些正的常数。

4. 使用定理5.25中的论证证明推论5.26。

5. （ ℓ_1-分析模型（3.83）是否有利于精确恢复？讨论实现这种恢复的条件。

6. **ℓ_1-最小化：在 RE 条件（Hastie 等人，2015年）下带噪声测量**。考虑带噪声测量的 ℓ_1-最小化问题（$\mathrm{P}_{1,\ \epsilon}$）：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \leq \epsilon, \tag{5.76}$$

其中，常数 $\epsilon > 0$ 是用户指定的参数。

(a) 假设选择 ϵ，使得 $\boldsymbol{\beta}^*$ 可行（即 $\|\boldsymbol{X}\boldsymbol{\beta}^* - \boldsymbol{y}\|_2 \leq \epsilon$）。证明^误差向量 $\widehat{\boldsymbol{e}} = \boldsymbol{\beta} - \boldsymbol{\beta}^*$ 必须满足约束

$$\|\widehat{\boldsymbol{e}}_{\mathbb{S}^c}\|_1 \leq \|\widehat{\boldsymbol{e}}_{\mathbb{S}}\|_1.$$

(b) 假设线性观测模型为 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，证明 $\widehat{\boldsymbol{e}}$ 满足不等式

$$\|\boldsymbol{X}\widehat{\boldsymbol{e}}\|_2^2 \leq 2\left\{\left\|\boldsymbol{X}^\top\boldsymbol{\epsilon}\right\|_\infty \|\widehat{\boldsymbol{e}}\|_1 + \left(\epsilon^2 - \|\boldsymbol{\epsilon}\|_2^2\right)\right\}.$$

(c) 假设 $\boldsymbol{X}$ 满足 μ-受限特征值性质，利用 (b) 推导 ℓ_2-误差 $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2$的上界。

7. **LASSO 解的唯一性（Tibshirani，2013年）**。对于任意 $\boldsymbol{y}$，$\boldsymbol{X}$，以及 $\lambda \geq 0$，证明 LASSO 问题（L_L）具有以下性质：

(a) 解要么唯一，要么有不可数无穷多个解。^(b) 所有 LASSO 解都得到相同的拟合值，即 $\boldsymbol{X}\boldsymbol{\beta}$ 对每一个^解 $\boldsymbol{\beta}$都相同。(c) 如果 $\lambda > 0$，那么每个 LASSO 解 $\widehat{\boldsymbol{\beta}}$ 的 ℓ_1 范数都相同，即 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1$。

提示：考虑 *LASSO* 与 ℓ_1-最小化之间的等价性。或者，也可以用反证法论证。

8. **LASSO KKT**。考虑拉格朗日 LASSO 问题（L_L），其中 $\lambda > 0$。证明

^ ^(a) $\widehat{\boldsymbol{\beta}}$ 是（L_L）的一个解，当且仅当 $\widehat{\boldsymbol{\beta}}$ 满足（5.77a）和（5.77b），其中某个$\boldsymbol{\nu} \in \mathbb{R}^p$:

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \lambda\boldsymbol{\nu}, \tag{5.77a}$$

$$\nu_i \in \begin{cases} \{\mathrm{sign}(\widehat{\beta}_i)\} & \text{if } \widehat{\beta}_i \neq 0; \\ [-1, 1] & \text{if } \widehat{\beta}_i = 0, \end{cases} \quad \text{for } i = 1, 2, \ldots p. \tag{5.77b}$$

^这里，$\boldsymbol{\nu} \in \mathbb{R}^p$ 是函数 $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ 在 $\boldsymbol{\beta} = \boldsymbol{\beta}$（定义2.57）处的一个次梯度。

(b) 满足上述条件的次梯度 $\boldsymbol{\nu}$ 是唯一的。

(c) 将等相关集 $\mathbb{S}$ 和等相关符号 $\boldsymbol{s}$ 定义为

$$\mathbb{S} \triangleq \left\{ i \in \{1, 2, \dots p\} \mid \left| \boldsymbol{x}_i^\top (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \right| = \lambda \right\}; \tag{5.77c}$$

$$\boldsymbol{s} \triangleq \operatorname{sign}\big(\boldsymbol{X}_{\mathbb{S}}^\top (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\big). \tag{5.77d}$$

证明 $\mathbb{S}$ 和 $\boldsymbol{s}$ 均是唯一确定的（即与所选的 LASSO 解无关）。

(d) 证明 $\widehat{\boldsymbol{\beta}}_{\mathbb{S}^c} = \boldsymbol{0}$ 对于任意 LASSO 解 $\widehat{\boldsymbol{\beta}}$ 均成立。因此，将式（5.77a）限制在索引 $\mathbb{S}$上可得

$$\boldsymbol{X}_{\mathbb{S}}^\top (\boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) = \lambda \boldsymbol{s}. \tag{5.77e}$$

这意味着 $\lambda \boldsymbol{s}$ 属于 $\boldsymbol{X}_{\mathbb{S}}$的行空间，因此 $\lambda \boldsymbol{s} = \boldsymbol{X}_{\mathbb{S}}^\top (\boldsymbol{X}_{\mathbb{S}}^\top)^+ \lambda \boldsymbol{s}$。利用这一事实并重新整理式（5.77e），可得

$$\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}} \widehat{\boldsymbol{\beta}}_{\mathbb{S}} = \boldsymbol{X}_{\mathbb{S}}^\top (\boldsymbol{y} - (\boldsymbol{X}_{\mathbb{S}}^\top)^+ \lambda \boldsymbol{s}).$$

提示：使用习题5.7。

9. **LASSO 解的唯一性。** 对于任意 $\boldsymbol{y}$, $\boldsymbol{X}$ 和 $\lambda > 0$，假设与等相关集 $\mathbb{S}$ 对应的子矩阵 $\boldsymbol{X}_{\mathbb{S}}$ 满足 $\mathcal{N}(\boldsymbol{X}_{\mathbb{S}}) = \{\boldsymbol{0}\}$，或者等价地，$\operatorname{rank}(\boldsymbol{X}_{\mathbb{S}}) = |\mathbb{S}|$。证明拉格朗日 LASSO 解唯一，且其形式为

$$\widehat{\boldsymbol{\beta}}_{\mathbb{S}^c} = \boldsymbol{0} \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{\mathbb{S}} = (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1} (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{y} - \lambda \boldsymbol{s}),$$

其中 $\mathbb{S}$和 $\boldsymbol{s}$ 分别是式（5.77c）和（5.77d）中定义的等相关集和符号函数。注意

10. **LASSO 解的唯一性。** 假设矩阵 $\boldsymbol{X}$ 的列 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、……、$\boldsymbol{x}_p \in \mathbb{R}^n$ 处于一般位置。这意味着，对于任意整数 $k < \min\{n, p\}$、索引 i_1、i_2、……、$i_{k+1} \in \{1, 2, \dots, p\}$ 以及符号函数 s_1、s_2、……、$s_{k+1} \in \{-1, +1\}$，向量 $s_1 \boldsymbol{x}_{i_1}$、$s_2 \boldsymbol{x}_{i_2}$、……、$s_{k+1} \boldsymbol{x}_{i_{k+1}}$ 的仿射张成不包含集合 $\{\pm \boldsymbol{x}_i \mid i \neq i_1、\dots\dots、i_{k+1}\}$ 中的任何向量。证明，对于任意 $\boldsymbol{y} \in \mathbb{R}^n$ 和 $\lambda > 0$，LASSO 问题 (L_L) 都有唯一解。

11. 考虑拉格朗日 LASSO（L_L），并令 $\widehat{\boldsymbol{\beta}}$ 表示它的任意一个解。证明对于任意系数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，以下不等式成立：

$$\left\| \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta} \right\|_2^2 \leq 2 \left\langle \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta} \right\rangle + 2\lambda (\|\boldsymbol{\beta}\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1),$$

提示：对于任意系数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，利用 $\widehat{\boldsymbol{\beta}}$ 的最优性：$\frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \right\|_2^2 + \lambda \|\widehat{\boldsymbol{\beta}}\|_1 \leq \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1$.

12. **LASSO 的预测误差。** 假设观测值服从线性模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$，其中 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$。考虑正则化参数为 $\lambda \geq \|\boldsymbol{X}^\top \boldsymbol{\epsilon}\|_\infty$的拉格朗日 LASSO（$\mathrm{L}_L$）。若 $\|\boldsymbol{\beta}^*\|_1 < R_1$，证明任意最优解 $\widehat{\boldsymbol{\beta}}$ 满足

$$\left\| \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \right\|_2^2 \leq 4 R_1 \lambda.$$

与定理5.30相比，这一结果给出了更精细的上界。值得注意的是，它所要求的条件 $\lambda \geq \|\boldsymbol{X}^\top \boldsymbol{\epsilon}\|_\infty$比 $\lambda \geq 2\|\boldsymbol{X}^\top \boldsymbol{\epsilon}\|_\infty$限制更少。提示：利用问题5.11 和 *Hölder* 不等式来界定预测损失。

第6章

使用高斯系综设计感知矩阵

目录

6.1 覆盖数.2536.2 高斯向量的集中性.2556.3 高斯随机矩阵下的随机投影 . . 2606.4 随机投影与 Johnson–Lindenstrauss 引理 . 2616.5 高斯随机矩阵的奇异值.2666.6 高斯随机矩阵下的稀疏优化 268第6章习题.271



S由于受限等距性质（RIP）、受限特征值（RE）条件、零空间性质（NSP）和受限强凸性（RSC）等性质对于保证稀疏恢复至关重要，因此有必要研究实际或构造的数据矩阵满足这些条件的问题设置。大量研究致力于显式构造能够证明满足 RIP 的矩阵——例如，参见Baraniuk 等人（2008年）；Bourgain 等人（2011年）；Jain 等人（2017年）。

为了建立直觉，我们首先介绍与高斯分布相关的必要概率工具。我们先分析使用随机高斯矩阵进行的稀疏恢复，然后在第7章将结果推广到次高斯随机矩阵。

6.1. 覆盖数

在讨论高斯随机变量及其在确保受限等距性质方面的应用之前，我们先介绍一些关于覆盖数的重要数学工具。

覆盖数是指完全覆盖给定空间或集合所需的集合的最小数量——通常是在度量空间中使用球，或在组合情形中使用子结构。相反，打包数衡量一个集合中能够放置多少个点$\mathbb{S}$，并且任意两点之间至少相距 ε。这刻画了该集合的“分散程度”或“彼此分离程度”。

形式化地，我们有如下定义。

定义 6.1（覆盖数、打包数）。设 $\mathbb{S}$是度量空间 $(\mathbb{X}, d)$ 的子集，其中 $d(\cdot,\cdot)$ 表示度量。对于 $\varepsilon > 0$，$\mathbb{S}$ 的 ε-覆盖数（或简称覆盖数）记作 $\mathcal{C}(\mathbb{S}, d, \varepsilon)$，定义为满足以下条件的最小整数 $\mathcal{C}$：$\mathbb{S}$ 可以被 $\mathcal{C}$ 个半径为 ε 的闭球覆盖，$\mathbb{B}_d[\boldsymbol{x}_i, \varepsilon] \triangleq \{\boldsymbol{y} \in \mathbb{X} \mid d(\boldsymbol{x}_i, \boldsymbol{y}) \leq \varepsilon\}$，$\boldsymbol{x}_i \in \mathbb{S}, i = 1, 2, \ldots, \mathcal{C}$。也就是说，

$$\mathcal{C} \equiv \mathcal{C}(\mathbb{S}, d, \varepsilon) = \min \left\{ k \in \mathbb{N} \mid \exists\, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{X} \text{ such that } \mathbb{S} \subset \bigcup_{i=1}^{k} \mathbb{B}[\boldsymbol{x}_i, \varepsilon] \right\}.$$

任何这样的集合 $\mathbb{C}_\varepsilon \triangleq \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathcal{C}\}$ 都称为 $\mathbb{S}$ 的 ε-覆盖。注意，中心 $\boldsymbol{x}_i$ 不一**定**要位于 $\mathbb{S}$ 中。

对于 $\varepsilon > 0$，$\mathbb{S}$ 的 ε-打包数（或简称打包数）记作 $\mathcal{P}(\mathbb{S}, d, \varepsilon)$，它是以 $\mathbb{S}$ 中的点为中心、半径为 ε 的互不相交闭球的最大数量 $\mathcal{P}$。等价地，它是满足 $d(\boldsymbol{x}_i, \boldsymbol{x}_i) > \varepsilon$ 对所有 $i \neq j$ 均成立的点 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathcal{P}$ 的最大数量 $\mathcal{P}$（也就是说，它是以 $\mathbb{S}$ 中的点为中心、半径为 $\varepsilon/2$ 的互不相交闭球的最大数量）。这样的集合 $\mathbb{P}_\varepsilon \triangleq \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathcal{P}\}$ 称为 ε-打包集或 ε-分离集。

随着 $\varepsilon \to 0$，需要更精细的分辨率，因此覆盖数 $\mathcal{C}(\mathbb{S}, d, \varepsilon)$ 通常会增加。

当 $\mathbb{X} = \mathbb{R}^n$ 是向量空间且度量由范数诱导时，即 $d(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{a} - \boldsymbol{b}\|$，为简便起见，我们通常将其写作：

$$\mathcal{C}(\mathbb{S}, \|\cdot\|, \varepsilon) \qquad \text{and} \qquad \mathcal{P}(\mathbb{S}, \|\cdot\|, \varepsilon).$$

现在我们列出覆盖数的若干基本性质；装填数也满足类似的性质。对于任意集合 $\mathbb{M}, \mathbb{S} \subset \mathbb{X}$, 以下“三角不等式”成立：

$$\mathcal{C}(\mathbb{M} \cup \mathbb{S}, d, \varepsilon) \leq \mathcal{C}(\mathbb{M}, d, \varepsilon) + \mathcal{C}(\mathbb{S}, d, \varepsilon). \tag{6.1a}$$

对于任意 $\eta > 0$,

$$\mathcal{C}(\mathbb{S}, \eta d, \varepsilon) = \mathcal{C}(\mathbb{S}, d, \varepsilon/\eta). \tag{6.1b}$$

如果 $\mathbb{X} = \mathbb{R}^n$ 和 $d(\cdot, \cdot)$ 是由范数 $\|\cdot\|$ 诱导的，那么

$$\mathcal{C}(\eta\mathbb{S}, d, \varepsilon) = \mathcal{C}(\mathbb{S}, d, \eta^{-1}\varepsilon). \tag{6.1c}$$

此外，如果 $d'(\cdot, \cdot)$ 是 $\mathbb{X}$ 上满足对所有 $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{X}$ 均有 $d'(\boldsymbol{a}, \boldsymbol{b}) \leq d(\boldsymbol{a}, \boldsymbol{b})$ 的另一种度量，那么

$$\mathcal{C}(\mathbb{S}, d', \varepsilon) \leq \mathcal{C}(\mathbb{S}, d, \varepsilon). \tag{6.1d}$$

覆盖数与打包数之间存在一个基本关系式：

引理 6.2：（覆盖数与打包数之间的关键关系）设 $\mathbb{S}$是度量空间 $(\mathbb{X}, d)$ 的子集，并设 $\varepsilon > 0$。那么，

$$\mathcal{P}(\mathbb{S}, d, 2\varepsilon) \leq \mathcal{C}(\mathbb{S}, d, \varepsilon) \leq \mathcal{P}(\mathbb{S}, d, \varepsilon).$$

证明 [引理 6.2] 设 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathcal{P}\}$ 是一个 2ε-打包集，使得每个 $\boldsymbol{x}_i \in \mathbb{S}$，且 $d(\boldsymbol{x}_i, \boldsymbol{x}_j) > 2\varepsilon$，其中 $i \neq j$；再设 $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_\mathcal{C}\}$ 是一个 ε-覆盖集，使得 $\mathbb{S} \subset \bigcup_{i=1}^{\mathcal{C}} \mathbb{B}[\boldsymbol{y}_i, \varepsilon]$。每个 $\boldsymbol{x}_j$ 都必须至少属于一个（闭）球 $\mathbb{B}_d[\boldsymbol{y}_i, \varepsilon]$。也就是说，我们可以为每个点 $\boldsymbol{y}_i$ 指定一个点 $\boldsymbol{x}_j$，使得 $d(\boldsymbol{y}_i, \boldsymbol{x}_j) \leq \varepsilon$。关键的是，每个球 $\mathbb{B}_d[\boldsymbol{y}_i, \varepsilon]$ 至多包含来自 2ε-打包集的一个点。事实上，如果两个不同的点 $\boldsymbol{x}_i$、$\boldsymbol{x}_k$ $(i \neq k)$ 都属于 $\mathbb{B}_d[\boldsymbol{y}_i, \varepsilon]$，满足

$$d(\boldsymbol{x}_i, \boldsymbol{y}_j) \leq \varepsilon \qquad \text{and} \qquad d(\boldsymbol{x}_k, \boldsymbol{y}_j) \leq \varepsilon,$$

根据三角不等式，

$$d(\boldsymbol{x}_i, \boldsymbol{x}_k) \leq d(\boldsymbol{x}_i, \boldsymbol{y}_j) + d(\boldsymbol{x}_k, \boldsymbol{y}_j) \leq 2\varepsilon,$$

这与 2ε-打包集的定义相矛盾。由此证明了第一个不等式 y.

对于第二个不等式，设 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathcal{P}\}$ 是 $\mathbb{S}$的一个 ε-打包集。由于它是极大的，对于任意点 $\boldsymbol{x} \in \mathbb{S}$，都存在某个 $\boldsymbol{x}_i$，使得 $d(\boldsymbol{x}_i, \boldsymbol{x}) \leq \varepsilon$。否则，我们就可以将$\boldsymbol{x}$ 加入该打包集，这与极大性相矛盾。因此，$\mathbb{S} \subset \bigcup_{i=1}^{\mathcal{P}} \mathbb{B}[\boldsymbol{y}_i, \varepsilon]$。这意味着 $\mathcal{P}$ 至少等于覆盖数。 ■

下面的结果给出了有限维范数诱导的单位球面（范数球面）的打包数（从而也是覆盖数）的上界。

定理 6.3：（范数球面的覆盖数与打包数）设 $\|\cdot\|$ besome 是 $\mathbb{R}^n$ 上的一个范数，且 $\mathbb{A}$ 是单位球 $\mathbb{B} \triangleq \mathbb{B}_{\|\cdot\|}[\boldsymbol{0}, 1] = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \leq 1\}$ 的子集。那么，n 维球面 $\mathbb{B}$ 在尺度 $\varepsilon < 1$ 下的打包数和覆盖数满足

$$\mathcal{C}(\mathbb{A}, \|\cdot\|, \varepsilon) \leq \mathcal{P}(\mathbb{A}, \|\cdot\|, \varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n. \tag{6.2}$$

证明 [定理6.3] 第一个不等式由引理6.2得出。令 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{\mathcal{P}}\} \subset \mathbb{A}$ 为 ε的一个极大 $\mathbb{A}$-打包。则球 $\mathbb{B}_{\|\cdot\|}[\boldsymbol{x}_\ell, \varepsilon/2]$ 两两不相交，且全部包含于半径为$1+\varepsilon/2$的缩放球（$\mathbb{B}$）中。比较体积可得$1+\varepsilon/2$。

$$\operatorname{Vol}\left(\bigcup_{\ell=1}^{\mathcal{P}} \mathbb{B}_{\|\cdot\|}[\boldsymbol{x}_\ell, \frac{\varepsilon}{2}]\right) = \mathcal{P}\operatorname{Vol}\left(\frac{\varepsilon}{2}\mathbb{B}\right) \leq \operatorname{Vol}\left((1+\frac{\varepsilon}{2})\mathbb{B}\right).$$

由于体积满足

$$\operatorname{Vol}(\varepsilon\mathbb{B}) = \varepsilon^n \operatorname{Vol}(\mathbb{B}) \quad \text{on } \mathbb{R}^n,$$

我们得到

$$\mathcal{P}(\frac{\varepsilon}{2})^n \operatorname{Vol}(\mathbb{B}) \leq (1+\frac{\varepsilon}{2})^n \operatorname{Vol}(\mathbb{B}) \qquad \Longrightarrow \qquad \mathcal{P} \leq (1+\frac{2}{\varepsilon})^n.$$

证明至此完成。 ■

具体而言，以下推论给出了 n维球面的覆盖数和装填数的上界

$$\mathbb{U}^{n-1} \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| = 1\},$$

其中包含 $\mathbb{R}^n$ 中的单位 (ℓ_2) 范数向量。请注意，对于给定的范数 $\|\cdot\|$: $\mathbb{R}^n \to \mathbb{R}$， $\mathbb{R}^n$ 中的单位球与 n 维球面之间的区别：

$$\mathbb{B}_{\|\cdot\|}[\mathbf{0}, 1] \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \leq 1\}; \tag{6.3a}$$

$$\mathbb{U}^{n-1} \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| = 1\}. \tag{6.3b}$$

推论 6.4：（球面的覆盖数） n维球面 $\mathbb{U}^{n-1}$ 在尺度 $\varepsilon < 1$ 下的覆盖数和打包数满足

$$\mathcal{C}\left(\mathbb{U}^{n-1}, \|\cdot\|, \varepsilon\right) \leq \mathcal{P}\left(\mathbb{U}^{n-1}, \|\cdot\|, \varepsilon\right) \leq \left(\frac{2+\varepsilon}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n. \tag{6.4}$$

我们递归地构造一个 ε-打包集 $\mathbb{P}_\varepsilon \subseteq \mathbb{U}^{n-1}$，如下所示：

- 我们将 $\mathbb{P}_\varepsilon$ 初始化为空集。选取一个点 $\boldsymbol{\alpha} \in \mathbb{U}^{n-1}$，使得对于任意 $\boldsymbol{\beta} \in \mathbb{P}_\varepsilon$都有
- $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\| > \varepsilon$。我们将 $\boldsymbol{\alpha}$ 加入 $\mathbb{P}_\varepsilon$，直到 $\mathbb{U}^{n-1}$ 中不存在距 $\mathbb{P}_\varepsilon$中任意点 ε 的点。

由于 $\mathbb{U}^{n-1}$ 是紧集，该过程经过有限步后终止；否则，无限序列将存在一个收敛子列，这与分离条件相矛盾。

6.2. 高斯向量的集中性

在低维情况下，独立同分布标准高斯向量的联合概率密度函数高度集中在原点附近。然而，随着环境空间的维数增加，这种行为会发生显著变化。具体来说，考虑一个 n 维向量 $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^\top$

其元素是独立同分布的标准高斯随机变量。[1] 其平方 ℓ_2范数为$\mathbf{x}$，记作 $\|\mathbf{x}\|_2^2 = \sum_{i=1}^n \mathrm{x}_i^2$，服从具有 χ^2 个 n 自由度的卡方（1.88）分布（见定义 n增大时，这个 χ^2 随机变量的概率密度函数会越来越集中在其均值附近，该均值等于 n:

$$\mathbb{E}[\|\mathbf{x}\|_2^2] = \mathbb{E}\left[\sum_{i=1}^n \mathrm{x}_i^2\right] = \sum_{i=1}^n \mathbb{E}[\mathrm{x}_i^2] = n. \tag{6.5}$$

此外，我们还给出了关于标准高斯随机向量的 ℓ_1-、 ℓ_2- 和ℓ_∞-范数期望值的简单结果。ℓ_∞-范数的上界依赖于第7.1节中讨论的次高斯随机变量的性质。

定理 6.5：（标准高斯向量的范数）设$\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^\top$ 是由（不一定相互独立的）标准高斯随机变量组成的向量。那么，

(i) $\mathbb{E}[\|\mathbf{x}\|_1] = \sqrt{\frac{2}{\pi}} n$. **(ii)** 有

$$\mathbb{E}[\|\mathbf{x}\|_2^2] = n \qquad \text{and} \qquad \sqrt{\frac{2}{\pi}}\sqrt{n} \le \mathbb{E}[\|\mathbf{x}\|_2] \le \sqrt{n}. \tag{6.6}$$

如果$\mathbf{x}$ 的元素相互独立，那么

$$\frac{n}{\sqrt{n+1}} \le \mathbb{E}[\|\mathbf{x}\|_2] \le \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \le \sqrt{n}, \tag{6.7}$$

因此 $\mathbb{E}[\|\mathbf{x}\|_2] \sim \sqrt{n}$ 当 $n \to \infty$时。

(iii) 成立：

$$\mathbb{E}[\max_{i\in\{1,2,\ldots,n\}} \mathrm{x}_i] \le \sqrt{2\ln(n)} \qquad \text{and} \qquad \mathbb{E}[\|\mathbf{x}\|_\infty] \le \sqrt{2\ln(2n)}. \tag{6.8}$$

若$\mathbf{x}$ 的元素相互独立，则对于 $n \ge 2$，有

$$\mathbb{E}[\|\mathbf{x}\|_\infty] \ge C\sqrt{\ln(n)}, \quad \text{with } C \approx 0.265. \tag{6.9}$$

证明 [定理 6.5] **(i).** 利用标准高斯随机变量的密度，

$$\mathbb{E}[|\mathrm{x}_i|] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |x| \exp(-\frac{x^2}{2})dx = \sqrt{\frac{2}{\pi}}\int_0^{\infty} x\exp(-\frac{x^2}{2})dx = \sqrt{\frac{2}{\pi}}.$$

根据期望的线性性，$\mathbb{E}[\|\mathbf{x}\|_1] = \sum_{i=1}^n \mathbb{E}[|\mathbf{x}_i|] = \sqrt{2/\pi}n$。

(ii). 显然，$\mathbb{E}[\|\mathbf{x}\|_2^2] = \sum_{i=1}^n \mathbb{E}[\mathrm{x}_i^2] = n$ 其中 $\mathbf{x}$ 为标准高斯随机变量。对平方函数的期望应用詹森不等式可得 $\mathbb{E}[\|\mathbf{x}\|_2] \le \sqrt{\mathbb{E}[\|\mathbf{x}\|_2^2]} = \sqrt{n}$，而对 $\mathbb{R}^n$ 上的内积应用柯西–施瓦茨不等式可得 $\mathbb{E}[\|\mathbf{x}\|_2] \ge \frac{1}{\sqrt{n}}\mathbb{E}[\|\mathbf{x}\|_1] = \sqrt{2/\pi}\sqrt{n}$；更多详情请参见习题 1.26 。

1. 再次注意，我们用粗体小写字母的正体表示随机向量，用粗体大写字母的正体表示随机矩阵。也就是说，$\mathbf{x}$、$\mathbf{a}$、$\mathbf{X}$ 是随机标量、向量或矩阵；而 x、$\boldsymbol{a}$、$\boldsymbol{X}$ 是标量、向量或矩阵。在许多情况下，这两个术语可以互换使用；也就是说，$\mathbf{x} = x$ 表示该变量的一个实现。

如果 $\mathbf{x}$ 的元素相互独立，那么 $\|\mathbf{x}\|_2^2$ 服从 $\chi^2_{(n)}$ 卡方分布（定义1.88）。因此，

$$\mathbb{E}[\|\mathbf{x}\|_2] = \int_0^\infty x^{\frac{1}{2}} \cdot \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} dx = \frac{2^{\frac{n}{2}+\frac{1}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty t^{\frac{n}{2}-\frac{1}{2}} e^{-t} dt = \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)},$$

其中，最后一个等式由伽马函数为 $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)dt$ 得出。令 $T_n \triangleq \mathbb{E}[\|\mathbf{x}\|_2] \leq \sqrt{n}$，其中长度为$n$ 的高斯向量 $\mathbf{x}$ 的上界已在上文给出。根据 $\Gamma(x+1) = x\Gamma(x)$ 对于 $x > 0$，可得

$$T_{n+1}T_n = 2\frac{\Gamma(n/2+1)}{\Gamma(n/2)} = n,$$

这意味着 $T_n = n/T_{n+1} \geq n/\sqrt{n+1}$，因为 $T_{n+1} \leq \sqrt{n+1}$。因此，$n/\sqrt{n+1} \leq T_n \leq \sqrt{n}$，这就证明了所需结果。

(iii). （6.8）中的界由命题 7.15（关于次高斯随机变量的最大值）得出。事实上，引理 1.85 表明 $\mathbb{E}[\exp(\theta \mathrm{x}_i)] \leq \exp(\theta^2/2)$，对所有 $\theta \in \mathbb{R}$ 均成立，因此参数 C 在针对高斯随机变量的命题 $C = 1/2$ 中为 7.15 。

对于独立性条件下的下界，由习题1.66

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}\|_\infty] &= \int_0^\infty \Pr\left(\max_{i\in\{1,2,\ldots,n\}} |\mathrm{x}_i| > x\right) dx = \int_0^\infty \left(1 - \Pr\left(\max_{i\in\{1,2,\ldots,n\}} |\mathrm{x}_i| \leq x\right)\right) dx \\ &= \int_0^\infty \left(1 - \prod_{i=1}^n \Pr(|\mathrm{x}_i| \leq x)\right) dx \geq \int_0^\delta \left(1 - \big(\Pr(|\mathrm{x}| \leq x)\big)^n\right) dx \\ &\geq \delta\big(1 - \big(1 - \Pr(|\mathrm{x}| > \delta)\big)^n\big),\end{aligned}$$

其中 $\delta > 0$。对于任意 $\delta > 0$，我们还有

$$\Pr(|\mathrm{x}| > \delta) = \sqrt{\frac{2}{\pi}} \int_\delta^\infty e^{-t^2/2} dt \geq \sqrt{\frac{2}{\pi}} \int_\delta^{2\delta} e^{-t^2/2} dt \geq \sqrt{\frac{2}{\pi}} \delta e^{-2\delta^2}.$$

取 $\delta = \sqrt{\ln n/2}$，对于 $n \geq 2$，

$$\mathbb{E}[\|\mathbf{x}\|_\infty] \geq \sqrt{\frac{\ln n}{2}} \left(1 - \left(1 - \sqrt{\frac{\ln n}{\pi}}\frac{1}{n}\right)^n\right) \geq \sqrt{\frac{\ln n}{2}} \left(1 - \exp\left(-\sqrt{\frac{\ln n}{\pi}}\right)\right),$$

其中，第二个不等式成立是因为当 $n > x > 0$ 时，$(1 - \frac{x}{n})^n \leq \exp(-x)$。对于 $n \geq 2$，圆括号中的因子有一个正的常数下界，从而得到所述下界，其中 $C = (1 - \exp(-\sqrt{\ln(2)/\pi}))/\sqrt{2} \approx 0.265$。 ■

我们还可以进一步推广前一个定理的 (iii) 部分，从而得到一列标准高斯随机向量的最大平方ℓ_2-范数的上界。

定理 6.6：（高斯向量范数的最大值）设$\mathbf{x}_1, \mathbf{x}_2, \ldots\ldots, \mathbf{x}_n \in \mathbb{R}^p$为一列（不一定相互独立的）标准高斯随机向量。那么，对于任意 $\nu > 0$，

$$\mathbb{E}\left[\max_{i\in\{1,2,\ldots,n\}} \|\mathbf{x}_i\|_2^2\right] \leq 2(1+\nu)\ln(n) + p(1+\nu)\ln(1+\nu^{-1}).$$

因此，

$$\mathbb{E}\left[\max_{i\in\{1,2,\ldots,n\}}\|\mathbf{x}_i\|_2^2\right]\leq(\sqrt{2\ln(n)}+\sqrt{p})^2.$$

证明 [定理 6。6] 根据对数的凹性和詹森不等式，对于任意$\theta>0$,

$$\begin{aligned}\mathbb{E}\left[\max_i\|\mathbf{x}_i\|_2^2\right]&=\theta^{-1}\mathbb{E}\left[\ln\max_i\exp(\theta\|\mathbf{x}_i\|_2^2)\right]\leq\theta^{-1}\ln\mathbb{E}\left[\max_i\exp(\theta\|\mathbf{x}_i\|_2^2)\right]\\&\leq\theta^{-1}\ln\left(\sum_{i=1}^n\mathbb{E}[\exp(\theta\|\mathbf{x}_i\|_2^2)]\right),\end{aligned}$$

对于每个标准高斯随机向量 $\mathbf{x}=\mathbf{x}_i$，根据 $\mathbf{x}$ 的分量之间的独立性以及引理 1.85，对于 $\theta<1/2$,

$$\mathbb{E}[\exp(\theta\|\mathbf{x}\|_2^2)]=\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^p\mathrm{x}_i^2\right)\right]=\mathbb{E}\left[\prod_{i=1}^p\exp(\theta\mathrm{x}_i^2)\right]=\prod_{i=1}^p\mathbb{E}[\exp(\theta\mathrm{x}_i^2)]\leq(1-2\theta)^{-\frac{p}{2}},$$

因此，

$$\mathbb{E}\left[\max_i\|\mathbf{x}_i\|_2^2\right]\leq\min_{0<\theta<1/2}\theta^{-1}\left(\ln n+\frac{p}{2}\ln\left((1-2\theta)^{-1}\right)\right).$$

选择 $\theta=(2+2\nu)^{-1}$ 即可得到第一个结论。利用 $\ln(1+\nu^{-1})\leq\nu^{-1}$ 这一事实，并选择 $\nu=\sqrt{p/(2\ln(n))}$，可得

$$\begin{aligned}\mathbb{E}\left[\max_i\|\mathbf{x}_i\|_2^2\right]&\leq2(1+\nu)\ln(n)+p(1+\nu^{-1})\\&=2\ln(n)+2\sqrt{2p\ln(n)}+p=(\sqrt{2\ln(n)}+\sqrt{p})^2.\end{aligned}$$

证明完毕。 ■

以下结果量化了标准高斯向量的平方 ℓ_2 范数的变异性。

命题 6.7：（高斯向量平方 ℓ_2范数的方差）设$\mathbf{x}\in\mathbb{R}^n$ 是一个独立同分布的标准高斯随机向量。$\|\mathbf{x}\|_2^2$ 的方差为 $2n$。

证明 [命题6.7] 回忆一下，$\mathbb{V}\mathrm{ar}[\|\mathbf{x}\|_2^2]=\mathbb{E}[(\|\mathbf{x}\|_2^2)^2]-(\mathbb{E}[\|\mathbf{x}\|_2^2])^2$。我们计算：

$$\begin{aligned}\mathbb{E}\left[\left(\|\mathbf{x}\|_2^2\right)^2\right]&=\mathbb{E}\left[\left(\sum_{i=1}^n\mathrm{x}_i^2\right)^2\right]=\mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^n\mathrm{x}_i^2\mathrm{x}_j^2\right]=\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}[\mathrm{x}_i^2\mathrm{x}_j^2]\\&=\sum_{i=1}^n\mathbb{E}[\mathrm{x}_i^4]+\sum_{i\neq j}^{n,n}\mathbb{E}[\mathrm{x}_i^2]\mathbb{E}[\mathrm{x}_j^2]\overset{\dagger}{=}3n+n(n-1)=n(n+2),\end{aligned}$$

其中等式（†）成立，是因为标准高斯分布的四阶矩等于 3。因此，$\mathbb{V}\mathrm{ar}[\|\mathbf{x}\|_2^2]=n(n+2)-n^2=2n$ 根据定理6.5。 ■

该结果表明，随着 n 增大，$\|\mathbf{x}\|_2$ 围绕其均值的相对波动按$1/\sqrt{n}$的速率衰减。换言之，平方范数越来越集中在n附近。下面的定理通过给出一个非渐近尾界，精确刻画了这种集中现象。

定理6.8：（独立同分布标准高斯向量的 ℓ_2-范数的切比雪夫尾界）设$\mathbf{x} \in \mathbb{R}^n$ 是一个独立同分布的标准高斯随机向量。则，对于任意 $\varepsilon > 0$,

$$\Pr\left(n(1-\varepsilon) \le \|\mathbf{x}\|_2^2 \le n(1+\varepsilon)\right) \ge 1 - \frac{2}{n\varepsilon^2}. \tag{6.10}$$

证明 [定理 6.8] 设 $\mathsf{y} \triangleq \|\mathbf{x}\|_2^2$。则定理 6.5 表明 $\mathsf{y} - n = \mathrm{y} - \mathbb{E}[\mathrm{y}]$。马尔可夫不等式（定理 1.68）表明

$$\Pr\left((\mathrm{y} - \mathbb{E}[\mathrm{y}])^2 \ge n^2\varepsilon^2\right) \le \frac{\mathbb{E}[(\mathrm{y} - \mathbb{E}[\mathrm{y}])^2]}{n^2\varepsilon^2} = \frac{\mathbb{V}\mathrm{ar}[\mathrm{y}]}{n^2\varepsilon^2} = \frac{2n}{n^2\varepsilon}.$$

取补集可得所需的上界（6.10）。 ■

定理 6.8 中的上界仅依赖方差来控制平方范数偏离其均值的概率。因此，它明显弱于下面的结果；该结果利用了标准高斯分布的高阶矩所表现出的良好性质。

定理 6.9：（独立同分布的标准高斯随机向量的 ℓ_2-范数的切尔诺夫尾界）设 $\mathbf{x} \in \mathbb{R}^n$ 为独立同分布的标准高斯随机向量。对于任意$\varepsilon \in (0, 1)$，有

$$\Pr\left(n(1-\varepsilon) \le \|\mathbf{x}\|_2^2 \le n(1+\varepsilon)\right) \ge 1 - 2\exp\left(-\frac{n\varepsilon^2}{8}\right). \tag{6.11}$$

证明 [定理69] 设 $\mathsf{y} \triangleq \|\mathbf{x}\|_2^2$。所需的不等式由以下两个单侧界得出：

$$\Pr(\mathrm{y} \ge n(1+\varepsilon)) \le \exp\left(-\frac{n\varepsilon^2}{8}\right);$$
$$\Pr(\mathrm{y} \ge n(1-\varepsilon)) \le \exp\left(-\frac{n\varepsilon^2}{8}\right).$$

我们证明第一个；第二个类似。固定 $\theta \in (0, 1/2)$。根据马尔可夫不等式（定理 1.68）以及 $\mathbf{x}$ 各分量的独立性，对于任意 $\tau > 0$:

$$\begin{aligned}\Pr(\mathrm{y} > \tau) &= \Pr(\exp(\theta y) > \exp(\theta\tau)) \le \exp(-\theta\tau)\,\mathbb{E}[\exp(\theta y)] \\ &\le \exp(-\theta\tau)\,\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^n \mathrm{x}_i^2\right)\right] = \exp(-\theta\tau)\prod_{i=1}^n \mathbb{E}[\exp(\theta\mathrm{x}_i^2)] = \frac{\exp(-\theta\tau)}{(1-2\theta)^{n/2}},\end{aligned}$$

其中，最后一个等式是标准高斯变量的矩引理（引理 1.85）的推论。令 $\tau \triangleq n(1+\varepsilon)$ 和 $\theta \triangleq \frac{1}{2} - \frac{1}{2(1+\varepsilon)}$。对$\theta \in (0, 1/2)$取下确界，得到

$$\Pr(\mathrm{y} > n(1+\varepsilon)) \le (1+\varepsilon)^{n/2}\exp\left(-\frac{n\varepsilon}{2}\right) = \exp\left(-\frac{n}{2}(\varepsilon - \ln(1+\varepsilon))\right) \le \exp\left(-\frac{n\varepsilon^2}{8}\right),$$

其中最后一个等式源于函数 $g(x) \triangleq x - \frac{x^2}{4} - \ln(1+x)$ 在 0 和 1 之间非负（其导数非负，且 $g(0) = 0$）。类似的论证可得下尾界，从而完成证明。 ■

▶ **投影后的尾界。** 众所周知，独立同分布的标准高斯噪声投影到固定子空间后，其 ℓ_2-范数集中在子空间维数的平方根附近。这是因为，在该子空间的任意标准正交基下，投影后的坐标仍然是独立同分布的标准高斯变量。

引理 6.10：（将独立同分布的高斯向量投影到子空间）设$\mathcal{V}$ 是 $\mathbb{R}^n$ 的一个 k 维子空间，且$\mathbf{x} \in \mathbb{R}^n$ 是一个元素为独立同分布标准高斯元素的向量。则$\mathbf{x}$ 在 $\mathcal{V}$上的正交投影的平方范数，记为$\|\mathcal{P}_{\mathcal{V}}\mathbf{x}\|_2^2$，服从自由度为 k 的 χ^2 分布。等价地，它与下式具有相同分布

$$\mathrm{y} \triangleq \sum_{i=1}^{k} \mathrm{y}_i^2 \tag{6.12}$$

其中 $\mathrm{y}_1, \mathrm{y}_2, \ldots\ldots, \mathrm{y}_k$ 是独立同分布的标准高斯随机变量。

证明 [引理 6.10] 设 $\boldsymbol{Q} \in \mathbb{R}^{n\times k}$ 是一个矩阵，其列构成 $\mathcal{V}$ （即 $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_k$）的标准正交基。则 $\boldsymbol{Q}\boldsymbol{Q}^\top$ 是子空间的正交投影矩阵 $\mathcal{V}$ （引理1.45）。因此，我们有

$$\|\mathcal{P}_{\mathcal{V}}\mathbf{x}\|_2^2 = \left\|\boldsymbol{Q}\boldsymbol{Q}^\top\mathbf{x}\right\|_2^2 = \mathbf{x}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\mathbf{x} = \mathbf{x}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\mathbf{x} \triangleq \mathbf{z}^\top\mathbf{z} = \sum_{i=1}^{k} \mathrm{z}_i^2,$$

根据引理 1.90，随机向量 $\mathbf{z} \triangleq \boldsymbol{Q}^\top\mathbf{x} \in \mathbb{R}^k$ 服从均值为零、协方差矩阵为

$$\Sigma_{\mathbf{z}} = \boldsymbol{Q}^\top\mathbb{C}\mathrm{ov}[\mathbf{x}]\boldsymbol{Q} = \boldsymbol{Q}^\top\boldsymbol{I}\boldsymbol{Q} = \boldsymbol{I}_k.$$

因此，这些元素是相互独立的标准高斯随机变量。 ■

由于投影后的范数平方服从 $\chi^2_{(k)}$ 随机变量分布，我们可以直接应用定理 6.9 ，将 n 替换为 k。

推论 6.11：（对独立同分布标准高斯向量的投影 ℓ_2-范数的切尔诺夫尾界）令 $\mathcal{V}$ 为 k 维子空间，且该子空间属于 $\mathbb{R}^n$；设 $\mathbf{x} \in \mathbb{R}^n$为一个元素服从独立同分布高斯分布的向量。对于任意 $\varepsilon \in (0,1)$，我们有

$$\Pr\left(\sqrt{k(1-\varepsilon)} \le \|\mathcal{P}_{\mathcal{V}}\mathbf{x}\|_2 \le \sqrt{k(1+\varepsilon)}\right) \ge 1 - 2\exp\left(-\frac{k\varepsilon^2}{8}\right).$$

这可由引理 6.10 和定理 6.9直接推出，只需将它们应用于 k 维高斯向量 $\boldsymbol{Q}^\top\mathbf{x}$。

6.3. 高斯随机矩阵下的随机投影

此前，我们证明了高斯随机向量的投影具有很强的集中性质。现在我们转而讨论高斯随机矩阵的一项关键性质。

在线性代数入门课程中，我们会学到，通过主成分分析（PCA）进行降维，涉及将数据投影到低维子空间上；这些子空间在尽可能保留能量（即方差）的意义下是最优的。

主方向对应于数据中变化最大的方向，可通过对数据矩阵进行奇异值分解（SVD）得到。然而，在处理流式数据时，或必须实时应用投影时，计算 SVD 的代价可能很高，甚至可能不可行。

在这类情形中，我们需要一种非自适应的主成分分析替代方法：在观测数据之前就选定投影矩阵。一种简单有效的方法是使用随机线性映射，以随机矩阵$\mathbf{X} \in \mathbb{R}^{n\times p}$ 表示，其元素从标准高斯分布中独立抽取。下面的引理表明，将这样的矩阵作用于任意固定的单位向量，会得到一个高斯随机向量。

引理 6.12：（独立同分布标准高斯矩阵下的随机投影）设$\mathbf{X} \in \mathbb{R}^{n\times p}$ 为一个元素独立同分布于标准高斯分布的矩阵。若 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是一个确定性向量，且 $\|\boldsymbol{\beta}\|_2 = 1$，则$\mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$ 是一个独立同分布的标准高斯随机向量。

证明 [引理 6.12] 对于每个 $i \in \{1, 2, \ldots, n\}$，$\mathbf{X}\boldsymbol{\beta}$的第 i个元素为$(\mathbf{X}\boldsymbol{\beta})_i = \langle \mathbf{x}^i, \boldsymbol{\beta}\rangle$，即 $\boldsymbol{\beta}$ 与第 i 行 $\mathbf{x}^i$ （视为 $\mathbb{R}^p$中的列向量）的内积。由于 $\mathbf{x}^i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$，且 $\boldsymbol{\beta}$是确定性的，根据引理 1.90，$\langle \mathbf{x}^i, \boldsymbol{\beta}\rangle$ 服从高斯分布。其均值为零，因为 $\mathbb{E}[\mathbf{x}^i] = \mathbf{0}$。方差为

$$\mathbb{V}\mathrm{ar}\left[\langle \mathbf{x}^i, \boldsymbol{\beta}\rangle\right] = \boldsymbol{\beta}^\top \mathbb{C}\mathrm{ov}[\mathbf{x}^i]\boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{I}\boldsymbol{\beta} = \|\boldsymbol{\beta}\|_2^2 = 1. \tag{6.13}$$

因此，$\mathbf{X}\boldsymbol{\beta}$ 的每个分量都是标准正态随机变量。此外，这些分量彼此独立，因为每个分量只取决于$\mathbf{X}$中各不相同的一行，而根据构造，这些行彼此独立。因此，$\mathbf{X}\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$。 ■

这一结果的一个直接推论是，$\mathbf{X}\boldsymbol{\beta}$ 的ℓ_2范数满足一个非渐近集中上界，其中 $\boldsymbol{\beta}$是任意固定的单位向量。

推论 6.13：（随机高斯投影的范数的切尔诺夫尾界）设$\mathbf{X} \in \mathbb{R}^{n\times p}$ 是一个元素独立同分布且服从标准高斯分布的矩阵。那么，对于任意具有单位范数的确定性 $\boldsymbol{\beta} \in \mathbb{R}^p$ 以及任意 $\varepsilon \in (0,1)$，有

$$\Pr\left(\sqrt{n(1-\varepsilon)} \le \|\mathbf{X}\boldsymbol{\beta}\|_2 \le \sqrt{n(1+\varepsilon)}\right) \ge 1 - 2\exp\left(-\frac{n\varepsilon^2}{8}\right). \tag{6.14}$$

这可由引理6.12直接得到；该引理表明$\mathbf{X}\boldsymbol{\beta}$ 是一个 n维的独立同分布标准高斯向量，再结合定理 6.9 应用于其平方范数即可。

6.4. 随机投影与 Johnson–Lindenstrauss 引理

当降维技术能够保留我们所关注的信息时，它们就很有用。在许多应用中，投影能够近似保留数据点之间的两两距离至关重要。该性质使我们能够在较低维度的空间中应用基于距离的算法——例如最近邻、聚类或核方法——而不会显著损失准确性。

以下结果称为 Johnson–Lindenstrauss（JL）引理，它保证随机投影能够恰好实现这一点：以高概率保持有限点集中所有点之间的两两距离，即使在非渐近情形下也是如此。值得注意的是，所需的目标维数 n 仅与点的数量 k 的对数成正比，而

完全不取决于环境维数 p。证明沿用了Johnson 等人 （1984年）的经典论证。

定理6.14：（Johnson–Lindenstrauss（JL）引理）设 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 为一个元素独立同分布且服从标准高斯分布的随机矩阵。设 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k \in \mathbb{R}^p$ 为任意一个固定的（确定性）向量集合。那么，对于任意的 $\delta \in (0,1)$ 以及任意一对 $i \neq j$,

$$(1-\delta)\left\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\right\|_2^2 \leq \left\|\frac{1}{\sqrt{n}}\mathbf{X}(\boldsymbol{\beta}_i - \boldsymbol{\beta}_j)\right\|_2^2 \leq (1+\delta)\left\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\right\|_2^2, \tag{6.15}$$

只要满足以下条件，事件发生的概率至少为 $1-\frac{1}{k}$

$$n \geq \frac{24\ln(k)}{\delta^2}. \tag{6.16}$$

定理 [的证明：6.14]我们分析 $\boldsymbol{X}$ 对归一化差向量的作用

$$\boldsymbol{\alpha}_{ij} \triangleq \frac{\boldsymbol{\beta}_i - \boldsymbol{\beta}_j}{\left\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\right\|_2}, \quad 1 \leq i < j \leq k.$$

这些向量满足 $\|\boldsymbol{\alpha}_{ij}\|_2 = 1$，其中 $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_j$ （如果 $\boldsymbol{\beta}_i = \boldsymbol{\beta}_j$，则距离为零，因此显然能保持不变）。定义集中事件

$$\mathcal{E}_{ij} = \left\{n(1-\delta) < \|\mathbf{X}\boldsymbol{\alpha}_{ij}\|_2^2 < n(1+\delta)\right\}, \quad 1 \leq i < j \leq k.$$

根据推论6.13，每个事件 $\mathcal{E}_{ij}$ 在条件（6.16）下以高概率成立：

$$\Pr\left(\mathcal{E}_{ij}^c\right) \leq 2\exp\left(-\frac{n\delta^2}{8}\right) \leq \frac{2}{k^3}, \quad \text{if } n \geq \frac{24\ln(k)}{\delta^2}.$$

Ther存在 $\binom{k}{2} = k(k-1)/2$ 个这样的数对。应用并集界（定理 1.61），

$$\Pr\left(\bigcup_{i,j}\mathcal{E}_{ij}^c\right) \leq \sum_{i,j}\Pr\left(\mathcal{E}_{ij}^c\right) \leq \frac{k(k-1)}{2}\cdot\frac{2}{k^3} \leq \frac{k-1}{k^2} = \frac{1}{k}.$$

因此，以至少 $1-1/k$ 的概率，不等式（6.15）对所有数对$i \neq j$同时成立。证明完毕。 ■

许多算法——例如最近邻搜索、聚类和核方法——都严重依赖数据的几何特性，尤其是两两距离。JL引理保证，将数据随机投影到维数为 $n \ll p$ 的空间后，这些距离仍基本保持不变。因此，尽管计算现在是在小得多的空间中进行，算法的行为仍得以保留。

本质上，JL引理将准确性与环境维度解耦：随机性使数据得以极度压缩，同时保持几何保真度。这使随机算法——例如高斯草图——兼具速度快、可扩展、易于实现且理论可靠等优点；这种组合在算法设计中既罕见又强大（Woodruff 等人，2014年；Lu，2021a，b）。

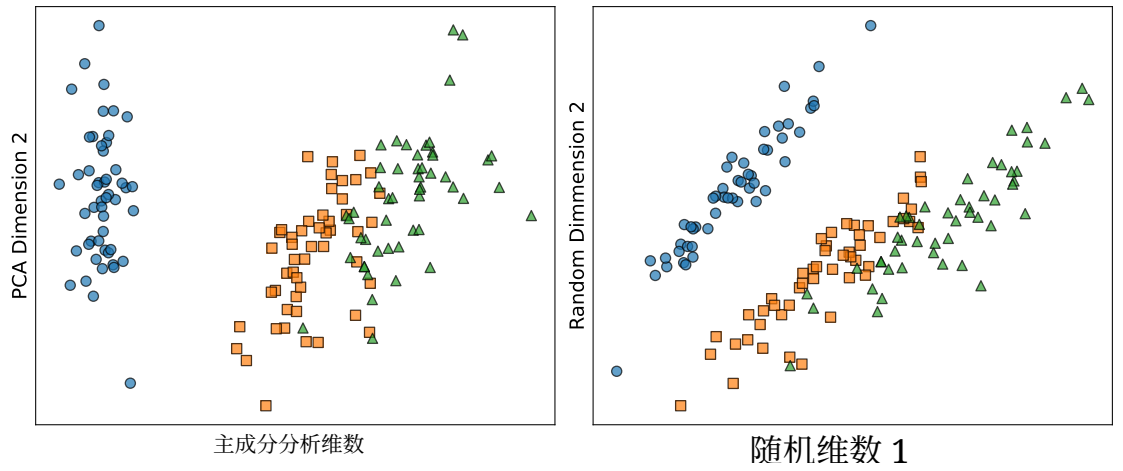


图6.1：使用（左）两个随机高斯方向和（右）主成分分析对鸢尾花数据集进行的近似二维投影。每个点代表一个鸢尾花样本，并按物种着色。

例6.15（通过随机投影降维）。我们考察由Fisher (1936）提出的鸢尾花数据集。这是一个基准多元数据集，包含来自三种鸢尾花物种的150个样本——Iris setosa、Iris versicolor和Iris virginica（每种50个样本）。每个样本由四个连续特征描述：萼片长度、萼片宽度、花瓣长度和花瓣宽度（单位均为厘米）。该数据集因其清晰的类别结构而广泛应用于机器学习和统计学：Iris setosa与另外两种物种线性可分，而Iris versicolor和Iris virginica则部分重叠，因此非常适合用于说明分类、聚类和降维技术。目标是将数据投影到二维空间以便可视化。在图6.1中，我们比较了随机投影数据与基于主成分分析的降维结果，随机投影的做法是应用一个 2×4独立同分布标准高斯矩阵。就保留不同鸢尾花物种类别而言，随机投影与主成分分析一样，能够有效保留数据中的结构。□

这一结果还揭示了集中不等式（6.14）与 JL引理之间的密切联系。由于（6.14）通过定理6.17 和6.19蕴含限制等距性质，因此从这个意义上说，可以认为 JL引理蕴含限制等距性质。

我们现在证明一个部分逆命题：在适当条件下，可以通过对矩阵各列的符号函数进行随机化，将满足限制等距性质的矩阵变成针对点集 $\mathbb{S}$ 的 JL嵌入。事实上，若不进行这种符号随机化，这一结论便不成立。对列符号进行随机化，可确保点集 $\mathbb{S}$ 落入（或几乎落入） $\boldsymbol{X}\boldsymbol{D}$ 的零空间的概率呈指数级小。

定理 6.16 设 $\mathbb{S} \subset \mathbb{R}^p$ 为基数为 $|\mathbb{S}| = S$的有限集。取定 $\delta \in (0,1)$。设$\mathbf{X} \in \mathbb{R}^{n\times p}$ 为一个满足阶为 $2k$ 、常数为 $\delta_{2k} \leq \delta/4$ 的限制等距性质的矩阵，其中 $k \geq 16$为某个整数，且 $k \geq 16 \ln(4S/\varepsilon)$。那么，概率至少为 $1-\varepsilon$

$$(1-\delta)\left\|\boldsymbol{\beta}\right\|_2^2 \leq \left\|\mathbf{X}\mathbf{D}\boldsymbol{\beta}\right\|_2^2 \leq (1+\delta)\left\|\boldsymbol{\beta}\right\|_2^2, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{S},$$

其中 $\mathbf{D} \triangleq \operatorname{diag}(\boldsymbol{\epsilon}) \in \mathbb{R}^{p\times p}$，且 $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_p]^\top$ 是一个拉德马赫向量（即，每个 ϵ_i都是一个独立随机变量，以相同的概率取值 $+1$ 和 -1 ）。

证明 [定理 6 的16] 不失一般性，假设 $\mathbb{S}$中的所有向量都已归一化：$\|\boldsymbol{\beta}\|_2 = 1$。任取一个 $\boldsymbol{\beta} \in \mathbb{S}$。我们将 $\boldsymbol{\beta}$ 按照其元素的幅值分成大小为 k 的若干块。具体而言，令 $\mathbb{T}_1 \subset \{1, 2, \ldots, p\}$ 索引 k中绝对值最大的 $\boldsymbol{\beta}$个元素，$\mathbb{T}_2 \subset \{1, 2, \ldots, p\} \setminus \mathbb{T}_1$ 索引剩余坐标中接下来最大的 k个元素，依此类推。由此得到分解 $\mathbf{XD}\boldsymbol{\beta} = \mathbf{XD}\sum_i \boldsymbol{\beta}(\mathbb{T}_i)$。[2]现在展开平方范数：

$$\begin{aligned}\|\mathbf{XD}\boldsymbol{\beta}\|_2^2 &= \left\|\mathbf{XD}\sum_i \boldsymbol{\beta}(\mathbb{T}_i)\right\|_2^2 \\ &= \sum_i \|\mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_i)\|_2^2 + 2\left\langle \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_1), \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_1^c)\right\rangle + \sum_{\substack{i,j\geq 2\\ i\neq j}} \left\langle \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_i), \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_j)\right\rangle . \end{aligned} \tag{6.17}$$

上界第一项。由于$\mathbf{X}$ 满足阶为 $2k$ 、常数为 $\delta_{2k} \leq \delta/4$的限制等距性质，并且 $\|\mathbf{D}\boldsymbol{\beta}(\mathbb{T}_i)\|_2 = \|\boldsymbol{\beta}_{\mathbb{T}_i}\|_2$，第一项满足

$$(1-\delta/4)\|\boldsymbol{\beta}\|_2^2 = (1-\delta/4)\sum_i \|\boldsymbol{\beta}_{\mathbb{T}_i}\|_2^2 \leq \sum_i \|\mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_i)\|_2^2 \leq (1+\delta/4)\|\boldsymbol{\beta}\|_2^2 .$$

上界第二项。为了上界（6.17）中的第二项，我们定义随机变量

$$\mathrm{y} \triangleq \left\langle \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_1), \mathbf{XD}\boldsymbol{\beta}(\mathbb{T}_1^c)\right\rangle = \left\langle \mathbf{u}, \boldsymbol{\epsilon}_{\mathbb{T}_1^c}\right\rangle = \sum_{i\notin\mathbb{T}_1} \epsilon_i \mathrm{u}_i,$$

其中$\mathbf{u} \triangleq \mathrm{diag}(\boldsymbol{\beta}_{\mathbb{T}_1^c})\mathbf{X}_{\mathbb{T}_1^c}^\top \mathbf{X}_{\mathbb{T}_1}\mathbf{D}_{\mathbb{T}_1}\boldsymbol{\beta}_{\mathbb{T}_1} \in \mathbb{R}^{|\mathbb{T}_1^c|}$。[3] 由于$\mathbf{u}$ 和$\boldsymbol{\epsilon}_{\mathbb{T}_1^c}$ 相互独立，拉德马赫序列的霍夫丁不等式（推论1.75）表明

$$\Pr(|\mathrm{y}| \geq \mu) \leq 2\exp\left(-\frac{\mu^2}{2\|\mathbf{u}\|_2^2}\right). \tag{6.18}$$

现在我们对 $\|\mathbf{u}\|_2$ 进行上界估计。根据对偶范数的定义（1.3）、柯西–施瓦茨不等式以及谱范数的次乘性（定理1.33），可得

$$\begin{aligned}\|\mathbf{u}\|_2 &= \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{a}, \mathbf{u}\rangle = \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{i\geq 2} \left\langle \boldsymbol{a}_{\mathbb{T}_i}, \mathrm{diag}(\boldsymbol{\beta}_{\mathbb{T}_i})\mathbf{X}_{\mathbb{T}_i}^\top \mathbf{X}_{\mathbb{T}_1}\mathbf{D}_{\mathbb{T}_1}\boldsymbol{\beta}_{\mathbb{T}_1}\right\rangle \\ &\overset{\dagger}{\leq} \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{i\geq 2} \left\|\mathbf{X}_{\mathbb{T}_i}^\top \mathbf{X}_{\mathbb{T}_1}\right\|_2 \|\boldsymbol{a}_{\mathbb{T}_i}\|_2 \|\boldsymbol{\beta}_{\mathbb{T}_i}\|_\infty \|\boldsymbol{\beta}_{\mathbb{T}_1}\|_2 \overset{\ddagger}{\leq} \frac{\delta_{2k}}{\sqrt{k}} \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{i\geq 2} \|\boldsymbol{a}_{\mathbb{T}_i}\|_2 \left\|\boldsymbol{\beta}_{\mathbb{T}_{i-1}}\right\|_2 \\ &\leq \frac{\delta_{2k}}{\sqrt{k}} \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{i\geq 2} \frac{1}{2}\left(\|\boldsymbol{a}_{\mathbb{T}_i}\|_2^2 + \left\|\boldsymbol{\beta}_{\mathbb{T}_{i-1}}\right\|_2^2\right) \leq \frac{\delta_{2k}}{\sqrt{k}},\end{aligned}$$

其中，不等式（†）成立是因为 $\|\mathrm{diag}(\boldsymbol{\beta})\|_2 = \|\boldsymbol{\beta}\|_\infty$ (对角矩阵 $\mathrm{diag}(\boldsymbol{\beta})$ 的最大奇异值为 $\|\boldsymbol{\beta}\|_\infty$)，且 $\|\boldsymbol{D}_{\mathbb{T}_1}\|_2 = \|\boldsymbol{\epsilon}\|_\infty = 1$。不等式（‡）由 $\|\boldsymbol{\beta}_{\mathbb{T}_i}\|_\infty \leq k^{-1/2}\|\boldsymbol{\beta}_{\mathbb{T}_{i-1}}\|_2$ 根据对有序集合进行划分的构造可得$\mathbb{T}_1$，$\mathbb{T}_2$，……（引理 4.47），$\|\mathbf{X}_{\mathbb{T}_i}^\top \mathbf{X}_{\mathbb{T}_1}\|_2 \leq \delta_{2k}$当 $i \geq 2$时由限制等距性质（命题 4.36）的标准推论得到，且 $\|\boldsymbol{\beta}_{\mathbb{T}_1}\|_2 \leq \|\boldsymbol{\beta}\|_2 = 1$。因此，（6.18）可写为

$$\Pr(|\mathrm{y}| \geq \mu) \leq 2\exp\left(-\frac{\mu^2}{2\|\mathbf{u}\|_2^2}\right) \leq 2\exp\left(-\frac{k\mu^2}{2\delta_{2k}^2}\right) \leq 2\exp\left(-\frac{8k\mu^2}{\delta^2}\right). \tag{6.19}$$

2。再次注意，$\boldsymbol{\beta}(\mathbb{S}) \in \mathbb{R}^p$ 和 $\boldsymbol{\beta}_{\mathbb{S}} \in \mathbb{R}^{|\mathbb{S}|}$；参见定义1.1。3。注意，$\boldsymbol{D}\boldsymbol{\beta} = \mathrm{diag}(\boldsymbol{\beta})\boldsymbol{D}$。

对第三项进行上界估计。考虑剩余的交叉项，我们定义如下随机变量

$$\mathrm{z} \triangleq \sum_{\substack{i,j\geq 2\\ i\neq j}} \langle \mathbf{X}\mathbf{D}\boldsymbol{\beta}(\mathbb{T}_i), \mathbf{X}\mathbf{D}\boldsymbol{\beta}(\mathbb{T}_j)\rangle = \sum_{i,j\in\{1,2,\ldots,p\}} \epsilon_i\epsilon_j \mathrm{a}_{ij} = \boldsymbol{\epsilon}^\top \mathbf{A}\boldsymbol{\epsilon},$$

其中 $\mathbf{A} \in \mathbb{R}^{p\times p}$ 是一个对角元素为零、非对角元素如下的对称矩阵：

$$\mathrm{a}_{ij} = \begin{cases} \beta_i \mathbf{x}_i^\top \mathbf{x}_j \beta_j & \text{if } i,j \in \mathbb{T}_1^{\mathrm{c}} \text{ and } i,j \text{ are contained in different blocks } \mathbb{T}_\ell; \\ 0 & \text{otherwise.} \end{cases}$$

表达式 $\boldsymbol{\epsilon}^\top\mathbf{A}\boldsymbol{\epsilon}$是二阶齐次拉德马赫混沌。为了给 $|\mathrm{z}|$确定上界，我们应用问题 6.7，这需要估计 $\|\mathbf{A}\|_2$ 和 $\|\mathbf{A}\|_F$。由于 $\mathbf{A}$ 是对称的，我们可以像之前一样利用 RIP 和块结构求得其谱范数：

$$\begin{aligned}
\|\mathbf{A}\|_2 &= \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \mathbf{A}\boldsymbol{a}, \boldsymbol{a}\rangle = \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{\substack{i,j\geq 2\\ i\neq j}} \left\langle \boldsymbol{a}_{\mathbb{T}_i}, \mathrm{diag}(\boldsymbol{\beta})_{\mathbb{T}_i} \mathbf{X}_{\mathbb{T}_i}^\top \mathbf{X}_{\mathbb{T}_j} \mathrm{diag}(\boldsymbol{\beta})_{\mathbb{T}_j} \boldsymbol{a}_{\mathbb{T}_j} \right\rangle \\
&\leq \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{\substack{i,j\geq 2\\ i\neq j}} \|\boldsymbol{a}_{\mathbb{T}_i}\|_2 \left\|\boldsymbol{a}_{\mathbb{T}_j}\right\|_2 \left\|\boldsymbol{\beta}_{\mathbb{T}_i}\right\|_\infty \left\|\boldsymbol{\beta}_{\mathbb{T}_j}\right\|_\infty \left\|\mathbf{X}_{\mathbb{T}_i}^\top \mathbf{X}_{\mathbb{T}_j}\right\|_2 \\
&\leq \delta_{2k} \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{\substack{i,j\geq 2\\ i\neq j}} \|\boldsymbol{a}_{\mathbb{T}_i}\|_2 \left\|\boldsymbol{a}_{\mathbb{T}_j}\right\|_2 k^{-1/2} \left\|\boldsymbol{\beta}_{\mathbb{T}_{i-1}}\right\|_2 k^{-1/2} \left\|\boldsymbol{\beta}_{\mathbb{T}_{j-1}}\right\|_2 \\
&\leq \frac{\delta_{2k}}{k} \max_{\|\boldsymbol{a}\|_2\leq 1} \sum_{\substack{i,j\geq 2\\ i\neq j}} \left(\frac{\left\|\boldsymbol{\beta}_{\mathbb{T}_{i-1}}\right\|_2^2 + \|\boldsymbol{a}_{\mathbb{T}_i}\|_2^2}{2}\right)\left(\frac{\left\|\boldsymbol{\beta}_{\mathbb{T}_{j-1}}\right\|_2^2 + \left\|\boldsymbol{a}_{\mathbb{T}_j}\right\|_2^2}{2}\right) \leq \frac{\delta_{2k}}{k} \leq \frac{\delta}{4k}.
\end{aligned}$$

对于弗罗贝尼乌斯范数，注意到

$$\begin{aligned}
\|\mathbf{A}\|_F^2 &= \sum_{\substack{s,t\geq 2\\ s\neq t}} \sum_{i\in\mathbb{T}_s} \sum_{j\in\mathbb{T}_t} (\beta_i \mathbf{x}_i^\top \mathbf{x}_j \beta_j)^2 = \sum_{\substack{s,t\geq 2\\ s\neq t}} \sum_{i\in\mathbb{T}_s} \beta_i^2 \mathbf{x}_i^\top \left(\sum_{j\in\mathbb{T}_t} \beta_j \mathbf{x}_j \beta_j \mathbf{x}_j^\top\right) \mathbf{x}_i \\
&= \sum_{\substack{s,t\geq 2\\ s\neq t}} \sum_{i\in\mathbb{T}_s} \beta_i^2 \left\|\mathrm{diag}(\boldsymbol{\beta}_{\mathbb{T}_t}) \mathbf{X}_{\mathbb{T}_t}^\top \mathbf{x}_i\right\|_2^2 \leq \sum_{\substack{s,t\geq 2\\ s\neq t}} \sum_{i\in\mathbb{T}_s} \beta_i^2 \left\|\boldsymbol{\beta}_{\mathbb{T}_t}\right\|_\infty^2 \left\|\mathbf{X}_{\mathbb{T}_t}^\top \mathbf{x}_i\right\|_2^2 \\
&\overset{\dagger}{\leq} \delta_{2k}^2 \sum_{\substack{s,t\geq 2\\ s\neq t}} \left\|\boldsymbol{\beta}_{\mathbb{T}_s}\right\|_2^2 k^{-1} \left\|\boldsymbol{\beta}_{\mathbb{T}_t}\right\|_2^2 \leq \frac{\delta_{2k}^2}{k} \leq \frac{\delta^2}{16k},
\end{aligned}$$

其中，不等式（†）源于以下事实：根据 RIP 的标准推论（命题 4.36），$\left\|\mathbf{X}_{\mathbb{T}_t}^\top \mathbf{x}_i\right\|_2 = \left\|\mathbf{X}_{\mathbb{T}_t}^\top \mathbf{x}_i\right\|_2 \leq \delta_{k+1} \leq \delta_{2k}$。问题6.7表明 $|\mathrm{z}|$ 满足

$$\Pr(|\mathrm{z}| \geq \tau) \leq 2\exp\left(-\min\left\{\frac{3\tau^2}{128\,\|\mathbf{A}\|_F^2}, \frac{\tau}{32\,\|\mathbf{A}\|_2}\right\}\right) \leq 2\exp\left(-k\min\left\{\frac{3\tau^2}{8\delta^2}, \frac{\tau}{8\delta}\right\}\right).$$

合并各个上界。选择 $\mu \triangleq \delta/4$ 和 $\tau \triangleq \delta/2$，并将三个估计值代入（6.17），可得

$$(1-\delta)\|\boldsymbol{\beta}\|_2^2 \leq \|\mathbf{X}\mathbf{D}\boldsymbol{\beta}\|_2^2 \leq (1+\delta)\|\boldsymbol{\beta}\|_2^2, \quad \text{for any } \boldsymbol{\beta} \in \mathbb{S} \tag{6.20}$$

以至少……的概率

$$1-2\exp(-k/2)-2\exp\left(-k\min\{3/32,1/16\}\right)\geq 1-4\exp(-k/16).$$

最后，对所有 S 点应用 $\mathbb{S}$上的并集界，可知该不等式以至少 $1-4S\exp(-k/16)\geq 1-\varepsilon$ 的概率同时对所有 $\boldsymbol{\beta}\in\mathbb{S}$ 成立，条件是$k\geq 16\ln(4S/\varepsilon)$。证明至此完成。 ■

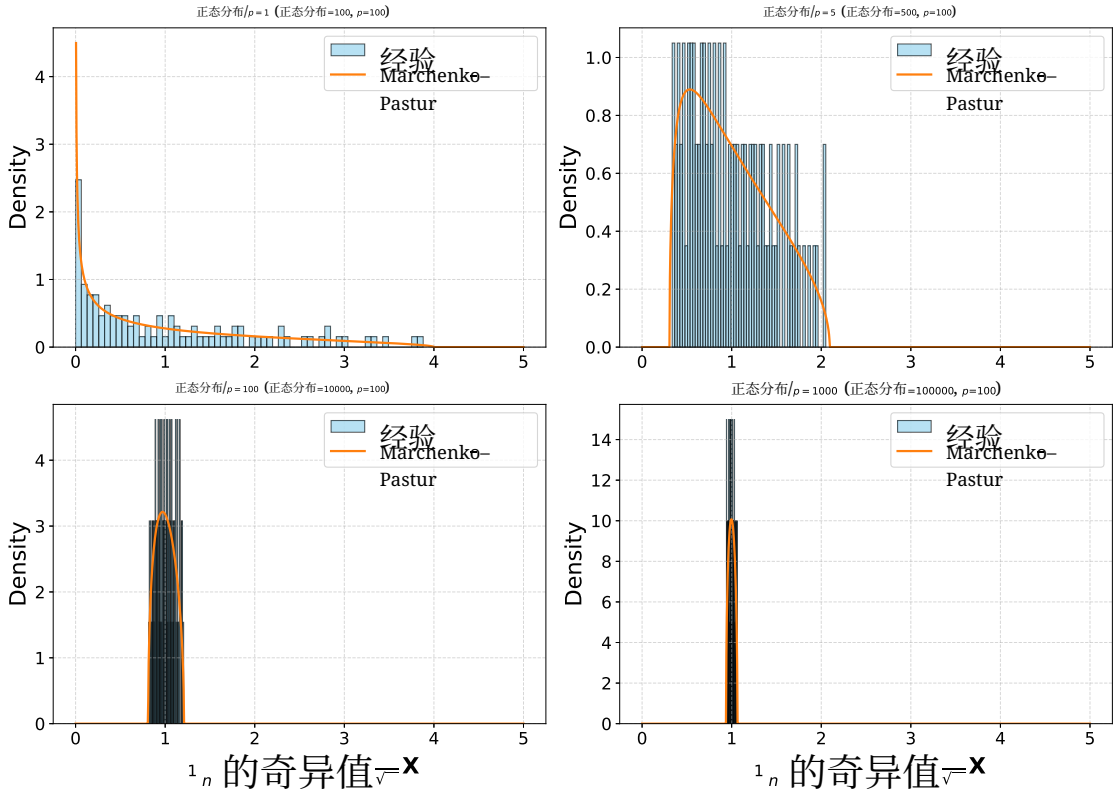


图 6.2： 随着长宽比 n/p 增大，具有独立同分布的标准高斯元素的 $n\times p$ 矩阵的奇异值经验分布。同时给出了每个经验分布对应的理论 Marchenko–Pastur 分布。

6.5. 高斯随机矩阵的奇异值

在本节中，我们分析元素为独立同分布标准高斯随机变量的随机矩阵的奇异值。具体而言，考虑一个 $n\times p$ 矩阵 $\boldsymbol{X}$ 满足 $n>p$。数值实验表明，随着 n 增大（且 p固定），所有 p 的奇异值 $\boldsymbol{X}$都紧密集中在 n附近，如图6.2所示。因此，

$$\boldsymbol{X}\approx\boldsymbol{U}(\sqrt{n}\boldsymbol{I})\boldsymbol{V}^\top=\sqrt{n}\boldsymbol{U}\boldsymbol{V}^\top, \tag{6.21}$$

这意味着 $\boldsymbol{X}$ 的行为类似于经过缩放的正交矩阵。从几何角度看，这表明如果我们在维数不断增大的高维环境空间中抽取固定数量的独立同分布高斯向量，那么随着维数增长，以高概率这些向量会变得近似正交。

以下结果为此类矩阵的极端奇异值提供了一个非渐近概率上界。证明采用了Vershynin（2010年）中的覆盖数论证，该论证足够灵活，可以推广到其他次高斯分布及相关情形。

定理 6.17：（高斯矩阵的奇异值）设$\mathbf{X} \in \mathbb{R}^{n\times k}$是一个具有独立同分布的标准高斯元素的随机矩阵，其中 $n > k$。对于任意固定的 $\varepsilon \in$ （0，1），$\mathbf{X}$ 的最小和最大奇异值满足

$$\sqrt{n}(1-\varepsilon) \le \sigma_k \le \sigma_1 \le \sqrt{n}(1+\varepsilon) \tag{6.22}$$

以至少 $1 - 2\left(\frac{12}{\varepsilon}\right)^k \exp\left(-\frac{n\varepsilon^2}{32}\right)$ 的概率，或者只要满足以下条件，以至少 $1 - 1/k$ 的概率

$$n > \frac{32}{\varepsilon^2} \ln\left(2k\left(\frac{12}{\varepsilon}\right)^k\right). \tag{6.23}$$

该定理给出了一个具体公式，说明 n 需要有多大（相对于 k 和 ε），才能以至少 $1-1/k$ 的概率保证这种行为。根据奇异值的变分刻画（见（1.29）），（6.22）中的界等价于

$$\sqrt{n}(1-\varepsilon) < \|\mathbf{X}\boldsymbol{\beta}\|_2 < \sqrt{n}(1+\varepsilon), \quad \text{for all } \boldsymbol{\beta} \in \mathbb{U}^{k-1}, \tag{6.24}$$

其中 $\mathbb{U}^{k-1} \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^k \mid \|\boldsymbol{\alpha}\|_2 = 1\}$ 是 $\mathbb{R}^k$中的 k维球面，包含 $\mathbb{R}^k$中的单位 ℓ_2范数向量。由于这个集合不可数无限，直接使用并集界——例如 JL引理的证明中所用的方法——并不适用。因此，我们采用基于网的论证。

定理 [6 的证明。17] 令 $\varepsilon_1 \triangleq \varepsilon/4$ 且 $\varepsilon_2 \triangleq \varepsilon/2$。考虑 ε_1-打包集 $\mathbb{P}_{\varepsilon_1}$，它由$\mathbb{U}^{k-1}$ 按推论 6.4下方所述方式构造。对于任意固定的 $\boldsymbol{\beta} \in \mathbb{U}^{k-1}$，定义事件

$$\mathcal{A}_{\boldsymbol{\beta},\varepsilon_2} \triangleq \left\{n(1-\varepsilon_2)\|\boldsymbol{\beta}\|_2^2 \le \|\mathbf{X}\boldsymbol{\beta}\|_2^2 \le n(1+\varepsilon_2)\|\boldsymbol{\beta}\|_2^2\right\}.$$

根据推论 6.13，对于任意固定的 $\boldsymbol{\beta} \in \mathbb{U}^{k-1}$，$\mathbf{Pr}\left(\mathcal{A}^c_{\boldsymbol{\beta},\varepsilon_2}\right) \le 2\exp\left(-n\varepsilon^2/32\right)$。对有限集 1.61)（定理 $\mathbb{P}_{\varepsilon_1}$ 应用并集界可得

$$\Pr\left(\bigcup_{\boldsymbol{\beta}\in\mathbb{P}_{\varepsilon_1}} \mathcal{A}^c_{\boldsymbol{\beta},\varepsilon_2}\right) \le \sum_{\boldsymbol{\beta}\in\mathbb{P}_{\varepsilon_1}} \Pr\left(\mathcal{A}^c_{\boldsymbol{\beta},\varepsilon_2}\right) \le |\mathbb{P}_{\varepsilon_1}| \cdot \Pr\left(\mathcal{A}^c_{\boldsymbol{\beta},\varepsilon_2}\right) \le 2\left(\frac{12}{\varepsilon}\right)^k \exp\left(-\frac{n\varepsilon^2}{32}\right) \le \frac{1}{k},$$

其中 $|\mathbb{P}_{\varepsilon_1}| \le (12/\varepsilon)^k$ 根据推论 6.4，最后一个不等式由条件（6.23）得出。

因此，以至少 $1 - 1/k$ 的概率，事件 $\mathcal{A}_{\boldsymbol{\beta},\,\varepsilon_2}$ 对所有$\boldsymbol{\beta} \in \mathbb{P}_{\varepsilon_1}$同时成立。为完成证明，我们需要说明，若 $\cup_{\boldsymbol{\beta}\in\mathbb{P}_{\varepsilon_1}} \mathcal{A}^c_{\boldsymbol{\beta},\,\varepsilon_2}$ 成立，则该上界对 $\mathbb{U}^{k-1}$ 中的每个元素都成立，而不仅仅对打包集中的元素成立 $\mathbb{P}_{\varepsilon_1}$。

对于球面上的任意向量 $\boldsymbol{\beta} \in \mathbb{U}^{k-1}$，在 $\varepsilon/4$-打包集中存在一个向量 $\boldsymbol{\alpha} \in \mathbb{P}_{\varepsilon_1}$ 使得 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 \le \varepsilon_1 = \varepsilon/4$。假设良好事件 $\left(\cup_{\boldsymbol{\alpha}\in\mathbb{P}_{\varepsilon_1}} \mathcal{A}^c_{\boldsymbol{\alpha},\,\varepsilon_2}\right)^c$ 成立（其概率至少为 $1-1/k$），根据三角不等式和谱范数的定义（定义1.31），可得

$$\|\mathbf{X}\boldsymbol{\beta}\|_2 \le \|\mathbf{X}\boldsymbol{\alpha}\|_2 + \|\mathbf{X}(\boldsymbol{\beta}-\boldsymbol{\alpha})\|_2 \le \sqrt{n}\left(1+\frac{\varepsilon}{2}\right) + \sigma_1\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2 \le \sqrt{n}\left(1+\frac{\varepsilon}{2}\right) + \frac{\sigma_1\varepsilon}{4},$$

第二个不等式成立是因为我们假设 $\cup_{\boldsymbol{\alpha}\in\mathbb{P}_{\varepsilon_1}}\mathcal{A}^c_{\boldsymbol{\alpha},\,\varepsilon_2}$ 成立。谱范数的定义还意味着，对于球面 $\mathbb{U}^{k-1}$ 上的所有 $\boldsymbol{\beta}$， σ_1 是 $\|\mathbf{X}\boldsymbol{\beta}\|_2$ 的最小上界，因此上述不等式中的上界不能更小：

$$\sigma_1 \le \sqrt{n}\left(1+\frac{\varepsilon}{2}\right)+\frac{\sigma_1\varepsilon}{4}.$$

整理后可得

$$\sigma_1 \le \sqrt{n}\left(\frac{1+\varepsilon/2}{1-\varepsilon/4}\right) = \sqrt{n}\left(1+\varepsilon-\frac{\varepsilon(1-\varepsilon)}{4-\varepsilon}\right) \le \sqrt{n}(1+\varepsilon). \tag{6.25}$$

对于下界，在同一良好事件 $(\cup_{\boldsymbol{\alpha}\in\mathbb{P}_{\varepsilon_1}}\mathcal{A}^c_{\boldsymbol{\alpha},\,\varepsilon_2})^{\,c}$ （其概率至少为 $1-1/k$）下，三角不等式还意味着

$$\begin{aligned}\|\mathbf{X}\boldsymbol{\beta}\|_2 &\ge \|\mathbf{X}\boldsymbol{\alpha}\|_2 - \|\mathbf{X}(\boldsymbol{\beta}-\boldsymbol{\alpha})\|_2 \ge \sqrt{n}\left(1-\frac{\varepsilon}{2}\right)-\sigma_1\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2\\ &\ge \sqrt{n}\left(1-\frac{\varepsilon}{2}\right)-\frac{\varepsilon}{4}\sqrt{n}(1+\varepsilon) \ge \sqrt{n}(1-\varepsilon),\end{aligned}$$

其中，倒数第二个不等式由（6.25）推出，最后一个不等式则由 $\varepsilon<1$ 得出。根据奇异值性质（1.29）， σ_k 是对球面 $\mathbb{U}^{k-1}$ 上所有 $\boldsymbol{\beta}$ 而言， $\|\mathbf{X}\boldsymbol{\beta}\|_2$ 的最大下界，因此只要（$\cup_{\boldsymbol{\alpha}\in\mathbb{P}_{\varepsilon_1}}\mathcal{A}^c_{\boldsymbol{\alpha},\,\varepsilon_2}$）^{c} 成立，就有 $\sigma_k \ge \sqrt{n}(1-\varepsilon)$。证明完毕。 ■

6.6. 高斯随机矩阵下的稀疏优化

有了这些工具，我们便能轻松建立压缩感知问题中高斯随机矩阵的恢复保证。

证明压缩感知对给定信号类有效的一种常见策略是构造一个对偶证书（参见定理3.19及其在定理5.13中的应用，该应用利用 RIP 找到证书），并验证它对该类中的每个信号都有效。我们使用高斯测量矩阵来说明这种方法。

定理 6.18：（标准高斯矩阵下 (P_1) 的精确恢复 （Bara- niuk 等人，2008年）） 设 $\mathbf{X}\in\mathbb{R}^{n\times p}$ 是一个元素服从独立同分布标准高斯分布的随机矩阵，并令测量值 $\boldsymbol{y}=\mathbf{X}\boldsymbol{\beta}^*\in\mathbb{R}^n$，其中 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 是 k-稀疏的（即至多有 k 个非零项）。那么， $\boldsymbol{\beta}^*$ 以至少 ℓ_1-最小化问题（P_1）（第 129 页）的概率，是它的唯一解；前提是测量数量满足

$$n \ge Ck\ln p, \tag{6.26}$$

其中 C 为某个通用常数。

定理 [的证明6.18] 该证明依赖于 ℓ_1-最小化的对偶证书条件（定理 3.19）。根据该结果，只需证明：对于任意满足 $\mathbb{S}\subseteq\{1,2,\dots,p\}$ 的支撑集 $|\mathbb{S}|=k$，以及任意可能的非零项符号模式 $\boldsymbol{w}\triangleq\operatorname{sign}(\boldsymbol{\beta}^*_{\mathbb{S}})\in\mathbb{R}^k$， $\boldsymbol{\beta}^*$都存在一个有效的对偶证书 $\boldsymbol{\lambda}\in\mathbb{R}^n$ 满足

$$\mathbf{X}_{\mathbb{S}}^{\top}\boldsymbol{\lambda}=\boldsymbol{w}.$$

我们将候选证书构造为最小二乘解（见（3.76））： $\boldsymbol{\lambda}_{\mathrm{LS}} = \mathbf{X}_{\mathbb{S}}(\mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}})^{-1}\boldsymbol{w}$。为了控制 $\boldsymbol{\lambda}_{\mathrm{LS}}$，我们借助定理$n \times k$ 6.17 给出的固定子矩阵奇异值上界。令 $\varepsilon \triangleq 0.5$，记 $\mathcal{E}$ 为事件

$$\mathcal{E} \triangleq \{0.5\sqrt{n} \le \sigma_k \le \sigma_1 \le 1.5\sqrt{n}\},$$

其中

$$\Pr(\mathcal{E}) \ge 1 - 2\left(\frac{12}{\varepsilon}\right)^{k} \exp\left(-\frac{n\varepsilon^2}{32}\right) \approx 1 - \exp\left(-C'\frac{n}{k}\right)$$

其中某个绝对常数为 $C' > 0$。在给定 $\mathcal{E}$ 的条件下，子矩阵 $\mathrm{X}_{\mathbb{S}}$ 满秩，且 $\mathrm{X}_{\mathbb{S}}^{\top}\mathrm{X}_{\mathbb{S}}$ 可逆，因此 $\boldsymbol{\lambda}_{\mathrm{LS}}$ 良态，并保证对于 $i \in \mathbb{S}$，$\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}} = \mathsf{sign}(\beta_i^*)$。为验证第二个条件 $\left\|\boldsymbol{X}_{\mathbb{S}^c}^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right\|_{\infty} < 1$ （参见定理 3.19），考虑任意索引$i \in \mathbb{S}^c$，并令 x_i 表示 X 的第 i 列。我们需要对所有索引 $i \in \mathbb{S}^c$ 给出 $\left|\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right|$ 的上界。令 $\mathrm{U}\boldsymbol{\Sigma}\mathrm{V}^{\top}$ 为 $\mathrm{X}_{\mathbb{S}}$ 的约化奇异值分解。在给定 $\mathcal{E}$ 的条件下，我们有

$$\|\boldsymbol{\lambda}_{\mathrm{LS}}\|_2 = \left\|\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\top}\boldsymbol{w}\right\|_2 \le \frac{\|\boldsymbol{w}\|_2}{\sigma_k} \le 2\sqrt{\frac{k}{n}}. \tag{6.27}$$

对于一个固定的 $i \in \mathbb{S}^c$ 以及一个固定向量 $\boldsymbol{v} \in \mathbb{R}^n$，$\mathrm{x}_i^{\top}\boldsymbol{v}/\|\boldsymbol{v}\|_2$ 根据高斯随机向量的定义和线性变换引理（引理1.63 或引理1.90），再由习题1.83可知

$$\Pr\left(\left|\mathrm{x}_i^{\top}\boldsymbol{v}\right| \ge 1\right) = \Pr\left(\frac{\left|\mathrm{x}_i^{\top}\boldsymbol{v}\right|}{\|\boldsymbol{v}\|_2} \ge \frac{1}{\|\boldsymbol{v}\|_2}\right) \le 2\exp\left(-\frac{1}{\|\boldsymbol{v}\|_2^2}/2\right).$$

注意，如果 $i \notin \mathbb{S}$，那么 x_i 和 $\boldsymbol{\lambda}_{\mathrm{LS}}$ 相互独立（它们依赖于$\mathbf{X}$中不同且因此相互独立的元素）。这意味着，由于（6.27），

$$\Pr\left(\left|\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right| \ge 1 \mid \mathcal{E}\right) = \Pr\left(\left|\mathrm{x}_i^{\top}\boldsymbol{v}\right| \ge 1\right) \le 2\exp\left(-\frac{n}{8k}\right), \quad \text{for } \|\boldsymbol{v}\|_2 \le 2\sqrt{\frac{k}{n}},\ i \notin \mathbb{S}.$$

因此，

$$\Pr\left(\left|\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right| \ge 1\right) \le \Pr\left(\left|\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right| \ge 1 \mid \mathcal{E}\right) + \Pr(\mathcal{E}^c) \le 2\exp\left(-\frac{n}{8k}\right) + \exp\left(-C'\frac{n}{k}\right).$$

对所有 $i \in \mathbb{S}^c$ （至多 p 个索引）应用并集界，

$$\Pr\left(\bigcup_{i\in\mathbb{S}^c}\left\{\left|\mathrm{x}_i^{\top}\boldsymbol{\lambda}_{\mathrm{LS}}\right| \ge 1\right\}\right) \le p\left(2\exp\left(-\frac{n}{8k}\right) + \exp\left(-C'\frac{n}{k}\right)\right).$$

选择常数 C，使得 $n \ge Ck\ln p$，即可确保该概率至多为 $\frac{1}{p}$。因此，以至少 $1-1/p$ 的概率，对所有 k-稀疏向量，对偶证书条件都成立，这意味着通过 ℓ_1-最小化可实现精确恢复。

■

除了构造对偶证书外，还可以使用 RIP 来保证精确恢复。虽然验证确定性矩阵的 RIP 在计算上难以处理——需要检查所有 $\binom{p}{k}$ 子矩阵——但可以证明，随机矩阵以高概率满足 RIP。接下来的两个定理将针对独立同分布高斯矩阵建立这一结论。

定理 6.19：（高斯矩阵的受限等距性质）设 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 具有独立同分布的标准高斯元素。对于任意 η，$\varepsilon \in (0,1)$，假设

$$n \geq 32\eta^{-2}\left(k\ln\left(\frac{ep}{k}\right)+k\ln\left(\frac{12}{\eta}\right)+\ln(2/\varepsilon)\right). \tag{6.28}$$

那么，至少以 $1-\varepsilon$ 的概率，缩放矩阵 $\frac{1}{\sqrt{n}}\mathbf{X}$ 的限制等距常数 δ_k 满足

$$\delta_k \leq \eta(2+\eta). \tag{6.29}$$

证明 [定理 6.19]。固定一个支撑集 $\mathbb{S}\subset\{1,2,\ldots,p\}$，使得 $|\mathbb{S}|=k$。子矩阵 $\mathbf{X}_{\mathbb{S}}$ 是一个 $n\times k$ 高斯矩阵。$\frac{1}{n}\mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}}-\boldsymbol{I}$ 的特征值位于 $[\sigma_{\min}^2-1,\sigma_{\max}^2-1]$，其中 $\sigma_{\min}$ 和 $\sigma_{\max}$ 分别是 $\frac{1}{\sqrt{n}}\mathbf{X}_{\mathbb{S}}$ 的最小和最大奇异值。记 $\widetilde{\mathbf{X}}_{\mathbb{S}} \triangleq \frac{1}{\sqrt{n}}\mathbf{X}_{\mathbb{S}}$。对于 $\eta>0$，定理 6.17 表明

$$\left\|\widetilde{\mathbf{X}}_{\mathbb{S}}^{\top}\widetilde{\mathbf{X}}_{\mathbb{S}}-\boldsymbol{I}\right\|_2 \leq \max\left\{(1+\eta)^2-1,1-(1-\eta)^2\right\}=2\eta+\eta^2,$$

至少以 $1-2\left(\frac{12}{\eta}\right)^k\exp\left(-\frac{n\eta^2}{32}\right)$ 的概率成立。对所有 $\binom{p}{k}$ 个可能的支撑集应用并集界，根据限制等距常数的定义（参见定义 4.30），$\overline{\delta_k}=\max_{\mathbb{S}\subset\{1,2,\ldots,p\},\ |\mathbb{S}|=k}\left\|\widetilde{\mathbf{X}}_{\mathbb{S}}^{\top}\widetilde{\mathbf{X}}_{\mathbb{S}}-\boldsymbol{I}\right\|_2$，可得

$$\Pr\left(\delta_k>2\eta+\eta^2\right)\leq 2\binom{p}{k}\left(\frac{12}{\eta}\right)^k\exp\left(-\frac{n\eta^2}{32}\right)\leq 2\left(\frac{ep}{k}\right)^k\left(\frac{12}{\eta}\right)^k\exp\left(-\frac{n\eta^2}{32}\right),$$

其中，第二个不等式由二项式系数的上界推出（见问题 6.2）。最后一项受 ε 支配，这是由条件（6.28）所致。证明至此完成。

■

由于 $\delta_k \leq 2\eta+\eta^2$，选取 $\eta=0.22$ 可得 $\delta_k \leq 0.4884 < 0.4931$（比较关于 ℓ_1 最小化的定理 5.21），在渐近情形下满足条件 $n \geq 662\big(k\ln(ep/k)+k\ln(12/\eta)+\ln(2/\varepsilon)\big)$。Foucart 和 Rauhut（2013年）通过改进定理 6.17 中的上界，将此上界改进为$n \geq 78.13\,(k\ln(ep/k)+\ln(2/\varepsilon))$；此处略去细节。

定理 6.20：（高斯矩阵下 (P_1) 的精确恢复）设 $\mathbf{X}\in\mathbb{R}^{n\times p}$为一个元素独立同分布且服从高斯分布的随机矩阵。假设 $k<p$ 且 $\varepsilon\in(0,1)$。若

$$n \geq C_1 k\ln(ep/k)+C_2k+C_3\ln(2/\varepsilon)$$

对于通用常数 C_1，C_2，$C_3>0$。那么，以至少 $1-\varepsilon$ 的概率，每个 k-稀疏向量 $\boldsymbol{\beta}$ 都可以通过求解 **测量值** $=\mathbf{X}\boldsymbol{\beta}$ 对应的 ℓ_1-最小化问题（P_1），从中精确恢复出来。

证明 [定理 6。20] 结合定理 6.19 （该定理建立了高斯矩阵的受限等距性质）与定理 5.14（该定理保证了在限制等距性质下的精确恢复）即可得出此结论。请注意，对测量矩阵进行重新缩放不会影响恢复，因此该结果对未归一化高斯矩阵同样成立。

■

一旦建立了限制等距性质，通过 ℓ_1-最小化进行恢复，不仅能够对稀疏信号实现精确恢复，而且在存在稀疏性缺陷时仍然稳定，并且对测量值中的加性噪声具有鲁棒性（参见定理 5.20 或定理5.21）。

此外，这些结果并不局限于高斯矩阵。它们也适用于任何满足集中不等式（7.38）的随机矩阵系综，包括具有独立各向同性次高斯行的矩阵（见第7章）。

第6章习题

1. 证明任何极小的 ε 覆盖集也必定是一个 ε 打包集。反过来，证明任何极大的 2ε 打包集必定是一个 ε 覆盖集。

2. **二项式系数的指数界**。对于整数 $p \geq k \geq 1$，证明

$$\left(\frac{p}{k}\right)^k \leq \binom{p}{k} \leq \left(\frac{ep}{k}\right)^k.$$

提示：利用熟知的泰勒展开 $\exp(k) = \sum_{i=0}^{\infty} k^i/(i!)$，求出其下界，并将其与 $\binom{p}{k}$ 的表达式联系起来。或者，使用斯特林近似：$k! \geq (k/e)^k$。

3.**高斯矩阵的受限特征值性质（Raskutti 等人，2010年）**。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为随机设计矩阵，其各行 $\boldsymbol{x}^{(i)} \in \mathbb{R}^p$ 独立同分布地取自 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$，其中协方差矩阵 $\boldsymbol{\Sigma}$ 严格正定。证明，当样本量满足 μ时，在集合 $\mathbb{C}[\mathbb{S}$上，以高概率满足 $\gamma]$ -RE 条件；若样本量满足 $n > C\,|\mathbb{S}|\log\, p$，且 C为足够大的通用常数。

4. **高斯矩阵的互相干性（Hastie 等人，2015年）**。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 为具有独立同分布的标准高斯元素的随机矩阵。证明，当样本量满足时，互不相干条件（定义4.18）以高概率成立，前提是样本量满足 $n > Ck\log p$，其中 C为足够大的数值常数。（已知事件 $\mathcal{E} = \left\{\sigma_{\min}\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_{\mathbb{S}}^{\top}\right) \geq \frac{1}{2}\right\}$ 以高概率成立；参见定理6.17。）

(i) 令 $\boldsymbol{u} \in \mathbb{R}^k$，并定义 $z_{i,\,\boldsymbol{u}} \triangleq \boldsymbol{x}_i^{\top}\boldsymbol{X}_{\mathbb{S}}(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{u}$。证明

$$\gamma = 1 - \max_{i\in\mathbb{S}^c}\max_{\boldsymbol{u}\in\{-1,+1\}^k} z_{i,\boldsymbol{u}}.$$

(ii) 利用上面定义的事件 $\mathcal{E}$，证明存在一个数值常数 $C_0 > 0$，使得

$$\Pr[z_{i,\boldsymbol{u}} \geq \tau] \leq e^{-C_0\frac{n\tau^2}{k}} + \Pr[\mathcal{E}^c], \quad \text{for any } \tau > 0,$$

对每个固定的索引 $i \in \mathbb{S}^c$ 和向量 $\boldsymbol{u} \in \{-1,+1\}^k$，

(iii) 将第 (ii) 部分与对所有 $i \in \mathbb{S}^c$ 和 $\boldsymbol{u} \in \{-1,+1\}^k$ 的并集界结合起来，完成互不相干条件成立的证明。

5. **随机化列空间近似（Halko 等人，2011年）**。随机投影能够高效估计低秩矩阵的列空间。考虑一个数据矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，它可以很好地用秩为$r$ 的矩阵近似，并考察以下算法：

(a). 生成一个随机矩阵 $\boldsymbol{A} \in \mathbb{R}^{p\times(r+k)}$，其元素为独立同分布的标准高斯元素，其中 k 是一个较小的过采样参数（例如，5）。(b). 计算草图 $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{A} \in \mathbb{R}^{n\times(r+k)}$。

(c). 为列空间 $\mathcal{C}(\boldsymbol{B})$计算一个标准正交基，并将得到的矩阵记为 $\widetilde{\boldsymbol{U}} \in \mathbb{R}^{n\times(r+k)}$。

~比较 $\boldsymbol{U}$ 的列空间与 $\boldsymbol{X}$的列空间，并比较这种随机化方法与通过完整奇异值分解计算 $\boldsymbol{X}$ 的精确列空间的计算成本。提示：参见 *JL* 引理（定理 *6.14*）的核心思想。考虑矩阵

$$\boldsymbol{B} = \boldsymbol{X}\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{A} \triangleq \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{C}, \tag{6.30}$$

其中 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ 是 $\boldsymbol{X}$的约化奇异值分解，且 $\boldsymbol{C} \triangleq \boldsymbol{V}^\top\boldsymbol{A} \in \mathbb{R}^{r\times(r+k)}$。如果 $\boldsymbol{X}$恰好是低秩的，那么 $\boldsymbol{C}$ 是一个独立同分布标准高斯矩阵，因为 $\boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}$ （参见关于高斯向量线性变换的引理 *1.90*）。在这种情况下， $\boldsymbol{B}$ 的列空间与 $\boldsymbol{X}$ 的列空间相同，因为 $\boldsymbol{C}$ 以高概率满秩。当 $\boldsymbol{X}$ 仅近似低秩时， $\boldsymbol{C}$ 则是一个大小为 $\min\{n,p\}\times(r+k)$ 的独立同分布标准高斯矩阵。令人惊讶的是，当 p 等于一个较小的整数时，与 $\boldsymbol{C}$ 相乘会以高概率保留对应于最大的 r 个奇异值的子空间，只要第 $(r+1)$ 个奇异值足够小。

6. **随机化奇异值分解**。给定一个矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，它可以很好地近似为一个秩为$r$的矩阵

~(a). 使用习题 6.5 中的算法，计算一个半正交矩阵 $\boldsymbol{U} \in \mathbb{R}^{n\times(r+k)}$，其列近似张成空间 $\mathcal{C}(\boldsymbol{X})$。(b). 构造由 $\boldsymbol{Y} \triangleq \widetilde{\boldsymbol{U}}^\top\boldsymbol{X}$定义的小矩阵 $\boldsymbol{Y} \in \mathbb{R}^{(r+k)\times p}$。(c). 计算 $\boldsymbol{Y} = \boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{V}_{\boldsymbol{Y}}^\top$的奇异值分解。(d). 输出 $\boldsymbol{U} \triangleq (\widetilde{\boldsymbol{U}}\boldsymbol{Y})_{:,1:r}$, $\boldsymbol{\Sigma} \triangleq (\boldsymbol{\Sigma}_{\boldsymbol{Y}})_{1:r,1:r}$ 和 $\boldsymbol{V} \triangleq (\boldsymbol{V}_{\boldsymbol{Y}})_{:,1:r}$，作为 $\boldsymbol{X}$的奇异值分解。

~证明：如果 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 的精确秩为 r，且 $\boldsymbol{U}$ 张成其列空间，那么该算法可以恢复 $\boldsymbol{X}$的精确奇异值分解。注意：计算 $\boldsymbol{Y}$的奇异值分解需要 $\mathcal{O}(pr^2)$，因此算法的整体复杂度由第二步决定，该步骤的计算量为$\mathcal{O}(npr)$。这比经典奇异值分解的计算量 $\mathcal{O}(np\min\{n,p\})$快得多，尤其当 $r \ll \min\{n,p\}$时。

7. **齐次拉德马赫混沌界**。设 $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ 是一个对角线元素全为零的对称矩阵，$\boldsymbol{\epsilon} \in \{-1,+1\}^p$ 是一个拉德马赫向量。证明齐次拉德马赫混沌 $\mathrm{x} = \boldsymbol{\epsilon}^\top\boldsymbol{A}\boldsymbol{\epsilon} = \sum_{i\neq j}\epsilon_i\epsilon_j a_{ij}$. 满足，

$$\Pr(|\mathrm{x}| \geq \tau) \leq 2\exp\left(-\min\left\{\frac{3\tau^2}{128\,\|\boldsymbol{A}\|_F^2}, \frac{\tau}{32\,\|\boldsymbol{A}\|_2}\right\}\right), \quad \text{for } \tau > 0.$$

提示：利用$\mathrm{x} = \boldsymbol{\epsilon}^\top\boldsymbol{A}\boldsymbol{\epsilon}$的矩母函数。找出一个解耦不等式来给 $\mathbb{E}[\exp(\theta\mathrm{x})]$, 作上界，然后利用$\exp(\theta x)$ 的凸性以及 $\boldsymbol{A}$ 的对称性来估计该二次型的矩母函数。

8. 考虑一个 $p\times k$ 矩阵 $\boldsymbol{G}$，其元素初始化为独立的拉德马赫随机变量（即以相同概率取 -1 或 $+1$），然后缩放 $\sqrt{p}$。讨论当 p 的值很大且阶为 10^6时，为什么 $\boldsymbol{G}$ 的各列将（大致）两两正交。这种技巧常用于机器学习，以快速生成 $n\times p$ 数据矩阵 $\boldsymbol{X}$ 的随机投影，记作 $\boldsymbol{X} = \boldsymbol{X}\boldsymbol{G}$。

9. **高斯矩阵的稳定 NSP**。讨论随机矩阵 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 在什么条件下，其独立同分布的高斯条目满足稳健 NSP。提示：使用定理*6.19* 和定理*4.56*。

10. **近似正交性**。从单位球面 $\mathbb{U}^{p-1} \subset \mathbb{R}^p$ 上均匀分布抽取的两个独立随机向量以很高的概率近似正交。具体而言，证明它们的内积满足 $|\langle\boldsymbol{u},\ \boldsymbol{v}\rangle| = \mathcal{O}(\sqrt{\ln(1/\varepsilon)\,p})$，其概率至少为 $1-\varepsilon$。

11. **随机化低秩近似（Drineas等人，2006年）**。令 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，并令 $\boldsymbol{C} \in \mathbb{R}^{n\times m}$ 包含 m 个列，来自 $\boldsymbol{X}$ 且 $m < p$。进一步假设，k 截断奇异值分解为 $\boldsymbol{C}$，即 $\boldsymbol{C} \approx \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^\top$（$k \leq m$）。证明

$$\left\| \boldsymbol{X} - \boldsymbol{U}_k \boldsymbol{U}_k^\top \boldsymbol{X} \right\|_2^2 \leq \|\boldsymbol{X} - \boldsymbol{X}_k\|_2^2 + 2 \left\| \boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{C}\boldsymbol{C}^\top \right\|_2 ; \tag{6.31a}$$

$$\left\| \boldsymbol{X} - \boldsymbol{U}_k \boldsymbol{U}_k^\top \boldsymbol{X} \right\|_F^2 \leq \|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 + 2\sqrt{k} \left\| \boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{C}\boldsymbol{C}^\top \right\|_F , \tag{6.31b}$$

其中 $\boldsymbol{X}_k$ 表示秩为k 的最佳近似，该近似由其截断 SVD（Lu，2022年）得到。

第7章

使用次高斯系综设计感知矩阵

目录

7.1 次高斯（SG）随机变量.275 7.2 严格次高斯（SSG）随机变量 . . . . . 279 7.3 各向同性次高斯随机向量.283 7.4 SSG随机变量的测度集中 . . . . . 285 7.5 次高斯随机矩阵下的设计性质 . . . 289 7.5.1 各向同性与次高斯随机系综下的限制等距性质 . . . . 290 7.5.2 SSG 随机系综下的限制等距性质.293 7.5.3 其他问题.295 **7.6 次高斯系综下的稀疏优化 295** 7.6.1 正交归一基下的稀疏性.296 7.6.2 严格次高斯条件下的稀疏恢复.297 **第7章习题.298**



W尽管高斯随机矩阵为压缩感知和稀疏恢复提供了强有力的理论保证，但其实际应用往往受计算和实现约束的限制。在本章中，我们将分析扩展到次高斯系综——其元素满足次高斯尾界的随机矩阵，并探讨.....

.....表明高斯矩阵的许多理想性质，例如受限等距性质和稳定稀疏恢复，在范围更广、灵活性更高的分布下仍然成立。

7.1. 次高斯（SG）随机变量

若干常见分布，尤其是高斯分布和拉德马赫分布，已知满足某些测度集中不等式。我们从更一般的视角分析这一现象，考察次高斯分布这一类（Buldygin andKozachenko，2000年；Vershynin，2010年）。

我们首先给出次高斯随机变量的一般定义，然后在下一节专门讨论零均值情形。

定义 7.1（利用集中不等式定义次高斯随机变量）。若随机变量 $\mathbf{x}$是次高斯的，则存在常数 Φ、$\xi > 0$ 使得

$$\Pr(|\mathrm{x}| \geq \tau) \leq \Phi e^{-\xi\tau^2}, \quad \text{for all } \tau > 0. \tag{7.1}$$

若满足次指数条件

$$\Pr(|\mathrm{x}| \geq \tau) \leq \Phi e^{-\xi\tau}, \quad \text{for all } \tau > 0. \tag{7.2}$$

显然，随机变量 $\mathbf{x}$ 是次高斯的，当且仅当 $\mathbf{x}^2$ 是次指数的。 [1]根据习题1.83 或引理1.84，标准高斯随机变量 $\mathbf{x} \sim \mathcal{N}(0,\ 1)$ 是次高斯的，其中 $\Phi = 1$ 且 $\xi = 1/2$。此外，拉德马赫变量（即对称伯努利分布）和有界随机变量也是次高斯的。根据定理1.73 和推论1.75，独立拉德马赫变量之和也具有次高斯性。

引理7.2 和7.3 互为逆命题，并建立了随机变量尾概率与矩之间的基本关系——这一联系对于理解次高斯随机变量的行为至关重要。这些引理以次高斯性的基本概念为基础，给出了能够严格控制均值附近集中程度的条件，从而让我们更深入地了解这类变量的概率结构。

引理 7.2：（基本集中引理）设 $\mathbf{x}$ 为满足以下条件的随机变量

$$\|\mathrm{x}\|_s \triangleq \left(\mathbb{E}[|\mathrm{x}|^s]\right)^{\frac{1}{s}} \leq \alpha\Phi^{\frac{1}{s}} s^{\frac{1}{\nu}}, \quad \text{for all } s_0 \leq s \leq s_1$$

对于某些常数 α，Φ，$\nu > 0$ 和 $s_1 > s_0 > 0$。于是

$$\Pr(|\mathrm{x}| \geq e^{\frac{1}{\nu}}\alpha\mu) \leq \Phi e^{-\frac{\mu^\nu}{\nu}}, \quad \text{for all } \mu \in [s_0^{1/\nu}, s_1^{1/\nu}]. \tag{7.3}$$

1。再次注意，我们使用粗体小写字母的正规字体表示随机向量，使用粗体大写字母的正规字体表示随机矩阵。也就是说，$\mathbf{x}$、$\mathbf{a}$、$\mathbf{X}$ 可以是随机标量、向量或矩阵；而 x、$\boldsymbol{a}$、$\boldsymbol{X}$ 则是标量、向量或矩阵。在许多情况下，这两个术语可以互换使用；也就是说，$\mathbf{x} = x$ 表示该变量的一个实现。

证明 [引理 7。2] 根据马尔可夫不等式（定理 1.68），对于任意 $\kappa > 0$，有

$$\Pr(|\mathrm{x}| \geq e^{\kappa}\alpha\mu) \leq \frac{\mathbb{E}[|\mathrm{x}|^s]}{(e^{\kappa}\alpha\mu)^s} \leq \Phi\left(\frac{\alpha s^{\frac{1}{\nu}}}{e^{\kappa}\alpha\mu}\right)^s.$$

取 $s = \mu^{\nu}$，可得 $\Pr(|\mathrm{x}| \geq e^{\kappa}\alpha\mu) \leq \Phi e^{-\kappa\mu^{\nu}}$；令 $\kappa = \frac{1}{\nu}$，即可得到所述结论。 ■

引理 7.2 的特殊情形 $\nu = 1$ 和 $\nu = 2$。事实上，若对所有 $\mathbb{E}[|\mathrm{x}|^s])^{\frac{1}{s}} \leq \alpha\Phi^{\frac{1}{s}}s$ 都有 $s \geq 2$，则

$$\Pr(|\mathrm{x}| \geq e\alpha\mu) \leq \Phi e^{-\mu}, \quad \text{for all } \mu \geq 2. \tag{7.4}$$

类似地，若对所有 $s \geq 2$，$(\mathbb{E}[|\mathrm{x}|^s])^{\frac{1}{s}} \leq \alpha\Phi^{\frac{1}{s}}\sqrt{s}$，则

$$\Pr(|\mathrm{x}| \geq e^{1/2}\alpha\mu) \leq \Phi e^{-\mu^2/2}, \quad \text{for all } \mu \geq \sqrt{2}. \tag{7.5}$$

引理 7.2的逆命题也成立。

引理 7.3：（基本矩界引理）假设随机变量 x 满足，对于某个 $\nu > 0$，

$$\Pr(|\mathrm{x}| \geq e^{\frac{1}{\nu}}\alpha\mu) \leq \Phi e^{-\frac{\mu^{\nu}}{\nu}}, \quad \text{for all } \mu > 0.$$

那么，对所有 $s > 0$，

$$\mathbb{E}[|\mathrm{x}|^s] \leq \Phi\alpha^s(e\nu)^{\frac{s}{\nu}}\Gamma\left(\frac{s}{\nu}+1\right), \tag{7.6a}$$

其中 $\Gamma(x) = \int_0^{\infty} t^{x-1}\exp(-t)dt$ 是伽马函数。因此，对于 $s \geq 1$，

$$\left(\mathbb{E}[|\mathrm{x}|^s]\right)^{\frac{1}{s}} \leq T_1\alpha(T_{2,\nu}\Phi)^{\frac{1}{s}}s^{\frac{1}{\nu}}, \tag{7.6b}$$

其中 $T_1 = e^{1/(2e)} \approx 1.2019$ 和 $T_{2,\ \nu} = \sqrt{\frac{2\pi}{\nu}}e^{\frac{\nu}{12}}$。

证明 [引理7.3]利用绝对矩引理（引理1.65）并进行两次变量替换（即 $\tau \triangleq e^{\frac{1}{\nu}}\alpha\mu$ 和 $x = \frac{\mu^{\nu}}{\nu}$），我们得到

$$\begin{aligned}
\mathbb{E}[|\mathrm{x}|^s] &= s\int_0^{\infty}\Pr(|\mathrm{x}| > \tau)\tau^{s-1}d\tau \xrightarrow{\tau \triangleq e^{\frac{1}{\nu}}\alpha\mu} s\alpha^s e^{\frac{s}{\nu}}\int_0^{\infty}\Pr(|\mathrm{x}| \geq e^{\frac{1}{\nu}}\alpha\mu)\mu^{s-1}d\mu \\
&\leq s\alpha^s e^{\frac{s}{\nu}}\int_0^{\infty}\Phi e^{-\frac{\mu^{\nu}}{\nu}}\mu^{s-1}d\mu \xrightarrow{x=\frac{\mu^{\nu}}{\nu}} s\Phi\alpha^s e^{\frac{s}{\nu}}\int_0^{\infty}e^{-x}(x)^{\frac{s}{\nu}-1}(\nu)^{\frac{s}{\nu}-1}dx \\
&= \Phi\alpha^s(e\nu)^{\frac{s}{\nu}}\frac{s}{\nu}\Gamma\left(\frac{s}{\nu}\right),
\end{aligned}$$

其中，不等式由假设得出，而最后一个等式则由伽马函数的定义得出（见（1.48））：$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ 对于 $x \geq 0$。根据伽马函数的性质，这就证明了（7.6a）。应用斯特林公式$\Gamma(x) = \sqrt{2\pi}x^{x-1/2}e^{-x}\exp\left(\frac{\theta(x)}{12x}\right)$ 对于正的 x 以及修正项 $0 \leq \theta(x) \leq 1$可得

$$\mathbb{E}[|\mathrm{x}|^s] \leq \Phi\alpha^s(e\nu)^{\frac{s}{\nu}}\sqrt{2\pi}\left(\frac{s}{\nu}\right)^{\frac{s}{\nu}+\frac{1}{2}}e^{-\frac{s}{\nu}}e^{\frac{\nu}{12s}} = \sqrt{2\pi}\Phi\alpha^s e^{\frac{\nu}{12s}}(s)^{\frac{s}{\nu}+\frac{1}{2}}\nu^{-\frac{1}{2}}.$$

对于 $s \geq 1$，取 s 次方根可得

$$\left(\mathbb{E}[|\mathrm{x}|^s]\right)^{\frac{1}{s}} \leq \left(\frac{\sqrt{2\pi}}{\sqrt{\nu}} e^{\frac{\nu}{12}} \Phi\right)^{\frac{1}{s}} \alpha s^{\frac{1}{2s}} s^{\frac{1}{\nu}}.$$

最后，函数 $s^{1/2s} = \sqrt{s^{1/s}}$ 在 $s = e$ 处取得最大值（见习题7.1），因此$s^{1/2s} \leq e^{1/2e}$。证明完毕。 ■

注7.4（次高斯变量和次指数变量的标准推论）。令 $\alpha \triangleq (2e\xi)^{-1/2}$ 以及 $\nu \triangleq 2$，并对引理 7.3 应用变量替换 $\tau = (2\xi)^{-1/2}\mu$，可得次高斯变量 x 的矩满足

$$\left(\mathbb{E}[|\mathrm{x}|^s]\right)^{1/s} \leq \widetilde{T}\xi^{-1/2}\Phi^{1/s}s^{1/2}, \quad \forall\, s \geq 1, \text{ with } \widetilde{T} \triangleq T_1 T_{2,2}/\sqrt{2e} \approx 1.0794. \tag{7.7}$$

类似地，令 $\alpha \triangleq (e\xi)^{-1}$ 和 $\nu \triangleq 1$，并使用变量替换 $\tau = \xi^{-1}\mu$，可得次指数变量 x 的矩满足

$$\left(\mathbb{E}[|\mathrm{x}|^s]\right)^{1/s} \leq \widehat{T}\xi^{-1}\Phi^{1/s}s, \quad \forall\, s \geq 1, \text{ with } \widehat{T} \triangleq T_1 T_{2,1}/e \approx 1.2047. \tag{7.8}$$

现在我们给出次高斯随机变量的一种等价刻画。

定理 7.5：（次高斯的等价表述）设 x 是一个随机变量。

(i) 若 x 是次高斯的，则存在常数 $c_1 > 0$ 和 $c_2 > 1$ 使得$\mathbb{E}[\exp(c_1\mathrm{x}^2)] \leq c_2$。

(ii) 反之，若对于某些常数 $c_1, c_2 > 0$，有 $\mathbb{E}[\exp(c_1\mathrm{x}^2)] \leq c_2$，则 x 是次高斯的。更确切地说，对于所有 $\tau > 0$，我们有 $\Pr(|\mathrm{x}| \geq \tau) \leq c_2\exp(-c_1\tau^2)$。

证明 [定理 7.5] **(i)**。利用众所周知的泰勒展开 $\exp(z) = \sum_{i=0}^{\infty} \frac{z^i}{i}!$（对所有 $z \in \mathbb{R}$，在 0 附近），可得

$$\mathbb{E}[\exp(c_1\mathrm{x}^2)] = 1 + \sum_{i=1}^{\infty} \frac{c_1^i \mathbb{E}[\mathrm{x}^{2i}]}{i!}$$

应用矩界（7.7）和斯特林不等式——$i! \geq \sqrt{2\pi i}(i/e)^i \geq \sqrt{2\pi}(i/e)^i$对于$i \geq 1$——可得

$$\mathbb{E}[\exp(c_1\mathrm{x}^2)] = 1 + \sum_{i=1}^{\infty} \frac{c_1^i \mathbb{E}[\mathrm{x}^{2i}]}{i!} \leq 1 + \sum_{i=1}^{\infty} \frac{c_1^i \widetilde{T}^{2i}\xi^{-i}\Phi(2i)^i}{\sqrt{2\pi}(i/e)^i} = 1 + \frac{\Phi}{\sqrt{2\pi}} \sum_{i=1}^{\infty} \left(\frac{c_1\widetilde{T}^2(2e)}{\xi}\right)^i.$$

若 $\left(c_1\widetilde{T}^2(2e)/\xi\right) < 1$，级数收敛，因此我们可以相应地定义 c_2：$c_2 \triangleq 1 + \frac{\Phi}{\sqrt{2\pi}} \sum_{i=1}^{\infty} \left(c_1\widetilde{T}^2(2e)/\xi\right)^i$。

(ii). 由马尔可夫不等式（定理1.68），

$$\Pr(|\mathrm{x}| \geq \tau) = \Pr\left(\exp(c_1\mathrm{x}^2) \geq \exp(c_1\tau^2)\right) \leq \mathbb{E}[\exp(c_1\mathrm{x}^2)]e^{-c_1\tau^2} \leq c_2 e^{-c_1\tau^2}.$$

证明至此完成。 ■

现在我们来考察零均值次高斯随机变量的矩母函数。

定理 7.6：（零均值次高斯变量的标准刻画）设 x 是一个随机变量。

(i) 若 x 是次高斯的，且 $\mathbb{E}[\mathrm{x}]=0$，则存在常数 $C>0$ 或常数$c>0$（仅取决于 Φ 和 ξ），使得

$$\mathbb{E}[\exp(\theta \mathrm{x})] \le \exp(\theta^2 \cdot C) = \exp(\theta^2 \cdot c^2/2), \quad \text{for all } \theta \in \mathbb{R}, \tag{7.9}$$

其中 $C = c^2/2$. [a]

（ii）反之，若（7.9）对某个 $C>0$成立，则 $\mathbb{E}[\mathrm{x}]=0$，且 x 是定义 7.1 中参数为 $\Phi=2$ 和 $\xi=\frac{1}{4C}$ 的次高斯随机变量。

[a]。满足（7.9）的任意常数 C 称为 x 的 *sub-Gaussian parameter*。通常选择这样的最小值 C。常数 c与标准差密切相关；参见定理 7.10。

定理 [的证明 7.6] **(i)**。只需考虑 $\theta \ge 0$，因为 $\theta<0$ 的情形可通过将 x 替换为 –x 得到，这样不会改变其次高斯性和零均值。利用众所周知的泰勒展开 $\exp(z)=\sum_{i=0}^{\infty}\frac{z^i}{i!}$（对所有 $z\in\mathbb{R}$，在 0 附近），可得

$$\mathbb{E}[\exp(\theta \mathrm{x})] = 1+\theta\mathbb{E}[\mathrm{x}]+\sum_{i=2}^{\infty}\frac{\theta^i\mathbb{E}[\mathrm{x}^i]}{i!} = 1+\sum_{i=2}^{\infty}\frac{\theta^i\mathbb{E}[|\mathrm{x}|^i]}{i!},$$

其中我们用到了 $\mathbb{E}[\mathrm{x}]=0$。首先，假设 $0\le\theta\le\theta_0$，其中 $\theta_0>0$ 将在稍后选取。应用矩界（7.7）和斯特林公式——$i! \ge \sqrt{2\pi i}(i/e)^i$ 对于 $i\ge 1$——可得

$$\begin{aligned}
\mathbb{E}[\exp(\theta \mathrm{x})] &\le 1+\Phi\sum_{i=2}^{\infty}\frac{\theta^i\widetilde{T}^i i^{i/2}\xi^{-i/2}}{i!} \le 1+\Phi\sum_{i=2}^{\infty}\frac{\theta^i\widetilde{T}^i i^{i/2}\xi^{-i/2}}{\sqrt{2\pi i}(i/e)^i}\\
&\le 1+\frac{\Phi(\widetilde{T}e\theta)^2}{\sqrt{2\pi\xi}}\sum_{i=0}^{\infty}\left(\widetilde{T}e\theta_0\xi^{-1/2}\right)^i \le 1+\theta^2\frac{\Phi(\widetilde{T}e)^2}{\sqrt{2\pi\xi}}\frac{1}{1-\widetilde{T}e\theta_0\xi^{-1/2}}\\
&= 1+C_1\theta^2 \le \exp(C_1\theta^2),
\end{aligned}$$

其中第三个不等式成立是因为 $i^{-(i+1)/2}<1$ 对于 $i\ge 2$，而倒数第二个不等式在 $\widetilde{T}e\theta_0\xi^{-1/2}<1$ 时成立（使用不等式 $\sum_{i=1}^{\infty}a^i \le \frac{a}{1-a}$ 对于 $0\le a<1$）。令下式成立即可满足最后一个不等式：

$$\theta_0 \triangleq (2\widetilde{T}e)^{-1}\sqrt{\xi},$$

这给出了常数 $C_1=\sqrt{2\Phi(\widetilde{T}e)^2\xi^{-1}}/\sqrt{\pi}$。

现在考虑 $\theta>\theta_0$。我们旨在证明，对于适当的 $C_2>0$，有 $\mathbb{E}[\exp(\theta \mathrm{x}-C_2\theta^2)]\le 1$。注意到

$$\theta \mathrm{x}-C_2\theta^2 = -(\sqrt{C_2}\theta-\frac{\mathrm{x}}{2\sqrt{C_2}})^2+\frac{\mathrm{x}^2}{4C_2} \le \frac{\mathrm{x}^2}{4C_2}.$$

应用定理7.5 (i)，取 $\widetilde{c}_1>0$ 和 $\widetilde{c}_2\ge 1$，并选择 $C_2=1/(4\widetilde{c}_1)$，可得

$$\mathbb{E}[\exp(\theta \mathrm{x}-C_2\theta^2)] \le \mathbb{E}[\exp(\widetilde{c}_1\mathrm{x}^2)] \le \widetilde{c}_2.$$

定义 $\nu \triangleq \ln(\widetilde{c}_2)\theta_0^{-2}$。上述不等式和条件 $\theta>\theta_0$ 表明

$$\begin{aligned}
\mathbb{E}[\exp(\theta \mathrm{x})] &\le \widetilde{c}_2\exp(C_2\theta^2) = \widetilde{c}_2\exp(-\nu\theta^2)\exp((C_2+\nu)\theta^2)\\
&\le \widetilde{c}_2\exp(-\nu\theta_0^2)\exp((C_2+\nu)\theta^2) \le \exp((C_2+\nu)\theta^2).
\end{aligned}$$

令 $C\triangleq\max\{C_1, C_2+\nu\}$，即可完成 (i) 部分的证明。

(ii). 反过来，假设对某个 7.9 成立。令 $C > 0$。令 θ，$\tau > 0$。根据马尔可夫不等式，

$$\Pr(\mathrm{x} \geq \tau) = \Pr(\exp(\theta \mathrm{x}) \geq \exp(\theta\tau)) \leq \frac{\mathbb{E}[\exp(\theta \mathrm{x})]}{\exp(\theta\tau)} \leq \exp(C\theta^2 - \theta\tau).$$

对 θ 进行优化，在 $\theta = \tau/(2C)$ 处取得最小值，从而得到

$$\Pr(\mathrm{x} \geq \tau) \leq \exp(-\tau^2/(4C)).$$

对 $-\mathrm{x}$ 应用相同的论证，可得 $\Pr(-\mathrm{x} \geq \tau) \leq \exp(-\tau^2/(4C))$。由并集界，

$$\Pr(|\mathrm{x}| \geq \tau) \leq 2\exp(-\tau^2/(4C)),$$

这表明 x 是次高斯的，且 $\Phi = 2$ 以及 $\xi = 1/(4C)$，如所声称。为验证$\mathbb{E}[\mathrm{x}] = 0$，注意对任意实数 θ，$1 + \theta \mathrm{x} \leq \exp(\theta \mathrm{x})$。取期望并使用（7.9），对于$|\theta| < 1$,

$$1 + \theta\mathbb{E}[\mathrm{x}] \leq \mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp(C\theta^2) \leq 1 + (C/2)\theta^2 + \mathcal{O}(\theta^4).$$

令 $\theta \to 0$，则必有 $\mathbb{E}[\mathrm{x}] = 0$，证明完毕。 ■

7.2. 严格次高斯（SSG）随机变量

如前所述，当次高斯随机变量的均值为零时，其矩母函数（MGF；参见定义 1.67）满足特定的上界（参见定理 7.6）。此外，关于集中不等式和设计性质的重要结果——例如定理 7.25中针对次高斯随机矩阵的受限等距性质（RIP）——均假设底层随机变量是零均值次高斯变量。因此，我们采用基于矩母函数的刻画来定义这一重要的次高斯随机变量子类，并探究它如何实现对设计矩阵行为的细粒度控制。

定义 7.7（（零均值）次高斯随机变量）。 随机变量 x 称为（零均值）次高斯（*SG*），记作 $\mathrm{x} \sim \mathcal{SG}(c^2)$，如果存在常数 $c > 0$ 使得 [a]

$$\mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp(\theta^2 c^2/2), \quad \forall \theta \in \mathbb{R}. \tag{7.10}$$

量 c^2 称为次高斯分布的方差代理。

[a] 更多等价的定义将在问题 7.4 中讨论。

该定义意味着，若随机变量 x 为（零均值）次高斯，则其拉普拉斯变换（即其矩母函数）逐点受方差为 c^2 的零均值高斯随机变量的矩母函数所支配。事实上，函数 $\mathbb{E}[\exp(\theta \mathrm{x})]$ 是 x 的矩母函数，而（ 7.10 ）右侧恰好是高斯随机变量 $\mathcal{N}(0, c^2)$[2] 的矩母函数。因此，（零均值）次高斯分布的尾部衰减速度至少与方差为 c^2 的高斯分布一样快（更多内容请参见定理 7.10 ）。

此类（零均值）次高斯分布的例子包括高斯分布本身、取值为 ± 1的拉德马赫分布（对称伯努利变量），以及更一般地，任何具有有界支撑集的零均值分布。

2. 回忆一下，高斯随机变量 $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$ 的矩母函数为 $\mathbb{M}(\theta) = \exp(\theta\mu + \theta^2\sigma^2/2)$。

例 7.8（高斯次高斯）。若 $\mathrm{x} \sim \mathcal{N}(0, \sigma^2)$，即 x 是一个方差为 σ^2 的零均值高斯随机变量，则 $\mathrm{x} \sim \mathcal{SG}(\sigma^2)$。事实上，如上所述，高斯分布的矩母函数为 $\mathbb{E}[\exp(\theta \mathrm{x})] = \exp(\theta^2\sigma^2/2)$，因此显然满足（7.10）。□

例 7.9（有界次高斯与拉德马赫）设 x 是一个零均值随机变量，且对某个 $B > 0$，几乎处处满足 $|\mathrm{x}| \leq B$。则 $\mathrm{x} \sim \mathcal{SG}(B^2)$。特别地，拉德马赫随机变量（即 $\Pr(\mathrm{x} = \pm 1) = 1/2$）满足 $\mathrm{x} \sim \mathcal{SG}(1)$ □

刻画（零均值）次高斯随机变量的一种常见方法是考察其矩。下面的定理由 Buldygin and Kozachenko（BuldyginandKozachenko，2000）提出，将前两个矩与方差代理联系起来。

定理 7.10：（布尔迪金–科扎琴科（Buldyginand Kozachenko，2000））若 $\mathrm{x} \sim \mathcal{SG}(c^2)$，则：

$$\mathbb{E}[\mathrm{x}] = 0 \qquad \text{and} \qquad \mathbb{E}[\mathrm{x}^2] \leq c^2. \tag{7.11}$$

换句话说，x 的方差不超过其方差代理 c^2。

证明 [定理 7。10] 利用指数函数的泰勒展开 $\exp(z) = \sum_{i=0}^{\infty} \frac{z^i}{i}$！对于任意 $\theta \in \mathbb{R}$，

$$\sum_{k=0}^{\infty} \frac{\theta^k}{k!}\mathbb{E}[\mathrm{x}^k] = \mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp(c^2\theta^2/2) = \sum_{k=0}^{\infty} \frac{c^{2k}\theta^{2k}}{2^k k!}.$$

当 $\theta \to 0$ 时，比较直到二阶的项，可得

$$\mathbb{E}[\mathrm{x}]\theta + \mathbb{E}[\mathrm{x}^2]\frac{\theta^2}{2!} \leq \frac{c^2\theta^2}{2} + o(\theta^2) \quad \text{as } \theta \to 0.$$

将两边除以 $\theta > 0$ 并令 $\theta \to 0^+$ 可得 $\mathbb{E}[\mathrm{x}] \leq 0$。将两边除以 $\theta < 0$并令 $\theta \to 0^-$ 可得 $\mathbb{E}[\mathrm{x}] \geq 0$。因此，$\mathbb{E}[\mathrm{x}] = 0$。（这也已在定理 7.6（ii）中证明。）确定均值为零后，将不等式除以 θ^2 并令 $\theta \to 0$。由此得到 $\mathbb{V}\mathrm{ar}[\mathrm{x}] \leq c^2$。

■

定理 7.10表明，如果 $\mathrm{x} \sim \mathcal{SG}(c^2)$，那么 $\mathbb{E}[\mathrm{x}^2] \leq c^2$。在某些应用中，考虑一类限制更严格的随机变量很有用，对于这类随机变量，该不等式成为等式。

定义 7.11（严格次高斯随机变量）。若随机变量 x 满足 $\mathrm{x} \sim \mathcal{SG}(\sigma^2)$，且其 $\sigma^2 = \mathbb{E}[\mathrm{x}^2]$，记为 $\mathrm{x} \sim \mathcal{SSG}(\sigma^2)$，则称 x 为严格次高斯（*SSG*），即，

$$\mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp(\theta^2\sigma^2/2), \text{ with } \sigma^2 = \mathbb{E}[\mathrm{x}^2], \quad \forall \theta \in \mathbb{R}. \tag{7.12}$$

例 7.12（严格次高斯随机变量）。以下是 SSG 随机变量的示例：

- 若 $\mathrm{x} \sim \mathcal{N}(0, \sigma^2)$，则 $\mathrm{x} \sim \mathcal{SSG}(\sigma^2)$。若 $\mathrm{x} \sim$ 服从均匀分布$(-1,1)$，则 $\mathbb{E}[\mathrm{x}^2] = 1/3$，
- 并且可以验证，对于所有 $\theta \in \mathbb{R}$，$\mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp(\theta^2/6)$。因此，$\mathrm{x} \sim \mathcal{SSG}(1/3)$。

- 考虑由下式定义的对称三点分布

$$\Pr(\mathrm{x}=1)=\Pr(\mathrm{x}=-1)=\frac{1-s}{2},\quad \Pr(\mathrm{x}=0)=s,\quad s\in[0,1). \tag{7.13}$$

其方差为 $\mathbb{E}[\mathrm{x}^2]=1-s$。当且仅当 $s\in[0,2/3]$ 时，该随机变量是严格次高斯的。对于 $s\in(2/3,\ 1)$，尾部行为不满足严格次高斯条件，因此 x 不是 SSG。

□

布尔迪金和科扎琴科（2000年）根据尾部衰减，对次高斯随机变量和严格次高斯随机变量给出了等价刻画，这直接反映了它们的测度集中性质。

定理 7.13：（布尔迪金–科扎琴科（Buldygin andKozachenko，2000年））随机变量 x 是方差代理为 c^2 的次高斯变量，即 $\mathrm{x}\sim\mathcal{SG}(c^2)$，当且仅当存在常数 $\tau_0\geq 0$ 和 $a\geq 0$ 使得

$$\Pr(|\mathrm{x}|\geq\tau)\leq 2\exp\left(-\frac{\tau^2}{2a^2}\right),\quad \text{for all } \tau\geq\tau_0. \tag{7.14}$$

此外，若 $\mathrm{x}\sim\mathcal{SSG}(\sigma^2)$，则上述不等式对所有满足 $a=\sigma$ 的 $\tau>0$ 都成立。

证明可直接由定理 7.6 (ii) 得出。这与直接对高斯变量应用切尔诺夫方法所得的上界一致（见习题 1.83）。事实上，对于高斯随机变量 $\mathrm{x}\sim\mathcal{N}(0,\ \sigma^2)$，引理 1.84 给出了一个稍紧的尾界：

$$\Pr(|\mathrm{x}|\geq\tau)\leq\exp\left(-\frac{\tau^2}{2\sigma^2}\right),\quad \text{for all } \tau\geq 0. \tag{7.15}$$

最后，次高斯分布与高斯分布具有一项重要的稳定性性质：相互独立的次高斯随机变量之和仍为次高斯分布。以下更一般的结果将这一闭合性质形式化。

命题 7.14 设 $\mathbf{x}=[\mathrm{x}_1,\mathrm{x}_2,\ldots,\mathrm{x}_p]$，其中每个 x_i 均为独立同分布，且 $\mathrm{x}_i\sim\mathcal{SG}(c^2)$。那么，

$$\langle\mathbf{x},\boldsymbol{\alpha}\rangle\sim\mathcal{SG}(c^2\|\boldsymbol{\alpha}\|_2^2),\quad \text{for any } \boldsymbol{\alpha}\in\mathbb{R}^p. \tag{7.16}$$

类似地，如果每个 $\mathrm{x}_i\sim\mathcal{SSG}(\sigma^2)$，那么

$$\langle\mathbf{x},\boldsymbol{\alpha}\rangle\sim\mathcal{SSG}(\sigma^2\|\boldsymbol{\alpha}\|_2^2),\quad \text{for any } \boldsymbol{\alpha}\in\mathbb{R}^p. \tag{7.17}$$

此外，由此可得

$$\Pr\left(|\langle\mathbf{x},\boldsymbol{\alpha}\rangle|\geq\tau\right)\leq 2\exp\left(-\frac{\tau^2}{4C\|\boldsymbol{\alpha}\|_2^2}\right),\quad \text{for all } \tau>0.$$

其中，$C=c^2/2$ 在次高斯和 $C=\sigma^2/2$ 在 SSG 情形下。

证明 [命题 7 的14]由于变量 x_i 是独立同分布的，联合分布可分解并简化为：

$$\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^{p}\alpha_i\mathrm{x}_i\right)\right]=\mathbb{E}\left[\prod_{i=1}^{p}\exp(\theta\alpha_i\mathrm{x}_i)\right]=\prod_{i=1}^{p}\mathbb{E}\left[\exp(\theta\alpha_i\mathrm{x}_i)\right]$$
$$\leq\prod_{i=1}^{p}\exp(\theta^2\alpha_i^2\cdot c^2/2)=\exp\left(\left(\sum_{i=1}^{p}\alpha_i^2\right)\theta^2c^2/2\right).$$

因此，$\langle\mathbf{x}\ \boldsymbol{\alpha}\rangle$ 是次高斯的，方差代理为 $c^2\left\|\boldsymbol{\alpha}\right\|_2^2$。如果变量 x_i 是严格次高斯的，则同样的论证适用，只需将 $c^2=\sigma^2$替换为 $\mathbb{E}[\langle\mathbf{x}\ \boldsymbol{\alpha}\rangle^2]=\sigma^2\left\|\boldsymbol{\alpha}\right\|_2^2$，从而确认其严格次高斯性。这证明了第一部分。尾界随后直接由定理7.6 （ii）得出。

■

举一个具体例子，设 $\boldsymbol{\epsilon}=[\epsilon_1,\ldots,\epsilon_p]^\top$ 是一个拉德马赫序列，且 $\mathrm{y}=\sum_{i=1}^{p}\alpha_i\epsilon_i$。那么

$$\mathbb{E}[\exp(\theta\mathrm{y})]\leq\exp(\theta^2\left\|\boldsymbol{\alpha}\right\|_2^2/2),$$

这证实 y 是次高斯的，方差代理为 $\left\|\boldsymbol{\alpha}\right\|_2^2$。

有限个（可能相依的）零均值次高斯随机变量的期望最大值可按如下方式给出上界。

命题 7.15：（零均值次高斯变量的最大值）设$\mathbf{x}=[\mathrm{x}_1,\mathrm{x}_2,\ldots,\mathrm{x}_p]$，其中每个 $\mathrm{x}_i\sim\mathcal{SG}(c_i^2)$（不一定相互独立）都满足 $\mathbb{E}[\exp(\theta\mathrm{x}_i)]\leq\exp(c_i^2\theta^2/2)$，$i\in\{1,2,\ldots,p\}$。定义 $C\triangleq\max_i c_i^2/2$。则

$$\mathbb{E}[\max_{i\in\{1,2,\ldots,p\}}\mathrm{x}_i]\leq\sqrt{4C\ln(p)},\qquad\text{and}\qquad\mathbb{E}[\max_{i\in\{1,2,\ldots,p\}}|\mathrm{x}_i|]\leq\sqrt{4C\ln(2p)}.\tag{7.18}$$

对于任意 $\tau>0$，都有

$$\Pr\left(\max_{i\in\{1,2,\ldots,p\}}\mathrm{x}_i\right)\leq p\exp(-\frac{\tau^2}{4C})\quad\text{and}\quad\Pr\left(\max_{i\in\{1,2,\ldots,p\}}|\mathrm{x}_i|\right)\leq 2p\exp(-\frac{\tau^2}{4C}).\tag{7.19}$$

证明 [命题7.15] 情形 $p=1$ 是显然的；假设 $p\geq 2$。令 $\Theta>0$ 为稍后选定的参数。利用指数函数的单调性和詹森不等式，

$$\Theta\mathbb{E}[\max_{i\in\{1,2,\ldots,p\}}\mathrm{x}_i]=\mathbb{E}[\ln\max_{i\in\{1,2,\ldots,p\}}\exp(\Theta\mathrm{x}_i)]\leq\mathbb{E}\left[\ln\left(\sum_{i=1}^{p}\exp(\Theta\mathrm{x}_i)\right)\right]$$
$$\leq\ln\left(\sum_{i=1}^{p}\mathbb{E}[\exp(\Theta\mathrm{x}_i)]\right)\leq\ln(p\exp(C\Theta^2))=C\Theta^2+\ln(p).$$

取 $\Theta\triangleq\sqrt{C^{-1}\ln(p)}$，得到

$$\sqrt{C^{-1}\ln(p)}\mathbb{E}[\max_{i\in\{1,2,\ldots,p\}}\mathrm{x}_i]\leq\ln(p)+\ln(p),$$

这意味着 $\mathbb{E}[\max^{i\in\{1,2,\ldots,p\}}\mathrm{x}_i]\leq\sqrt{4C\ln(p)}$。

对于绝对最大值，我们可以将期望改写为 $\mathbb{E}[\max_{i\in\{1,2,\dots,p\}} |\mathrm{x}_i|] = \mathbb{E}[\max\{\mathrm{x}_1,\dots,\mathrm{x}_p,-\mathrm{x}_1,\dots,-\mathrm{x}_p\}]$。将前面的上界应用于这组 $2p$ 变量，可得第二个期望不等式。

尾界由并集界与次高斯尾部不等式共同推出。

■

该命题中的界几乎是紧的：对于一列相互独立的标准高斯随机变量，其期望最大值按 $\sqrt{2\ln(p}$ 的量级增长，与上述结果相差至多一个常数因子（见定理6.5 (iii)）。

7.3. 各向同性次高斯随机向量

次高斯分布与测度集中现象密切相关（Ledoux，2001年）。例如，结合定理7.13和命题7.14，可得次高斯随机变量加权和的偏差界。

更广泛地说，次高斯随机矩阵属于一类更大的随机矩阵，它们具有很强的集中性质——具体而言，在经过线性变换后，它们近似保持向量的欧几里得范数。这一性质对于压缩感知和高维统计等应用至关重要。

我们首先介绍几个重要定义。

定义 7.16（各向同性和次高斯随机向量）。设$\mathbf{x}$是一个取值于 $\mathbb{R}^p$的随机向量。

(i) 如果 $\mathbb{E}[\langle \mathbf{x}\ \boldsymbol{\beta}\rangle^2] = \|\boldsymbol{\beta}\|_2^2$ 对所有 $\boldsymbol{\beta}\in\mathbb{R}^p$都成立，那么 $\mathbf{x}$ 称为各向同性。(ii) 如果对于每个单位向量 $\boldsymbol{\beta}\in\mathbb{R}^p$ （即 $\|\boldsymbol{\beta}\|_2 = 1$），标量随机变量$\langle \mathbf{x}\ \boldsymbol{\beta}\rangle$ 是次高斯的，次高斯参数为 $C \equiv c^2/2$ （参见定理7.6中的定义或定义7.10），且该参数独立于 $\boldsymbol{\beta}$（理想情况下也独立于 p）；也就是说，

$$\mathbb{E}[\exp(\theta\,\langle \mathbf{x},\boldsymbol{\beta}\rangle)] \le \exp(\theta^2\cdot C), \quad \text{for all } \theta\in\mathbb{R}, \quad \|\boldsymbol{\beta}\|_2 = 1, \tag{7.20}$$

那么$\mathbf{x}$ 称为次高斯随机向量。

注意，各向同性次高斯随机向量的元素不一定相互独立。

以下命题表明，元素相互独立、均值为零且方差为单位的次高斯随机向量，既是各向同性的，也是次高斯的；进一步讨论请参见习题7.6 。

命题 7.17：（各向同性次高斯随机向量）设$\mathbf{x}\in\mathbb{R}^p$ 是一个随机向量，其元素为相互独立、均值为零、方差为 1 且具有相同次高斯参数 $C\equiv c^2/2$ （如式（7.20）所示）。5那么，$\mathbf{x}$ 是一个各向同性的次高斯随机向量，其常数与维数无关p。

证明 [命题 7 的17] 设 $\boldsymbol{\beta}\in\mathbb{R}^p$ 满足 $\|\boldsymbol{\beta}\|_2 = 1$。由于 $\mathbf{x}$ 的各个元素$_i$ 属于 $\mathbf{x}$ 且相互独立、均值为零、方差为单位方差，

$$\mathbb{E}[\langle \mathbf{x},\boldsymbol{\beta}\rangle^2] = \sum_{i,j=1}^{p} \beta_i\beta_j\mathbb{E}[\mathrm{x}_i\mathrm{x}_j] = \sum_{i=1}^{p}\beta_i^2 = \|\boldsymbol{\beta}\|_2^2 .$$

因此，$\mathbf{x}$ 是各向同性的。此外，根据命题 7.14 和定理 7.6，随机变量 $\mathrm{y} = \langle \mathbf{x}\ \boldsymbol{\beta} \rangle$ 是次高斯的 $\mathcal{SG}(c^2/2)$（或者参数为 $\Phi = 2$ 和$\xi = 1/(4C\,\|\boldsymbol{\beta}\|_2^2) = 1/(4C)$）。因此，$\mathbf{x}$ 是一个参数不依赖于 p的次高斯随机向量。 ■

现在考虑一个矩阵 $\mathbf{X} \in \mathbb{R}^{n\times p}$，其元素是（单位方差次高斯）随机变量。这样的矩阵称为（次高斯）随机矩阵或（次高斯）随机矩阵系综。更具体地说，我们关注行向量相互独立、各向同性且为次高斯随机向量的随机矩阵$\mathbf{X} \in \mathbb{R}^{n\times p}$，其中 $\mathbb{R}^p$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}.$$

其中每个 $\mathbf{x}_i \in \mathbb{R}^p$ 都满足定义 7.1 中的条件。

例 7.18（次高斯随机矩阵）。设 $\mathbf{X}$ 是一个 $n\times p$ 随机矩阵。

（i）如果 $\mathbf{X}$ 的元素是独立拉德马赫变量（即以相同的概率取值 ± 1），那么 $\mathbf{X}$称为拉德马赫随机矩阵。（ii）如果 $\mathbf{X}$的元素是独立的标准正态分布随机变量，那么 $\mathbf{X}$ 称为高斯随机矩阵。（iii）更一般地，如果 $\mathbf{X}$ 的所有元素都是相互独立、均值为零、方差为 1 的次高斯随机变量，且具有定义 Φ，ξ 中关于次高斯随机变量7.1 的相同常数，也就是说，

$$\Pr(|\mathrm{x}_{ij}| \geq \tau) \leq \Phi e^{-\xi\tau^2}, \quad \text{对所有 } \tau > 0,\ i \in \{1,2,\ldots,n\}, j \in \{1,2,\ldots,p\}, \tag{7.21}$$

$\mathbf{X}$称为次高斯随机矩阵。等价地，也可以要求矩母函数界

$$\mathbb{E}[\exp(\theta \mathrm{x}_{ij})] \leq \exp(\theta^2 C), \quad \text{for all } \theta \in \mathbb{R}, \quad i \in \{1,2,\ldots,n\}, j \in \{1,2,\ldots,p\}, \tag{7.22}$$

存在某个常数 $C \equiv c^2/2$，它与 i、j和 p 无关（参见定理 7.6）。

高斯矩阵和拉德马赫矩阵都是次高斯随机矩阵的特殊情形。注意，次高斯矩阵的元素不必服从相同的分布——只需它们具有一致的次高斯参数。□

具有次高斯行的随机矩阵的一个关键性质是其范数保持行为，这构成了受限等距性质（RIP）的基础，我们将在后文讨论这一性质。

定理 7.19：（次高斯行下的范数保持）设$\mathbf{X} \in \mathbb{R}^{n\times p}$是一个随机矩阵，其各行相互独立、各向同性且为次高斯，具有相同的次高斯参数 $C \equiv c^2/2$，见（7.20）。那么，

$$\Pr\left(\left|n^{-1}\,\|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| \geq \varepsilon\,\|\boldsymbol{\beta}\|_2^2\right) \leq 2\exp(-\widetilde{C}\varepsilon^2 n), \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \text{ all } \varepsilon \in (0,1),$$

~其中 C 仅取决于 C。

定理 [的7.19] 设 $\boldsymbol{\beta} \in \mathbb{R}^p$。不失一般性，假设 $\|\boldsymbol{\beta}\|_2 = 1$。令$\mathbf{x}^{(1)}$、$\mathbf{x}^{(2)}$、……、$\mathbf{x}^{(n)} \in \mathbb{R}^p$表示矩阵$\mathbf{X}$的各行，并定义随机变量

$$\mathrm{y}_i = \left\langle \mathbf{x}^{(i)}, \boldsymbol{\beta} \right\rangle^2 - \|\boldsymbol{\beta}\|_2^2 = \left\langle \mathbf{x}^{(i)}, \boldsymbol{\beta} \right\rangle^2 - 1, \quad i \in \{1, 2, \ldots, n\}.$$

由于 $\mathbf{x}^{(i)}$ 是各向同性的，因此有 $\mathbb{E}[\mathrm{y}_i] = 0$。此外，由于 $\langle \mathbf{x}^{(i)}, \boldsymbol{\beta} \rangle$ 是次高斯的，y_i 是次指数的；也就是说，存在常数 Φ，ξ （仅取决于 C），使得 $\Pr(|\mathrm{y}_i| \geq \varepsilon) \leq \Phi \exp(-\xi\varepsilon)$。注意到

$$n^{-1} \|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\langle \mathbf{x}^{(i)}, \boldsymbol{\beta} \rangle^2 - \|\boldsymbol{\beta}\|_2^2 \right) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{y}_i.$$

由于 y_i 相互独立，根据次指数随机变量的伯恩斯坦不等式（推论1.78），可得

$$\begin{aligned} \Pr\left(\left| n^{-1} \sum_{i=1}^{n} \mathrm{y}_i \right| \geq \varepsilon \right) &= \Pr\left(\left| \sum_{i=1}^{n} \mathrm{y}_i \right| \geq \varepsilon n \right) \\ &\leq 2 \exp\left(-\frac{\xi^2 \varepsilon^2 n^2 / 2}{2\Phi n + \xi \varepsilon n} \right) \leq 2 \exp\left(-\frac{\xi^2 \varepsilon^2 n}{4\Phi + 2\xi} \right), \end{aligned}$$

其中，最后一个不等式成立是因为 $\varepsilon \in (0, 1)$。因此，结论由 $\widetilde{C} \triangleq \frac{\xi^2}{4\Phi + 2\xi}$ 得出。

——— ■

最后，注意缩放后的矩阵 $\widetilde{\mathbf{X}} \triangleq \frac{1}{\sqrt{n}} \mathbf{X}$，其中 $\mathbf{X}$ 满足前一个结果的假设，满足

$$\Pr\left(\left| \left\| \widetilde{\mathbf{X}} \boldsymbol{\beta} \right\|_2^2 - \|\boldsymbol{\beta}\|_2^2 \right| \geq \varepsilon \|\boldsymbol{\beta}\|_2^2 \right) \leq 2 \exp(-\widetilde{C} \varepsilon^2 n), \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \text{ all } \varepsilon \in (0, 1). \quad (7.23)$$

这个集中不等式是证明次高斯随机矩阵的受限等距性质的基础；参见命题7.24。

7.4. SSG随机变量的测度集中

~本节将说明，当元素严格次高斯时，式 C （7.23）中的常数可以明确给出。此外，就我们的目的而言，研究由次高斯随机变量构成的向量的欧几里得范数尤其有启发性。具体来说，设$\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^\top$，其中每个 x_i 都独立同分布，且 $\mathrm{x}_i \sim \mathcal{SG}(c)$。我们关注 $\|\mathbf{x}\|_2$ 与其期望值的偏差。

为了建立主要结果，我们将使用非负随机变量的马尔可夫不等式（定理1.68），以及以下关于次高斯随机变量指数矩的上界。

引理 7.20 假设 $\mathrm{x} \sim \mathcal{SG}(c^2)$。那么，对于任意 $\lambda \in [0, 1)$，

$$\mathbb{E}[\exp\left(\lambda \mathrm{x}^2 / 2c^2\right)] \leq \frac{1}{\sqrt{1 - \lambda}},$$

证明 [引理 7.20] 当 $\lambda = 0$ 时，该结论显然成立。现在假设 $\lambda \in (0,1)$。令 $f(x)$ 表示 $\mathbf{x}$ 的概率密度函数。根据次高斯性的定义，

$$\int_{-\infty}^{\infty} \exp(\theta x) f(x)\, dx \leq \exp(\theta^2 c^2/2), \quad \forall\, \theta \in \mathbb{R}.$$

两边乘以 $\exp(-\theta^2 c^2/2\lambda)$，得到

$$\int_{-\infty}^{\infty} \exp\left(\theta x - \theta^2 c^2/2\lambda\right) f(x)\, dx \leq \exp(\theta^2 c^2(\lambda-1)/2\lambda).$$

对两边关于 $\theta \in \mathbb{R}$ 积分，得到

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \exp\left(\theta x - \theta^2 c^2/2\lambda\right) d\theta \right) f(x)\, dx \leq \int_{-\infty}^{\infty} \exp(\theta^2 c^2(\lambda-1)/2\lambda)\, d\theta.$$

计算两边的高斯积分，得到

$$\frac{1}{c}\sqrt{2\pi\lambda} \int_{-\infty}^{\infty} \exp\left(\lambda x^2/2c^2\right) f(x)\, dx \leq \frac{1}{c}\sqrt{\frac{2\pi\lambda}{1-\lambda}},$$

化简后即得所需的不等式。证明完毕。 ■

该引理进一步阐明了零均值次高斯变量的性质，并补充了定理7.5的第 (i) 部分。

现在我们给出一个关键的集中结果，它推广了Dasgupta 和 Gupta (2003年的研究成果，并拓展了高斯向量的著名性质（参见第 6.2节）。

定理 7.21：（次高斯变量的集中（Dasgupta 和 Gupta，2003；Baraniuk 等人，2011年）） 设 $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^\top$，其中每个 x_i 均为独立同分布，且 $\mathrm{x}_i \sim \mathcal{SG}(c^2)$，并且 $\mathbb{E}[\mathrm{x}_i^2] = \sigma^2$。那么，

$$\mathbb{E}[\|\mathbf{x}\|_2^2] = n\sigma^2. \tag{7.24}$$

此外，对于任意 $\mu \in (0,\ 1)$ 和任意 $\nu \in [c^2/\sigma^2, \nu_{\max}]$，存在一个常数 $\kappa \geq 4$，它仅取决于 $\nu_{\max}$ 和比值 σ^2/c^2，使得

$$\Pr(\|\mathbf{x}\|_2^2 \leq \mu n\sigma^2) \leq \exp\left(-n(1-\mu)^2/\kappa\right) \tag{7.25}$$

and

$$\Pr(\|\mathbf{x}\|_2^2 \geq \nu n\sigma^2) \leq \exp\left(-n(\nu-1)^2/\kappa\right). \tag{7.26}$$

定理 [的证明 7。21]方程（7.24）可由独立性立即推出：

$$\mathbb{E}[\|\mathbf{x}\|_2^2] = \sum_{i=1}^{n} \mathbb{E}[\mathrm{x}_i^2] = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$$

由于 $\|\mathbf{x}\|_2^2$ 非负且指数函数单调，因此我们可以应用马尔可夫不等式（定理 1.68）：

$$\begin{aligned}\Pr(\|\mathbf{x}\|_2^2 \geq \nu n\sigma^2) &= \Pr\left(\exp\left(\lambda \|\mathbf{x}\|_2^2\right) \geq \exp\left(\lambda\nu n\sigma^2\right)\right) \\ &\leq \frac{\mathbb{E}\left[\exp\left(\lambda \|\mathbf{x}\|_2^2\right)\right]}{\exp\left(\lambda\nu n\sigma^2\right)} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\lambda \mathrm{x}_i^2\right)\right]}{\exp\left(\lambda\nu n\sigma^2\right)}.\end{aligned}$$

由于 $\mathbf{x}_i \sim \mathcal{SG}(c^2)$，对于 $\lambda \in [0, 1/(2c^2))$，由引理 7.20 可得

$$\mathbb{E}\left[\exp\left(\lambda \mathrm{x}_i^2\right)\right] = \mathbb{E}\left[\exp\left(2c^2\lambda \mathrm{x}_i^2/2c^2\right)\right] \leq \frac{1}{\sqrt{1-2c^2\lambda}}$$
$$\Longrightarrow \prod_{i=1}^{n} \mathbb{E}\left(\exp\left(\lambda \mathrm{x}_i^2\right)\right) \leq \left(\frac{1}{1-2c^2\lambda}\right)^{n/2}.$$

结合前面的两个结果可得

$$\Pr(\|\mathbf{x}\|_2^2 \geq \nu n\sigma^2) \leq \left(\frac{\exp\left(-2\lambda\nu\sigma^2\right)}{1-2c^2\lambda}\right)^{n/2}.$$

对 λ进行优化可得如下选择

$$\lambda = \frac{\nu\sigma^2 - c^2}{2c^2\sigma^2\nu},$$

只要满足 $\nu \in [c^2/\sigma^2, \nu_{\max}]$，该选择就属于 $[0, 1/(2c^2)]$。代入这个最优的 λ 可得

$$\Pr(\|\mathbf{x}\|_2^2 \geq \nu n\sigma^2) \leq \left(\nu\frac{\sigma^2}{c^2}\exp\left(1-\nu\frac{\sigma^2}{c^2}\right)\right)^{n/2}. \tag{7.27}$$

类似的计算可得

$$\Pr(\|\mathbf{x}\|_2^2 \leq \mu n\sigma^2) \leq \left(\mu\frac{\sigma^2}{c^2}\exp\left(1-\mu\frac{\sigma^2}{c^2}\right)\right)^{n/2}. \tag{7.28}$$

为简化这些界，定义

$$\kappa \triangleq \max\left\{4, 2\frac{(\nu_{\max}\sigma^2/c^2 - 1)^2}{(\nu_{\max}\sigma^2/c^2 - 1) - \ln(\nu_{\max}\sigma^2/c^2)}\right\}.$$

那么，对于所有 $\gamma \in [0, \nu_{\max}\sigma^2/c^2]$，都有

$$\ln(\gamma) \leq (\gamma-1) - \frac{2(\gamma-1)^2}{\kappa} \quad \Longrightarrow \quad \gamma \leq \exp\left((\gamma-1) - \frac{2(\gamma-1)^2}{\kappa}\right).$$

由于 $\sigma^2 \leq c^2$ 根据定理 7.10，将 $\gamma = \nu\sigma^2/c^2 \leq \nu \leq \nu_{\max}$ 代入（7.27）可得（7.26）。类似地，将 $\gamma = \mu\sigma^2/c^2$ 代入（7.28）可得（7.25）。 ■

该结果表明，次高斯向量的平方范数会紧密集中于其均值附近 $n\sigma^2$，且随着 n 增大，尾概率呈指数衰减。然而请注意，上尾界（7.26）仅适用于 $\nu \geq c^2/\sigma^2 \geq 1$，正如定理 7.10所示。因此，对于一般的次高斯分布，集中程度可能无法任意紧密。

关键的是，对于严格次高斯分布（其中 $c^2 = \sigma^2$），这一限制便不复存在。在这一有利情形下，我们得到如下简化且对称的集中不等式。

定理7.22：（SSG变量的集中性）设 $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^\top$，其中每个 x_i 均独立同分布，且 $\mathrm{x}_i \sim \mathcal{SSG}(\sigma^2)$。那么，

$$\mathbb{E}[\|\mathbf{x}\|_2^2] = n\sigma^2 \tag{7.29}$$

且对于任意 $\varepsilon \in (0,1)$，

$$\Pr\left(\left|\|\mathbf{x}\|_2^2 - n\sigma^2\right| \geq \varepsilon n\sigma^2\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{\kappa}\right) \tag{7.30}$$

其中 $\kappa = 2/(1-\ln(2)) \approx 6.52$。

证明 [定理 7.22] 。由于每个 $\mathrm{x}_i \sim \mathcal{SSG}(\sigma^2)$，我们有 $\mathrm{x}_i \sim \mathcal{SG}(\sigma^2)$，且$\mathbb{E}[\mathrm{x}_i^2] = \sigma^2$。利用定理 7.21，并取 $\mu = 1-\varepsilon$ 和 $\nu = 1+\varepsilon$，我们可以简化并合并 (7.25) 和 (7.26) 中的界，从而得到 (7.30)。由于对于 $\varepsilon \in (0,1)$，$\nu < 2$，我们可以取 $\nu_{\max} = 2$。因此，$\kappa = 2/(1-\ln(2)) \approx 6.52$。证明完毕。 ■

作为一个直接应用，考虑一个噪声向量 $\boldsymbol{\epsilon} \in \mathbb{R}^n$，其元素独立同分布 $\epsilon_i \sim \mathcal{N}(0,\sigma^2)$。由于高斯分布是严格次高斯的，定理 7.22 保证存在一个常数 $c_0 > 0$，使得对于任意 $e > 0$,

$$\Pr\left(\|\boldsymbol{\epsilon}\|_2 \geq (1+e)\sqrt{n}\sigma\right) \leq \exp\left(-c_0 n e^2\right). \tag{7.31}$$

特别地，令 $e = 1$ 可得

$$\Pr\left(\|\boldsymbol{\epsilon}\|_2 \geq 2\sqrt{n}\sigma\right) \leq \exp\left(-c_0 n\right). \tag{7.32}$$

最后，定理 7.22 还推出了以下有用的推论，说明随机矩阵对向量的变换会呈现集中现象。该结果是定理 7.19 在常数确定时的一个特例。

推论 7.23：（SSG矩阵的范数保持）设$\mathbf{X} \in \mathbb{R}^{n\times p}$ 是一个元素独立同分布、元素为 $\mathrm{x}_{ij} \sim \mathcal{SSG}(1/n)$ 的随机矩阵。对于任意固定的 $\boldsymbol{\beta} \in \mathbb{R}^p$，定义$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$。那么

$$\mathbb{E}[\|\mathbf{y}\|_2^2] = \|\boldsymbol{\beta}\|_2^2 \tag{7.33}$$

并且对于任意 $\varepsilon \in (0,1)$,

$$\Pr\left(\left|\|\mathbf{y}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| \geq \varepsilon\|\boldsymbol{\beta}\|_2^2\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{\kappa}\right) \tag{7.34}$$

$$\Longrightarrow \Pr\left((1-\varepsilon)\|\boldsymbol{\beta}\|_2^2 \leq \|\mathbf{y}\|_2^2 \leq (1+\varepsilon)\|\boldsymbol{\beta}\|_2^2\right) \geq 1 - 2\exp\left(-\frac{n\varepsilon^2}{\kappa}\right), \tag{7.35}$$

其中 $\kappa = 2/(1-\ln(2)) \approx 6.52$。

证明 [推论7。23] 设 $\boldsymbol{x}^{(i)}$ 表示 $i\mathbf{X}$ 的第$\mathbf{X}$行。那么 $\mathrm{y}_i = \langle \boldsymbol{x}^{(i)}, \boldsymbol{\beta}\rangle$因此根据命题7.14，$\mathrm{y}_i \sim \mathcal{SSG}(\|\boldsymbol{\beta}\|_2^2/n)$ 。将定理7.22 应用于 n维随机向量$\mathbf{y}$，我们得到（7.34）。 ■

7.5 次高斯随机矩阵下的设计性质

现在我们再次回到如何构造满足 RIP 的矩阵这一问题。虽然可以确定性地构造 $n \times p$ 满足阶 RIP 的矩阵 k，早期的方法依赖于互相干性。若矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 的列均为单位范数，则称其为 μ-相干的，如果 $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \leq \mu$ 对所有 $i \neq j \in \{1, 2, \ldots, p\}$（见定义4.16）。这样的矩阵满足 阶 RIP，其常数为 k（ $\delta = (k-1)\mu$（见命题4.22）。

自 Kashin（1975）的研究以来，具有 $\mu = \mathcal{O}\left(\ln \sqrt{n}^{p} \ln n\right)$ 的 μ-相干矩阵的确定性构造便已为人所知。然而，这些构造需要 $n = \widetilde{O}\left(k^2 \ln_{\delta^2}{}^{3} p\right)$，比随机构造所需的行数大出二次量级。其他显式构造也好不了多少：DeVore（2007）实现了 $n = \mathcal{O}(k^2 \ln p)$，而 Indyk（2008）在某个常数 $\mu > 0$ 下需要 $n = \mathcal{O}(kp^{\mu})$。在许多实际应用中，这类对 n 的要求大得令人难以承受。Bourgain 等（2011）带来了首次重大改进，他们提出了满足限制等距性质（RIP）的确定性组合构造，其中 $n = \widetilde{O}\left(k^{(2-\epsilon)}_{\delta^2}\right)$，且常数为 $\epsilon > 0$。尽管这降低了对 k 的二次依赖，但这些构造仍然复杂，也未能达到随机矩阵所实现的近线性缩放。

幸运的是，通过对矩阵构造进行随机化，可以克服这些局限。一种简单而有效的方法如下：给定维数 n 和 p，通过从适当的概率分布中独立采样其元素 x_{ij}，生成随机矩阵 $\mathbf{X} \in \mathbb{R}^{n \times p}$。首先可以注意到，如果我们只要求 $\delta_{2k} > 0$（即对 $2k$-稀疏向量具有单射性），那么令 $n = 2k$ 并从连续分布（例如高斯分布）中抽取 $\mathbf{X}$ 就足够了。以概率 1，$2k$ 列的每个子集都是线性无关的，因此 RIP 对某个 $\delta_{2k} \in (0,1)$ 成立。

然而，在实践中，我们通常需要指定的 RIP 常数 δ_{2k}（例如 $\delta_{2k} < 0.1$）。为了保证这样的上界，我们必须控制与 k-稀疏支撑集对应的所有 $\binom{p}{k}$ 子矩阵的奇异值——对于 p 和 k 的现实取值而言，这在计算上是不可行的。此外，当 $n = 2k$ 时，所得的 δ_{2k} 通常非常接近 1，使该矩阵不适用于压缩感知等应用。

因此，我们转而聚焦于构造满足以下阶数的限制等距性质（RIP）的矩阵： $2k$其常数为指定的 $\delta_{2k} \in (0,1)$。我们采用的方法遵循 Baraniuk 等人（2008）提出的简单框架，Mendelson 等人（2008）后来将其扩展到更广泛的随机矩阵类别。

本节将说明，次高斯分布的测度集中性质如何带来一个直接的证明，表明随机次高斯矩阵满足限制等距性质（RIP）。具体而言，我们旨在证明：如果随机抽取一个 $n \times p$ 矩阵$\mathbf{X}$，且 n 足够大，那么以高概率存在一个常数 $\delta_k \in (0,1)$，使得

$$(1 - \delta_k) \|\boldsymbol{\beta}\|_2^2 \leq \|\mathbf{X}\boldsymbol{\beta}\|_2^2 \leq (1 + \delta_k) \|\boldsymbol{\beta}\|_2^2 \tag{7.36}$$

对所有 $\boldsymbol{\beta} \in \mathbb{B}_0[k] \triangleq \{\boldsymbol{\alpha} \mid \|\boldsymbol{\alpha}\|_0 \leq k\}$均成立；也就是说，对所有 k稀疏向量均成立。高斯情形也有类似结果（见定理6.19）。

为确保满足限制等距性质（RIP），我们对底层分布施加两个条件：

- 次高斯性。 我们假设 $\mathbf{X}$ 的元素服从次高斯分布。这样我们就可以应用定理7.19 或定理 7.21中的简洁集中上界。利用推论 7.23，可以将这一论证推广到严格次高斯分布，并得到显式常数。

- 近似范数保持。 我们要求矩阵在期望意义下保持欧几里得范数。可以通过要求其各向同性，或令

$$\mathbb{E}[\mathrm{x}_{ij}^2] = \frac{1}{n}, \tag{7.37}$$

使每个元素的方差为 $1/n$。在这种归一化下，对于任意固定向量 $\boldsymbol{\beta}$，我们有 $\mathbb{E}[\|\mathbf{X}\boldsymbol{\beta}\|_2^2] = \|\boldsymbol{\beta}\|_2^2$，因此 RIP 是关于均值附近集中的陈述。

7.5.1 各向同性与次高斯随机系综下的 RIP

现在我们证明，满足集中不等式（7.23）的随机矩阵——当其行各向同性且次高斯时，该不等式成立——或满足不等式（7.34）的随机矩阵——当其元素严格次高斯时，该不等式成立——也满足 k 阶的受限等距性质（RIP），前提是行数至少按 k 乘以一个对数因子的规模增长。我们首先证明，在维数满足适当条件时，此类随机矩阵的任意 k 列子矩阵都是良态的。这是对高斯条目矩阵奇异值定理的推广；参见定理 6.17。

命题 7.24：（通过尾界得到限制等距性质） 设 $\mathbb{S} \subset \{1, 2, \dots, p\}$ 满足 $|\mathbb{S}| = k$。假设一个 $n \times p$ 随机矩阵 $\mathbf{X}$ 取自满足集中不等式（7.23）的分布，即

$$\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| > \tau \|\boldsymbol{\beta}\|_2^2\right) \le 2\exp(-\widetilde{C}\tau^2 n), \quad \text{for all } \tau \in (0,1),\ \boldsymbol{\beta} \in \mathbb{R}^p.^a \tag{7.38}$$

等价地，定理7.19 表明，$\sqrt{n}\mathbf{X}$具有独立、各向同性且次高斯的行，并具有相同的次高斯参数 C（见7.20）。如果

$$n \ge \frac{2}{3\widetilde{C}\delta^2}(7k + 2\ln(2/\varepsilon)), \quad \text{for } \varepsilon, \delta \in (0,1), \tag{7.39}$$

则以至少 $1 - \varepsilon$ 的概率，

$$\left\|\mathbf{X}_{\mathbb{S}}^\top \mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \le \delta.$$

a。或者，集中不等式（7.34）以 $\widetilde{C} = 1/\kappa = (1 - \ln(2))/2$成立。

证明 [命题 7.24] 根据定理 6.3，对于 $\rho \in (0, 1/2)$，存在单位球面 $\mathbb{A}$的有限子集 $\mathbb{U} = \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \mathrm{supp}(\boldsymbol{\beta}) \subset \mathbb{S},\ \|\boldsymbol{\beta}\|_2 = 1\}$ 使得

$$|\mathbb{A}| \le \left(1 + \frac{2}{\rho}\right)^k \qquad \text{and} \qquad \min_{\boldsymbol{\alpha} \in \mathbb{A}} \|\mathbf{z} - \boldsymbol{\alpha}\|_2 \le \rho \quad \text{for all } \mathbf{z} \in \mathbb{U}.$$

也就是说，$\mathbb{A}$ 是其自身的 ρ-打包集（定义6.1）。对于任意 $\boldsymbol{\alpha} \in \mathbb{A}$，集中不等式（7.38）给出，当 $\tau \in (0, 1)$ 取决于 δ 和 ρ，后者将在稍后确定时，

$$\begin{aligned}\Pr\left(\left|\|\mathbf{X}\boldsymbol{\alpha}\|_2^2 - \|\boldsymbol{\alpha}\|_2^2\right| > \tau \|\boldsymbol{\alpha}\|_2^2\right) &\le \sum_{\boldsymbol{a} \in \mathbb{A}} \Pr\left(\left|\|\mathbf{X}\boldsymbol{a}\|_2^2 - \|\boldsymbol{a}\|_2^2\right| > \tau \|\boldsymbol{a}\|_2^2\right) \\ &\le 2\,|\mathbb{A}|\exp(-\widetilde{C}\tau^2 n) \le 2\left(1 + \frac{2}{\rho}\right)^k \exp(-\widetilde{C}\tau^2 n).\end{aligned}$$

由于 $\mathbb{A}$ 是单位球面 $\mathbb{U}$的子集，上述不等式表明，随机矩阵$\mathbf{X}$ 满足

$$\left|\|\mathbf{X}\boldsymbol{\alpha}\|_2^2 - \|\boldsymbol{\alpha}\|_2^2\right| \le \tau, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{A}. \tag{7.40}$$

以至少.....的概率

$$D \triangleq 1 - 2\left(1+\frac{2}{\rho}\right)^k \exp(-\widetilde{C}\tau^2 n). \tag{7.41}$$

现在我们证明，对于所有 $\boldsymbol{\beta} \in \mathbb{U}$，(7.40) 蕴含 $\left|\|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| \le \delta$；也就是说，只要适当选取 ρ、τ，就有 $\left\|\mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \le \delta$。令 $\mathrm{H} \triangleq \mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}$。那么，(7.40) 表示，对于所有 $\boldsymbol{\alpha} \in \mathbb{A}$，都有$|\langle \mathrm{H}\boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\alpha}_{\mathbb{S}}\rangle| \le \tau$，且其概率大于 D。现在考虑一个向量$\boldsymbol{\beta} \in \mathbb{U}$，为此我们选择另一个满足 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 \le \rho < 1/2$ 的向量 $\boldsymbol{\alpha} \in \mathbb{A}$。利用三角不等式和柯西–施瓦茨不等式（定理 1.59），可得

$$\begin{aligned}
|\langle \mathbf{H}\boldsymbol{\beta}_{\mathbb{S}}, \boldsymbol{\beta}_{\mathbb{S}}\rangle| &= |\langle \mathbf{H}\boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\alpha}_{\mathbb{S}}\rangle + \langle \mathbf{H}(\boldsymbol{\beta}_{\mathbb{S}} + \boldsymbol{\alpha}_{\mathbb{S}}), \boldsymbol{\beta}_{\mathbb{S}} - \boldsymbol{\alpha}_{\mathbb{S}}\rangle| \\
&\le |\langle \mathbf{H}\boldsymbol{\alpha}_{\mathbb{S}}, \boldsymbol{\alpha}_{\mathbb{S}}\rangle| + |\langle \mathbf{H}(\boldsymbol{\beta}_{\mathbb{S}} + \boldsymbol{\alpha}_{\mathbb{S}}), \boldsymbol{\beta}_{\mathbb{S}} - \boldsymbol{\alpha}_{\mathbb{S}}\rangle| \\
&\le \tau + \|\mathbf{H}\|_2 \|\boldsymbol{\beta}_{\mathbb{S}} + \boldsymbol{\alpha}_{\mathbb{S}}\|_2 \|\boldsymbol{\beta}_{\mathbb{S}} - \boldsymbol{\alpha}_{\mathbb{S}}\|_2 \le \tau + 2\|\mathbf{H}\|_2 \rho.
\end{aligned}$$

根据谱范数的定义（定义1.31），并对所有$\boldsymbol{\beta} \in \mathbb{U}$取上确界，可得

$$\|\mathbf{H}\|_2 \le \tau + 2\|\mathbf{H}\|_2 \rho \quad \Longrightarrow \quad \|\mathbf{H}\|_2 \le \frac{\tau}{1-2\rho}.$$

选取 $\tau \triangleq (1-2\rho)\delta \in (0,1)$ 可确保 $\|\mathbf{H}\|_2 \le \delta$。由（7.41）可得

$$\Pr\left(\left\|\mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \le \delta\right) \ge 1 - 2\left(1+\frac{2}{\rho}\right)^k \exp(-\widetilde{C}(1-2\rho)^2\delta^2 n). \tag{7.42}$$

Thus，事件 $\left\|\mathbf{X}_{\mathbb{S}}^{\top}\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \le \delta$ 发生的概率至少为 $1-\varepsilon$，只要

$$n \ge \frac{1}{\widetilde{C}(1-2\rho)^2\delta^2}\left(k\ln(1+2/\rho) + \ln(2/\varepsilon)\right).$$

选择 $\rho \triangleq 0.2/(e^{3/2}-1) < 0.5$ 可使 $(\ln(1+2/\rho))/(1-2\rho)^2 \le 14/3$ 且 $1/(1-2\rho)^2 \le 4/3$。因此，当满足以下条件时，关于 n 的所需条件即可成立

$$n \ge \frac{2}{3\widetilde{C}\delta^2}\left(7k + 2\ln(2/\varepsilon)\right),$$

至此，证明完毕。 ■

将以上论证中的坐标子空间替换为由 $\mathbb{S}$索引的任意 k维子空间（属于 $\mathbb{R}^p$），论证依然成立。此外，结果仅取决于由$\mathbf{X}$ 的 $\mathbb{S}$索引的列；因此，对于 $n \times k$ 次高斯随机矩阵$\mathbf{Z}$，结论同样成立（参见例7.18）。特别地，对于这样的矩阵，

$$\left\|\frac{1}{n}\mathbf{Z}^{\top}\mathbf{Z} - \boldsymbol{I}\right\|_2 \le \delta$$

以至少 $1-\varepsilon$ 为概率，前提是满足条件（7.39）。

现在给出本节的主要结果：元素表现出次高斯行为的随机矩阵的限制等距性质。

定理 7.25：（通过尾界得到的限制等距性质）设 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 是从满足集中不等式（7.23）的分布中抽取的随机矩阵：

$$\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| > \tau\|\boldsymbol{\beta}\|_2^2\right) \le 2\exp(-\widetilde{C}\tau^2 n), \quad \text{for all } \tau \in (0,1) \text{ and } \boldsymbol{\beta} \in \mathbb{R}^p.^{a}$$

等价地，定理7.19 表明 $\sqrt{n}\mathbf{X}$具有独立、各向同性且次高斯的行，且具有相同的次高斯参数 C （7.20）。则以下结论成立：

- If
$$n \ge \frac{2}{3\widetilde{C}\delta^2}\left[k\ln(e^9(p/k)^2) + 2\ln(2/\varepsilon)\right], \quad \text{for } \delta,\varepsilon \in (0,1),$$
那么，以至少 $1-\varepsilon$ 的概率，$\mathbf{X}$ 的 k 阶限制等距常数 δ_k 满足 $\delta_k \le \delta$。
- 令 $\varepsilon =$exp(−3 $\widetilde{C}\delta^2 n/8)$ 得到条件
$$n \ge \frac{8}{3\widetilde{C}\delta^2}k\ln\left(\frac{ep}{k}\right) + \frac{12k}{\widetilde{C}\delta^2}, \tag{7.43}$$
这保证了 $\delta_k \le \delta$，其概率至少为 $1 - 2\text{exp}\left(-\delta^2 n/(2C)\right)$.

a。或者集中不等式（7.34）以 $\widetilde{C} = 1/\kappa = (1-\ln(2))/2$成立。

定理 [的证明 7.25] 单个 k 列子矩阵 $\mathbf{X}_{\mathbb{S}}$ （其中 $|\mathbb{S}| = k$）的条件数已在命题 7.24中进行了分析；这里采用相同的符号。回顾定义 4.30 ，阶数为 k 的限制等距常数定义为

$$\delta_k = \max_{\mathbb{S}\subset\{1,2,\ldots,p\},|\mathbb{S}|=k}\left\|\mathbf{X}_{\mathbb{S}}^\top\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2.$$

对所有大小为 $\binom{p}{k}$ 的子集 $\mathbb{S}\subset\{1,2,\ldots,p\}$ 进行并集界估计 k，并利用（7.42），得到

$$\begin{aligned}\Pr(\delta_k > \delta) &\le \sum_{\mathbb{S}\subset\{1,2,\ldots,p\},|\mathbb{S}|=k}\Pr\left(\left\|\mathbf{X}_{\mathbb{S}}^\top\mathbf{X}_{\mathbb{S}} - \boldsymbol{I}\right\|_2 \ge \delta\right)\\ &\le 2\binom{p}{k}\left(1+\frac{2}{\rho}\right)^k\exp(-\widetilde{C}(1-2\rho)^2\delta^2 n)\\ &\le 2\left(\frac{ep}{k}\right)^k\left(1+\frac{2}{\rho}\right)^k\exp(-\widetilde{C}(1-2\rho)^2\delta^2 n),\end{aligned}$$

其中，最后一个不等式由问题 6.2 得出。这表明，事件 $\delta_k \le \delta$ 以至少 $1-\varepsilon$ 的概率成立，条件是

$$n \ge \frac{1}{\widetilde{C}(1-2\rho)^2\delta^2}\left(k\ln(1+2/\rho) + k\ln(ep/k) + \ln(2/\varepsilon)\right).$$

再一次，取 $\rho \triangleq 0.2/(e^{3/2}-1) < 0.5$ ，可得 $\delta_k \le \delta$ 以至少$1-\varepsilon$ 的概率成立，条件是

$$n \ge \frac{1}{\widetilde{C}\delta^2}\left(\frac{4}{3}k\ln(ep/k) + \frac{14}{3}k + \frac{4}{3}\ln(2/\varepsilon)\right).$$

证明至此完成。 ■

由于高斯分布矩阵和拉德马赫随机矩阵的元素都是方差为 1 的次高斯变量（参见例 7.18 和习题7.2），因此可立即得出，这些矩阵在条件（7.43）下满足 RIP。

推论 7.26 设 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 为高斯分布或拉德马赫随机矩阵。那么存在一个通用常数 $C>0$，使得 $\frac{1}{\sqrt{n}}\mathbf{X}$的受限等距常数满足 $\delta_k \le \delta$，其成立概率至少为 $1-\varepsilon$，条件是

$$n \ge C\delta^{-2}(k\ln(ep/k) + \frac{7}{2}k + \ln(2/\varepsilon)). \tag{7.44}$$

归一化 $\frac{1}{\sqrt{n}}\mathbf{X}$ 是自然的：对于任意固定向量 $\boldsymbol{\beta}\in\mathbb{R}^p$，我们有

$$\mathbb{E}[\|\frac{1}{\sqrt{n}}\mathbf{X}\boldsymbol{\beta}\|_2^2] = \|\boldsymbol{\beta}\|_2^2,$$

条件是 $\mathbf{X}$ 的元素相互独立、均值为零且具有单位方差（高斯系综、拉德马赫系综或 $\mathcal{SSG}(1/n)$ 变量均满足这些条件）。因此，受限等距常数 δ_k 量化了 $\|\frac{1}{\sqrt{n}}\mathbf{X}\boldsymbol{\beta}\|_2^2$ 偏离其期望的最大程度，且这一结论对所有 k-稀疏向量 $\boldsymbol{\beta}$均成立。

对于高斯矩阵，我们将在下一小节中通过明确给出常数，对（7.44）稍作改进。

7.5.2 SSG 随机系综下的限制等距性质

现在，我们将说明如何将推论 7.23 中的测度集中不等式（该推论指出，次高斯随机向量的范数会强烈集中于其均值附近）与覆盖论证结合起来，以证明次高斯随机矩阵的限制等距性质。同样，我们的一般方法仍是：在每个 k维子空间中构造点网，通过并集界将（7.34）应用于所有这些点，然后将结果从有限点集推广到所有可能的 k维信号。文献中常见如下结论。

定理 7.27：（SSG 下的限制等距性质，Baraniuk 等人（2008年）的定理 5.2）取定$\delta \in (0,1)$。设 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 为一个随机矩阵，其元素 x_{ij} 独立同分布，且 $\mathrm{x}_{ij} \sim \mathcal{SSG}(1/n)$。若

$$n \ge \kappa_1 k \ln\left(\frac{p}{k}\right), \tag{7.45}$$

则 $\mathbf{X}$ 以超过 $1-2\exp(-\kappa_2 n)$ 的概率满足具有给定常数 δ 的 k 阶限制等距性质（RIP）。这里，$\kappa_1 > 1$ 是任意常数，且 $\kappa_2 = \delta^2/2\kappa - 1/\kappa_1 - \ln(42e/\delta)$，其中 $\kappa = 2/(1-\ln(2)) \approx 6.52$。

证明 [定理 7 的27] 首先，注意到只需验证满足 7.36 的向量所满足的不等式（$\boldsymbol{\beta}$ $\|\boldsymbol{\beta}\|_2 = 1$，因为 $\mathbf{X}$ 是线性的。接下来，固定一个指标集 $\mathbb{S}\subset\{1,2,\ldots,p\}$，其$|\mathbb{S}| = k$，并令 $\mathbb{C}_{\mathbb{S}}$ 表示由 k 维子空间，该子空间由 $\mathbf{X}_{\mathbb{S}}$ 的列张成。我们构造一个有限点集 $\mathbb{A}_{\mathbb{S}} \subseteq \mathbb{C}_{\mathbb{S}}$，使得 $\|\boldsymbol{\alpha}\|_2 = 1$ 对所有 $\boldsymbol{\alpha}\in\mathbb{A}_{\mathbb{S}}$，且对每个$\boldsymbol{\beta}\in\mathbb{C}_{\mathbb{S}}$ 满足 $\|\boldsymbol{\beta}\|_2 = 1$的，都存在某个 $\boldsymbol{\alpha}\in\mathbb{A}_{\mathbb{S}}$ 满足：

$$\min_{\boldsymbol{\alpha}\in\mathbb{A}_{\mathbb{S}}} \|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2 \le \delta/14.$$

根据定理 6.3，这样的集合 $\mathbb{A}_{\mathbb{S}}$ 可以选取得基数至多为 $|\mathbb{A}_{\mathbb{S}}| \leq (42/\delta)^k$。对每个大小为 $\mathbb{S}$ 的可能指标集重复这一构造，我们定义并集

$$\mathbb{A} = \bigcup_{\mathbb{S}:|\mathbb{S}|=k} \mathbb{A}_{\mathbb{S}}.$$

共有 $\binom{p}{k}$ 个这样的指标集 $\mathbb{S}$。利用问题 6.2，可以将这个数量上界估计为$\binom{p}{k} \leq \left(\frac{ep}{k}\right)^k$。因此，$|\mathbb{A}| \leq (42ep/\delta k)^k$。由于 $\mathbf{X}$ 的元素服从 SSG 分布，推论 7.23 推出不等式（7.34）：

$$\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}\|_2^2 - \|\boldsymbol{\beta}\|_2^2\right| \geq \varepsilon \|\boldsymbol{\beta}\|_2^2\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{\kappa}\right), \quad \text{for any } \boldsymbol{\beta} \in \mathbb{R}^p.$$

对所有满足 $\varepsilon = \delta/\sqrt{2}$ 的 $\boldsymbol{\alpha} \in \mathbb{S}$ 应用并集界，可知该结论成立的概率至少为

$$\widetilde{p} \triangleq 1 - 2\left(\frac{42ep}{\delta k}\right)^k \exp\left(-\frac{n\delta^2}{2\kappa}\right), \tag{7.46}$$

以下结论对所有 $\boldsymbol{\alpha} \in \mathbb{A}$ 均成立：

$$(1-\delta/\sqrt{2})\|\boldsymbol{\alpha}\|_2^2 \leq \|\mathbf{X}\boldsymbol{\alpha}\|_2^2 \leq (1+\delta/\sqrt{2})\|\boldsymbol{\alpha}\|_2^2, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{A}. \tag{7.47}$$

现在，若 n 满足条件（7.45），则

$$\ln\left(\frac{42ep}{\delta k}\right)^k \leq k\left(\ln\left(\frac{p}{k}\right) + \ln\left(\frac{42e}{\delta}\right)\right) \leq \frac{n}{\kappa_1} + n\ln\left(\frac{42e}{\delta}\right).$$

因此，$\widetilde{p}$ 在（7.46）中的概率大于 $1 - 2\exp(-\kappa_2 n)$，正如所需.

然而，(7.47) 对所有 $\boldsymbol{\alpha} \in \mathbb{S}$均成立。现在我们将 D 定义为满足以下条件的最小数

$$\|\mathbf{X}\boldsymbol{\beta}\|_2 \leq \|\mathbf{X}\|_2 \|\boldsymbol{\beta}\|_2 \leq \sqrt{1+D}, \quad \text{for all } \boldsymbol{\beta} \in \mathbb{B}_0[k],\ \|\boldsymbol{\beta}\|_2 = 1, \tag{7.48}$$

其中，第一个不等式由谱范数的定义得出。为了证明 RIP 条件，我们的目标是证明 $D \leq \delta$。为此，根据集合 $\mathbb{A}$ 的构造，对于任意满足 $\|\boldsymbol{\beta}\|_2 = 1$ 的 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$，存在一个向量 $\boldsymbol{\alpha} \in \mathbb{A}$，使得 $\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2 \leq \delta/14$ 且 $\boldsymbol{\beta} - \boldsymbol{\alpha} \in \mathbb{B}_0[k]$。应用三角不等式，

$$\|\mathbf{X}\boldsymbol{\beta}\|_2 \leq \|\mathbf{X}\boldsymbol{\alpha}\|_2 + \|\mathbf{X}(\boldsymbol{\beta}-\boldsymbol{\alpha})\|_2 \leq \sqrt{1+\delta/\sqrt{2}} + \sqrt{1+D}\cdot\delta/14.$$

由于 D 被定义为满足 (7.48) 的最小常数，因此 $\sqrt{1+D} \leq \sqrt{1+\delta/\sqrt{2}} + \sqrt{1+D}\cdot\delta/14$。重新整理可得

$$\sqrt{1+D} \leq \frac{\sqrt{1+\delta/\sqrt{2}}}{1-\delta/14} \leq \sqrt{1+\delta}.$$

这就得到了 RIP 条件的上界。下界的证明类似：

$$\|\mathbf{X}\boldsymbol{\beta}\|_2 \geq \|\mathbf{X}\boldsymbol{\alpha}\|_2 - \|\mathbf{X}(\boldsymbol{\beta}-\boldsymbol{\alpha})\|_2 \geq \sqrt{1-\delta/\sqrt{2}} - \sqrt{1+\delta}\cdot\delta/14 \geq \sqrt{1-\delta},$$

至此，证明完成。 ■

注意到，鉴于定理 4.44 所确立的 RIP 测量数量上界，条件（7.45）实现了最优的测量数量（相差一个常数因子）。

7.5.3 其他问题

使用随机矩阵构造$\mathbf{X}$ 还有若干额外的好处。为了说明这些好处，我们再次聚焦于 RIP 条件。

首先，随机构造会产生民主测量：只要测量值的子集足够大，就足以实现稳定信号恢复（Davenport 等人，2009年；Laska 等人，2011年）。因此，基于随机$\mathbf{X}$的系统天然能够抵御少量测量值的丢失或损坏。

其次——或许更重要的是——在实践中，我们通常关注的信号并非在标准基下稀疏，而是在某个其他正交归一基下稀疏 $\boldsymbol{Q}$；参见下一节。在这种情况下，我们实际需要的是乘积$\mathbf{X}\boldsymbol{Q}$ 满足限制等距性质（RIP）。如果$\mathbf{X}$ 是通过确定性方式构造的，我们就需要根据 $\boldsymbol{Q}$ 的结构专门设计它。然而，当$\mathbf{X}$ 是随机抽取的，这种依赖关系便不复存在。例如，如果 $\mathbf{X}$ 的元素为独立同分布的高斯分布，且 $\boldsymbol{Q}$ 是任意固定的正交归一基，那么$\mathbf{X}\boldsymbol{Q}$ 的元素也服从独立同分布的高斯分布（见引理1.90）。因此，只要n 足够大，$\mathbf{X}\boldsymbol{Q}$就以高概率满足限制等距性质（RIP）——正如$\mathbf{X}$ 本身所满足的那样。尽管不那么显而易见，次高斯系综也有类似的结果（Baraniuk 等人，2008年）。这一理想特征通常称为普适性，它是随机矩阵构造的一大优势。

另一方面，Agarwal 等人 (2010年) 的工作表明，只要X 的各行独立地从定义在 $\mathbb{R}^p$上的次高斯分布中抽取，且其协方差矩阵非奇异，受限强凸性（RSC）和限制强光滑性（RSS）性质（定义 4.33）便成立。这一结果尤其有价值，因为许多现实数据集（例如基因表达分析中产生的数据）都可以合理地建模为次高斯随机向量。因此，以高概率，这类数据矩阵满足 RSC/RSS，从而能够在实际应用中使用稀疏恢复算法。

如果愿意接受 X 的行数略有增加，就可以采用效率更高的构造，尤其是那些支持快速矩阵-向量乘法的构造。Candes 和 Tao（2005年）的开创性工作表明，以概率 $\mathcal{O}\left(k \ln^4 p/p\right)$ 独立选取傅里叶矩阵的每一行，所得矩阵以高概率满足限制等距性质（RIP）。随后，Haviv 和 Regev（2017年）改进了这一结果，将所需的抽样速率降至 $\mathcal{O}\left(k \ln^2 k \ln(p)/p\right)$。关键的是，将这样的结构化矩阵与一个 k-稀疏向量相乘只需 $\mathcal{O}(k \ln^2 p)$ 时间；相比之下，稠密高斯矩阵需要 $\mathcal{O}(k^2 \ln p)$ 时间（采用典型缩放 $n \sim k\ln(p/k)$ 时）。

7.6. 次高斯系综下的稀疏优化

为此，请注意，迄今为止介绍的所有 ℓ_1-最小化结果都是确定性的：只要测量矩阵 $\boldsymbol{\beta}$ 满足限制等距性质（见第 $\mathbf{X}$5.3节），它们便对所有向量同时成立。这是一项强大的理论性质。然而，如前所述，在实践中通常无法验证或保证给定的确定性矩阵满足限制等距性质。事实上，目前已知的、能够证明满足限制等距性质的构造，都依赖随机性来以高概率实现这一点。例如，回想定理 7.25或定理 7.27，它们说明随机性如何带来略弱一些、但仍极具实用价值的概率保证。

现在，我们将本章的结果结合起来，陈述通过 ℓ_1-最小化问题（P_1）（第129页）利用随机测量进行稀疏恢复的主要定理。

定理 7.28：（在次高斯条件下恢复（P_1））设X是一个具有独立各向同性次高斯行的 $n\times p$ 随机矩阵。设 $k<p$，$\varepsilon\in$（0，1），并假设

$$n\geq C_1k\ln(ep/k)+C_2k+C_3\ln(2/\varepsilon),$$

其中常数 C_1、C_2、$C_3>0$ 仅取决于次高斯参数 Φ、ξ。那么，以至少 $1-\varepsilon$ 的概率，每个 k-稀疏向量 $\boldsymbol{\beta}$ 都可以通过求解$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}$ 得到精确恢复，方法是求解 ℓ_1-最小化问题（P_1）。

定理 [7. 的28] 证明。该结论结合定理 7.25 （该定理保证次高斯矩阵以高概率满足 RIP）和定理 5.14 （该定理保证在 RIP 条件下能够精确恢复）即可得出。注意，精确稀疏恢复对测量矩阵的重新缩放具有不变性，因此归一化不会影响结论。

■

7.6.1 正交归一基下的稀疏性

在许多应用中，稀疏性并非相对于典范基表示，而是相对于某个其他标准正交基表示。也就是说，感兴趣信号具有 $\boldsymbol{\alpha}=\boldsymbol{Q}\boldsymbol{\beta}$ 的形式，其中$\boldsymbol{Q}\in\mathbb{R}^{p\times p}$ 是一个正交矩阵，而 $\boldsymbol{\beta}\in\mathbb{R}^p$ 是 k-稀疏的。用随机矩阵 $\mathbf{X}$ 对 $\boldsymbol{\alpha}$ 进行测量，得到

$$\mathbf{y}=\mathbf{X}\boldsymbol{\alpha}=\mathbf{X}\boldsymbol{Q}\boldsymbol{\beta}.$$

要恢复 $\boldsymbol{\alpha}$，只需先恢复稀疏系数向量 $\boldsymbol{\beta}$，然后计算 $\boldsymbol{\alpha}=\boldsymbol{Q}\boldsymbol{\beta}$。因此，这一广义情形可归结为具有有效测量矩阵 $\mathbf{X}'=\mathbf{X}\boldsymbol{Q}$ 的标准压缩感知问题。

因此，我们考虑这样的模型：$\mathbf{X}$ 是一个随机 $n\times p$ 矩阵，而 $\boldsymbol{Q}$ 是一个固定的（确定性）正交矩阵。重要的是，前几节中建立的分析自然适用于这一情形。

推论 7.29 设 $\boldsymbol{Q}\in\mathbb{R}^{p\times p}$ 为一个（固定的）正交矩阵。假设 $n\times p$ 随机矩阵 $\mathbf{X}$ 满足集中不等式

$$\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}\|_2^2-\|\boldsymbol{\beta}\|_2^2\right|>\tau\|\boldsymbol{\beta}\|_2^2\right)\leq 2\exp(-\widetilde{C}\tau^2n),\quad \text{for all } \tau\in(0,1) \text{ and } \boldsymbol{\beta}\in\mathbb{R}^p, \tag{7.49}$$

其中某个常数为 $C>0$。设 δ，$\varepsilon\in$（0，1）。则$\mathbf{X}\boldsymbol{Q}$ 的受限等距常数 δ_k 满足 $\delta_k\leq\delta$，其概率至少为 $1-\varepsilon$，前提是

$$n\geq\frac{1}{\widetilde{C}\delta^2}\left(\frac{4}{3}k\ln(ep/k)+\frac{14}{3}k+\frac{4}{3}\ln(2/\varepsilon)\right).$$

证明 [推论 7.29] 关键观察是，集中不等式（7.49）对矩阵 $\mathbf{X}\boldsymbol{Q}$也成立。事实上，对于任意 $\boldsymbol{\beta}\in\mathbb{R}^p$，定义 $\boldsymbol{\beta}'\triangleq\boldsymbol{Q}\boldsymbol{\beta}$。正交性$\boldsymbol{Q}$ 表明

$$\begin{aligned}\Pr\left(\left|\|\mathbf{X}\boldsymbol{Q}\boldsymbol{\beta}\|_2^2-\|\boldsymbol{\beta}\|_2^2\right|>\tau\|\boldsymbol{\beta}\|_2^2\right)&=\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}'\|_2^2-\|\boldsymbol{Q}^{-1}\boldsymbol{\beta}'\|_2^2\right|>\tau\|\boldsymbol{Q}^{-1}\boldsymbol{\beta}'\|_2^2\right)\\&=\Pr\left(\left|\|\mathbf{X}\boldsymbol{\beta}'\|_2^2-\|\boldsymbol{\beta}'\|_2^2\right|>\tau\|\boldsymbol{\beta}'\|_2^2\right)\leq 2\exp(-\widetilde{C}\tau^2n).\end{aligned}$$

因此，该结论可直接由定理 7.25得出。

■

这一结果表明，使用次高斯随机矩阵进行稀疏恢复时，无论信号在哪个标准正交基下是稀疏的，恢复效果都具有普适性。关键在于，正交矩阵 $\boldsymbol{Q}$ 可以是任意的——尤其是，在获取测量值$\mathbf{y} = \mathbf{X}\boldsymbol{Q}\boldsymbol{\beta}$ 的编码阶段，无须知道该矩阵。只有在解码时求解 ℓ_1-最小化问题，才需要知道 $\boldsymbol{Q}$ 。

然而，需要强调的是，这种普适性是概率性的，并且依赖于基：这并不意味着同一个固定矩阵 $\boldsymbol{X}$ 能够同时适用于所有可能的基。事实上，对于任意给定的确定性 $\boldsymbol{X}$，都可以构造一个使稀疏恢复失败的标准正交基$\boldsymbol{Q}$ 。该推论仅保证：对于任意固定的 $\boldsymbol{Q}$，随机抽取的 $\mathbf{X}$ 以高概率实现成功恢复。

7.6.2 严格次高斯条件下的稀疏恢复

为了建立鲁棒恢复保证，我们依赖这样一个事实：严格次高斯随机矩阵以高概率保持任意固定向量的欧几里得范数。具体而言，对推论 7.23 稍作修改可得，对于任意 $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\Pr\left(\|\mathbf{X}\boldsymbol{\beta}\|_2^2 \geq 2\,\|\boldsymbol{\beta}\|_2^2\right) \leq \exp(-\kappa_3 n), \quad \text{with } \kappa_3 \triangleq 1/\kappa \text{ and } \kappa = 2/(1-\ln(2)) \approx 6.52. \tag{7.50}$$

现在我们利用这一集中性质推导以下结果。

定理 7.30：（SSG 条件下（P_1）的误差界）设 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 为一个固定向量，并假设 $\delta_{2k} < \sqrt{2}-1$。设 $\mathrm{X} \in \mathbb{R}^{n\times p}$ 为严格次高斯随机矩阵（$\mathrm{x}^{ij} \sim \mathcal{SSG}(1/n)$），满足 $n \geq 2\kappa_1 k\ln(p/2k)$。假设我们观测到测量值 $\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta}^*$。那么，以至少 $1-2\exp(-\kappa_2 n) - \exp(-\kappa_3 n)$ 的概率，问题(P_1)的解 $\widehat{\boldsymbol{\beta}}$ 可归约为问题（P1,ϵ）（第131页），其中 $\epsilon = 2\sigma_k(\boldsymbol{\beta}^*)_2,$，且 $\boldsymbol{\beta}$ 满足

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 \leq \frac{8\sqrt{1+\delta_{2k}} - (1+\sqrt{2})\delta_{2k} + 1}{1-(1+\sqrt{2})\delta_{2k}}\sigma_k(\boldsymbol{\beta}^*)_2. \tag{7.51}$$

证明 [定理 7.30] 首先，根据定理 7.27，只要 $n \geq 2\kappa_1 k\ln(p/2k)$，矩阵 X 就以至少 $1-2\exp(-\kappa_2 n)$ 的概率满足 $2k$ 阶 RIP。接下来，令 $\mathbb{S}$ 表示 $\boldsymbol{\beta}^*$ 中大小最大的 k 个元素所对应的指标集，并将$\boldsymbol{\beta}^*$ 分解为 $\boldsymbol{\beta}^* = \boldsymbol{\beta}^*(\mathbb{S}) + \boldsymbol{\beta}^*(\mathbb{S}^c)$；注意，$\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_2 = \sigma_k(\boldsymbol{\beta}^*)_2$ 是最佳 k 项近似误差。由于 X 是严格次高斯的，命题 7.14 表明向量 $\mathrm{X}_{\mathbb{S}^c}\boldsymbol{\beta}^*_{\mathbb{S}^c}$ 也是严格次高斯的。应用式 (7.50)，可得以至少$1-\exp(-\kappa_3 n)$ 的概率，$\|\mathrm{X}_{\mathbb{S}^c}\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_2 \leq 2\,\|\boldsymbol{\beta}^*_{\mathbb{S}^c}\|_2 = 2\sigma_k(\boldsymbol{\beta}^*)_2$。现在注意到 $\boldsymbol{\beta}^*(\mathbb{S}) \in \mathbb{B}_0[k]$，因此可以将测量向量写为 $\mathrm{y} = \mathrm{X}\boldsymbol{\beta}^* = \mathrm{X}_{\mathbb{S}}\boldsymbol{\beta}^*_{\mathbb{S}} + \mathrm{X}_{\mathbb{S}^c}\boldsymbol{\beta}^*_{\mathbb{S}^c} = \mathrm{X}_{\mathbb{S}}\boldsymbol{\beta}^*_{\mathbb{S}} + \boldsymbol{\epsilon}$，其中$\|\boldsymbol{\epsilon}\|_2 \leq \epsilon$。根据并集界，RIP 条件（对于 $2k$ 阶）和上述噪声上界同时成立的概率至少为 $1-2\exp(-\kappa_2 n) - \exp(-\kappa_3 n)$。在这些条件下，我们可以应用定理 5.20，从含噪测量 $\mathrm{y} = \mathrm{X}_{\mathbb{S}}\boldsymbol{\beta}^*_{\mathbb{S}} + \boldsymbol{\epsilon}$ 中恢复 $\boldsymbol{\beta}^*(\mathbb{S})$。由于 $\boldsymbol{\beta}^*(\mathbb{S})$ 恰好是 k-稀疏的，其最佳 k 项 ℓ_1 近似误差为零，即 $\sigma_k(\boldsymbol{\beta}^*(\mathbb{S}))_1 = 0$。因此，

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*(\mathbb{S})\right\|_2 \leq C_2\epsilon \leq 2C_2\sigma_k(\boldsymbol{\beta}^*)_2,$$

其中 $C_2 = 4\frac{\sqrt{1+\delta_{2k}}}{1-\sqrt{1+\sqrt{2}}()\delta_{2k}}$，如定理 5.20中所定义。最后，根据三角不等式，

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2 = \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*(\mathbb{S}) + \boldsymbol{\beta}^*(\mathbb{S}) - \boldsymbol{\beta}^*\right\|_2 \leq \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*(\mathbb{S})\right\|_2 + \|\boldsymbol{\beta}^*(\mathbb{S}) - \boldsymbol{\beta}^*\|_2 \leq (2C_2+1)\sigma_k(\boldsymbol{\beta}^*)_2,$$

将 C_2 的表达式代入并化简分子后，即可得到所声称的误差上界。

■

因此，尽管要获得（5.30）形式的确定性保证，需要的测量数量会大得令人难以承受，但我们仍有可能以很高的概率，用少得多的测量值实现相近的性能——所需数量显著少于定理 5.19 所暗示的数量。

不过，我们注意到，上述结果假设噪声水平参数 ϵ 选取为 $2\sigma_k(\boldsymbol{\beta}^*)_2$，，这需要对 $\boldsymbol{\beta}^*$ 有部分了解。在实践中，可以采用交叉验证（Ward，2009年）等数据驱动的参数选择方法来克服这一限制。此外，对 ℓ_1-最小化进行更精细的分析表明，无需预言机知识 $\sigma_k(\boldsymbol{\beta}^*)_2$ (Wojtaszczyk，2010年），也能获得类似的恢复保证。

第7章习题

1. 证明函数 $s^{1/s}$ 在 $s = e$处取得最大值，因此对所有 $s > 0$，都有 $s^{1/s} \leq e^{1/e}$。提示：对该函数取对数并求其导数。
2. 证明拉德马赫随机变量的均值为零、方差为一 e.
3. 陈述并证明 JL 引理（定理6.14）的次高斯版本。
4. 次高斯定义。设 x 是 $\mathbb{R}$上的一个中心化随机变量。证明以下每个命题都蕴含下一个命题（假设第一个条件中的 $\sigma^2 > 0$ 是一个方差代理）。

   - 拉普拉斯变换。对于任意 $\theta \in \mathbb{R}, \mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp\left(\frac{\sigma^2\theta^2}{2}\right)$.尾部估计。对于任意 $t > 0, \max\{\Pr(\mathrm{x} \geq t), \Pr(\mathrm{x} \leq -t)\} \leq \exp\left(\frac{-t^2}{2\sigma^2}\right)$.矩条件。对于任意正整数 $k, \mathbb{E}\left[\mathrm{x}^{2k}\right] \leq k!(4\sigma^2)^k$.奥利奇条件。 $\mathbb{E}\left[\exp\left(\frac{\mathrm{x}^2}{8\sigma^2}\right)\right] \leq 2$.拉普拉斯变换。对于任意 $\tau \in \mathbb{R}, \mathbb{E}[\exp(\tau \mathrm{x})] \leq \exp\left(\frac{24\sigma^2\tau^2}{2}\right)$.

5. 设 x 是一个实值中心化随机变量，且 $\mathrm{x} \in [a, b]$ 几乎必然成立。证明

$$\mathbb{E}[\exp(\theta \mathrm{x})] \leq \exp\left(\theta^2 \frac{(b-a)^2}{8}\right)$$

对于任意 $\theta \in \mathbb{R}$，并得出结论：x 是方差代理为 $(b-a)^2/4$ 的次高斯随机变量。

6. **次高斯向量**。若随机向量 $\mathbf{x} \in \mathbb{R}^p$ 是方差代理为 c^2 的次高斯向量，记作 $\mathbf{x} \sim \mathcal{SG}(c^2)$，则它是中心化的，并且对于任意 $\boldsymbol{a} \in \mathbb{R}^p$ 满足$\|\boldsymbol{a}\|_2 = 1$的 $\boldsymbol{a}^\top \mathbf{x}$，标量随机变量 x 是方差代理为该值的次高斯随机变量 c^2（定义 7.16）。

   设 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_p$ 是相互独立的 $\mathcal{SG}(c^2)$ 随机变量。证明向量$\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_p]^\top$ 是方差代理为 c^2 的次高斯向量。
7. **样本均值的集中性**。设 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_p$是相互独立的 $\mathcal{SG}(c^2)$ 随机变量。证明

$$\Pr\left(\frac{1}{p}\sum_{i=1}^{p} \mathrm{x}_i \geq \tau\right) \leq \exp\left(\frac{-\tau^2 p}{2c^2}\right).$$

8. 设 $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n \sim \mathcal{SG}(c^2)$ 是一个零均值次高斯随机变量序列。对于任意 $\delta > 0$，证明以下结论以至少 $1-\delta$ 的概率成立，

$$\max_{i=1,2\ldots,n} |\mathrm{x}_i| \leq c\sqrt{2\ln(2n/\delta)}.$$

提示：使用命题*7.15*。

9. 设$\mathbf{x} \sim \mathcal{SG}(c^2)$为一个 p维随机向量。证明：

$$\mathbb{E}\left[\max_{\|\boldsymbol{a}\|_2 \leq 1} \boldsymbol{a}^\top \mathbf{x}\right] = \mathbb{E}\left[\max_{\|\boldsymbol{a}\|_2 \leq 1} \left|\boldsymbol{a}^\top \mathbf{x}\right|\right] \leq 4c\sqrt{p}.$$

此外，对于任意 $\tau > 0$，证明：

$$\Pr\left[\max_{\|\boldsymbol{a}\|_2 \leq 1} \left|\boldsymbol{a}^\top \mathbf{x}\right| > \tau\right] = \Pr\left[\max_{\|\boldsymbol{a}\|_2 \leq 1} \boldsymbol{a}^\top \mathbf{x} > \tau\right] \leq 6^p \exp\left(-\frac{\tau^2}{8c^2}\right).$$

对于任意 $\delta > 0$，取$\tau = \sqrt{8\ln(6)c\sqrt{p} + 2c\sqrt{2\ln(1/\delta)}}$，证明以概率$1-\delta$有：

$$\max_{\|\boldsymbol{a}\|_2 \leq 1} \boldsymbol{a}^\top \mathbf{x} = \max_{\|\boldsymbol{a}\|_2 \leq 1} \left|\boldsymbol{a}^\top \mathbf{x}\right| \leq 4c\sqrt{p} + 2c\sqrt{2\ln(1/\delta)} = 4c\sqrt{p}\left(1 + \sqrt{\frac{\ln(1/\delta)}{2p}}\right).$$

提示：使用定理*6.3*，半径为 $\varepsilon = 1/2$。

10. 设 $\mathrm{x}_1 \sim \mathcal{SG}(c_1^2)$ 和 $\mathrm{x}_2 \sim \mathcal{SG}(c_2^2)$。证明 $|\gamma|\mathrm{x}_1 \sim \mathcal{SG}(\gamma^2 c_1^2)$ 以及 $\mathrm{x}_1 + \mathrm{x}_2 \sim \mathcal{SG}((c_1 + c_2)^2)$。证明如果 x_1 和 x_2 独立，则 $\mathrm{x}_1 + \mathrm{x}_2 \sim \mathcal{SG}(c_1^2 + c_2^2)$。11. 确定推论 7.26 中的常数 C。提示：使用推论 7.23。

第三部分

算法

第8章

稀疏优化问题的统一算法

目录

8.1 Fenchel 对偶定理：再探.303**8.2 原始-对偶算法.**304**8.3 原始-对偶框架下的不同问题 . . . . .** 306 8.3.1 拉格朗日 LASSO.306 8.3.2 约束 LASSO.307 8.3.3 ℓ_1-最小化.308 8.3.4 ℓ_1-最小化与含噪测量值.309**8.4 原始-对偶算法的最优性条件 . . . . .** 309**8.5 原始-对偶算法的收敛性.**310**8.6 原始-对偶算法的收敛速度 . . . . . .** 315**第8章习题 .**316

NW我们将在第9章和第10章中更详细地讨论用于求解稀疏回归和稀疏恢复问题的算法。本章主要介绍一种统一而简单的算法——原始-对偶算法——用于求解本书涵盖的四个核心优化问题，即拉格朗日 LASSO、约束 LASSO、 ℓ_1-最小化和使用含噪测量值的 ℓ_1-最小化。

8.1. Fenchel 对偶定理：再探

受这些考虑的启发，我们现在考察以下凸优化问题，该问题是本书的核心：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{X}\boldsymbol{\beta}) + G(\boldsymbol{\beta}), \tag{PG}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和 $F\colon \mathbb{R}^n \to (-\infty, \infty]$, $G\colon \mathbb{R}^p \to (-\infty, \infty]$ 是适当的凸函数。引入辅助变量 $\boldsymbol{\alpha} \triangleq \boldsymbol{X}\boldsymbol{\beta}$，可得到等价的约束形式：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p,\boldsymbol{\alpha}\in\mathbb{R}^n} F(\boldsymbol{\alpha}) + G(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\alpha} = \boldsymbol{0}. \tag{PG$'$}$$

该问题的拉格朗日对偶函数为

$$\begin{aligned} D(\boldsymbol{\lambda}) &= \min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left\{ F(\boldsymbol{\alpha}) + G(\boldsymbol{\beta}) + \left\langle \boldsymbol{X}^\top\boldsymbol{\lambda}, \boldsymbol{\beta} \right\rangle - \langle \boldsymbol{\lambda}, \boldsymbol{\alpha} \rangle \right\} \\ &= -\max_{\boldsymbol{\alpha}\in\mathbb{R}^n} \{\langle \boldsymbol{\alpha}, \boldsymbol{\lambda}\rangle - F(\boldsymbol{\alpha})\} - \max_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \left\langle \boldsymbol{\beta}, -\boldsymbol{X}^\top\boldsymbol{\lambda} \right\rangle - G(\boldsymbol{\beta}) \right\} \\ &= -F^*(\boldsymbol{\lambda}) - G^*(-\boldsymbol{X}^\top\boldsymbol{\lambda}), \end{aligned} \tag{8.1}$$

其中 F^* 和 G^* 分别表示 F 和 G的凸共轭函数（定义2.63）。因此，(PG) 的对偶问题为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} D(\boldsymbol{\lambda}) \equiv -F^*(\boldsymbol{\lambda}) - G^*(-\boldsymbol{X}^\top\boldsymbol{\lambda}). \tag{DG}$$

严格来说，(DG) 并非原始无约束问题 (PG) 的拉格朗日对偶，因为无约束问题不会自然地产生对偶变量。然而，由于 (PG) 及其约束重构形式 (PG$'$) 等价且具有相同的最优值，我们将 (DG) 称为 (PG) 的对偶问题。[1]

下面给出的 Fenchel 对偶定理（定理 2.119）的一个变体，为原始问题 (PG) 及其对偶问题 (DG) 之间强对偶性成立提供了充分条件。

推论 8.1：（Fenchel 对偶定理；Rockafellar（1970年），定理 31.1）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，且 $F\colon \mathbb{R}^n \to (-\infty, \infty]$ 和 $G\colon \mathbb{R}^p \to (-\infty, \infty]$ 为适当凸函数。假设 $\operatorname{dom}(F) = \mathbb{R}^n$ 或 $\operatorname{dom}(G) = \mathbb{R}^p$，并且存在某个$\boldsymbol{\beta} \in \mathbb{R}^p$ 使得 $\boldsymbol{X}\boldsymbol{\beta} \in \operatorname{dom}(F)$。如果 (PG) 和 (DG) 的最优值均能取到，则强对偶性成立：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{X}\boldsymbol{\beta}) + G(\boldsymbol{\beta}) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^n} -F^*(\boldsymbol{\lambda}) - G^*(-\boldsymbol{X}^\top\boldsymbol{\lambda}). \tag{8.2}$$

1. 一般而言，等价的原始形式可能会导出不同的对偶问题。

在与上述定理相同的条件下，强对偶性意味着拉格朗日函数具有鞍点性质(2.62)：

$$L((\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*), \boldsymbol{\lambda}) \leq L((\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*), \boldsymbol{\lambda}^*) \leq L((\boldsymbol{\beta}, \boldsymbol{\alpha}), \boldsymbol{\lambda}^*) \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \mathbb{R}^n, \tag{8.3}$$

其中（$\boldsymbol{\beta}^*$， $\boldsymbol{\alpha}^*$）和 $\boldsymbol{\lambda}^*$ 分别表示原问题和对偶问题的最优点。由（8.1）可知，最优值也可以表示为以下极小极大问题的解：

$$\begin{aligned}
&\min_{\boldsymbol{\beta}, \boldsymbol{\alpha} \in \mathbb{R}^p} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} F(\boldsymbol{\alpha}) + G(\boldsymbol{\beta}) + \left\langle \boldsymbol{X}^\top \boldsymbol{\lambda}, \boldsymbol{\beta} \right\rangle - \langle \boldsymbol{\lambda}, \boldsymbol{\alpha} \rangle \\
&= \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \left(\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} - \langle \boldsymbol{\lambda}, \boldsymbol{\alpha} \rangle + F(\boldsymbol{\alpha}) \right) + \left\langle \boldsymbol{X}^\top \boldsymbol{\lambda}, \boldsymbol{\beta} \right\rangle + G(\boldsymbol{\beta}) \\
&= \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\lambda} \rangle + G(\boldsymbol{\beta}) - F^*(\boldsymbol{\lambda}),
\end{aligned} \tag{8.4}$$

其中，最后一个等式直接由凸共轭的定义得出。交换最小值和最大值的顺序是合理的，因为如果（（$\boldsymbol{\beta}^*$， $\boldsymbol{\alpha}^*$） $\boldsymbol{\lambda}^*$）是完整拉格朗日函数 L（（$\boldsymbol{\beta}$， $\boldsymbol{\alpha}$） $\boldsymbol{\lambda}$）的一个鞍点，那么（$\boldsymbol{\beta}^*$, $\boldsymbol{\lambda}^*$）就是约化拉格朗日函数 $\mathcal{L}$（$\boldsymbol{\beta}$， $\boldsymbol{\lambda}$） $\triangleq \min_{\boldsymbol{\alpha}} L$（（$\boldsymbol{\beta}$， $\boldsymbol{\alpha}$） $\boldsymbol{\lambda}$）的一个鞍点。因此，任何原始-对偶最优对（$\boldsymbol{\beta}^*$， $\boldsymbol{\lambda}^*$）都满足鞍点问题

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\lambda} \rangle + G(\boldsymbol{\beta}) - F^*(\boldsymbol{\lambda}). \tag{8.5}$$

此外，约化拉格朗日函数 $\mathcal{L}$满足鞍点不等式：

$$\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\lambda}^*) \leq \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}^*) \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^n. \tag{8.6}$$

8.2. 原始-对偶算法

我们考虑如下形式问题的一般优化框架

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} F(\boldsymbol{X}\boldsymbol{\beta}) + G(\boldsymbol{\beta}), \tag{8.7}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，且 F： $\mathbb{R}^n \to (-\infty, \infty]$， G： $\mathbb{R}^p \to (-\infty, \infty]$ 均为真闭凸函数。由式（8.1），与式（8.7）相关的对偶问题为

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \left\{ D(\boldsymbol{\lambda}) \triangleq -F^*(\boldsymbol{\lambda}) - G^*(-\boldsymbol{X}^\top \boldsymbol{\lambda}) \right\}. \tag{8.8}$$

其中 F^* 和 G^* 分别表示 F 和 G的凸共轭函数（定义 2.63）。

根据 Fenchel 对偶定理（推论 8.1），在 F 和 G满足适当正则性条件时，原始-对偶对（8.7）和（8.8）具有强对偶性。本书所关注的特殊情形总是满足这些条件。此外，如式（8.4）所示，同时求解原始问题和对偶问题等价于求拉格朗日函数的一个鞍点：

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\lambda} \rangle + G(\boldsymbol{\beta}) - F^*(\boldsymbol{\lambda}). \tag{8.9}$$

下面介绍的算法依赖于 F^* 和 G （定义2.90）的近端映射。为方便起见，我们引入步长参数 $\gamma > 0$ 和 $\phi > 0$，并定义

$$\mathcal{P}\mathrm{rox}_G(\gamma; \boldsymbol{\alpha}) \triangleq \mathcal{P}\mathrm{rox}_{\gamma G}(\boldsymbol{\alpha}) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \gamma G(\boldsymbol{\beta}) + \frac{1}{2} \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 \right\}, \quad \boldsymbol{\alpha} \in \mathbb{R}^p, \tag{8.10}$$

类似地，

$$\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\alpha}) \triangleq \mathcal{P}\mathrm{rox}_{\phi F^*}(\boldsymbol{\alpha}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^n} \left\{ \phi F^*(\boldsymbol{\beta}) + \frac{1}{2}\|\boldsymbol{\beta}-\boldsymbol{\alpha}\|_2^2 \right\}, \quad \boldsymbol{\alpha}\in\mathbb{R}^n. \tag{8.11}$$

我们假设 $\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\alpha})$ 和 $\mathcal{P}\mathrm{rox}_{G}(\gamma;\boldsymbol{\alpha})$ 均可高效计算——无论是通过闭式表达式，还是通过简单的线性时间算法。注意，根据莫罗分解（定理 2.106 或定理 2.107），只要已知 F^* 的近端算子，就能轻松计算 F 的近端算子。尽管对于任意凸 F 和 G，该算法都有明确定义，但其实用效率取决于能否反复且低成本地计算这些近端映射。因此，以下算法仅适用于近端映射计算开销较低的情形。

Alg算法 1 原始-对偶方法：一般情形（Chambolle 和 Pock，2011年）

Require: Functions $F(\boldsymbol{X}\boldsymbol{\beta})$ and $G(\boldsymbol{\beta})$;
input: $\xi\in[0,1]$, $\gamma,\phi>0$ such that $\gamma\phi\|\boldsymbol{X}\|_2<1$;
initialize: $\boldsymbol{\beta}^{(0)}\in\mathbb{R}^p$, $\boldsymbol{\lambda}^{(0)}\in\mathbb{R}^n$, $\widetilde{\boldsymbol{\beta}}^{(0)}=\boldsymbol{\beta}^{(0)}$;
for $t=0,1,2,\ldots$ **do**
  $\boldsymbol{\lambda}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}^{(t)}+\phi\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)})$; ▷ ($\mathrm{PDM}_1$), dual proximal step
  $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{G}(\gamma;\boldsymbol{\beta}^{(t)}-\gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)})$; ▷ ($\mathrm{PDM}_2$), primal proximal step
  $\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)}+\xi(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})$; ▷ ($\mathrm{PDM}_3$), extrapolation
  Exit if stopping criterion is met at $t=\bar{t}$;
end for
return: Approximation $\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}^{(\bar{t})}$ to the solution of primal problem (8.7);
return: Approximation $\widehat{\boldsymbol{\lambda}}=\boldsymbol{\lambda}^{(\bar{t})}$ to the solution of dual problem (8.8).

算法 1概述了用于求解（8.7）的原始-对偶方法。[2] 参数 ξ 控制施加于原始迭代点 $\boldsymbol{\beta}$的外推量。具体而言：

- 当 $\xi=1$时，我们有 $\widetilde{\boldsymbol{\beta}}^{(t+1)}=2\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}$（完全外推）。当 $\xi=0$时，
- $\widetilde{\boldsymbol{\beta}}^{(t+1)}=\boldsymbol{\beta}^{(t+1)}$（无外推）。该方法便简化为经典的 Arrow–Hurwicz 算法，该算法最初于 1958 年提出（Arrow 等人，1958年）。

步长 γ 和 ϕ 的选择会显著影响收敛；详情请参见定理 8.7 。本文分析的是算法 1 ，其中 $\xi=1$。如果 F^* 或 G 是一致凸的，则可以在迭代过程中自适应地调整参数 ξ、γ和 ϕ，从而进一步加速；参见 Chambolle 和 Pock（2011年）。

▶ **变体**。 通过交换原问题与对偶更新的顺序（即 $\boldsymbol{\lambda}^{(t+1)}$ 和 $\boldsymbol{\beta}^{(t+1)}$），并引入辅助对偶变量 $\widetilde{\boldsymbol{\lambda}}^{(t)}$，可得到该算法的一个对称变体：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{G}(\gamma;\boldsymbol{\beta}^{(t)}-\gamma\boldsymbol{X}^\top\widetilde{\boldsymbol{\lambda}}^{(t)}); \tag{8.12a}$$
$$\boldsymbol{\lambda}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}^{(t)}+\phi\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}); \tag{8.12b}$$
$$\widetilde{\boldsymbol{\lambda}}^{(t+1)} \leftarrow \boldsymbol{\lambda}^{(t+1)}+\xi(\boldsymbol{\lambda}^{(t+1)}-\boldsymbol{\lambda}^{(t)}). \tag{8.12c}$$

2. 直观来说，该算法交替执行对偶变量 $\boldsymbol{\lambda}$ 的近端梯度上升步和原问题变量 $\boldsymbol{\beta}$ 的近端梯度下降步（参见第 9.2.1 节）。

▶ **停止准则**。 一种自然的停止准则基于原始-对偶间隙（见（2.54）），在此情形下其形式为

$$H(\boldsymbol{\beta},\boldsymbol{\lambda}) = F(\boldsymbol{X}\boldsymbol{\beta}) + G(\boldsymbol{\beta}) + F^*(\boldsymbol{\lambda}) + G^*(-\boldsymbol{X}^\top\boldsymbol{\lambda}) \geq 0.$$

$\widehat{\ }\ \widehat{\ }$ $\widehat{\ }\ \widehat{\ }$在达到最优性时（$\boldsymbol{\beta},\boldsymbol{\lambda}$），有 $H(\boldsymbol{\beta},\boldsymbol{\lambda}) = 0$。因此，当 $H(\boldsymbol{\beta}^{(t)}, \boldsymbol{\lambda}^{(t)}) \leq \varepsilon$ 达到预设容差 $\varepsilon > 0$时，可以在第 t 次迭代终止算法。

然而，在下一节介绍的四个关键问题中，函数 F 可能取值为 $+\infty$，这可能导致原始-对偶间隙 H 在中间迭代过程中变为无穷大，从而无法反映当前迭代点的质量。在这种情况下，可以修改算法（例如，将迭代点限制在 F的有效定义域内），以避免出现无穷值。

8.3. 原始-对偶框架下的不同问题

在继续分析算法的收敛性之前，我们将说明原始-对偶框架如何应用于若干稀疏优化问题。

8.3.1 拉格朗日 LASSO

拉格朗日 LASSO（也称为 ℓ_1-正则化最小二乘）问题（L_L）（第126页）定义如下：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 + \frac{\eta}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2, \tag{8.13}$$

其中正则化参数为 $\eta > 0$。经过重新参数化 $\lambda = \eta^{-1}$后，该形式与标准拉格朗日 LASSO（L_L）等价。它符合一般形式（8.7），其中 $G(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ 且 $F(\boldsymbol{X}\boldsymbol{\beta}) = \frac{\eta}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$。显然，$F$不仅是凸函数，而且是光滑函数。$F$ 的凸共轭（见例 2.69）为：

$$F^*(\boldsymbol{\lambda}) = \langle \boldsymbol{y}, \boldsymbol{\lambda}\rangle + \frac{1}{2\eta}\|\boldsymbol{\lambda}\|_2^2.$$

凸共轭 的凸共轭 G（见习题 8.5）是单位 ℓ_∞-球的示性函数 ll:

$$G^*(\boldsymbol{\nu}) = \delta_{\mathbb{B}_\infty[1]}(\boldsymbol{\nu}) = \begin{cases} 0, & \text{if } \|\boldsymbol{\nu}\|_\infty \leq 1; \\ \infty, & \text{otherwise,} \end{cases} \tag{8.14}$$

其中 $\mathbb{B}_\infty[1]$表示 ℓ_∞-范数下的闭单位球（定义 1.29）。因此，与（8.13）相关的对偶问题为：

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} -\langle \boldsymbol{y}, \boldsymbol{\lambda}\rangle - \frac{1}{2\eta}\|\boldsymbol{\lambda}\|_2^2 \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \leq 1,$$

相应的鞍点问题变为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}, \boldsymbol{\lambda}\rangle - \frac{1}{2\eta}\|\boldsymbol{\lambda}\|_2^2 + \|\boldsymbol{\beta}\|_1.$$

通过直接计算可得 F 的近端算子：

$$\mathcal{P}\mathrm{rox}_F(\phi;\boldsymbol{\lambda}) = \frac{\phi\eta}{\phi\eta+1}\boldsymbol{y} + \frac{1}{\phi\eta+1}\boldsymbol{\lambda}. \tag{8.15}$$

根据扩展 Moreau 分解（定理 2.107），F^* 的近端映射为

$$\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}) = \left(1-\frac{\eta}{\eta+\phi}\right)\boldsymbol{\lambda} - \frac{\phi^2}{\eta+\phi}\boldsymbol{y}.$$

$\gamma G(\boldsymbol{\beta}) = \gamma\|\boldsymbol{\beta}\|_1$ 的近端映射是众所周知的软阈值算子（参见例 2.104）：

$$\mathcal{P}\mathrm{rox}_G(\gamma;\boldsymbol{z}) = \mathcal{T}_\gamma(\boldsymbol{z}) = \{\mathcal{T}_\gamma(z_i)\}_{i=1}^p = [|\boldsymbol{z}| - \gamma\mathbf{1}]_+ \circ \mathrm{sign}(\boldsymbol{z}),$$

其中，$\circ$ 表示逐元素乘法，$[\cdot]_+$ 逐分量取正部。

结合这些组成部分，拉格朗日 LASSO（8.13）的原始-对偶算法形式如下：

$$\boldsymbol{\lambda}^{(t+1)} \leftarrow \left(1-\frac{\eta}{\eta+\phi}\right)\boldsymbol{\lambda}^{(t)} + \frac{\phi^2}{\eta+\phi}(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} - \boldsymbol{y}); \tag{8.16a}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{T}_\gamma(\boldsymbol{\beta}^{(t)} - \gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)}); \tag{8.16b}$$

$$\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)} + \xi(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}). \tag{8.16c}$$

8.3.2 约束 LASSO

该 co 约束 LASSO（或 ℓ_1约束最小二乘）问题（L_C）（第 126 页）i s

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_1 \le \Sigma. \tag{8.17}$$

通过设置以下各项，可将其写成形式（8.7）：$G(\boldsymbol{\beta}) = \delta_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta})$，以及 $F(\boldsymbol{X}\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2$，其中 $\mathbb{B}_1[\Sigma]$ 表示半径为 ℓ_1的球。函数 Σ是闭凸函数，因为集合 G 是闭且凸的；参见习题 $\mathbb{B}_1[\Sigma]$ 2.36 和 2.53。与之前一样，在此情形下仍然连续。F 的凸共轭为 F：

$$F^*(\boldsymbol{\lambda}) = \langle \boldsymbol{y},\boldsymbol{\lambda}\rangle + \frac{1}{2}\|\boldsymbol{\lambda}\|_2^2.$$

对于 G，其凸共轭为

$$G^*(\boldsymbol{z}) = \max_{\boldsymbol{\beta}\in\mathbb{R}^p} \langle \boldsymbol{z},\boldsymbol{\beta}\rangle - G(\boldsymbol{\beta}) = \max_{\boldsymbol{\beta}\in\mathbb{B}_1[\Sigma]} \langle \boldsymbol{z},\boldsymbol{\beta}\rangle = \|\boldsymbol{\beta}\|_1\|\boldsymbol{z}\|_\infty,$$

其中，最后一个等号由霍尔德不等式（定理 1.60）得出。因此，对偶问题为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} -\langle \boldsymbol{y},\boldsymbol{\lambda}\rangle - \frac{1}{2}\|\boldsymbol{\lambda}\|_2^2 - \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty.$$

相应的鞍点问题为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y},\boldsymbol{\lambda}\rangle - \frac{1}{2}\|\boldsymbol{\lambda}\|_2^2 + \delta_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta}).$$

由（8.15）和 $\eta=1$，可得

$$\mathcal{P}\mathrm{rox}_F(\phi;\boldsymbol{\lambda}) = \frac{\phi}{\phi+1}\boldsymbol{y} + \frac{1}{\phi+1}\boldsymbol{\lambda}.$$

根据扩展 Moreau 分解（定理 2.107）， F^* 的近端映射为

$$\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}) = \left(1-\frac{1}{1+\phi}\right)\boldsymbol{\lambda} - \frac{\phi^2}{1+\phi}\boldsymbol{y}.$$

近端映射 G 等价于投影到 ℓ_1-球 $\mathbb{B}_1[\Sigma]$(见定义 2.90)，这将在第9.2 节中介绍，用于投影梯度下降方法。

综合以上内容，约束 LASSO 的原始-对偶方法（8.17）如下：

$$\boldsymbol{\lambda}^{(t+1)} \leftarrow \left(1-\frac{1}{1+\phi}\right)\boldsymbol{\lambda}^{(t)} + \frac{\phi^2}{1+\phi}(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} - \boldsymbol{y}); \tag{8.18a}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta}^{(t)} - \gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)}); \tag{8.18b}$$

$$\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)} + \xi(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}). \tag{8.18c}$$

8.3.3 ℓ_1-最小化

该 ℓ_1-最小化问题（P_1）（第 129页）为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{8.19}$$

这个问题符合一般的原始-对偶形式（8.7），其中 $G(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ 且

$$F(\boldsymbol{X}\boldsymbol{\beta}) = \delta_{\{\boldsymbol{y}\}}(\boldsymbol{X}\boldsymbol{\beta}) = \begin{cases} 0, & \text{if } \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}; \\ \infty, & \text{if } \boldsymbol{X}\boldsymbol{\beta} \neq \boldsymbol{y}, \end{cases}$$

也就是说， F 是单点集 $\{\boldsymbol{y}\}$ 的示性函数（参见习题2.36 中的定义）。注意， F 显然是闭的（习题2.53）。 G的凸共轭由式（8.14）给出，而 F 的凸共轭为

$$F^*(\boldsymbol{\lambda}) = \langle\boldsymbol{\lambda}, \boldsymbol{y}\rangle.$$

由于目标函数取值为 $-\infty$ 的项在最大化时可以忽略，因此我们可以显式地施加由 G^*的定义域所产生的约束。于是，对偶问题（8.8）变为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} -\langle\boldsymbol{y}, \boldsymbol{\lambda}\rangle \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \leq 1.$$

相应的鞍点问题为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\lambda}\rangle + \|\boldsymbol{\beta}\|_1 - \langle\boldsymbol{\lambda}, \boldsymbol{y}\rangle. \tag{8.20}$$

根据扩展 Moreau 分解（定理2.107）， F^*的近端映射因此为

$$\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}) = \boldsymbol{\lambda} - \phi\mathcal{P}\mathrm{rox}_{\phi^{-1}F}(\boldsymbol{\lambda}/\phi) = \boldsymbol{\lambda} - \phi\boldsymbol{y},$$

其中， F 的近端映射是到 $\{\boldsymbol{y}\}$上的投影，也就是说，对所有 $\boldsymbol{\lambda}\in\mathbb{R}^n$，它都是常值映射$\mathcal{P}\mathrm{rox}_F(\phi;\boldsymbol{\lambda}) = \boldsymbol{y}$。

综合这些组成部分， ℓ_1-最小化问题（8.19）的原始-对偶算法如下：

$$\boldsymbol{\lambda}^{(t+1)} \leftarrow \boldsymbol{\lambda}^{(t)} + \phi(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} - \boldsymbol{y}); \tag{8.21a}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{T}_\gamma(\boldsymbol{\beta}^{(t)} - \gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)}); \tag{8.21b}$$

$$\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)} + \xi(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}). \tag{8.21c}$$

8.3.4 ℓ_1-最小化与含噪测量值

二次约束 ℓ_1-最小化问题（$\mathrm{P}_{1,\ \epsilon}$）（第131页）为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2 \le \epsilon, \tag{8.22}$$

其中 $\epsilon>0$ 表示噪声水平。它可以表示为以下形式：8.7），其中 $G(\boldsymbol{\beta}) =\|\boldsymbol{\beta}\|_1$ 并且

$$F(\boldsymbol{X}\boldsymbol{\beta}) = \delta_{\mathbb{B}[\boldsymbol{y},\epsilon]}(\boldsymbol{X}\boldsymbol{\beta}) = \begin{cases} 0, & \text{if } \|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2 \le \epsilon; \\ \infty, & \text{otherwise.} \end{cases}$$

函数 F 是闭的，因为集合 $\mathbb{B}[\boldsymbol{y},\epsilon]$ 是闭的（习题 2.53）。其凸共轭（参见问题 8.4）是球的支撑函数：

$$F^*(\boldsymbol{\lambda}) = \max_{\boldsymbol{z}:\,\|\boldsymbol{z}-\boldsymbol{y}\|_2\le\epsilon} \langle \boldsymbol{z},\boldsymbol{\lambda}\rangle = \langle \boldsymbol{y},\boldsymbol{\lambda}\rangle + \epsilon\|\boldsymbol{\lambda}\|_2 .$$

的凸共轭 G 仍如（8.14）所示。因此，对偶问题变为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} -\langle \boldsymbol{y},\boldsymbol{\lambda}\rangle - \epsilon\|\boldsymbol{\lambda}\|_2 \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{\lambda}\right\|_\infty \le 1,$$

相应的鞍点表述为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \langle \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y},\boldsymbol{\lambda}\rangle - \epsilon\|\boldsymbol{\lambda}\|_2 + \|\boldsymbol{\beta}\|_1 .$$

例2.97 表明， F 的近端映射是到球 $\mathbb{B}[\boldsymbol{y},\epsilon]$的正交投影，

$$\begin{aligned}\mathcal{P}\mathrm{rox}_F(\phi;\boldsymbol{\lambda}) &= \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathbb{R}^n:\,\|\boldsymbol{z}-\boldsymbol{y}\|_2\le\epsilon} \|\boldsymbol{z}-\boldsymbol{\lambda}\|_2 = \boldsymbol{y} + \frac{\epsilon}{\max\{\|\boldsymbol{\lambda}-\boldsymbol{y}\|_2,\epsilon\}}(\boldsymbol{\lambda}-\boldsymbol{y}) \\ &= \begin{cases} \boldsymbol{\lambda} & \text{if } \|\boldsymbol{\lambda}-\boldsymbol{y}\|_2 \le \epsilon; \\ \boldsymbol{y} + \dfrac{\epsilon}{\|\boldsymbol{\lambda}-\boldsymbol{y}\|_2}(\boldsymbol{\lambda}-\boldsymbol{y}) & \text{otherwise.} \end{cases}\end{aligned}$$

再次应用扩展 Moreau 分解，可得 F^* 的近端映射：

$$\mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}) = \begin{cases} \boldsymbol{0}, & \text{if } \|\boldsymbol{\lambda}-\phi\boldsymbol{y}\|_2 \le \epsilon\phi; \\ \left(1-\dfrac{\epsilon\phi}{\|\boldsymbol{\lambda}-\phi\boldsymbol{y}\|_2}\right)(\boldsymbol{\lambda}-\phi\boldsymbol{y}), & \text{otherwise.} \end{cases}$$

最后，二次约束 ℓ_1-最小化问题（8.22）的原始-对偶算法为：

$$\boldsymbol{\lambda}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{F^*}(\phi;\boldsymbol{\lambda}^{(t)} + \phi\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)}); \tag{8.23a}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{T}_\gamma(\boldsymbol{\beta}^{(t)} - \gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)}); \tag{8.23b}$$

$$\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)} + \xi(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}). \tag{8.23c}$$

8.4. 原始-对偶算法的最优性条件

该算法可以解释为不动点迭代：

定理 8.2：（原始-对偶方法的最优性条件）点（$\widehat{\boldsymbol{\beta}}$，$\widehat{\boldsymbol{\lambda}}$）是迭代（$\mathrm{PDM}_1$）、（$\mathrm{PDM}_2$）、（$\mathrm{PDM}_3$）（对于 ξ的任意选择）当且仅当$\widehat{\boldsymbol{\beta}}$ $\widehat{\boldsymbol{\lambda}}$（$\widehat{\boldsymbol{\beta}}$，$\widehat{\boldsymbol{\lambda}}$）是（8.9）的鞍点，即问题（8.7）和（8.8）的原始-对偶最优点。

证明 [定理8.2] 由算法1可知，不动点必须满足

$$\widehat{\boldsymbol{\lambda}} = \mathcal{P}\mathrm{rox}_{F^*}(\phi; \widehat{\boldsymbol{\lambda}} + \phi \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \qquad \text{and} \qquad \widehat{\boldsymbol{\beta}} = \mathcal{P}\mathrm{rox}_{G}(\gamma; \widehat{\boldsymbol{\beta}} - \gamma \boldsymbol{X}^\top \widehat{\boldsymbol{\lambda}}).$$

由于 F^*和 G是真闭凸函数（F^*根据引理2.64和习题2.67，是正常闭凸函数），由近端性质 I（引理2.100）$\widehat{\boldsymbol{\beta}}$ $\widehat{\boldsymbol{\lambda}}$可知，不动点（$\widehat{\boldsymbol{\beta}}$，$\widehat{\boldsymbol{\lambda}}$）满足

$$\phi \boldsymbol{X}\widehat{\boldsymbol{\beta}} \in \phi \partial F^*(\widehat{\boldsymbol{\lambda}}) \qquad \text{and} \qquad -\gamma \boldsymbol{X}^\top \widehat{\boldsymbol{\lambda}} \in \gamma \partial G(\widehat{\boldsymbol{\beta}}),$$

其中 ∂F^*和 ∂G分别表示 F^*和 G的次微分（定义2.57）。将两个包含式分别除以正标量 ϕ和 γ，可得到等价条件

$$\mathbf{0} \in -\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \partial F^*(\widehat{\boldsymbol{\lambda}}) \qquad \text{and} \qquad \mathbf{0} \in \boldsymbol{X}^\top \widehat{\boldsymbol{\lambda}} + \partial G(\widehat{\boldsymbol{\beta}}).$$

根据次微分最优性条件（定理2.62），这些关系式等价于：$\widehat{\boldsymbol{\beta}}$是函数 $\boldsymbol{\beta} \mapsto \langle \boldsymbol{\beta}, \boldsymbol{X}^\top \widehat{\boldsymbol{\lambda}} \rangle + G(\boldsymbol{\beta}) - F^*(\widehat{\boldsymbol{\lambda}})$的最小值，且 $\widehat{\boldsymbol{\lambda}}$是函数 $\boldsymbol{\lambda} \mapsto \langle \boldsymbol{X}\widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda} \rangle + G(\widehat{\boldsymbol{\beta}}) - F^*(\boldsymbol{\lambda})$的最大值。综上，这意味着$\widehat{\boldsymbol{\beta}}$ $\widehat{\boldsymbol{\lambda}}$(β̂, λ̂)满足（8.9）中拉格朗日函数的鞍点条件。因此，它是该问题的鞍点。

反过来，8.9) 的任一鞍点都满足上述次微分包含关系；根据同样的近端刻画，这意味着它是原始-对偶迭代的不动点。证明完毕。

■

原始-对偶算法通常通过拆分原问题与对偶问题的更新，以启发式方式推导出来。该定理验证了这种设计：针对 G 和 F^* 的特定近端步骤并非随意设定，而是精确地编码了鞍点所需的次微分最优性条件。

8.5. 原始-对偶算法的收敛性

现在讨论原始-对偶方法的收敛性。回顾一下，对于鞍点问题（8.9），我们定义约化的拉格朗日函数

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \triangleq \langle \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\lambda} \rangle + G(\boldsymbol{\beta}) - F^*(\boldsymbol{\lambda}),$$

注意，这个拉格朗日函数引出了原始-对偶间隙，其定义如下：

$$H(\boldsymbol{\beta}, \boldsymbol{\lambda}) \triangleq \max_{\boldsymbol{\lambda}' \in \mathbb{R}^n} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}') - \min_{\boldsymbol{\beta}' \in \mathbb{R}^p} \mathcal{L}(\boldsymbol{\beta}', \boldsymbol{\lambda}). \tag{8.24}$$

由此可得 $H(\boldsymbol{\beta}, \boldsymbol{\lambda}) \geq 0$对所有 $\boldsymbol{\beta}, \boldsymbol{\lambda}$，且 $H(\boldsymbol{\beta}, \boldsymbol{\lambda}) = 0$当且仅当（$\boldsymbol{\beta}, \boldsymbol{\lambda}$）是 $\mathcal{L}$的鞍点；参见（8.3）中的鞍点刻画。

我们首先确立原始-对偶算法的关键性质，随后将用这些性质证明其收敛性。为简化符号，考虑一个序列 $(\boldsymbol{a}^{(t)})_{t \in \mathbb{N}_0}$ (由

标量或向量组成）。我们引入差商算子:

$$\Delta_\theta \boldsymbol{a}^{(t)} \triangleq \frac{\boldsymbol{a}^{(t)} - \boldsymbol{a}^{(t-1)}}{\theta}; \tag{8.25}$$

$$\Delta_\theta \left\|\boldsymbol{a}^{(t)}\right\|_2^2 \triangleq \frac{\left\|\boldsymbol{a}^{(t)}\right\|_2^2 - \left\|\boldsymbol{a}^{(t-1)}\right\|_2^2}{\theta}, \quad t \in \mathbb{N}. \tag{8.26}$$

这个量可以解释为步长为 θ 的离散导数。以下恒等式与熟悉的连续微积分法则非常相似。

引理 8.3 设 $\boldsymbol{a}, \boldsymbol{a}^{(t)} \in \mathbb{R}^p$, $t \in \mathbb{N}_0$。则

$$2\left\langle \Delta_\theta \boldsymbol{a}^{(t)}, \boldsymbol{a}^{(t)} - \boldsymbol{a} \right\rangle = \Delta_\theta \left\|\boldsymbol{a}^{(t)} - \boldsymbol{a}\right\|_2^2 + \theta \left\|\Delta_\theta \boldsymbol{a}^{(t)}\right\|_2^2. \tag{8.27}$$

此外，若 $\boldsymbol{b}^{(t)}$, $t \in \mathbb{N}_0$ 是另一个向量序列，则对任意 $T \in \mathbb{N}$，以下离散分部求和公式成立，

$$\theta \sum_{t=1}^{T} \left(\left\langle \Delta_\theta \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)} \right\rangle + \left\langle \boldsymbol{a}^{(t-1)}, \Delta_\theta \boldsymbol{b}^{(t)} \right\rangle \right) = \left\langle \boldsymbol{a}^{(T)}, \boldsymbol{b}^{(T)} \right\rangle - \left\langle \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)} \right\rangle. \tag{8.28}$$

证明 [引理 8.3] 设 $\widetilde{\boldsymbol{a}}^{(t)} \triangleq \boldsymbol{a}^{(t)} - \boldsymbol{a}$。则 $\Delta_\theta \widetilde{\boldsymbol{a}}^{(t)} = \Delta_\theta \boldsymbol{a}^{(t)}$，且

$$\begin{aligned} 2\theta \left\langle \Delta_\theta \boldsymbol{a}^{(t)}, \boldsymbol{a}^{(t)} - \boldsymbol{a} \right\rangle &= 2\theta \left\langle \Delta_\theta \widetilde{\boldsymbol{a}}^{(t)}, \widetilde{\boldsymbol{a}}^{(t)} \right\rangle = 2\left\langle \widetilde{\boldsymbol{a}}^{(t)} - \widetilde{\boldsymbol{a}}^{(t-1)}, \widetilde{\boldsymbol{a}}^{(t)} \right\rangle \\ &= \left\langle \widetilde{\boldsymbol{a}}^{(t)} - \widetilde{\boldsymbol{a}}^{(t-1)}, \widetilde{\boldsymbol{a}}^{(t)} - \widetilde{\boldsymbol{a}}^{(t-1)} \right\rangle + \left\langle \widetilde{\boldsymbol{a}}^{(t)} - \widetilde{\boldsymbol{a}}^{(t-1)}, \widetilde{\boldsymbol{a}}^{(t)} + \widetilde{\boldsymbol{a}}^{(t-1)} \right\rangle \\ &= \left\|\widetilde{\boldsymbol{a}}^{(t)} - \widetilde{\boldsymbol{a}}^{(t-1)}\right\|_2^2 + \theta \Delta_\theta \left\|\widetilde{\boldsymbol{a}}^{(t)}\right\|_2^2. \end{aligned}$$

这就证明了（8.27）。

对于第二部分，我们观察到

$$\Delta_\theta \left\langle \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)} \right\rangle = \frac{\left\langle \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)} \right\rangle - \left\langle \boldsymbol{a}^{(t-1)}, \boldsymbol{b}^{(t-1)} \right\rangle}{\theta} = \left\langle \Delta_\theta \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)} \right\rangle + \left\langle \boldsymbol{a}^{(t-1)}, \Delta_\theta \boldsymbol{b}^{(t)} \right\rangle.$$

对 $t = 1$、2、.....、T 求和并消去相邻项，即可证明所需结果。 ■

引理 8.4：（原始-对偶算法的迭代点不等式） 设 $\{\boldsymbol{\beta}^{(t)}, \widetilde{\boldsymbol{\beta}}^{(t)}, \boldsymbol{\lambda}^{(t)}\}_{t\geq 0}$为算法1、2、3在算法1中生成的序列，并设 $\boldsymbol{\beta} \in \mathbb{R}^p$、$\boldsymbol{\lambda} \in \mathbb{R}^n$ 为任意值。那么，对于 $t \in \mathbb{N}$,

$$\begin{aligned} &\frac{1}{2}\Delta_\phi \left\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(t)}\right\|_2^2 + \frac{1}{2}\Delta_\gamma \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}\right\|_2^2 + \frac{\phi}{2}\left\|\Delta_\phi \boldsymbol{\lambda}^{(t)}\right\|_2^2 + \frac{\gamma}{2}\left\|\Delta_\gamma \boldsymbol{\beta}^{(t)}\right\|_2^2 \\ &\leq \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}^{(t)}) - \mathcal{L}(\boldsymbol{\beta}^{(t)}, \boldsymbol{\lambda}) + \left\langle \boldsymbol{X}(\boldsymbol{\beta}^{(t)} - \widetilde{\boldsymbol{\beta}}^{(t-1)}), \boldsymbol{\lambda} - \boldsymbol{\lambda}^{(t)} \right\rangle. \end{aligned} \tag{8.29}$$

此外，假设 $\xi = 1$ 满足算法1中的条件。将不等式（8.29）在 $t = 1$，2，……，T上求和并利用望远镜消去，得到

$$\begin{aligned}
&\frac{1}{2\phi}\left(\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\|_2^2-\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(0)}\right\|_2^2\right)+\frac{1}{2\gamma}\left(\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(T)}\right\|_2^2-\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right\|_2^2\right)\\
&+\frac{1}{2\phi}\sum_{t=1}^{T}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2+\frac{1}{2\gamma}\sum_{t=1}^{T}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2\\
&\leq\sum_{t=1}^{T}\left(\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\lambda}^{(t)})-\mathcal{L}(\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda})\right)+\gamma^2\sum_{t=1}^{T}\left\langle\boldsymbol{X}\Delta_\gamma^2\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle,
\end{aligned}\tag{8.30}$$

其中 $\Delta_\gamma^2\boldsymbol{\beta}^{(t)} \triangleq \Delta_\gamma(\Delta_\gamma\boldsymbol{\beta}^{(t)})$.

证明 [引理8.4] 由 (PDM$_1$), (PDM$_2$) 以及近端性质 I（引理 2.100），迭代点满足以下关系式（将 $t+1$ 替换为 t）

$$\boldsymbol{\lambda}^{(t-1)}+\phi\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t-1)}-\boldsymbol{\lambda}^{(t)}\in\phi\partial F^*(\boldsymbol{\lambda}^{(t)}),\qquad\boldsymbol{\beta}^{(t-1)}-\gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t)}-\boldsymbol{\beta}^{(t)}\in\gamma\partial G(\boldsymbol{\beta}^{(t)}),$$

其中 ∂F^*和 ∂G分别表示 F^*和 G的次微分。应用次梯度不等式（定义2.57），得到:

$$\begin{aligned}
\phi F^*(\boldsymbol{\lambda})-\phi F^*(\boldsymbol{\lambda}^{(t)})&\geq\left\langle-\boldsymbol{\lambda}^{(t)}+\boldsymbol{\lambda}^{(t-1)}+\phi\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t-1)},\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle;\\
\gamma G(\boldsymbol{\beta})-\gamma G(\boldsymbol{\beta}^{(t)})&\geq\left\langle-\boldsymbol{\beta}^{(t)}+\boldsymbol{\beta}^{(t-1)}-\gamma\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t)},\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}\right\rangle.
\end{aligned}$$

分别除以 ϕ 和 γ，并根据差商的定义改写，得到:

$$\begin{aligned}
F^*(\boldsymbol{\lambda})-F^*(\boldsymbol{\lambda}^{(t)})&\geq\left\langle\Delta_\phi\boldsymbol{\lambda}^{(t)},\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}\right\rangle+\left\langle\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t-1)},\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle;\\
G(\boldsymbol{\beta})-G(\boldsymbol{\beta}^{(t)})&\geq\left\langle\Delta_\gamma\boldsymbol{\beta}^{(t)},\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}\right\rangle-\left\langle\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}),\boldsymbol{\lambda}^{(t)}\right\rangle.
\end{aligned}$$

将这两个不等式相加，并应用引理 8.3 (8.27)，得到:

$$\begin{aligned}
&\left(\frac{1}{2}\Delta_\phi\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\|_2^2+\frac{\phi}{2}\left\|\Delta_\phi\boldsymbol{\lambda}^{(t)}\right\|_2^2\right)+\left(\frac{1}{2}\Delta_\gamma\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}\right\|_2^2+\frac{\gamma}{2}\left\|\Delta_\gamma\boldsymbol{\beta}^{(t)}\right\|_2^2\right)\\
&\leq F^*(\boldsymbol{\lambda})-F^*(\boldsymbol{\lambda}^{(t)})+G(\boldsymbol{\beta})-G(\boldsymbol{\beta}^{(t)})+\left\langle\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}),\boldsymbol{\lambda}^{(t)}\right\rangle-\left\langle\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t-1)},\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle\\
&=\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\lambda}^{(t)})-\mathcal{L}(\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda})+\left\langle\boldsymbol{X}(\boldsymbol{\beta}^{(t)}-\widetilde{\boldsymbol{\beta}}^{(t-1)}),\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle.
\end{aligned}$$

这就证明了（8.29）。

现在假设 $\xi = 1$。在这种情况下，外推满足 $\widetilde{\boldsymbol{\beta}}^{(t-1)} = 2\beta^{(t-1)} - \boldsymbol{\beta}^{(t-2)}$，因此

$$\boldsymbol{\beta}^{(t)}-\widetilde{\boldsymbol{\beta}}^{(t-1)}=\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}-(\boldsymbol{\beta}^{(t-1)}-\boldsymbol{\beta}^{(t-2)})=\gamma^2\Delta_\gamma(\Delta_\gamma\boldsymbol{\beta}^{(t)})\triangleq\gamma^2\Delta_\gamma^2\boldsymbol{\beta}^{(t)},\quad\text{for } t\geq 2.$$

该公式在 $t = 1$ 的情况下成立，若令 $\boldsymbol{\beta}^{(-1)} = \boldsymbol{\beta}^{(0)}$，则可推出 $\widetilde{\boldsymbol{\beta}}^{(0)} = \boldsymbol{\beta}^{(0)}$ 以及 $\Delta_\gamma\boldsymbol{\beta}^{(0)} = \mathbf{0}$。将不等式（8.29）对 $t = 1$、2、……、T求和并进行裂项，可证明（8.30）。 ■

不等式（8.29）表明，理想的选择应为 $\widetilde{\boldsymbol{\beta}}^{(t-1)} = \boldsymbol{\beta}^{(t)}$，这样可以消去最后一个内积项。然而，这将导致一个隐式方案，其中每次迭代都需要求解一个与原问题同样困难的耦合系统。

利用望远镜式迭代不等式（8.30），现在我们给出累计原始-对偶间隙的一个上界。

引理 8.5：（原始-对偶间隙不等式） 设 $\{\boldsymbol{\beta}^{(t)}, \widetilde{\boldsymbol{\beta}}^{(t)}, \boldsymbol{\lambda}^{(t)}\}_{t\geq 0}$ 为算法$_1$中由 (PDM$_2$)、(PDM$_3$)、(PDM1 在参数选择 $\xi = 1$下生成的序列，并设 $\boldsymbol{\beta} \in \mathbb{R}^p$、 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 为任意值。那么，对于任意 $T \in \mathbb{N}$,

$$\begin{aligned}
&\sum_{t=1}^{T}\Big(\mathcal{L}(\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}) - \mathcal{L}(\boldsymbol{\beta},\boldsymbol{\lambda}^{(t)})\Big) + \frac{1-\gamma\phi\|\boldsymbol{X}\|_2^2}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\|_2^2 + \frac{1}{2\gamma}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(T)}\right\|_2^2 \\
&\quad + \frac{1-\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\phi}\sum_{t=1}^{T}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2 + \frac{1-\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\gamma}\sum_{t=1}^{T-1}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2 \\
&\leq \frac{1}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(0)}\right\|_2^2 + \frac{1}{2\gamma}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right\|_2^2. \qquad (8.31)
\end{aligned}$$

引理 [的证明8.5] 我们从求和后的迭代不等式（引理8.4 (8.30)）开始。利用离散分部求和公式（引理 8.3 (8.28)），结合初始化 $\Delta_\gamma\boldsymbol{\beta}^{(0)} = \mathbf{0}$ （通过初始化 $\boldsymbol{\beta}^{(-1)} = \boldsymbol{\beta}^{(0)}$）以及定义 $\Delta_\gamma^2\boldsymbol{\beta}^{(t)} \triangleq \Delta_\gamma(\Delta_\gamma\boldsymbol{\beta}^{(t)})$，我们将二阶项改写为:

$$\begin{aligned}
\gamma^2\sum_{t=1}^{T}\left\langle \boldsymbol{X}\Delta_\gamma^2\boldsymbol{\beta}^{(t)}, \boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle &= \gamma^2\sum_{t=1}^{T}\left\langle \boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(t-1)}, \Delta_\gamma\boldsymbol{\lambda}^{(t)}\right\rangle + \gamma\left\langle \boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(T)}, \boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\rangle \\
&= \gamma\phi\sum_{t=1}^{T}\left\langle \boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(t-1)}, \Delta_\phi\boldsymbol{\lambda}^{(t)}\right\rangle + \gamma\left\langle \Delta_\gamma\boldsymbol{\beta}^{(T)}, \boldsymbol{X}^\top(\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)})\right\rangle. \qquad (8.32)
\end{aligned}$$

对于右侧的第一项，令 $\eta \triangleq \sqrt{\gamma/\phi}$，我们有

$$\begin{aligned}
\gamma\phi\left\langle \boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(t-1)}, \Delta_\phi\boldsymbol{\lambda}^{(t)}\right\rangle &\leq \gamma\phi\|\boldsymbol{X}\|_2\left\|\Delta_\gamma\boldsymbol{\beta}^{(t-1)}\right\|_2\left\|\Delta_\phi\boldsymbol{\lambda}^{(t)}\right\|_2 \\
&\leq \frac{\gamma\phi\|\boldsymbol{X}\|_2}{2}\left(\eta\left\|\Delta_\gamma\boldsymbol{\beta}^{(t-1)}\right\|_2^2 + \eta^{-1}\left\|\Delta_\phi\boldsymbol{\lambda}^{(t)}\right\|_2^2\right) \\
&= \frac{\phi\eta\|\boldsymbol{X}\|_2}{2\gamma}\left\|\boldsymbol{\beta}^{(t-1)}-\boldsymbol{\beta}^{(t-2)}\right\|_2^2 + \frac{\gamma\|\boldsymbol{X}\|_2}{2\eta\phi}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2 \\
&\overset{\eta=\sqrt{\gamma/\phi}}{=\!=} \frac{\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\gamma}\left\|\boldsymbol{\beta}^{(t-1)}-\boldsymbol{\beta}^{(t-2)}\right\|_2^2 + \frac{\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\phi}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2, \qquad (8.33)
\end{aligned}$$

其中，第一个不等式由柯西–施瓦茨不等式（定理1.59）得出，第二个不等式则由以下事实得出： $2ab \leq \eta a^2 + b^2/\eta$对于正的 a， b， η。类似地，第二项满足

$$\begin{aligned}
\gamma\left\langle \Delta_\gamma\boldsymbol{\beta}^{(T)}, \boldsymbol{X}^\top(\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)})\right\rangle &\leq \frac{\gamma}{2}\left(\left\|\Delta_\gamma\boldsymbol{\beta}^{(T)}\right\|_2^2 + \|\boldsymbol{X}\|_2^2\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\|_2^2\right) \\
&= \frac{1}{2\gamma}\left\|\boldsymbol{\beta}^{(T)}-\boldsymbol{\beta}^{(T-1)}\right\|_2^2 + \frac{\gamma\phi\|\boldsymbol{X}\|_2^2}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\|_2^2.
\end{aligned}$$

将这些估计值代入（8.32），并利用 $\boldsymbol{\beta}^{(-1)}=\boldsymbol{\beta}^{(0)}$，我们得到：

$$\gamma^2\sum_{t=1}^{T}\left\langle \boldsymbol{X}\Delta_\gamma^2\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(t)}\right\rangle \le \frac{\sqrt{\gamma\phi}\,\|\boldsymbol{X}\|_2}{2\gamma}\sum_{t=1}^{T-1}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2 + \frac{\sqrt{\gamma\phi}\,\|\boldsymbol{X}\|_2}{2\phi}\sum_{t=1}^{T}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2+\frac{1}{2\gamma}\left\|\boldsymbol{\beta}^{(T)}-\boldsymbol{\beta}^{(T-1)}\right\|_2^2+\frac{\gamma\phi\,\|\boldsymbol{X}\|_2^2}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(T)}\right\|_2^2.$$

将此上界代回求和后的迭代点不等式（引理8.4(8.30）中，并重新整理各项，可得所需的不等式（8.31）。 ■

利用上述结果，我们现在为原始-对偶算法建立一个下降引理。

引理 8.6：（原始-对偶算法的下降引理） 设 $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}})$ 是一个原始-对偶最优解，即鞍点问题（8.9）的一个鞍点。设 $\{\boldsymbol{\beta}^{(t)},\widetilde{\boldsymbol{\beta}}^{(t)},\boldsymbol{\lambda}^{(t)}\}_{t\ge0}$为由算法 1 在参数选择 $\xi=1$ 和 $\gamma\phi\,\|\boldsymbol{X}\|_2^2<1$下，通过 (PDM$_1$)、(PDM$_2$)、(PDM$_3$) 生成的序列。那么，对任意 $T\in\mathbb{N}$,

$$\frac{1}{2\phi}\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^{(T)}\right\|_2^2+\frac{1}{2\gamma}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(T)}\right\|_2^2\le C\left(\frac{1}{2\phi}\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^{(0)}\right\|_2^2+\frac{1}{2\gamma}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(0)}\right\|_2^2\right),$$

其中 $C\triangleq(1-\gamma\phi\,\|\boldsymbol{X}\|_2^2)^{-1}$。特别地，迭代点 $\{\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}^{(t)}\}$ 保持有界。

证明 [引理 8.6] 对于一个鞍点 $(\widehat{\boldsymbol{\beta}},\ \widehat{\boldsymbol{\lambda}})$，各项 $\mathcal{L}(\widehat{\boldsymbol{\beta}},\ \widehat{\boldsymbol{\lambda}})-\mathcal{L}(\widehat{\boldsymbol{\beta}},\ \boldsymbol{\lambda}^{(t)})$ 均非负，因此 (8.31) 左侧的所有项均为正。因此

$$\frac{1-\gamma\phi\,\|\boldsymbol{X}\|_2^2}{2\phi}\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^{(T)}\right\|_2^2+\frac{1}{2\gamma}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(T)}\right\|_2^2\le\frac{1}{2\phi}\left\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^{(0)}\right\|_2^2+\frac{1}{2\gamma}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(0)}\right\|_2^2.$$

证明至此完成。 ■

收　　敛性随后将在下述定理中得到证明。

定理 8.7：（原始-对偶算法的收敛性） 设 $\{\boldsymbol{\beta}^{(t)},\widetilde{\boldsymbol{\beta}}^{(t)},\boldsymbol{\lambda}^{(t)}\}_{t\ge0}$ 为算法$_1$生成的序列（PDM$_2$）、（PDM$_3$），其参数取为 $\xi=1$ 和$\gamma\phi\,\|\boldsymbol{X}\|_2^2<1$。假设鞍点问题（8.9）存在一个鞍点。那么，迭代点 $\{(\boldsymbol{\beta}^{(t)},\ \boldsymbol{\lambda}^{(t)})\}$ 收敛到 $(\widehat{\boldsymbol{\beta}},\ \widehat{\boldsymbol{\lambda}})$ 这个（8.9）的鞍点，使得$\boldsymbol{\beta}^{(t)}$ 收敛到问题（8.7）的一个极小化解。

定理 8.7] 的证明 [由引理 8.6 可知，序列 $\{(\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}^{(t)})\}$ 有界。因此，存在一个收敛子列 $\{(\boldsymbol{\beta}^{(t_k)}\ \boldsymbol{\lambda}^{(t_k)})\}_k$，使得 $(\boldsymbol{\beta}^{(t_k)}\ \boldsymbol{\lambda}^{(t_k)})\to(\boldsymbol{\beta}^\star,\boldsymbol{\lambda}^\star)$ 当 $\frown\frown k\to\infty$。在 (8.31) 中选取 $(\boldsymbol{\beta},\boldsymbol{\lambda})$ 作为鞍点 $(\boldsymbol{\beta},\boldsymbol{\lambda})$，则所有项均非负。由于左侧所有项均非负，我们特别得到：

$$\frac{1-\sqrt{\gamma\phi}\,\|\boldsymbol{X}\|_2}{2\gamma}\sum_{t=1}^{T-1}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2\le\frac{1}{2\gamma}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right\|_2^2+\frac{1}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(0)}\right\|_2^2.$$

由于右侧与 T 和 $\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2 < 1$ 无关，因此我们得出级数 $\sum_{t=1}^{\infty}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2$ 收敛；于是，$\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2 \to 0$ 当 $t\to\infty$ 时。类似的论证表明 $\lim_{t\to\infty}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2 = 0$。特别地，子序列 $(\boldsymbol{\beta}^{(t_k-1)}\ \boldsymbol{\lambda}^{(t_k-1)})$ 也收敛到 $(\boldsymbol{\beta}^\star,\ \boldsymbol{\lambda}^\star)$。因此，$(\boldsymbol{\beta}^\star,\ \boldsymbol{\lambda}^\star)$ 是原始-对偶算法的不动点。根据定理 8.2，它是原始-对偶最优的（一个鞍点）。

代入 $(\boldsymbol{\beta},\ \boldsymbol{\lambda}) = (\boldsymbol{\beta}^\star,\ \boldsymbol{\lambda}^\star)$ 到（8.29）中，我们由原始-对偶间隙性质得到 $\mathcal{L}(\boldsymbol{\beta}^{(t)},\ \boldsymbol{\lambda}^\star) - \mathcal{L}(\boldsymbol{\beta}^\star, \boldsymbol{\lambda}^{(t)}) \geq 0$ 类似于（8.30）的证明，将（8.29）从 $t=t_k$ 到 $t=T>t_k$ 求和并进行望远镜求和，得到

$$
\begin{aligned}
&\frac{1}{2\phi}\left(\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(T)}\right\|_2^2-\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t_k)}\right\|_2^2\right)+\frac{1}{2\gamma}\left(\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(T)}\right\|_2^2-\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(t_k)}\right\|_2^2\right)\\
&+\frac{1}{2\phi}\sum_{t=t_k}^{T}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2+\frac{1}{2\gamma}\sum_{t=t_k}^{T}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2\leq\gamma^2\sum_{t=t_k}^{T}\left\langle\boldsymbol{X}\Delta_\gamma^2\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t)}\right\rangle\\
&=\gamma\phi\sum_{t=t_k}^{T}\left\langle\boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(t-1)},\Delta_\phi\boldsymbol{\lambda}^{(t)}\right\rangle+\gamma\left\langle\boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(T)},\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(T)}\right\rangle-\gamma\left\langle\boldsymbol{X}\Delta_\gamma\boldsymbol{\beta}^{(t_k)},\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t_k)}\right\rangle,
\end{aligned}
$$

其中，最后一个等式由离散分部求和引理8.3（8.28）得出。因此，不等式（8.33）蕴含

$$
\begin{aligned}
&\frac{1}{2\phi}\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(T)}\right\|_2^2+\frac{1}{2\gamma}\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(T)}\right\|_2^2+\frac{1-\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\phi}\sum_{t=t_k}^{T}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2^2\\
&+\frac{1-\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2}{2\gamma}\sum_{t=t_k}^{T-1}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2^2\\
&+\frac{1}{2\gamma}\left(\left\|\boldsymbol{\beta}^{(T)}-\boldsymbol{\beta}^{(T-1)}\right\|_2^2-\sqrt{\gamma\phi}\|\boldsymbol{X}\|_2\left\|\boldsymbol{\beta}^{(t_k-1)}-\boldsymbol{\beta}^{(t_k-2)}\right\|_2^2\right)\\
&+\left\langle\boldsymbol{X}(\boldsymbol{\beta}^{(T)}-\boldsymbol{\beta}^{(T-1)}),\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(T)}\right\rangle-\left\langle\boldsymbol{X}(\boldsymbol{\beta}^{(t_k)}-\boldsymbol{\beta}^{(t_k-1)}),\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t_k)}\right\rangle\\
&\leq\frac{1}{2\phi}\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t_k)}\right\|_2^2+\frac{1}{2\gamma}\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(t_k)}\right\|_2^2.
\end{aligned}
$$

由于 $\lim_{t\to\infty}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|_2=\lim_{t\to\infty}\left\|\boldsymbol{\lambda}^{(t)}-\boldsymbol{\lambda}^{(t-1)}\right\|_2=0$ 且 $\lim_{k\to\infty}\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(t_k)}\right\|_2=\lim_{k\to\infty}\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(t_k)}\right\|_2=0$，可得 $\lim_{T\to\infty}\left\|\boldsymbol{\beta}^\star-\boldsymbol{\beta}^{(T)}\right\|_2=\lim_{T\to\infty}\left\|\boldsymbol{\lambda}^\star-\boldsymbol{\lambda}^{(T)}\right\|_2=0$。证明完毕。 ■

8.6. 原始-对偶算法的收敛速度

为了分析原始-对偶算法的收敛速度，我们引入部分原始-对偶间隙。对于两个集合 $\mathbb{S}_1\subset\mathbb{R}^p$ 和 $\mathbb{S}_2\subset\mathbb{R}^n$，该量定义为

$$
\mathcal{P}_{\mathbb{S}_1,\mathbb{S}_2}(\boldsymbol{\beta},\boldsymbol{\lambda})\triangleq\max_{\boldsymbol{\lambda}'\in\mathbb{S}_2}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\lambda}')-\min_{\boldsymbol{\beta}'\in\mathbb{S}_1}\mathcal{L}(\boldsymbol{\beta}',\boldsymbol{\lambda}). \tag{8.34}
$$

这是对式 (2.54) 中定义的标准原始-对偶间隙的一种变体。在我们的情形中，它尤为方便，其自然动机来自鞍点问题 (8.9)。只要乘积集 $\mathbb{S}_1\times\mathbb{S}_2$ 包含一个鞍点 $(\boldsymbol{\beta},\boldsymbol{\lambda})$，部分间隙就满足 $\mathcal{P}_{\mathbb{S}_1,\mathbb{S}_2}(\boldsymbol{\beta},\boldsymbol{\lambda})\geq 0$ 对所有 $\boldsymbol{\beta},\boldsymbol{\lambda}$ 都成立；并且 $\mathcal{P}_{\mathbb{S}_1,\mathbb{S}_2}(\boldsymbol{\beta},\boldsymbol{\lambda})=0$ 当且仅当 $(\boldsymbol{\beta},\boldsymbol{\lambda})$ 本身是一个鞍点时成立；参见鞍点性质 (8.3)。因此，$\mathcal{P}_{\mathbb{S}_1,\mathbb{S}_2}(\boldsymbol{\beta},\boldsymbol{\lambda})$ 可作为衡量点对 $(\boldsymbol{\beta},\boldsymbol{\lambda})$ 距离鞍点问题 (8.9) 最优解有多远的自然指标。

定理 8.8：（原始-对偶算法的收敛速率）假设与定理 8.7相同的条件。定义平均迭代 $\overline{\boldsymbol{\beta}}^{(T)} \triangleq T^{-1}\sum_{t=1}^{T}\boldsymbol{\beta}^{(t)}$和$\overline{\boldsymbol{\lambda}}^{(T)} \triangleq T^{-1}\sum_{t=1}^{T}\boldsymbol{\lambda}^{(t)}$。设 $\mathbb{S}_1 \subset \mathbb{R}^p$、$\mathbb{S}_2 \subset \mathbb{R}^n$为有界集合，且 $\mathbb{S}_1 \times \mathbb{S}_2$包含$\widehat{\ }\ \widehat{\ }$至少一个鞍点（$\widehat{\boldsymbol{\beta}}$，$\widehat{\boldsymbol{\lambda}}$）。那么，

$$\mathcal{P}_{\mathbb{S}_1,\mathbb{S}_2}(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\lambda}}^{(T)}) \leq \frac{D(\mathbb{S}_1,\mathbb{S}_2)}{T}, \tag{8.35}$$

其中 $D(\mathbb{S}_1, \mathbb{S}_2) \triangleq (2\phi)^{-1}$最大值$_{\boldsymbol{\lambda}\in\mathbb{S}_2}\left\|\boldsymbol{\lambda}-\widehat{\boldsymbol{\lambda}}\right\|_2^2+(2\gamma)^{-1}$最大值$_{\boldsymbol{\beta}\in\mathbb{S}_1}\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|_2^2$，且 $\mathcal{P}_{\mathbb{S}_1,\,\mathbb{S}_2}$表示式（8.34）中定义的部分原始-对偶间隙。

定理 8.8] 的证明：由 G 的凸性和凸共轭 F^*，并利用原始-对偶间隙不等式（8.31），对于任意（$\boldsymbol{\beta}$，$\boldsymbol{\lambda}$），有

$$\begin{aligned}
&\mathcal{L}(\overline{\boldsymbol{\beta}}^{(T)},\boldsymbol{\lambda}) - \mathcal{L}(\boldsymbol{\beta},\overline{\boldsymbol{\lambda}}^{(T)}) \\
&\quad= \left(\left\langle \boldsymbol{X}\overline{\boldsymbol{\beta}}^{(T)},\boldsymbol{\lambda}\right\rangle + G(\overline{\boldsymbol{\beta}}^{(T)}) - F^*(\boldsymbol{\lambda})\right) - \left(\left\langle \boldsymbol{X}\boldsymbol{\beta},\overline{\boldsymbol{\lambda}}^{(T)}\right\rangle + G(\boldsymbol{\beta}) - F^*(\overline{\boldsymbol{\lambda}}^{(T)})\right) \\
&\quad\leq \frac{1}{T}\sum_{t=1}^{T}\left(\left\langle \boldsymbol{X}\boldsymbol{\beta}^{(t)},\boldsymbol{\lambda}\right\rangle + G(\boldsymbol{\beta}^{(t)}) - F^*(\boldsymbol{\lambda})\right) - \frac{1}{T}\sum_{t=1}^{T}\left(\left\langle \boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\lambda}^{(t)}\right\rangle + G(\boldsymbol{\beta}) - F^*(\boldsymbol{\lambda}^{(t)})\right) \\
&\quad\leq \frac{1}{T}\left(\frac{1}{2\phi}\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{(0)}\right\|_2^2 + \frac{1}{2\gamma}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right\|_2^2\right).
\end{aligned}$$

对所有（$\boldsymbol{\beta}$，$\boldsymbol{\lambda}$）$\in \mathbb{S}_1 \times \mathbb{S}_2$取上确界即可证明（8.35）。 ■

不等式（8.35）表明，原始-对偶算法就部分原始-对偶间隙而言达到了 $\mathcal{O}(1/T)$ 的收敛速率。请注意，即使 $\mathbb{S}_1 \times \mathbb{S}_2$ 不包含鞍点，该上界仍然成立；然而，在这种情况下，部分间隙 $\mathcal{P}_{\mathbb{S}_1,\,\mathbb{S}_2}$ 可能变为负值，因而不再能解释为有意义的误差度量。

重要的是，不等式（8.35）并不意味着各个序列 $\mathcal{O}(1/T)$ 具有 $\left\|\boldsymbol{\beta}^{(t)}-\widehat{\boldsymbol{\beta}}\right\|_2$ 或 $\left\|\boldsymbol{\lambda}^{(t)}-\widehat{\boldsymbol{\lambda}}\right\|_2$速率。尽管如此，在实践中，这些范数下的迭代点通常也能以相当快的速度收敛。

第8章 问题

1. **合页损失**。考虑一维函数 $f:\ \mathbb{R}\to\mathbb{R}$ 由 $f(\beta)$ =max$\{1-\beta,0\}$ 定义。证明其凸共轭满足：对于任意 $\alpha\in\mathbb{R}$,

$$f^*(\alpha) = \max_{\beta}\left[\alpha\beta - \max\{1-\beta,0\}\right] = \alpha + \delta_{[-1,0]}(\alpha).$$

2. L令 $f:\ \mathbb{R}\to\mathbb{R}$ 由 $f(\beta)=\frac{1}{s}|\beta|^s$ 给出，其中 $s>1$。证明对于任意 $\alpha\in\mathbb{R}$,

$$f^*(\alpha) = \max_{\beta}\left\{\beta\alpha - \frac{1}{s}|\beta|^s\right\} = \frac{1}{t}|\alpha|^t, \quad \text{with } \frac{1}{s}+\frac{1}{t}=1.$$

3. **严格凸二次函数**。设 $f:\ \mathbb{R}^p\to\mathbb{R}$ 定义为

$$f(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{y}^\top\boldsymbol{\beta} + z,$$

其中 $\boldsymbol{X}$ 为严格正定矩阵，$\boldsymbol{y} \in \mathbb{R}^p$，且 $z \in \mathbb{R}$。证明对于任意 $\boldsymbol{\alpha} \in \mathbb{R}^p$，

$$f^*(\boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{y})^\top \boldsymbol{X}^{-1}(\boldsymbol{\alpha} - \boldsymbol{y}) - z.$$

如果 $\boldsymbol{X}$ 仅为半正定（即不可逆），会发生什么?

4. **示性函数**。设 $f = \delta_{\mathbb{S}}$ 为凸集 $\mathbb{S}$ 的示性函数（见习题 2.36）。证明其凸共轭为

$$f^*(\boldsymbol{\alpha}) = \sup_{\boldsymbol{\beta} \in \mathbb{S}} \langle \boldsymbol{\beta}, \boldsymbol{\alpha} \rangle = \sigma_{\mathbb{S}}(\boldsymbol{\alpha}),$$

其中 $\sigma_{\mathbb{S}}(\boldsymbol{\alpha}) = \max_{\boldsymbol{\beta} \in \mathbb{S}} \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle$ 表示集合 $\mathbb{S}$ 的支撑函数。

5. **范数**。设 $f: \mathbb{R}^p \to \mathbb{R}$ 由 $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|$给定，其中 $\|\cdot\|$ 为任意范数。首先，证明

$$f(\boldsymbol{\beta}) = \sigma_{\mathbb{B}_{\|\cdot\|_*}[\mathbf{0},1]}(\boldsymbol{\beta}),$$

其中 $\sigma_{\mathbb{S}}(\boldsymbol{\alpha}) = \max_{\boldsymbol{\beta} \in \mathbb{S}} \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle$ 表示 $\mathbb{S}$ 的支撑函数，而 $\|\cdot\|_*$ 表示对偶范数。然后证明

$$f^*(\boldsymbol{\alpha}) = \delta_{\mathrm{cl}(\mathrm{conv}(\mathbb{B}_{\|\cdot\|_*}[\mathbf{0},1]))}(\boldsymbol{\alpha}),$$

由于 $\mathbb{B}_{\|\cdot\|_*}[\mathbf{0}, 1]$ 是闭且凸的，$\mathrm{cl}(\mathrm{conv}(\mathbb{B}_{\|\cdot\|_*}[\mathbf{0}, 1])) = \mathbb{B}_{\|\cdot\|_*}[\mathbf{0}, 1]$，因此对于任意 $\boldsymbol{\alpha} \in \mathbb{R}^p$，

$$f^*(\boldsymbol{\alpha}) = \delta_{\mathbb{B}_{\|\cdot\|_*}[\mathbf{0},1]}(\boldsymbol{\alpha}) = \begin{cases} 0, & \|\boldsymbol{\alpha}\|_* \le 1; \\ \infty, & \text{else.} \end{cases}$$

6. **合页损失与支持向量机**。给定数据对 $(\boldsymbol{x}_1, y_1)$，$(\boldsymbol{x}_2, y_2) \ldots$，$(\boldsymbol{x}_n, y_n)$，考虑软间隔支持向量机（SVM）问题：

$$\min_{\boldsymbol{\beta}, b} \|\boldsymbol{\beta}\|_2^2 + \sum_{i=1}^{n} \max\{0, 1 - y_i(\langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle + b)\}.$$

讨论如何应用原始-对偶算法求解此优化问题。

7. 描述如何使用原始-对偶算法求解以下线性规划问题：

$$\min_{\boldsymbol{\beta}} \langle \boldsymbol{\beta}, \boldsymbol{c} \rangle \quad \text{s.t.} \quad \begin{cases} \boldsymbol{X}\boldsymbol{\beta} & = \boldsymbol{y}; \\ \boldsymbol{\beta} & \ge \mathbf{0}. \end{cases}$$

8. (Goldluecke 和 Cremers，2010年) 二维网格上的 Potts 模型可写成如下凸松弛：

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{D}\boldsymbol{\beta}\|_1 + \langle \boldsymbol{\beta}, \boldsymbol{c} \rangle \quad \text{s.t.} \quad 0 \le \beta_i \le 1, \quad \forall\, i,$$

其中，$\boldsymbol{D}$ 是一个用于编码相邻像素之间有限差分的矩阵（例如水平方向和垂直方向的梯度），$\|\boldsymbol{D}\boldsymbol{\beta}\|_1$ 近似表示相邻像素绝对差之和，$\boldsymbol{c}$ 包含数据保真项。讨论原始-对偶算法在此模型中的应用。

第9章

稀疏回归问题算法

目录

9.1 算法： ℓ_0-约束问题.319

9.1.1 驻点条件下的最优性条件 L-驻点条件.320 9.1.2 迭代硬阈值（IHT）方法.322

9.2 约束 LASSO 问题的算法.329

9.2.1 投影梯度下降（PGD）.329 9.2.2 近端梯度法.332 9.2.3 条件梯度（Frank–Wolfe）法.336 9.2.4 最小角回归（LARS）.339

9.3 拉格朗日套索问题算法.353

9.3.1 Huber损失平滑.354 9.3.2 近端梯度与主化最小化方法 . . . 355 9.3.3 快速近端梯度法—FISTA.357 9.3.4 罚函数法.361 9.3.5 交替方向乘子法（ADMM） . . . . . 365 9.3.6 块坐标下降 .371 9.3.7 Barzilai–Borwein 梯度投影算法 .372

9.4 其他稀疏优化建模形式.374 9.4.1 广义 LASSO 问题.374 9.4.2 组 LASSO 与对偶多面体投影.375 9.4.3 稀疏噪声.380 **第9章习题.381**



T本章全面概述了用于求解稀疏优化问题的算法，重点介绍 ℓ_0约束形式和 LASSO 建模形式。我们首先考察用于 ℓ_0约束问题的迭代方法，例如迭代硬阈值（IHT）方法，它直接施加稀疏性约束。接下来，我们将讨论约束 LASSO 问题，介绍其中的关键方法，包括投影梯度下降、近端梯度方法、条件梯度法（Frank–Wolfe）和最小角回归（LARS）。本章还将进一步探讨 LASSO 的拉格朗日形式，涵盖 Huber 损失平滑、FISTA、罚函数法、交替方向乘子法（ADMM）和块坐标下降等高级技术。最后，我们将这一框架扩展到其他稀疏模型，包括广义 LASSO、组 LASSO 和稀疏噪声模型。这些算法共同构成了现代机器学习和信号处理领域稀疏恢复与特征选择方法的基础。

9.1. l_0约束问题的算法

在第9.2节中，我们将介绍几种求解约束 LASSO 问题（L_C）（第 126 页）的方法，包括 投影梯度下降、近端梯度法和条件梯度法。只要可行集 $\mathbb{S}$ 是闭且凸的，这些方法也适用于形式为 $\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$ s.t. $\boldsymbol{\beta} \in \mathbb{S}$的一般约束优化问题。不过，本节关注的是约束集为非凸集的稀疏优化问题，具体而言，是 ℓ_0-约束问题。因此，前述凸优化技术无法直接应用。尽管凸情形与非凸情形之间存在一些表面上的相似之处，但二者的理论保证和算法行为有显著差异。为解决这类问题，我们将介绍迭代硬阈值算法（IHT）。

特别地，我们将讨论在稀疏性约束下最小化一般连续可微目标函数这一挑战，见（S_0）（第 125页）：

$$(\mathrm{S}_0) \qquad \min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_0 \le k, \tag{9.1}$$

其中 $f: \ \mathbb{R}^p \to \mathbb{R}$ 是 L_b-光滑的（见定义 2.78 和定理 2.80）， $k > 0$ 是满足 $k < p$的整数，并且 $\|\boldsymbol{\beta}\|_0$ 表示 $\boldsymbol{\beta}$的所谓“ℓ_0-范数”，该范数统计 $\boldsymbol{\beta}$中的非零分量数量：

$$\|\boldsymbol{\beta}\|_0 = \big|\{i \mid \beta_i \neq 0, \forall\, i \in \{1, 2, \ldots, p\}\}\big|.$$

需要注意的是， ℓ_0-范数并不是真正的范数，因为它不满足齐次性：对于任意标量 $\lambda \neq 0$ （定义 1.24），我们有 $\|\lambda\boldsymbol{\beta}\|_0 = \|\boldsymbol{\beta}\|_0$ （见定义 1.24）。此外，我们不假设 f 是凸的，且约束集本质上是非凸的。因此，许多标准的凸优化理论并不适用于问题（S_0）。然而，该问题在压缩感知等应用中具有核心重要性，这推动了专用算法和适用分析工具的发展。

我们回顾一些有用的符号说明。对于任意向量 $\boldsymbol{\beta} \in \mathbb{R}^p$，将其支撑集及其补集定义为

$$\mathbb{I}_1(\boldsymbol{\beta}) \triangleq \operatorname{supp}(\boldsymbol{\beta}) \triangleq \{i \mid \beta_i \neq 0\} \qquad \text{and} \qquad \mathbb{I}_0(\boldsymbol{\beta}) \triangleq \{i \mid \beta_i = 0\}. \tag{9.2}$$

一个向量 $\boldsymbol{\beta}$ 称为 k-稀疏的，如果 $|\operatorname{supp}(\boldsymbol{\beta})| \le k$；或者等价地，如果 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ （以原点为中心、半径为ℓ_0的闭 k 球： $\mathbb{B}_0[k] = \mathbb{B}_0[\mathbf{0}, k]$ 为简便起见；参见定义1.29）。

使用这一符号，问题（S_0）可改写为：

$$\min \ f(\boldsymbol{\beta}) \qquad \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{B}_0[k]. \tag{9.3}$$

对于向量 $\boldsymbol{\beta} \in \mathbb{R}^p$和 $i \in \{1, 2, \ldots, p\}$，令 $[\boldsymbol{\beta}]_i$表示其各分量绝对值中第 i大的值，因此

$$[\boldsymbol{\beta}]_1 \geq [\boldsymbol{\beta}]_2 \geq \ldots \geq [\boldsymbol{\beta}]_p, \tag{9.4}$$

其中

$$[\boldsymbol{\beta}]_1 = \max_{i=1,\ldots,p} |\beta_i| \quad \text{and} \qquad [\boldsymbol{\beta}]_p = \min_{i=1,\ldots,p} |\beta_i| \,. \tag{9.5}$$

9.1.1 L-驻点条件下的最优性条件

现在，我们的目标是探讨问题（S_0）的必要最优性条件。由于ℓ_0约束是非凸的，凸优化中的经典驻点条件通常并不适用。不过，我们可以针对稀疏性约束的结构，提出并分析其他最优性概念。

我们首先利用投影算子来刻画驻点条件，从而将驻点概念推广到问题（S_0）。回顾一下，对于在闭且凸集上最小化连续可微函数的问题，点 $\boldsymbol{\beta}^*$ 是驻点，当且仅当 $\mathbb{S}$

$$\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}^* - \eta \nabla f(\boldsymbol{\beta}^*)), \tag{9.6}$$

其中 $\eta > 0$ 是任意正标量（参见推论 2.113）。尽管条件（9.6）是用 η表示的，但实际上它与 η的选择无关。然而，当 $\mathbb{S} = \mathbb{B}_0[k]$时，这一刻画不能直接推广，因为向非凸集（例如 $\mathbb{B}_0[k]$ ）作正交投影并非唯一。相反，投影成为由下式定义的集值映射：

$$\mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\alpha}} \{\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 \mid \boldsymbol{\alpha} \in \mathbb{B}_0[k]\}.$$

根据魏尔斯特拉斯定理（定理 2.56(c)），由 $\mathcal{P}_{\mathbb{B}_0[k]}$($\boldsymbol{\beta}$的闭性以及函数 $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$（相对于 $\boldsymbol{\alpha}$）的强制性，可知 $\mathcal{P}_{\mathbb{B}_0[k]}$($\boldsymbol{\beta}$中元素的存在性成立。为了将条件（9.6）推广到稀疏约束问题（S_0），我们引入“L-驻点条件”的概念。

定义 9.1（L-驻点条件）。 如果向量 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k]$ 满足问题（S_0），则称其为该问题的 L-驻点 点：

$$\boldsymbol{\beta}^* \in \mathcal{P}_{\mathbb{B}_0[k]} \left(\boldsymbol{\beta}^* - \frac{1}{L} \nabla f(\boldsymbol{\beta}^*) \right). \tag{9.7}$$

由于投影算子 $\mathcal{P}_{\mathbb{B}_0[k]}(\cdot)$是一个集合值映射，其输出由所有满足以下条件的向量构成：保留 $\boldsymbol{\beta}$ 中绝对值最大的 k 个元素，并将其余元素置零。例如，

$$\mathcal{P}_{\mathbb{B}_0[2]}([3, 1, 1]^\top) = \{[3, 1, 0]^\top, [3, 0, 1]^\top\}.$$

现在我们证明，在假设 f 是 L_b-光滑的条件下，只要 $L > L_b$， L-驻点条件就是最优性的必要条件。在证明这一结果之前，我们先对 L-驻点条件给出一个更显式的刻画。

引理9.2：（L-驻点条件）对于任意 $L>0$，向量 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k]$ 满足 L-驻点条件，当且仅当

$$\left|\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \begin{cases} \leq L \cdot [\boldsymbol{\beta}^*]_k, & \text{if } i \in \mathbb{I}_0(\boldsymbol{\beta}^*); \\ = 0, & \text{if } i \in \mathbb{I}_1(\boldsymbol{\beta}^*), \end{cases} \tag{9.8}$$

w这里， $[\boldsymbol{\beta}^*]_k$ 表示 $\boldsymbol{\beta}$各分量中绝对值第 k大的值 *.

引理 9.2] 的证明。假设 $\boldsymbol{\beta}^*$ 满足 L-驻点条件。注意，对于任意索引$j \in \{1,2,\ldots,p\}$， $\mathcal{P}_{\mathbb{B}_0[k]}$ $(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*))$中任意向量的第 j个分量要么为零，要么等于 $x_j^* - \frac{1}{L}\frac{\partial f}{\partial x_j}(\boldsymbol{\beta}^*)$。另一方面，由于 $\boldsymbol{\beta}^* \in \mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*))$，可知若 $i \in \mathbb{I}_1(\boldsymbol{\beta}^*)$，则 $\beta_i^* = \beta_i^* - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)$，因此 $\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*) = 0$。若 $i \in \mathbb{I}_0(\boldsymbol{\beta}^*)$，则 $\left|\beta_i^* - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \leq [\boldsymbol{\beta}^*]_k$ 且 $\beta_i^* = 0$，从而得到 $\left|\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \leq L \cdot [\boldsymbol{\beta}^*]_k$。因此，条件 (9.8) 成立。

反过来，假设 $\boldsymbol{\beta}^*$ 满足 (9.8)。若 $\|\boldsymbol{\beta}^*\|_0 < k$，则 $[\boldsymbol{\beta}^*]_k = 0$，由 (9.8) 可得 $\nabla f(\boldsymbol{\beta}^*) = \mathbf{0}$；因此，在这种情况下， $\mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)\right) = \mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta}^*)$ 是集合 $\{\boldsymbol{\beta}^*\}$，显然满足 L- 驻点条件。另一方面，若 $\|\boldsymbol{\beta}^*\|_0 = k$，则 $[\boldsymbol{\beta}^*]_k \neq 0$ 且 $|\mathbb{I}_1(\boldsymbol{\beta}^*)| = k$。由 (9.8)，

$$\left|\beta_i^* - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \begin{cases} \leq [\boldsymbol{\beta}^*]_k, & i \in \mathbb{I}_0(\boldsymbol{\beta}^*), \\ = |\beta_i^*|, & i \in \mathbb{I}_1(\boldsymbol{\beta}^*). \end{cases}$$

因此，向量 $\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)$ 包含 $\boldsymbol{\beta}^*$ 中绝对值最大的 k个分量，其余所有分量的大小都小于或等于它们。这也意味着 L-驻点条件成立，并完成了证明。

■

显然，条件（9.8）取决于常数 L；随着 L减小，条件会变得更加严格。这凸显了非凸与凸情形之间的一个根本区别。在探究 L-驻点条件构成必要最优性条件的条件之前，我们先证明一个虽具技术性但很有用的结果。

引理 9.3：（光滑条件下稀疏投影的下降引理）设 $f:\mathbb{R}^p \to \mathbb{R}$ 是一个 L_b-光滑函数，且 $L > L_b$。那么，对于任意 $\boldsymbol{\beta} \in \mathbb{B}_0[k]$ 以及任意满足$\boldsymbol{\alpha} \in \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta})\right)$的 $\boldsymbol{\alpha} \in \mathbb{R}^p$，有

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \geq \frac{L - L_b}{2}\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2.$$

证明 [引理 9.3] 由于 $\boldsymbol{\alpha} \in \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta})\right)$，可得

$$\begin{aligned} \boldsymbol{\alpha} &\in \arg\min_{\boldsymbol{u} \in \mathbb{B}_0[k]} \left\|\boldsymbol{u} - \left(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta})\right)\right\|_2^2 = \arg\min_{\boldsymbol{u} \in \mathbb{B}_0[k]} \frac{L}{2}\left\|\boldsymbol{u} - \left(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta})\right)\right\|_2^2 + C_{\boldsymbol{u}} \\ &= \arg\min_{\boldsymbol{u} \in \mathbb{B}_0[k]} \left\{ f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta}), \boldsymbol{u} - \boldsymbol{\beta}\rangle + \frac{L}{2}\|\boldsymbol{u} - \boldsymbol{\beta}\|_2^2 \triangleq \Psi_L(\boldsymbol{u}, \boldsymbol{\beta}) \right\}, \end{aligned} \tag{9.9}$$

其中 $C_{\boldsymbol{u}} \triangleq f(\boldsymbol{\beta}) - \frac{1}{2L}\|\nabla f(\boldsymbol{\beta})\|_2^2$表示一个相对于 $\boldsymbol{u}$ 的常数。这意味着

$$\Psi_L(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq \Psi_L(\boldsymbol{\beta}, \boldsymbol{\beta}) = f(\boldsymbol{\beta}).$$

根据 L_b的光滑性 f （定义2.78），我们还有

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\beta}) - \Psi_{L_b}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Psi_L(\boldsymbol{\beta}, \boldsymbol{\beta}) - \Psi_{L_b}(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

结合前述性质以及 $\Psi_{L_b}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Psi_L(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \frac{L-L_b}{2}\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2$这一事实，即可得到所需结果。 ■

现在，我们可以证明当$L > L_b$ 且 f 是 L_b-光滑时， L-驻点条件对于最优性是必要的。

定理 9.4：（L-驻点条件与光滑性下的必要性）设 f： $\mathbb{R}^p \to \mathbb{R}$是一个 L_b-光滑函数（定义 2.78），并设 $L > L_b$。若 $\boldsymbol{\beta}^*$ 是问题（S_0）的一个最优解，则：

(i) $\boldsymbol{\beta}^*$ 是一个 L-驻点。(ii) 集合 $\mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)\right)$是一个单元素集合。

证明 [定理 9.4] 反设投影集包含一个向量$\boldsymbol{\alpha} \neq \boldsymbol{\beta}^*$，即 $\boldsymbol{\alpha} \in \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)\right)$。由引理 9.3 可得， $f(\boldsymbol{\beta}^*) - f(\boldsymbol{\alpha}) \geq \frac{L-L_b}{2}\|\boldsymbol{\beta}^* - \boldsymbol{\alpha}\|_2^2$，这与 $\boldsymbol{\beta}^*$ 的最优性矛盾。因此， $\boldsymbol{\beta}^*$ 是集合 $\mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)\right)$中的唯一向量，并且 $\boldsymbol{\beta}^*$是 L-驻点。 ■

总之，我们已经证明，对于任意 L_b-光滑目标函数， L-驻点条件与任意 $L > L_b$ 都是问题 (S_0) 中最优性的必要条件。

9.1.2 迭代硬阈值（IHT）方法

算法 2 迭代硬阈值（IHT）方法

Require: Sparsity level s; 1: 输入 (a)。一个可微函数 f; 2: 输入 (b)。 $f(\boldsymbol{\beta}) = \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2$: 设计矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和响应 $\boldsymbol{y} \in \mathbb{R}^n$; 3: 初始化： $\boldsymbol{\beta}^{(1)} \in \mathbb{B}_0[s]$; 4: 对于 $t = 1, 2, \ldots,$ 执行 5: 选择步长 η_t; 6: $\boldsymbol{\alpha}^{(t+1)} (\overset{a}{\leftarrow} \boldsymbol{\beta}^{(t)} - \eta_t \nabla f)$ $(\boldsymbol{\beta}^{(t)})$ 或 $(b) \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t \frac{1}{n}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})$; 7: $\boldsymbol{\beta}^{(t+1)} \in \mathcal{P}_{\mathbb{B}_0[s]}(\boldsymbol{\alpha}^{(t+1)})$; ▷ 选择 $\boldsymbol{\alpha}^{(t+1)}$ 中大小最大的 s 个元素; 8: 结束循环 9: 返回最终的 $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^{(t)}$;

我们之前介绍了一般 ℓ_0-约束优化问题在第9.1.1节：

$$(S_0) \qquad \min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_0 \leq k, \tag{9.10}$$

求解问题（S_0）的一种自然方法是推广投影梯度下降（PGD）框架，在梯度下降步与向非凸稀疏集 $\mathbb{B}_0[k] = \{\boldsymbol{\alpha} \mid \|\boldsymbol{\alpha}\|_0 \leq k\}$ 的投影之间交替进行（详见第 9.2.1 节）。由此得到迭代方案

$$\boldsymbol{\beta}^{(t+1)} \in \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^{(t)} - \frac{1}{L}\nabla f(\boldsymbol{\beta}^{(t)})\right), \quad t = 1, 2, \ldots. \tag{9.11a}$$

文献中将该方法称为 迭代硬阈值（*IHT*）方法（见算法2），本文沿用这一术语。该方法既适用于一般的光滑函数f（在满足引理 9.5中所述假设的条件下），也适用于以下形式的最小二乘目标函数$f(\boldsymbol{\beta}) = \frac{1}{2n}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2$（该目标函数满足前文讨论的某些设计性质）。可以证明，IHT 方法的一般步骤等价于以下关系

$$\boldsymbol{\beta}^{(t+1)} \in \underset{\boldsymbol{\beta}\in\mathbb{B}_0[k]}{\arg\min}\, \Psi_L(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}), \quad t=1,2,\ldots, \tag{9.11b}$$

其中，代理函数定义为 $\Psi_L(\boldsymbol{u},\ \boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta})\ \boldsymbol{u}-\boldsymbol{\beta}\rangle + \frac{L}{2}\|\boldsymbol{u}-\boldsymbol{\beta}\|_2^2$，如式（9.9）所示。

当约束集 $\mathbb{S}$ 为闭凸集时，投影梯度下降（参见算法3）可被解释为求解 $\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}\big(\boldsymbol{\beta}^* - \eta\nabla f(\boldsymbol{\beta}^*)\big)$（推论2.113）的不动点方法。尽管 $\mathbb{B}_0[k]$ 是非凸的，迭代硬阈值方法同样可以看作一种不动点迭代，旨在满足 L-驻点条件（定义9.1）。具体而言，假设 f 是 L_b-光滑且 $L > L_b$，则 IHT 迭代序列的任何极限点都满足

$$\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}\big(\boldsymbol{\beta}^* - \frac{1}{L}\nabla f(\boldsymbol{\beta}^*)\big)$$

根据定理9.4，这是最优性的必要条件。

以下引理总结了迭代硬阈值方法的若干基本收敛性质。

引理9.5：（光滑性条件下迭代硬阈值的性质）设 $f:\ \mathbb{R}^p \to \mathbb{R}$ 是一个 L_b-光滑且下有界的函数。设 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 是由迭代硬阈值方法（算法2(a)）以恒定步长 $\eta \triangleq \eta_t \triangleq \frac{1}{L}$ 生成的序列，其中 $L > L_b$，投影稀疏度为 $s = k$。则：

(i) $f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}^{(t+1)}) \geq \frac{L-L_b}{2}\big\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t+1)}\big\|_2^2$。(ii) 序列 $\{f(\boldsymbol{\beta}^{(t)})\}_{t>0}$ 单调不增。(iii) 当 $t\to\infty$时，$\big\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t+1)}\big\|_2 \to 0$。(iv) 对每个 $t=1,\ 2,\ \ldots,$ 若 $\boldsymbol{\beta}^{(t)} \neq \boldsymbol{\beta}^{(t+1)}$，则 $f(\boldsymbol{\beta}^{(t+1)}) < f(\boldsymbol{\beta}^{(t)})$。

证明 [引理 95] 。第 (i) 部分可直接由引理 9.3 令 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$ 且$\boldsymbol{\alpha} = \boldsymbol{\beta}^{(t+1)}$即可得到。第 (ii) 部分由 (i) 立即推出，因为右侧非负。为证明 (iii)，注意到 $\{f(\boldsymbol{\beta}^{(t)})\}_{t>0}$ 单调不增且有下界，因此收敛到某个极限 $B \in \mathbb{R}$。因此，$f(\boldsymbol{\beta}^{(t)})) - f(\boldsymbol{\beta}^{(t+1)}) \to B - B = 0$ 当 $t\to\infty$ 时。由第 (i) 部分以及 $L > L_b$这一事实可知，$\big\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t+1)}\big\|_2 \to 0$。最后，(iv) 可直接由 (i) 得出：若 $\boldsymbol{\beta}^{(t)} \neq \boldsymbol{\beta}^{(t+1)}$，则范数差为正，因此目标函数严格递减。

■

如前所述，IHT算法可解释为一种用于满足 L-驻点条件的不动点方法。以下定理证明，采用恒定步长 $\frac{1}{L}$ 生成的序列的每个聚点确实都是 L-驻点。

定理 9.6：（迭代硬阈值的收敛性（Beck，2014年））假设与引理 9.5中的条件相同。设 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 为由迭代硬阈值方法生成的序列（算法

2(a))，其步长为恒定步长 $\eta \triangleq \eta_t \triangleq \frac{1}{L}$，其中 $L > L_b$，稀疏度为 $s = k$。那么，$\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 的任意聚点都是一个 L-驻点。

证明 [定理 9.6]设 $\boldsymbol{\beta}^*$ 是序列 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$的一个聚点。则存在一个子序列 $\{\boldsymbol{\beta}^{(t_j)}\}_{j>0}$ 使得 $\boldsymbol{\beta}^{(t_j)} \to \boldsymbol{\beta}^*$ 当 $j \to \infty$时。由引理 9.5，可得

$$f(\boldsymbol{\beta}^{(t_j)}) - f(\boldsymbol{\beta}^{(t_j+1)}) \geq \frac{L - L_b}{2} \left\| \boldsymbol{\beta}^{(t_j)} - \boldsymbol{\beta}^{(t_j+1)} \right\|_2^2. \tag{9.12}$$

由于 $\{f(\boldsymbol{\beta}^{(t_j)})\}_{j>0}$ 和 $\{f(\boldsymbol{\beta}^{(t_j+1)})\}_{j>0}$ 是单调不增且有下界的序列，因此它们都收敛到同一个极限。因此，$f(\boldsymbol{\beta}^{(t_j)}) - f(\boldsymbol{\beta}^{(t_j+1)}) \to 0$ 当 $j \to \infty$ 时，结合式（9.12）可得 $\boldsymbol{\beta}^{(t_j+1)} \to \boldsymbol{\beta}^*$ 当 $j \to \infty$. 时。回顾一下，对所有 $j > 0$, $\boldsymbol{\beta}^{(t_j+1)} \in \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}^{(t_j)} - \frac{1}{L}\nabla f(\boldsymbol{\beta}^{(t_j)})\right)$.接下来考虑以下两种情况。

令 $i \in \mathbb{I}_1(\boldsymbol{\beta}^*)$（即属于支撑集；$\mathbb{I}_1(\boldsymbol{\beta}) \triangleq \text{supp}(\boldsymbol{\beta}) \triangleq \{i \mid \beta_i \neq 0\}$）。由 $\boldsymbol{\beta}^{(t_j)}$ 和 $\boldsymbol{\beta}^{(t_j+1)}$ 收敛到 $\boldsymbol{\beta}^*$，可知存在一个整数 J，使得

$$\beta_i^{(t_j)}, \beta_i^{(t_j+1)} \neq 0 \text{ for all } j > J,$$

因此，对于 $j > J$,

$$\beta_i^{(t_j+1)} = \beta_i^{(t_j)} - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^{(t_j)}).$$

令 $j \to \infty$ 取极限，并利用 f 的连续性，可得 $\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*) = 0$，对 $i \in \mathbb{I}_1(\boldsymbol{\beta}^*)$ 成立。

现在令 $i \in \mathbb{I}_0(\boldsymbol{\beta}^*)$。若存在无穷多个索引 t_j 使得 $\beta^{(t_j+1)}{}_i \neq 0$，那么与前一种情况一样，可得对于这些索引，$\beta^{(t_j+1)}{}_i = \beta^{(t_j)}{}_i - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^{(t_j)})$，这意味着（取极限可得）$\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*) = 0$。特别地，$\left|\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \leq L[\boldsymbol{\beta}^*]_k$。另一方面，若存在一个 $M > 0$，使得对所有 $j > M$，$\beta^{(t_j+1)}{}_i = 0$，那么

$$\left|\beta_i^{(t_j)} - \frac{1}{L}\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^{(t_j)})\right| \leq \left[\boldsymbol{\beta}^{(t_j)} - \frac{1}{L}\nabla f(\boldsymbol{\beta}^{(t_j)})\right]_k = \left[\boldsymbol{\beta}^{(t_j+1)}\right]_k.$$

令 $j \to \infty$ 取极限，并利用映射 $[\cdot]_k$ 的连续性（当第 k 大和第 $(k+1)$ 大幅值不相等时，该映射连续；由于该子序列收敛到 $\boldsymbol{\beta}^*$，这一条件沿该子序列成立），我们得到

$$\left|\frac{\partial f}{\partial \beta_i}(\boldsymbol{\beta}^*)\right| \leq L[\boldsymbol{\beta}^*]_k, \quad i \in \mathbb{I}_0(\boldsymbol{\beta}^*).$$

综合这两种情况，我们得出 $\boldsymbol{\beta}^*$ 满足引理 L中的驻点条件9.2。 ■

现在我们证明一个重要结果：如果设计矩阵满足限制等距性质（RIP；参见定义4.30），且其常数满足适当条件，那么对于最小二乘目标函数，IHT算法可保证普适稀疏恢复：$f(\boldsymbol{\beta}) = \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2$。

定理 9.7：（RIP 条件下 IHT 的收敛速度（Jain 等人，2014年，2017年））假设 $\frac{1}{\sqrt{n}}\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 $3k$的 RIP，常数为 $\delta_{3k} < \frac{1}{2}$。设 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k] \subset \mathbb{R}^p$ 是任意一个 k-稀疏向量，并设 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$。那么，采用恒定步长 2(b) 中的 IHT 算法在步长为 $\eta_t = 1$ 、稀疏度为 $s = k$时，

会生成一个序列 $\{\boldsymbol{\beta}^{(t)}\}$，使得 $\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\|_2 \leq \epsilon$，且最多经过 $t=\mathcal{O}\left(\ln \frac{\|\boldsymbol{\beta}^*\|_2}{\epsilon}\right)$次迭代。

证明 [定理97]设 $\mathbb{S}^* \triangleq \operatorname{supp}(\boldsymbol{\beta}^*)$ 和 $\mathbb{S}^t \triangleq \operatorname{supp}(\boldsymbol{\beta}^{(t)})$。定义指标集$\mathbb{I}^t \triangleq \mathbb{S}^t \cup \mathbb{S}^{t+1} \cup \mathbb{S}^*$，其中包含两个连续迭代点的支撑集和真实支撑集。由于每个集合的基数至多为 k，因此有 $\left|\mathbb{I}^t\right| \leq 3k$。关键在于，两个误差向量 $\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*$ 和 $\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^*$ 都支撑于 $\mathbb{I}^t$。

步长为 $\eta_t=1$时，算法 2(b) 中的梯度步得到 $\boldsymbol{\alpha}^{(t+1)}=\boldsymbol{\beta}^{(t)}-\frac{1}{n}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y})$。尽管 $\mathbb{B}_0[k]$ 是非凸的，投影 $\boldsymbol{\beta}^{(t+1)}=\mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\alpha}^{(t+1)})$ 仍满足基本的最优性性质（投影性质-O；见引理 2.93）：

$$
\begin{aligned}
&\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\alpha}^{(t+1)}\right\|_2^2 \leq \left\|\boldsymbol{\beta}^*-\boldsymbol{\alpha}^{(t+1)}\right\|_2^2 \\
\Longrightarrow \quad &\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\alpha}_{\mathbb{I}}^{(t+1)}\right\|_2^2+\left\|\boldsymbol{\beta}_{\mathbb{I}^c}^{(t+1)}-\boldsymbol{\alpha}_{\mathbb{I}^c}^{(t+1)}\right\|_2^2 \leq \left\|\boldsymbol{\beta}_{\mathbb{I}}^*-\boldsymbol{\alpha}_{\mathbb{I}}^{(t+1)}\right\|_2^2+\left\|\boldsymbol{\beta}_{\mathbb{I}^c}^*-\boldsymbol{\alpha}_{\mathbb{I}^c}^{(t+1)}\right\|_2^2,
\end{aligned}
$$

其中，为简洁起见，我们记 $\mathbb{I} \triangleq \mathbb{I}^t$ 使得 $\boldsymbol{\beta}_{\mathbb{I}^c}^{(t+1)}=\boldsymbol{\beta}_{\mathbb{I}^c}^*=\mathbf{0}$。利用 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^*$这一事实，并记 $\widetilde{\boldsymbol{X}} \triangleq \frac{1}{\sqrt{n}}\boldsymbol{X}$和 $\widetilde{\boldsymbol{X}}_{\mathbb{I}} \triangleq \widetilde{\boldsymbol{X}}[:,\mathbb{I}]$，可得

$$
\begin{aligned}
\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^*\right\|_2=\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\beta}_{\mathbb{I}}^*\right\|_2 &\leq \left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\alpha}_{\mathbb{I}}^{(t+1)}\right\|_2+\left\|\boldsymbol{\alpha}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\beta}_{\mathbb{I}}^*\right\|_2 \leq 2\left\|\boldsymbol{\alpha}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\beta}_{\mathbb{I}}^*\right\|_2 \\
&=2\left\|\left(\boldsymbol{\beta}_{\mathbb{I}}^{(t)}-\widetilde{\boldsymbol{X}}_{\mathbb{I}}^\top\widetilde{\boldsymbol{X}}(\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*)\right)-\boldsymbol{\beta}_{\mathbb{I}}^*\right\|_2 \overset{\dagger}{=} 2\left\|(\boldsymbol{I}-\widetilde{\boldsymbol{X}}_{\mathbb{I}}^\top\widetilde{\boldsymbol{X}}_{\mathbb{I}})(\boldsymbol{\beta}_{\mathbb{I}}^{(t)}-\boldsymbol{\beta}_{\mathbb{I}}^*)\right\|_2 \\
&\leq 2\left\|\widetilde{\boldsymbol{X}}_{\mathbb{I}}^\top\widetilde{\boldsymbol{X}}_{\mathbb{I}}-\boldsymbol{I}\right\|_2\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t)}-\boldsymbol{\beta}_{\mathbb{I}}^*\right\|_2 \leq 2\delta_{3k}\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\|_2,
\end{aligned}
$$

其中，等式（†）成立是因为 $\boldsymbol{\beta}_{\mathbb{I}^c}^{(t)}=\boldsymbol{\beta}_{\mathbb{I}^c}^*=\mathbf{0}$，最后一个不等式则由 RIP 条件（定义 4.30 或备注 4.31）推出。为达到 $\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\|_2 \leq \epsilon$，只需取 $t=\mathcal{O}\left(\ln \frac{\|\boldsymbol{\beta}^*\|_2}{\epsilon}\right)$，并以 $\boldsymbol{\beta}^{(1)}=\mathbf{0}$ 初始化，即完成证明。 ■

上述定理确立了最小二乘目标函数下 IHT算法的收敛速率。为将这一结果推广到一般的光滑目标函数，我们首先需要以下关于稀疏投影算子的引理。

引理 9.8：（稀疏投影性质） 令 $\boldsymbol{\beta} \in \mathbb{R}^p$，并令 $\widetilde{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta})$ 表示其最佳 k-稀疏近似。于是，对于任意满足 $k^* \leq k \leq p$ 的 k^*-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k^*] \subseteq \mathbb{R}^p$，以下不等式成立：

$$
\frac{\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2^2}{p-k} \leq \frac{\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|_2^2}{p-k^*} \qquad \Longrightarrow \qquad \left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2^2 \leq \frac{p-k}{p-k^*}\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|_2^2.
$$

证明 [引理 9。8]不失一般性，假设 $\boldsymbol{\beta}$ 的坐标按如下顺序排列：$|\beta_1| \geq |\beta_2| \geq \ldots \geq |\beta_p|$。由于投影算子 $\mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta})$ 的运作方式是~ ~ ~按大小选取最大的元素，因此有 $\beta_1=\beta_1$，$\beta_2=\beta_2\ldots$，$\beta_k=\beta_k$且~ ~ ~$\beta_{k+1}=\beta_{k+2}=\ldots=\beta_p=0$。

类似地，定义 $\widehat{\boldsymbol{\beta}} \triangleq \mathcal{P}_{\mathbb{B}_0[k^*]}(\boldsymbol{\beta})$，使得 $\widehat{\beta}_1 = \beta_1$，$\widehat{\beta}_2 = \beta_2$，..，$\widehat{\beta}_{k^*} = \beta_{k^*}$ 且 $\widehat{\beta}_{k^*+1} = \widehat{\beta}_{k^*+2} = \ldots = \beta_p = 0$。由于 $k^* \leq k$，我们计算得

$$\begin{aligned}(p-k)(p-k^*)\left(\frac{\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2^2}{p-k^*}-\frac{\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_2^2}{p-k}\right) &= (p-k)\sum_{i=k^*+1}^{p}\beta_i^2-(p-k^*)\sum_{i=k+1}^{p}\beta_i^2 \\ &= (p-k)\sum_{i=k^*+1}^{k}\beta_i^2+\big((p-k)-(p-k^*)\big)\sum_{i=k+1}^{p}\beta_i^2 \\ &\geq (p-k)(k-k^*)\beta_k^2+(k^*-k)(p-k)\beta_{k+1}^2 \geq 0.\end{aligned} \tag{9.13}$$

最后，根据投影的最优性（投影性质-O；参见引理 2.93），我们有 $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2 \leq \|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|_2$ 对于任意 $\boldsymbol{\beta}^* \in \mathbb{B}_0[k^*]$，结合上述不等式即可得到所需结果。

■

对于更一般的目标函数 f（不一定是最小二乘），假设存在一个最优的 k^*—稀疏参数 $\boldsymbol{\beta}^* =$ 最小值点集合$_{\|\boldsymbol{\beta}\|_0 \leq k^*} f(\boldsymbol{\beta})$。我们的分析将上述投影性质与关于 f 的受限强凸性（RSC）和受限光滑性（RSS）假设相结合，以建立 IHT 的几何收敛性。

定理 9.9：（RSC 和 RSS 条件下 IHT 的收敛速度（Jain 等人，2014年））假设 f 满足受限强凸性（RSC）和受限光滑性（RSS），其参数分别为 $\mu_{2k+k^*}(f) = L_a > 0$ 和$\nu_{2k+k^*}(f) = L_b > 0$（定义 2.86 或定义 4.33）。令算法 2 以 f、$s \triangleq k \geq 32\left(\frac{L_b}{L_a}\right)^2 k^*$ 以及恒定步长 $\eta_t = \frac{2}{3L_b}$ 运行。另令 $\overline{\boldsymbol{\beta}^*}$ =最小值点集合$_{\|\boldsymbol{\beta}\|_0 \leq k^*} f(\boldsymbol{\beta})$。则对于 $\overline{T} = \mathcal{O}\left(\frac{L_b}{L_a}\cdot\ln\left(\frac{f(\boldsymbol{\beta}^{(1)})}{\epsilon}\right)\right)$，算法 2(a) 的第 T 次迭代满足：$f(\boldsymbol{\beta}^{(T)}) - f(\boldsymbol{\beta}^*) \leq \epsilon$.

定理 9.9] 的证明 [记作 $\boldsymbol{g}^{(t)} \triangleq \nabla f(\boldsymbol{\beta}^{(t)})$，回忆一下，$\boldsymbol{\beta}^{(t+1)} = \mathcal{P}_{\mathbb{B}_0[k]}(\boldsymbol{\beta}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}^{(t)})$，其中 $\gamma = \frac{2}{3} < 1$。令 $\mathbb{S}^t \triangleq$supp$(\boldsymbol{\beta}^{(t)})$、$\mathbb{S}^* \triangleq$supp$(\boldsymbol{\beta}^*)$ 和 $\mathbb{S}^{t+1} \triangleq$supp$(\boldsymbol{\beta}^{(t+1)})$。此外，令 $\mathbb{I} \triangleq \mathbb{S}^* \cup \mathbb{S}^t \cup \mathbb{S}^{t+1}$。根据 RSS 性质（定义 2.86），以及 supp$(\boldsymbol{\beta}^{(t)}) \subseteq \mathbb{I}$且 supp$(\boldsymbol{\beta}^{(t+1)}) \subseteq \mathbb{I}$ 这一事实（这意味着 $\boldsymbol{\beta}^{(t+1)}_{\mathbb{I}^c} = \boldsymbol{\beta}^{(t)}_{\mathbb{I}^c} = \mathbf{0}$），再结合对任意向量 $\boldsymbol{u} \in \mathbb{R}^p$，都有 $\|\boldsymbol{u}\|_2^2 = \|\boldsymbol{u}_{\mathbb{I}}\|_2^2 + \|\boldsymbol{u}_{\mathbb{I}^c}\|_2^2$，我们得到：

$$\begin{aligned}f(\boldsymbol{\beta}^{(t+1)}) - f(\boldsymbol{\beta}^{(t)}) &\leq \left\langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)} \right\rangle + \frac{L_b}{2}\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right\|_2^2, \\ &= \frac{L_b}{2}\left\|\boldsymbol{\beta}^{(t+1)}_{\mathbb{I}} - \boldsymbol{\beta}^{(t)}_{\mathbb{I}} + \frac{\gamma}{L_b}\cdot\boldsymbol{g}^{(t)}_{\mathbb{I}}\right\|_2^2 - \frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{I}}\right\|_2^2 + (1-\gamma)\left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}, \boldsymbol{g}^{(t)} \right\rangle.\end{aligned} \tag{9.14}$$

由于 $\mathbb{S}^t \setminus \mathbb{S}^{t+1}$ 和 $\mathbb{S}^{t+1}$ 不相交，因此有：

$$\begin{aligned}\left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}, \boldsymbol{g}^{(t)} \right\rangle &= \cancel{\left\langle \boldsymbol{\beta}^{(t+1)}_{\mathbb{S}^t \setminus \mathbb{S}^{t+1}} - \boldsymbol{\beta}^{(t)}_{\mathbb{S}^t \setminus \mathbb{S}^{t+1}}, \boldsymbol{g}^{(t)}_{\mathbb{S}^t \setminus \mathbb{S}^{t+1}} \right\rangle} + \left\langle \boldsymbol{\beta}^{(t+1)}_{\mathbb{S}^{t+1}} - \boldsymbol{\beta}^{(t)}_{\mathbb{S}^{t+1}}, \boldsymbol{g}^{(t)}_{\mathbb{S}^{t+1}} \right\rangle \\ &\overset{\dagger}{=} -\left\langle \boldsymbol{\beta}^{(t)}_{\mathbb{S}^t \setminus \mathbb{S}^{t+1}}, \boldsymbol{g}^{(t)}_{\mathbb{S}^t \setminus \mathbb{S}^{t+1}} \right\rangle - \frac{\gamma}{L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^{t+1}}\right\|_2^2 \\ &\overset{\ddagger}{\leq} \frac{\gamma}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^{t+1}\setminus\mathbb{S}^t}\right\|_2^2 - \frac{\gamma}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^t\setminus\mathbb{S}^{t+1}}\right\|_2^2 - \frac{\gamma}{L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^{t+1}}\right\|_2^2 - \cancel{\left(\frac{L_b}{2\gamma}\left\|\boldsymbol{\beta}_{\mathbb{S}^t\setminus\mathbb{S}^{t+1}}\right\|_2^2\right)} \\ &\overset{*}{=} -\frac{\gamma}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^{t+1}\setminus\mathbb{S}^t}\right\|_2^2 - \frac{\gamma}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^t\setminus\mathbb{S}^{t+1}}\right\|_2^2 - \frac{\gamma}{L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^t\cap\mathbb{S}^{t+1}}\right\|_2^2 \leq -\frac{\gamma}{2L_b}\left\|\boldsymbol{g}^{(t)}_{\mathbb{S}^t\cup\mathbb{S}^{t+1}}\right\|_2^2,\end{aligned} \tag{9.15}$$

其中，等式（†）由梯度步得到，即 $\boldsymbol{\beta}_{\mathbb{S}^{t+1}}^{(t+1)} = \boldsymbol{\beta}_{\mathbb{S}^{t+1}}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{S}^{t+1}}^{(t)}$；不等式（‡）利用以下事实得到：$\boldsymbol{\beta}^{(t+1)}$ 通过硬阈值化得到（$\mathbb{S}^{t+1}$包含幅值最大的$\left|\mathbb{S}^{t+1}\right| = \left|\mathbb{S}^{t}\right|$的索引），并利用以下事实：$\left|\mathbb{S}^{t} \setminus \mathbb{S}^{t+1}\right| = \left|\mathbb{S}^{t+1} \setminus \mathbb{S}^{t}\right|$，使得 $\left\|\boldsymbol{\beta}_{\mathbb{S}^{t} \setminus \mathbb{S}^{t+1}}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{S}^{t} \setminus \mathbb{S}^{t+1}}^{(t)}\right\|_2^2 \leq \left\|\boldsymbol{\beta}_{\mathbb{S}^{t+1} \setminus \mathbb{S}^{t}}^{(t+1)}\right\|_2^2 = \frac{\gamma^2}{L_b^2}\left\|\boldsymbol{g}_{\mathbb{S}^{t+1} \setminus \mathbb{S}^{t}}^{(t)}\right\|_2^2$；等式（*）由 $\left\|\boldsymbol{g}_{\mathbb{S}^{t+1}}^{(t)}\right\|_2^2 = \left\|\boldsymbol{g}_{\mathbb{S}^{t+1} \setminus \mathbb{S}^{t}}^{(t)}\right\|_2^2 + \left\|\boldsymbol{g}_{\mathbb{S}^{t} \cap \mathbb{S}^{t+1}}^{(t)}\right\|_2^2$得到。结合（9.14）和（9.15）可得

$$
\begin{aligned}
& f(\boldsymbol{\beta}^{(t+1)}) - f(\boldsymbol{\beta}^{(t)}) \\
\leq & \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{I}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2 - \frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2 - \frac{\gamma(1-\gamma)}{2L_b}\left\|\boldsymbol{g}_{\mathbb{S}^{t} \cup \mathbb{S}^{t+1}}^{(t)}\right\|_2^2 \\
= & \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{I}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2 - \frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)}\right\|_2^2 \\
& - \frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{S}^{t} \cup \mathbb{S}^{*}}^{(t)}\right\|_2^2 - \frac{\gamma(1-\gamma)}{2L_b}\left\|\boldsymbol{g}_{\mathbb{S}^{t} \cup \mathbb{S}^{t+1}}^{(t)}\right\|_2^2 .
\end{aligned}
\tag{9.16}
$$

接下来，我们将对上述不等式右侧的前两项给出上界。由于 $\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*}) = \mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*}) \subseteq \mathbb{S}^{t+1}$，我们有 $\boldsymbol{\beta}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t+1)} = \boldsymbol{\beta}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)}$。然而，由于 $\boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{S}^{t}}^{(t)} = \mathbf{0}$，前面的等式可化简为 $\boldsymbol{\beta}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t+1)} = -\frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)}$。令 $\mathbb{T} \subseteq \mathbb{S}^{t} \setminus \mathbb{S}^{t+1}$，使得 $|\mathbb{T}| = \left|\mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})\right|$。这样的选择是可行的，因为$\left|\mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})\right| = \left|\mathbb{S}^{t} \setminus \mathbb{S}^{t+1}\right| - \left|(\mathbb{S}^{t+1} \cap \mathbb{S}^{*}) \setminus \mathbb{S}^{t}\right|$（这本身是 $\left|\mathbb{S}^{t+1}\right| = \left|\mathbb{S}^{t}\right|$ 使得 $\left|\mathbb{S}^{t} \setminus \mathbb{S}^{t+1}\right| = \left|\mathbb{S}^{t+1} \setminus \mathbb{S}^{t}\right|$的结果）。此外，由于 $\boldsymbol{\beta}^{(t+1)}$ 是通过对 $\left(\boldsymbol{\beta}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}^{(t)}\right)$ 进行硬阈值化得到的，对于上文任取的 $\mathbb{T}$，我们有：

$$
\frac{\gamma^2}{L_b^2}\left\|\boldsymbol{g}_{\mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)}\right\|_2^2 = \left\|\boldsymbol{\beta}_{\mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t+1)}\right\|_2^2 \geq \left\|\boldsymbol{\beta}_{\mathbb{T}}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{T}}^{(t)}\right\|_2^2 .
\tag{9.17}
$$

利用上述方程，由于 $\boldsymbol{\beta}_{\mathbb{T}}^{(t+1)} = \mathbf{0}$ (因为 $\mathbb{T} \not\subseteq \mathbb{S}^{t+1}$)，并且由于$\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*}) = \mathbb{S}^{t+1} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})$，等式（9.16）的前两项变为

$$
\begin{aligned}
& \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{I}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2 - \frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I} \setminus (\mathbb{S}^{t} \cup \mathbb{S}^{*})}^{(t)}\right\|_2^2 \\
\leq & \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{I}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2 - \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{T}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{T}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{T}}^{(t)}\right\|_2^2 \\
= & \frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t+1)} - \boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I} \setminus \mathbb{T}}^{(t)}\right\|_2^2 .
\end{aligned}
\tag{9.18}
$$

由于我们构造集合 $\mathbb{T}$ 满足 $\mathbb{T} \subseteq \mathbb{S}^{t} \setminus \mathbb{S}^{t+1}$ 且 $|\mathbb{T}| = \left|\mathbb{S}^{t+1} \setminus \mathbb{S}^{t} \cup \mathbb{S}^{*}\right|$，使其满足 $\left|\mathbb{I} \setminus \mathbb{S}^{t} \cup \mathbb{S}^{*}\right|$ 且 $|\mathbb{I} \setminus \mathbb{T}| \leq \left|\mathbb{S}^{t}\right| + \left|\mathbb{S}^{t} \setminus \mathbb{S}^{t+1} \setminus \mathbb{T}\right| + |\mathbb{S}^{*}| \leq$ （）（），因此可以将 的大小上界设为（）$k + \left|(\mathbb{S}^{t+1} \cap \mathbb{S}^{*}) \setminus \mathbb{S}^{t}\right| + k^{*} \leq k + 2k^{*}$。此外，由于 $\mathbb{S}^{t+1} \subseteq (\mathbb{I} \setminus \mathbb{T})$，我们有 $\boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t+1)} = \mathcal{P}_{\mathbb{B}_0[k]}\left(\boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t)} - \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I} \setminus \mathbb{T}}^{(t)}\right)$。

根据上述观察（9.18），并假设 $k \geq 32\left(\frac{L_b}{L_a}\right)^2 k^{*}$，引理 9.8 在 $\widetilde{\boldsymbol{\beta}} \triangleq \boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t+1)} + \frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I} \setminus \mathbb{T}}^{(t)}$ 和 $\boldsymbol{\beta} \triangleq \boldsymbol{\beta}_{\mathbb{I} \setminus \mathbb{T}}^{(t)}$成立，因此我们将前两个 __

等式项（9.16）化为

$$\begin{aligned}
&\frac{L_b}{2}\left\|\boldsymbol{\beta}_{\mathbb{I}}^{(t+1)}-\boldsymbol{\beta}_{\mathbb{I}}^{(t)}+\frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2-\frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)}^{(t)}\right\|_2^2\leq\frac{L_b}{2}\cdot\frac{|\mathbb{I}\setminus\mathbb{T}|-k}{|\mathbb{I}\setminus\mathbb{T}|-k^*}\left\|\boldsymbol{\beta}_{\mathbb{I}\setminus\mathbb{T}}^*-\boldsymbol{\beta}_{\mathbb{I}\setminus\mathbb{T}}^{(t)}+\frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}\setminus\mathbb{T}}^{(t)}\right\|_2^2\\
&\overset{\dagger}{\leq}\frac{L_b}{2}\cdot\frac{2k^*}{k+k^*}\left\|\boldsymbol{\beta}_{\mathbb{I}}^*-\boldsymbol{\beta}_{\mathbb{I}}^{(t)}+\frac{\gamma}{L_b}\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2=\frac{2k^*}{k+k^*}\cdot\left(\gamma\left\langle\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)},\boldsymbol{g}^{(t)}\right\rangle+\frac{L_b}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2+\frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2\right)\\
&\overset{\ddagger}{\leq}\frac{2k^*}{k+k^*}\cdot\left(\gamma f(\boldsymbol{\beta}^*)-\gamma f(\boldsymbol{\beta}^{(t)})+\frac{L_b-\gamma L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2+\frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2\right),
\end{aligned}\tag{9.19}$$

其中，不等式（†）由 $|\mathbb{I}\setminus\mathbb{T}|\leq k+2k^*$ 如前所示，以及$\frac{z-a}{z-b}$ 在区间 $z\geq a$ （当 $a\geq b\geq 0$时）是一个正的递增函数这一观察可得。注意，由于我们已经有 $\mathbb{S}^{t+1}\subseteq(\mathbb{I}\setminus\mathbb{T})$，因此得到 $|\mathbb{I}\setminus\mathbb{T}|\geq k$。不等式（‡由函数的 RSC 性质（定义2.86）可得。

将（9.19）代入等式（9.16）的前两项，并利用以下事实：$\mathbb{S}^{t+1}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)\subseteq(\mathbb{S}^{t+1}\cup\mathbb{S}^t)$，得到：

$$\begin{aligned}
f(\boldsymbol{\beta}^{(t+1)})-f(\boldsymbol{\beta}^{(t)})\leq&\frac{2k^*}{k+k^*}\cdot\left(\gamma f(\boldsymbol{\beta}^*)-\gamma f(\boldsymbol{\beta}^{(t)})+\frac{L_b-\gamma L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2+\frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2\right)\\
&-\frac{\gamma^2}{2L_b}\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2-\frac{\gamma(1-\gamma)}{2L_b}\left\|\boldsymbol{g}_{\mathbb{S}^{t+1}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)}^{(t)}\right\|_2^2.
\end{aligned}\tag{9.20}$$

由于 $\gamma=2/3$ 且 $k\geq 32\left(\frac{L_b}{L_a}\right)^2k^*$，因此有 $\frac{2k^*}{k+k^*}\leq\frac{L_a^2}{16L_b(L_b-\gamma L_a)}$。由于 $L_b\geq L_a$，我们也有 $\frac{L_a^2}{16L_b(L_b-\gamma L_a)}\leq\frac{3}{16}$。利用这些不等式，对上文（9.19）中的各项重新排列：

$$\begin{aligned}
f(\boldsymbol{\beta}^{(t+1)})-f(\boldsymbol{\beta}^{(t)})\leq&\frac{2k^*}{k+k^*}\cdot\gamma\cdot\left(f(\boldsymbol{\beta}^*)-f(\boldsymbol{\beta}^{(t)})\right)+\frac{L_a^2}{32L_b}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2+\frac{1}{24L_b}\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2\\
&-\frac{2}{9L_b}\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2-\frac{1}{9L_b}\left\|\boldsymbol{g}_{\mathbb{S}^{t+1}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)}^{(t)}\right\|_2^2.
\end{aligned}\tag{9.21}$$

分解 $\left\|\boldsymbol{g}_{\mathbb{I}}^{(t)}\right\|_2^2=\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2+\left\|\boldsymbol{g}_{\mathbb{S}^{t+1}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)}^{(t)}\right\|_2^2$ 可得

$$\begin{aligned}
f(\boldsymbol{\beta}^{(t+1)})-f(\boldsymbol{\beta}^{(t)})\leq&\frac{2k^*}{k+k^*}\cdot\gamma\cdot\left(f(\boldsymbol{\beta}^*)-f(\boldsymbol{\beta}^{(t)})\right)-\frac{1}{2L_b}\left(\frac{13}{36}\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2-\frac{L_a^2}{16}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2\right)\\
&-\frac{1}{2L_b}\cdot\left(\frac{4}{9}-\frac{1}{12}\right)\left\|\boldsymbol{g}_{\mathbb{S}^{t+1}\setminus(\mathbb{S}^t\cup\mathbb{S}^*)}^{(t)}\right\|_2^2\\
\leq&\frac{2k^*}{k+k^*}\cdot\gamma\cdot\left(f(\boldsymbol{\beta}^*)-f(\boldsymbol{\beta}^{(t)})\right)-\frac{13}{72L_b}\left(\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2-\frac{L_a^2}{4}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2\right).
\end{aligned}\tag{9.22}$$

利用 RSC 性质，我们有：

$$\begin{aligned}
f(\boldsymbol{\beta}^{(t)})-f(\boldsymbol{\beta}^*)\leq&\left\langle\boldsymbol{g}^{(t)},\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\rangle-\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2=\left\langle\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)},\boldsymbol{\beta}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}-\boldsymbol{\beta}_{\mathbb{S}^t\cup\mathbb{S}^*}^*\right\rangle-\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2\\
\leq&\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\|_2-\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2.
\end{aligned}$$

这表明

$$\begin{aligned}
\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2^2-\frac{L_a^2}{4}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2^2=&\left(\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2-\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2\right)\left(\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2+\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2\right)\\
\geq&\frac{(f(\boldsymbol{\beta}^{(t)})-f(\boldsymbol{\beta}^*))}{\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^*\right\|_2}\cdot\left(\left\|\boldsymbol{g}_{\mathbb{S}^t\cup\mathbb{S}^*}^{(t)}\right\|_2+\frac{L_a}{2}\left\|\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)}\right\|_2\right)\geq\frac{L_a}{2}\cdot(f(\boldsymbol{\beta}^{(t)})-f(\boldsymbol{\beta}^*)).
\end{aligned}$$

将这个不等式代入（9.22）可得

$$\begin{aligned} f(\boldsymbol{\beta}^{(t+1)}) - f(\boldsymbol{\beta}^{(t)}) &\leq \frac{2k^*}{k+k^*} \cdot \gamma \cdot \left(f(\boldsymbol{\beta}^*) - f(\boldsymbol{\beta}^{(t)})\right) - \frac{L_a}{12L_b}\left(f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}^*)\right) \\ &= -\left(\frac{2k^*}{k+k^*} \cdot \gamma + \frac{L_a}{12L_b}\right)\left(f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}^*)\right). \end{aligned} \tag{9.23}$$

观察到 $\frac{2k^*}{k+k^*} \geq 0$，即可得出该结果。——■

9.2. 约束 LASSO 问题的算法

与迭代硬阈值方法不同，约束 LASSO（ℓ_0-约束）的原始形式并未施加 L_C）（第 126 页），而是定义如下：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} f(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \quad \text{s.t.} \quad \{\|\boldsymbol{\beta}\|_1 \leq \Sigma\} \equiv \{\boldsymbol{\beta} \in \mathbb{B}_1[\Sigma]\}, \tag{9.24}$$

其中，$\mathbb{B}_1[\Sigma] \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^p \mid \|\boldsymbol{\alpha}\|_1 \leq \Sigma\}$ 表示半径为 Σ 的闭 ℓ_1-球。由于 $\mathbb{B}_1[\Sigma]$是凸的，求解该问题的一种自然方法是投影梯度下降（*PGD*），该方法交替执行梯度下降步和向可行集$\mathbb{B}_1[\Sigma]$的投影。

算法3 投影（次）梯度下降（PGD）方法

需要： 一个函数 $f(\boldsymbol{\beta})$ 和一个集合 $\mathbb{S}$;

initialize: $\boldsymbol{\beta}^{(1)}$;
for $t = 1, 2, \ldots$ **do**
  Choose a stepsize η_t;
  $\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)})$ or $\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t f'(\boldsymbol{\beta}^{(t)})$, where $f'(\boldsymbol{\beta}^{(t)}) \in \partial f(\boldsymbol{\beta}^{(t)})$;
  $\boldsymbol{\beta}^{(t+1)} \in \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}^{(t+1)})$;
  Stop if a stopping criterion is satisfied at iteration $t = T$;
end或
(Output Option 1) Output $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$;
(Output Option 2) Output $\boldsymbol{\beta}_{\text{avg}} \leftarrow \frac{1}{T}(\sum_{t=1}^{t} \boldsymbol{\beta}^{(t)})$ or $\sum_{t=1}^{T} \frac{2t}{T(T+1)}\boldsymbol{\beta}^{(t)}$;
（输出选项 3）输出 $\boldsymbol{\beta}_{\text{best}}$最小值点集合$_{t\in\{1,2,\ldots,T\}} f(\boldsymbol{\beta}^{t})$;

9.2.1 投影梯度下降（PGD）

标准梯度下降是为无约束优化问题设计的。为处理约束，投影梯度下降（PGD）方法——如算法 3所述——解决以下形式的问题：

$$(\mathrm{C1}): \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{S}. \tag{9.25}$$

节 2.9.1 讨论此类约束问题的最优性条件。PGD 是求解 (C1) 最直接的一阶方法，也是约束优化的基础构件。许多高级算法（包括近端梯度法）都是 PGD 的自然扩展；有关进一步讨论，请参见 Lu (2025年) 第4章。

当 $\mathbb{S}$ 是闭且凸的，且步长 $\eta_t \equiv \eta$ 为常数时，PGD 更新的一个不动点 $\boldsymbol{\beta}^*$ 满足驻点条件：

$$\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}^* - \eta \nabla f(\boldsymbol{\beta}^*))$$

正如推论 2.113 中所述，由于 PGD 与近端梯度法非常相似，我们将在接下来的几节中借助近端框架分析其收敛性。

More 具体而言，从迭代 t 到 $t+1$ 的 PGD 更新可以等价地写为 s:

$$\begin{aligned}(\mathbf{PGD}): \quad \boldsymbol{\beta}^{(t+1)} &\leftarrow \mathcal{P}_{\mathbb{S}}(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)})) \\ &= \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\| \boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)})\right)\right\|_2^2 \\ &= \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\{ \frac{1}{2\eta_t} \left\| \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\|_2^2 + f(\boldsymbol{\beta}^{(t)}) + \left\langle \nabla f(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle \right\}.\end{aligned} \tag{9.26}$$

在第三行中，常数项 $f(\boldsymbol{\beta}^{(t)})$ 不影响极小化解，仅为便于解释而加入。这种表述揭示了，每一步 PGD 都是在最小化目标函数 f 的局部线性模型，并通过二次邻近项进行正则化，以惩罚偏离当前迭代点 $\boldsymbol{\beta}^{(t)}$ 的较大偏差。

线性项推动迭代朝着使 f 递减的方向前进，而二次项则通过使下一迭代点接近当前迭代点来确保稳定性。这种分解为设计更复杂的一阶方法提供了关键直觉，例如近端梯度法、条件梯度法（Frank–Wolfe）和镜像下降法；这些方法都在局部近似与针对问题几何结构设计的正则化之间寻求平衡。

投影到 ℓ_1-球

要将 PGD 应用于约束 LASSO（L_C），我们必须在每次迭代时求解以下子问题：给定向量 $\boldsymbol{v} \in \mathbb{R}^p$，计算其到半径为 ℓ_1 的 Σ -球的欧几里得投影，即求解：

$$\mathcal{P}_{\mathbb{B}_1[\Sigma]}(\boldsymbol{v}) \equiv \left\{ \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2} \|\boldsymbol{\beta} - \boldsymbol{v}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_1 \le \Sigma \right\}.$$

这是一个凸优化问题，具有唯一解，且可高效计算（Duchietal.，2008年）。

投影 $\boldsymbol{\beta}^* = \mathcal{P}_{\mathbb{B}_1[\Sigma]}(\boldsymbol{v})$ 通过 软阈值处理 （见例 2.104）得到，具体如下：

(i) 若 $\|\boldsymbol{v}\|_1 \le \Sigma$，则 $\boldsymbol{v}$ 已经位于 ℓ_1-球内，因此 $\boldsymbol{\beta}^* = \boldsymbol{v}$。(ii) 否则（即 $\|\boldsymbol{v}\|_1 > \Sigma$），投影为

$$\beta_i^* = \operatorname{sign}(v_i) \cdot [|v_i| - \lambda]_+, \quad \forall i \in \{1, 2, \ldots, p\},$$

其中 $[\cdot]_+ \triangleq \max(\cdot, 0)$，且 $\lambda \ge 0$ 的选取满足：

$$\sum_{i=1}^{p} [|v_i| - \lambda]_+ = \Sigma.$$

换句话说， λ 是唯一一个非负标量，它能确保软阈值化向量的 ℓ_1-范数恰好等于 Σ。直观而言，我们对 $\boldsymbol{v}$ 的所有分量幅值进行均匀收缩，直到总 ℓ_1-范数达到约束边界 Σ。

等价地，我们要求满足 $\|\mathcal{T}_\lambda(\boldsymbol{v})\|_1 = \Sigma$ 的 λ，其中软阈值算子$\mathcal{T}_\lambda(\boldsymbol{v})$按分量作用如下：$\big(\mathcal{T}_\lambda(\boldsymbol{v})\big)_i =$ 符号函数$(v_i)[|v_i| - \lambda]_+$。这可归结为求解标量方程：$\sum_{i=1}^{p}[|v_i| - \lambda]_+ = \Sigma$。令 $b_i \triangleq |v_i|$，并按降序排列：$b_{(1)} \geq b_{(2)} \geq \ldots \geq b_{(p)}$。计算 λ 的高效算法如下：

1. 对所有 i 计算 $b_i = |v_i|$。2. 按降序排列 $\boldsymbol{b}$：$b_{(1)} \geq b_{(2)} \geq \ldots \geq b_{(p)}$。3. 计算累积和 $S_t \triangleq \sum_{i=1}^{t} b_{(i)}$。4. 找到满足 $b_{(t)} > \frac{S_t - \Sigma}{t}$ 的最大 $t \in \{1, 2, \ldots, p\}$。这确保了 $\lambda < b_{(t)}$。5. 设

$$\lambda \triangleq \frac{S_t - \Sigma}{t}. \tag{9.27}$$

6. 应用软阈值处理：

$$\beta_i^* = \operatorname{sign}(v_i) \cdot [|v_i| - \lambda]_+. \tag{9.28}$$

例9.10。 设 $\boldsymbol{v} = [3, -2, 1]$, $\Sigma = 3$。由于 $\|\boldsymbol{v}\|_1 = 6 > 3$，因此需要进行投影。绝对值为 $\boldsymbol{b} = [3, 2, 1]$，且已按降序排列。

- 尝试 $t = 1$： $S_1 = 3$，因此 $\lambda = (3-3)/1 = 0$，且 $b_{(1)} = 3 > 0$，所以 $t = 1$ 有效。尝试 $t = 2$:
- $S_2 = 3 + 2 = 5$，因此 $\lambda = (5-3)/2 = 1$。检查 $b_{(2)} = 2 > 1$，所以 $t = 2$ 有效。尝试 $t = 3$:
- $S_3 = 3 + 2 + 1 = 6$，因此 $\lambda = (6-3)/3 = 1$。检查检查 $b_{(3)} = 1 \not> 1$，所以$t = 3$ 无效。

因此， $\lambda = 1$ 我们进行投影：

$$\begin{aligned}
\beta_1^* &= \operatorname{sign}(3)(3-1) = 2; \\
\beta_2^* &= \operatorname{sign}(-2)(2-1) = -1; \\
\beta_3^* &= \operatorname{sign}(1)(1-1) = 0.
\end{aligned}$$

投影结果为 $\boldsymbol{\beta}^* = [2, -1, 0]$，其中 $\|\boldsymbol{\beta}^*\|_1 = 3 = \Sigma$。 □

使用 PGD 求解约束 LASSO

将算法 3应用于约束集 $\mathbb{S} \triangleq \mathbb{B}_1[\Sigma]$和最小二乘目标函数$f(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$，可得到一种实用的数值方法来求解约束 LASSO 问题。只要我们能够高效地投影到 ℓ_1-球上，投影梯度下降的行为就类似于将标准梯度下降应用于无约束最小二乘问题。特别是，当格拉姆矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$ 良态时，投影梯度下降具有线性收敛速率；条件数较小时，收敛速度很快。

为了正式证明这一收敛保证，我们将在下一小节介绍近端梯度法，并阐明它与投影梯度下降之间的关系。

算法 4 近端梯度法

要求： 函数 $f(\boldsymbol{\beta})$ 和闭凸函数 g （通常非光滑）满足定理中的 (A1) 和 (A2) 9.12；

初始化： $\boldsymbol{\beta}^{(1)}$；

对于 $t = 1, 2, \cdots$ **执行**

  选择一个步长 η_t；

  $\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)})$； ▷ (PROX_1)

  $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t+1)}) \triangleq \mathcal{T}^{f,\ g}_{L_t}(\boldsymbol{\beta}^{(t)})$； ▷ (PROX_2)

  如果在第 $t = T$ 次迭代时满足停止准则，则停止；

结束循环

（输出选项 1）输出 $\boldsymbol{\beta}_{\mathrm{final}} \leftarrow \boldsymbol{\beta}^{(T)}$；（输出选项 2）输出 $\boldsymbol{\beta}_{\mathrm{avg}} \leftarrow \frac{1}{T}(\sum_{t=1}^{t} \boldsymbol{\beta}^{(t)})$ 或 $\sum_{t=1}^{T} {}_{T}{}^{2t}{}_{(T+1)}\boldsymbol{\beta}^{(t)}$；（输出选项 3）输出 $\boldsymbol{\beta}_{\mathrm{best}} \leftarrow$ 最小值点集合 $_{t\in\{1,2,\ldots,T\}} f(\boldsymbol{\beta}^{(t)})$；

9.2.2 近端梯度法

▶ **近端梯度法**。在第2.8节中，我们强调了投影与近端算子之间的密切关系。具体来说，回顾以下恒等式

$$\mathcal{P}\mathrm{rox}_{\delta_{\mathbb{S}}}(\boldsymbol{\alpha}) = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}),$$

其中，$\delta_{\mathbb{S}}$ 表示集合 $\mathbb{S}$的示性函数。这使我们可以将投影梯度下降中的投影步重新解释为近端操作：

$$\left\{\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}^{(t+1)})\right\} \quad \rightsquigarrow \quad \left\{\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\delta_{\mathbb{S}}}(\boldsymbol{\alpha}^{(t+1)})\right\}.$$

近端梯度法（算法4）通过用任意闭、凸且可能非光滑的函数 g替代示性函数，对这一思想进行了推广：

$$\left\{\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\delta_{\mathbb{S}}}(\boldsymbol{\alpha}^{(t+1)})\right\} \quad \rightsquigarrow \quad \left\{\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t+1)})\right\}, \tag{9.29}$$

其中，$\eta_t > 0$ 是第 t 次迭代的步长（其作用很快就会明了）。我们假设 g 闭且凸，这确保近端算子满足标准性质（近端性质 O、I、II、III、IV；参见引理2。99∼2.102）。当 $g \equiv \delta_{\mathbb{S}}$时，近端梯度法恰好化为投影梯度下降。

回顾一下，在 PGD 中，从第 t 次迭代到第 $t+1$ 次迭代的更新可以写为：

$$\textbf{(PGD)}: \quad \begin{aligned} \boldsymbol{\beta}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\|\boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)})\right)\right\|_2^2 \\ &= \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\{\frac{1}{2\eta_t}\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}\right\|_2^2 + f(\boldsymbol{\beta}^{(t)}) + \left\langle \nabla f(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}\right\rangle\right\}. \end{aligned} \tag{9.30}$$

这里，常数项 $f(\boldsymbol{\beta}^{(t)})$ 不会影响极小化解，仅为便于解释而引入。因此，PGD 更新最小化了光滑函数 f 在 $\boldsymbol{\beta}^{(t)}$ 附近的线性近似，并通过二次邻近项进行正则化。线性项推动迭代点向目标函数值更低的方向前进，而二次项使新的迭代点保持接近当前迭代点，从而确保稳定性。

近端梯度法通过将非光滑惩罚项 g 直接纳入目标函数，将这一框架扩展到无约束复合优化：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{2\eta_t} \left\| \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\|_2^2 + f(\boldsymbol{\beta}^{(t)}) + \left\langle \nabla f(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle + g(\boldsymbol{\beta}) \right\} \tag{9.31a}$$

（近端算子）：
$$= \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \eta_t g(\boldsymbol{\beta}) + \frac{1}{2} \left\| \boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)}) \right) \right\|_2^2 \tag{9.31b}$$

$$= \mathcal{P}\mathrm{rox}_{\eta_t g} \left(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)}) \right) \triangleq \mathcal{T}_{L_t}^{f,g}(\boldsymbol{\beta}^{(t)}), \tag{9.31c}$$

其中 $L_t \triangleq \frac{1}{\eta_t}$，且 $\mathcal{T}^{f,\ g}_{L}(\boldsymbol{\beta}) \triangleq \mathcal{P}\mathrm{rox}_{\frac{1}{L}g}\left(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta})\right)$ 表示近端梯度算子。我们还将梯度映射 定义为

$$\mathcal{G}_L^{f,g}(\boldsymbol{\beta}) \triangleq L\big(\boldsymbol{\beta} - \mathcal{T}_L^{f,g}(\boldsymbol{\beta})\big). \tag{9.31d}$$

只要近端算子 $\mathcal{P}\mathrm{rox}_{\eta_t g}$具有闭式解或能够快速计算（例如，对于 ℓ_1-正则化，可通过软阈值处理实现），更新（9.31c）在计算上就是高效的。这一形式（9.31c）特别适用于采用惩罚目标函数的统计学习问题——例如拉格朗日函数（无约束）LASSO——而不是具有如下形式的硬约束 $g(\boldsymbol{\beta}) \leq \Sigma$ （详见第9.3.2节）。

此外，更新（9.31c）可以改写为

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g} \left(\boldsymbol{\beta}^{(t)} - \eta_t \nabla f(\boldsymbol{\beta}^{(t)}) \right) \triangleq \mathcal{T}_{L_t}^{f,g}(\boldsymbol{\beta}^{(t)}) \triangleq \boldsymbol{\beta}^{(t)} - \frac{1}{L_t}\mathcal{G}_{L_t}^{f,g}(\boldsymbol{\beta}^{(t)}), \tag{9.31e}$$

这与标准梯度下降（GD）步骤非常相似。因此，$\mathcal{G}^{f,\ g}_{L_t}(\boldsymbol{\beta}^{(t)})$ 被称为梯度映射。当函数 f 和 g 在上下文中含义明确时，我们简化记号，使用 $\mathcal{T}_{L_t}(\cdot)$ 和 $\mathcal{G}_{L_t}(\cdot)$，而不再使用 $\mathcal{T}^{f,\ g}_{L_t}(\cdot)$ 和 $\mathcal{G}^{f,\ g}_{L_t}(\cdot)$。

方程（9.31）证明了在 η_t （9.29）中进行缩放的合理性。总之，近端梯度法旨在最小化复合目标函数

$$\min\{F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})\}, \tag{9.32}$$

其中，f 是光滑的，而 g 是闭、凸且可能非光滑的。

在给出收敛结果之前，我们先陈述近端梯度法的以下下降引理。

引理 9.11：（近端梯度法的下降引理）设 f 是一个真闭且 L_b-光滑的函数，g 是一个真闭凸函数。假设以下条件成立：

(A1) g： $\mathbb{R}^p \to (-\infty, \infty]$ 是适当、闭且凸的。(A2) f： $\mathbb{R}^p \to (-\infty, \infty]$ 是适当、闭且凸的，$\mathrm{dom}(f)$ 是凸集，$\mathrm{dom}(g)$ ⊆包含于$(\mathrm{dom}(f))$ 的内部，且 f 在 L_b-光滑于内部$(\mathrm{dom}(f)$ 上。

考虑复合函数 $F \triangleq f + g$ 和近端梯度算子 $\mathcal{T}_L \triangleq \mathcal{T}^{f,\ g}_{L}$。那么，对于任意 $\boldsymbol{\beta} \in$ int$(\mathrm{dom}(f))$和 $L \in \left(\frac{L_b}{2}, \infty\right)$，以下不等式成立：

$$F(\mathcal{T}_L(\boldsymbol{\beta})) - F(\boldsymbol{\beta}) \leq \frac{\frac{L_b}{2} - L}{L^2} \left\| \mathcal{G}_L^{f,g}(\boldsymbol{\beta}) \right\|_2^2 \leq 0, \tag{9.33}$$

其中 $\mathcal{G}^{f,\ g}_{L}: \mathrm{int}(\mathrm{dom}(f)) \to \mathbb{R}^p$ 是由 $\mathcal{G}^{f,\ g}_{L}(\boldsymbol{\beta}) \triangleq L\big(\boldsymbol{\beta} - \mathcal{T}_L(\boldsymbol{\beta})\big)$ 对所有$\boldsymbol{\beta} \in \mathrm{int}(\mathrm{dom}(f))$ 定义的算子。

此外，对于任意 $\boldsymbol{\alpha} \in \mathbb{R}^p$, $\boldsymbol{\beta} \in \operatorname{int}(\operatorname{dom}(f))$，考虑近端间隙 $\mathcal{D}_f(\boldsymbol{\alpha}, \boldsymbol{\beta}) \triangleq f(\boldsymbol{\alpha}) - f(\boldsymbol{\beta}) - \langle \nabla f(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle$ [a] 以及 $L \triangleq L_b$。则有

$$F(\mathcal{T}_L(\boldsymbol{\beta})) - F(\boldsymbol{\alpha}) \leq \frac{L}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 - \frac{L}{2} \|\boldsymbol{\alpha} - \mathcal{T}_L(\boldsymbol{\beta})\|_2^2 - \mathcal{D}_f(\boldsymbol{\alpha}, \boldsymbol{\beta}).^{b} \tag{9.34}$$

当 $\boldsymbol{\alpha} = \boldsymbol{\beta}$ 时，(9.34) 通过设定 $L \triangleq L_b$，可化为 (9.33)。

a。当 f 为凸函数时，这也称为 Bregman 距离（定义 2.87）。b。我们可以将 $\boldsymbol{\alpha}$ 设为最优点 $\boldsymbol{\alpha} \triangleq \boldsymbol{\beta}^*$，以获得收敛结果。

引理 [的证明 9.11] **方程** (9.33)。**为简便起见，记** $\boldsymbol{\beta}^+ \triangleq \mathcal{T}_L(\boldsymbol{\beta})$。根据 f 的光滑性（定义 2.78），

$$f(\boldsymbol{\beta}^+) \leq f(\boldsymbol{\beta}) + \left\langle \nabla f(\boldsymbol{\beta}), \boldsymbol{\beta}^+ - \boldsymbol{\beta} \right\rangle + \frac{L_b}{2} \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^+\right\|_2^2.$$

根据近端 性质-I（引理 2.100），由于 $\boldsymbol{\beta}^+ = \mathcal{P}\mathrm{rox}_{\frac{1}{L}g}(\boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta}))$，我们有 e

$$\begin{aligned} &\left\langle \boldsymbol{\beta} - \frac{1}{L}\nabla f(\boldsymbol{\beta}) - \boldsymbol{\beta}^+, \boldsymbol{\beta} - \boldsymbol{\beta}^+ \right\rangle \leq \frac{1}{L} g(\boldsymbol{\beta}) - \frac{1}{L} g(\boldsymbol{\beta}^+), \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p \\ \Longrightarrow \quad & g(\boldsymbol{\beta}^+) \leq g(\boldsymbol{\beta}) - L \left\|\boldsymbol{\beta}^+ - \boldsymbol{\beta}\right\|_2^2 - \left\langle \nabla f(\boldsymbol{\beta}), \boldsymbol{\beta}^+ - \boldsymbol{\beta} \right\rangle. \end{aligned}$$

结合前面的两个不等式，我们得到

$$f(\boldsymbol{\beta}^+) + g(\boldsymbol{\beta}^+) \leq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta}) + \left(-L + \frac{L_b}{2}\right) \left\|\boldsymbol{\beta}^+ - \boldsymbol{\beta}\right\|_2^2.$$

因此，考虑到 $\boldsymbol{\beta}^+$、$\mathcal{G}^{f,\ g}_{\ L}(\boldsymbol{\beta})$ 的定义，以及恒等式 $F(\boldsymbol{\beta}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$、$F(\boldsymbol{\beta}^+) = f(\boldsymbol{\beta}^+) + g(\boldsymbol{\beta}^+)$，所需结果即得。

方程 (9.34)。定义辅助函数

$$\varphi(\boldsymbol{u}) \triangleq f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta}), \boldsymbol{u} - \boldsymbol{\beta} \rangle + g(\boldsymbol{u}) + \frac{L}{2} \|\boldsymbol{u} - \boldsymbol{\beta}\|_2^2.$$

这正是近端梯度更新（9.31）中极小化的第一个目标函数。由于 φ 是一个 L-强凸函数，且 $\mathcal{T}_L(\boldsymbol{\beta})$ = 极小化算子$_{\boldsymbol{u} \in \mathbb{R}^p} \varphi(\boldsymbol{u})$，由定理2.82(ii)可得

$$\varphi(\boldsymbol{\alpha}) - \varphi(\mathcal{T}_L(\boldsymbol{\beta})) \geq \frac{L}{2} \|\boldsymbol{\alpha} - \mathcal{T}_L(\boldsymbol{\beta})\|_2^2, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{R}^p. \tag{9.35}$$

由此可知，根据 f 的光滑性，

$$\begin{aligned} \varphi(\mathcal{T}_L(\boldsymbol{\beta})) &= f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta}), \mathcal{T}_L(\boldsymbol{\beta}) - \boldsymbol{\beta} \rangle + \frac{L}{2} \|\mathcal{T}_L(\boldsymbol{\beta}) - \boldsymbol{\beta}\|_2^2 + g(\mathcal{T}_L(\boldsymbol{\beta})) \\ &\geq f(\mathcal{T}_L(\boldsymbol{\beta})) + g(\mathcal{T}_L(\boldsymbol{\beta})) = F(\mathcal{T}_L(\boldsymbol{\beta})). \end{aligned} \tag{9.36}$$

因此，结合（9.35）和（9.36），可得

$$\underbrace{f(\boldsymbol{\beta}) + \langle \nabla f(\boldsymbol{\beta}), \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle + g(\boldsymbol{\alpha}) + \frac{L}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2}_{=\varphi(\boldsymbol{\alpha})} - F(\mathcal{T}_L(\boldsymbol{\beta})) \geq \frac{L}{2} \|\boldsymbol{\alpha} - \mathcal{T}_L(\boldsymbol{\beta})\|_2^2, \quad \forall \boldsymbol{\alpha} \in \mathbb{R}^p.$$

这就证明了所需的结论。 ■

当 f 为凸且强光滑时，近端梯度法具有如下收敛保证。

定理 9.12：（凸函数和强光滑函数的近端梯度收敛 f： $\mathcal{O}(1/T)$ （Lu，2025年）设 f 为真闭且 L_b光滑函数，并设 g 为真闭凸函数。假设以下条件成立：

(A1) g：$\mathbb{R}^p \to (-\infty, \infty]$ 是适当的、闭的且凸的。
(A2) f： $\mathbb{R}^p \to (-\infty, \infty]$ 是适当的、闭的且凸的，$\mathrm{dom}(f)$ 是凸的，$\mathrm{dom}(g) \subseteq$的内部$(\mathrm{dom}(f))$，并且 f 在 $L_b$int（$\mathrm{dom}(f)$）上是 L_b-光滑的。
设 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 为采用恒定步长规则应用于复合问题 $\min\{F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$的近端梯度法（算法 4）生成的序列，其中 $L_t \triangleq L_b$ 对所有 $t > 0$均成立。那么，对于任意最优解 $\boldsymbol{\beta}^*$ 以及 $F \triangleq f + g$ 的任意整数 $T > 1$，以下上界成立：

$$F(\boldsymbol{\beta}^{(T)}) - F(\boldsymbol{\beta}^*) \leq \frac{L_b \left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\right\|_2^2}{2(T-1)}.$$

因此，对于任意容差 $\epsilon > 0$，有

$$F(\boldsymbol{\beta}^{(T)}) \leq F(\boldsymbol{\beta}^*) + \epsilon, \quad \text{for all } T \geq \frac{L_b \left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\right\|_2^2}{2\epsilon} + 1.$$

定理 [的证明 9.12]由于 f 是凸的，近端间隙 $\mathcal{D}_f$ 构成一个 Bregman 距离，使得 $\mathcal{D}_f(\boldsymbol{\beta}^*, \boldsymbol{\beta}^{(t)}) \geq 0$（注释 2.88）。对于任意 $t > 0$，利用式（9.34）中的下降引理，可得

$$\begin{aligned}\frac{2}{L_b}\left(F(\boldsymbol{\beta}^{(t+1)}) - F(\boldsymbol{\beta}^*)\right) &\leq \left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)}\right\|_2^2 - \left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t+1)}\right\|_2^2 - \frac{2}{L_b}\mathcal{D}_f(\boldsymbol{\beta}^*, \boldsymbol{\beta}^{(t)}) \\ &\leq \left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)}\right\|_2^2 - \left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t+1)}\right\|_2^2.\end{aligned}$$

S对上述不等式关于 $t = \{1, 2, \ldots, T-1\}$ 求和，得到一个望远镜求和 ：

$$\sum_{t=1}^{T-1}\left(F(\boldsymbol{\beta}^{(t+1)}) - F(\boldsymbol{\beta}^*)\right) \leq \frac{L_b}{2}\left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}\right\|_2^2 - \frac{L_b}{2}\left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(T)}\right\|_2^2 \leq \frac{L_b}{2}\left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}\right\|_2^2.$$

根据 $\{F(\boldsymbol{\beta}^{(t)})\}_{t>0}$ (方程 (9.33) 和引理 9.11) 的单调性，我们可以得出

$$(T-1)\left(F(\boldsymbol{\beta}^{(T)}) - F(\boldsymbol{\beta}^*)\right) \leq \sum_{t=1}^{T-1}\left(F(\boldsymbol{\beta}^{(t+1)}) - F(\boldsymbol{\beta}^*)\right) \leq \frac{L_b}{2}\left\|\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}\right\|_2^2,$$

由此可得所需结果。 ■

▶ **带近端算子的约束 LASSO**。 利用示性函数，可以将 (9.24) 中的约束 LASSO 问题等价地改写为以下复合优化问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \delta_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta}). \tag{9.37}$$

这里， $\delta_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta})$ 表示缩放后的 ℓ_1-球 $\mathbb{B}_1[\Sigma]$的示性函数。 $F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$ 和 $g(\boldsymbol{\beta}) \triangleq \delta_{\mathbb{B}_1[\Sigma]}(\boldsymbol{\beta})$ 是示性函数。由于该集合是闭且凸的，示性函数 g 是适当、闭且凸的（参见习题2.53 和 2.36）。

将近端梯度法（算法4）应用于此表述，可得到第 t次迭代的以下更新规则：

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}); \tag{9.38a}$$

$$\lambda^{(t+1)} \leftarrow \text{Using the projection onto } \ell_1\text{-ball in (9.27)}; \tag{9.38b}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t+1)}) = \mathcal{T}_{\eta_t \lambda^{(t+1)}}(\boldsymbol{\alpha}^{(t+1)}) = \left[\left|\boldsymbol{\alpha}^{(t+1)}\right| - \eta_t \lambda^{(t+1)} \mathbf{1}\right]_+ \circ \text{sign}(\boldsymbol{\alpha}^{(t+1)}), \tag{9.38c}$$

其中 $\mathcal{T}_\lambda(\cdot)$ 表示软阈值算子（例 2.104）。这等价于约束 LASSO 问题的 PGD 更新。换言之，该算法首先对光滑损失 f执行梯度下降步，然后应用近端（收缩）步骤，通过投影到可行集来满足 ℓ_1-范数约束。这种收缩能促进迭代点的稀疏性，因此近端梯度法特别适用于约束 LASSO 等稀疏恢复问题。

定理 9.12 保证了该方案的收敛性。具体而言，由于 f 是 L_b-光滑的，且 $L_b = \left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2$（例 2.72），恒定步长 $\eta_t = 1/L_b = \frac{1}{\left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2}$可确保达到定理中确立的 $\mathcal{O}(1/T)$ 收敛速率。

算法 5 条件梯度法

Require: A function $f(\boldsymbol{\beta})$ and a set $\mathbb{S}$;
initialize: $\boldsymbol{\beta}^{(0)}$;
for $t = 0, 1, 2, \ldots$ **do**
  Choose a stepsize $\eta_t \in [0, 1]$;
  $\widetilde{\boldsymbol{\beta}}^{(t)} \leftarrow \arg\min_{\boldsymbol{\beta} \in \mathbb{S}} \left\langle \nabla f(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} \right\rangle$; ▷ 线性优化
  $\boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} + \eta_t(\widetilde{\boldsymbol{\beta}}^{(t)} - \boldsymbol{\beta}^{(t)}) = (1 - \eta_t)\boldsymbol{\beta}^{(t)} + \eta_t \widetilde{\boldsymbol{\beta}}^{(t)}$; ▷ “凸”更新
  Stop if a stopping criterion is satisfied at iteration $t = T$;
end for
return $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$;

9.2.3 条件梯度（Frank–Wolfe）法

▶ **条件梯度法（CG，Frank–Wolfe）**我们来研究条件梯度法，它通常也称为*Frank–Wolfe*（*FW*）算法。这种方法为求解约束优化问题提供了一种颇具吸引力的替代方案，尤其是因为它避免了 PGD 方法所需的、可能代价高昂的投影步。再来考虑以下约束优化问题：

$$\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{\beta} \in \mathbb{S}. \tag{9.39}$$

在条件梯度法中，下一个迭代点由当前迭代点与目标函数在线性近似下于可行集上的一个极小化解的凸组合计算得到 $\mathbb{S}$ （参见算法 5）。

回顾一下，在投影梯度下降方法中，从第 t 次迭代到第 $(t+1)$ 次迭代的更新由下式给出：

$$\text{（投影梯度下降）：}\quad \begin{aligned}\boldsymbol{\beta}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\|\boldsymbol{\beta}-\left(\boldsymbol{\beta}^{(t)}-\eta_t\nabla f(\boldsymbol{\beta}^{(t)})\right)\right\|_2^2 \\ &= \arg\min_{\boldsymbol{\beta}\in\mathbb{S}} \left\{\frac{1}{2}\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}\|_2^2+\eta_t\left\langle\nabla f(\boldsymbol{\beta}^{(t)}),\boldsymbol{\beta}\right\rangle\right\},\end{aligned} \tag{9.40}$$

其中，二次项充当正则化项，确保新的迭代点保持接近当前迭代点 $\boldsymbol{\beta}^{(t)}$。

相比之下，条件梯度法虽然与 PGD 更新密切相关，但它将下一迭代点计算为线性子问题最小值点集合 $\arg\min_{\boldsymbol{\beta}\in\mathbb{S}}\langle\nabla f(\boldsymbol{\beta}^{(t)}),\boldsymbol{\beta}\rangle$ 的一个解与当前迭代点 $\boldsymbol{\beta}^{(t)}$的凸组合。该方法完全避免了二次正则化。不过，凸组合本身起到了隐式正则化的作用，防止偏离当前迭代点过大：

$$\textbf{(CG)}:\quad \boldsymbol{\beta}^{(t+1)} \leftarrow (1-\eta_t)\boldsymbol{\beta}^{(t)}+\eta_t\arg\min_{\boldsymbol{\beta}\in\mathbb{S}}\left\langle\nabla f(\boldsymbol{\beta}^{(t)}),\boldsymbol{\beta}\right\rangle. \tag{9.41}$$

与 PGD 和近端梯度法类似，条件梯度法对于凸且光滑的目标函数可达到 $\mathcal{O}(T)$ 的收敛速率。

定理9.13：（凸且 SS：$\mathcal{O}(1/T)$ 的 FW）设 $f:\mathbb{S}\to\mathbb{R}$ 为定义在凸域 $\mathbb{S}\subseteq\mathbb{R}^p$上的适当、可微、凸且 L_b-光滑函数。假设 f 在某个点 $\boldsymbol{\beta}^*\in\mathbb{S}$处取得其全局最小值，并令 D 表示 $\mathbb{S}$ 的直径，定义为 $D\triangleq\max_{\boldsymbol{\beta},\boldsymbol{u}\in\mathbb{S}}\|\boldsymbol{\beta}-\boldsymbol{u}\|_2$。设 $\{\boldsymbol{\beta}^{(t)}\}_{t\geq 0}$ [a] 为 Frank–Wolfe 方法（算法 5）在动态步长规则下求解问题 $\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$ subject to $\boldsymbol{\beta}\in\mathbb{S}$ 所生成的序列，其中 $\eta_t=\frac{2}{t+2}$ 在第 t次迭代时。则对于任意 $T\geq 0$,

$$f(\boldsymbol{\beta}^{(T)})-f(\boldsymbol{\beta}^*)\leq\frac{2L_bD^2}{T+2}.$$

[a]。为简单起见，在条件梯度法中，我们从 $t=0$ 开始为迭代点编号。

证明 [定理 9.13]根据条件梯度法的光滑性、凸性和进展规则，我们有

$$\begin{aligned} f(\boldsymbol{\beta}^{(t+1)}) &\leq f(\boldsymbol{\beta}^{(t)})+\nabla f(\boldsymbol{\beta}^{(t)})^\top(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})+\frac{L_b}{2}\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}\right\|_2^2 && \text{(smoothness)}\\ &= f(\boldsymbol{\beta}^{(t)})+\eta_t\nabla f(\boldsymbol{\beta}^{(t)})^\top(\widetilde{\boldsymbol{\beta}}^{(t)}-\boldsymbol{\beta}^{(t)})+\frac{\eta_t^2L_b}{2}\left\|\widetilde{\boldsymbol{\beta}}^{(t)}-\boldsymbol{\beta}^{(t)}\right\|_2^2 && \text{(update rule)}\\ &\leq f(\boldsymbol{\beta}^{(t)})+\eta_t\nabla f(\boldsymbol{\beta}^{(t)})^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)})+\frac{\eta_t^2L_bD^2}{2}. && \text{(linear update rule)}\end{aligned}$$

由 f 的凸性，我们还有：$\nabla f(\boldsymbol{\beta}^{(t)})^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)})\leq f(\boldsymbol{\beta}^*)-f(\boldsymbol{\beta}^{(t)})$（参见定理 2.45）；合并这两个不等式可得：

$$f(\boldsymbol{\beta}^{(t+1)})-f(\boldsymbol{\beta}^*)\leq(1-\eta_t)(f(\boldsymbol{\beta}^{(t)})-f(\boldsymbol{\beta}^*))+\frac{\eta_t^2L_bD^2}{2}. \tag{9.42}$$

现在我们用归纳法证明，对于所有 $t\geq 0$，都有 $f(\boldsymbol{\beta}^{(t)})-f(\boldsymbol{\beta}^*)\leq\frac{2L_bD^2}{t+2}$。

基例 $t=0$。当 $t=0$时，我们有 $\eta_0=\frac{2}{2}=1$ 且 $\boldsymbol{\beta}^{(0)}=\mathbf{0}$，并且（9.42）化简为

$$f(\boldsymbol{\beta}^{(1)})-f(\boldsymbol{\beta}^*)\leq(1-\eta_0)(f(\boldsymbol{\beta}^{(0)})-f(\boldsymbol{\beta}^*))+\frac{L_b}{2}\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^*\right\|_2^2=\frac{L_bD^2}{2}\leq\frac{2L_bD^2}{2}.$$

归纳步骤。假设该结论对某个 $t \geq 0$成立，即 $f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}^*) \leq \frac{2L_b D^2}{t+2}$对所有不超过 t 的整数均成立，下面证明该结论对 $t+1$成立。利用（9.42）并取 $\eta_t = \frac{2}{t+2}$，得到：

$$
\begin{aligned}
f(\boldsymbol{\beta}^{(t+1)}) - f(\boldsymbol{\beta}^*) &\leq \left(1 - \frac{2}{t+2}\right)\left(f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}^*)\right) + \frac{2L_b D^2}{(t+2)^2} \\
&\leq \left(1 - \frac{2}{t+2}\right)\frac{2L_b D^2}{t+2} + \frac{2L_b D^2}{(t+2)^2} = 2L_b D^2 \frac{t+1}{(t+2)^2} \\
&= 2L_b D^2 \frac{t+1}{t+2} \cdot \frac{1}{t+2} \leq 2L_b D^2 \frac{t+2}{t+3} \cdot \frac{1}{t+2} = 2L_b D^2 \frac{1}{t+3}.
\end{aligned}
$$

因此，该不等式对 $t+1$ 情形也成立，归纳证明完成。 ■

算法 6 条件梯度法

要求：矩阵 $\boldsymbol{X}$、向量 $\boldsymbol{y}$、标量 $\Sigma > 0$、容差 δ；1：目标：求 $\boldsymbol{\beta}$，使得在满足 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$ 的条件下 $\frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 最小化；2：初始化：$\boldsymbol{\beta}^{(0)} = \boldsymbol{0}$（或 $\mathbb{S}$ 中任意一点）；3：对于 $t = 0, 1, 2, \ldots\ldots$，执行以下步骤：4：计算梯度：$\boldsymbol{g}^{(t)} \leftarrow \nabla f(\boldsymbol{\beta}^{(t)}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})$；5：找到梯度绝对值最大的坐标：$i^* = \arg\max_{i=1,2\ldots,p} \left|g_i^{(t)}\right|$；6：计算线性极小化解：$\widetilde{\boldsymbol{\beta}}^{(t)} \leftarrow -\text{sign}(g_{i^*}^{(t)})\Sigma \boldsymbol{e}_{i^*}$（即，坐标 i^* 上为 $-\Sigma$，其余坐标为零）；7：选择步长 $\eta_t \in [0,1]$（例如 $\eta_t = \frac{2}{t+2}$）；8：更新：$\boldsymbol{\beta}^{(t+1)} \leftarrow (1-\eta_t)\boldsymbol{\beta}^{(t)} + \eta_t \widetilde{\boldsymbol{\beta}}^{(t)}$；9：如果在迭代 $t = T$ 时满足停止准则，则停止；10：结束循环；11：返回 $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$；

▶ **使用 FW 方法求解约束 LASSO**。我们将条件梯度法应用于约束 LASSO 问题（L_C）；见算法 6。在第 t 次迭代中，线性子问题需要计算梯度 $\boldsymbol{g} \triangleq \nabla f(\boldsymbol{\beta}^{(t)}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})$。然后，我们在约束 $\boldsymbol{g}^\top\boldsymbol{\beta}$ 下最小化线性函数 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$。根据 ℓ_1-范数与 ℓ_∞-范数之间的对偶性，该子问题的最优值为

$$
\widetilde{\boldsymbol{\beta}} = \min_{\|\boldsymbol{\beta}\|_1 \leq \Sigma} \boldsymbol{g}^\top\boldsymbol{\beta} = -\Sigma\|\boldsymbol{g}\|_\infty . \tag{9.43}
$$

一个最优的 $\boldsymbol{\beta}$ 可通过将全部 ℓ_1 预算 Σ 分配给绝对梯度分量最大的单个坐标 j 来获得，即满足 $|g_j| = \|\boldsymbol{g}\|_\infty$的坐标，并选择与$g_j$相反的符号函数：

$$
\widetilde{\beta}_i = \begin{cases} -\Sigma\text{sign}(g_j), & i = j, \\ 0, & i \neq j. \end{cases}
$$

这种选择得到 $\boldsymbol{g}^\top\widetilde{\boldsymbol{\beta}} = -\Sigma\|\boldsymbol{g}\|_\infty$。如果多个坐标的绝对值都达到最大值（即 $\|\boldsymbol{g}\|_\infty$ 中存在并列），则最优解不唯一。在这种情况下，可以将 ℓ_1 预算分配给这些坐标的任意子集，前提是每个

选定的坐标 i 所取的符号与 g_i 相反。如果 $\boldsymbol{g} = \mathbf{0}$，那么每个可行的 $\boldsymbol{\beta}$ （即每个满足 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$ 的 $\boldsymbol{\beta}$ ）都是最优解，且目标函数值为零。

由于可行集是 ℓ_1球 $\mathbb{B}_1[\Sigma]$，其直径（按欧几里得范数计算）满足$D = \max_{\boldsymbol{u}\in\mathbb{S}} \|\boldsymbol{\beta} - \boldsymbol{u}\|_2 \leq 2\Sigma$，因为 ℓ_1球中任意两点之间的距离至多为 2Σ ，按ℓ_1范数计算；并且 $\|\cdot\|_2 \leq \|\cdot\|_1$ （习题1.26）。因此，该方法达到 $\mathcal{O}(1/T)$ 收敛速率，其常数取决于 Σ。

9.2.4 最小角回归（LARS）

最小角回归（*LARS*）是一种高效的模型拟合算法，为理解几种关键回归技术提供了统一框架，包括前向选择、前向逐步回归和 LASSO。它以几何直观的方式构建线性模型：迭代地沿一个方向添加预测变量，使其与当前残差相关性的夹角保持相等，因此得名“最小角回归”。

该算法按顺序进行，从与响应相关性最高的预测变量开始。它不会立即将该变量完全纳入模型，而是逐渐增大其系数，使其连续地趋近于最小二乘估计值。随着系数增大，该预测变量与当前残差的相关性会降低。一旦另一个预测变量“迎头赶上”——也就是说，它与残差的相关性大小相等——这一过程便会暂停。此时，第二个预测变量加入活跃集，两个系数则一同更新，同时与残差保持相等的绝对相关性（即相等的角距离）。这种协同更新会持续进行，直至所有变量都已纳入，最终收敛到完整最小二乘解。

在介绍 LARS 算法之前，我们先回顾两种相关方法：前向选择和前向逐步回归，它们是 LARS 的概念先导（Efron 等人，2004年；Weisberg，2005年）。

前向选择算法

LARS 与一种称为前向选择算法 （又称前向逐步回归）的经典模型选择方法密切相关。前向选择是一种贪心算法，它通过在每一步依次加入信息量最大的预测变量来构建模型。

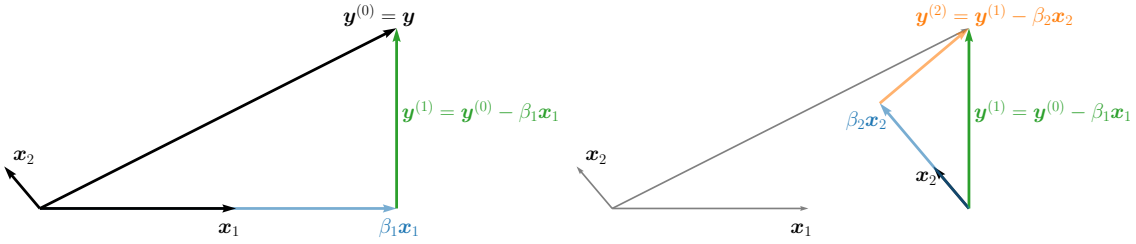
考虑线性模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$，其中 $\boldsymbol{y} \in \mathbb{R}^n$ 是响应向量， $\boldsymbol{X} \in \mathbb{R}^{n\times p}$是包含 p 个预测变量的设计矩阵。算法从所有系数均设为零开始，并将初始残差定义为 $\boldsymbol{y}^{(0)} \triangleq \boldsymbol{y}$。在 $\boldsymbol{X}$中的所有候选预测变量中，算法选择与当前残差相关性最高的变量——例如 $\boldsymbol{x}_k$——（等价地说，它与残差的夹角最小，或具有最大的绝对余弦相似度）。随后，算法按以下步骤进行：算法按以下步骤进行：

（i）计算 $\boldsymbol{y}^{(0)}$ 在 $\boldsymbol{x}_k$上的投影，得到拟合分量 $\beta_k\boldsymbol{x}_k$，并更新残差：$\boldsymbol{y}^{(1)} = \boldsymbol{y}^{(0)} - \beta_k\boldsymbol{x}_k$。根据构造， $\boldsymbol{y}^{(1)}$ 与 $\boldsymbol{x}_k$ 正交。（ii）将 $\boldsymbol{y}^{(1)}$ 视为新的响应变量，从可用预测变量池中移除 $\boldsymbol{x}_k$ ，并使用剩余特征 $\{\boldsymbol{x}_i \mid i \neq k\}$重复该过程。（iii）持续迭代，直到满足停止准则——例如所有预测变量均已用尽、残差足够小，或达到预先指定的模型规模。最终估计值为 $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]^\top$，其中未使用的预测变量的系数为零。

举例来说，考虑一个简单的二维示例（见图 9.1）。第一步中， $\boldsymbol{x}_1$ 与 $\boldsymbol{y}^{(0)}$的夹角最小。其投影 $\beta_1\boldsymbol{x}_1$作为 $\boldsymbol{y}$的初始近似，而残差为 $\boldsymbol{y}^{(1)} = \boldsymbol{y}^{(0)} - \beta_1\boldsymbol{x}_1$。下一步中， $\boldsymbol{y}^{(1)}$ 变为

新的目标，此时唯一剩下的预测变量是 $\boldsymbol{x}_2$。投影到 $\boldsymbol{x}_2$ 上后，算法终止，得到解 $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$。

由于前向选择每选入一个变量只更新一次，因此计算效率高、速度快。然而，这种简便性是以粗糙为代价的：一旦加入某个变量，其系数便固定不变；与 LARS 或 LASSO 等更精细的方法相比，这可能导致次优拟合。



(a) 步骤 1。　　(b) 步骤 2。

图9.1： 前向选择算法示例。

前向逐步算法

与前向选择算法类似，前向逐步算法首先在特征变量集合 $\boldsymbol{X}$ 中确定与初始残差向量 $\boldsymbol{y}^{(0)} \equiv \boldsymbol{y}$ 相关性最高的预测变量 $\boldsymbol{x}_k$ ——也就是与 $\boldsymbol{y}$ 形成最小角（或者等价地，与其具有最大绝对余弦相似度）的预测变量。将步长记为 $\eta > 0$。算法过程如下：

(i) 沿 $\boldsymbol{x}_k$ 的方向迈出大小为 η 的一小步：通过添加 $\eta \boldsymbol{x}_k$ 来更新拟合结果，并计算新的残差 $\boldsymbol{y}^{(1)} = \boldsymbol{y}^{(0)} - \eta \boldsymbol{x}_k$。(ii) 将 $\boldsymbol{y}^{(1)}$ 作为新的响应（残差），并保留所有特征变量 $\{\boldsymbol{x}_i\}_{i=1}^p$——包括已经使用过的变量——作为下一步的候选变量。重复步骤 (i)，每次选择与当前残差相关性最高的预测变量，直到满足终止条件。当残差变得足够小时，或无法再取得有意义的进展时，算法终止。

图9.2 展示了这一过程在二维情形下的表现。假设 $\boldsymbol{x}_1$ 最初与 $\boldsymbol{y}^{(0)}$形成的角度最小。算法沿着 $\boldsymbol{x}_1$的方向反复迈出小步，每一步的步长为 $\eta \boldsymbol{x}_1$，相应的残差以橙色向量表示。随着残差不断变化，它与 $\boldsymbol{x}_2$ 的相关性逐渐增大。一旦 $\boldsymbol{x}_2$ 与残差的方向比 $\boldsymbol{x}_1$ 更一致（即形成的角度更小），算法便会切换方向，开始沿着 $\boldsymbol{x}_2$ 迈步。

该过程以迭代方式持续进行：每一步中，所有预测变量都仍可供选择，任何变量都可以被重复使用。当步长 η 选得非常小时，该算法可以非常接近最优解，但代价是需要大量迭代，从而增加计算成本。

由此得到的路径看起来像一段阶梯，这是逐步法的典型特征——因此得名前向逐步算法。

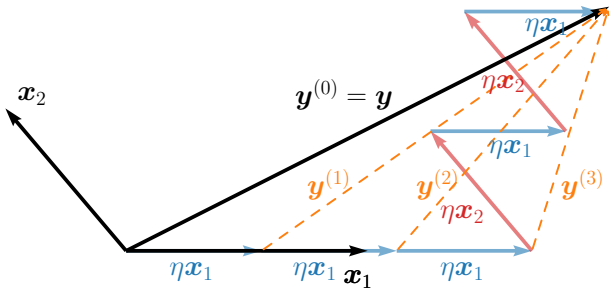


图9.2： 二维情形下前向逐步算法的若干步骤。蓝色线段表示沿 $\boldsymbol{x}_1$方向移动，红色线段表示沿$\boldsymbol{x}_2$方向移动，橙色向量表示每一步之后的残差。阶梯状轨迹反映了逐步路径增量式、分段式的特点。

LARS

最小角回归（*LARS*）算法是上述两种算法之间的一种折中：它比逐步回归的计算效率更高，也比前向选择更准确。

令 $\{\widehat{\boldsymbol{y}}_0, \widehat{\boldsymbol{y}}_1, \widehat{\boldsymbol{y}}_2, \ldots, \widehat{\boldsymbol{y}}_p\}$ 表示 LARS 在逐步逼近响应向量 $\boldsymbol{y}$的过程中生成的拟合值向量序列。算法以 $\widehat{\boldsymbol{y}}_0 = \boldsymbol{0}$ 为初始值（对应于仅含截距的模型，我们假定截距已作处理）。当满足以下任一条件时，过程终止：没有剩余残差需要解释；所有可用的预测变量均已纳入；或者残差范数低于预先设定的容差。

在一维情形下，LARS 与前向选择完全一致。我们计算 $\widehat{\boldsymbol{y}}_1 = \widehat{\boldsymbol{y}}_0 + \widehat{\gamma}_1\boldsymbol{u}_1$，其中 $\boldsymbol{u}_1$ 是沿所选预测变量 $\boldsymbol{x}_1$ 方向的单位向量。通过适当选择 $\widehat{\gamma}_1$，$\widehat{\boldsymbol{y}}_1$ 与 $\boldsymbol{y}_1$ 对齐（响应变量 $\boldsymbol{y}$ 在由 $\boldsymbol{x}_1$ 张成的一维子空间上的正交投影），算法在满足终止条件后停止。

在二维情形下，假设有两个预测变量 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$。令 $\boldsymbol{y}_2$ 表示 $\boldsymbol{y}$ 在列空间 $\mathcal{C}$（$[\boldsymbol{x}_1, \boldsymbol{x}_2]$）上的投影。LARS 首先找出与当前残差 $\boldsymbol{y}_2$ 相关性最高的预测变量——在此例中，设为 $\boldsymbol{x}_1$。LARS 不会沿着 $\boldsymbol{x}_1$ 一直移动到 $\boldsymbol{y}_2$，而是沿着 $\boldsymbol{u}_1$（沿 $\boldsymbol{x}_1$ 的单位向量）方向前进大小为 $\widehat{\gamma}_1$ 的一步，从而得到

$$\widehat{\boldsymbol{y}}_1 = \widehat{\boldsymbol{y}}_0 + \widehat{\gamma}_1\boldsymbol{u}_1,$$

选择它，使新的残差 $\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_1$ 位于 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$之间的角平分线上。接下来，算法沿着这条角平分线方向前进。设 $\boldsymbol{u}_2$ 为沿着 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$之间的角平分线方向的单位向量。那么

$$\widehat{\boldsymbol{y}}_2 = \widehat{\boldsymbol{y}}_1 + \widehat{\gamma}_2\boldsymbol{u}_2.$$

通过调整 $\widehat{\gamma}_2$，我们可以直接移动到 $\boldsymbol{y}_2$，在二维空间中完成拟合；参见图 9.3(a)。

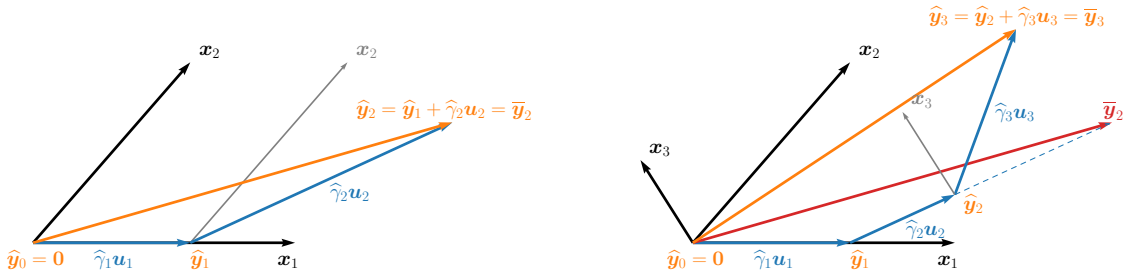
对于一般的高维情形，考虑算法的第 t次迭代：

(i) 沿当前活跃预测变量之间的等角（即平分角）方向移动步长 $\widehat{\gamma}_t$。具体而言，令

$$\widehat{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \widehat{\gamma}_t\boldsymbol{u}_t,$$

其中， $\boldsymbol{u}_t$ 是由活跃集定义的等角方向上的单位向量。选择这一方向，使更新后的残差与所有活跃预测变量的绝对相关程度相同。下一个进入活跃集的预测变量——记为 $\boldsymbol{x}_{t+1}$——是在此步骤结束时与残差相关程度变得相同的那个预测变量。也就是说，当前残差 $\boldsymbol{y}_{t+1} - \widehat{\boldsymbol{y}}_t$ 位于 $\boldsymbol{u}_t$ 和$\boldsymbol{x}_{t+1}$之间的角平分线上。

(ii) 重复步骤 (i)，直到满足终止条件。



(a) LARS 的二维情形。 (b) LARS 的三维情形。

图 9.3：LARS 算法示意图。当 $\boldsymbol{X}$ 是一个包含两个协变量（或特征、预测变量） $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的二维矩阵时，我们找到与 $\boldsymbol{y}_2$ 最接近的分量（即夹角最小、余弦值最大，并且 $\boldsymbol{y}_2$ 是 $\boldsymbol{y}$ 在$\mathcal{C}$（$[\boldsymbol{x}_1, \boldsymbol{x}_2]$）上的投影），记作 $\boldsymbol{x}_1$。我们沿着 $\boldsymbol{x}_1$ 的方向移动步长 $\widehat{\gamma}_1\boldsymbol{u}_1$ ，使得到的残差 $\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_1$ 落在特征向量 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$之间的角平分线上（其中 $\boldsymbol{u}_1$ 是沿着 $\boldsymbol{x}_1$方向的单位向量）。令 $\widehat{\boldsymbol{y}}_2 = \widehat{\boldsymbol{y}}_1 + \widehat{\gamma}_2\boldsymbol{u}_2$，其中 $\boldsymbol{u}_2$ 是沿着角平分线方向的单位向量。通过调整 $\widehat{\gamma}_2$，我们可以直接使 $\widehat{\boldsymbol{y}}_2$ 与 $\boldsymbol{y}_2$ 对齐，直到满足终止条件。请注意，在 $p = 2$的情形下，$\widehat{\boldsymbol{y}}_2 = \boldsymbol{y}_2$ 成立，但在 $p > 2$的情形下则不成立，例如图 9.3(b) 所示。

LARS 的关键创新出现在选定第一个预测变量之后。与前向选择——在考虑其他变量之前，先对选定变量进行完全拟合——或逐步回归——采取微小的增量步骤——不同，LARS 沿着一个方向移动，使当前残差与所有活跃预测变量之间保持相等的角（或相等的绝对相关性）。这一等角方向确保所选变量能够均衡地发挥作用。算法沿着这一路径继续前进，直到另一个预测变量与残差的相关性“赶上”当前水平；此时，它便加入活跃集，并重新计算一个等角方向。这一过程生成了分段线性解路径，既高效，在几何上也十分优雅。

▶ **示例说明**。 为进一步阐明 LARS 算法，我们作出以下标准预处理假设：

- 每个预测变量 $\boldsymbol{x}_i$ 都经过标准化，使其均值为零且欧几里得范数为 1。响应向
- 量 $\boldsymbol{y}$ 经过中心化，使其均值为零。

这些步骤简化了计算，并确保所有变量的尺度具有可比性。在这种标准化下，任意预测变量与残差之间的（绝对）相关性都与它们的内积成正比，因此可以直接通过点积来解读角度。$\widehat{}$令 $\boldsymbol{\beta}$ 表示当前系数向量， $\widehat{\boldsymbol{y}}$ 表示当前估计值。LARS逐步构建估计值

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{9.44}$$

每一步都向模型中添加一个协变量，因此经过 t 步后，只有系数 β_i 非零。令 $\boldsymbol{c}(\widehat{\boldsymbol{y}})$表示当前相关性向量

$$\widehat{\boldsymbol{c}} \triangleq \boldsymbol{c}(\widehat{\boldsymbol{y}}) = \boldsymbol{X}^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \equiv \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{9.45}$$

使得 $\widehat{c}_i = c_i(\widehat{\boldsymbol{y}})$ 表示第 i 个协变量 $\boldsymbol{x}_i$ 与当前残差向量 $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ 之间的相关性。

图 9.3(a) 展示了在有 $p = 2$ 个预测变量时的算法，$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2]$。由于这些预测变量张成一个二维子空间，当前相关性仅取决于 $\boldsymbol{y}$ 在该空间上的投影，记作 $\boldsymbol{y}_2 = \mathcal{P}_{[\boldsymbol{x}_1,\boldsymbol{x}_2]}(\boldsymbol{y})$。因此，

$$\boldsymbol{c}(\widehat{\boldsymbol{y}}) \triangleq \boldsymbol{X}^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}) = \boldsymbol{X}^\top(\overline{\boldsymbol{y}}_2 - \widehat{\boldsymbol{y}}), \tag{9.46}$$

其中，第 i 个分量 $c_i(\widehat{\boldsymbol{y}})$ 恰好是 $\boldsymbol{x}_i$ 与残差（$\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}$）之间夹角的余弦，因为协变量 $\boldsymbol{x}_i$ 具有单位范数。

算法从 $\widehat{\boldsymbol{y}}_0 = \boldsymbol{0}$开始。在图 9.3(a) （图 9.3(b)中情况类似）中，初始残差 $\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_0 = \boldsymbol{y}_2$与 $\boldsymbol{x}_1$所成的角小于与 $\boldsymbol{x}_2$所成的角，即 $c_1(\widehat{\boldsymbol{y}}_0) > c_2(\widehat{\boldsymbol{y}}_0)$。随后，LARS 从 $\widehat{\boldsymbol{y}}_0$ 沿着 $\boldsymbol{x}_1$方向移动：

$$\widehat{\boldsymbol{y}}_1 = \widehat{\boldsymbol{y}}_0 + \widehat{\gamma}_1 \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|_2} = \widehat{\boldsymbol{y}}_0 + \widehat{\gamma}_1 \boldsymbol{u}_1. \tag{9.47}$$

经典前向选择算法会取足够大的 $\widehat{\gamma}_1$，使 $\widehat{\boldsymbol{y}}_1$ 等于 $\boldsymbol{y}_1 \triangleq \mathcal{P}_{[\boldsymbol{x}_1]}(\boldsymbol{y})$，即将 $\boldsymbol{y}$ 投影到 $\mathcal{C}([\boldsymbol{x}_1])$ 上所得的正交投影；而前向逐步算法建议将 $\widehat{\gamma}_1$ 取为某个较小的值 η，然后多次重复这一过程。LARS 采用一种折中策略：它选择 $\widehat{\gamma}_1$，使新的残差 $\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_1$ 与 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$的相关性**相同**。从几何上看，这意味着残差$\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_1$ 位于 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$之间的角平分线上，因此

$$c_1(\widehat{\boldsymbol{y}}_1) = c_2(\widehat{\boldsymbol{y}}_1).$$

令 $\boldsymbol{u}_2 \triangleq \frac{\boldsymbol{y}_2 - \boldsymbol{y}_1}{\|\boldsymbol{y}_2 - \widehat{\boldsymbol{y}}_1\|_2}$ 表示沿该角平分线方向的单位向量。下一次 LARS 更新为

$$\widehat{\boldsymbol{y}}_2 = \widehat{\boldsymbol{y}}_1 + \widehat{\gamma}_2 \boldsymbol{u}_2, \tag{9.48}$$

取 $\widehat{\gamma}_2$ 使其满足 $\widehat{\boldsymbol{y}}_2 = \boldsymbol{y}_2$ 在 $\overline{p} = 2$ 的情况下（图 9.3(a)）。然而，当有 $p > 2$ 个协变量时，$\widehat{\gamma}_2$ 会**更小**，从而导致方向再次改变，如图 9.3(b) 所示，使新的残差 $\boldsymbol{y}_3 - \widehat{\boldsymbol{y}}_2$ 位于 $\boldsymbol{u}_2$ 和$\boldsymbol{x}_3$ 之间的角平分线上，其中 $\boldsymbol{x}_3$为给定值。

后续的 LARS 步将这一思想推广开来：当 k 个预测变量处于活跃状态时，算法会沿等角方向前进——这是一个独特的方向，使其与所有活跃预测变量的绝对相关性保持相等。这个方向将角平分线的几何概念推广到了更高维空间，如图 $\boldsymbol{u}_3$ 中的向量所示，见9.3(b)。

最小角回归算法的数学公式

在建立了对 LARS 算法的几何直觉之后，我们现在转向其精确的数学表述。

▶ **求单位范数的角平分线。** 假设协变量向量$\boldsymbol{x}_1$、$\boldsymbol{x}_2$、……、$\boldsymbol{x}_p$线性无关。假设 LARS 刚完成第 $t-1$ 步，得到当前估计值 $\widehat{\boldsymbol{y}}_{t-1}$，并即将开始第 t 步。令 $\mathbb{S}$ 表示活跃集——即当前纳入模型的预测变量的索引集合——其基数为 $|\mathbb{S}|=t$。将活跃变量的带符号设计矩阵定义为

$$\boldsymbol{X}_t \triangleq [\ldots, s_i\boldsymbol{x}_i, \ldots]_{i\in\mathbb{S}}, \tag{9.49a}$$

其中符号 s_i 等于 ± 1。[1] 接下来，定义格拉姆矩阵和一个缩放常数：

$$\boldsymbol{G}_t \triangleq \boldsymbol{X}_t^\top \boldsymbol{X}_t \qquad \text{and} \qquad \sigma_t \triangleq (\mathbf{1}_t^\top \boldsymbol{G}_t^{-1}\mathbf{1}_t)^{-1/2}, \tag{9.49b}$$

其中 $\mathbf{1}_t$ 是一个长度为 $|\mathbb{S}|=t$的全 $\mathbf{1}$ 向量。于是，向量

$$\boldsymbol{u}_t = \boldsymbol{X}_t\boldsymbol{w}_t, \quad \text{where } \boldsymbol{w}_t \triangleq \sigma_t \boldsymbol{G}_t^{-1}\mathbf{1}_t, \tag{9.49c}$$

是一个单位向量，满足

$$\boldsymbol{X}_t^\top \boldsymbol{u}_t = \boldsymbol{X}_t^\top \boldsymbol{X}_t \boldsymbol{w}_t = \sigma_t \mathbf{1}_t. \tag{9.49d}$$

等式（9.49c）表明 $\|\boldsymbol{u}_t\|_2 = 1$ 使得 $\boldsymbol{u}_t$ 是一个单位长度向量；而（9.49d）则意味着 $\boldsymbol{u}_t$ 与每个带符号的预测变量 $s_i\boldsymbol{x}_i$ （$i\in\mathbb{S}$）之间的内积（因此也包括夹角的余弦）都相同，且等于 $\sigma_t > 0$ （因此， $\boldsymbol{x}_i$、 $i\in\mathbb{S}$和 $\boldsymbol{u}_t$ 之间的夹角相等且小于 $90°$）。因此， $\boldsymbol{u}_t$ 与所有活跃预测变量都形成相等的锐角。这样的向量 $\boldsymbol{u}_t$ 称为 等角向量 或 角平分线，它是 $\boldsymbol{X}_t$的列向量。

▶ **LARS 算法的一般步骤。** 现在我们可以描述完整的 LARS 过程。从 $\widehat{\boldsymbol{y}}_0 = \mathbf{0}$出发，算法通过一系列步骤构造拟合值 $\widehat{\boldsymbol{y}}$，每一步通常都比逐步回归中的步长更大。假设$\widehat{\boldsymbol{y}}_{t-1}$ 是当前的 LARS 估计，位于对应于活跃集 $\mathbb{S}_-$ 的 $\boldsymbol{X}_{t-1}$ 的列空间中，其中 $|\mathbb{S}_-| = |\mathbb{S}| - 1 = t-1$。令当前相关向量为

$$\widehat{\boldsymbol{c}} = \boldsymbol{X}^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}), \tag{9.50}$$

其中 $\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}$ 是当前残差。一般步骤由两部分组成：**(i)** 沿当前活动子空间（即 $\boldsymbol{X}_t$的列空间）中的等角方向移动，以更新估计值；**(ii)** 确定何时应有新的预测变量加入活跃集，从而将活跃集更新为 $\mathbb{S}_+$，使得 $|\mathbb{S}_+| = |\mathbb{S}| + 1$。

第 t 步的活跃集 $\mathbb{S}\subseteq\{1,2,\ldots,p\}$ 由所有使当前相关性的绝对值达到最大值的索引组成：

$$\widehat{C}_t = \max_i\{|\widehat{c}_i|\} \qquad \text{and} \qquad \mathbb{S} = \left\{ i \mid |\widehat{c}_i| = \widehat{C}_t \right\}. \tag{9.51}$$

对于每个 $i\in\mathbb{S}$，定义符号函数

$$s_i \triangleq \operatorname{sign}\{\widehat{c}_i\}, \quad \text{for } i\in\mathbb{S}, \tag{9.52}$$

并构造 $\boldsymbol{X}_t$、 σ_t和 $\boldsymbol{u}_t$，如式（9.49）所示。此外，计算方向相关性向量：

$$\boldsymbol{z} \triangleq \boldsymbol{X}^\top \boldsymbol{u}_t. \tag{9.53}$$

1。如果残差向量与 $\boldsymbol{x}_i$ 的相关性为正，则 $s_i = 1$；否则 $s_i = -1$；更多细节见（9.52）。这些带符号的协变量使回归向量在算法结束时为正，即满足符号约束；参见引理 9.15。

下一个 LARS 估计值通过沿着 $\boldsymbol{u}_t$方向从 $\widehat{\boldsymbol{y}}_{t-1}$ 移动得到:

$$\widehat{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \widehat{\gamma}_t \boldsymbol{u}_t, \tag{9.54}$$

其中

$$\widehat{\gamma}_t = \min_{i\in\mathbb{S}^c}{}^{+} \left\{ \frac{\widehat{C}_t - \widehat{c}_i}{\sigma_t - z_i}, \frac{\widehat{C}_t + \widehat{c}_i}{\sigma_t + z_i} \right\}; \tag{9.55}$$

“最小值$^+$”表示最小值仅在每个选择的正分量上取得，取值范围为 i在（9.55）中。

为理解步长 $\widehat{\gamma}_t$ （9.55），请考虑任意$\gamma > 0$的候选更新。我们定义潜在的下一个 LARS 估计值

$$\boldsymbol{y}(\gamma) \triangleq \widehat{\boldsymbol{y}}_{t-1} + \gamma \boldsymbol{u}_t, \tag{9.56}$$

这将得到预测变量 i的更新后相关性

$$c_i(\gamma) = \boldsymbol{x}_i^\top \big(\boldsymbol{y} - \boldsymbol{y}(\gamma)\big) = \widehat{c}_i - \gamma z_i, \quad \text{for all } i. \tag{9.57}$$

对于活动变量 活跃索引 $i \in \mathbb{S}$, (9.51) 以及 $\boldsymbol{X}_t^\top \boldsymbol{u}_t = \sigma_t \mathbf{1}_t$ 在 (9.49d) 这一事实可得

$$|c_i(\gamma)| = \widehat{C}_t - \gamma\sigma_t, \quad \text{for all } i \in \mathbb{S}, \tag{9.58}$$

这表明所有最大的绝对当前相关性都以相同幅度下降。对于 $i \in \mathbb{S}^c$，令 (9.57) 与 (9.58) 相等可知， $c_i(\gamma)$ 在 $\gamma = (C_t - \widehat{c}_i)/(\sigma_t - z_i)$ 处达到最大值。同样， $-c_i(\gamma)$，即反向协变量 $-\boldsymbol{x}_i$ 的当前相关性，在 $(C_t + \widehat{c}_i)/(\sigma_t + z_i)$ 处达到 $\widehat{}$ 极大性。因此，(9.55) 中的 $\widehat{\gamma}_t$ 是使某个新索引 i_+ 加入活跃集的最小正值； i_+ 是 (9.55) 中使其取到最小值的索引，而新的活跃集 $\mathbb{S}_+$ 为 $\mathbb{S}_+ = \mathbb{S} \cup \{i_+\}$；对于 $i \in \mathbb{S}_+$，新的最大绝对相关性为 $\widehat{}$ $\widehat{C}_{t+1} = C_t - \widehat{\gamma}_t\sigma_t$。

▸LARS 路径的几何性质。设

$$\widehat{C}_t \mathbf{1}_t \triangleq \boldsymbol{X}_t^\top (\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}) \tag{9.59}$$

表示步骤 t 处的相关性向量，此时根据构造，所有活跃预测变量都具有相同的$\widehat{}$绝对相关性 C_t。设 $\boldsymbol{y}_t$ 表示 $\boldsymbol{y}$ 在列空间 $\mathcal{C}(\boldsymbol{X}_t)$ 上的正交投影。由于 $\widehat{\boldsymbol{y}}_{t-1} \in \mathcal{C}(\boldsymbol{X}_{t-1}) \subseteq \mathcal{C}(\boldsymbol{X}_t)$，该投影可写为

$$\overline{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \boldsymbol{X}_t \boldsymbol{G}_t^{-1} \boldsymbol{X}_t^\top (\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}) = \widehat{\boldsymbol{y}}_{t-1} + \frac{\widehat{C}_t}{\sigma_t} \boldsymbol{u}_t, \tag{9.60}$$

其中，$\widehat{\boldsymbol{y}}_{t-1} + \boldsymbol{X}_t \boldsymbol{G}_t^{-1} \boldsymbol{X}_t^\top (\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}$表示将残差 $\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}$ 投影到$\mathcal{C}(\boldsymbol{X}_t$上（引理1.44），最后一个等式由（9.49c）以及由符号函数确定的$\widehat{}$当前相关性均等于 C_t 这一事实得出（见（9.59））。由于 $\boldsymbol{u}_t$ 是单位向量，（9.60）说明$\boldsymbol{y}_t - \widehat{\boldsymbol{y}}_{t-1}$ 的长度为

$$\overline{\gamma}_t \triangleq \frac{\widehat{C}_t}{\sigma_t} \quad \Longrightarrow \quad \overline{\boldsymbol{y}}_t - \widehat{\boldsymbol{y}}_{t-1} = \overline{\gamma}_t \boldsymbol{u}_t. \tag{9.61}$$

注 9.14（最小角回归路径的几何性质）。与（9.54）——即估计值的更新 $\widehat{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \widehat{\gamma}_t \boldsymbol{u}_t$——进行比较可知，LARS 估计 $\widehat{\boldsymbol{y}}_t$ 位于从 $\widehat{\boldsymbol{y}}_{t-1}$ 到 $\boldsymbol{y}_t$的直线上，

$$\widehat{\boldsymbol{y}}_t - \widehat{\boldsymbol{y}}_{t-1} = \widehat{\gamma}_t \boldsymbol{u}_t = \frac{\widehat{\gamma}_t}{\overline{\gamma}_t}(\overline{\boldsymbol{y}}_t - \widehat{\boldsymbol{y}}_{t-1}). \tag{9.62}$$

根据（9.55），很容易看出 $\widehat{\gamma}_t$ 总是小于 γ_t，因此 $\widehat{\boldsymbol{y}}_t$ 比 $\boldsymbol{y}_t$更接近 $\widehat{\boldsymbol{y}}_{t-1}$。因此，连续的 LARS 估计 $\widehat{\boldsymbol{y}}_t$ 总是在逐渐接近、但永远不会达到 OLS 估计 $\boldsymbol{y}_t$。

唯一的例外出现在最后一步。当活跃集包含全部 p预测变量时，即 $\mathbb{S} = \{1, 2, \ldots, p\}$，步长公式（9.55）不再适用（因为不存在非活跃变量）。按照惯例，算法取 $\widehat{\gamma}_p = \gamma_p = C_p/\sigma_p$，使当前估计值 $\widehat{\boldsymbol{y}}_p = \boldsymbol{y}_p$ 以及当前系数 $\boldsymbol{\beta}_p$ 等于包含全部 p 协变量的集合所对应的 OLS 估计。

为使符号清晰，我们定义：$\boldsymbol{c}_t \triangleq \boldsymbol{X}_t^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1})$ 表示第 t 步的相关性向量，其中 c_{ti} 表示 $\boldsymbol{c}_t$ 的第 i 个元素。并令 $\boldsymbol{\beta}_t$ 表示满足 $\widehat{\boldsymbol{y}}_t = \boldsymbol{X}\boldsymbol{\beta}_t$ 的完整回归系数向量，令 β_{ti} 表示 $\boldsymbol{\beta}_t$ 的第 i 个元素。现在我们证明 LARS 算法的一个重要性质：新变量进入模型的方向与其当前残差相关性一致。

引理 9.15：（LARS 的符号限制）假设 LARS 已完成第 $t-1$ 步，得到估计值 $\widehat{\boldsymbol{y}}_{t-1}$ 和第 t 步的活跃集 $\mathbb{S}_t$，其中预测变量 $\boldsymbol{x}_t$ 是最新加入活跃集的变量（即 $\mathbb{S}_t = \mathbb{S}_{t-1} \cup \{t\}$）。如果在第 $t-1$ 步结束时，只有 $\boldsymbol{x}_t$ 加入活跃集，那么等角向量 $\boldsymbol{u}_t = \boldsymbol{X}_t \boldsymbol{w}_t$ (9.49c) 的系数向量 $\boldsymbol{w}_t = \sigma_t \boldsymbol{G}_t^{-1}\mathbf{1}_t$ 的第 t 个分量 w_{tt} 的符号与当前相关性 $c_{tt} = \boldsymbol{x}_t^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1})$一致。此外，$\widehat{\boldsymbol{y}}_t = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_t$ 的回归向量 $\widehat{\boldsymbol{\beta}}_t$ 的第 t 个分量 β_{tt} 的符号与 c_{tt} 一致。

证明 [引理 9.15] 情形 $t = 1$ 显然成立，因为第一个预测变量沿着其与 $\boldsymbol{y}$ 的相关性方向进入。对于 $t \geq 2$，回顾式（9.59）可知 $bX_t^\top (\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1}) = \widehat{C}_t \mathbf{1}_t$。利用式（9.49c），我们可以将其表示为

$$\boldsymbol{w}_t = \sigma_t \boldsymbol{G}_t^{-1}\mathbf{1}_t = \sigma_t \widehat{C}_t^{-1}[(\boldsymbol{X}_t^\top \boldsymbol{X}_t)^{-1}\boldsymbol{X}_t^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1})] \triangleq \sigma_t \widehat{C}_t^{-1}\widehat{\boldsymbol{w}}_t.$$

方括号中的项是将当前残差对 $\boldsymbol{X}_t$ 进行回归所得的最小二乘系数向量，而其前面的项为正。

现在注意到

$$\boldsymbol{X}_t^\top(\boldsymbol{y} - \overline{\boldsymbol{y}}_{t-1}) = [\mathbf{0}, \delta]^\top, \quad \text{with } \delta > 0, \tag{9.63}$$

这是因为根据定义，$\boldsymbol{X}_{t-1}^\top(\boldsymbol{y} - \boldsymbol{y}_{t-1}) = \mathbf{0}_{t-1}$（因为 $\boldsymbol{y}_{t-1}$ 是 $\boldsymbol{y}$ 在 $\mathcal{C}(\boldsymbol{X}_{t-1})$上的投影），并且当 $i \in \mathbb{S}_{t-1}$时，$c_t(\gamma) = \boldsymbol{x}_t^\top(\boldsymbol{y} - \gamma \boldsymbol{u}_{t-1})$随 γ 的递减速度比 $c_i(\gamma)$更慢：

$$c_t(\gamma) \begin{cases} < c_i(\gamma), & \text{if } \gamma < \widehat{\gamma}_{t-1}; \\ = c_i(\gamma) = \widehat{C}_t, & \text{if } \gamma = \widehat{\gamma}_{t-1}; \\ > c_i(\gamma), & \text{if } \widehat{\gamma}_{t-1} < \gamma < \overline{\gamma}_{t-1}. \end{cases}$$

因此，

$$\begin{aligned} \widehat{\boldsymbol{w}}_t &= (\boldsymbol{X}_t^\top \boldsymbol{X}_t)^{-1}\boldsymbol{X}_t^\top(\boldsymbol{y} - \overline{\boldsymbol{y}}_{t-1} + \overline{\boldsymbol{y}}_{t-1} - \widehat{\boldsymbol{y}}_{t-1}) \\ &= (\boldsymbol{X}_t^\top \boldsymbol{X}_t)^{-1}\begin{bmatrix}\mathbf{0} \\ \delta\end{bmatrix} + (\boldsymbol{X}_t^\top \boldsymbol{X}_t)^{-1}\boldsymbol{X}_t^\top[(\overline{\gamma}_{t-1} - \widehat{\gamma}_{t-1})\boldsymbol{u}_{t-1}], \end{aligned} \tag{9.64}$$

其中，最后一个等式由（9.61）和（9.62）推出。阶元素 $\widehat{w}_t$ 为正，因为它出现在（9.64）中的第一项里（由于 $\boldsymbol{X}_t$ 具有满列秩，因此 $\boldsymbol{X}_t^\top \boldsymbol{X}_t$ 是正定的），而在第二项中它为 0，因为 $\boldsymbol{u}_{t-1} = \boldsymbol{X}_{t-1}\boldsymbol{w}_{t-1} \in \mathcal{C}(\boldsymbol{X}_{t-1})$。这证明了第一个结论。

对于第二个结论，注意到完整的系数向量按如下方式更新

$$\widehat{\beta}_{tt} = \widehat{\beta}_{t-1,t} + \widehat{\gamma}_t w_{tt}, \tag{9.65}$$

其中 $\widehat{\beta}_{t-1,\,t} = 0$，因为 $\boldsymbol{x}_t$ 在步骤 t之前并未处于激活状态。由于 $\widehat{\gamma}_t > 0$ 和 w_{tt} 都与 c_{tt}具有相同的符号，因此 β_{tt} 也具有相同的符号。 ■

引理 9.15 表明，新变量以“正确”的方向进入 LARS 模型——也就是说，其系数符号与其和残差的当前相关性相匹配。这一性质较弱，但至关重要；与 LASSO 所要求的更严格的符号一致性相比（见条件（9.75））。正如我们将看到的，这种一致性是 LARS 与 LASSO 之间深层联系的关键要素。

约束 LASSO 问题的 LARS

LARS 沿着最自然的折中方向前进——即（9.49c）中定义的等角向量——而约束 LASSO 则对这一等角策略施加了额外限制。

现在我们简要说明，只需对标准 LARS 算法进行少许修改，即可生成整个 LASSO 解集合（9.50）∼(9.55).

▶ LARS–LASSO 中拟合值的分段线性更新。我们首先讨论 LARS 与 LASSO 之间的关系。令 $\boldsymbol{\beta} = \boldsymbol{\beta}(\Sigma)$ 表示约束 LASSO 问题 (L_C) 的一个解，并将相应的拟合值定义为 $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{y}}(\Sigma) = \boldsymbol{X}\boldsymbol{\beta}(\Sigma)$，其中 $\Sigma \geq 0$ 是满足 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$的约束半径。由于损失函数 $f(\widehat{\boldsymbol{\beta}}) = \frac{1}{2}\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{y}\|_2^2$和约束 $g(\widehat{\boldsymbol{\beta}}) = \|\widehat{\boldsymbol{\beta}}\|_1$都是关于 $\boldsymbol{\beta}$ 的凸函数——且 f 是严格凸的——因此根据标准凸分析可知，$\boldsymbol{\beta}(\Sigma)$ 和$\widehat{\boldsymbol{y}}(\Sigma)$ 都是约束半径 Σ的唯一连续函数。

对于固定的 Σ，将活跃集定义为

$$\mathbb{S} \triangleq \{i \mid \widehat{\beta}_i(\Sigma) \neq 0\}.$$

事实上，$\mathbb{S}$ 也是在 LARS–LASSO 计算中确定等角方向 $\boldsymbol{u}_{\mathbb{S}}$ 的活跃集。

我们希望刻画 LASSO 解的轨迹 $\widehat{\boldsymbol{\beta}}(\Sigma)$，或等价地，$\widehat{\boldsymbol{y}}(\Sigma)$ 随着 Σ 从 0 增加到其最大有效值时的轨迹。设 $\mathbb{M}$ 为 Σ 轴上的一个开区间，其下确界为 $\Sigma_0 \triangleq \inf(\mathbb{M})$，在此区间内，非零 LASSO 系数 $\widehat{\beta}_i(\Sigma)$ 的集合 $\mathbb{S}$ 保持不变：

$$\mathbb{S} \equiv \{i \mid \widehat{\beta}_i(\Phi) \neq 0\}, \quad \text{for all } \Phi \in \mathbb{M}.$$

引理 9.16：（LARS–LASSO 中拟合值的分段线性更新）对于 $\Sigma \in \mathbb{M}$，LASSO 估计值 $\widehat{\boldsymbol{y}}(\Sigma)$ 满足

$$\widehat{\boldsymbol{y}}(\Sigma) = \widehat{\boldsymbol{y}}(\Sigma_0) + \sigma_{\mathbb{S}}(\Sigma - \Sigma_0)\boldsymbol{u}_{\mathbb{S}} \tag{9.66}$$

其中 $\boldsymbol{u}_{\mathbb{S}}$ 是等角向量 $\boldsymbol{u}_{\mathbb{S}} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{w}_{\mathbb{S}}$, $\boldsymbol{w}_{\mathbb{S}} = \sigma_{\mathbb{S}}\boldsymbol{G}_{\mathbb{S}}^{-1}\mathbf{1}_{\mathbb{S}}$。等价地，关于 $\widehat{}$ 非零回归系数 $\boldsymbol{\beta}_{\mathbb{S}}(\Sigma)$,

$$\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma) = \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_0) + \boldsymbol{S}_{\mathbb{S}}\sigma_{\mathbb{S}}(\Sigma - \Sigma_0)\boldsymbol{w}_{\mathbb{S}}, \tag{9.67}$$

其中， $\boldsymbol{S}_{\mathbb{S}} \triangleq \operatorname{diag}(\boldsymbol{s}_{\mathbb{S}})$ 是对角元素为 $s_i = \operatorname{sign}(\widehat{\beta}_i)$、 $i \in \mathbb{S}$ 和 $\boldsymbol{s}_{\mathbb{S}}$ 的对角矩阵，而 [a] 是包含这些符号函数元素的向量。

[a]。矩阵 $\boldsymbol{S}_{\mathbb{S}}$ 之所以出现，是因为根据定义（9.49a）和（9.52），我们有 $\widehat{\boldsymbol{y}}(\Sigma) = \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\Sigma) = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma)$。

$\widehat{}$ $\widehat{}$注意，（9.66）意味着最大绝对相关性 $C(\Sigma)$ 等于 $C(\Sigma_0)$ _ $\sigma_{\mathbb{S}}^2$ $\Sigma - \Sigma_0$ 9 49 $\widehat{C}$ Σ ()，根据（.d），因此（）是关于 LASSO 参数 Σ 的分段线性递减函数。

证明 [引理 9。16] 该引理指出，对于 Σ 属于 $\mathbb{M}$ 且活跃集 $\mathbb{S}$ 及符号函数 $\boldsymbol{s}_{\mathbb{S}}$保持不变时， $\widehat{\boldsymbol{y}}(\Sigma)$ 沿由 $\boldsymbol{u}_{\mathbb{S}}$ 确定的等角向量 $\mathbb{S}$线性移动。由于 $\frown\boldsymbol{\beta}(\Sigma)$ 满足约束 LASSO 问题（L_C），且其支撑集为 $\mathbb{S}$，它也使.....最小化

$$f(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) = \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}\right\|_2^2$$

满足约束

$$\sum_{i\in\mathbb{S}} s_i\widehat{\beta}_i = \Sigma \qquad \text{and} \qquad \operatorname{sign}(\widehat{\beta}_i) = s_i \quad \text{for } i \in \mathbb{S}. \tag{9.68}$$

只要 Σ 小于 $\sum|\widehat{\beta}_i|$ 对于完整的 p变量 OLS 解 $\widehat{\boldsymbol{\beta}}_p$，约束 LASSO 问题（$\mathrm{L}_C$）中的不等式约束就可以替换为 $\|\widehat{\boldsymbol{\beta}}\|_1 = \Sigma$。此外，最小化点 $\boldsymbol{\beta}_{\mathbb{S}}(\Sigma$严格位于单纯形（9.68）内部这一事实，结合 $\widehat{}$的严格凸性， $f(\boldsymbol{\beta}_{\mathbb{S}}$意味着我们可以去掉（9.68）中的第二个条件，从而使 $\widehat{}$ $\boldsymbol{\beta}_{\mathbb{S}}(\Sigma)$ 成为其解

$$\min \quad \{f(\widehat{\boldsymbol{\beta}}_{\mathbb{S}})\} \quad \text{s.t.} \quad \sum_{i\in\mathbb{S}} s_i\widehat{\beta}_i = \Sigma. \tag{9.69}$$

引入拉格朗日乘子 λ，式（9.69）变为

$$\min \quad \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}\right\|_2^2 + \lambda\sum_{i\in\mathbb{S}} s_i\widehat{\beta}_i.$$

对 $\boldsymbol{\beta}_{\mathbb{S}}$ 求导并令梯度为零，得到

$$-\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^{\top}(\boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}}\widehat{\boldsymbol{\beta}}_{\mathbb{S}}) + \lambda\boldsymbol{S}_{\mathbb{S}}\mathbf{1}_{\mathbb{S}} = \mathbf{0} \tag{9.70}$$

$$\Longrightarrow \quad \boldsymbol{\beta}_{\mathbb{S}} = (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}(\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{y} - \lambda\boldsymbol{s}_{\mathbb{S}}). \tag{9.71}$$

现在考虑两个值 Σ_1、$\Sigma_2 \in \mathbb{M}$，其中 $\Sigma_0 < \Sigma_1 < \Sigma_2$，并令 $\lambda_1 > \lambda_2$ 为相应的拉格朗日乘子（因为较大的 Σ意味着较小的惩罚权重）。将 Σ_1 和 Σ_2 的最优性条件相减，得到

$$\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}}\big(\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_2) - \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_1)\big) = (\lambda_1 - \lambda_2)\mathbf{1}_{\mathbb{S}},$$

这意味着

$$\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_2) - \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_1) = (\lambda_1 - \lambda_2)\boldsymbol{S}_{\mathbb{S}}\boldsymbol{G}_{\mathbb{S}}^{-1}\mathbf{1}_{\mathbb{S}}.$$

不过，我们也可以根据 LASSO 的定义证明 $\boldsymbol{s}_{\mathbb{S}}^{\top}[\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_2) - \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_1)] = \Sigma_2 - \Sigma_1$：

$\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_2) - \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_1)$ 与 $\Sigma_2 - \Sigma_1$ 之间的关系式：
对于活跃集，$\Sigma = \|\boldsymbol{\beta}\|_1 = \sum_{i\in\mathbb{S}} |\beta_i| = \boldsymbol{s}_{\mathbb{S}}^\top \boldsymbol{\beta}_{\mathbb{S}}$。代入（9.71）：

$$\Sigma = \boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}(\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{y} - \lambda \boldsymbol{s}_{\mathbb{S}}).$$

定义 $a \triangleq \boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{y}$ 和 $b \triangleq \boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}$。那么

$$\Sigma = a - \lambda b \quad \Longrightarrow \quad \lambda = \frac{a - \Sigma}{b}.$$

因此，λ 是关于 Σ的仿射（线性）函数，而 $\mathbb{S}$ 和 $\boldsymbol{s}_{\mathbb{S}}$ 保持不变。对（9.71）关于 Σ求导：$\frac{\partial \boldsymbol{\beta}_{\mathbb{S}}}{\partial \Sigma} = -(\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}} \cdot \frac{\partial \lambda}{\partial \Sigma}$。由 $\lambda = (a - \Sigma)/b$，可得 $\frac{\partial \lambda}{\partial \Sigma} = -\frac{1}{b}$。因此

$$\frac{\partial \boldsymbol{\beta}_{\mathbb{S}}}{\partial \Sigma} = \frac{(\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}{\boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}, \qquad \frac{\partial \boldsymbol{\beta}_{\mathbb{S}^c}}{\partial \Sigma} = \boldsymbol{0}.$$

如果 Σ_1 和 Σ_2 位于同一区域内，且活跃集 $\mathbb{S}$ 和符号向量 $\boldsymbol{s}_{\mathbb{S}}$ 保持不变，那么

$$\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = (\Sigma_2 - \Sigma_1)\frac{(\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}{\boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{S}_{\mathbb{S}}\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}}\boldsymbol{S}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}. \tag{9.72}$$

也就是说，$\boldsymbol{\beta}$ 沿固定方向 $\boldsymbol{v} \triangleq \frac{(\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}{\boldsymbol{s}_{\mathbb{S}}^\top (\boldsymbol{X}_{\mathbb{S}}^\top \boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{s}_{\mathbb{S}}}$随 Σ 线性变化。

因此，

$$\Sigma_2 - \Sigma_1 = (\lambda_1 - \lambda_2)\boldsymbol{s}_{\mathbb{S}}^\top \boldsymbol{S}_{\mathbb{S}}\boldsymbol{G}_{\mathbb{S}}^{-1}\boldsymbol{1}_{\mathbb{S}} = (\lambda_1 - \lambda_2)\boldsymbol{1}_{\mathbb{S}}^\top \boldsymbol{G}_{\mathbb{S}}^{-1}\boldsymbol{1}_{\mathbb{S}} = (\lambda_1 - \lambda_2)\sigma_{\mathbb{S}}^{-2},$$

因而

$$\widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_2) - \widehat{\boldsymbol{\beta}}_{\mathbb{S}}(\Sigma_1) = \boldsymbol{S}_{\mathbb{S}}\sigma_{\mathbb{S}}^2(\Sigma_2 - \Sigma_1)\boldsymbol{G}_{\mathbb{S}}^{-1}\boldsymbol{1}_{\mathbb{S}} = \boldsymbol{S}_{\mathbb{S}}\sigma_{\mathbb{S}}(\Sigma_2 - \Sigma_1)\boldsymbol{w}_{\mathbb{S}}. \tag{9.73}$$

利用 $\Sigma_2 = \Sigma$ 和 $\Sigma_1 \to \Sigma_0$，由 9.67 的连续性可得（$\widehat{\boldsymbol{\beta}}(\Sigma)$，从而推出所需结果（9.66）。 ■

引理 9.16 (9.66) 表明，对于 $\Sigma \in \mathbb{M}$，约束 LASSO 回归的估计值$\widehat{\boldsymbol{y}}(\Sigma)$ 沿着 LARS 等角方向 $\boldsymbol{u}_{\mathbb{S}}$ 移动，该方向定义见（9.49c）。此外，LASSO 解 $\widehat{\boldsymbol{\beta}}(\Sigma)$ 位于菱形凸多胞体的边界上（即半径为 Σ 的“ℓ_1- 范数”球）：

$$\mathbb{B}_1[\boldsymbol{0}, \Sigma] = \left\{ \boldsymbol{\alpha} \mid \sum |\alpha_i| \le \Sigma \right\}, \tag{9.74}$$

该集合 $\mathbb{B}_1[\boldsymbol{0}, \Sigma]$ 随着 Σ 增大而单调扩张。引理 9.16 (9.67) 进一步表明，对于 $\Sigma \in \mathbb{M}$，回归系数 $\widehat{\boldsymbol{\beta}}(\Sigma$沿多面体的边 $\mathbb{S}$ 线性移动，也就是说，移动于满足 $\beta_i = 0$（对所有 $i \notin \mathbb{S}$）的面上。

符号约束。设 $\boldsymbol{\beta}$ 是约束 LASSO 问题（L_C）的一个解，且令 $\widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$ 表示相应的拟合值。很容易证明

对于任意非零分量 β_i，其符号函数必须与当前相关性$\widehat{c}_i = \boldsymbol{x}_i^\top (\boldsymbol{y} - \widehat{\boldsymbol{y}})$ 的符号函数 s_i 一致，

$$\operatorname{sign}(\widehat{\beta}_i) = \operatorname{sign}(\widehat{c}_i) = s_i. \tag{9.75}$$

引理 9.17：（LARS–LASSO 的符号约束）对于任意 LASSO 解 $\boldsymbol{\beta}$，当前相关性满足

$$\widehat{c}_i = \widehat{C} \cdot \operatorname{sign}(\widehat{\beta}_i), \quad \text{for } i \in \mathbb{S}, \tag{9.76}$$

其中，$\widehat{c}_i$ 是当前相关性 $\boldsymbol{x}_i^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}) = \boldsymbol{x}_i^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$。特别地，这意味着

$$\operatorname{sign}(\widehat{\beta}_i) = \operatorname{sign}(\widehat{c}_i), \quad \text{for } i \in \mathbb{S}. \tag{9.77}$$

证明 [引理9。17] 这直接由驻点条件（9.70）得到，注意到左侧的第 i 个元素为 $\widehat{c}_i$，而右侧为 $\lambda \cdot \operatorname{sign}(\beta_i)$对于 $i \in \mathbb{S}$。此外，$\lambda = |\widehat{c}_i| = C$由此即完成证明。

■

标准最小角回归算法不强制执行符号约束（9.75），因此不能直接用于计算 LASSO 解。不过，可以很容易地对其进行修改，以纳入这一约束。

▶ **LASSO 修正**。假设我们刚完成一个 LARS 步，得到新的活跃集 $\mathbb{S}$，它由之前的活跃集 $\mathbb{S}_-$得出，如式（9.51）所示，并且当前的 LARS 估计值$\widehat{\boldsymbol{y}}_{\mathbb{S}_-}$ 恰好是一个有效的 LASSO 解 $\widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$。在这个 LARS 步开始时，我们知道（假设）符号约束（9.75）成立。然而，沿当前等角方向继续前进后，这一条件可能不再成立。为分析这一点，定义

$$\boldsymbol{w}_{\mathbb{S}} = \sigma_{\mathbb{S}} \boldsymbol{G}_{\mathbb{S}}^{-1} \mathbf{1}_{\mathbb{S}}, \tag{9.78}$$

一个长度等于 $|\mathbb{S}|$大小的向量。现在构造 p维方向向量 $\boldsymbol{d}$，令

$$\widehat{d}_i = \begin{cases} s_i w_{\mathbb{S},i}, & i \in \mathbb{S}; \\ 0, & i \notin \mathbb{S}. \end{cases}$$

沿最小角回归路径（9.56）的正 γ 方向移动时，系数路径为

$$\boldsymbol{y}(\gamma) = \boldsymbol{X}\boldsymbol{\beta}(\gamma), \quad \text{where } \beta_i(\gamma) \triangleq \widehat{\beta}_i + \gamma \widehat{d}_i, \text{ for } i \in \mathbb{S}, \tag{9.79}$$

另见式（9.65）。因此，β_i（γ）将在

$$\gamma_i \triangleq -\widehat{\beta}_i / \widehat{d}_i; \tag{9.80}$$

首次发生符号变化时

$$\widetilde{\gamma} \triangleq \min_{\gamma_i > 0} \{\gamma_i\}, \tag{9.81}$$

例如，对于协变量 β_{i_+}。如果不存在这样的 γ_i（即，所有 $\gamma_i \le 0$），我们定义 $\widetilde{\gamma} = \infty$。

如果 $\widetilde{\gamma} < \widehat{\gamma}_{\mathbb{S}}$——其中 $\widehat{\gamma}_{\mathbb{S}}$ 是下一个加入活跃集的变量所对应的步长，如（9.55）中所定义的——那么路径 $\beta_i(\gamma)$ 对于任意 $\gamma > \widetilde{\gamma}$ 都不再是有效的 LASSO 解，因为符号约束（9.75）将被违反：$\beta_i(\gamma)$ 改变符号，但相应的相关性 $c_i(\gamma)$ 却没有改变。这是因为连续函数 $c_i(\gamma)$ 在单个 LARS 步内不能改变符号，因为 $|c_i(\gamma)| = C_{\mathbb{S}} - \gamma\sigma_{\mathbb{S}} > 0$ 根据（9.58）成立。因此，计算 LASSO 路径时需要作如下修正：

注9.18（LASSO 修正）。如果 $\tilde{\gamma} < \hat{\gamma}_{\mathbb{S}}$，则停止正在进行的 LARS 步，并从下一个等角方向的计算中移除索引 i_+（即取得 $\tilde{\gamma}$ 的索引）。具体而言，将拟合值和活跃集更新为

$$\hat{\boldsymbol{y}}_{\mathbb{S}} = \hat{\boldsymbol{y}}_{\mathbb{S}_-} + \tilde{\gamma}\boldsymbol{u}_{\mathbb{S}} \qquad \text{and} \qquad \mathbb{S}_+ = \mathbb{S} - \{i_+\} \tag{9.82}$$

而不是执行标准更新（9.54）。否则，如果 $\tilde{\gamma} \geq \hat{\gamma}_{\mathbb{S}}$，则执行通常的 LARS 更新（9.54）。

在原始 LARS 算法中，活跃集 $\mathbb{S}$ 单调增长。然而，LASSO 修正允许在变量的系数变为零时将其从 $\mathbb{S}$ 中移除。LASSO 修正还假设满足“一次一个”的条件，这意味着每一步最多只有一个变量被添加到活跃集或从中移除。对于连续（定量）数据，这一条件通常成立；必要时，也可以通过对响应向量 $\boldsymbol{y}$ 添加无穷小的随机扰动来确保该条件成立。完整流程总结于算法7中。

定理 9.19：（LARS–LASSO 关系）在上述 LASSO 修正下，并假设满足“一次一个”的条件，LARS 算法会生成完整的 LASSO 解路径。

证明 参见Efron 等人(2004年)。 ■

LARS–LASSO 关系

现在简要讨论定理 9.19，该定理确立了改进的 LARS 算法与约束 LASSO 问题之间的等价性。这种等价性源于这样一个事实：改进的 LARS 算法逐步执行时所生成的系数估计序列（即所谓的解路径），恰好与约束参数 Σ 从 0 增加到 $\left\|\widehat{\boldsymbol{\beta}}^{OLS}\right\|_1$（即 ℓ_1普通最小二乘（OLS）解的范数）时，约束 LASSO 问题的解集完全一致。

标准的约束 LASSO 问题为 $\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$，满足约束 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$，其中 Σ 是一个非负标量，用于控制系数绝对值之和的总“预算”。随着 Σ 从 0 开始增加，解 $\boldsymbol{\beta}(\Sigma)$ 的变化如下：

- 当 $\Sigma = 0$ 时，唯一的解是 $\boldsymbol{\beta} = \mathbf{0}$。^随着 Σ 略微增大，最优解 $\boldsymbol{\beta}(\Sigma)$ 将具有一
- 个或少数几个非零系数，这些系数的选择会使每单位“ℓ_1-成本”所能带来的最小二乘损失降低幅度尽可能大。^随着 $\Sigma \to \infty$，约束变为非活跃，且 $\boldsymbol{\beta}(\Sigma)$ 趋近于 OLS 解。
- ^解路径 $\boldsymbol{\beta}(\Sigma)$ 是 Σ 变化时所有最优解构成的集合。

修正的 LARS 与约束 LASSO 之间的等价性，源于 LASSO 问题的最优性条件（KKT 条件），以及 LARS 在每一步如何满足^这些条件。对于给定约束 $\boldsymbol{\beta}$ 下的最优解 Σ，必须满足以下 KKT 条件（定理 2.115）：

1. 次梯度条件。 最小二乘损失的梯度 $(-\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}))$ 必须属于 $\lambda\left\|\widehat{\boldsymbol{\beta}}\right\|_1$ 的次微分，其中 λ 是拉格朗日对偶变量。由此可得：

算法 7 带 LASSO 修正的 LARS 算法（Efron 等人，2004年）

要求： 数据向量 $\boldsymbol{y} \in \mathbb{R}^n$ 和输入矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$。

初始化： $\mathbb{S}_0 = \varnothing$, $\widehat{\boldsymbol{y}} = \boldsymbol{0}$ 和 $\boldsymbol{X}_{\mathbb{S}_0} = \boldsymbol{X}$; **对于** $t = 1, 2, \ldots$ **执行** 计算当前相关向量 $\widehat{\boldsymbol{c}}_t = \boldsymbol{X}^\top(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{t-1})$; 用 $C = \max\{|\widehat{c}_{t1}|, |\widehat{c}_{t2}|, \ldots\}$ 更新活跃集 $\mathbb{S}_t = \mathbb{S}_{t-1} \cup \{i \mid |\widehat{c}_{ti}| = C\}$，且 $\widehat{c}_{ti}$ 是 $\widehat{\boldsymbol{c}}_t$ 的第 i 个元素；更新输入矩阵 $\boldsymbol{X}_{\mathbb{S}_t} = [\ldots, s_i \boldsymbol{x}_i, \ldots]_{i \in \mathbb{S}_t}$，其中 $s_i = \operatorname{sign}(\widehat{c}_{ti})$; 求当前最小角的方向：

$$\boldsymbol{G}_{\mathbb{S}_t} = \boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{X}_{\mathbb{S}_t} \in \mathbb{R}^{|\mathbb{S}_t| \times |\mathbb{S}_t|}, \qquad \sigma_{\mathbb{S}_t} = (\boldsymbol{1}_t^\top \boldsymbol{G}_{\mathbb{S}_t}^{-1} \boldsymbol{1}_t)^{-1/2},$$
$$\boldsymbol{w}_{\mathbb{S}_t} = \sigma_{\mathbb{S}_t} \boldsymbol{G}_{\mathbb{S}_t}^{-1} \boldsymbol{1}_t \in \mathbb{R}^{|\mathbb{S}_t|}, \qquad \boldsymbol{u}_t = \boldsymbol{X}_{\mathbb{S}_t} \boldsymbol{w}_{\mathbb{S}_t} \in \mathbb{R}^n.$$

计算 $\boldsymbol{z} = \boldsymbol{X}^\top \boldsymbol{u}_t = [z_1, z_2, \ldots, z_p]^\top$ 并估计系数向量

$$\widehat{\boldsymbol{\beta}}_t = (\boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{X}_{\mathbb{S}_t})^{-1} \boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{y} = \boldsymbol{G}_{\mathbb{S}_t}^{-1} \boldsymbol{X}_{\mathbb{S}_t}^\top \boldsymbol{y}.$$

根据（9.55），计算

$$\widehat{\gamma} = \min_{i \in \mathbb{S}_t^c}{}^{+} \left\{ \frac{C - \widehat{c}_{ti}}{\sigma_{\mathbb{S}_t} - z_i}, \frac{C + \widehat{c}_{ti}}{\sigma_{\mathbb{S}_t} + z_i} \right\}, \qquad \widetilde{\gamma} = \min_{i \in \mathbb{S}_t} \left\{ -\frac{z_i}{s_i w_{\mathbb{S}_t, i}} \right\}^{+},$$

其中 $w_{\mathbb{S}_t,\, i}$ 是 $\boldsymbol{w}_{\mathbb{S}_t} = [w_1, w_2, \ldots]^\top$ 的 i 个元素，且 $(\cdot)^+$ 表示正最小项。如果不存在正项，则取 $\min^+\{\cdot\} = \infty$;

如果 $\widetilde{\gamma} < \widehat{\gamma}$ **then**

拟合向量 $\widehat{\boldsymbol{y}}_t$ 和活跃集 $\mathbb{S}_t$ 修改如下：

$$\widehat{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \widetilde{\gamma} \boldsymbol{u}_t, \qquad \mathbb{S}_t = \mathbb{S}_t \setminus \{i_+\},$$

其中，被移除的索引 i_+ 是满足 $\widetilde{\gamma}$ 为最小值的索引 $i \in \mathbb{S}_t$；

else

拟合向量 $\widehat{\boldsymbol{y}}_t$ 和活跃集 $\mathbb{S}_t$ 按如下方式修改：

$$\widehat{\boldsymbol{y}}_t = \widehat{\boldsymbol{y}}_{t-1} + \widehat{\gamma} \boldsymbol{u}_t, \qquad \mathbb{S}_t = \mathbb{S}_t \cup \{i_+\},$$

其中，添加的索引 i_+ 是满足 $\widehat{\gamma}$ 为最小值的索引 $i \in \mathbb{S}_t^c$；**结束 if** 若满足某个停止准则，则退出；**结束 for** **返回** $\boldsymbol{\beta} = \boldsymbol{\beta}_t$;

- 对于任意满足 $\widehat{\beta}_i \neq 0$ 的 i：$\left|\boldsymbol{x}_i^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\right| = \lambda$（预测变量 i 与残差的绝对相关性恰好为 λ）。对于任意满足 $\widehat{\beta}_i = 0$ 的 i：$\left|\boldsymbol{x}_i^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\right| \leq \lambda$（绝对相
- 关性小于或等于 λ）。

2. 互补松弛。$\lambda(\left\|\widehat{\boldsymbol{\beta}}\right\|_1 - \Sigma) = 0$。如果 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1 < \Sigma$，则 $\lambda = 0$（无正则化）。如果 $\lambda > 0$，则 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1 = \Sigma$（约束处于活跃状态）。

修正的 LARS 算法通过以下方式满足这些 KKT 条件：

- 相关性相等。 每个 LARS 步开始时，算法都会确保活跃集中的所有预测变量与当前残差向量的**绝对相关性相等** 。这符合 KKT 条件：非零系数必须满足 $\left|\boldsymbol{x}_i^\top \boldsymbol{r}\right| = \lambda$，其中 λ，而 $\boldsymbol{r}$ 表示残差向量。最大相关性。 算法还确保没有非活跃预测变量与残差向量的绝对相关性高于活跃预测变量。这符合 KKT 条件 $\left|\boldsymbol{x}_i^\top \boldsymbol{r}\right| \leq \lambda$，其中 i 不在活跃集中。递增 Σ。 随着 LARS 每步推进并调整系数， ℓ_1范数 $\|\boldsymbol{\beta}\|_1$会增大。LARS 改变方向的点（即新变量进入或已有变量离开时）对应于 $\|\boldsymbol{\beta}\|_1 = \Sigma_t$ 的特定取值。随着 Σ 增大而变化的路径。 LARS 生成的分段线性路径，以步数或 $\|\boldsymbol{\beta}\|_1$ 为参数，在路径上的每个点都满足约束 LASSO 问题的 KKT 条件。KKT 条件中的“λ”由每一步中活跃预测变量共同的绝对相关性值隐式确定。

重要的是，无论是标准的还是修正的 LARS 算法（算法7），都不将约束参数 Σ 作为显式输入。相反，该算法依据预测变量与残差之间的相关性，沿几何步骤推进。尽管如此，在每个 LARS 步 $\curlywedge t$结束时，得到的系数向量 $\boldsymbol{\beta}_t$ 恰好是 $\Sigma_t = \left\|\widehat{\boldsymbol{\beta}}_t\right\|_1$对应的约束 LASSO 问题的解：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_1 \leq \Sigma_t.$$

换言之，LARS 路径就是LASSO 解路径。尽管 Σ 并非算法的输入，但它是为路径建立索引的自然参数：LARS 轨迹上的每个点都对应 $\Sigma = \|\boldsymbol{\beta}\|_1$ 的唯一取值。因此，完整的正则化路径会自动得到，随后便可选出最优的Σ——例如，通过交叉验证。

9.3. 拉格朗日套索问题算法

我们介绍了几种求解约束 LASSO 问题（L_C）的算法。现在转而讨论 LASSO 问题的拉格朗日形式（或惩罚形式）。这种形式更常用于机器学习，因为它简单明了：它将问题表述为只有一个目标函数的无约束优化，相较于（L_C）这类约束问题，在算法上更容易处理。此外，它还能自然地融入正则化框架。

在机器学习中，正则化是防止过拟合并促使模型具备理想属性（例如稀疏性）的一项基本技术，参见第3.2.3节。拉格朗日 LASSO 明确刻画了数据保真度（由损失衡量）与模型复杂度（通过 ℓ_1范数惩罚）之间的权衡。

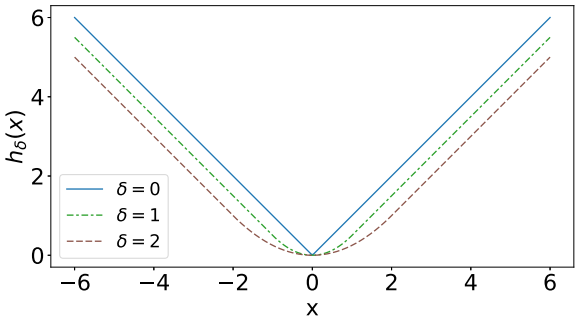
拉格朗日 LASSO 的标准形式（L_L）（第126页）为：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \quad F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\,, \tag{9.83}$$

其中 $g(\boldsymbol{\beta}) \triangleq \lambda\|\boldsymbol{\beta}\|_1$。 [2] 由于 f 和 g 都是凸函数（习题2.35），因此可以考虑将梯度下降作为一种直接的方法。然而，目标函数

2。再次注意，对于稀疏信号恢复问题，拉格朗日 LASSO 等价于惩罚式 ℓ_1-最小化（$\mathrm{P}_{1,\,\sigma}$）。

图 9.4： 不同 δ 取值下的 Huber 损失函数 $h_\delta(x)$ 的示意图。



$F(\boldsymbol{\beta})$ 在 $\boldsymbol{\beta}$ 的任一分量为零的点处都不光滑（这是由于 ℓ_1-范数不可微所致）。因此，经典梯度并非处处存在，无法直接应用标准梯度下降。

对此，可以采用次梯度方法。虽然 f 是光滑的——具体而言，$\left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2$-光滑（见例 2.72）——且 g 是 Lipschitz 连续的，但它们的和 $F = f + g$ 既不光滑也不可微。因此，对 F 应用次梯度方法通常会表现出较慢的收敛速度，且缺乏强有力的理论保证（例如，在没有额外结构的情况下，不存在次线性甚至加速收敛速率）。

9.3.1 Huber损失平滑

处理 LASSO 目标函数非光滑性的另一种方法是认识到，ℓ_1 范数 $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^p |\beta_i|$ 是绝对值函数之和——每个绝对值函数在零点处都不可微。如果用光滑近似函数替换每个绝对值，那么就可以应用标准的基于梯度的优化方法。一种常用的近似函数是***Huber*** 损失函数，其定义如下：

$$h_\delta(x) = \begin{cases} \frac{1}{2\delta}x^2, & |x| < \delta; \\ |x| - \frac{\delta}{2}, & \text{otherwise.} \end{cases} \tag{9.84}$$

当 $\delta \to 0$时，光滑函数 $h_\delta(x)$ 趋近于绝对值函数 $|x|$。图9.4 展示了 $h_\delta(x)$ 在 δ的不同取值下的图像。

将 $|\beta_i|$ 替换为 $h_\delta(\beta_i)$ 后，我们得到一个光滑的 *LASSO* 问题：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \quad F_\delta(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda H_\delta(\boldsymbol{\beta}), \quad \text{with } H_\delta(\boldsymbol{\beta}) \triangleq \sum_{i=1}^p h_\delta(\beta_i), \tag{9.85}$$

其中，δ 是给定的平滑参数。目标函数 $F_\delta(\boldsymbol{\beta})$ 现在是连续可微的，其梯度可计算为：

$$\nabla F_\delta(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \lambda \nabla H_\delta(\boldsymbol{\beta}), \tag{9.86}$$

其中，$\nabla H_\delta(\boldsymbol{\beta})$ 按分量定义为：

$$(\nabla H_\delta(\boldsymbol{\beta}))_i = \begin{cases} \operatorname{sign}(\beta_i), & |\beta_i| > \delta; \\ \beta_i/\delta, & |\beta_i| \le \delta, \end{cases} \quad i \in \{1, 2, \ldots, p\}. \tag{9.87}$$

由于 $F_\delta(\boldsymbol{\beta})$ 中的两项都是凸函数，且梯度 $\nabla F_\delta(\boldsymbol{\beta})$ 是 Lipschitz 连续的，因此函数 $F_\delta(\boldsymbol{\beta})$ 是 L_b-光滑的，其 Lipschitz 常数为

$$L_b = \left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2 + \frac{\lambda}{\delta}$$

（参见例2.72 和习题2.79）。因此，根据问题9.2，采用恒定步长 $\eta = 1/L_b$ 的梯度下降法保证收敛到 F_δ的全局最小值。

然而，当 δ 非常小时——也就是说，平滑后的问题与原始 LASSO 问题非常接近时——由于 λ/δ 项，Lipschitz 常数 L_b 会变大。在这种情况下，所需的步长 $\eta = 1/L_b$ 会非常小，这可能导致实践中的收敛速度变慢。因此，使用这种平滑方法时，仔细调整 δ（或采用自适应步长策略）至关重要。

9.3.2 近端梯度与主化最小化方法

鉴于拉格朗日 LASSO 问题（ f 和 g （9.83）中的两个函数都是凸函数，其中 f 可微，而 g 不可微，问题（ 9.83 ）可视为一个复合模型。我们已经介绍过用于处理此类复合问题的近端梯度法（第 9.2.2节）。尽管也有其他方法，例如广义条件梯度法和广义镜像下降方法（Lu，2025年），但这些方法通常要求在每次迭代时求解如下形式的子问题：最小值点集合$_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\langle \nabla f(\boldsymbol{\beta}^{(t)})\ \boldsymbol{\beta}\right\rangle + g(\boldsymbol{\beta})$。对于 LASSO，这个子问题的求解难度基本与原问题（9.83）相当。因此，近端梯度法仍然是最实用、最高效的选择。

根据算法4，应用于拉格朗日 LASSO 问题（9.83）的近端梯度方法在第 t 次迭代时的更新规则为：

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}); \tag{9.88a}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t+1)}) = \mathcal{T}_{\eta_t\lambda}(\boldsymbol{\alpha}^{(t+1)}) = \left[\left|\boldsymbol{\alpha}^{(t+1)}\right| - \eta_t\lambda\mathbf{1}\right]_+ \circ \mathrm{sign}(\boldsymbol{\alpha}^{(t+1)}), \tag{9.88b}$$

其中 $\mathcal{T}_\lambda(\cdot)$ 表示软阈值算子（例 2.104）。这一两步过程首先对光滑损失 f 执行标准梯度下降步，然后进行收缩（阈值处理）操作，以促进迭代点的稀疏性。使用恒定步长 $\eta_t = \frac{1}{\|\boldsymbol{X}^\top\boldsymbol{X}\|_2}$ 时，该方案的收敛性由定理 9.12 保证，因为 f 是 $\|\boldsymbol{X}^\top\boldsymbol{X}\|_2$-光滑的（例 2.72）。

注意，这一更新与约束 LASSO 问题（9.38）的近端梯度步非常相似；不过，在约束形式中，正则化参数 λ 并非固定不变，而是会随着约束水平的变化，在每次迭代中自适应地调整。

最后，当近端梯度法专门应用于拉格朗日 LASSO 问题时，文献中通常称之为迭代收缩阈值算法（*ISTA*）——这个名称反映了每次迭代都会将较小的系数向零收缩，从而促进稀疏解的产生。

下界最大化方法

接下来，我们进一步介绍一类方法，它们称为下界最大化或上界最小化（*MM*）算法。这类算法对于非凸问题尤其有用，属于辅助变量方法：它们引入一个辅助函数，在局部为原目标函数提供上界（即构造其上界），从而将困难的最小化问题转化为一系列较简单的问题。

尽管 MM 方法适用于约束和无约束情形，这里我们针对如下形式的无约束问题来描述它们：$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} f(\boldsymbol{\beta})$，其中 $f\colon\ \mathbb{R}^p \to \mathbb{R}$ 可能是非凸的。

定义 9.20（辅助函数（上界函数））。如果函数 $\Psi(\boldsymbol{\beta},\ \boldsymbol{\beta})$ 满足以下条件：[a]，则称其为 $F(\boldsymbol{\beta})$ 的辅助函数（或称其为 F 在 $\boldsymbol{\beta}$ 处的上界函数）。

$$\Psi(\boldsymbol{\beta},\widetilde{\boldsymbol{\beta}}) \geq F(\boldsymbol{\beta}) \quad \text{and} \quad \Psi(\boldsymbol{\beta},\boldsymbol{\beta}) = F(\boldsymbol{\beta}). \tag{9.89}$$

换句话说，辅助函数 $\Psi(\cdot,\ \boldsymbol{\beta})$ 是 $F(\cdot)$ 的上界，并且该上界在 $\boldsymbol{\beta}=\boldsymbol{\beta}$ 处取等。

[a] $\boldsymbol{\beta}$ 可以是标量、向量或矩阵。

引理 9.21：（辅助函数下的非增性）如果 Ψ是 F的辅助函数，则由下式生成的序列

$$\boldsymbol{\beta}^{(t+1)} = \arg\min_{\boldsymbol{\beta}} \Psi(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}), \quad \text{for } t=0,1,2,\ldots \tag{9.90}$$

满足 $F(\boldsymbol{\beta}^{(t+1)}) \leq F(\boldsymbol{\beta}^{(t)})$。

证明 [引理 9.21] 根据辅助函数的定义和更新规则：$F(\boldsymbol{\beta}^{(t+1)}) \leq \Psi(\boldsymbol{\beta}^{(t+1)},\ \boldsymbol{\beta}^{(t)}) \leq \Psi(\boldsymbol{\beta}^{(t)},\ \boldsymbol{\beta}^{(t)}) = F(\boldsymbol{\beta}^{(t)})$。 ■

MM 算法通过初始化 $\boldsymbol{\beta}^{(0)}$，并反复应用更新式（9.90）来执行。目标函数值单调递减：

$$\boldsymbol{\beta}^{(t+1)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \Psi\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right), \quad \text{for } t=0,1,2,\ldots \tag{9.91}$$

注意，$F\left(\boldsymbol{\beta}^{(t+1)}\right) = F\left(\boldsymbol{\beta}^{(t)}\right)$ 仅当 $\boldsymbol{\beta}^{(t)}$ 是 $\Psi\left(\boldsymbol{\beta},\ \boldsymbol{\beta}^{(t)}\right)$ 关于 $\boldsymbol{\beta}$的局部最小值时才成立。如果 F 的偏导数存在，且在 $\boldsymbol{\beta}^{(t)}$的一个小邻域内连续，这也意味着梯度 $\nabla F\left(\boldsymbol{\beta}^{(t)}\right) = \mathbf{0}$。因此，通过迭代应用（9.90）中的更新式，我们得到一列估计值，并收敛到目标函数的局部最小值 $\boldsymbol{\beta}^* =$ 最小值点集合$_{\boldsymbol{\beta}} F\left(\boldsymbol{\beta}\right)$：

$$F(\boldsymbol{\beta}^{(0)}) \geq F(\boldsymbol{\beta}^{(1)}) \geq F(\boldsymbol{\beta}^{(2)}) \geq \ldots \geq F(\boldsymbol{\beta}^{(t)}) \geq F(\boldsymbol{\beta}^{(t+1)}) \geq \ldots \geq F(\boldsymbol{\beta}^*). \tag{9.92}$$

定义 9.20 求得 Ψ 作为 F的一个上界函数，而引理 9.21 说明了 Ψ中的最小化性质，因此该算法通常称为上界最小化（*MM*）框架。MM 框架的主要优点在于，如果代理函数$\Psi\left(\cdot,\ \boldsymbol{\beta}^{(t)}\right)$ 存在闭式极小化解，或可以高效地最小化，那么每次迭代在计算上都易于处理——即使无法直接最小化 F。此外，如果 F 严格凸，MM 算法就会收敛到全局极小点。

▶作为拉格朗日 LASSO 问题的 MM 算法的近端梯度法。回顾第 9.2.2 节和 9.3.2 节，近端梯度法适用于如下形式的目标函数 $F = f + g$，其中 $f = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2$ 是凸且可微的，而 $g = \lambda\|\boldsymbol{\beta}\|_1$ 是凸且不可微的。对 应用线性近似定理（定理 2.13），可得 f:

$$F(\boldsymbol{\beta}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\beta}) = f(\widetilde{\boldsymbol{\beta}}) + \left\langle \nabla f(\widetilde{\boldsymbol{\beta}}), \widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta} \right\rangle + \frac{1}{2}\left\langle \widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}, \nabla^2 f(\boldsymbol{\xi})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) \right\rangle + g(\boldsymbol{\beta}),$$

其中 $\boldsymbol{\xi}=\gamma\boldsymbol{\beta}+(1-\gamma)\widetilde{\boldsymbol{\beta}}$，对于某个 $\gamma\in[0,1]$。由于 f 是 L_b-光滑的，且 $L_b\triangleq\left\|\boldsymbol{X}^\top\boldsymbol{X}\right\|_2$（例2.72），海森矩阵有一个上界，即 $\nabla^2 f(\boldsymbol{\xi})\preceq L_b\boldsymbol{I}_{p\times p}$，由此可得不等式

$$F(\boldsymbol{\beta})\leq f(\widetilde{\boldsymbol{\beta}})+\left\langle\nabla f(\widetilde{\boldsymbol{\beta}}),\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\rangle+\frac{L_b}{2}\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2^2+g(\boldsymbol{\beta})\triangleq\Psi(\boldsymbol{\beta},\widetilde{\boldsymbol{\beta}}),$$

当 $\widetilde{\boldsymbol{\beta}}=\boldsymbol{\beta}$ 时取等号。在 $\widetilde{\boldsymbol{\beta}}=\boldsymbol{\beta}^{(t)}$ 处，第 t 次迭代的更新变为：

$$\boldsymbol{\beta}^{(t+1)}=\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\text{最小值点集合}}\ \Psi\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)}\right) \tag{9.93}$$

$$\Longrightarrow\ \frac{\lambda}{L_b}\operatorname{sign}(\boldsymbol{\beta})+\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}-\frac{1}{L_b}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}), \tag{9.94}$$

这正是步长为 $\eta_t=1/L_b$ 的近端梯度步。因此，拉格朗日 LASSO 的近端梯度法可以解释为 MM 框架的一个实例，其中代理函数 Ψ 是通过对光滑部分 f 构造二次上界得到的。

算法 8 FISTA-V1

1: $\boldsymbol{\alpha}^{(1)}=\boldsymbol{\beta}^{(1)}\in\mathbb{R}^p$，$\gamma_1=\frac{1}{\zeta_1}=1$;
2: **针对** $t=1,2,\cdots$ **执行**
3: 选择 η_t 和 $\zeta_t=\frac{1}{\gamma_t}$;
4: $\boldsymbol{\beta}^{(t+1)}\leftarrow\mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)}-\eta_t\nabla f(\boldsymbol{\alpha}^{(t)}))$;
5: $\boldsymbol{\alpha}^{(t+1)}\leftarrow\boldsymbol{\beta}^{(t+1)}+\frac{\zeta_t-1}{\zeta_{t+1}}(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})$;
$=\boldsymbol{\beta}^{(t+1)}+(\frac{\gamma_{t+1}}{\gamma_t}-\gamma_{t+1})(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})$;
6: **结束循环**
7: 输出 $\boldsymbol{\beta}_{\text{final}}\leftarrow\boldsymbol{\beta}^{(T)}$;

算法 9 FISTA-V2

1: $\boldsymbol{u}^{(1)}=\boldsymbol{\alpha}^{(1)}=\boldsymbol{\beta}^{(1)}\in\mathbb{R}^p$, $\gamma_1=1$;
2: **对于** $t=1,2,\cdots$ **执行**
3: 选择 η_t 和 γ_t;
4: $\boldsymbol{\beta}^{(t+1)}\leftarrow\mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)}-\eta_t\nabla f(\boldsymbol{\alpha}^{(t)}))$;
5: $\boldsymbol{u}^{(t+1)}\leftarrow\boldsymbol{\beta}^{(t)}+\frac{1}{\gamma_t}(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})$;
6: $\boldsymbol{\alpha}^{(t+1)}\leftarrow(1-\gamma_{t+1})\boldsymbol{\beta}^{(t+1)}+\gamma_{t+1}\boldsymbol{u}^{(t+1)}$;
7: **结束循环**
8: 输出 $\boldsymbol{\beta}_{\text{final}}\leftarrow\boldsymbol{\beta}^{(T)}$;

9.3.3 快速近端梯度法—FISTA

该快速近端梯度法也称为快速迭代收缩阈值算法（*FISTA*）（算法9），是由Beck和Teboulle（2009年）提出的近端梯度法加速变体。回顾一下，标准近端梯度法在最小化光滑凸函数与非光滑凸正则项之和时，其收敛速率为 $\mathcal{O}$（$1/T$），（定理9.12）。一个自然的问题是能否改进这一速率。FISTA通过引入动量来解决这一问题，从而实现更快的收敛 $\mathcal{O}\left(1/T^2\right)$。

FISTA在每次迭代中分两步进行 t:

- 它利用最近两个迭代点的信息构造一个外推点 $\boldsymbol{\alpha}^{(t)}$，并在该外推点处执行一次近端梯度步。
-

具体而言，更新形式为：

$$\boldsymbol{\alpha}^{(t)}\leftarrow\boldsymbol{\beta}^{(t)}+\frac{t-2}{t+1}(\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}); \tag{9.95a}$$

$$\boldsymbol{\beta}^{(t+1)}\leftarrow\mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)}-\eta_t\nabla f(\boldsymbol{\alpha}^{(t)})). \tag{9.95b}$$

与标准近端梯度法的关键区别在于动量项，$\frac{t-2}{t+1}\left(\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right)$，该项利用历史迭代点来加速收敛。重要的是，这种加速不会显著增加每次迭代的计算成本，却能在经验和理论性能上实现显著更快的表现。如果步长 η_t 固定且满足$\eta_t = 1/L_b$（其中 f 假设为 L_b-光滑），则 FISTA 保证收敛速率为$\mathcal{O}\left(1/T^2\right)$；参见定理9.23。基于这一理解，完整的 FISTA 算法在算法8中给出，其中我们对所有 $\zeta_t = \frac{1}{\gamma_t}$ 定义 $t>0$。当$\frac{1}{\gamma_t} = \zeta_t \triangleq \frac{t+1}{2}$ 时，算法8 等价于公式（9.95）中的直观更新。

为便于理论分析，通常会采用 FISTA 的一种等价改写形式，如算法 9所示。虽然算法 8 明确展示了动量机制，因此更容易理解，但算法9 以一种便于进行收敛性证明的方式组织更新（见定理 9.23）。

▶ **收敛要求**。为达到加速后的 $\mathcal{O}\left(1/T^2\right)$ 速率，步长η_t 和$\gamma_t = \frac{1}{\zeta_t}$ 必须满足某些条件。具体而言，以下三个要求可确保获得所需的收敛行为：

$$f(\boldsymbol{\beta}^{(t)}) \le f(\boldsymbol{\alpha}^{(t)}) + \left\langle \nabla f(\boldsymbol{\alpha}^{(t)}), \boldsymbol{\beta}^{(t)} - \boldsymbol{\alpha}^{(t)} \right\rangle + \frac{1}{2\eta_t}\left\|\boldsymbol{\beta}^{(t)} - \boldsymbol{\alpha}^{(t)}\right\|_2^2, \text{ i.e., } \eta_t \le \frac{1}{L_b}; \tag{9.96a}$$

$$\gamma_1 = 1, \quad \frac{(1-\gamma_{t+1})\eta_{t+1}}{\gamma_{t+1}^2} \le \frac{\eta_t}{\gamma_t^2}, \quad t>1; \tag{9.96b}$$

$$\frac{\gamma_t^2}{\eta_t} = \mathcal{O}\left(\frac{1}{t^2}\right). \tag{9.96c}$$

▶ **的选择**。一种常见且有效的选择是令 $\eta_t = \frac{1}{L_b}$ 且 $\frac{1}{\zeta_t} = \gamma_t = \frac{2}{t+1}$，其中 $t>0$。此选择满足（9.96）中的所有条件，并得到（9.95）中直观的 FISTA 更新。不过，这种选择并非唯一。另一个有效的序列可递归定义如下：

$$\gamma_1 = 1, \quad \frac{1}{\gamma_{t+1}} = \frac{1}{2}\left(1+\sqrt{1+\frac{4}{\gamma_t^2}}\right) \iff \zeta_{t+1} = \frac{1+\sqrt{1+4\zeta_t^2}}{2}, \quad t>0.$$

其中 $\{\gamma_t\} = \{\frac{1}{\zeta_t}\}$。该序列还给出了一个 $\mathcal{O}\left(1/T^2\right)$ 收敛速率。下一个引理证实，这种递归选择满足（9.96）中的条件。

引理 9.22 设 $\{\frac{1}{\gamma_t}\}_{t>0} = \{\zeta_t\}_{t>0}$ 为序列（见算法 8），其定义为 $\zeta_1 = 1$，$\zeta_{t+1} = \frac{1+\sqrt{1+4\zeta_t^2}}{2}$，$t \ge 0$. 则 $\zeta_t \ge \frac{t+1}{2}$ 对所有 $t>0$均成立。

引理的证明 [: 9.22] 我们用归纳法证明。显然，当 $t=1$，$\zeta_1 = 1 \ge \frac{1+1}{2}$ 时，该结论成立。假设当 $t>1$ 且 $\zeta_t \ge \frac{t+1}{2}$ 时，该结论成立。我们将证明 $\zeta_{t+1} \ge \frac{t+2}{2}$。根据定义该序列的递归关系以及归纳假设，$\zeta_{t+1} = \frac{1+\sqrt{1+4\zeta_t^2}}{2} \ge \frac{1+\sqrt{1+(t+1)^2}}{2} \ge \frac{1+\sqrt{(t+1)^2}}{2} = \frac{t+2}{2}$。归纳证明至此完成。 ■

▶ **单调性问题**。原始 FISTA 算法并非下降法——也就是说，目标函数值 $F(\boldsymbol{\beta}^{(t)})$ 在某些迭代中可能会增加。为了解决这一问题，我们提出了 FISTA 的单调（下降）变体，只需对算法 8 做少量修改：

我们不总是接受新的迭代点 $\boldsymbol{\beta}^{(t+1)}$，而是选择

$$\boldsymbol{\beta}^{(t+1)} \in \arg\min\left\{F(\boldsymbol{\beta}) \mid \boldsymbol{\beta} \in \{\boldsymbol{\beta}^{(t)}, \boldsymbol{\xi}^{(t)}\}\right\}$$

确保 $F(\boldsymbol{\beta}^{(t+1)}) \leq$最小值$\{F(\boldsymbol{\xi}^{(t)})\ F(\boldsymbol{\beta}^{(t)})\}$。这一修改后的过程见算法 10。实践中，计算出近端更新 $\boldsymbol{\xi}^{(t)}$后，我们将其目标函数值与当前迭代点 $\boldsymbol{\beta}^{(t)}$的目标函数值进行比较。只有当 $F(\boldsymbol{\xi}^{(t)}) \leq F(\boldsymbol{\beta}^{(t)})$ 时，我们才接受 $\boldsymbol{\xi}^{(t)}$作为下一个迭代点；否则，保持 $\boldsymbol{\beta}^{(t)}$ 不变。更简洁地说，算法 10 的步骤6至10可等价地表示为：

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\beta}^{(t+1)} + \frac{\zeta_t}{\zeta_{t+1}}(\boldsymbol{\xi}^{(t)} - \boldsymbol{\beta}^{(t+1)}) + \frac{\zeta_t - 1}{\zeta_{t+1}}(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}).$$

算法 法 10 单调 FISTA（Beck和Teboulle，2009年），与算法 8比较

Require: A function $f(\boldsymbol{\beta})$ and a closed convex function g (usually non-sm滑）满足 (A1) and (A2) in Theorem 9.23;

initialize: $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\beta}^{(1)} \in \mathbb{R}^p$, $\gamma_1 = \frac{1}{\zeta_1} = 1$;
for $t = 1, 2, \ldots$ **do**
  Choose η_t and $\zeta_t = \frac{1}{\gamma_t}$;
  $\boldsymbol{\xi}^{(t)} \leftarrow \mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}))$;
  Choose $\boldsymbol{\beta}^{(t+1)}$ such that $F(\boldsymbol{\beta}^{(t+1)}) \leq \min\{F(\boldsymbol{\xi}^{(t)}), F(\boldsymbol{\beta}^{(t)})\}$;
  if $\boldsymbol{\beta}^{(t+1)}$ is $\boldsymbol{\xi}^{(t)}$ **then**
    $\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)} + \frac{\zeta_t - 1}{\zeta_{t+1}}(\boldsymbol{\xi}^{(t)} - \boldsymbol{\beta}^{(t)})$ ▷ Same as FISTA
  else if $\boldsymbol{\beta}^{(t+1)}$ is $\boldsymbol{\beta}^{(t)}$ **then**
    $\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} + \frac{\zeta_t}{\zeta_{t+1}}(\boldsymbol{\xi}^{(t)} - \boldsymbol{\beta}^{(t)})$;
  end if
  Stop if a stopping criterion is satisfied at iteration $t = T$;
end for
return $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$;

算法 11 第 t 次迭代的线搜索算法

输入： $\eta_t = \eta_{t-1} > 0$, $\rho < 1$。参考点 $\boldsymbol{\alpha}^{(t)}$ 及其梯度 $\nabla f(\boldsymbol{\alpha}^{(t)})$；计算候选更新 $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}))$;**当** 条件 (9.96a) 对 $\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t)}$ **不满足时，执行**减小步长 $\eta_t \leftarrow \rho\eta_t$；重新计算 $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\text{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}))$;**结束循环** **返回**更新 $\boldsymbol{\beta}^{(t+1)}$，步长 η_t;

▶ **线搜索方法**。在算法 8和 9中，步长必须满足 $\eta_t \leq \frac{1}{L_b}$ 以确保条件（9.96a）成立。然而，在许多实际问题中，Lipschitz 常数L_b 的 ∇f 是未知的。在这种情况下，我们仍可通过使用回溯线搜索，自适应地选择每次迭代中的 η_t。同时，序列 $\{\gamma_t\}$可以选取得同时满足条件（9.96b）和（9.96c），从而保持 $\mathcal{O}\left(1/T^2\right)$的加速收敛速率。

算法 11实现了这种线搜索策略。在迭代 t时，它用前一步的步长 η_{t-1} 初始化 η_t，随后按几何级数递减 η_t，直到条件（9.96a）满足。由于足够小的 η_t 总能使该条件成立（这是因为 f的光滑性），线搜索总能在有限步内终止。此外，不难验证，当适当地更新 $\{\gamma_t\}$ 时（例如，采用引理 9.22中的递归规则），在整个迭代过程中，另外两个条件——（9.96b）和（9.96c——仍然成立。

下面的定理给出了固定步长下 FISTA 的收敛速率。自适应（线搜索）变体的分析类似，故为简洁起见省略。

定理 9.23：（凸且光滑函数的 FISTA f， $\mathcal{O}(1/T^2)$：与定理 9.12比较）设 f 是适当闭凸且 L_b-光滑函数，并设 g 是适当闭凸函数。假设满足以下条件：

(A1) $g: \mathbb{R}^p \to (-\infty, \infty]$ 是适当、闭且凸的。
(A2) $f: \mathbb{R}^p \to (-\infty, \infty]$ 是适当、闭且凸的，$\mathrm{dom}(f)$ 是凸集，$\mathrm{dom}(g)$ ⊆包含于$(\mathrm{dom}(f))$ 的内部，且 f 在 L_b的定义域内部上是 光滑$(\mathrm{dom}(f)$。
设 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 为采用恒定步长规则、通过 FISTA 方法（算法 9）求解问题 $\min\{F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})\}$ 所生成的序列，其中对所有$\eta_t \triangleq \frac{1}{L_b}$均有$t > 0$。那么，对于任意优化器 $\boldsymbol{\beta}^*$ （对应 $F \triangleq f + g$ 和 $T > 1$），可得

$$F(\boldsymbol{\beta}^{(T)}) - F(\boldsymbol{\beta}^*) \leq \frac{2L_b}{T^2} \left\| \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^* \right\|_2^2 . \ ^a$$

^{a}。如果事先不知道 Lipschitz 常数 L_b，可以应用线搜索算法11 以达到相同的收敛速率。

定理 [的证明 9。23] 由于 $\boldsymbol{\beta}^{(t+1)} = \mathcal{P}\mathrm{rox}_{\eta_t g}(\boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}))$，根据近端性质 I（引理 2.100），可得

$$-\boldsymbol{\beta}^{(t+1)} + \boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}) \in \eta_t \partial g(\boldsymbol{\beta}^{(t+1)}).$$

根据次梯度不等式（定义 2.57），对于任意 $\boldsymbol{\beta} \in \mathbb{R}^p$，有

$$\eta_t g(\boldsymbol{\beta}) \geq \eta_t g(\boldsymbol{\beta}^{(t+1)}) + \left\langle -\boldsymbol{\beta}^{(t+1)} + \boldsymbol{\alpha}^{(t)} - \eta_t \nabla f(\boldsymbol{\alpha}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t+1)} \right\rangle .$$

利用 L_b-光滑性（定义 f）2.78以及 $\eta_t = \frac{1}{L_b}$，可得

$$f(\boldsymbol{\beta}^{(t+1)}) \leq f(\boldsymbol{\alpha}^{(t)}) + \left\langle \nabla f(\boldsymbol{\alpha}^{(t)}), \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)} \right\rangle + \frac{1}{2\eta_t} \|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\|_2^2.$$

结合上述两个不等式，对于任意 $\boldsymbol{\beta}$，我们有

$$\begin{aligned}
&F(\boldsymbol{\beta}^{(t+1)}) = f(\boldsymbol{\beta}^{(t+1)}) + g(\boldsymbol{\beta}^{(t+1)}) \\
&\leq g(\boldsymbol{\beta}) + f(\boldsymbol{\alpha}^{(t)}) + \left\langle \nabla f(\boldsymbol{\alpha}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\alpha}^{(t)} \right\rangle + \frac{1}{\eta_t} \left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t+1)} \right\rangle + \frac{1}{2\eta_t} \|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\|_2^2 \\
&\overset{\dagger}{\leq} g(\boldsymbol{\beta}) + f(\boldsymbol{\beta}) + \frac{1}{\eta_t} \left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t+1)} \right\rangle + \frac{1}{2\eta_t} \|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\|_2^2 \\
&= F(\boldsymbol{\beta}) + \frac{1}{\eta_t} \left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t+1)} \right\rangle + \frac{1}{2\eta_t} \|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\|_2^2,
\end{aligned} \tag{9.97}$$

其中不等式（†）由 f 的凸性得出。将（9.97）分别应用于 $\boldsymbol{\beta} \triangleq \boldsymbol{\beta}^{(t)}$ 和$\boldsymbol{\beta} \triangleq \boldsymbol{\beta}^*$，并分别乘以 $1-\gamma_t$ 和 γ_t，再将两式相加，得到

$$\begin{aligned}
&F(\boldsymbol{\beta}^{(t+1)}) - F(\boldsymbol{\beta}^*) - (1-\gamma_t)\big(F(\boldsymbol{\beta}^{(t)}) - F(\boldsymbol{\beta}^*)\big) \\
&\leq \frac{1}{\eta_t}\left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}, (1-\gamma_t)\boldsymbol{\beta}^{(t)} + \gamma_t\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t+1)} \right\rangle + \frac{1}{2\eta_t}\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\right\|_2^2 \\
&\overset{\dagger}{=} \frac{1}{2\eta_t}\left\{ \left\|\boldsymbol{\alpha}^{(t)} - (1-\gamma_t)\boldsymbol{\beta}^{(t)} - \gamma_t\boldsymbol{\beta}^*\right\|_2^2 - \left\|\boldsymbol{\beta}^{(t+1)} - (1-\gamma_t)\boldsymbol{\beta}^{(t)} - \gamma_t\boldsymbol{\beta}^*\right\|_2^2 \right\} \\
&= \frac{\gamma_t^2}{2\eta_t}\left\{ \left\|\boldsymbol{u}^{(t)} - \boldsymbol{\beta}^*\right\|_2^2 - \left\|\boldsymbol{u}^{(t+1)} - \boldsymbol{\beta}^*\right\|_2^2 \right\},
\end{aligned} \tag{9.98}$$

其中，等式（$\|\boldsymbol{a}-\boldsymbol{c}\|_2^2$ $\overset{\dagger}{\boldsymbol{a}\ \boldsymbol{b}\ \boldsymbol{c} \in \mathbb{R}^p}$ $2\langle \boldsymbol{a}-\boldsymbol{b}\ \boldsymbol{c}-\boldsymbol{a}\rangle + \|\boldsymbol{a}-\boldsymbol{b}\|_2^2 = \|\boldsymbol{b}-\boldsymbol{c}\|_2^2 -$）由以下事实可得：，对于任意,,，最后一个等式则由更新规则得出：

$$\boldsymbol{u}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \frac{1}{\gamma_t}(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}) \qquad \text{and} \qquad \boldsymbol{\alpha}^{(t)} = (1-\gamma_t)\boldsymbol{\beta}^{(t)} + \gamma_t\boldsymbol{u}^{(t)}.$$

回顾一下，η_t 和 γ_t 的选取满足 $\frac{1-\gamma_t}{\gamma_t^2}\eta_t \leq \frac{1}{\gamma_{t-1}^2}\eta_{t-1}$，由此可得相邻迭代点之间的不等式

$$\begin{aligned}
&\frac{\eta_t}{\gamma_t^2}\big(F(\boldsymbol{\beta}^{(t+1)}) - F(\boldsymbol{\beta}^*)\big) + \frac{1}{2}\left\|\boldsymbol{u}^{(t+1)} - \boldsymbol{\beta}^*\right\|_2^2 \leq (1-\gamma_t)\frac{\eta_t}{\gamma_t^2}\big(F(\boldsymbol{\beta}^{(t)}) - F(\boldsymbol{\beta}^*)\big) + \frac{1}{2}\left\|\boldsymbol{u}^{(t)} - \boldsymbol{\beta}^*\right\|_2^2 \\
&\leq \frac{\eta_{t-1}}{\gamma_{t-1}^2}\big(F(\boldsymbol{\beta}^{(t)}) - F(\boldsymbol{\beta}^*)\big) + \frac{1}{2}\left\|\boldsymbol{u}^{(t)} - \boldsymbol{\beta}^*\right\|_2^2 \leq \frac{\eta_1}{\gamma_1^2}\big(F(\boldsymbol{\beta}^{(2)}) - F(\boldsymbol{\beta}^*)\big) + \frac{1}{2}\left\|\boldsymbol{u}^{(2)} - \boldsymbol{\beta}^*\right\|_2^2 \\
&\leq \frac{(1-\gamma_1)\eta_1}{\gamma_1^2}\big(F(\boldsymbol{\beta}^{(1)}) - F(\boldsymbol{\beta}^*)\big) + \frac{1}{2}\left\|\boldsymbol{u}^{(1)} - \boldsymbol{\beta}^*\right\|_2^2 = \frac{1}{2}\left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\right\|_2^2,
\end{aligned}$$

其中，最后一个等式由以下事实可得：$\gamma_1 = 1$且 $\boldsymbol{u}^{(1)} = \boldsymbol{\beta}^{(1)}$。应用引理9.22即可得证。

■

▶ **应用于 LASSO**。请注意，FISTA 方法（算法8 和引理9.22）是对标准近端梯度法的轻微加速。将其应用于 LASSO 问题时

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

FISTA 的更新形式如下：

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \mathcal{T}_{\eta_t\lambda}\left(\boldsymbol{\beta}^{(t)} - \eta_t\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\right); \tag{9.99a}$$

$$\zeta_{t+1} \leftarrow \frac{1+\sqrt{1+4\zeta_t^2}}{2}; \tag{9.99b}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\alpha}^{(t+1)} + \left(\frac{\zeta_t - 1}{\zeta_{t+1}}\right)(\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)}), \tag{9.99c}$$

其中 $\eta_t = 1/\left\|\boldsymbol{X}^\top\boldsymbol{X}\right\|_2$ 和 $\zeta_1 = 1$。

9.3.4 罚函数法

我们现在已经介绍了几种求解拉格朗日 LASSO 问题的方法。请注意，当正则化参数 λ较小时，目标函数主要由

光滑损失项 $f = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 主导。近端梯度法通常采用固定步长 $\eta = 1/L_b = 1/\left\|\boldsymbol{X}^\top \boldsymbol{X}\right\|_2$，它仅取决于数据矩阵 $\boldsymbol{X}$，与 λ无关。更重要的是，当 λ 非常小时，近端（软阈值处理）步骤只会进行非常弱的收缩。因此，稀疏性几乎没有得到强化，迭代点的行为几乎就像对病态最小二乘问题应用标准梯度下降一样，导致收敛速度极其缓慢。为了解决这一问题，我们转而采用罚函数法。

二次惩罚函数法

考虑带等式约束的优化问题：

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & f(\boldsymbol{\beta}), \\ \text{s.t.} \quad & g_i(\boldsymbol{\beta}) = 0,\ i \in \mathcal{E},\ \text{with } |\mathcal{E}| = k. \end{aligned} \tag{9.100}$$

在各种可能的罚函数形式中，二次惩罚最简单，也最常用。其定义如下：

定义 9.24（等式约束的二次罚函数）。对于等式约束优化问题(9.100)，二次罚函数定义为

$$f_\sigma(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + \frac{1}{2}\sigma \sum_{i\in\mathcal{E}} g_i^2(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + \frac{1}{2}\sigma \|\boldsymbol{g}(\boldsymbol{\beta})\|_2^2, \tag{9.101}$$

其中，$\boldsymbol{g}(\boldsymbol{\beta}): \mathbb{R}^p \to \mathbb{R}^k$ 是向量值约束函数，其第 i 个分量为 $g_i(\boldsymbol{\beta})$。项 $\frac{1}{2}\sigma\|\boldsymbol{g}(\boldsymbol{\beta})\|_2^2$ 称为惩罚项，而 $\sigma > 0$ 是惩罚参数（或罚因子）。

该罚函数会对违反约束条件的点赋予高代价。在优化过程中，迭代点通常位于可行域之外。随着 σ 增大，惩罚项会支配目标函数，推动 $f_\sigma(\boldsymbol{\beta})$ 的极小化解更接近可行集。

相反，在可行域内（即 $\boldsymbol{g}(\boldsymbol{\beta}) = \mathbf{0}$ 的区域），惩罚项为零，而最小化 $f_\sigma(\boldsymbol{\beta})$ 等价于最小化原目标函数 $f(\boldsymbol{\beta})$。因此，当 $\sigma \to \infty$ 时，f_σ 的极小化点会在温和的正则性条件下收敛到原约束问题（9.100）的解。

算法 12 模型（9.100）的二次惩罚函数法

要求： 函数 $f(\boldsymbol{\beta})$ 以及约束条件集合 $\{g_i(\boldsymbol{\beta})\}$；**输入：** 选择惩罚因子增长系数 $\rho > 1$；**输入：** 初始化 $\boldsymbol{\beta}^{(1)}$ 和 $\sigma_1 > 0$；**对于** $t = 1, 2, \ldots$ **执行** 以 $\boldsymbol{\beta}^{(t)}$ 作为初始点，求解 $\boldsymbol{\beta}^{(t+1)} =$ 最小值点集合$_{\boldsymbol{\beta}}$ $f_{\sigma_t}(\boldsymbol{\beta})$；更新惩罚参数：$\sigma_{t+1} = \rho\sigma_t$；如果在迭代 $t = T$ 处满足停止准则，则停止；**结束循环** **返回：** 最终 $\boldsymbol{\beta}^{(T)}$。

▷ (QPF_1)
▷ (QPF_2)
▷ (QPF_3)

二次罚函数法（如算法 12 所述）按如下方式进行：生成一个呈指数增长的惩罚参数序列 $\{\sigma_t\}$（例如，$\sigma_{t+1} = \rho\sigma_t$，其中 $\rho > 1$），并针对每个 σ_t，近似最小化惩罚目标函数 $f_{\sigma_t}(\boldsymbol{\beta})$。

步骤 (QPF$_1$) 中的子问题——一个无约束优化问题——可以使用任何适用于光滑目标函数的迭代方法求解，例如梯度下降、拟牛顿方法或牛顿型算法。

重要的是，步骤 (QPF$_1$) 中 argmin 的解释可能因问题和求解器而异：

- $\boldsymbol{\beta}^{(t+1)}$ 是 $f_{\sigma_t}(\boldsymbol{\beta})$ 的全局极小点；$\boldsymbol{\beta}^{(t+1)}$ 是 $f_{\sigma_t}(\boldsymbol{\beta})$ 的局部极小点；$\boldsymbol{\beta}^{(t+1)}$ 不一定是严
- 格极小化解，但近似满足一阶最优性条件，即 $\nabla f_{\sigma_t}(\boldsymbol{\beta}^{(t+1)}) \approx \mathbf{0}$。
-

基于此，在实现算法 12 时，以下三点实践考量至关重要：

- 惩罚参数调度的选择。 σ_t 增加的速率会显著影响性能。如果 σ_t 增长过快，子问题会变得越来越病态，难以求解。相反，如果 σ_t 增加过慢，算法就需要进行许多外层迭代，从而导致整体收敛缓慢。一种更具自适应性的策略是根据当前子问题的求解难度来调整 σ_t 的增长：如果前一个子问题快速收敛，则可以更积极地增加 σ_{t+1}；否则，建议采用更保守的更新。较小时无界的风险 σ_t。 当 σ_t 过小时，惩罚项可能不足以保证可行性，且 $f_{\sigma_t}(\boldsymbol{\beta})$ 可能向下无界。在这种情况下，子问题求解器可能发
- 散。如果在步骤 (QPF$_1$) 中检测到发散，算法应停止当前内层求解，增大 σ_t，然后重试。子问题解所需的精度。 为确保整体方法收敛，必须以充分的精度求解子问题。特别是，满足一阶最优性条件时的误差应随着 $t \to \infty$ 趋于零。
-

收敛性分析

现在，我们来考察二次罚函数法用于等式约束问题时的收敛性质。为简单起见，我们假设对于每个 $\sigma_t > 0$，惩罚函数 $f_{\sigma_t}(\boldsymbol{\beta})$ 至少存在一个极小化解。请注意，对于某些问题，这一假设可能不成立——例如，如果惩罚力度太弱，无法阻止目标函数沿不可行方向无限减小。在这种情况下，二次罚函数法不适用，因此我们将分析范围限定在满足这一基本正则性条件的问题上。

定理 9.25：（具有最优子问题的罚函数法）设 $\boldsymbol{\beta}^{(t+1)}$是 $f_{\sigma_t}(\boldsymbol{\beta})$ 的一个全局极小点，[a] 并假设惩罚参数 $\{\sigma_t\}$ 在算法 $\sigma_t \to \infty$中单调满足12。则序列 $\boldsymbol{\beta}^*$ 的每个极限点 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 都是原约束问题（9.100）的一个全局极小点。

[a]。问题9.11讨论了子问题次优解的一种改进方法。

定理的证明 [9.25] 设 $\boldsymbol{\beta}$是原问题的一个全局最小值（9.100），即 $f(\widehat{\boldsymbol{\beta}}) \leq f(\boldsymbol{\beta})$，对于所有 $\boldsymbol{\beta}$满足 $g_i(\boldsymbol{\beta}) = 0, \quad i \in \mathcal{E}$。由于 $\boldsymbol{\beta}^{(t+1)}$使 $f_{\sigma_t}(\boldsymbol{\beta})$ 达到最小值，我们有

$f_{\sigma_t}(\boldsymbol{\beta}^{(t+1)}) \leq f_{\sigma_t}(\widehat{\boldsymbol{\beta}})$，即，

$$f(\boldsymbol{\beta}^{(t+1)}) + \frac{\sigma_t}{2}\sum_{i\in\mathcal{E}} g_i(\boldsymbol{\beta}^{(t+1)})^2 \leq f(\widehat{\boldsymbol{\beta}}) + \frac{\sigma_t}{2}\sum_{i\in\mathcal{E}} g_i(\widehat{\boldsymbol{\beta}})^2 \tag{9.102}$$

$$\Longrightarrow \quad \sum_{i\in\mathcal{E}} g_i(\boldsymbol{\beta}^{(t+1)})^2 \leq \frac{2}{\sigma_t}\big(f(\widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta}^{(t+1)})\big). \tag{9.103}$$

令 $\boldsymbol{\beta}^*$ 为序列 $\{\boldsymbol{\beta}^{(t)}\}$ 的任意极限点。对 (9.103) 中的 $t\to\infty$ 取极限，并利用 $f(\boldsymbol{\beta})$ 和 $g_i(\boldsymbol{\beta})$ 的连续性以及 $\sigma_t \to +\infty$，可得 $\sum_{i\in\mathcal{E}} g_i(\boldsymbol{\beta}^*)^2 = 0$，因此 $\boldsymbol{\beta}^*$ 对原问题是可行的。此外，由 (9.102) 可知，对于所有 t，都有 $f(\boldsymbol{\beta}^{(t+1)}) \leq f(\widehat{\boldsymbol{\beta}})$ ^ ^。取极限可得 $f(\boldsymbol{\beta}^*) \leq f(\boldsymbol{\beta})$。根据 $\boldsymbol{\beta}$ 的最优性可知，$\backsim f(\boldsymbol{\beta}^*) = f(\boldsymbol{\beta})$，因此 $\boldsymbol{\beta}^*$ 也是一个全局极小点。

■

算法 13 通过罚函数法求解拉格朗日 LASSO 问题

1: **输入：** 给定初始值 $\boldsymbol{\beta}^{(1)}$、正则化参数 λ、初始参数 $\zeta_1 \gg \lambda$、缩减因子 $\gamma \in$ （0, 1）。2: **对于** $t = 1,\ 2,$ ..... **执行**3: 使用上一次迭代的 $\boldsymbol{\beta}^{(t)}$ 作为初始猜测，通过次梯度下降法或其他无约束优化方法求解问题 $\boldsymbol{\beta}^{(t+1)}$ =arg $\min_{\boldsymbol{\beta}} \left\{\frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \zeta_t\|\boldsymbol{\beta}\|_1\right\}$；4: **如果** $\zeta_t = \lambda$ **那么**5: 停止迭代，输出 $\boldsymbol{\beta}^{(t+1)}$；6: **否则**7: 更新/缩减罚因子 $\zeta_{t+1} = \max\{\lambda, \gamma\zeta_t\}$；8: **结束 if**9: 如果在迭代 $t = T$ 时满足停止准则，则停止；10: **结束 for**11: **返回** 最终 $\boldsymbol{\beta}^{(T)}$;

应用于拉格朗日 LASSO 的罚函数法

回顾拉格朗日 LASSO 问题（L_L）：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

其中，$\lambda > 0$ 是一个正则化参数。求解 LASSO 问题可以看作是对以下基追踪（BP）问题的近似求解：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y},$$

BP 问题是一个带有线性等式约束的非光滑优化问题。通过应用二次罚函数来强制满足约束 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$，我们得到以下惩罚形式：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 + \frac{\sigma}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2.$$

设定 $\lambda \triangleq \frac{1}{\sigma}$ 后，我们便能看出二者之间存在直接对应关系：上述惩罚问题恰好就是正则化参数为 λ 的 LASSO 问题。这一观察带来了两个重要见解：

- LASSO 问题和基追踪问题的解通常并不相同。然而，当 $\lambda \to 0^+$（等价地，当 $\sigma \to \infty$）时，LASSO 问题的解会收敛到基追踪问题的一个解——前提是存在可行解。
- 当 λ 很小时（即 σ 很大时），二次项会占主导地位，并且对约束违反高度敏感；但由于 ℓ_1 范数，整体目标函数仍然是非光滑的。此外，如果尝试通过罚函数法直接求解基追踪问题，并固定一个较大的 σ，所得子问题可能会变得病态，从而导致收敛缓慢。

在经典的罚函数框架中，惩罚参数 σ 会逐渐增大至无穷大，从而在后续迭代中更加严格地满足可行性要求。将这一思路转化到 LASSO 问题中，相当于从较大的 λ（即较弱地施加数据拟合项）开始，并逐步减小 λ，直至达到所需值。这种类似同伦的方法能够提高数值稳定性和收敛性能。具体步骤详见算法 13。

9.3.5 交替方向乘子法（ADMM）

交替方向乘子法（*ADMM*）广泛用于求解 LASSO、稀疏逻辑回归和基追踪等问题。这些应用受益于 ADMM 高效处理稀疏性约束的能力。此外，ADMM 还可用于求解支持向量机（SVM；参见问题8.6）训练中出现的优化问题，尤其是在处理大规模数据集时（Boyd 等人，2011年）。我们首先简要介绍 ADMM，然后讨论其在 LASSO 问题中的应用。

▶ **固定增广拉格朗日法。** 考虑以下凸优化问题：

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\alpha}) \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} = \boldsymbol{f}, \tag{9.104a}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、$\boldsymbol{Z} \in \mathbb{R}^{n\times q}$、$\boldsymbol{f} \in \mathbb{R}^n$ 和 $f\colon \mathbb{R}^p \to (-\infty, \infty]$ 和 $g\colon \mathbb{R}^q \to (-\infty, \infty]$ 是适当闭凸函数。拉格朗日函数为

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\alpha}) + \langle \boldsymbol{\nu}, \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f} \rangle . \tag{9.104b}$$

对偶目标函数通过对原变量 $\boldsymbol{\beta}$、$\boldsymbol{\alpha}$ 最小化拉格朗日函数得到（见第 2.5节）：

$$Q(\boldsymbol{\nu}) = \min_{\boldsymbol{\beta},\boldsymbol{\alpha}} L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}) = -f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}) - g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}) - \langle \boldsymbol{\nu}, \boldsymbol{f} \rangle .$$

因此，对偶问题变为

$$\min_{\boldsymbol{\nu}\in\mathbb{R}^n} f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}) + \langle \boldsymbol{\nu}, \boldsymbol{f} \rangle . \tag{9.104c}$$

若 f 和 g 均为真闭凸函数，则由于共轭函数的适当性、闭性和凸性（引理*2.64*，习题2.67），可以使用近端点法（问题9.6，这是原始梯度法的推论）求解对偶问题。给定参数 $\rho > 0$，第 t次迭代的更新为

$$\boldsymbol{\nu}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\nu}\in\mathbb{R}^n} \left\{ G(\boldsymbol{\nu}) \triangleq f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}) + \langle \boldsymbol{\nu}, \boldsymbol{f} \rangle + \frac{1}{2\rho} \left\| \boldsymbol{\nu} - \boldsymbol{\nu}^{(t)} \right\|_2^2 \right\} .$$

根据凸函数的一阶充要条件（定理2.62），可得

$$\mathbf{0} \in \partial G(\boldsymbol{\nu}^{(t+1)}) = -\boldsymbol{X}\partial f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}^{(t+1)}) - \boldsymbol{Z}\partial g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}^{(t+1)}) + \boldsymbol{f} + \frac{\boldsymbol{\nu}^{(t+1)} - \boldsymbol{\nu}^{(t)}}{\rho}.$$

根据共轭函数的定义（定义2.63），可得

$$\begin{aligned}
\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\langle \boldsymbol{X}^\top \boldsymbol{\nu}^{(t+1)}, \boldsymbol{\beta} \right\rangle + f(\boldsymbol{\beta}) = f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}^{(t+1)}); \\
\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \left\langle \boldsymbol{Z}^\top \boldsymbol{\nu}^{(t+1)}, \boldsymbol{\alpha} \right\rangle + g(\boldsymbol{\alpha}) = g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}^{(t+1)}).
\end{aligned}$$

根据共轭次梯度定理（定理2.70），这等价于通过以下方式更新 $\boldsymbol{\beta}$、 $\boldsymbol{\alpha}$和 $\boldsymbol{\nu}$：

$$\begin{aligned}
\boldsymbol{\beta}^{(t+1)} &\in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\langle \boldsymbol{X}^\top \boldsymbol{\nu}^{(t+1)}, \boldsymbol{\beta} \right\rangle + f(\boldsymbol{\beta}) = \partial f^*(-\boldsymbol{X}^\top \boldsymbol{\nu}^{(t+1)}); \\
\boldsymbol{\alpha}^{(t+1)} &\in \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \left\langle \boldsymbol{Z}^\top \boldsymbol{\nu}^{(t+1)}, \boldsymbol{\alpha} \right\rangle + g(\boldsymbol{\alpha}) = \partial g^*(-\boldsymbol{Z}^\top \boldsymbol{\nu}^{(t+1)}); \\
\boldsymbol{\nu}^{(t+1)} &\leftarrow \boldsymbol{\nu}^{(t)} + \rho\left(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}\right).
\end{aligned}$$

将 $\boldsymbol{\nu}^{(t+1)}$ 的表达式代入前两个等式，并再次应用凸函数的一阶必要且充分条件（定理2.62）（因为 f和 g 是真闭凸函数），可得

$$\mathbf{0} \in \boldsymbol{X}^\top \left(\boldsymbol{\nu}^{(t)} + \rho\big(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}\big)\right) + \partial f(\boldsymbol{\beta}^{(t+1)}); \tag{9.105a}$$

$$\mathbf{0} \in \boldsymbol{Z}^\top \left(\boldsymbol{\nu}^{(t)} + \rho\big(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}\big)\right) + \partial g(\boldsymbol{\alpha}^{(t+1)}); \tag{9.105b}$$

$$\boldsymbol{\nu}^{(t+1)} \leftarrow \boldsymbol{\nu}^{(t)} + \rho\left(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}\right). \tag{9.105c}$$

给定 $\boldsymbol{\nu}^{(t)}$，式（9.105b）和式（9.105c）对应于函数的最优性条件

$$L_\rho(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}^{(t)}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\alpha}) + \left\langle \boldsymbol{\nu}^{(t)}, \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f} \right\rangle + \frac{\rho}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f}\right\|_2^2, \tag{9.106}$$

该函数称为增广拉格朗日函数（Lu，2025年）。当 $\rho = 0$时，增广拉格朗日函数退化为标准拉格朗日函数；当 $\rho > 0$时，它充当拉格朗日函数的惩罚形式。增广拉格朗日方法进行迭代（在第 t次迭代中）如下：

$$\text{augmented Lagrangian:} \quad \begin{cases} (\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}) \in \arg\min\limits_{\boldsymbol{\beta},\boldsymbol{\alpha}} L_\rho(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}^{(t)}); \\ \qquad\qquad \boldsymbol{\nu}^{(t+1)} \leftarrow \boldsymbol{\nu}^{(t)} + \rho(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}). \end{cases}$$

然而，与标准的增广拉格朗日法（Lu，2025年）不同，参数 ρ由于近端点法（问题9.6）的收敛保证，可以在各次迭代中保持固定。

▶ **交替方向乘子法**。求解该问题的一项重要挑战，是变量 $\boldsymbol{\beta}$ 与 $\boldsymbol{\alpha}$之间存在耦合项 $\rho(\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{Z}\boldsymbol{\alpha})$。交替方向乘子法通过一次交替最小化迭代，取代对（$\boldsymbol{\beta}$, $\boldsymbol{\alpha}$）进行联合最小化，从而解决这一问题。具体而言，在第 t 次迭代中，交替方向乘子法依次更新 $\boldsymbol{\beta}$、$\boldsymbol{\alpha}$和对偶变量 $\boldsymbol{\nu}$，其更新依据为（9.106），具体如下：

$$\text{ADMM:}\quad \begin{cases} \boldsymbol{\beta}^{(t+1)} \in \arg\min\limits_{\boldsymbol{\beta}} \left\{ f(\boldsymbol{\beta}) + \dfrac{\rho}{2} \left\| \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f} + \dfrac{1}{\rho}\boldsymbol{\nu}^{(t)} \right\|_2^2 \right\}; \\ \boldsymbol{\alpha}^{(t+1)} \in \arg\min\limits_{\boldsymbol{\alpha}} \left\{ g(\boldsymbol{\alpha}) + \dfrac{\rho}{2} \left\| \boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f} + \dfrac{1}{\rho}\boldsymbol{\nu}^{(t)} \right\|_2^2 \right\}; \\ \boldsymbol{\nu}^{(t+1)} \leftarrow \boldsymbol{\nu}^{(t)} + \rho(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}). \end{cases} \tag{9.107a}$$

令 $\widetilde{\boldsymbol{\nu}} \triangleq \frac{1}{\rho}\boldsymbol{\nu}$，该方案可用缩放对偶变量 $\widetilde{\boldsymbol{\nu}}$ 等价地重写为：

$$\text{ADMM:}\quad \begin{cases} \boldsymbol{\beta}^{(t+1)} \in \arg\min\limits_{\boldsymbol{\beta}} \left\{ f(\boldsymbol{\beta}) + \dfrac{\rho}{2} \left\| \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f} + \widetilde{\boldsymbol{\nu}}^{(t)} \right\|_2^2 \right\}; \\ \boldsymbol{\alpha}^{(t+1)} \in \arg\min\limits_{\boldsymbol{\alpha}} \left\{ g(\boldsymbol{\alpha}) + \dfrac{\rho}{2} \left\| \boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f} + \widetilde{\boldsymbol{\nu}}^{(t)} \right\|_2^2 \right\}; \\ \widetilde{\boldsymbol{\nu}}^{(t+1)} \leftarrow \widetilde{\boldsymbol{\nu}}^{(t)} + (\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f}). \end{cases} \tag{9.107b}$$

注意，式（9.107b）中的缩放形式 ADMM 更新可由以下用缩放对偶变量表示的增广拉格朗日函数推导得到 $\widetilde{\boldsymbol{\nu}}$：

$$L_\rho(\boldsymbol{\beta}, \boldsymbol{\alpha}, \widetilde{\boldsymbol{\nu}}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\alpha}) + \frac{\rho}{2} \|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} - \boldsymbol{f} + \widetilde{\boldsymbol{\nu}}\|_2^2 - \frac{\rho}{2} \|\widetilde{\boldsymbol{\nu}}\|_2^2 . \tag{9.108}$$

这一迭代过程使算法能够收敛到一个满足约束条件且使目标函数最小化的解。ADMM 的主要优势在于能够高效处理复杂的目标函数和约束条件，尤其是在现代统计学习和大规模优化问题中常见的高维场景下。

▶ **停止准则**。ADMM 迭代会持续进行，直到原问题和对偶问题的残差都低于预设的容差水平。这些残差定义如下：

$$\begin{aligned} \boldsymbol{r}^{(t+1)} &= \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t+1)} \quad \text{(primal residual)}, \\ \boldsymbol{s}^{(t+1)} &= \rho(\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)}) \quad \text{(dual residual)}. \end{aligned}$$

算法在以下条件下终止

$$\left\| \boldsymbol{r}^{(t+1)} \right\|_2 \le \varepsilon_{\text{pri}} \quad \text{and} \quad \left\| \boldsymbol{s}^{(t+1)} \right\|_2 \le \varepsilon_{\text{dual}}.$$

其中 ε_{pri} 和 $\varepsilon_{\text{dual}}$ 是用户指定的容差。

▶ **ADMM 的收敛**。以下定理说明了 ADMM 生成的序列具有 $\mathcal{O}(1/T)$ 收敛速度。为陈述这一结果，我们作出以下假设。对于任意 $\boldsymbol{u} \in \mathbb{R}^p$ 和 $\boldsymbol{v} \in \mathbb{R}^q$，问题的最优解集

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ f(\boldsymbol{\beta}) + \frac{\rho}{2} \|\boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \langle \boldsymbol{u}, \boldsymbol{\beta} \rangle \right\}$$

and

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \left\{ g(\boldsymbol{\alpha}) + \frac{\rho}{2} \|\boldsymbol{Z}\boldsymbol{\alpha}\|_2^2 + \langle \boldsymbol{v}, \boldsymbol{\alpha} \rangle \right\}$$

~非空。此外，我们假设存在 $\boldsymbol{\beta} \in \mathrm{dom}(f)$ 和 $\widetilde{\boldsymbol{\alpha}} \in \mathrm{dom}(g)$，使得 ~ $\boldsymbol{D}\boldsymbol{\beta} + \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}} = \boldsymbol{f}$。这些条件确保 ADMM 迭代序列定义良好。

定理 9.26：（ADMM 的收敛性，$\mathcal{O}(1/T)$）假设上述假设成立，并令 $f\colon \mathbb{R}^p \to (-\infty, \infty]$ 和 $g\colon \mathbb{R}^q \to (-\infty, \infty]$ 为适当闭凸函数。令 $\{(\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t)})\}_{t>0}$ 为 ADMM 求解问题 (9.104a) 所生成的序列。令 $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)$ 为问题 (9.104a) 的一个最优解，并令 $\boldsymbol{\nu}^*$ 为对偶问题 (9.104c) 的一个最优解。假设 $\zeta > 0$ 是满足 $\zeta \geq \|\boldsymbol{\nu}^*\|_2$的任意常数。那么对于所有 $T > 0$,

$$F(\overline{\boldsymbol{\beta}}^{(T)}, \overline{\boldsymbol{\alpha}}^{(T)}) - F(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*) \leq \frac{\left\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}^{(1)}\right\|_{\boldsymbol{C}}^2 + \frac{1}{\rho}\left(\zeta + \left\|\boldsymbol{\nu}^{(1)}\right\|_2\right)^2}{2T}, \tag{9.109}$$

w此处 $\boldsymbol{C} \triangleq \rho \boldsymbol{Z}^\top \boldsymbol{Z}$, $\|\boldsymbol{a}\|_{\boldsymbol{C}} = (\boldsymbol{a}^\top \boldsymbol{C} \boldsymbol{a})^{1/2}$，平均迭代定义为 a s

$$\overline{\boldsymbol{\beta}}^{(T)} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\beta}^{(t+1)} \qquad \text{and} \qquad \overline{\boldsymbol{\alpha}}^{(T)} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\alpha}^{(t+1)}.$$

证明 [定理9.26]根据凸函数的一阶最优性条件（定理 2.62）和 ADMM 更新规则（9.107a），可知 $\boldsymbol{\beta}^{(t+1)}$ 和 $\boldsymbol{\alpha}^{(t+1)}$满足

$$-\rho \boldsymbol{X}^\top \left(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)} \right) \in \partial f(\boldsymbol{\beta}^{(t+1)}), \tag{9.110}$$

$$-\rho \boldsymbol{Z}^\top \left(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{f} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)} \right) \in \partial g(\boldsymbol{\alpha}^{(t+1)}). \tag{9.111}$$

定义辅助变量

$$\widetilde{\boldsymbol{\beta}}^{(t)} \triangleq \boldsymbol{\beta}^{(t+1)}, \qquad \widetilde{\boldsymbol{\alpha}}^{(t)} \triangleq \boldsymbol{\alpha}^{(t+1)}, \qquad \widetilde{\boldsymbol{\nu}}^{(t)} \triangleq \boldsymbol{\nu}^{(t)} + \rho(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f}).$$

利用上述最优性条件和次梯度不等式（定义 2.57），可得对于任意 $\boldsymbol{\beta} \in \mathrm{dom}(f)$ 和 $\boldsymbol{\alpha} \in \mathrm{do\,m}(g)$,

$$f(\boldsymbol{\beta}) - f(\widetilde{\boldsymbol{\beta}}^{(t)}) + \left\langle \rho \boldsymbol{X}^\top \left(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)} \right), \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}^{(t)} \right\rangle \geq 0,$$

$$g(\boldsymbol{\alpha}) - g(\widetilde{\boldsymbol{\alpha}}^{(t)}) + \left\langle \rho \boldsymbol{Z}^\top \left(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} + \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}}^{(t)} - \boldsymbol{f} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)} \right), \boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)} \right\rangle \geq 0.$$

根据 $\widetilde{\boldsymbol{\nu}}^{(t)}$ 的定义，这些不等式可以改写为

$$f(\boldsymbol{\beta}) - f(\widetilde{\boldsymbol{\beta}}^{(t)}) + \left\langle \boldsymbol{X}^\top \widetilde{\boldsymbol{\nu}}^{(t)}, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}^{(t)} \right\rangle \geq 0,$$
$$g(\boldsymbol{\alpha}) - g(\widetilde{\boldsymbol{\alpha}}^{(t)}) + \left\langle \boldsymbol{Z}^\top \widetilde{\boldsymbol{\nu}}^{(t)} + \boldsymbol{C}(\widetilde{\boldsymbol{\alpha}}^{(t)} - \boldsymbol{\alpha}^{(t)}), \boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)} \right\rangle \geq 0,$$

其中 $\boldsymbol{C} = \rho \boldsymbol{Z}^\top \boldsymbol{Z}$。将两个不等式相加，并使用对偶更新 $\boldsymbol{\nu}^{(t+1)}$：$\boldsymbol{\nu}^{(t+1)} - \boldsymbol{\nu}^{(t)} = \rho(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} + \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}}^{(t)} - \boldsymbol{f})$，可得，对于任意 $\boldsymbol{\beta} \in \mathrm{dom}(f)$、$\boldsymbol{\alpha} \in \mathrm{dom}(g)$ 以及 $\boldsymbol{\nu} \in \mathbb{R}^n$，

$$F(\boldsymbol{\beta}, \boldsymbol{\alpha}) - F(\widetilde{\boldsymbol{\beta}}^{(t)}, \widetilde{\boldsymbol{\alpha}}^{(t)}) + \left\langle \begin{bmatrix} \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}^{(t)} \\ \boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)} \\ \boldsymbol{\nu} - \widetilde{\boldsymbol{\nu}}^{(t)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{X}^\top \widetilde{\boldsymbol{\nu}}^{(t)} \\ \boldsymbol{Z}^\top \widetilde{\boldsymbol{\nu}}^{(t)} \\ -\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} - \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}}^{(t)} + \boldsymbol{f} \end{bmatrix} - \begin{bmatrix} \boldsymbol{0}(\boldsymbol{\beta}^{(t)} - \widetilde{\boldsymbol{\beta}}^{(t)}) \\ \boldsymbol{C}(\boldsymbol{\alpha}^{(t)} - \widetilde{\boldsymbol{\alpha}}^{(t)}) \\ \frac{1}{\rho}(\boldsymbol{\nu}^{(t)} - \boldsymbol{\nu}^{(t+1)}) \end{bmatrix} \right\rangle \geq 0. \quad (9.112)$$

对于任意向量 $\boldsymbol{a}$，$\boldsymbol{b}$，$\boldsymbol{c}$，$\boldsymbol{d}$，以及半正定矩阵 $\boldsymbol{H}$，有：

$$(\boldsymbol{a} - \boldsymbol{b})^\top \boldsymbol{H}(\boldsymbol{c} - \boldsymbol{d}) = \frac{1}{2}\left(\|\boldsymbol{a} - \boldsymbol{d}\|_{\boldsymbol{H}}^2 - \|\boldsymbol{a} - \boldsymbol{c}\|_{\boldsymbol{H}}^2 + \|\boldsymbol{b} - \boldsymbol{c}\|_{\boldsymbol{H}}^2 - \|\boldsymbol{b} - \boldsymbol{d}\|_{\boldsymbol{H}}^2 \right).$$

将此结果应用于 $\boldsymbol{H} = \boldsymbol{C}$，可得

$$(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)})^\top \boldsymbol{C}(\boldsymbol{\alpha}^{(t)} - \widetilde{\boldsymbol{\alpha}}^{(t)}) = \frac{1}{2}\left\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)}\right\|_{\boldsymbol{C}}^2 - \frac{1}{2}\left\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^{(t)}\right\|_{\boldsymbol{C}}^2 + \frac{1}{2}\left\|\widetilde{\boldsymbol{\alpha}}^{(t)} - \boldsymbol{\alpha}^{(t)}\right\|_{\boldsymbol{C}}^2. \quad (9.113)$$

类似地，利用 $\widetilde{\boldsymbol{\nu}}^{(t)}$ 和 $\boldsymbol{\nu}^{(t+1)}$ 的定义，可推得

$$\begin{aligned} 2(\boldsymbol{\nu} - \widetilde{\boldsymbol{\nu}}^{(t)})^\top (\boldsymbol{\nu}^{(t)} - \boldsymbol{\nu}^{(t+1)}) &= \|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t+1)}\|_2^2 - \|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)}\|_2^2 + \|\widetilde{\boldsymbol{\nu}}^{(t)} - \boldsymbol{\nu}^{(t)}\|_2^2 - \|\widetilde{\boldsymbol{\nu}}^{(t)} - \boldsymbol{\nu}^{(t+1)}\|_2^2 \\ &= \|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t+1)}\|_2^2 - \|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)}\|_2^2 + \rho^2 \left\|\boldsymbol{X}\widetilde{\boldsymbol{\beta}}^{(t)} + \boldsymbol{Z}\boldsymbol{\alpha}^{(t)} - \boldsymbol{f}\right\|_2^2 - \rho^2 \left\|\boldsymbol{Z}(\boldsymbol{\alpha}^{(t)} - \widetilde{\boldsymbol{\alpha}}^{(t)})\right\|_2^2. \end{aligned}$$

因此，

$$\frac{1}{\rho}(\boldsymbol{\nu} - \widetilde{\boldsymbol{\nu}}^{(t)})^\top (\boldsymbol{\nu}^{(t)} - \boldsymbol{\nu}^{(t+1)}) \geq \frac{1}{2\rho}\left(\|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t+1)}\|_2^2 - \|\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)}\|_2^2 \right) - \frac{\rho}{2}\|\boldsymbol{Z}(\boldsymbol{\alpha}^{(t)} - \widetilde{\boldsymbol{\alpha}}^{(t)})\|_2^2. \quad (9.114)$$

现在定义块对角矩阵和堆叠向量：

$$\boldsymbol{A} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\rho}\boldsymbol{I} \end{bmatrix}, \qquad \boldsymbol{u} \triangleq \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \\ \boldsymbol{\nu} \end{bmatrix}, \qquad \boldsymbol{u}^{(t)} \triangleq \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \boldsymbol{\alpha}^{(t)} \\ \boldsymbol{\nu}^{(t)} \end{bmatrix}, \quad \text{and} \quad \widetilde{\boldsymbol{u}}^{(t)} \triangleq \begin{bmatrix} \widetilde{\boldsymbol{\beta}}^{(t)} \\ \widetilde{\boldsymbol{\alpha}}^{(t)} \\ \widetilde{\boldsymbol{\nu}}^{(t)} \end{bmatrix}.$$

结合式（9.113）和（9.114），得到

$$\left\langle \begin{bmatrix} \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}^{(t)} \\ \boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}^{(t)} \\ \boldsymbol{\nu} - \widetilde{\boldsymbol{\nu}}^{(t)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{0}(\boldsymbol{\beta}^{(t)} - \widetilde{\boldsymbol{\beta}}^{(t)}) \\ \boldsymbol{C}(\boldsymbol{\alpha}^{(t)} - \widetilde{\boldsymbol{\alpha}}^{(t)}) \\ \frac{1}{\rho}(\boldsymbol{\nu}^{(t)} - \boldsymbol{\nu}^{(t+1)}) \end{bmatrix} \right\rangle \geq \frac{1}{2}\left\|\boldsymbol{u} - \boldsymbol{u}^{(t+1)}\right\|_{\boldsymbol{A}}^2 - \frac{1}{2}\left\|\boldsymbol{u} - \boldsymbol{u}^{(t)}\right\|_{\boldsymbol{A}}^2.$$

将前面的不等式与式（9.112）结合，我们得到，对于任意 $\boldsymbol{\beta} \in \mathrm{dom}(f)$，$\boldsymbol{\alpha} \in \mathrm{dom}(g)$，以及 $\boldsymbol{\nu} \in \mathbb{R}^n$，

$$F(\boldsymbol{\beta}, \boldsymbol{\alpha}) - F(\widetilde{\boldsymbol{\beta}}^{(t)}, \widetilde{\boldsymbol{\alpha}}^{(t)}) + \left\langle \boldsymbol{u} - \widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}\widetilde{\boldsymbol{u}}^{(t)} + \widetilde{\boldsymbol{f}} \right\rangle \geq \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{u}^{(t+1)}\|_{\boldsymbol{A}}^2 - \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{u}^{(t)}\|_{\boldsymbol{A}}^2, \quad (9.115)$$

其中

$$\boldsymbol{G} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X}^\top \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}^\top \\ -\boldsymbol{X} & -\boldsymbol{Z} & \boldsymbol{0} \end{bmatrix} \qquad \text{and} \qquad \widetilde{\boldsymbol{f}} \triangleq \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{f} \end{bmatrix}.$$

注意

$$\left\langle \boldsymbol{u}-\widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}\widetilde{\boldsymbol{u}}^{(t)}+\widetilde{\boldsymbol{f}}\right\rangle = \left\langle \boldsymbol{u}-\widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}(\widetilde{\boldsymbol{u}}^{(t)}-\boldsymbol{u})+\boldsymbol{G}\boldsymbol{u}+\widetilde{\boldsymbol{f}}\right\rangle = \left\langle \boldsymbol{u}-\widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}\boldsymbol{u}+\widetilde{\boldsymbol{f}}\right\rangle,$$

其中，最后一个等式由以下事实可得： $\boldsymbol{G}$ 是斜对称的（即 $\boldsymbol{G}^{\top}=-\boldsymbol{G}$）。因此，式（9.115）可改写为

$$F(\boldsymbol{\beta},\boldsymbol{\alpha})-F(\widetilde{\boldsymbol{\beta}}^{(t)},\widetilde{\boldsymbol{\alpha}}^{(t)})+\left\langle \boldsymbol{u}-\widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}\boldsymbol{u}+\widetilde{\boldsymbol{f}}\right\rangle \geq \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{u}^{(t+1)}\|_{\boldsymbol{A}}^2-\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{u}^{(t)}\|_{\boldsymbol{A}}^2.$$

P对上述不等式在 $t=1$，2，……，T 上求和，得到不等式

$$T\cdot F(\boldsymbol{\beta},\boldsymbol{\alpha})-\sum_{t=1}^{T}F(\widetilde{\boldsymbol{\beta}}^{(t)},\widetilde{\boldsymbol{\alpha}}^{(t)})+\left\langle T\cdot\boldsymbol{u}-\sum_{t=1}^{T}\widetilde{\boldsymbol{u}}^{(t)}, \boldsymbol{G}\boldsymbol{u}+\widetilde{\boldsymbol{f}}\right\rangle \geq -\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{u}^{(1)}\|_{\boldsymbol{A}}^2.$$

定义

$$\overline{\boldsymbol{u}}^{(T)}\triangleq\frac{1}{T}\sum_{t=1}^{T}\widetilde{\boldsymbol{u}}^{(t)},\qquad \overline{\boldsymbol{\beta}}^{(T)}\triangleq\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}^{(t+1)},\qquad \overline{\boldsymbol{\alpha}}^{(T)}\triangleq\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\alpha}^{(t+1)},$$

并利用 F 的凸性，可得

$$\begin{aligned}&F(\boldsymbol{\beta},\boldsymbol{\alpha})-F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})+\left\langle \boldsymbol{u}-\overline{\boldsymbol{u}}^{(T)}, \boldsymbol{G}\boldsymbol{u}+\widetilde{\boldsymbol{f}}\right\rangle+\frac{1}{2T}\|\boldsymbol{u}-\boldsymbol{u}^{(1)}\|_{\boldsymbol{A}}^2\geq 0\\ \Longrightarrow\ &F(\boldsymbol{\beta},\boldsymbol{\alpha})-F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})+\left\langle \boldsymbol{u}-\overline{\boldsymbol{u}}^{(T)}, \boldsymbol{G}\overline{\boldsymbol{u}}^{(T)}+\widetilde{\boldsymbol{f}}\right\rangle+\frac{1}{2T}\|\boldsymbol{u}-\boldsymbol{u}^{(1)}\|_{\boldsymbol{A}}^2\geq 0,\end{aligned}$$

其中，该推论同样由 $\boldsymbol{G}$ 的斜对称性得出。换言之，对于任意 $\boldsymbol{\beta}\in\mathrm{dom}(f)$ 和 $\boldsymbol{\alpha}\in\mathrm{dom}(g)$,

$$F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})-F(\boldsymbol{\beta},\boldsymbol{\alpha})+\left\langle \overline{\boldsymbol{u}}^{(T)}-\boldsymbol{u}, \boldsymbol{G}\overline{\boldsymbol{u}}^{(T)}+\widetilde{\boldsymbol{f}}\right\rangle \leq \frac{1}{2T}\|\boldsymbol{u}-\boldsymbol{u}^{(1)}\|_{\boldsymbol{A}}^2. \tag{9.116}$$

令 $(\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*)$ 为问题 (9.104a) 的一个最优解。那么 $\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{Z}\boldsymbol{\alpha}^*=\boldsymbol{f}$。令 $\overline{\boldsymbol{\nu}}^{(T)}\triangleq\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}^{(t)}$。将 $\boldsymbol{\beta}=\boldsymbol{\beta}^*$, $\boldsymbol{\alpha}=\boldsymbol{\alpha}^*$ 以及 $\boldsymbol{u}^{(T)}$, $\boldsymbol{u}$, $\boldsymbol{u}^{(1)}$, $\boldsymbol{G}$,~$\boldsymbol{Z}$ 的表达式代入 (9.116)，得到

$$\begin{aligned}&F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})-F(\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*)+\left\langle \overline{\boldsymbol{\beta}}^{(T)}-\boldsymbol{\beta}^*, \boldsymbol{X}^{\top}\overline{\boldsymbol{\nu}}^{(T)}\right\rangle+\left\langle \overline{\boldsymbol{\alpha}}^{(T)}-\boldsymbol{\alpha}^*, \boldsymbol{Z}^{\top}\overline{\boldsymbol{\nu}}^{(T)}\right\rangle\\ &+\left\langle \overline{\boldsymbol{\nu}}^{(T)}-\boldsymbol{\nu}, -\boldsymbol{X}\overline{\boldsymbol{\beta}}^{(T)}-\boldsymbol{Z}\overline{\boldsymbol{\alpha}}^{(T)}+\boldsymbol{f}\right\rangle \quad\leq\quad \frac{1}{2T}\left\{\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}^{(1)}\|_{\boldsymbol{C}}^2+\frac{1}{\rho}\|\boldsymbol{\nu}-\boldsymbol{\nu}^{(1)}\|_2^2\right\}.\end{aligned}$$

消去 t 项，并利用 $\boldsymbol{X}\boldsymbol{\beta}^*+\boldsymbol{Z}\boldsymbol{\alpha}^*=\boldsymbol{f}$ 这一事实，上述不等式化简为

$$F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})-F(\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*)+\left\langle \boldsymbol{\nu}, \boldsymbol{X}\overline{\boldsymbol{\beta}}^{(T)}+\boldsymbol{Z}\overline{\boldsymbol{\alpha}}^{(T)}-\boldsymbol{f}\right\rangle \leq \frac{\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}^{(1)}\|_{\boldsymbol{C}}^2+\frac{1}{\rho}\|\boldsymbol{\nu}-\boldsymbol{\nu}^{(1)}\|_2^2}{2T}.$$

取 $\boldsymbol{\nu}\in\mathbb{B}[\boldsymbol{0},\zeta]$ 为两边的最大值，得到

$$F(\overline{\boldsymbol{\beta}}^{(T)},\overline{\boldsymbol{\alpha}}^{(T)})-F(\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*)+\zeta\left\|\boldsymbol{X}\overline{\boldsymbol{\beta}}^{(T)}+\boldsymbol{Z}\overline{\boldsymbol{\alpha}}^{(T)}-\boldsymbol{f}\right\|_2 \leq \frac{\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}^{(1)}\|_{\boldsymbol{C}}^2+\frac{1}{\rho}(\zeta+\|\boldsymbol{\nu}^{(1)}\|_2)^2}{2T}.$$

证明至此完成。 ■

事实上，利用原问题（9.104a）与其对偶问题（9.104c）之间的强对偶性，并假设 $\zeta \geq 2\|\boldsymbol{\nu}^*\|_2$，还可以建立以下可行性残差上界：

$$\left\|\boldsymbol{D}\overline{\boldsymbol{\beta}}^{(T)} + \boldsymbol{Z}\overline{\boldsymbol{\alpha}}^{(T)} - \boldsymbol{f}\right\|_2 \leq \frac{\left\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}^{(1)}\right\|_{\boldsymbol{C}}^2 + \frac{1}{\rho}\left(\zeta + \left\|\boldsymbol{\nu}^{(1)}\right\|_2\right)^2}{\zeta T},$$

参见Beck（2017年）了解更多详情。

▶ **将交替方向乘子法应用于 LASSO**。注意，拉格朗日 LASSO问题（L_L）可通过引入辅助变量等价地改写如下：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}) \triangleq f(\boldsymbol{\beta}) + g(\boldsymbol{\alpha}) \triangleq \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \quad \text{s.t. } \boldsymbol{\beta} = \boldsymbol{\alpha}.$$

该表述符合标准的交替方向乘子法框架。给定平滑参数 $\rho > 0$，第 t 次迭代的更新如下：

$$\begin{aligned}
\boldsymbol{\beta}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\beta} - \boldsymbol{\alpha}^{(t)} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right\|_2^2 \right\} \\
&= \left(\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^\top\boldsymbol{y} + \rho\boldsymbol{\alpha}^{(t)} - \boldsymbol{\nu}^{(t)}\right); \\
\boldsymbol{\alpha}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{\alpha}} \left\{ \lambda\|\boldsymbol{\alpha}\|_1 + \frac{\rho}{2}\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right\|_2^2 \right\} \\
&= \mathcal{P}\mathrm{rox}_{(\lambda/\rho)\|\cdot\|_1}\left(\boldsymbol{\beta}^{(t+1)} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right) = \mathcal{T}_{(\lambda/\rho)}(\boldsymbol{\beta}); \\
\boldsymbol{\nu}^{(t+1)} &\leftarrow \boldsymbol{\nu}^{(t)} + \rho(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t+1)}),
\end{aligned}$$

其中 $\mathcal{T}_\lambda(\cdot)$ 表示软阈值算子（例 2.104）。

由于 $\rho > 0$，$\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{I}$ 始终可逆。对 $\boldsymbol{\beta}$ 的更新本质上是在求解一个岭回归问题（一个带 ℓ_2-范数平方正则项的最小二乘问题）；而对 $\boldsymbol{\alpha}$ 的更新涉及一个 ℓ_1-范数近端算子，可通过软阈值处理得到闭式解。在求解 $\boldsymbol{\beta}$-子问题时，如果在所有迭代中使用固定的惩罚参数 ρ，则可以预先计算并缓存 $\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{I}$的矩阵分解（例如 Cholesky分解）；参见例如 Lu （2022年）。这显著降低了后续迭代中的计算成本，因为昂贵的分解只需执行一次。

9.3.6 块坐标下降

我们简要介绍如何使用块坐标下降（*BCD*）方法求解 LASSO 问题。由于目标函数中的 $\|\boldsymbol{\beta}\|_1$ 项具有可分离性，每个块变量对应 $\boldsymbol{\beta}$ 的一个坐标。为方便起见，在更新第 i 个块时，我们将 $\boldsymbol{\beta}$ 表示为：

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_i \\ \boldsymbol{\beta}_{-i} \end{bmatrix},$$

其中，β_i 表示 $\boldsymbol{\beta}$的第 i个分量，而 $\boldsymbol{\beta}_{-i} \in \mathbb{R}^{p-1}$ 包含除第 i个分量之外的所有分量。类似地，我们将设计矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 划分为：

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_i & \boldsymbol{X}_{-i} \end{bmatrix},$$

其中，$\boldsymbol{x}_i \in \mathbb{R}^n$ 是 $\boldsymbol{X}$的第 i列，而 $\boldsymbol{X}_{-i} \in \mathbb{R}^{n\times p-1}$ 由除第 i列之外的所有列组成。这种重新排序不会改变问题，只是将正在优化的变量单独分离出来。

更新第 i 个坐标时，LASSO 目标函数变为：

$$\min_{\beta_i} \lambda|\beta_i| + \lambda\left\|\boldsymbol{\beta}_{-i}\right\|_1 + \frac{1}{2}\left\|\boldsymbol{x}_i\beta_i - (\boldsymbol{y} - \boldsymbol{X}_{-i}\boldsymbol{\beta}_{-i})\right\|_2^2.$$

由于在此步骤中 $\boldsymbol{\beta}_{-i}$ 固定，因此项 $\lambda\left\|\boldsymbol{\beta}_{-i}\right\|_1$ 是常数，在优化时可以忽略。定义残差向量 $\boldsymbol{r}_i \triangleq \boldsymbol{y} - \boldsymbol{X}_{-i}\boldsymbol{\beta}_{-i}$。于是子问题可化简为：

$$\min_{\beta_i} f_i(\beta_i) \triangleq \lambda|\beta_i| + \frac{1}{2}\|\boldsymbol{x}_i\|_2^2\beta_i^2 - \boldsymbol{x}_i^\top\boldsymbol{r}_i\beta_i. \tag{9.117}$$

(9.117) 的极小化解有闭式解，可通过软阈值求得：

$$\beta_i^{(t)} \leftarrow \arg\min_{\beta_i} f_i(\beta_i) = \begin{cases} \frac{\boldsymbol{x}_i^\top\boldsymbol{r}_i - \lambda}{\|\boldsymbol{x}_i\|_2^2}, & \boldsymbol{x}_i^\top\boldsymbol{r}_i > \lambda; \\ \frac{\boldsymbol{x}_i^\top\boldsymbol{r}_i + \lambda}{\|\boldsymbol{x}_i\|_2^2}, & \boldsymbol{x}_i^\top\boldsymbol{r}_i < -\lambda; \\ 0, & \text{otherwise.} \end{cases} \tag{9.118}$$

因此，LASSO 问题的块坐标下降算法通过循环更新每个坐标来执行，更新时使用 (9.118)。完整过程总结于算法 14。

算法 14 LASSO 块坐标下降法

1: **输入：** $\boldsymbol{X}$，$\boldsymbol{y}$，参数 λ；2: **初始化：** $\boldsymbol{\beta}^{(1)}$；3: **对于** $t = 1$，2，...**执行**4: **对于** $i = 1$，2，...，p **执行**5: 根据定义计算 $\boldsymbol{\beta}_{-i}$，$\boldsymbol{r}_i$；6: 根据（9.118）计算 $\beta_i^{(t)}$；7: **结束循环**8: 如果在第 $t = T$ 次迭代时满足停止准则，则停止；9: **结束循环**10: **返回：** 最终 $\boldsymbol{\beta}^{(T)}$;

9.3.7 Barzilai–Borwein 梯度投影算法

考虑标准的有界约束二次规划（*BCQP*）问题

$$\min_{\boldsymbol{\alpha}} \left\{ Q(\boldsymbol{\alpha}) \triangleq \frac{1}{2}\boldsymbol{\alpha}^\top\boldsymbol{Y}\boldsymbol{\alpha} - \boldsymbol{z}^\top\boldsymbol{\alpha} \right\} \quad \text{s.t.} \quad \boldsymbol{\ell} \le \boldsymbol{\alpha} \le \boldsymbol{u}. \tag{9.119}$$

投影梯度下降（PGD；参见算法3）方法提供了一种求解大型有界约束二次规划问题的方法。令 $\mathbb{S}$ 表示（ 9.119 ）的可行集：

$$\mathbb{S} \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^p \mid \boldsymbol{\ell} \le \boldsymbol{\alpha} \le \boldsymbol{u}\} \quad \Longrightarrow \quad \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) = \operatorname{mid}(\boldsymbol{\ell}, \boldsymbol{\alpha}, \boldsymbol{u}), \tag{9.120}$$

其中，$\mathcal{P}_{\mathbb{S}}$ 表示到 $\mathbb{S}$上的投影算子，其具有简单的闭式表达；即，$\mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}) = \operatorname{mid}(\boldsymbol{\ell}, \boldsymbol{\alpha}, \boldsymbol{u})$ 是一个向量，其第 i 个分量是集合的中位数

$\{\ell_i, \alpha_i, u_i\}$。给定当前的可行迭代点 $\boldsymbol{\alpha}^{(t)}$，投影梯度法计算下一个迭代点为

$$\boldsymbol{\alpha}^{(t+1)} = \mathcal{P}_{\mathbb{S}}(\boldsymbol{\alpha}^{(t)} - \eta_t \boldsymbol{g}^{(t)}),$$

其中 $\eta_t > 0$ 是某个步长，$\boldsymbol{g}^{(t)} \triangleq \boldsymbol{Y}\boldsymbol{\alpha}^{(t)} - \boldsymbol{z}$ 表示梯度；参见第 9.2.1节。Barzilai 和 Borwein (1988) 提出了一种非常有效的步长选择策略η_t，如今称为*Barzilai–Borwein*（*BB*）规则。两种常见变体如下：

$$\eta_t^{BB1} = \frac{\boldsymbol{s}^{(t-1)\top}\boldsymbol{s}^{(t-1)}}{\boldsymbol{s}^{(t-1)\top}\boldsymbol{t}^{(t-1)}} \qquad \text{and} \qquad \eta_t^{BB2} = \frac{\boldsymbol{s}^{(t-1)\top}\boldsymbol{t}^{(t-1)}}{\boldsymbol{t}^{(t-1)\top}\boldsymbol{t}^{(t-1)}}, \tag{9.121}$$

其中 $\boldsymbol{s}^{(t-1)} = \boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}^{(t-1)}$ 和 $\boldsymbol{t}^{(t-1)} = \boldsymbol{g}^{(t)} - \boldsymbol{g}^{(t-1)}$。

结合 BB 步长时，投影梯度法称为*Barzilai–Borwein* 梯度投影（*BBGP*）法。BBGP 的一个显著特点是它具有非单调性：目标函数值 $Q(\boldsymbol{\alpha}^{(t)})$ 在某些迭代中可能增加。尽管如此，该方法简单易用，无需进行矩阵分解，而且在实际应用中通常表现出较快的收敛速度。

▶ **采用 BB 梯度投影的拉格朗日 LASSO**。有趣的是，拉格朗日 LASSO 问题（L_L）最初是一个无约束凸优化问题，可以改写为有界约束二次规划问题。具体来说，考虑

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \quad \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{X} \in \mathbb{R}^{n\times p}, \boldsymbol{y} \in \mathbb{R}^n. \tag{9.122}$$

为了将其转化为有界约束问题，我们采用标准的变量拆分技术：对于每个分量 β_i，引入非负变量 $u_i = (\beta_i)_+$ 和 $\ell_i = (-\beta_i)_+$，对所有 $i = 1, 2, \ldots, p$，其中 $(\beta)_+ = \max(\beta, 0)$ 表示正部算子。然后，系数向量 $\boldsymbol{\beta}$ 可以拆分为两个非负向量之差：

$$\boldsymbol{\beta} = \boldsymbol{u} - \boldsymbol{\ell}, \quad \boldsymbol{u} \geq \boldsymbol{0}, \quad \boldsymbol{\ell} \geq \boldsymbol{0}. \tag{9.123}$$

因此，$\|\boldsymbol{\beta}\|_1 = \boldsymbol{1}_p^\top \boldsymbol{u} + \boldsymbol{1}_p^\top \boldsymbol{\ell}$，其中 $\boldsymbol{1}_p = [1, 1, \ldots, 1]^\top$ 是一个由 $p \times 1$ 个 1 组成的求和向量。令 p，拉格朗日 LASSO 问题（9.122）可改写为以下有界约束二次规划问题（Figueiredo 等人，2008年）：

$$\min_{\boldsymbol{\alpha}} \left\{ F(\boldsymbol{\alpha}) \triangleq \boldsymbol{c}^\top \boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{Y}\boldsymbol{\alpha} \right\} \quad \text{s.t.} \quad \boldsymbol{\alpha} \geq \boldsymbol{0}, \tag{9.124}$$

其中

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}, \qquad \boldsymbol{c} = \lambda \boldsymbol{1}_{2p} + \begin{bmatrix} -\boldsymbol{X}^\top \boldsymbol{y} \\ \boldsymbol{X}^\top \boldsymbol{y} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{X}^\top \boldsymbol{X} & -\boldsymbol{X}^\top \boldsymbol{X} \\ -\boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{X} \end{bmatrix}. \tag{9.125}$$

为求解这个带有非负约束的有界约束二次规划问题（9.124），Figueiredo 等人（2008年）提出了以下两步更新：

$$\boldsymbol{w}^{(t)} \leftarrow \left(\boldsymbol{\alpha}^{(t)} - \eta_t \nabla F(\boldsymbol{\alpha}^{(t)})\right)_+; \tag{9.126a}$$

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \boldsymbol{\alpha}^{(t)} + \lambda_t (\boldsymbol{w}^{(t)} - \boldsymbol{\alpha}^{(t)}), \tag{9.126b}$$

其中 $\lambda_t \in [0, 1]$。步骤（9.126a）对应于标准的投影梯度下降步（投影到非负正交象限）。步骤（9.126b）引入了一种外推（或松弛），其思路与第 9.2.3 节讨论的条件梯度法类似。参数 λ_t 通常通过回溯线搜索来选取，以确保目标函数 F（Bertsekas，1997 年）充分下降。

9.4. 其他稀疏优化建模形式

除了标准 LASSO 问题外，我们还简要介绍三种重要的变体：广义 LASSO、组 LASSO 和稀疏噪声 LASSO。

在广义 LASSO 中，我们假设回归系数呈现特定的结构模式，例如光滑性、分段常数性或潜在趋势。该建模形式不只是惩罚系数的大小，还会惩罚系数之间的差异或高阶变化。它广泛应用于信号处理、图像去噪和时间序列分析等领域，因为这些领域中此类结构性假设是自然存在的。

组 LASSO 适用于预测变量自然形成分组的场景，例如生物通路中的基因、用于编码分类特征的虚拟变量，或多通道传感器测量值。在此类情况下，组 LASSO 会同时选择或舍弃整个变量组。这种方法保留了模型的可解释性，并遵循数据固有的分组结构。

稀疏噪声 LASSO（也称为稳健 LASSO）用于处理观测值受到偶发大误差或离群值污染的场景。它不假设所有噪声都较小且服从高斯分布，而是对数据中的稀疏但显著的（即“粗大”）误差进行显式建模。在金融数据分析、传感器网络或临床测量值等现实应用中，这一变体尤其有用，因为罕见但严重的异常值可能损害数据完整性。

9.4.1 广义 LASSO 问题

广义 LASSO 问题定义如下：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{Z}\boldsymbol{\beta}\|_1, \tag{9.127}$$

其中 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$、$\boldsymbol{Z}\in\mathbb{R}^{l\times p}$ 和 $\boldsymbol{y}\in\mathbb{R}^n$。在标准 LASSO 中，$\|\boldsymbol{\beta}\|_1$ 惩罚项促使 $\boldsymbol{\beta}$具有稀疏性。然而，在许多应用中，$\boldsymbol{\beta}$ 本身可能并不稀疏，但经过线性变换后会变得稀疏。一个典型例子是 $\boldsymbol{Z}\in\mathbb{R}^{(p-1)\times p}$ 为一阶差分矩阵：

$$z_{ij}=\begin{cases}1, & j=i+1,\\ -1, & j=i,\\ 0, & \text{otherwise},\end{cases}$$

且 $\boldsymbol{X}=\boldsymbol{I}$。在这种情况下，广义 LASSO 可化简为：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{\beta}-\boldsymbol{y}\|_2^2+\lambda\sum_{i=1}^{p-1}|\beta_{i+1}-\beta_i|,$$

这种形式对应于图像去噪中使用的全变差（*TV*）模型。当 $\boldsymbol{X}=\boldsymbol{I}$ 和 $\boldsymbol{Z}$ 为二阶差分矩阵时，问题（9.127）称为趋势滤波（Rudin 等人，1992年；Kim 等人，2009年）。[3]

▶ **使用交替方向乘子法求解广义 LASSO。** 在第9.3.5节中，我们介绍了使用交替方向乘子法求解标准（拉格朗日函数）LASSO 问题的方法。这里，我们将该方法扩展到广义 LASSO。为了将交替方向乘子法应用于问题（9.127），我们引入一个

3。关于其去噪效果的示例，请参见问题9.14。

辅助变量 $\boldsymbol{\alpha}$ 并施加约束 $\boldsymbol{Z}\boldsymbol{\beta} = \boldsymbol{\alpha}$。由此得到等价的约束问题：

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \quad \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{\alpha} = \boldsymbol{0}.$$

令 $\boldsymbol{\nu} \in \mathbb{R}^l$ 为与该约束相关联的拉格朗日乘子（假设 $\boldsymbol{Z} \in \mathbb{R}^{l\times p}$）。那么增广拉格朗日函数为：

$$L_\rho(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\nu}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 + \boldsymbol{\nu}^\top(\boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{\alpha}) + \frac{\rho}{2}\|\boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2,$$

其中 $\rho > 0$ 为惩罚参数。与标准 LASSO 情形一样，交替方向乘子法（ADMM）的更新将光滑二次项与非光滑的 ℓ_1 惩罚项分离开来。具体而言，$\boldsymbol{\beta}$ 更新需要求解一个线性方程组：

$$\left(\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{Z}^\top\boldsymbol{Z}\right)\boldsymbol{\beta} = \boldsymbol{X}^\top\boldsymbol{y} + \rho\boldsymbol{Z}^\top\left(\boldsymbol{\alpha}^{(t)} - \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right),$$

而 $\boldsymbol{\alpha}$ 更新则通过 ℓ_1 范数的近端算子来完成。因此，对于给定的 $\rho > 0$，ADMM 迭代（参见第 9.3.5 节）如下：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \left(\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{X}^\top\boldsymbol{y} + \rho\boldsymbol{Z}^\top\left(\boldsymbol{\alpha}^{(t)} - \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right)\right); \tag{9.128a}$$

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{(\lambda/\rho)\|\cdot\|_1}\left(\boldsymbol{Z}\boldsymbol{\beta}^{(t+1)} + \frac{1}{\rho}\boldsymbol{\nu}^{(t)}\right); \tag{9.128b}$$

$$\boldsymbol{\nu}^{(t+1)} \leftarrow \boldsymbol{\nu}^{(t)} + \rho(\boldsymbol{Z}\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t+1)}). \tag{9.128c}$$

9.4.2 组 LASSO 与对偶多面体投影

当 $\boldsymbol{X}$ 的列可以自然地划分为互不相交的子集 $\mathbb{G} = \{g_1, g_2, \ldots, g_G\}$，且每个子集对应一组连贯的特征（例如，通路中的基因或分类预测变量的虚拟变量）时，组 LASSO 会在组级而非单个系数层面施加稀疏性（Yuan 和 Lin，2006年）。具体而言，这些组涵盖了 p 的所有列 $\boldsymbol{X}$，即

$$\bigcup_{g\in\mathbb{G}} g = \{1, 2, \ldots, p\},$$

通常假设各组之间互不重叠（不过也存在针对重叠情形的扩展）。为了实现组级稀疏性，我们对 ℓ_1 中的一个向量施加$\mathbb{R}^{|\mathbb{G}|}$惩罚，该向量的各个元素是每组内系数子向量的 ℓ_2范数。（在组内使用 ℓ_2范数本身并不会促使模型按组进行选择。）这种策略避免直接惩罚单个系数，而是促使整组系数共同取零或非零。

当领域知识表明预测变量构成有意义的组时，这种方法尤其有价值；它能让模型选择或舍弃整组相关特征，从而提升可解释性和统计效率。

更正式地，令 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 和 $\boldsymbol{y} \in \mathbb{R}^n$ 分别表示基于 n 个观测值构成的设计矩阵和响应向量。参数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 被划分为 $G = |\mathbb{G}|$ 个块：

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \ldots, \boldsymbol{\beta}_G^\top]^\top \in \mathbb{R}^p,$$

其中，$\boldsymbol{\beta}_g \in \mathbb{R}^{p_g}$ 对应于第 g组，且 $\sum_{g=1}^{G} p_g = p$。组 LASSO 假设只有少数几个 $\boldsymbol{\beta}_g$ 非零。相应的优化问题为：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \lambda\sum_{g=1}^{G}\sqrt{p_g}\,\|\boldsymbol{\beta}_g\|_2,$$

具有特征组 $\boldsymbol{X}\boldsymbol{\beta} = \sum_{g=1}^{G}\boldsymbol{X}_g\boldsymbol{\beta}_g$，其中 $\boldsymbol{X}_g \in \mathbb{R}^{n\times p_g}$ 表示$\boldsymbol{X}$ 包含组 g中各列的子矩阵。因子 $\sqrt{p_g}$ 用于补偿不同组的大小差异，确保对不同维数的组进行公平惩罚。正则化参数 $\lambda > 0$ 控制数据保真度与组稀疏性之间的权衡。请注意，该惩罚项对每个组使用（未平方的）欧几里得范数 $\|\boldsymbol{\beta}_g\|_2$，因此这是一种范数和正则化（Ohlsson 等人，2010年）。

使用交替方向乘子法的组 LASSO

遵循交替方向乘子法框架（参见第9.3.5节），我们引入辅助变量 $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \ldots, \boldsymbol{\alpha}_G^\top]^\top \in \mathbb{R}^p$，以将光滑损失与非光滑组惩罚项解耦。该问题可等价地重写为：

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \lambda\sum_{g=1}^{G}\sqrt{p_g}\,\|\boldsymbol{\alpha}_g\|_2 \quad \text{s.t.} \quad \boldsymbol{\beta}_g = \boldsymbol{\alpha}_g,\ \forall g.$$

令 $\boldsymbol{\nu} = [\boldsymbol{\nu}_1^\top, \boldsymbol{\nu}_2^\top, \ldots, \boldsymbol{\nu}_G^\top]^\top \in \mathbb{R}^p$ 表示与这些约束条件相关的缩放对偶变量。缩放形式的增广拉格朗日函数为：

$$L_\rho(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\nu}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \lambda\sum_{g=1}^{G}\sqrt{p_g}\,\|\boldsymbol{\alpha}_g\|_2 + \langle\boldsymbol{\nu},\boldsymbol{\beta}-\boldsymbol{\alpha}\rangle + \frac{\rho}{2}\sum_{g=1}^{G}\|\boldsymbol{\beta}_g-\boldsymbol{\alpha}_g\|_2^2,$$

其中，$\boldsymbol{\alpha}_g$ 是与等式约束 $\boldsymbol{\beta}_g = \boldsymbol{\alpha}_g$相关的缩放拉格朗日乘子，$\rho > 0$ 是惩罚参数。

交替方向乘子法通过迭代更新 $\boldsymbol{\beta}$、$\boldsymbol{\alpha}$和 $\boldsymbol{\nu}$来进行；参见（9.107b）。$\boldsymbol{\beta}$的更新为

$$\boldsymbol{\beta}^{(t+1)} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\beta}-\boldsymbol{\alpha}^{(t)}+\boldsymbol{\nu}^{(t)}\right\|_2^2.$$

这是一个岭回归类型的问题，其闭式解为：

$$\boldsymbol{\beta}^{(t+1)} = (\boldsymbol{X}^\top\boldsymbol{X} + \rho\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{y} + \rho(\boldsymbol{\alpha}^{(t)} - \boldsymbol{\nu}^{(t)})).$$

对于大规模问题，可以使用共轭梯度法等迭代求解器，或预先计算的因式分解来加速计算（Lu，2025年）。

$\boldsymbol{\alpha}$的更新为

$$\boldsymbol{\alpha}^{(t+1)} = \arg\min_{\boldsymbol{\alpha}} \lambda\sum_{g=1}^{G}\sqrt{p_g}\,\|\boldsymbol{\alpha}_g\|_2 + \frac{\rho}{2}\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\alpha}+\boldsymbol{\nu}^{(t)}\right\|_2^2$$

这会按组进行分解。对于每个组 $g \in \{1, 2, \ldots, G\}$，更新由缩放后的 ℓ_2 范数的近端算子给出

$$\begin{aligned}\boldsymbol{\alpha}_g^{(t+1)} &= \mathcal{P}\mathrm{rox}_{(\lambda\sqrt{p_g}/\rho)\|\cdot\|_2}(\boldsymbol{\beta}_g^{(t+1)} + \boldsymbol{\nu}_g^{(t)}) \\ &= \begin{cases} \left(1 - \dfrac{\lambda\sqrt{p_g}}{\rho\left\|\boldsymbol{\beta}_g^{(t+1)} + \boldsymbol{\nu}_g^{(t)}\right\|_2}\right)(\boldsymbol{\beta}_g^{(t+1)} + \boldsymbol{\nu}_g^{(t)}), & \text{if } \left\|\boldsymbol{\beta}_g^{(t+1)} + \boldsymbol{\nu}_g^{(t)}\right\|_2 > \dfrac{\lambda\sqrt{p_g}}{\rho}; \\ 0, & \text{otherwise,} \end{cases}\end{aligned}$$

这称为块软阈值算子 （参见例 2.105）。

缩放对偶变量 $\boldsymbol{\nu}$更新为

$$\boldsymbol{\nu}^{(t+1)} = \boldsymbol{\nu}^{(t)} + (\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t+1)})$$

综合这些步骤，在步骤 t处的完整ADMM迭代为：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow (\boldsymbol{X}^\top \boldsymbol{X} + \rho \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{y} + \rho(\boldsymbol{\alpha}^{(t)} - \boldsymbol{\nu}^{(t)})); \tag{9.129a}$$

$$\boldsymbol{\alpha}_g^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{(\lambda\sqrt{p_g}/\rho)\|\cdot\|_2}(\boldsymbol{\beta}_g^{(t+1)} + \boldsymbol{\nu}_g^{(t)}), \quad \forall\, g; \tag{9.129b}$$

$$\boldsymbol{\nu}^{(t+1)} \leftarrow \boldsymbol{\nu}^{(t)} + (\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\alpha}^{(t+1)}). \tag{9.129c}$$

组LASSO对偶问题

再次令设计矩阵划分为 G组，即 $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_g]$，其中$\boldsymbol{X}_g \in \mathbb{R}^{n\times p_g}$ 和 $p = \sum_{g=1}^G p_g$。组 LASSO 问题等价地表示为：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\left\|\boldsymbol{y} - \sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g\right\|_2^2 + \lambda\sum_{g=1}^G \sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2. \tag{9.130}$$

引入残差向量 $\boldsymbol{r} \triangleq \boldsymbol{y} - \sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g$，我们将问题重新表述为：

$$\min_{\boldsymbol{\beta},\boldsymbol{r}} \frac{1}{2}\|\boldsymbol{r}\|_2^2 + \lambda\sum_{g=1}^G \sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2 \quad \text{s.t.} \quad \boldsymbol{r} = \boldsymbol{y} - \sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g. \tag{9.131}$$

令 $\boldsymbol{\nu} \in \mathbb{R}^n$ 表示与等式约束对应的对偶变量。问题（9.131）的拉格朗日函数为：

$$L(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\nu}) = \frac{1}{2}\|\boldsymbol{r}\|_2^2 + \lambda\sum_{g=1}^G \sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2 + \boldsymbol{\nu}^\top\left(\boldsymbol{y} - \sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g - \boldsymbol{r}\right). \tag{9.132}$$

T对拉格朗日函数关于 $\boldsymbol{\beta}$ 和 $\boldsymbol{r}$求最小值，即可得到对偶函数 $D(\boldsymbol{\nu})$ ：

$$\begin{aligned} D(\boldsymbol{\nu}) &= \min_{\boldsymbol{\beta},\boldsymbol{r}} L(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\nu}) \\ &= \boldsymbol{\nu}^\top\boldsymbol{y} + \min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left(-\boldsymbol{\nu}^\top\sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g + \lambda\sum_{g=1}^G \sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2\right) + \min_{\boldsymbol{r}\in\mathbb{R}^n}\left(\frac{1}{2}\|\boldsymbol{r}\|_2^2 - \boldsymbol{\nu}^\top\boldsymbol{r}\right). \end{aligned}$$

最后一项与标准 LASSO 对偶问题（3.38）一致，结果为 $-\frac{1}{2}\|\boldsymbol{\nu}\|_2^2$。现在我们来分析第二项。定义

$$\begin{aligned} h(\boldsymbol{\beta}) &= -\boldsymbol{\nu}^\top\sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g + \lambda\sum_{g=1}^G \sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2; \\ h_g(\boldsymbol{\beta}_g) &= -\boldsymbol{\nu}^\top\boldsymbol{X}_g\boldsymbol{\beta}_g + \lambda\sqrt{p_g}\left\|\boldsymbol{\beta}_g\right\|_2, \qquad g = 1, 2, \ldots, G. \end{aligned}$$

由于各组互不相交，且最小化问题可在各组之间解耦，因此我们可以将 $h(\boldsymbol{\beta})$ 拆分为一组子问题。显然， $h_g(\boldsymbol{\beta}_g)$ 是凸的，但不是光滑的，因为它在 $\mathbf{0}$ 处有一个奇异点。考虑 h_g 的次梯度，

$$\partial h_g(\boldsymbol{\beta}_g) = -\boldsymbol{X}_g^\top\boldsymbol{\nu} + \lambda\sqrt{p_g}\boldsymbol{u}_g, \quad \boldsymbol{u}_g \in \partial\left\|\boldsymbol{\beta}_g\right\|_2, \qquad g = 1, 2, \ldots, G,$$

其中 $\boldsymbol{u}_g$ 是 $\|\boldsymbol{\beta}_g\|_2$ (习题 2.58):

$$\boldsymbol{u}_g \in \begin{cases} \boldsymbol{\beta}_g / \|\boldsymbol{\beta}_g\|_2, & \text{if } \boldsymbol{\beta}_g \neq \mathbf{0}; \\ \mathbb{B}_2[\mathbf{0}, 1], & \text{if } \boldsymbol{\beta}_g = \mathbf{0}, \end{cases} \tag{9.133}$$

且 $\mathbb{B}_2[\mathbf{0}, 1]$ 表示单位 ℓ_1-球。 h_g 的最优解 $\boldsymbol{\beta}_g$ 满足

$$\exists \widehat{\boldsymbol{u}}_g \in \partial \|\widehat{\boldsymbol{\beta}}_g\|_2, \quad -\boldsymbol{X}_g^\top \boldsymbol{\nu} + \lambda \sqrt{p_g} \widehat{\boldsymbol{u}}_g = \mathbf{0}.$$

如果 $\widehat{\boldsymbol{\beta}}_g = \mathbf{0}$, 显然 $h_g(\widehat{\boldsymbol{\beta}}_g) = 0$。否则，由于 $\lambda \sqrt{p_g} \widehat{\boldsymbol{u}}_g = \boldsymbol{X}_g^\top \boldsymbol{\nu}$ 且 $\widehat{\boldsymbol{u}}_g = \widehat{\boldsymbol{\beta}}_g / \|\widehat{\boldsymbol{\beta}}_g\|_2$, 我们有

$$h_g(\widehat{\boldsymbol{\beta}}_g) = -\lambda \sqrt{p_g} \frac{\widehat{\boldsymbol{\beta}}_g^\top}{\|\widehat{\boldsymbol{\beta}}_g\|_2} \boldsymbol{\beta}_g + \lambda \sqrt{p_g} \|\widehat{\boldsymbol{\beta}}_g\|_2 = 0.$$

因此，这表明 $\min_{\boldsymbol{\beta}_g} h_g(\boldsymbol{\beta}_g) = 0, \quad g = 1, \ 2, \ \ldots\ldots, \ G$ 从而

$$\min_{\boldsymbol{\beta}} h(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \sum_{g=1}^{G} h_g(\boldsymbol{\beta}_g) = \sum_{g=1}^{G} \min_{\boldsymbol{\beta}_g} h_g(\boldsymbol{\beta}_g) = 0.$$

由（9.133），容易看出 $\|\boldsymbol{u}_g\|_2 \leq 1$。由于 $\lambda \sqrt{p_g} \widehat{\boldsymbol{u}}_g = \boldsymbol{X}_g^\top \boldsymbol{\nu}$, 我们得到关于 $\boldsymbol{\nu}$的一个约束，即 $\boldsymbol{\nu}$ 应满足:

$$\left\| \boldsymbol{X}_g^\top \boldsymbol{\nu} \right\|_2 \leq \lambda \sqrt{p_g}, \quad g = 1, 2, \ldots, G. \tag{9.134}$$

因此，对偶函数 $D(\boldsymbol{\nu})$为:

$$D(\boldsymbol{\nu}) = \boldsymbol{\nu}^\top \boldsymbol{y} - \frac{1}{2} \|\boldsymbol{\nu}\|_2^2 = \frac{1}{2} \|\boldsymbol{y}\|_2^2 - \frac{1}{2} \|\boldsymbol{\nu} - \boldsymbol{y}\|_2^2 .$$

组 LASSO 问题的对偶形式为:

$$\max_{\boldsymbol{\nu} \in \mathbb{R}^n} D(\boldsymbol{\nu}) = \frac{1}{2} \|\boldsymbol{y}\|_2^2 - \frac{1}{2} \|\boldsymbol{\nu} - \boldsymbol{y}\|_2^2 \quad \text{s.t.} \quad \left\| \boldsymbol{X}_g^\top \boldsymbol{\nu} \right\|_2 \leq \lambda \sqrt{p_g}, \ \forall g. \tag{9.135}$$

与标准 LASSO 一样，对偶变量 $\boldsymbol{\mu} \triangleq \frac{\boldsymbol{\nu}}{\lambda}$ 的简单缩放可得到一个等价形式:

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^n} D(\boldsymbol{\mu}) = \frac{1}{2} \|\boldsymbol{y}\|_2^2 - \frac{\lambda^2}{2} \left\| \boldsymbol{\mu} - \frac{\boldsymbol{y}}{\lambda} \right\|_2^2 \quad \text{s.t.} \quad \left\| \boldsymbol{X}_g^\top \boldsymbol{\mu} \right\|_2 \leq \sqrt{p_g}, \ \forall g. \tag{9.136}$$

▶ **稀疏性的对偶可行性条件**。显然，问题（9.131）是凸问题，且其约束条件均为仿射约束。根据 Slater 条件，只要原问题可行，强对偶性就成立。设（$\boldsymbol{\beta}$, $\widehat{\boldsymbol{r}}$, $\widehat{\boldsymbol{\nu}}$）是最优原始-对偶解。由拉格朗日函数（9.132）和 KKT 条件（定理 2.115），可得

$$\mathbf{0} \in \partial_{\boldsymbol{\beta}_g} L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}, \widehat{\boldsymbol{\nu}}) = -\boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}} + \lambda \sqrt{p_g} \boldsymbol{u}_g, \quad \text{where } \boldsymbol{u}_g \in \partial \|\widehat{\boldsymbol{\beta}}_g\|_2, \quad \forall g; \tag{9.137a}$$

$$\nabla_{\boldsymbol{r}} L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}, \widehat{\boldsymbol{\nu}}) = \widehat{\boldsymbol{r}} - \widehat{\boldsymbol{\nu}} = \mathbf{0}; \tag{9.137b}$$

$$\nabla_{\boldsymbol{\nu}} L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}, \widehat{\boldsymbol{\nu}}) = \boldsymbol{y} - \sum_{g=1}^{G} \boldsymbol{X}_g \widehat{\boldsymbol{\beta}}_g - \widehat{\boldsymbol{r}} = \mathbf{0}. \tag{9.137c}$$

由（9.137b）和（9.137c），可得：

$$\boldsymbol{y} = \sum_{g=1}^{G} \boldsymbol{X}_g \widehat{\boldsymbol{\beta}}_g + \widehat{\boldsymbol{\nu}}. \tag{9.138}$$

由（9.137a）可知，存在 $\widehat{\boldsymbol{u}}_g \in \partial \left\| \widehat{\boldsymbol{\beta}}_g \right\|_2$ 使得

$$\boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}} = \lambda \sqrt{p_g} \widehat{\boldsymbol{u}}_g. \tag{9.139}$$

于是，对于每个组 $g = 1, 2, \ldots, G$，都有：

$$\boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}} \in \begin{cases} \lambda \sqrt{p_g} \dfrac{\widehat{\boldsymbol{\beta}}_g}{\left\| \widehat{\boldsymbol{\beta}}_g \right\|_2}, & \text{if } \widehat{\boldsymbol{\beta}}_g \neq \boldsymbol{0}; \\ \lambda \sqrt{p_g} \mathbb{B}[\boldsymbol{0}, 1], & \text{if } \widehat{\boldsymbol{\beta}}_g = \boldsymbol{0}. \end{cases}$$

对偶问题之所以重要，是因为它导出了组级稀疏性筛选规则。也就是说，通过检查最优性条件，已知 $\widehat{\boldsymbol{\nu}}$ 即可识别出 $\boldsymbol{\beta}_g$ 中的零项：

$$\left\| \boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}} \right\|_2 < \lambda \sqrt{p_g} \quad \Longrightarrow \quad \widehat{\boldsymbol{\beta}}_g = \boldsymbol{0}. \tag{9.140}$$

该条件类似于标准 LASSO 对偶可行性条件（3.41），但它是在组级而非单个系数层面施加稀疏性。请注意，该条件是充分但非必要的：即使 $\widehat{\boldsymbol{\beta}}_g = \boldsymbol{0}$，也可能出现 $\left\| \boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}} \right\|_2 = \lambda \sqrt{p_g}$（即处于对偶可行集的边界上）。

采用对偶多面体投影的组 LASSO

在第 3.3.2 节中，我们介绍了对偶多面体投影（DPP）规则。该规则利用对偶空间中的几何性质，在求解标准 LASSO 问题之前，高效地识别并消除非活跃预测变量——即最终解中系数为零的预测变量。现在，我们将这一思想扩展到组 LASSO 问题（9.130）。

考虑对偶形式（9.135）或（9.136）。不难看出，对偶最优解 $\widehat{\boldsymbol{\nu}}$ 是 $\boldsymbol{y}$ 在可行集上的投影。对于每个组 g，约束 $\left\| \boldsymbol{X}_g^\top \boldsymbol{\nu} \right\|_2 \leq \lambda \sqrt{p_g}$ 将 $\boldsymbol{\nu}$ 限制在一个闭且凸的椭球体内。因此，对偶问题（9.135）的可行集是椭球体的交集，因而也是闭且凸的。因此，在惩罚参数 λ 下，对偶问题解 $\widehat{\boldsymbol{\nu}}(\lambda)$ 对组 LASSO 问题同样是非扩张的。

与标准 LASSO 问题类似，令 $\lambda_{\max} \triangleq \max_g \left\| \boldsymbol{X}_g^\top \boldsymbol{y} \right\|_2 / \sqrt{p_g}$，我们可以看到 $\frac{\boldsymbol{y}}{\lambda_{\max}}$ 其本身是可行的，并且 $\lambda_{\max}$ 是使问题（9.130）具有非零解的最大参数。也就是说，如果 $\lambda > \lambda_{\max}$，原始解 $\boldsymbol{\beta}_g(\lambda) = \boldsymbol{0}$，$\forall g = 1, 2, \ldots\ldots, G$。显然，在 $\lambda = \lambda_{\max}$ 下，问题（9.135）的对偶问题解为 $\widehat{\boldsymbol{\nu}}(\lambda_{\max}) = \boldsymbol{y}$。

基于标准 LASSO 的 DPP定理（定理3.12），我们得到以下关于组 LASSO 的结果。

定理 9.27：（组 LASSO 的对偶多胞体投影（DPP）） 假设给定一个解 $\widehat{\boldsymbol{\nu}}(\lambda')$ 作为组 LASSO 对偶问题（9.135）的解，以及一个针对特定 $\boldsymbol{\beta}(\lambda')$ 的原始解 $0 < \lambda' \leq \lambda_{\max}$。令 λ 为一个非负且不等于 λ'的值。对于每个

组 $g \in \{1,2,\ldots,G\}$，如果满足以下条件：

$$\left\|\boldsymbol{X}_g^\top \widehat{\boldsymbol{\nu}}(\lambda')\right\|_2 < \lambda'\sqrt{p_g} - \|\boldsymbol{X}_g\|_F \|\boldsymbol{y}\|_2 \left|1 - \frac{\lambda'}{\lambda}\right|. \tag{9.141}$$

那么，第 g 个组系数向量满足 $\widehat{\boldsymbol{\beta}}_g(\lambda) = \mathbf{0}$。

上述定理表明，如果以下条件成立：

$$\left\|\boldsymbol{X}_g^\top \left(\boldsymbol{y} - \sum_g \boldsymbol{X}_g \widehat{\boldsymbol{\beta}}_g(\lambda')\right)\right\|_2 < \lambda'\sqrt{p_g} - \|\boldsymbol{X}_g\|_F \|\boldsymbol{y}\|_2 \left|1 - \frac{\lambda'}{\lambda}\right|, \tag{9.142a}$$

那么，可以保证组 g 在 λ处处于非活动状态，即 $\widehat{\boldsymbol{\beta}}_g(\lambda) = \mathbf{0}$。特别地，当$\lambda' = \lambda_{\max} \equiv \max_g \left\|\boldsymbol{X}_g^\top \boldsymbol{y}\right\|_2 / \sqrt{p_g}$时，对于所有 g，我们都有 $\widehat{\boldsymbol{\beta}}_g(\lambda_{\max}) = \mathbf{0}$，因此残差等于 $\boldsymbol{y}$。上述不等式变为

$$\left\|\boldsymbol{X}_g^\top \boldsymbol{y}\right\|_2 < \lambda_{\max}\sqrt{p_g} - \|\boldsymbol{X}_g\|_F \|\boldsymbol{y}\|_2 \left(\frac{\lambda_{\max}}{\lambda} - 1\right). \tag{9.142b}$$

▶ **组 LASSO 的序贯 DPP**。 与推论3.14类似，我们可以推导出组 LASSO 的序贯 DPP 规则。考虑正则化参数递减序列上的组 LASSO 问题（9.130）：$\lambda_{\max} = \lambda_0 > \lambda_1 > \ldots > \lambda_m$。对于任意 整数 $0 \le t < m$，我们有 $\widehat{\boldsymbol{\beta}}_g(\lambda_{t+1}) = \mathbf{0}$，如果 $\widehat{\boldsymbol{\beta}}(\lambda_t)$已知且以下条件成立：

$$\left\|\boldsymbol{X}_g^\top \left(\boldsymbol{y} - \sum_{g=1}^{G} \boldsymbol{X}_g \widehat{\boldsymbol{\beta}}_g(\lambda_t)\right)\right\|_2 < \lambda_t\sqrt{p_g} - \|\boldsymbol{X}_g\|_F \|\boldsymbol{y}\|_2 \left(\frac{\lambda_t}{\lambda_{t+1}} - 1\right). \tag{9.143}$$

这一顺序筛选规则使我们能够随着 λ 减小，逐步剔除无关的特征组，从而显著降低路径跟踪算法的计算成本。

9.4.3 稀疏噪声

在标准 LASSO 公式（L_L）或（L_C）（第 126 页）中，假设残差（或噪声）项$\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ 是稠密的。相比之下，以下问题明确考虑了稀疏噪声：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{F(\boldsymbol{\beta}) \triangleq \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_1 + \lambda \|\boldsymbol{\beta}\|_1\right\}, \tag{9.144}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、 $\boldsymbol{y} \in \mathbb{R}^n$ 和 $\lambda > 0$。我们考虑两种优化方法来求解此问题：

▶ **次梯度下降**。 对（9.144）应用次梯度法时，我们使用 ℓ_1-范数的次梯度，形式为 $\partial \|\boldsymbol{\beta}\|_1 \ni \operatorname{sign}(\boldsymbol{\beta})$（参见习题 2.58）。更新规则为：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} - \eta_t\big(\boldsymbol{X}^\top \operatorname{sign}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}) + \lambda \operatorname{sign}(\boldsymbol{\beta})\big).$$

这里， η_t 表示步长。根据问题 9.3 中的指导原则以及例 ℓ_1中 2.72 对 -范数的利普希茨性，可选取

$$\eta_t = \frac{1}{\left\|F'(\boldsymbol{\beta}^{(t)})\right\|_2 \sqrt{t+1}},$$

其中， $F'(\boldsymbol{\beta}^{(t)})$ 是 F 在 $\boldsymbol{\beta}^{(t)}$处的任意次梯度。由于定义域不受约束（$\mathbb{S} = \mathbb{R}^p$），因此该方法对应于一种（无约束）次梯度下降算法。

▶**近端次梯度**。我们将目标函数分解为 $F(\boldsymbol{\beta}) = f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$，其中$f(\boldsymbol{\beta}) \triangleq \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_1$且 $g(\boldsymbol{\beta}) \triangleq \lambda\|\boldsymbol{\beta}\|_1$。近端（次）梯度法的迭代步骤如下：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g}\big(\boldsymbol{\beta}^{(t)} - \eta_t \boldsymbol{X}^\top \operatorname{sign}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\big).$$

由于 $g(\boldsymbol{\beta}) = \lambda\|\boldsymbol{\beta}\|_1$，其近端算子 $\mathcal{P}\mathrm{rox}_{\eta_t g}$ 是众所周知的软阈值算子。具体而言，根据例 2.104， $\mathcal{P}\mathrm{rox}_{\eta_t g} = \mathcal{T}_{\lambda\eta_t}$，因此一般的更新规则为

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{T}_{\lambda\eta_t}\big(\boldsymbol{\beta}^{(t)} - \eta_t \boldsymbol{X}^\top \operatorname{sign}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\big).$$

步长的一种常见选择是 $\eta_t = \frac{1}{\|f'\ \boldsymbol{\beta}^{(t)}()\|_2\sqrt{t+1}}$。

第9章习题

1. 利用定理 9.12 中的近端梯度法收敛性，证明当投影梯度下降法用于在闭凸集 L_b上最小化一个适当、闭、凸且 f -光滑函数 $\mathbb{S}$时的收敛性。如果进一步假设 f 还是 L_a-强凸的，会有什么变化?

2. **凸且光滑函数的梯度下降：** $\mathcal{O}(1/T)$ **(Lu，2025年)**。设 $f: \mathbb{R}^p \to \mathbb{R}$ 是定义在 L_b 上的可微、凸且 $\mathbb{R}^p$-光滑函数。设 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 是梯度下降法（算法 3，不含投影）生成的序列，用于求解 $\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$，步长为恒定步长 $\eta = 1/L_b$。设 $\boldsymbol{\beta}^*$ 是f的任意一个极小化解。证明，对于任意整数 $T > 0$,

$$f(\boldsymbol{\beta}^{(T)}) - f(\boldsymbol{\beta}^*) \leq \frac{2L_b\left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\right\|_2^2}{T-1}.$$

3. 凸利普希茨函数的投影梯度下降： $\mathcal{O}(1/\sqrt{T})$（Lu，2025）设 $f: \mathbb{S} \to \mathbb{R}$是定义在闭、凸且有界集 $\mathbb{S} \subseteq \mathbb{R}^p$ 上的适当、凸、可微且 L-利普希茨函数。设 $\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots,$ $\boldsymbol{\beta}^{(T)}$ 是投影梯度下降法（算法3）生成的 T步序列，用于求解 $\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$，满足约束 $\boldsymbol{\beta} \in \mathbb{S}$，并采用与 $\eta = \frac{R}{L\sqrt{T}}$ 相对应的恒定步长，其中 R 是从初始点 $\boldsymbol{\beta}^{(1)}$ 到某个最优解 $\boldsymbol{\beta}^* \in$最小值点集合$_{\boldsymbol{\beta}\in\mathbb{S}}$ $f(\boldsymbol{\beta})$ 的距离上界。证明对任意 $T \geq 1$,

$$f\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}^{(t)}\right) - f\left(\boldsymbol{\beta}^*\right) \leqslant \frac{RL}{\sqrt{T}}.$$

S说明这如何推出：对满足 $T \geq \frac{R^2L^2}{\epsilon^2}$ 的任意整数 T，可得 t

$$f\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}^{(t)}\right) - f\left(\boldsymbol{\beta}^*\right) \leq \epsilon.$$

此外，证明如果 R 未知，仍可得到类似的保证：

$$f\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}^{(t)}\right) - f\left(\boldsymbol{\beta}^*\right) \leq \epsilon, \quad \text{for } T = \mathcal{O}(\frac{1}{\epsilon^2}) \text{ and } \eta = \frac{1}{\sqrt{T}}.$$

4. 讨论 *Nesterov* 加速方法 算法 15 中所描述的方法。以定理 9.23 为指导，证明当步长固定为 $\mathcal{O}(1/T^2)$ 且动量参数设为 L_b-光滑函数 f时，该方法在最小化凸且$\eta_t = 1/L_b$ 方面可达到收敛速率 $\gamma_t = 2/(t+1)$。

算法15 Nesterov 加速方法

1: $\boldsymbol{\beta}^{(1)} = \boldsymbol{\alpha}^{(1)} = \widetilde{\boldsymbol{\beta}}^{(1)} \in \mathbb{R}^p$, $\gamma_1 = 1$;
2: **for** $t = 1, 2, \ldots$ **do**
3: Choose η_t and γ_{t+1};
4: $\widetilde{\boldsymbol{\beta}}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\frac{\eta_t}{\gamma_t} g}\left(\widetilde{\boldsymbol{\beta}}^{(t)} - \frac{\eta_t}{\gamma_t}\nabla f(\boldsymbol{\alpha}^{(t)})\right)$.
5: $\boldsymbol{\beta}^{(t+1)} \leftarrow (1-\gamma_t)\boldsymbol{\beta}^{(t)} + \gamma_t \widetilde{\boldsymbol{\beta}}^{(t+1)}$.
6: $\boldsymbol{\alpha}^{(t+1)} \leftarrow (1-\gamma_{t+1})\boldsymbol{\beta}^{(t+1)} + \gamma_{t+1}\widetilde{\boldsymbol{\beta}}^{(t+1)}$.
7: **end for**
8: Output $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$;

5. **核范数正则化的近端梯度法**。考虑复合目标函数 $F(\boldsymbol{X}) = f(\boldsymbol{X}) + g(\boldsymbol{X})$，其中 $f(\boldsymbol{X}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2$和$\boldsymbol{Y} \in \mathbb{R}^{n\times p}$是给定矩阵。假设 $g(\boldsymbol{X}) = \lambda \sum_i \sigma_i(\boldsymbol{X})$ 是核范数（即 $\boldsymbol{X}$ 的奇异值之和）。证明近端梯度更新可通过以下步骤得到：

(a) 对输入矩阵 $\boldsymbol{Y}$ 进行奇异值分解，即 $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$，其中 $\boldsymbol{\Sigma} = \mathrm{diag}([\sigma_1(\boldsymbol{Y}) \ \sigma_2(\boldsymbol{Y}) \ \ldots])$ 是由奇异值构成的对角矩阵。

(b) 应用软阈值算子（例 2.104）计算“收缩后”的奇异值

$$\zeta_j \triangleq \mathcal{T}_\lambda(\sigma_j(\boldsymbol{Y})), \quad \text{for } j = 1, 2, \ldots$$

(c) 返回矩阵 $\widehat{\boldsymbol{Y}} = \boldsymbol{U}\mathrm{diag}([\zeta_1, \zeta_2, \ldots])\boldsymbol{V}^\top$。

算法 16 近端点法

要求： </style}一个适当闭凸函数 g；
1: **输入：** 初始化 $\boldsymbol{\beta}^{(1)}$； 2: **对**
于 $t = 1$， 2， ..... **执行**3: 选
择步长 η_t； 4: $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}$
$\mathrm{rox}_{\eta_t g}(\boldsymbol{\beta}^{(t)})$； 5: **结束循环**6:
输出 $\boldsymbol{\beta}_{\text{final}} \leftarrow \boldsymbol{\beta}^{(T)}$;

6. **近端点法**。考虑无约束最小化问题

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} g(\boldsymbol{\beta}), \tag{9.145}$$

其中 g： $\mathbb{R}^p \to$ （$-\infty$， ∞] 是适当、闭且凸的。这是复合问题（9.32）在$f \equiv 0$时的特殊情形。此时，近端梯度更新可化简为

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_t g}(\boldsymbol{\beta}^{(t)}).$$

令 $\eta_t = \eta > 0$ 对所有 t成立，即可得到近端点法，如算法 16所述。假设（9.145）的最优解集非空。令 $\boldsymbol{\beta}^*$ 为任意一个最优解，并令 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 为采用固定步长 $\eta > 0$ 的近端点法所生成的序列。证明

(a) $g(\boldsymbol{\beta}^{(t)}) - g(\boldsymbol{\beta}^*) \leq \frac{\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\|_2^2}{2\eta t}$ 对于 $t > 0$；(b) 序列 $\{\boldsymbol{\beta}^{(t)}\}_{t>0}$ 收敛到 g的某个最优点 $\boldsymbol{\beta}^*$。

提示：应用定理 *9.12* 并结合 $f \equiv 0$，它是一个 *0*-光滑函数。

7. 严格证明，将 Frank–Wolfe（FW）方法应用于约束 LASSO 问题时，其子问题（9.43）具有所述形式。

8. 基于组 LASSO 的 DPP 规则（定理 9.27）和标准 LASSO 的增强型 DPP 规则（推论 3.13），推导组 LASSO 问题的增强型 DPP 筛选规则。

9. **幂迭代法（Lu，2021a，2025年）**。FW 方法为经典的幂迭代法提供了有益的见解，该方法用于计算矩阵的主特征对。考虑以下具体问题

$$\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) = -\frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{X}\boldsymbol{\beta} \quad \text{s.t.} \quad \mathbb{S} = \{\|\boldsymbol{\beta}\|_2 \leq 1\},$$

其中 $\boldsymbol{X}$ 是半正定的。在幂迭代法中，我们迭代更新主特征向量的估计值，并计算相应的瑞利商：

$$\lambda^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t+1)\top} \boldsymbol{X} \boldsymbol{\beta}^{(t+1)}.$$

讨论这一过程与应用于上述约束问题的条件梯度法之间的联系。

10. **罚函数法的失效**。考虑约束优化问题

$$\min \quad -2x^2 + 9y^2 \quad \text{s.t.} \quad x = 1.$$

代入约束 x后，容易验证唯一最优解为 $[1, 0]^\top$。现在考虑二次罚函数

$$f_\sigma(x, y) = -2x^2 + 9y^2 + \frac{\sigma}{2}(x-1)^2,$$

其中 $\sigma > 0$ 是惩罚参数。证明对于任意 $\sigma \leq 4$，该罚函数都无下界。这说明了罚函数法（算法 12）的一个关键局限：如果惩罚参数 σ 过小，不可行点处原目标函数的下降幅度可能超过违反约束所招致的惩罚。因此，惩罚问题可能不存在极小化解，或者得到的解可能远离真正的最优解。因此，初始选择的 σ 必须足够大，才能确保算法具有合理的表现。

11. 带有非最优子问题的罚函数法（Lu，2025）。设 $f(\boldsymbol{\beta})$ 和 $g_i(\boldsymbol{\beta})$，$i \in \mathcal{E}$ 是连续可微函数，并设 $\sigma_t \to +\infty$ 属于算法 12。假设子问题的解 $\boldsymbol{\beta}^{(t+1)}$ 满足 $\left\|\nabla f_{\sigma_t}(\boldsymbol{\beta}^{(t+1)})\right\|_2 \leq \varepsilon_t$，其中 $\varepsilon_t \to 0$。此外，假设对于序列 $\{\boldsymbol{\beta}^{(t)}\}$ 的任意极限点 $\boldsymbol{\beta}^*$，梯度 $\{\nabla g_i(\boldsymbol{\beta}^*)\}_{i\in\mathcal{E}}$ 线性无关。证明 $\boldsymbol{\beta}^*$ 是等式约束优化问题 (9.100) 的一个 KKT 点，并

$$\lim_{t\to\infty} (\sigma_t g_i(\boldsymbol{\beta}^{(t+1)})) = \lambda_i^*, \quad \forall i \in \mathcal{E},$$

其中，λ_i^* 是与约束 $g_i(\boldsymbol{\beta}^*) = 0$对应的拉格朗日乘子。

12. **将交替方向乘子法应用于矩阵分解（Lu，2022年）**。考虑正则化矩阵分解问题

$$\min_{\boldsymbol{Z}} \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{Z}\|_F^2 + r(\boldsymbol{Z})$$

其中，$\boldsymbol{X}\in\mathbb{R}^{n\times p}, \boldsymbol{W}\in\mathbb{R}^{n\times k}$ 是已知的，$r(\cdot)$ 是一个（可能非光滑的）正则化函数。引入辅助变量 $\widetilde{\boldsymbol{Z}}\in\mathbb{R}^{k\times p}$，将光滑最小二乘项与正则化项分离。问题变为

$$\min_{\boldsymbol{Z}} \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{Z}\|_F^2 + r(\widetilde{\boldsymbol{Z}}) \quad \text{s.t.} \quad \boldsymbol{Z}=\widetilde{\boldsymbol{Z}}. \tag{9.146}$$

推导求解 (9.146) 的交替方向乘子法更新步骤。

13. 考虑二次函数 $f(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}^\top\boldsymbol{\beta} + z$, 的无约束最小化，其中 $\boldsymbol{X}\succ\mathbf{0}$ 是一个对称正定矩阵，$\boldsymbol{y}\in\mathbb{R}^p$，且 $z\in\mathbb{R}$ 为常数。

(a). 证明最优解 $\boldsymbol{\beta}^*$ 存在且唯一，并用 $\boldsymbol{X}$ 和 $\boldsymbol{y}$显式表示该解。(b). 写出以恒定步长 $\eta>0$ 最小化该函数的梯度下降更新规则。

提示：使用习题*2.35*。

14. **通过RLS去噪**。假设我们观测到一个含噪信号 $\boldsymbol{y}$，它由 $\boldsymbol{y}=\boldsymbol{\beta}+\boldsymbol{e}$生成，其中 $\boldsymbol{\beta}\in\mathbb{R}^p$ 是真实（未知）信号，$\boldsymbol{e}$ 是噪声向量。朴素重建会求解 $\min\|\boldsymbol{\beta}-\boldsymbol{y}\|_2^2$，其解就是 $\boldsymbol{\beta}=\boldsymbol{y}$。尽管从数学上看，这对于该问题的表述是最优的，但这种估计值在实践中毫无用处，因为它没有对观察进行去噪——只是返回了含噪数据。为了得到更光滑、更可信的估计值，我们加入一个正则化项，对相邻元素之间的突变进行惩罚：$R(\boldsymbol{\beta}) = \sum_{i=1}^{p-1}(\beta_i-\beta_{i+1})^2$. 这会促成分段光滑解。请完成以下任务：

- 将去噪问题重新表述为一个约束最小二乘问题（例如，通过引入辅助变量），并推导其闭式解。指出该模型的实际应用。例如，在金融时间序列分析中，某
- 项资产的损益（P&L）通常会随连续日期平滑变化；因此，通过 $R(\boldsymbol{\beta})$施加光滑性约束，比原始观测值能得到更可信的估计值。

15. 讨论交替方向乘子法与原始-对偶方法之间的关系（见第8章）。特别是，解释如何将交替方向乘子法理解为应用于具有可分离结构和线性约束的问题的一种原始-对偶算法变体。

第10章

稀疏恢复问题算法

目录

10.1 用于 ℓ_0-最小化的算法. 387 10.1.1 贪心算法. 387 10.1.2 正交匹配追踪算法. 390 10.1.3 基于阈值的算法. 394 **10.2 用于 ℓ_1-最小化的算法.** 398 10.2.1 同伦算法. 398 10.2.2 迭代重加权最小二乘法（IRLS）. 402 10.2.3 一阶增广拉格朗日算法. 412 10.2.4 布雷格曼迭代算法. 423 **第10章 问题.** 429



I在第9章中，我们探讨了一大类用于稀疏回归问题的算法——这类任务旨在根据含噪观测估计稀疏系数向量。我们考察了迭代等基础方法硬阈值化（IHT）和近端梯度下降，以及更高级的技术，如交替方向乘子法和坐标下降。这些算法专门用于求解涉及 LASSO、组 LASSO 和稀疏噪声模型等惩罚项的正则化优化问题。它们在高维统计学习中发挥着核心作用，并广泛应用于信号处理、机器学习和计量经济学。

在此基础上，我们现在转向稀疏恢复问题。它们与前述问题密切相关，但在表述和目标上有所不同。第9 章聚焦于回归情境——其中响应变量是连续的，并被建模为预测变量的线性组合——而本章探讨的则是更广泛的挑战：如何从不完整或间接的测量值中重建稀疏信号。

在此，我们深入探讨利用稀疏性从欠定线性系统中恢复信号的专用算法。这是压缩感知、像重建和逆问题中的核心问题。我们首先介绍匹配追踪（MP）和正交匹配追踪（OMP）等贪心方法，它们通过迭代从字典中选取原子来近似目标信号。随后，我们讨论基于阈值的方法——包括硬阈值和软阈值——这些方法与第9章中介绍的近端算子密切相关。

此外，我们还介绍用于 ℓ_0最小化及其 ℓ_1松弛的迭代算法，例如同伦算法、迭代重加权最小二乘（IRLS）和Bregman迭代。这些方法通过连续更新来改进信号估计，并拓展了前文讨论的正则化与近端优化原理，使其专门适用于信号恢复问题。

因此，第9 章着重讨论回归框架中的估计，而本章则聚焦于传感与逆问题中的重建——这两个领域都高度依赖稀疏性假设以及凸优化和非凸优化工具。第5 章和第9 章中提出的理论见解和算法策略，为理解本文介绍的更专门的恢复算法奠定了自然基础。这些章节共同构成了对现代稀疏建模技术在不同应用领域中的连贯论述。

10.1. ℓ_0最小化算法

10.1.1 贪心算法

我们首先介绍几种压缩感知中常用的贪心算法。

在稀疏性约束 k 下，求解欠定线性系统 $\boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1}=\boldsymbol{y}_{n\times 1}$ （其中 $n\ll p$）精确稀疏解的一般方法是：对 $\boldsymbol{X}$的每个可能的 k 列子集求解最小二乘问题 $\boldsymbol{X}_{n\times k}\boldsymbol{\beta}_{k\times 1}=\boldsymbol{y}$ （通常为 $n\gg k$），然后从中选出最佳解。由于超定方程有$\binom{p}{k}$ 种可能的组合，这种蛮力策略即使对于规模中等的 p和 k，计算开销也难以承受。

贪心算法 通过在每一步寻求局部最优选择，而非全局最优解，来规避这种难解性（Foucart 和 Rauhut，2013年）。尽管它们并不保证在所有情况下都能达到全局最优性，但在实践中，贪心方法通常能够得到精确稀疏解或高质量的近似解。

典型的贪心算法采用匹配追踪策略，包括：

(i) 基本匹配追踪（*MP*）方法（*Mallat 和 Zhang，1993 年*）。MP 不再最小化全局代价函数，而是通过将观测信号 $\boldsymbol{\beta}$ 表示为字典矩阵 $\boldsymbol{y}$ 中少数几个选定列（称为原子）的线性组合，迭代构造稀疏近似 $\boldsymbol{X}$。每次迭代中，与当前残差 $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ 相关性最高的原子都会被加入活跃（支撑）集。如果残差范数单调递减，则可以保证 MP 算法收敛。参见算法17。

（ii）正交匹配追踪（*OMP*）（*Pati 等，1993 年*）。标准 MP 不会强制残差与先前选定的原子保持正交，而 OMP 则通过确保在每次迭代中残差都与当前支撑集中的所有原子正交来改进这一点。这是通过对所选索引求解一个最小二乘问题实现的。因此，OMP 保证每一步的系数具有最优性，通常比 MP 需要更少的迭代次数，并且表现出更高的数值稳定性。OMP 的计算复杂度为 $\mathcal{O}(np)$/每次迭代，并且在满足 k-稀疏信号条件 $k \leq n/(2\ln p)$（测量矩阵满足适当条件时），能够精确恢复任意 20。另见算法。

（iii）正则化正交匹配追踪（*ROMP*）（*Needell 和 Vershynin，2009年，2010年*）。ROMP 在 OMP 的基础上引入了正则化步骤，以增强鲁棒性。在每次迭代中，它首先识别出一组与残差具有显著相关性的候选原子，然后依据正则化原则从这些候选原子中选择一个子集（例如，将相关性幅值相近的原子分组）。这种结构化选择有助于确保在噪声条件下实现稳定恢复，并满足类似于凸优化方法的理论保证。

(iv) 分阶段正交匹配追踪（*StOMP*）（*Donoho 等人，2007年*）。StOMP 是一种更快速的 OMP 近似变体，专为大规模问题设计。它不是每次迭代只选择一个原子，而是添加多个相关性超过数据相关阈值的原子。这减少了迭代次数，但代价是近似精度略低。StOMP 的单次迭代复杂度为 $\mathcal{O}(p)$，适用于优先考虑速度而非精确恢复的高维场景。

(v) 压缩采样匹配追踪（*CoSaMP*）（*Needell 和 Tropp，2009年*）。CoSaMP 算法通过引入识别和剪枝步骤，在整个迭代过程中保持固定的稀疏度，从而改进了 ROMP。它在保持较低计算复杂度 $\mathcal{O}(p\ln^2 p)$ 的同时，实现了比 OMP 更高的重构精度。在测量矩阵满足适当条件时，CoSaMP 可以恢复任意 k-稀疏信号，所需测量值为 $k \leq n/\big(2\ln(1+p/k)\big)$。参见算法 18。

(vi) 子空间基追踪（*SP*）（*Dai 和 Milenkovic，2009年*）。算法19 介绍了用于压缩感知信号重建的子空间追踪方法。与 OMP、StOMP 和 ROMP 不同，后者会将新选取的索引永久加入支撑集中，而 SP 始终保持固定的支撑集大小 k（对于一个 k-稀疏信号），并在迭代过程中保持不变。每一步都会将先前的估计值与新的候选项合并，以更新整个支撑集，然后通过最小二乘细化进行剪枝，将其缩减至 k 个元素。这种动态更新方式使 SP 能够纠正先前误判的支撑集元素，通常可提升恢复性能。

算法 17 匹配追踪（MP）要求：字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、测量向量 $\boldsymbol{y} \in \mathbb{R}^n$、稀疏度 k （或容差ε）；确保：近似求解（P_0）：最小值点集合$_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0$，满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$；1：初始化：$\boldsymbol{r}^{(0)} \leftarrow \boldsymbol{y}$, $\mathbb{S} \leftarrow \emptyset$, $\boldsymbol{\beta} \leftarrow \boldsymbol{0}$；2：对于 $t = 0$，1，2，...，执行 3：$i_t \leftarrow$最小值点集合$_{i\in\{1,\ldots,p\}} \left|\left\langle \boldsymbol{r}^{(t)}, \boldsymbol{x}_i \right\rangle\right|$；▷(MP$_1$) 4：$\mathbb{S} \leftarrow \mathbb{S} \cup \{i_t\}$；5：$\beta_{i_t} \leftarrow \langle \boldsymbol{r}^{(t)}, \boldsymbol{x}_{i_t} \rangle$；▷(MP$_3$) 6：$\boldsymbol{r}^{(t+1)} \leftarrow \boldsymbol{r}^{(t)} - \beta_{i_t}\boldsymbol{x}_{i_t}$；▷(MP$_4$) 7：若 $t \geq k$ 或$\left\|\boldsymbol{r}^{(t)}\right\|_2 \leq \varepsilon$则退出；8：结束循环 9：返回 $\boldsymbol{\beta}$；

▷ (MP$_2$)

算法18 压缩采样匹配追踪 (CoSaMP)

Require: D字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，测量向量 $\boldsymbol{y} \in \mathbb{R}^n$，稀疏性 level k；1：初始化：$\boldsymbol{\beta}^{(0)} \in \mathbb{B}_0[k]$ 或简单地设为 $\boldsymbol{\beta}^{(0)} = \boldsymbol{0}$；2：定义：$\mathcal{L}_k(\boldsymbol{\beta}) \triangleq$个 k 的绝对值最大的元素的指标集；3：定义：$\mathcal{P}_k(\boldsymbol{\beta}) \triangleq \boldsymbol{\beta}_{\mathcal{L}_k(\boldsymbol{\beta})}$；4：对于 $t = 0$，1，2，...，执行 5：$\mathbb{S}^{(t+1)} \leftarrow \mathrm{supp}(\boldsymbol{\beta}^{(t)}) \cup \mathcal{L}_{2k}\left(\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)})\right)$；▷(CoSaMP$_1$) 6：$\boldsymbol{\alpha}^{(t+1)} \leftarrow$最小值点集合$_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2, \mathrm{supp}(\boldsymbol{\beta}) \subseteq \mathbb{S}^{(t+1)}\}$；▷(CoSaMP$_2$) 7：$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}_k(\boldsymbol{\alpha}^{(t+1)})$；▷(CoSaMP$_3$) 8：如果在迭代 $t = T$处满足停止准则，则停止；9：结束循环 10：返回最终的 $\boldsymbol{\beta}^{(T)}$。

算法19 子空间追踪（SP）算法

要求： 字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、测量向量 $\boldsymbol{y} \in \mathbb{R}^n$、稀疏度 k；**确保：** 近似求解（P_0）的解：最小值点集合$_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0$ 满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$；1：**初始化：**

(i) $\mathbb{S}^{(0)} \leftarrow \{k$ 中绝对值最大的元素对应的索引；

(ii) $\boldsymbol{r}^{(0)} \leftarrow \boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}^{(0)}}\boldsymbol{X}^+_{\mathbb{S}^{(0)}}\boldsymbol{y}$，其中 $\boldsymbol{A}^+$ 表示 $\boldsymbol{A}$的伪逆；

2: **对于** $t = 1, 2, \cdots$ **执行**

3: $\mathbb{T}^{(t)} \leftarrow \mathbb{S}^{(t-1)} \cup \{k$ 中绝对值最大的元素对应的索引；

4: $\boldsymbol{\alpha}^{(t)} \leftarrow$支撑集的最小值点集合$_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2, (\boldsymbol{\beta}) \subseteq \mathbb{T}^{(t)}\}$，即计算 $\boldsymbol{\alpha}^{(t)} \leftarrow \boldsymbol{X}^+_{\mathbb{T}^{(t)}}\boldsymbol{y}$；

5: $\mathbb{S}^{(t)} \leftarrow \{k \ \boldsymbol{\alpha}\}$中绝对值最大的元素所对应的索引；

6: $\boldsymbol{\beta}^{(t)} \leftarrow$支撑集的最小值点集合$_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2, (\boldsymbol{\beta}) \subseteq \mathbb{S}^{(t)}\}$，即计算 $\boldsymbol{\beta}^{(t)} \leftarrow \boldsymbol{X}^+_{\mathbb{S}^{(t)}}\boldsymbol{y}$；

7: $\boldsymbol{r}^{(t)} \leftarrow \boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}^{(t)}}\boldsymbol{X}^+_{\mathbb{S}^{(t)}}\boldsymbol{y}$；

8: 若 $\|\boldsymbol{r}^{(t)}\|_2$则退出 $> \left\|\boldsymbol{r}^{(t-1)}\right\|_2$，并令 $\mathbb{S}^{(t)} \leftarrow \mathbb{S}^{(t-1)}$；

9: **循环结束**

10: **返回** $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^{(t)}$。

10.1.2 正交匹配追踪算法

如前所述，正交匹配追踪（*OMP*）方法是一种用于稀疏近似的贪心、逐步最小二乘算法。在第 t次迭代中，OMP 向当前支撑估计中添加一个索引 $\mathbb{S}^{(t)}$ ，并将系数向量 $\boldsymbol{\beta}^{(t)}$ 更新为支撑于 $\mathbb{S}^{(t)}$ 且最符合测量值 $\boldsymbol{y}$的最小二乘解。算法20给出了 OMP 的形式化描述，我们将在本小节中对其进行分析。

算法 20 正交匹配追踪（OMP）

要求： 字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$，测量向量 $\boldsymbol{y} \in \mathbb{R}^n$；**确保：** 近似求解（$\mathrm{P}_0$）的解：最小值点集合$_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0$ 满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$;

1: **初始化：** $\mathbb{S}^{(0)} \leftarrow \varnothing$, $\boldsymbol{\beta}^{(0)} \leftarrow \mathbf{0}$;2: **对于** $t = 0,$ 1、2、..... **执行**

3: $\mathbb{S}^{(t+1)} \leftarrow \mathbb{S}^{(t)} \cup \left\{ i_{t+1} = \text{最大值点}_i \left\{ \left| \left(\boldsymbol{X}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}) \right)_i \right|, \ \forall i \right\} \right\}$; ▷ (OMP_1)

4: $\boldsymbol{\beta}^{(t+1)} \leftarrow$ 最小值点集合 $_{\boldsymbol{\beta}} \{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 , \text{支撑集}(\boldsymbol{\beta}) \subseteq \mathbb{S}^{(t+1)} \}$; ▷ (OMP_2)

5: 如果在第 $t = T$ 次迭代时满足停止准则，则停止；6: **结束循环**

7: **返回** 最终 $\boldsymbol{\beta}^{(T)}$。

▶ **停止准则（Tropp 和 Wright，2010年）**。OMP 的一个自然停止条件是精确恢复：$\boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{y}$然而，在实践中，由于测量噪声或数值误差，当残差能量低于容差时停止会更加稳健$\varepsilon > 0$，即

$$\left\| \boldsymbol{r}^{(t)} \right\|_2 \triangleq \left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)} \right\|_2 \leq \varepsilon.$$

另一种方法是，当 $\boldsymbol{X}$ 中没有任何列与残差表现出显著相关性时停止，即

$$\left\| \boldsymbol{X}^\top \boldsymbol{r}^{(t)} \right\|_\infty \leq \varepsilon.$$

如果可以获得真实稀疏度的估计值 k ，则第三种常见的停止规则是在恰好进行 k 次迭代后终止，从而得到一个 k-稀疏估计值 $\boldsymbol{\beta}^{(k)}$。例如，如果 $\boldsymbol{X}$ 是方形正交矩阵，那么采用此停止准则的 OMP 可以从 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$ 中精确恢复任意 k-稀疏向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 。在这种情况下，迭代点 $\boldsymbol{\beta}^{(t)}$ 与这样一个向量重合：该向量保留 $\boldsymbol{\beta}$ 中大小最大的 t 个元素，其余元素均设为零。

▶ **最小二乘**。第（OMP_2）步中的更新求解一个限制在当前支撑集 $\mathbb{S}^{(t+1)}$ 上的最小二乘问题：

$$\min_{\alpha} \left\| \boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}^{(t+1)}} \boldsymbol{\alpha} \right\|_2 , \quad \boldsymbol{\alpha} \in \mathbb{R}^{|\mathbb{S}^{(t+1)}|}. \tag{10.1}$$

解满足正规方程（见（3.16））：

$$\boldsymbol{X}_{\mathbb{S}^{(t+1)}}^\top \boldsymbol{X}_{\mathbb{S}^{(t+1)}} \boldsymbol{\beta}_{\mathbb{S}^{(t+1)}} = \boldsymbol{X}_{\mathbb{S}^{(t+1)}}^\top \boldsymbol{y} \quad \Longrightarrow \quad \boldsymbol{X}_{\mathbb{S}^{(t+1)}}^\top \left(\boldsymbol{X}_{\mathbb{S}^{(t+1)}} \boldsymbol{\beta}_{\mathbb{S}^{(t+1)}} - \boldsymbol{y} \right) = \mathbf{0}. \tag{10.2}$$

这个正交条件——即残差与$\boldsymbol{X}_{\mathbb{S}^{(t+1)}}$ 的列空间正交——解释了正交匹配追踪这一名称。这也表明，向量$\boldsymbol{X}_{\mathbb{S}^{(t+1)}} \boldsymbol{\beta}_{\mathbb{S}^{(t+1)}}$ 是 $\boldsymbol{y}$ 在$\boldsymbol{X}_{\mathbb{S}^{(t+1)}}$的列空间上的正交投影；参见第1.2.3节。

注意，(10.2)还表明，只要尚未满足任何停止准则，在 (OMP i_{t+1}) 中选取的索引$_1$就不属于 $\mathbb{S}^{(t)}$，从而确保每次迭代中支撑集恰好增加一个新元素。

步骤（OMP$_2$）是正交匹配追踪算法中计算开销最大的部分。不过，可以通过求解以下问题来高效实现：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}_{\mathbb{S}^{(t+1)}}\boldsymbol{\beta}_{\mathbb{S}^{(t+1)}}\|_2, \tag{10.3}$$

这实际上是一个标准的最小二乘问题。通过维护 $\boldsymbol{X}_{\mathbb{S}^{(t+1)}}$ 的 QR 分解或 Cholesky分解，并在添加新列时增量更新分解，还可以进一步提高计算效率（Lu，2021b）。

索引 i_{t+1} 的选择遵循贪心策略：在每次迭代中，OMP 都会选择与当前残差 $\boldsymbol{r}^{(t)} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}$相关性最高的原子，即使索引最大化 $\left|(\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}))_i\right|$。下面的命题应用于 $\mathbb{S} = \mathbb{S}^{(t)}$ 和 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}^{(t)}$时，揭示了这一选择为何能显著降低残差的 ℓ_2-范数。

命题10.1：（迭代次数不增加）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个具有 ℓ_2—归一化列的矩阵。给定一个支撑集 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 和一个索引 $i \in \{1, 2, \ldots, p\}$，定义

$$\boldsymbol{\beta}_1 = \arg\min_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2, \operatorname{supp}(\boldsymbol{\beta}) \subseteq \mathbb{S}\},$$
$$\boldsymbol{\beta}_2 = \arg\min_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2, \operatorname{supp}(\boldsymbol{\beta}) \subseteq \mathbb{S} \cup \{i\}\}.$$

则残差范数满足

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_2\|_2^2 \le \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1\|_2^2 - \left(\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1)\right)_i^2.$$

证明 [命题10.1] 由于任意形如 $\boldsymbol{\beta}_1 + \mu\boldsymbol{e}_i$ 且满足 $\mu \in \mathbb{R}$ 的向量都支撑于 $\mathbb{S} \cup \{i\}$，根据 $\boldsymbol{\beta}_2$ 的最优性可得

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_2\|_2^2 \le \min_{\mu\in\mathbb{R}} \|\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\beta}_1 + \mu\boldsymbol{e}_i)\|_2^2.$$

展开右侧可得

$$\|\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\beta}_1 + \mu\boldsymbol{e}_i)\|_2^2 = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1\|_2^2 + \mu^2\|\boldsymbol{X}\boldsymbol{e}_i\|_2^2 - 2\mu\left(\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1)\right)_i.$$

这是关于 μ的二次函数，在 $\mu^* = \left(\boldsymbol{X}^\top\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1\right)\right)_i$处取得最小值。这表明

$$\min_{\mu\in\mathbb{R}} \|\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\beta}_1 + \mu\boldsymbol{e}_i)\|_2^2 = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1\|_2^2 - \left(\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_1)\right)_i^2.$$

结合前面的不等式，证明得证。 ■

更一般地，使用恰好 k 次正交匹配追踪算法迭代来恢复 k-稀疏向量的成功与否，可由以下结果刻画。

定理 10.2：（OMP 下的恢复）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$。每个非零向量$\boldsymbol{\beta}^* \in \mathbb{R}^p$ 支撑于集合 $\mathbb{S}$，且满足 $|\mathbb{S}| = k$，都可以从 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 中在至多k 次迭代内精确恢复，所用的是 OMP 算法（算法20）；当且仅当子矩阵 $\boldsymbol{X}_{\mathbb{S}}$ 具有满列秩，并且

$$\max_{i\in\mathbb{S}} \left|(\boldsymbol{X}^\top \boldsymbol{z})_i\right| > \max_{j\in\mathbb{S}^c} \left|(\boldsymbol{X}^\top \boldsymbol{z})_j\right|, \tag{10.4}$$

对于所有非零的 $\boldsymbol{z} \in \{\boldsymbol{X}\boldsymbol{\alpha}, \text{supp}(\boldsymbol{\alpha}) \subseteq \mathbb{S}\}$。

证明 [定理 10.2] 假设 OMP 能在 $\mathbb{S}$ 次迭代内$k = |\mathbb{S}|$ 恢复所有支撑于 $\mathbb{S}$ 的向量。首先注意，如果两个不同的支撑于 $\boldsymbol{y}$的向量产生相同的测量值，OMP 就无法区分它们——然而假设它能精确恢复这两个向量。因此，$\boldsymbol{X}_{\mathbb{S}}$ 必须是单射，即具有满列秩；否则，两个向量之差将属于零空间，因此零空间非平凡。此外，考虑任意非零的 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$，其支撑集 $\text{supp}(\boldsymbol{\beta}^*) = \mathbb{S}$。由于 OMP 在第一次迭代中只能从 $\mathbb{S}$ 中选择索引，因此此时不能选择任何索引 $j \in \mathbb{S}^c$。这意味着

$$\max_{i\in\mathbb{S}} \left|(\boldsymbol{X}^\top \boldsymbol{y})_i\right| > \max_{j\in\mathbb{S}^c} \left|(\boldsymbol{X}^\top \boldsymbol{y})_j\right|.$$

因为每个非零的 $\boldsymbol{z}$ 属于 $\{\boldsymbol{x}_i \mid i \in \mathbb{S}\}$ 的张成空间的向量，都可以写成 $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\alpha}$，其中 $\boldsymbol{\alpha}$支撑于 $\mathbb{S}$，所以不等式 (10.4) 对所有此类 $\boldsymbol{z}$都成立。

反过来，假设 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩，且条件（10.4）成立。我们用归纳法证明，对于所有 $0 \le t \le k$，支撑估计 $\mathbb{S}^{(t)}$ 满足 $\mathbb{S}^{(t)} \subseteq \mathbb{S}$ 和$\left|\mathbb{S}^{(t)}\right| = t$。由此可得 $\mathbb{S}^{(k)} = \mathbb{S}$，进而由（$\text{OMP}_2$）可得$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$，最后由 $\boldsymbol{X}_{\mathbb{S}}$的满列秩可得$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$。

基例 $t = 0$ 显然成立，因为 $\mathbb{S}^{(0)} = \varnothing$。现在假设该结论对某个 $t < k$成立。定义迭代 t 时的残差为

$$\boldsymbol{z}^{(t)} \triangleq \boldsymbol{y} - \boldsymbol{X}(\mathbb{S}^{(t)})\boldsymbol{\beta}^{(t)}.$$

[1] 由于 $\mathbb{S}^{(t)} \subseteq \mathbb{S}$ 且 $\boldsymbol{\beta}^{(t)}$ 是 $\mathbb{S}^{(t)}$上的最小二乘解，因此 $\boldsymbol{z}^{(t)} \in$ 张成空间$\{\boldsymbol{x}_i \,|i \in \mathbb{S}\}$，即 $\boldsymbol{z}^{(t)} = \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\alpha}$ 其中 $\boldsymbol{\alpha}$ 支撑于 $\mathbb{S}$（因为 $\boldsymbol{\beta}^*$ 支撑于 $\mathbb{S}$，且 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩）。由（10.4），在第 $t+1$步选取的索引为

$$i_{t+1} = \arg\max_i \left|(\boldsymbol{X}^\top \boldsymbol{z}^{(t)})_i\right|,$$

必须属于 $\mathbb{S}$。因此，$\mathbb{S}^{(t+1)} = \mathbb{S}^{(t)} \cup \{i_{t+1}\} \subseteq \mathbb{S}$。

为证明 $\left|\mathbb{S}^{(t+1)}\right| = t+1$，注意到根据正规方程（10.2），残差$\boldsymbol{z}^{(t)}$ 与所有列 $\{\boldsymbol{x}_i \mid i \in \mathbb{S}^{(t)}\}$正交。因此，$(\boldsymbol{X}^\top \boldsymbol{z}^{(t)})_i = 0$ 对所有 $i \in \mathbb{S}^{(t)}$均成立，这意味着 $i_{t+1} \notin \mathbb{S}^{(t)}$。因此，支撑集在每一步都严格增加一个新索引。

由归纳法，$\mathbb{S}^{(k)} = \mathbb{S}$。接着，根据最小二乘更新（$\text{OMP}_2$），有 $\boldsymbol{X}(\mathbb{S})\boldsymbol{\beta}^{(k)} = \boldsymbol{y}$；又由于 $\boldsymbol{X}_{\mathbb{S}}$ 满列秩，因此可得 $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$。证明完毕。 ■

备注 10.3（精确恢复条件）。定理10.2 中的必要且充分条件可以更简洁地表述为精确恢复条件：

$$\left\|\boldsymbol{X}_{\mathbb{S}}^{+}\boldsymbol{X}_{\mathbb{S}^c}\right\|_1 < 1, \tag{10.5}$$

[1]。再次注意，$\boldsymbol{X}(\mathbb{S}) \in \mathbb{R}^{n\times p}$ 以及 $\boldsymbol{X}_{\mathbb{S}} \in \mathbb{R}^{n\times|\mathbb{S}|}$；参见定义 1.1。

其中$\boldsymbol{X}_{\mathbb{S}}^{+}$ 表示 $\boldsymbol{X}_{\mathbb{S}}$的伪逆，且$\|\cdot\|_1 \equiv \|\cdot\|_{1,\,1}$ 表示诱导矩阵范数（定义1.34）。伪逆 $\boldsymbol{X}_{\mathbb{S}}^{+}=(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{X}_{\mathbb{S}}^{\top}$存在等价于 $\boldsymbol{X}_{\mathbb{S}}$ 具有满列秩。此外，条件（10.4）等价于

$$\left\|\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\alpha}\right\|_{\infty} > \left\|\boldsymbol{X}_{\mathbb{S}^c}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\alpha}\right\|_{\infty}, \quad \text{for all } \boldsymbol{\alpha}\in\mathbb{R}^k\setminus\{\boldsymbol{0}\}. \tag{10.6}$$

令 $\boldsymbol{w}\triangleq\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}}\boldsymbol{\alpha}$，则有

$$\|\boldsymbol{w}\|_{\infty} > \left\|\boldsymbol{X}_{\mathbb{S}^c}^{\top}\boldsymbol{X}_{\mathbb{S}}(\boldsymbol{X}_{\mathbb{S}}^{\top}\boldsymbol{X}_{\mathbb{S}})^{-1}\boldsymbol{w}\right\|_{\infty} = \left\|\boldsymbol{X}_{\mathbb{S}^c}^{\top}(\boldsymbol{X}_{\mathbb{S}}^{+})^{\top}\boldsymbol{w}\right\|_{\infty}, \quad \text{for all } \boldsymbol{w}\in\mathbb{R}^k\setminus\{\boldsymbol{0}\}.$$

由该不等式可得 $\left\|\boldsymbol{X}_{\mathbb{S}^c}^{\top}(\boldsymbol{X}_{\mathbb{S}}^{+})^{\top}\right\|_{\infty}<1$ 根据矩阵向量乘积不等式（1.13），这等价于条件 $\left\|\boldsymbol{X}_{\mathbb{S}}^{+}\boldsymbol{X}_{\mathbb{S}^c}^{\top}\right\|_1<1$ 根据诱导范数的对偶性（见（1.14）和（1.15））。

▶ **CoSaMP**。 OMP算法的一个关键局限是，一旦将错误索引选入估计支撑集 $\mathbb{S}^{(t)}$，它就会保留在后续所有支撑集 $\mathbb{S}^{(s)}$ 中，直到 $s\geq t$。因此，如果在最初的 k 次迭代中哪怕选错一个索引， k-稀疏向量也可能无法实现精确恢复——因为OMP从不从支撑集中移除索引。

一种可能的补救方法是执行多于 k 次迭代。然而，这并未解决根本问题：缺少支撑集修正。若能获得稀疏度 k 的估计值，更有效的方法是采用允许向支撑集中添加和移除索引的算法。其中一种方法是压缩采样匹配追踪（*CoSaMP*）算法，本章前文曾介绍过（Needell 和 Tropp，2009年）。

为描述 CoSaMP，我们先介绍两个标准算子。对于任意向量 $\boldsymbol{\beta}\in\mathbb{R}^p$，令

- $\mathcal{L}_k(\boldsymbol{\beta})$ 表示 $\boldsymbol{\beta}\in\mathbb{R}^p$ 中绝对值最大的 k 个元素所对应的指标集。$\mathcal{P}_k(\boldsymbol{\beta})=\boldsymbol{\beta}_{\mathcal{L}_k(\boldsymbol{\beta})}$ 表
- 示相应的 k 项近似，即仅保留这些 k 个元素、并将其余元素置为零所得到的向量。

该映射 $\mathcal{P}_k\triangleq\mathcal{P}_{\mathbb{B}_0[k]}$ 称为阶数为 k [2]的硬阈值算子。给定向量 $\boldsymbol{\beta}\in\mathbb{R}^p$，算子 $\mathcal{P}_k$ 保留其 k 个绝对值最大的元素，并将其他元素置零。当多个元素的大小相同时，该算子的定义可能不唯一。为确保定义良好，我们采用确定性的平局裁决规则（例如按索引的字典序排列），从所有符合条件的候选项中选出 $\mathcal{L}_k(\boldsymbol{\beta})$。根据这些符号，CoSaMP 算法只需通过以下步骤进行迭代更新：

$$\mathbb{S}^{(t+1)} \leftarrow \operatorname{supp}(\boldsymbol{\beta}^{(t)})\cup\mathcal{L}_{2k}\big(\boldsymbol{X}^{\top}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(t)})\big); \tag{10.7a}$$

$$\boldsymbol{\alpha}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}}\{\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2, \operatorname{supp}(\boldsymbol{\beta})\subseteq\mathbb{S}^{(t+1)}\}; \tag{10.7b}$$

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}_k(\boldsymbol{\alpha}^{(t+1)}). \tag{10.7c}$$

在第 4.2.2、 4.2.3 和 5.2 节中，我们说明了当感知矩阵的相干性较小时，稀疏恢复算法的性能会得到提升。现在，我们将在正交匹配追踪算法的背景下证明这一论断。

定理 10.4：（OMP 在 ℓ_1-相干性条件下）设 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 为一个列经过 ℓ_2-归一化的矩阵。如果

$$\mu_1(\boldsymbol{X},k)+\mu_1(\boldsymbol{X},k-1)<1, \tag{10.8}$$

2。注意， $\mathcal{P}_{\mathbb{S}}$ 表示向量在集合 $\mathbb{S}$上的投影；参见定义 2.90。

那么，每个 k-稀疏向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 都可以通过测量向量$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$ 在 OMP算法的至多 k 次迭代内被精确恢复。

证明 [定理 10.4] 根据定理10.2，只需验证对于任意支撑集$\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 满足 $|\mathbb{S}| = k$，子矩阵 $\boldsymbol{X}_{\mathbb{S}}$具有满列秩，并且

$$\max_{i\in\mathbb{S}} |\langle \boldsymbol{z}, \boldsymbol{x}_i\rangle| > \max_{j\in\mathbb{S}^c} |\langle \boldsymbol{z}, \boldsymbol{x}_j\rangle|, \quad \forall\, \boldsymbol{z} \in \{\boldsymbol{X}\boldsymbol{\alpha}, \mathrm{supp}(\boldsymbol{\alpha}) \subseteq \mathbb{S}\} \setminus \{\boldsymbol{0}\}. \tag{10.9}$$

由推论 $\boldsymbol{X}_{\mathbb{S}}$ 4.25可知其具有满列秩。因此，接下来只需证明不等式（10.9）。令 $\boldsymbol{z} \triangleq \sum_{i\in\mathbb{S}} \alpha_i \boldsymbol{x}_i$ 为张成空间 $\boldsymbol{X}_{\mathbb{S}}$中的非零向量，并令 $q \in \mathbb{S}$为满足 $|\alpha_q| = \max_{i\in\mathbb{S}} |\alpha_i| > 0$ 的索引。对于任意 $j \in \mathbb{S}^c$，根据 $\boldsymbol{z}$ 的定义以及ℓ_1-相干性，我们得到

$$|\langle \boldsymbol{z}, \boldsymbol{x}_j\rangle| = \left|\sum_{i\in\mathbb{S}} \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle\right| \le \sum_{i\in\mathbb{S}} |\alpha_i|\, |\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle| \le |\alpha_q|\, \mu_1(\boldsymbol{X}, k), \quad \text{for } j \in \mathbb{S}^c.$$

另一方面，对于索引 $q \in \mathbb{S}$，我们有

$$|\langle \boldsymbol{z}, \boldsymbol{x}_q\rangle| = \left|\sum_{i\in\mathbb{S}} \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x}_q\rangle\right| \ge |\alpha_q|\, |\langle \boldsymbol{x}_q, \boldsymbol{x}_q\rangle| - \sum_{i\in\mathbb{S}, i\neq q} |\alpha_i|\, |\langle \boldsymbol{x}_i, \boldsymbol{x}_q\rangle| \ge |\alpha_q| - |\alpha_q|\, \mu_1(\boldsymbol{X}, k-1).$$

根据假设（10.8），我们有 $1 - \mu_1(\boldsymbol{X},\ k-1) > \mu_1(\boldsymbol{X},\ k)$，这意味着（10.9）成立。证明完毕。 ■

10.1.3 基于阈值的算法

如果稀疏度 k 非常小，OMP 的速度会极快，因为它的运行时间主要取决于迭代次数——算法成功时，迭代次数等于 k。然而，如果 k 相对于环境维数 p来说适中偏大，OMP 的计算开销可能会很高。不过，基于阈值化的算法族中的方法不需要预先知道稀疏度 k。

本节将介绍其他一些采用硬阈值算子 $\mathcal{L}_k$ 或 $\mathcal{P}_k$的算法；其作用见（10.7）。这些方法构成了一个独特的算法族，其出发点是一个简单而有力的直觉：对稀疏向量施加测量矩阵 $\boldsymbol{X}$ 的作用，可以通过对其应用转置 $\boldsymbol{X}^\top$来近似求逆。

这类方法中最简单的是基本阈值法（*BT*）算法。给定测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} \in \mathbb{R}^n$，其中 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是 k-稀疏的，BT 将与相关向量中 $\boldsymbol{\beta}$ 绝对值最大的 k 个元素对应的索引组成的集合，作为 $\boldsymbol{X}^\top\boldsymbol{y}$的支撑集估计值。一旦确定支撑集，算法便在该支撑集上计算最优的最小二乘拟合。该过程总结于算法 21中。注意，BT 无需初始化，因为估计值只需一步即可计算得出。

算法 21 基本阈值法（基本阈值法）

需要： 字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、测量向量 $\boldsymbol{y} \in \mathbb{R}^n$、稀疏度 k；

初始化： $\boldsymbol{\beta}^{(0)} \in \mathbb{B}_0[k]$ 或者简单地说 $\boldsymbol{\beta}^{(0)} = \boldsymbol{0}$；

$\mathbb{S}^* \leftarrow \mathcal{L}_k(\boldsymbol{X}^\top\boldsymbol{y})$； ▷ (BT_1)

$\boldsymbol{\beta}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2$,支撑集$(\boldsymbol{\beta}) \subseteq \mathbb{S}^*\}$； **返回** 最终 $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^*$。 ▷ (BT_2)

通过 BT 算法精确恢复 k-稀疏向量的充要条件，与 OMP 的条件十分相似（参见定理 10.2 (10.4)）。陈述如下。

定理 10.5：（BT 下的恢复）设 $\boldsymbol{\beta} \in \mathbb{R}^p$ 是支撑于集合 $\mathbb{S}$ 的向量，并设 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$。当且仅当 $\boldsymbol{\beta}$ 可由 BT 算法精确恢复时，

$$\min_{i \in \mathbb{S}} \left| (\boldsymbol{X}^\top \boldsymbol{y})_i \right| > \max_{j \in \mathbb{S}^c} \left| (\boldsymbol{X}^\top \boldsymbol{y})_j \right|. \tag{10.10}$$

证明 [定理10.5]该算法精确恢复 $\boldsymbol{\beta}$ 当且仅当估计支撑集 $\mathbb{S}^*$（在步骤（BT_1）中计算得到）等于真实支撑集 $\mathbb{S}$。当且仅当 $\boldsymbol{X}^\top \boldsymbol{y}$ 中由 $\mathbb{S}$ 索引的每个元素的大小都严格大于补集 $\mathbb{S}^c$中由索引的每个元素时，这一条件才成立，这正是条件（10.10）。结论立即得证。 ■

尽管 BT 的恢复条件很简单，但它相当严格。一种更为复杂的方法是迭代硬阈值法（*IHT*）算法——这是一种迭代方法，旨在求解欠定线性系统 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$，其假设是解 $\boldsymbol{\beta}$ 为 k-稀疏的。[3]

与其直接求解 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$，IHT 对最小二乘问题的正规方程进行处理（见（3.16））：

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y}.$$

该系统可以改写为不动点方程 $\boldsymbol{\beta}^{(t+1)} = (\boldsymbol{I} - \boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{\beta}^{(t)} + \boldsymbol{X}^\top \boldsymbol{y}$，经典的不动点迭代由此给出如下更新规则

$$\boldsymbol{\beta}^{(t+1)} \leftarrow (\boldsymbol{I} - \boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{\beta}^{(t)} + \boldsymbol{X}^\top \boldsymbol{y}.$$

然而，由于我们要寻找一个 k-稀疏解，因此在每一步都应用硬阈值算子 $\mathcal{P}_k$，仅保留大小最大的 k 个元素，并将其余元素置零。由此得到 IHT 更新规则。或者，也可以将其视为标准 IHT 算法（算法2）应用于损失函数 $f(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$ 和约束集$\mathbb{S} = \mathbb{B}_0[k]$。所得过程见算法22。

算法 22 迭代硬阈值法（IHT）

输入： 字典 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，测量向量 $\boldsymbol{y} \in \mathbb{R}^n$，稀疏度 k

;1: **初始化：** $\boldsymbol{\beta}^{(0)} \in \mathbb{B}_0[k]$ 或简单地使用 $\boldsymbol{\beta}^{(0)} = \boldsymbol{0}$; 2: **对于** $t = 0$, 1、2、..... **执行**3: $\boldsymbol{\beta}^{(t+1)} \leftarrow \mathcal{P}_k(\boldsymbol{\beta}^{(t)} - \eta_t \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}))$;▷ $\eta_t = 1$ 是常用的选择。4: 如果在第 $t = T$ 次迭代时满足停止准则，则停止；5: **结束循环**6: **返回** 最终的 $\boldsymbol{\beta}^{(T)}$。

IHT 的一项关键优势在于，它避免了计算投影到由 $\boldsymbol{X}$ 的列子集张成的子空间上的正交投影（即最小二乘拟合），因为这类计算的开销可能很大。不过，如果我们愿意承担这类投影的计算成本——正如 OMP 等贪心方法所做的那样——自然会产生这样一个问题：一旦选定支撑集，为什么不在该支撑集上计算最佳拟合？这一思路引出了硬阈值化

3。其在 ℓ_0-约束问题中的作用见算法 2 。

追踪（*HTP*）算法，它将 IHT 的支撑集选择与最小二乘细化步骤结合起来。

算法 23 硬阈值追踪（HTP）

要求： 字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、测量向量 $\boldsymbol{y} \in \mathbb{R}^n$、稀疏度 k；**初始化：** $\boldsymbol{\beta}^{(0)} \in \mathbb{B}_0[k]$ 或简单地使用 $\boldsymbol{\beta}^{(0)} = \mathbf{0}$；**对于** $t = 0,\ \mathbf{1},\ 2,\ \ldots\ldots$ **执行**：$\mathbb{S}^{(t+1)} \leftarrow \mathcal{L}_k\big(\boldsymbol{\beta}^{(t)} + \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)})\big)$； ▷ (HTP_1)

$\boldsymbol{\beta}^{(t+1)} \leftarrow$ 最小值点集合$_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2$,支撑集$(\boldsymbol{\beta}) \subseteq \mathbb{S}^{(t+1)}\}$；如果在第 $t = T$ 次迭代时满足停止准则，则停止； **结束循环**：**返回** 最终的 $\boldsymbol{\beta}^{(T)}$。 ▷ (HTP_2)

为了分析这些算法的稀疏恢复性质，我们遵循 Foucart 和 Rauhut （2013年），并继续使用 ℓ_1-相干性条件。RIP 下的恢复保证将在问题 10.1∼10.2 中进一步讨论。

定理 10.6：（在 ℓ_1-相干性下 BT 的精确恢复） 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个列经 ℓ_2归一化的矩阵，$\boldsymbol{\beta}^* \in \mathbb{R}^p$ 是一个支撑于集合 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 上的 k-稀疏向量，且满足 $|\mathbb{S}| = k$。若

$$\mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, k-1) < \frac{\min_{i\in\mathbb{S}} |\beta_i^*|}{\max_{i\in\mathbb{S}} |\beta_i^*|}, \tag{10.11}$$

则 BT 算法能够从测量向量 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 中精确恢复向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 。

证明 [定理 10.6]根据定理10.5，只需证明对每个 $i \in \mathbb{S}$和每个$j \in \mathbb{S}^c$,

$$|\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_i\rangle| > |\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_j\rangle|\,. \tag{10.12}$$

由于 $\boldsymbol{\beta}^*$ 支撑于 $\mathbb{S}$，因此对任意 $i \in \mathbb{S}$,

$$|\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_i\rangle| = \left|\sum_{q\in\mathbb{S}} \beta_q^* \langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle\right| \geq |\beta_i^*| - \sum_{q\in\mathbb{S}, q\neq i} \left|\beta_q^*\right| |\langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle| \geq |\beta_i^*| - \mu_1(\boldsymbol{X}, k-1) \max_{q\in\mathbb{S}} \left|\beta_q^*\right|.$$

对任意 $j \in \mathbb{S}^c$，我们得到:

$$|\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_j\rangle| = |\langle \boldsymbol{X}_{\mathbb{S}}\boldsymbol{\beta}_{\mathbb{S}}^*, \boldsymbol{x}_j\rangle| = \left|\sum_{q\in\mathbb{S}} \beta_q^* \langle \boldsymbol{x}_q, \boldsymbol{x}_j\rangle\right| \leq \sum_{q\in\mathbb{S}} \left|\beta_q^*\right| |\langle \boldsymbol{x}_q, \boldsymbol{x}_j\rangle| \leq \mu_1(\boldsymbol{X}, k) \max_{q\in\mathbb{S}} \left|\beta_q^*\right|.$$

结合这两个上界并利用假设（10.11），可得:

$$|\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_i\rangle| - |\langle \boldsymbol{X}\boldsymbol{\beta}^*, \boldsymbol{x}_j\rangle| \geq \min_{q\in\mathbb{S}} \left|\beta_q^*\right| - (\mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, k-1)) \max_{i\in\mathbb{S}} \left|\beta_q^*\right| > 0.$$

这就证明了（10.12），并完成了证明。 ■

现在我们来讨论更复杂的硬阈值追踪算法。与正交匹配追踪算法一样，我们证明，在类似于（ k ）的条件下，至多经过 k次迭代，就足以恢复任意10.8稀疏向量。

定理10.7：（在 ℓ_1-相干性条件下硬阈值追踪的精确恢复）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个矩阵，其 ℓ_2 列均已归一化。若

$$2\mu_1(\boldsymbol{X}, k) + \mu_1(\boldsymbol{X}, k-1) < 1,$$

那么，在 HTP 算法的大多数 k 次迭代中，每个支撑于 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 且满足 $|\mathbb{S}| = k$ 的 k-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 都能从测量向量 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 中精确恢复。

证明 [定理 10.7] 重新排列索引，使得 i_1，i_2，.. .，i_p 满足

$$|\beta_{i_1}| \geq |\beta_{i_2}| \geq \ldots \geq |\beta_{i_k}| > \left|\beta_{i_{k+1}}\right| = \ldots = \left|\beta_{i_p}\right| = 0.$$

我们用归纳法证明，对于满足 $0 \leq t \leq k-1$ 的每次迭代 t，集合 $\{i_1, i_2, \ldots, i_{t+1}\}$ 都包含在由式 (HTP$_1$) 定义的 $\mathbb{S}^{(t+1)}$ 中，其中测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 是绝对值最大的元素

$$\boldsymbol{z}^{(t+1)} \triangleq \boldsymbol{\beta}^{(t)} + \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}) = \boldsymbol{\beta}^{(t)} + \boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)}). \tag{10.13}$$

这将意味着 $\mathbb{S}^{(k)} = \mathbb{S} = \operatorname{supp}(\boldsymbol{\beta}^*)$，而随后执行的步骤 (HTP$_2$)——它在 $\mathbb{S}^{(k)}$ 上计算最小二乘解——会得到 $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$。为此，只需证明关键不等式：

$$\min_{1\leq q\leq t+1} \left|z_{i_q}^{(t+1)}\right| > \max_{j\in\mathbb{S}^c} \left|z_j^{(t+1)}\right|. \tag{10.14}$$

也就是说，(HTP$_1$) 选取最显著的系数。我们注意到，对于每个 $i \in \{1, 2, \ldots, p\}$,

$$z_i^{(t+1)} = \beta_i^{(t)} + \sum_{q=1}^{p}(\beta_q^* - \beta_q^{(t)}) \langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle = \beta_i^* + \sum_{q\neq i}(\beta_q^* - \beta_q^{(t)}) \langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle \tag{10.15}$$

$$\Longrightarrow \left|z_i^{(t+1)} - \beta_i^*\right| \leq \sum_{q\in\mathbb{S}^{(t)}, q\neq i} \left|\beta_q^* - \beta_q^{(t)}\right| |\langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle| + \sum_{q\in\mathbb{S}\setminus\mathbb{S}^{(t)}, q\neq i} \left|\beta_q^*\right| |\langle \boldsymbol{x}_q, \boldsymbol{x}_i\rangle|. \tag{10.16}$$

该不等式成立，因为 $\beta_q^{(t)} = 0$ 对于 $q \in \mathbb{S} \setminus \mathbb{S}^{(t)}$ 根据 (HTP$_2$)。由于 $t \leq k-1$ 且 $|\mathbb{S}| = k$, (10.16) 的右侧可以拆分为

$$\leq \begin{cases} \sum_{q\in\mathbb{S}^{(t)}, q\neq i} \left|\beta_q^* - \beta_q^{(t)}\right| \mu_1(\boldsymbol{X}, k-1) + \sum_{q\in\mathbb{S}\setminus\mathbb{S}^{(t)}} \left|\beta_q^*\right| \mu_1(\boldsymbol{X}, k-1), & i \in \mathbb{S}^{(t)}; \\ \sum_{q\in\mathbb{S}^{(t)}} \left|\beta_q^* - \beta_q^{(t)}\right| \mu_1(\boldsymbol{X}, k) + \sum_{q\in\mathbb{S}\setminus\mathbb{S}^{(t)}, q\neq i} \left|\beta_q^*\right| \mu_1(\boldsymbol{X}, k-1), & i \notin \mathbb{S}^{(t)}. \end{cases} \tag{10.17}$$

现在我们用归纳法证明 (10.14)。初始时，对于迭代 $t = 0$，$\boldsymbol{\beta}^{(0)}_{\mathbb{S}^{(0)}} = \boldsymbol{0}$ 且 $\mathbb{S}^{(0)} = \varnothing$。将 $\left\|(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(0)})_{\mathbb{S}^{(0)}}\right\|_\infty = 0$ 代入 (10.13) 和 (10.16)，并利用以下事实：$2\mu_1(\boldsymbol{X}, k) < 1$ 由假设可得，我们有

$$\left|z_{i_1}^{(1)}\right| \geq (1 - \mu_1(\boldsymbol{X}, k)) \left\|\boldsymbol{\beta}^*\right\|_\infty > \mu_1(\boldsymbol{X}, k) \left\|\boldsymbol{\beta}^*\right\|_\infty \geq \left|z_j^{(1)}\right|, \quad \text{for all } j \in \mathbb{S}^c.$$

对于 10.14)，其基例适用于 $t = 0$。假设 10.14) 对 $t-1$在 $t \geq 1$的条件下成立。这意味着 $\{i_1, i_2, \ldots, i_t\} \subseteq \mathbb{S}^{(t)}$。根据步骤（HTP$_2$），残差 $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}$ 与空间 $\{\boldsymbol{X}\boldsymbol{z} \,|\, \operatorname{supp}(\boldsymbol{z}) \subseteq \mathbb{S}^{(t)}\}$正交，即 $\langle \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}, \boldsymbol{X}\boldsymbol{z}\rangle = \langle \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}) \boldsymbol{z}\rangle = 0$ 对于支撑于 $\boldsymbol{z} \in \mathbb{R}^p$ 上的向量都成立 $\mathbb{S}^{(t)}$。代入 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 可得

$$\left(\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)})\right)_{\mathbb{S}^{(t)}} = \boldsymbol{0}.$$

因此，对于所有 $i \in \mathbb{S}^{(t)}$，根据（10.13），有 $z_i^{(t+1)} = \beta_i$。利用（10.16）或（10.17），得到

$$\left|\beta_i^{(t)} - \beta_i^*\right| \leq \mu_1(\boldsymbol{X}, k-1) \left\|(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)})_{\mathbb{S}^{(t)}}\right\|_\infty + \mu_1(\boldsymbol{X}, k-1) \left\|\boldsymbol{\beta}^*_{\mathbb{S}\setminus\mathbb{S}^{(t)}}\right\|_\infty, \quad \forall i \in \mathbb{S}^{(t)}.$$

对 $i \in \mathbb{S}^{(t)}$ 取最大值并重新整理各项：

$$\left\|(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(t)})_{\mathbb{S}^{(t)}}\right\|_\infty \leq \frac{\mu_1(\boldsymbol{X}, k-1)}{1 - \mu_1(\boldsymbol{X}, k-1)} \left\|\boldsymbol{\beta}^*_{\mathbb{S}\setminus\mathbb{S}^{(t)}}\right\|_\infty. \tag{10.18}$$

将其与式（10.13）和（10.16）结合，可得，对于 $1 \leq q \leq t+1$ 和 $j \in \mathbb{S}^c$，

$$\left|z_{i_q}^{(t+1)}\right| \geq \left(1 - \frac{\mu_1(\boldsymbol{X}, k)}{1 - \mu_1(\boldsymbol{X}, k-1)}\right) \left|\beta^*_{i_{t+1}}\right|,$$
$$\left|z_j^{(t+1)}\right| \leq \frac{\mu_1(\boldsymbol{X}, k)}{1 - \mu_1(\boldsymbol{X}, k-1)} \left|\beta^*_{i_{t+1}}\right|.$$

根据假设，$\mu_1(\boldsymbol{X}, k)/(1 - \mu_1(\boldsymbol{X}, k-1)) < 1/2$。因此，对于迭代 10.14)，式（ t成立，从而完成归纳法。 ■

10.2. ℓ_1 最小化算法

现在我们转向 ℓ_1-最小化问题，它们是压缩感知中信号恢复的基石。正如模型（P_1'）（第 129 页）所述，在实数情形下，ℓ_1-最小化可以重新表述为一个线性规划（LP）。

基于内点法或经典单纯形法的标准线性规划求解器现成可用，且应用广泛（Boyd and Vandenberghe，2004年；Beck，2014年）。然而，在实践中，直接求解线性规划形式（P_1'）往往颇具挑战。在典型的压缩感知应用中，测量矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 规模很大（其中 $p \approx 10^6$），且是稠密矩阵，由此得到的线性规划也常常病态。此外，内点法需要对涉及 $\boldsymbol{X}$ 的矩阵进行分解，例如 $\boldsymbol{X}^\top \boldsymbol{X}$，或对规模与 $n + p$成正比的增广系统进行分解；在这种规模下，这些操作的计算成本高得难以承受。因此，对于大规模压缩感知问题，通用内点求解器并不适用于求解（P_1'）。

尽管这类通用求解器可靠且易于使用，但它们并未针对 ℓ_1-最小化问题的特定结构进行设计。事实上，专门为 ℓ_1-最小化设计的算法，速度可能显著快于通用线性规划方法。本节将介绍并分析几种此类专用算法。

10.2.1 同伦算法

我们从<style id='1'>同伦方法开始。它在思想上类似于 OMP，但与 OMP 不同的是，它能够保证恢复 ℓ_1-最小化问题的精确解。在拓扑学中，同伦描述了两个对象之间的连续变形。同伦算法采用了这一思想：它从一个简单且易于计算的解出发，通过一系列迭代更新逐步对其进行变形，直到得到所需的（通常更为复杂的）解。因此，任何同伦方法的一个关键方面，就是选择一个合适的初始解；该解既要简单，又要通过一条性质良好的路径与目标解相连。

具体而言，同伦方法求解 ℓ_1-最小化问题（P_1）（第 129 页）：

$$(\mathrm{P}_1) \qquad \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}, \tag{10.19a}$$

其中 $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 且 $\boldsymbol{y}\in\mathbb{R}^n$。为了解决这个问题，对于 $\lambda>0$，考虑称为 LASSO 的拉格朗日形式（参见（L_L），第 126页）：

$$F_\lambda(\boldsymbol{\beta})=\frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{\beta}\|_1, \quad \boldsymbol{\beta}\in\mathbb{R}^p, \tag{10.19b}$$

并令 $\boldsymbol{\beta}_\lambda$ 表示 F_λ的一个极小化解。当 $\lambda=\lambda_{\max}$ 充分大时，有 $\boldsymbol{\beta}_{\lambda_{\max}}=\mathbf{0}$；参见第 3.3节。此外，当 $\lambda\to 0$，$\boldsymbol{\beta}_\lambda$ 收敛到 $\boldsymbol{\beta}^*$，其中 $\boldsymbol{\beta}^*$ 是（P_1）的一个极小化解。以下定理给出了精确表述。

定理 10.8：（从 LASSO 恢复 ℓ_1-最小化）假设系统 $\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y}$ 相容。若（P_1）的极小化解 $\boldsymbol{\beta}^*$ 唯一，则

$$\lim_{\lambda\to 0}\boldsymbol{\beta}_\lambda=\boldsymbol{\beta}^*.$$

更一般地，若（10.19a）的极小化解不唯一，则集合 $\{\boldsymbol{\beta}_\lambda\}_{\lambda>0}$ 有界，并且 $\boldsymbol{\beta}_\lambda$ 当 $\lambda\to 0$ 时的每个聚点都是（P_1）的一个极小化解。

证明 [定理 10。8] 为证明有界性，注意到 $F_\lambda(\boldsymbol{\beta}_\lambda)\le F_\lambda(\mathbf{0})=\|\boldsymbol{y}\|_2^2/2$。令 $\{\lambda_i\}_{i\in\mathbb{N}}\subset(0,\infty)$ 为一个单调递减且趋于 0 的序列，为简便起见，记 $\boldsymbol{\beta}^i=\boldsymbol{\beta}_{\lambda_i}$。令 $\boldsymbol{\beta}^*$ 为 ℓ_1-最小化问题 (P_1) 的任意一个极小化解。于是，由于$\boldsymbol{X}\boldsymbol{\beta}^*=\boldsymbol{y}$ 根据 (P_1)，可得

$$\left\|\boldsymbol{\beta}^i\right\|_1\le\frac{1}{\lambda_i}F_{\lambda_i}(\boldsymbol{\beta}^i)\le\frac{1}{\lambda_i}F_{\lambda_i}(\boldsymbol{\beta}^*)=\|\boldsymbol{\beta}^*\|_1. \tag{10.20}$$

因此，所有 $\boldsymbol{\beta}^i$ 都属于紧集

$$\mathbb{S}\triangleq\{\boldsymbol{\beta}\in\mathbb{R}^p,\|\boldsymbol{\beta}\|_1\le\|\boldsymbol{\beta}^*\|_1\}.$$

定义 $F\colon\ \mathbb{S}\to\mathbb{R}$ 由 $F(\boldsymbol{\beta})=\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2/2$ 给出。由于 $\mathbb{S}$ 是紧集，F 连续，并根据魏尔斯特拉斯定理（见定理 2.56）在 $\mathbb{S}$ 上取得最小值。此外，记 $\sim F_{\lambda_i}$ 为函数 F_{λ_i} 在 $\mathbb{S}$ 上的限制函数，我们注意到，序列 F_{λ_i} 在 i 上单调递减，并逐点收敛到 F。根据问题 10.3，序列 $\{\boldsymbol{\beta}^i\}_i$ 的任意聚点都是 F 的极小化解 $\widehat{\boldsymbol{\beta}}$，因此 $\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{y}$。根据 $\boldsymbol{\beta}^*$ 对 ℓ_1-最小化问题的最优性，有 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1\ge\|\boldsymbol{\beta}^*\|_1$。而根据集合 $\mathbb{S}$ 的定义以及 (10.20)，我们还有 $\left\|\widehat{\boldsymbol{\beta}}\right\|_1=\lim_{i\to\infty}\left\|\boldsymbol{\beta}^i\right\|_1\le\|\boldsymbol{\beta}^*\|_1$。因此，

$$\left\|\widehat{\boldsymbol{\beta}}\right\|_1=\|\boldsymbol{\beta}^*\|_1,$$

这意味着 $\{\boldsymbol{\beta}^i\}_i$ 的任意聚点都是（P_1）的极小化解。如果极小化解唯一，那么 $\{\boldsymbol{\beta}^i\}_i$ 的每个子序列都收敛于 $\boldsymbol{\beta}^*$，因此整个序列收敛于 $\boldsymbol{\beta}^*$。 ■

注释 10.9. 请注意，我们使用一个逆命题，通过算法13中的惩罚函数方法来计算拉格朗日 LASSO 解。

因此， $\boldsymbol{\beta}_{\lambda_{\max}} = \mathbf{0}$ 是求解 ℓ_1-最小化问题（P_1）时同伦算法的自然起点。同伦方法的核心思想是追踪解路径 $\lambda \to \boldsymbol{\beta}_\lambda$ ，从 $\boldsymbol{\beta}_{\lambda_{\max}} = \mathbf{0}$ 一直到 $\boldsymbol{\beta}^*$。如下文所示，该解路径是分段线性的，因此只需计算断点——即各线性段的端点（Osborne 等人，2000年；Efron 等人，2004年；Foucart 和 Rauhut，2013年）。

根据无约束凸问题的最优性条件（定理 2.62），可利用次微分（定义2.57）刻画（10.19b）的极小化解。 F_λ 的次微分为

$$\partial F_\lambda(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \lambda \partial \|\boldsymbol{\beta}\|_1 ,$$

其中， ℓ_1范数的次微分（习题 2.58）为

$$\partial \|\boldsymbol{\beta}\|_1 = \{\boldsymbol{u} \in \mathbb{R}^p \mid u_i \in \partial |\beta_i| , i \in \{1, 2, \dots, p\}\}, \quad \text{where } \partial |x| = \begin{cases} \{\operatorname{sign}(x)\}, & \text{if } x \neq 0, \\ [-1, 1], & \text{if } x = 0. \end{cases}$$

向量 $\boldsymbol{\beta}$ 是 F_λ 的极小化解，当且仅当 $\mathbf{0} \in \partial F_\lambda(\boldsymbol{\beta})$，参见定理 2.62。利用上面的表达式，该条件等价于对每个坐标$i \in \{1, 2, \dots, p\}$满足以下两个关系式：

$$(\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}))_i = -\lambda \operatorname{sign}(\beta_i), \qquad \text{if } \beta_i \neq 0, \tag{10.21a}$$

$$\left|(\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}))_i\right| \leq \lambda, \qquad \text{if } \beta_i = 0, \quad \text{for } i \in \{1, 2, \dots, p\}. \tag{10.21b}$$

同伦方法从初始解 $\boldsymbol{\beta}_\lambda = \boldsymbol{\beta}^{(0)} = \mathbf{0}$开始。根据条件（10.21b），相应的正则化参数选为 $\lambda = \lambda^{(0)} = \left\|\boldsymbol{X}^\top \boldsymbol{y}\right\|_\infty$；另见第 3.3 (3.43)。

在后续步骤 $t = 1$、2、.....中，算法减小 λ，计算相应的极小化解 $\boldsymbol{\beta}^{(1)}$， $\boldsymbol{\beta}^{(2)}$、.....，并在每一步 t维护活跃（支撑）集 $\mathbb{T}_t$。将步骤 t 的残差向量记为

$$\boldsymbol{r}^{(t)} \triangleq \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t-1)} - \boldsymbol{y}).$$

同伦方法的步骤 1。对于步骤 1，由于 $\boldsymbol{\beta}^{(0)} = \mathbf{0}$，定义

$$i^{(1)} \triangleq \operatorname*{arg\,max}_{i \in \{1,2,\dots,p\}} \left|(\boldsymbol{X}^\top \boldsymbol{y})_i\right| = \operatorname*{arg\,max}_{i \in \{1,2,\dots,p\}} \left|r_i^{(1)}\right| . \tag{10.22}$$

我们始终假设这个最大值在唯一的索引 i处取得。（尽管理论上最大值可能在多个索引处同时取得，但这种情况在实践中极为罕见，且会带来一些技术上的复杂性，因此此处不予讨论。）

将初始活跃集设为 $\mathbb{T}_1 \triangleq \{i^{(1)}\}$。现在定义描述解路径初始线段的方向向量 $\boldsymbol{d}^{(1)} \in \mathbb{R}^p$:

$$d^{(1)}_{i^{(1)}} \triangleq \|\boldsymbol{x}_{i^{(1)}}\|_2^{-2} \operatorname{sign}((\boldsymbol{X}^\top \boldsymbol{y})_{i^{(1)}}) \qquad \text{and} \qquad d_i^{(1)} = 0, \quad \text{for } i \neq i^{(1)}.$$

解路径的第一个线性段为

$$\boldsymbol{\beta} = \boldsymbol{\beta}(\eta) = \boldsymbol{\beta}^{(0)} + \eta \boldsymbol{d}^{(1)}, \quad \eta \in [0, \eta^{(1)}], \tag{10.23}$$

其中上界 $\eta^{(1)}$ 稍后将确定。根据 $\boldsymbol{d}^{(1)}$ 的定义可直接验证，条件（10.21a）对所有 $\boldsymbol{\beta}=\boldsymbol{\beta}$（$\eta$）均成立，只要我们更新参数$\lambda=\lambda$（$\eta$） $=\lambda^{(0)}-\eta$， $\eta\in[0,\lambda^{(0)}]$，即：

$$\left(\boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})\right)_{i^{(1)}}=-\lambda\mathrm{sign}(\beta_{i^{(1)}}).$$

下一个断点出现在满足以下条件的最大值处： $\eta=\eta^{(1)}>0$ 条件（10.21b）对所有非活跃坐标 $i\notin\mathbb{T}_1$均成立。令（a）$_+=\max\{a,0\}$。由此确定（ $\eta^{(1)}$ 10.23）中的上界：

$$\eta^{(1)}=\min_{i\notin\{i^{(1)}\}}\left\{\left(\frac{\lambda^{(0)}+r_i^{(1)}}{1-(\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{d}^{(1)})_i}\right)_+,\left(\frac{\lambda^{(0)}-r_i^{(1)}}{1+(\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{d}^{(1)})_i}\right)_+\right\}. \tag{10.24}$$

那么，下一个解为 $\boldsymbol{\beta}^{(1)}=\boldsymbol{\beta}(\eta^{(1)})=\eta^{(1)}\boldsymbol{d}^{(1)}$，它使 F_λ 在 $\lambda=\lambda^{(1)}\triangleq\lambda^{(0)}-\eta^{(1)}$时达到最小值。根据构造，该参数 $\lambda^{(1)}$ 满足 $\lambda^{(1)}=\left\|\boldsymbol{r}^{(2)}\right\|_\infty$。令 $i^{(2)}$ 表示（10.24）中达到最小值的索引（这里同样假设最小值只在一个索引处取得），并将活跃集更新为 $\mathbb{T}_2\triangleq\{i^{(1)},i^{(2)}\}$。步骤 1 至此完成。

同伦方法的 t 步骤。在步骤 t，我们从步骤 $t-1$获得活跃集 $\mathbb{T}_t$。通过求解线性方程组，可以确定同伦路径的新方向 $\boldsymbol{d}^{(t)}$

$$\boldsymbol{X}_{\mathbb{T}_t}^{\top}\boldsymbol{X}_{\mathbb{T}_t}\boldsymbol{d}_{\mathbb{T}_t}^{(t)}=\mathrm{sign}(\boldsymbol{r}_{\mathbb{T}_t}^{(t)}). \tag{10.25}$$

其中， $\boldsymbol{X}_{\mathbb{T}_t}$ 表示由索引属于 $\mathbb{T}_t$ 的列组成的 $\boldsymbol{X}$ 的子矩阵， $\boldsymbol{r}_{\mathbb{T}_t}^{(t)}$ 是残差对应的子向量。这是一个规模为 $|\mathbb{T}_t|\times|\mathbb{T}_t|$的线性方程组，其中$|\mathbb{T}_t|\le t$。对于活跃集之外的索引，对所有 $i\notin\mathbb{T}_t$，我们令 $d_i^{(t)}=0$。解路径的下一个线性段为

$$\boldsymbol{\beta}(\eta)=\boldsymbol{\beta}^{(t-1)}+\eta\boldsymbol{d}^{(t)},\quad \eta\in[0,\eta^{(t)}]. \tag{10.26}$$

使条件（10.21b）对所有非活跃坐标均保持成立的最大步长为 η

$$\eta_+^{(t)}=\min_{i\notin\mathbb{T}_t}\left\{\left(\frac{\lambda^{(t-1)}+r_i^{(t)}}{1-(\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{d}^{(t)})_i}\right)_+,\left(\frac{\lambda^{(t-1)}-r_i^{(t)}}{1+(\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{d}^{(t)})_i}\right)_+\right\}. \tag{10.27}$$

使条件（10.21a）对所有活跃坐标继续成立（即 $\boldsymbol{\beta}$ 中没有分量提前变号或经过零点）的最大步长为

$$\eta_-^{(t)}=\min_{i\in\mathbb{T}_t}\left\{\left(-\beta_i^{(t-1)}/d_i^{(t)}\right)_+\right\}. \tag{10.28}$$

下一个断点出现在 $\eta^{(t)}=\min\{\eta_+^{(t)},\eta_-^{(t)}\}$，更新后的解为 $\boldsymbol{\beta}^{(t)}=\boldsymbol{\beta}(\eta^{(t)})$:

- 如果 $\eta^{(t)}=\eta_+^{(t)}$，则使目标函数最小的索引 $i^{(t)}\notin\mathbb{T}_t$（来自（10.27））会被加入活跃集： $\mathbb{T}_{j+1}=\mathbb{T}_t\cup\{i^{(t)}\}$。如果 $\eta^{(t)}=\eta_-^{(t)}$，则对应的索引 $i_-^{(t)}\in\mathbb{T}_t$ 会被移除：
- $\mathbb{T}_{j+1}=\mathbb{T}_t\setminus\{i_-^{(t)}\}$。

这与 LARS 的 LASSO 修正相同；参见注释9.18。最后，按如下方式更新正则化参数 $\lambda^{(t)}=\lambda^{(t-1)}-\eta^{(t)}$。根据构造，这满足 $\lambda^{(t)}=\left\|\boldsymbol{r}^{(t+1)}\right\|_\infty$。

算法在满足以下条件时终止： $\lambda^{(t)}=\left\|\boldsymbol{r}^{(t+1)}\right\|_\infty=0$，也就是说，此时残差为零。在这一点上，当前迭代点 $\boldsymbol{\beta}^{(t)}$ 就是原始 $\boldsymbol{\beta}^*$ -最小化问题（10.19a）的精确解，算法输出 $\boldsymbol{\beta}^*=\boldsymbol{\beta}^{(t)}$。

定理 10.10：（同伦的精确恢复）假设 $\boldsymbol{\beta}^*$ 是 ℓ_1-最小化问题（P_1）的唯一极小化解。如果在算法的每一步中，（10.27）和（10.28）中的极小值都在唯一索引 i处取得，那么上述同伦算法将以输出 $\boldsymbol{\beta}^*$终止。

定理 [的证明10.10] 按照算法的描述，只需证明该算法在有限步后终止。为此，注意到在每次迭代中，符号模式 $\operatorname{sign}(\boldsymbol{\beta}_{\lambda^{(a)}})$ 都不相同 a。事实上，假设对于两个参数 $\lambda^{(a)}$ 和 $\lambda^{(b)}$，满足 $b > a$，有 $\operatorname{sign}(\boldsymbol{\beta}_{\lambda^{(a)}})_i = \operatorname{sign}(\boldsymbol{\beta}_{\lambda^{(b)}})_i$。令 $0 \neq \sigma_i \triangleq \operatorname{sign}(\boldsymbol{\beta}_{\lambda^{(a)}})_i = \operatorname{sign}(\boldsymbol{\beta}_{\lambda^{(b)}})_i$ 对支撑集中的所有索引 i 成立。那么，由条件 (10.21a)，可得

$$(\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\beta}_{\lambda^{(a)}} - \boldsymbol{\beta}_{\lambda^{(b)}}))_i = (\lambda^{(a)} - \lambda^{(b)})\sigma_i,$$

这意味着 $\boldsymbol{\beta}_{\lambda^{(b)}}$ 位于同伦路径上从 $\boldsymbol{\beta}_{\lambda^{(a)}}$开始的同一线性段上。然而，根据构造，每个 $\boldsymbol{\beta}_{\lambda^{(a)}}$ 都对应一个断点——也就是线性片段的端点，而不是任何线性段的内部点。

由于 $\mathbb{R}^p$ 中可能的符号模式数量有限（考虑正、负和零元素，最多为 3^p），算法不可能生成无限个不同的断点。因此，它必定会在有限步后终止。

终止时，我们有 $\lambda^{(t)} = 0$，这意味着 $\boldsymbol{\beta}_{\lambda^{(t)}} = \boldsymbol{\beta}^*$ 根据定理 10.8。因此，算法输出唯一极小化解 $\boldsymbol{\beta}^*$。 ■

如果算法在某次迭代 t 时提前终止，它显然会返回 $F_\lambda = F_{\lambda^{(t)}}$ 的极小化解，即参数为 $\lambda = \lambda^{(t)}$ 的拉格朗日 LASSO 问题的解。此外，可以在算法执行过程中应用简单的停止准则，以求解以下约束

$$\min \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \leq \epsilon \tag{10.29}$$

$$\text{or} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_1 \leq \Sigma. \tag{10.30}$$

第一种形式对应于含噪测量下的 ℓ_1-最小化，而第二种则是约束 LASSO。

10.2.2 迭代重加权最小二乘法（IRLS）

现在我们介绍迭代重加权最小二乘法（*IRLS*）方法。尽管 IRLS 仅作为 ℓ_1-最小化的替代方法，但其形式直接受此启发；在某些条件下，IRLS 确实能够恢复 ℓ_1-极小化解。假设满足稳定零空间性质——这一条件已知与通过 ℓ_1-最小化实现精确且稳定的稀疏恢复等价——我们将证明 IRLS 具有与 ℓ_1-最小化相同的误差保证。然而，一般而言，IRLS 的输出可能不同于真正的 ℓ_1-极小化解。

该算法源自初等恒等式 $|x| = \frac{|x|^2}{|x|}$，适用于 $x \neq 0$。这一观察表明，ℓ_1极小化可以改写为加权 ℓ_2极小化。具体而言，设 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，且满足 $n \leq p$，并假设 $\boldsymbol{\beta}^*$ 是以下问题的极小化解：

$$(\mathrm{P}_1) \quad \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{10.31}$$

若对所有 $i \in \{1,2,\dots,p\}$ 都有 $\beta_i^* \neq 0$，则 $\boldsymbol{\beta}^*$ 也是以下加权最小二乘问题的极小化解：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{p} \frac{\beta_i^2}{|\beta_i^*|} \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{10.32}$$

这种重新表述的优势在于，最小化光滑二次函数 $|x|^2$ 在计算上比最小化非光滑绝对值 $|x|$ 更容易处理。然而，这种方法有两个主要缺点：首先，真正的极小化解 $\boldsymbol{\beta}^*$ 事先并不已知；其次，我们通常无法指望 $\beta_i^* \neq 0$ 对所有 i都成立，因为我们通常寻求 k-稀疏解。事实上，根据问题 3.2，任何唯一的 ℓ_1-极小化解至多是 n-稀疏的。

尽管存在这些困难，上述恒等式仍启发了一种迭代策略：每一步中，我们求解一个加权的 ℓ_2-最小化问题，并根据当前解更新下一次迭代所用的权重。具体而言，我们可以定义 $F(\boldsymbol{\beta},\ \boldsymbol{w}) = \sum_i^p \beta_i^2 w_i$，并迭代地更新 $\boldsymbol{\beta}$ 和 $\boldsymbol{w}$：

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} F(\boldsymbol{\beta}, \boldsymbol{w}^{(t)});\\ \boldsymbol{w}^{(t+1)} &\leftarrow \arg\min_{\boldsymbol{w}\in\mathbb{R}^p} F(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{w}).\end{aligned}$$

出于稳定性和收敛方面的考虑，在算法的公式表述和分析中，采用下面这个更稳健的泛函：

$$\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{w}, \varepsilon) = \frac{1}{2}\left[\sum_{i=1}^{p} \beta_i^2 w_i + \sum_{i=1}^{p} (\varepsilon^2 w_i + w_i^{-1})\right], \tag{10.33}$$

其中，$\boldsymbol{\beta} \in \mathbb{R}^p$、$\varepsilon \geq 0$ 和 $\boldsymbol{w} \in \mathbb{R}^p$ 是严格正的权重向量（对所有 $i \in \{1,2,\dots,p\}$，$w_i > 0$）。下面的算法会用到当前迭代点 $\boldsymbol{\beta}^{(t)} \in \mathbb{R}^p$ 的非递增重排，记为

$$[\boldsymbol{\beta}^{(t)}]^{\downarrow} \in \mathbb{R}^p, \tag{10.34}$$

它满足 $[\boldsymbol{\beta}^{(t)}]_1^{\downarrow} \geq [\boldsymbol{\beta}^{(t)}]_2^{\downarrow} \geq \ldots \geq [\boldsymbol{\beta}^{(t)}]_p^{\downarrow}$。完整的过程见算法 24，其中需要指定一个稀疏度参数 k。

算法 24 迭代重加权最小二乘法（IRLS）（Daubechies 等人，2010年）要求：字典 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、测量向量 $\boldsymbol{y} \in \mathbb{R}^n$、$\eta > 0$、$k \in \{1,2,\dots,p\}$；1：初始化：$\boldsymbol{w}^{(0)} = [1,1,\dots,1]^\top \in \mathbb{R}^p$、$\varepsilon_0 = 1$；2：对于 $t = 1$、2、.....，执行3：$\boldsymbol{\beta}^{(t+1)} \leftarrow$最小值点集合$_{\boldsymbol{\beta}\in\mathbb{R}^p}\ \mathcal{F}(\boldsymbol{\beta}$、$\boldsymbol{w}^{(t)}$、$\varepsilon_t)$，满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$；4：$\varepsilon_{t+1} \leftarrow$ 最小值$\{\varepsilon_t, \eta[\boldsymbol{\beta}^{(t+1)}]_{k+1}^{\downarrow}\}$； ▷ (IRLS$_1$)
5：$\boldsymbol{w}^{(t+1)} \leftarrow$最小值点集合$_{\boldsymbol{w}>0}\ \mathcal{F}(\boldsymbol{\beta}^{(t+1)}$、$\boldsymbol{w}$、$\varepsilon_{t+1})$；6：若满足停止准则或 $\varepsilon_{t+1} = 0$，则停止；7：结束循环8：返回最终的 $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^{(t)}$。 ▷ (IRLS$_2$)
▷ (IRLS$_3$)
▷ (IRLS$_4$)

▶ **更新** $\boldsymbol{\beta}^{(t+1)}$。由于 $\boldsymbol{w}^{(t)}$ 和 ε_t 在最小化问题（IRLS$_1$）中是固定的，定义式（10.33）中的第二个求和项 $\mathcal{F}$ 是常数。因此，$\boldsymbol{\beta}^{(t+1)}$ 将一个加权最小二乘问题 最小化为如下形式（见（L2, $\boldsymbol{w}$），第 137页）：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_{2,\boldsymbol{w}^{(t)}} = \left(\sum_{i=1}^{p} \beta_i^2 w_i^{(t)}\right)^{1/2} \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{10.35a}$$

根据（3.19），极小化解 $\boldsymbol{\beta}^{(t+1)}$ 具有显式表达式

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{D}_t^{-1/2}(\boldsymbol{X}\boldsymbol{D}_t^{-1/2})^{+}\boldsymbol{y}, \tag{10.35b}$$

其中，$(\boldsymbol{X}\boldsymbol{D}_t^{-1/2})^{+}$ 表示 $\boldsymbol{X}\boldsymbol{D}_t^{-1/2}$ 的伪逆，而 $\boldsymbol{D}_t = \operatorname{diag}(\boldsymbol{w}^{(t)})$是对角矩阵，其元素由权重向量 $\boldsymbol{w}^{(t)}$给出。如果 $\boldsymbol{X}$ 满行秩——这在压缩感知中通常如此——那么，由于权重 $w^{(t)} > 0$具有正性，矩阵 $\boldsymbol{X}\boldsymbol{D}_t$ 也满行秩。在这种情况下，伪逆可简化，解可写为

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{D}_t^{-1}\boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{D}_t^{-1}\boldsymbol{X}^{\top})^{-1}\boldsymbol{y}. \tag{10.35c}$$

或者，定义 $\boldsymbol{\xi} \triangleq (\boldsymbol{X}\boldsymbol{D}_t^{-1}\boldsymbol{X}^{\top})^{-1}\boldsymbol{y}$，则解也可表示为

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{D}_t^{-1}\boldsymbol{X}^{\top}\boldsymbol{\xi}, \quad \text{where} \quad \boldsymbol{X}\boldsymbol{D}_t^{-1}\boldsymbol{X}^{\top}\boldsymbol{\xi} = \boldsymbol{y}. \tag{10.35d}$$

因此，计算 $\boldsymbol{\beta}^{(t+1)}$ 可归结为求解上面的线性方程组以得到 $\boldsymbol{\xi}$。关于最小二乘和加权最小二乘问题的更多背景知识，请参阅Lu(2021b)。

▶ **对 $\boldsymbol{w}^{(t+1)}$的更新。**(IRLS_3) 中的 $\boldsymbol{w}^{(t+1)}$ 最小化在各个坐标之间相互独立，可以显式地完成：

$$w_i^{(t+1)} = \frac{1}{\sqrt{(\beta_i^{(t+1)})^2 + \varepsilon_{t+1}^2}}, \quad i \in \{1, 2, \ldots, p\}. \tag{10.36}$$

该公式还说明了平滑参数 ε_t的作用。在朴素表述$w_i^{(t+1)} = \left|\beta_i^{(t+1)}\right|^{-1}$中，当 $\beta_i^{(t+1)}$ 趋近于零时，权重会变得无界。引入 ε_{t+1} 后，权重 $\boldsymbol{w}^{(t+1)}$得到了正则化；具体而言，它确保 $\left\|\boldsymbol{w}^{(t+1)}\right\|_{\infty} \leq \varepsilon_{t+1}^{-2}$.。然而，为了在极限情况下恢复 ℓ_1-极小化解，我们要求当 $t \to \infty$时 $\varepsilon_t \to 0$。更新规则（IRLS_2）恰好保证了这一行为：$\varepsilon_{t+1} = \min\{\varepsilon_t, \eta[\boldsymbol{\beta}^{(t+1)}]_{k+1}^{\downarrow}\} \leq \varepsilon_t$，因此 ε_t 是非递增的。此外，如果迭代点 $\boldsymbol{\beta}^{(t+1)}$ 收敛到一个 k-稀疏向量，则 $[\boldsymbol{\beta}^{(t+1)}]_{k+1}^{\downarrow} \to 0$，从而 $\varepsilon_{t+1} \to 0$。因此，稀疏度参数 k 直接控制解的目标稀疏度。

注 10.11. 对于两者，也可以采用其他更新规则 ε_{t+1} 以及权重 $\boldsymbol{w}^{(t+1)}$。例如，可以考虑截断更新 $w_j^{(t+1)} = \min\{\left|\beta_j^{(t+1)}\right|^{-1}, \varepsilon_{t+1}^{-1}\}$. 这一变体的收敛结果已经得到确立；详情请参见 Fornasier 等人 (2011年)。

为陈述 IRLS 算法的主要收敛结果，对于任意$\varepsilon > 0$，我们引入辅助泛函

$$F_{\varepsilon}(\boldsymbol{\beta}) \triangleq \sum_{i=1}^{p} \sqrt{\beta_i^2 + \varepsilon^2}. \tag{10.37}$$

证明　收敛速率依赖于 IRLS迭代序列的以下关键性质 .

引理 10.12：（IRLS算法的性质） 令 $\{\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}\}_{t\in\mathbb{N}}$ 为由 IRLS 算法 24生成的迭代点序列。则，对于 $t \in \mathbb{N}$,

$$\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) = \sum_{i=1}^{p} \sqrt{(\beta_i^{(t)})^2 + \varepsilon_t^2} = F_{\varepsilon_t}(\boldsymbol{\beta}^{(t)}), \tag{10.38a}$$

and

$$\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) \le \mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t-1)}, \varepsilon_t) \le \mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t-1)}, \varepsilon_{t-1}) \tag{10.38b}$$

$$\le \mathcal{F}(\boldsymbol{\beta}^{(t-1)}, \boldsymbol{w}^{(t-1)}, \varepsilon_{t-1}). \tag{10.38c}$$

此外，序列 $\{\boldsymbol{\beta}^{(t)}\}$有界，

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \mathcal{F}(\boldsymbol{\beta}^{(1)}, \boldsymbol{w}^{(0)}, \varepsilon_0) \triangleq B, \quad t \in \mathbb{N}, \tag{10.38d}$$

并且权重 $\boldsymbol{w}^{(t)}$ 一致地远离零:

$$w_i^{(t)} \ge B^{-1}, \quad i \in \{1, 2, \ldots, p\},\ t \in \mathbb{N}. \tag{10.38e}$$

IRLS 算法的迭代点24 也满足

$$\sum_{t=1}^{\infty} \left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right\|_2^2 \le 2B^2, \tag{10.38f}$$

其中， B 是（10.38d）中定义的常数。因此，$\lim_{t\to\infty}(\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}) = \mathbf{0}$。

证明 [引理10.12] 方程（10.38a）直接由（ $\mathcal{F}$ ）中10.33的定义以及权重的显式更新（10.36）得出。

(10.38b) 中的第一个不等式成立，是因为 $\boldsymbol{w}^{(t)}$ 根据构造（IRLS$_3$）使 $\mathcal{F}(\boldsymbol{\beta}^{(t)}$、 、$\varepsilon_t)$ 最小化。第二个不等式由 $\varepsilon_{t+1} \le \varepsilon_t$ 的单调性以及 $\boldsymbol{w}^{(t-1)}$的正性得到。最后一个不等式（10.38c）则是 $\boldsymbol{\beta}^{(t)}$在加权最小二乘子问题（IRLS$_1$）中最优性的结果。

由（10.38a）以及（10.38b）中的单调性，我们得到

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \sum_{i=1}^{p} \sqrt{(\beta_i^{(t)})^2 + \varepsilon_t^2} = F_{\varepsilon_t}(\boldsymbol{\beta}^{(t)}) = \mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) \le \mathcal{F}(\boldsymbol{\beta}^{(1)}, \boldsymbol{w}^{(0)}, \varepsilon_0) = B,$$

这就证明了（ 10 38 10 36 $w_i^{(t)\ -1} = \sqrt{\beta_i^{(t)\ 2} + \varepsilon_t^2} \le B \quad i \in$。
$\{1, 2, \ldots, p\}$ 10 38
d）。利用（。），我们进一步得到（） （）、从而证明了（。e）。

为证明（10.38f），注意到（10.38b）中的单调性意味着

$$2\left(\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) - \mathcal{F}(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{w}^{(t+1)}, \varepsilon_{t+1})\right) \ge 2\left(\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) - \mathcal{F}(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{w}^{(t)}, \varepsilon_t)\right)$$

$$= \sum_{i=1}^{p} \left((\beta_i^{(t)})^2 - (\beta_i^{(t+1)})^2\right) w_i^{(t)} = \left\langle \boldsymbol{\beta}^{(t)} + \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}}$$

$$\overset{\dagger}{=} \left\langle \boldsymbol{\beta}^{(t)} + \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}} - 2\left\langle \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}}$$

$$= \left\langle \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}} = \left\| \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\|_{\boldsymbol{w}^{(t)}}^2 = \sum_{i=1}^{p} w_i^{(t)} \left| \beta_i^{(t)} - \beta_i^{(t+1)} \right|^2$$

$$\geq B^{-1} \left\| \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\|_2^2,$$

其中 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{\boldsymbol{w}} = \sum_{i=1}^{p} a_i b_i w_i$ 表示加权内积。由于 $\boldsymbol{\beta}^{(t)}$ 和$\boldsymbol{\beta}^{(t+1)}$ 都满足 $\boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^{(t+1)}$，根据（IRLS$_1$），可得 $\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \in \mathcal{N}(\boldsymbol{X})$。等式（†由加权最小二乘解的正交性性质（定理 3.8）可得：$\langle \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \rangle_{\boldsymbol{w}^{(t)}} = 0$。最后一个不等式由（10.38e）可得。对 t 求和并利用这些不等式的望远镜相消性质，可得

$$\sum_{t=1}^{\infty} \left\| \boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)} \right\|_2^2 \leq 2B \sum_{t=1}^{\infty} \left(\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) - \mathcal{F}(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{w}^{(t+1)}, \varepsilon_{t+1}) \right)$$

$$\leq 2B\mathcal{F}(\boldsymbol{\beta}^{(1)}, \boldsymbol{w}^{(1)}, \varepsilon_1) \leq 2B^2,$$

证明完毕。 ■

注意，(10.38c)表明目标函数值 $\mathcal{F}$ 在每次迭代时都会下降，而(10.38f)保证相邻迭代点渐近地趋于驻点。

对于约束优化问题，我们还需要刻画 $\boldsymbol{\beta}^{(\varepsilon)}$ 平滑目标函数$F_\varepsilon(\boldsymbol{\beta}) = \sum_{i=1}^{p} \sqrt{\beta_i^2 + \varepsilon^2}$的极小化解，该函数在(10.37)中定义为

$$\boldsymbol{\beta}^{(\varepsilon)} \triangleq \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} F_\varepsilon(\boldsymbol{\beta}) \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}. \tag{10.39}$$

极小化解 $\boldsymbol{\beta}^{(\varepsilon)}$ 由于**严格凸性而**唯一；参见问题 F_ε。10.4。

引理 10.13：（$F_\varepsilon(\boldsymbol{\beta})$**的最小化的最优性条件**）设 $\varepsilon > 0$，并令 $\boldsymbol{\beta} \in \mathbb{R}^p$ 满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$。则 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\varepsilon)}$ 当且仅当 $\langle \boldsymbol{\beta}, \boldsymbol{n} \rangle_{\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon}} = 0$ 对所有 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$均成立，其中权重向量 $\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon} \in \mathbb{R}^p$ 按分量定义为 $(\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon})_i = (\beta_i^2 + \varepsilon^2)^{-1/2}$，$i = 1, 2, \ldots\ldots, p$。

证明 [引理 10.13] 设 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\varepsilon)}$ 是（10.39）的极小化解。令 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$ 为 $\boldsymbol{X}$的任意零空间向量，并定义函数

$$g(\nu) = F_\varepsilon(\boldsymbol{\beta} + \nu\boldsymbol{n}) - F_\varepsilon(\boldsymbol{\beta}), \quad \nu \in \mathbb{R}.$$

由于 $\boldsymbol{X}(\boldsymbol{\beta} + \nu\boldsymbol{n}) = \boldsymbol{y}$ 对所有 $\nu \in \mathbb{R}$均成立，扰动点 $\boldsymbol{\beta} + \nu\boldsymbol{n}$ 仍然可行。根据 $\boldsymbol{\beta}$的最优性，我们有 $g(\nu) \geq 0$ 对所有 $\nu \in \mathbb{R}$均成立，并且显然 $g(0) = 0$。因此，g 在 $\nu = 0$处取得其最小值，所以 $g'(0) = 0$。直接计算可得

$$g'(0) = \sum_{i=1}^{p} \frac{\beta_i n_i}{\sqrt{\beta_i^2 + \varepsilon^2}} = \langle \boldsymbol{\beta}, \boldsymbol{n} \rangle_{\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon}},$$

这意味着 $\langle \boldsymbol{\beta}, \boldsymbol{n} \rangle_{\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon}} = 0$ 对所有 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$ 均成立。

反之，假设 $\boldsymbol{\beta}$满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$和 $\langle \boldsymbol{\beta}, \boldsymbol{n} \rangle_{\boldsymbol{w}_{\boldsymbol{\beta}, \varepsilon}} = 0$ 对所有 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$均成立。考虑标量函数 $f(u) \triangleq \sqrt{u^2 + \varepsilon^2}$，它在 $\mathbb{R}$上是凸函数。根据梯度

不等式（定理 2.45），对于任意 u，$u_0 \in \mathbb{R}$，我们有 $\sqrt{u^2+\varepsilon^2} \geq \sqrt{u_0^2+\varepsilon^2} + \frac{u_0(u-u_0)}{\sqrt{u_0^2+\varepsilon^2}}$。因此，对于任意 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$，我们有

$$F_\varepsilon(\boldsymbol{\beta}+\boldsymbol{n}) \geq F_\varepsilon(\boldsymbol{\beta}) + \sum_{i=1}^{p} \frac{\beta_i n_i}{\sqrt{\beta_i^2+\varepsilon^2}} = F_\varepsilon(\boldsymbol{\beta}) + \langle \boldsymbol{\beta}, \boldsymbol{n} \rangle_{\boldsymbol{w}_{\boldsymbol{\beta},\varepsilon}} = F_\varepsilon(\boldsymbol{\beta}).$$

由于 $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{X})$ 是任意的，因此可知 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\varepsilon)}$ 是（10.39）的极小化解。 ■

现在，我们可以证明 IRLS 算法的收敛性了。下面的恢复定理基于稳定 NSP 的概念（定义 4.6），其内容与 ℓ_1-最小化的类似结果十分相似。回顾一下，在第 6.6 节和7.5节中，我们直接证明了在适当条件下，高斯随机矩阵和次高斯随机矩阵以高概率满足 RIP。此外，由定理 4.56可知，RIP 蕴含 NSP。因此，本书讨论的各种矩阵系综在适当假设下也满足稳定 NSP。

定理 10.14：（稳定 NSP 下 IRLS 的收敛性）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 是一个满足稳定 NSP（定义 4.6）的设计矩阵，其阶为 k，参数为$\rho < 1$。设 $\boldsymbol{\beta}^* \in \mathbb{R}^p$，并考虑测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$。考虑稀疏度参数为 k、平滑参数更新规则采用$\eta = 1/p$的 IRLS 算法 24 [a]。那么，迭代序列 $\{\boldsymbol{\beta}^{(t)}\}_t$ 收敛到一个向量 $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^p$，当 $t \to \infty$时，其非递增重排 $[\widehat{\boldsymbol{\beta}}]^{\downarrow}$ 满足 $[\widehat{\boldsymbol{\beta}}]_k^{\downarrow} = p$ $\lim_{t\to\infty} \varepsilon_t$。此外，以下结论成立：

(i) 如果 $\lim_{t\to\infty} \varepsilon_t = 0$，那么 $\boldsymbol{\beta}$ 是 k-稀疏的，并且求解 ℓ_1-最小化问题 (P_1)。此外，如果真实向量 $\boldsymbol{\beta}^*$ 本身也是 k-稀疏的，那么 $\boldsymbol{\beta}^* = \boldsymbol{\beta}$。更一般地，我们有误差界

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{1-\rho} \sigma_k(\boldsymbol{\beta}^*)_1, \tag{10.40}$$

其中 $\sigma_k(\boldsymbol{\beta}^*)_1$ 表示 k项 ℓ_1对 $\boldsymbol{\beta}^*$的最佳 $\boldsymbol{\beta}^*$-逼近误差；参见 (4.4)。

(ii) 如果 $\varepsilon \triangleq \lim_{t\to\infty} \varepsilon_t > 0$，那么 $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\varepsilon)}$，即，该极限与在 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 条件下平滑泛函 F_ε 的极小化解一致。在这种情况下，如果 ρ 满足更紧的上界 $\rho < 1 - \frac{2}{k+2}$，那么对于任意 $\widetilde{k} < k - \frac{2\rho}{1-\rho}$，都有

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq C\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1 \quad \text{with} \quad C \triangleq \frac{2(1+\rho)}{1-\rho} \cdot \frac{k-\widetilde{k}+3/2}{k-\widetilde{k}-\frac{2\rho}{1-\rho}}. \tag{10.41}$$

特别地，如果 $\boldsymbol{\beta}^*$ 对某个 $\widetilde{k} < k - \frac{2\rho}{1-\rho}$ 是 $\widetilde{k}$-稀疏的，那么 $\lim_{t\to\infty} \varepsilon_t = 0$，因此情形 $\varepsilon > 0$ 不可能发生。

[a] 选择 $\eta = 1/p$ 是为了便于理论分析，并确保收敛结果有效。然而在实践中，较大的取值（例如 $\eta = 1$）往往能带来更快的收敛速度和更好的经验性能。目前尚无针对这类替代选择的严格收敛保证。

证明 [定理 10。14] 注意，由于 $0 \leq \varepsilon_{t+1} \leq \varepsilon_t$，序列 $\{\varepsilon_t\}_{t\in\mathbb{N}}$ 总是收敛。将其极限记为 $\varepsilon \triangleq \lim_{t\to\infty} \varepsilon_t$。

(i) 当 $\lim_{t\to\infty}\varepsilon_t = 0$ 时。情形 (a)。首先，假设对于某个 $t_0 \in \mathbb{N}$， $\varepsilon_{t_0} = 0$ 。那么根据定义，算法终止（不执行下一步 (IRLS$_3$)；参见 (IRLS$_4$)），并且我们可以为 $t \geq t_0$ 设置 $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t_0)}$，使得 $\lim_{t\to\infty}\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t_0)} = \widehat{\boldsymbol{\beta}}$。根据 (IRLS$_2$) 中 ε_t 的定义可知，非递增重排 $[\boldsymbol{\beta}^{(t_0)}]^{\downarrow}_{k+1} = 0$，这意味着 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(t_0)}$ 是 k-稀疏的。由于 $\boldsymbol{X}$ 满足 k 阶稳定 NSP（因此也满足 k 阶标准 NSP；参见定义 4.1 和 4.6），推论 5.3 保证 $\boldsymbol{\beta}$ 是 ℓ_1-最小化问题 (P$_1$) 的唯一解。因此，若真实向量 $\boldsymbol{\beta}^*$（其中 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$）是 k-稀疏的，则 $\boldsymbol{\beta}$ 也是唯一的 ℓ_1-极小化解，即 $\boldsymbol{\beta}^* = \boldsymbol{\beta}$。对于满足 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 的一般 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ （不一定是 k-稀疏的），定理 5.7 给出误差界

$$\left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{2(1+\rho)}{1-\rho}\sigma_k(\boldsymbol{\beta}^*)_1.$$

情形 (b)。 现在假设对所有 $t \in \mathbb{N}$ 都有 $\varepsilon_t > 0$，但仍有 $\lim_{t\to\infty}\varepsilon_t = 0$。那么存在一个递增的索引序列 $\{t_i\}$，使得对所有 $i \in \mathbb{N}$ 都有 $\varepsilon_{t_i} < \varepsilon_{t_{i-1}}$。根据 (IRLS$_2$) 中 ε_t 的定义以及 $\eta = 1/p$ 的选取，这意味着 $\boldsymbol{\beta}^{(t_i}$ 的非递增重排 $)$ 满足

$$[\boldsymbol{\beta}^{(t_i)}]^{\downarrow}_{k+1} < p\varepsilon_{t_{i-1}}, \quad i \in \mathbb{N}.$$

由性质 (10.38d，序列 $\{\boldsymbol{\beta}^{(t)}\}$ 有界。因此，存在 $\{t_i\}$ 的一个子序列 $\{t_{i_\ell}\}$，使得 $\{\boldsymbol{\beta}^{(t_{i_\ell})}\}_\ell$ 收敛到某个 $\widehat{\boldsymbol{\beta}}$，且满足 $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}$。由于 $\left\|[\widehat{\boldsymbol{\beta}}]^{\downarrow} - [\boldsymbol{\beta}^{(t_{i_\ell})}]^{\downarrow}\right\|_\infty \leq \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(t_{i_\ell})}\right\|_\infty$ [4]，可知 $[\boldsymbol{\beta}^{(t_{i_\ell})}]^{\downarrow}$ 收敛到 $[\widehat{\boldsymbol{\beta}}]^{\downarrow}$。因此，我们有

$$[\widehat{\boldsymbol{\beta}}]^{\downarrow}_{k+1} = \lim_{\ell\to\infty}[\boldsymbol{\beta}^{(t_{i_\ell})}]^{\downarrow}_{k+1} \leq \lim_{i\to\infty} p\varepsilon_{t_{i_\ell}} = 0,$$

这表明 $\widehat{\boldsymbol{\beta}}$ 是 k-稀疏的。与之前一样，阶为 k 的零空间性质（NSP）意味着 $\boldsymbol{\beta}$ 是唯一的 ℓ_1-极小化解。接下来需要证明整个序列 $\{\boldsymbol{\beta}^{(t)}\}$ 收敛到 $\widehat{\boldsymbol{\beta}}$。由于 $\boldsymbol{\beta}^{(t_{i_\ell})} \to \widehat{\boldsymbol{\beta}}$ 且 $\varepsilon_{t_{i_\ell}} \to 0$ 当 $\ell \to \infty$时，恒等式（10.38a）表明

$$\lim_{\ell\to\infty}\mathcal{F}(\boldsymbol{\beta}^{(t_{i_\ell})}, \boldsymbol{w}^{(t_{i_\ell})}, \varepsilon_{t_{i_\ell}}) = \left\|\widehat{\boldsymbol{\beta}}\right\|_1.$$

根据（10.38b）和（10.38c）中的单调性，我们得出 $\lim_{t\to\infty}\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) = \left\|\widehat{\boldsymbol{\beta}}\right\|_1$。此外，由（10.38a）和不等式 $\sqrt{(\beta_i^{(t)})^2 + \varepsilon_t} - \varepsilon_t \leq |\beta_i^{(t)}|$，可得

$$\mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t) - p\varepsilon_t \leq \left\|\boldsymbol{\beta}^{(t)}\right\|_1 \leq \mathcal{F}(\boldsymbol{\beta}^{(t)}, \boldsymbol{w}^{(t)}, \varepsilon_t),$$

使得极限 $\lim_{t\to\infty}\left\|\boldsymbol{\beta}^{(t)}\right\|_1 = \left\|\widehat{\boldsymbol{\beta}}\right\|_1$。最后，应用稳定 NSP 及稳定 NSP 的标准推论（定理 4.7），可得

$$\limsup_{t\to\infty}\left\|\boldsymbol{\beta}^{(t)} - \widehat{\boldsymbol{\beta}}\right\|_1 \leq \frac{1+\rho}{1-\rho}\left(\lim_{t\to\infty}\left\|\boldsymbol{\beta}^{(t)}\right\|_1 - \left\|\widehat{\boldsymbol{\beta}}\right\|_1 + 2\sigma_k(\widehat{\boldsymbol{\beta}})_1\right) = 0,$$

这表明 $\boldsymbol{\beta}^{(t)} \to \widehat{\boldsymbol{\beta}}$ 因为 $\widehat{\boldsymbol{\beta}}$ 是 k-稀疏的。误差估计（10.40）可按上述方法类似地推出。

4。参见习题 10.5。

(ii) 当 $\varepsilon \triangleq \lim_{t\to\infty}\varepsilon_t > 0$。我们首先证明 $\boldsymbol{\beta}^{(t)} \to \boldsymbol{\beta}^{(\varepsilon)}$，其中 $\boldsymbol{\beta}^{(\varepsilon)}$ 是问题（10.39）中平滑泛函 F_ε 的极小化解 F_ε。根据性质（10.38d），序列 $\{\boldsymbol{\beta}^{(t)}\}$ 有界，因而有聚点。令 $\{\boldsymbol{\beta}^{(t_i)}\}_i$ 为一个收敛子列，其极限为 $\widehat{\boldsymbol{\beta}}$。我们断言 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(\varepsilon)}$，由 $\boldsymbol{\beta}^{(\varepsilon)}$ 的唯一性可知，每个收敛子列都收敛到 $\boldsymbol{\beta}$。这又意味着 $\boldsymbol{\beta}$ 是 $\{\boldsymbol{\beta}^{(t)}\}_t$ 的唯一聚点，因此，当 $t\to\infty$ 时，$\boldsymbol{\beta}^{(t)}$ 收敛到 $\widehat{\boldsymbol{\beta}}$。

注意到，当 $t\to\infty$ 时，由（10.36）可得 $w_i^{(t)} = ((\beta_i^{(t)})^2+\varepsilon^2)^{-1/2} \le \varepsilon^{-1}$，因此我们有

$$\lim_{i\to\infty} w_i^{(t_i)} = \lim_{i\to\infty}((\beta_i^{(t_i)})^2+\varepsilon^2)^{-1/2} = (\boldsymbol{w}_{\widehat{\boldsymbol{\beta}},\varepsilon})_i \triangleq \widehat{w}_i, \quad i\in\{1,2,\dots,p\},$$

其中 $(\boldsymbol{w}_{\boldsymbol{\beta},\varepsilon})_i = (\beta_i^2+\varepsilon^2)^{-1/2}$——与引理 10.13 中的记号相同。由（10.38f）可知，$\{\boldsymbol{\beta}^{(t_i+1)}\}_i$ 也收敛到 $\widehat{\boldsymbol{\beta}}$。根据定理 3.8 中的加权最小二乘正交关系以及加权最小二乘优化问题 $\boldsymbol{\beta}^{(t_i+1)} = $最小值点集合$_{\boldsymbol{\beta}}\,\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{w}^{(t_i)}\varepsilon_{t_i})$ s.t. $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 的最小化性质（IRLS$_1$），对于每个 $\boldsymbol{n}\in\mathcal{N}(\boldsymbol{X})$，我们有：

$$\left\langle \widehat{\boldsymbol{\beta}}, \boldsymbol{n}\right\rangle_{\widehat{\boldsymbol{w}}} = \lim_{i\to\infty}\left\langle \boldsymbol{\beta}^{(t_i+1)}, \boldsymbol{n}\right\rangle_{\boldsymbol{w}^{(t_i)}} = 0.$$

由引理 10.13可知，这个正交条件刻画了 $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(\varepsilon)}$。

现在我们来证明误差界（10.41）。对于满足 $\boldsymbol{\beta}^*\in\mathbb{R}^p$ 的真实向量 $\boldsymbol{X}\boldsymbol{\beta}^* = \boldsymbol{y}$，由 $\boldsymbol{\beta}^{(\varepsilon)}$ 关于 $F_\varepsilon(\boldsymbol{\beta})$ 的最小化性质可得

$$\left\|\boldsymbol{\beta}^{(\varepsilon)}\right\|_1 \le F_\varepsilon(\boldsymbol{\beta}^{(\varepsilon)}) \le F_\varepsilon(\boldsymbol{\beta}^*) = \sum_{i=1}^p\sqrt{\beta_i^2+\varepsilon^2} \le p\varepsilon + \sum_{i=1}^p|\beta_i| = p\varepsilon + \|\boldsymbol{\beta}^*\|_1.$$

~应用阶为 k 且满足 $k\le k$（即 $\sigma_k(\boldsymbol{\beta}^*)_1 \le \sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1$的稳定 NSP，并结合稳定 NSP 的标准推论（定理4.7），我们得到

$$\left\|\boldsymbol{\beta}^{(\varepsilon)} - \boldsymbol{\beta}^*\right\|_1 \le \frac{1+\rho}{1-\rho}\left(\left\|\boldsymbol{\beta}^{(\varepsilon)}\right\|_1 - \|\boldsymbol{\beta}^*\|_1 + 2\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1\right) \le \frac{1+\rho}{1-\rho}\left(p\varepsilon + 2\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1\right). \quad (10.42)$$

根据 $_2$的定义（IRLS ε_t），满足 $\eta = 1/p$，再利用非递增重排的连续性，$\left\|[\boldsymbol{\beta}^{(t)}]^\downarrow - [\boldsymbol{\beta}^{(\varepsilon)}]^\downarrow\right\|_\infty \le \left\|\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(\varepsilon)}\right\|_\infty$ [5]，可得

$$p\varepsilon = \lim_{t\to\infty} p\varepsilon_t \le \lim_{t\to\infty}[\boldsymbol{\beta}^{(t)}]^\downarrow_{k+1} = [\boldsymbol{\beta}^{(\varepsilon)}]^\downarrow_{k+1},$$

~因此，将上述不等式乘以 $(k+1-k)$，可得：

$$\begin{aligned}(k+1-\widetilde{k})p\varepsilon \le (k+1-\widetilde{k})[\boldsymbol{\beta}^{(\varepsilon)}]^\downarrow_{k+1} &\overset{\dagger}{\le} \left\|\boldsymbol{\beta}^{(\varepsilon)} - \boldsymbol{\beta}^*\right\|_1 + \sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1 \\ &\overset{\ddagger}{\le} \frac{1+\rho}{1-\rho}\left(p\varepsilon + 2\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1\right) + \sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1,\end{aligned}$$

其中不等式（†）由（iii）问题10.5中的不等式（‡）则由（10.42）得出。重新整理各项可得

$$\left(k+1-\widetilde{k}-\frac{1+\rho}{1-\rho}\right)\left(p\varepsilon + 2\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1\right) \le 2\left(\frac{1}{2}+k+1-\widetilde{k}\right)\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1.$$

5. 参见习题 10.5。

在假设 $k - \frac{2\rho}{1-\rho} > \widetilde{k}$ 下，我们有 $k+1-\widetilde{k} > (1+\rho)/(1-\rho)$，因此左侧的系数为正。所以，

$$p\varepsilon + 2\sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1 \leq \frac{2(k-\widetilde{k})+3}{(k-\widetilde{k}) - \frac{2\rho}{1-\rho}} \sigma_{\widetilde{k}}(\boldsymbol{\beta}^*)_1.$$

将其代入（10.42），即可得到所需的上界（10.41）。

最后，假设 $\boldsymbol{\beta}^*$ 是 k-稀疏的，且 $k \leq k$。那么由（10.41）可得 $\boldsymbol{\beta} = \boldsymbol{\beta}^*$，因此 $\widehat{\boldsymbol{\beta}}$ 也 $\widetilde{k}$是 -稀疏的。因此，$[\widehat{\boldsymbol{\beta}}]^{\downarrow}_{k+1} = 0$，根据 $_2$的定义（IRLS ε_t），我们有$\varepsilon = \lim_{t\to\infty} \varepsilon_t = 0$。因此，当 $\varepsilon > 0$ 足够稀疏时，$\boldsymbol{\beta}^*$ 情形不可能发生。■

该定理证明了信号 $\boldsymbol{\beta}$的稳定恢复，与定理 5.7中的结果相呼应。

我们以对收敛速率的估计作为本节 IRLS 算法的结尾。不过，只有当迭代点充分接近其极限时，该速率才开始发挥作用。算法初始阶段的表现没有任何保证，尽管实践经验表明，这一过渡阶段通常很短（Daubechies 等人，2010年；Foucart 和 Rauhut，2013年）。下面的结果证明，在精确稀疏情形下，ℓ_1-范数意义下的线性收敛。

定理 10.15：（稳定 NSP 下 IRLS 的收敛速度）设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$满足阶数为 k、常数为 $\rho < 1 - \frac{2}{k+2}$的稳定 NSP，其中 $\widetilde{k} < k - \frac{2\rho}{1-\rho}$，并考虑测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$，其中 $\boldsymbol{\beta}^* \in \mathbb{R}^p$ 是 k-稀疏的，支撑集为 $\mathbb{S} = \operatorname{supp}(\boldsymbol{\beta}^*)$。设$0 < \nu < 1$ 满足

$$\mu \triangleq \frac{\rho(1+\rho)}{1-\nu}\left(1 + \frac{1}{k+1-\widetilde{k}}\right) < 1.$$

考虑由下列迭代点序列 $\{\boldsymbol{\beta}^{(t)}\}_t$ 通过 IRLS 算法 24 生成，且参数 k 和平滑更新规则 $\eta = 1/p$，并令 $\boldsymbol{\beta}$ 表示其极限。根据定理 10.14，我们有 $\boldsymbol{\beta} = \boldsymbol{\beta}^*$。假设 $t_0 \in \mathbb{N}$ 满足

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(t_0)}\right\|_1 \leq R \triangleq \nu \min_{i\in\mathbb{S}} |\beta_i^*|.$$

那么对所有 $t \geq t_0$，

$$\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(t+1)}\right\|_1 \leq \mu \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(t)}\right\|_1. \tag{10.43}$$

因此，迭代点线性收敛到 $\widehat{\boldsymbol{\beta}}$；具体而言 $\left\|\boldsymbol{\beta}^{(t)} - \widehat{\boldsymbol{\beta}}\right\|_1 \leq \mu^{(t-t_0)} \left\|\boldsymbol{\beta}^{(t_0)} - \widehat{\boldsymbol{\beta}}\right\|_1$ 对所有 $t > t_0$均成立。

证明 [定理 10。15]证明与Daubechies 等人 （2010年）Foucart 和 Rauhut （2013年）的证明基本相同，仅作了少许调整。对于每个 $t > 0$，记误差向量为 $\boldsymbol{n}^{(t)} \triangleq \boldsymbol{\beta}^{(t)} - \widehat{\boldsymbol{\beta}} \in \mathcal{N}$（$\boldsymbol{X}$）（因为 $\boldsymbol{\beta}^{(t)}$和 $\widehat{\boldsymbol{\beta}}$均满足 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$）。根据（IRLS$_1$）的最小化性质以及加权最小二乘极小化解的正交条件（定理3.8），有

$$0 = \left\langle \boldsymbol{\beta}^{(t+1)}, \boldsymbol{n}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}} = \left\langle \widehat{\boldsymbol{\beta}} + \boldsymbol{n}^{(t+1)}, \boldsymbol{n}^{(t+1)} \right\rangle_{\boldsymbol{w}^{(t)}}.$$

由于 $\operatorname{supp}(\boldsymbol{\beta}) = \operatorname{supp}(\boldsymbol{\beta}^*) = \mathbb{S}$，重新整理得

$$\sum_{i=1}^{p} \left(n_i^{(t+1)}\right)^2 w_i^{(t)} = -\sum_{i\in\mathbb{S}} \widehat{\beta}_i n_i^{(t+1)} w_i^{(t)} = -\sum_{i\in\mathbb{S}} \frac{\widehat{\beta}_i n_i^{(t+1)}}{\sqrt{(\beta_i^{(t)})^2+\varepsilon_t^2}}, \tag{10.44}$$

其中，最后一个等式由（10.36）得出。

现在固定 $t \geq t_0$，并记 $C_t \triangleq \left\|\boldsymbol{\beta}^{(t)} - \widehat{\boldsymbol{\beta}}\right\|_1 = \left\|\boldsymbol{n}^{(t)}\right\|_1 \leq R$（见下文的论证）。对于任意 $i \in \mathbb{S}$，我们有 $\left|n_i^{(t)}\right| \leq \left\|\boldsymbol{n}^{(t)}\right\|_1 = C_t \leq \nu\left|\widehat{\beta}_i\right|$，这意味着

$$\frac{\left|\widehat{\beta}_i\right|}{\sqrt{(\beta_i^{(t)})^2+\varepsilon_t^2}} \leq \frac{\left|\widehat{\beta}_i\right|}{\left|\beta_i^{(t)}\right|} = \frac{\left|\widehat{\beta}_i\right|}{\left|\widehat{\beta}_i + n_i^{(t)}\right|} \leq \frac{1}{1-\nu}. \tag{10.45}$$

结合（10.44）和（10.45），可得

$$\sum_{i=1}^{p} \left(n_i^{(t+1)}\right)^2 w_i^{(t)} \leq \frac{1}{1-\nu}\left\|\boldsymbol{n}_{\mathbb{S}}^{(t+1)}\right\|_1 \leq \frac{\rho}{1-\nu}\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1,$$

其中，最后一个不等式由稳定 NSP 的定义（定义 4.6）得出。利用上述不等式和柯西–施瓦茨不等式，可得 [6]

$$\begin{aligned}\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1^2 &\leq \left(\sum_{i\in\mathbb{S}^c}\left(n_i^{(t+1)}\right)^2 w_i^{(t)}\right)\left(\sum_{i\in\mathbb{S}^c}\sqrt{\left(\beta_i^{(t)}\right)^2+\varepsilon_t^2}\right)\\ &\leq \left(\sum_{i=1}^{p}\left(n_i^{(t+1)}\right)^2 w_i^{(t)}\right)\left(\left\|\boldsymbol{\beta}_{\mathbb{S}^c}^{(t)}\right\|_1 + p\varepsilon_t\right) \leq \frac{\rho}{1-\nu}\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1\left(\left\|\boldsymbol{n}^{(t)}\right\|_1 + p\varepsilon_t\right).\end{aligned} \tag{10.46}$$

如果 $\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)} = \mathbf{0}$，那么 $\boldsymbol{\beta}_{\mathbb{S}^c}^{(t+1)} = \mathbf{0}$，因此 $\boldsymbol{\beta}^{(t+1)}$ 是 $\widetilde{k}$-稀疏的，算法根据定义终止。由于 $\boldsymbol{\beta}^{(t+1)} - \widehat{\boldsymbol{\beta}} \in \mathcal{N}(\boldsymbol{X})$，且 NSP 保证 $\mathcal{N}(\boldsymbol{X})$ 不包含非零 $\widetilde{k}$-稀疏向量，因此可得 $\boldsymbol{\beta}^{(t+1)} = \widehat{\boldsymbol{\beta}}$。于是，$C_{t+1} = 0$，且 (10.43) 显然成立。

否则，若 $\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)} \neq \mathbf{0}$，我们可以将 (10.46) 两边除以 $\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1$，得到

$$\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1 \leq \frac{\rho}{1-\nu}\left(\left\|\boldsymbol{n}^{(t)}\right\|_1 + p\varepsilon_t\right).$$

再次利用稳定 NSP（定义 4.6），我们对总误差进行上界估计：

$$\left\|\boldsymbol{n}^{(t+1)}\right\|_1 = \left\|\boldsymbol{n}_{\mathbb{S}}^{(t+1)}\right\|_1 + \left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1 \leq (1+\rho)\left\|\boldsymbol{n}_{\mathbb{S}^c}^{(t+1)}\right\|_1 \leq \frac{\rho(1+\rho)}{1-\nu}\left(\left\|\boldsymbol{n}^{(t)}\right\|_1 + p\varepsilon_t\right).$$

根据 平滑更新规则（IRLS$_2$）针对 ε_{t+1}，假设 $\eta = 1/p$，以及问题10.5中的(iii)，可得

$$p\varepsilon_t \leq [\boldsymbol{\beta}^{(t)}]_{k+1}^{\downarrow} \leq \frac{1}{k+1-\widetilde{k}}\left(\left\|\boldsymbol{\beta}^{(t)} - \widehat{\boldsymbol{\beta}}\right\|_1 + \sigma_k(\widehat{\boldsymbol{\beta}})_1\right) = \frac{\left\|\boldsymbol{n}^{(t)}\right\|_1}{k+1-\widetilde{k}},$$

6. 由于 $\boldsymbol{x}^\top\boldsymbol{y} \leq \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$ $\sum_i a_i^2 \leq \sum a_i^2 b_i \sum \frac{1}{b_i}$ $a_i\ b_i > 0$ $x_i \triangleq |a_i|\sqrt{b_i}$ $y_i \triangleq \frac{1}{\sqrt{b_i}}$ 若，令，对所有

其中 $\sigma_k(\boldsymbol{\beta})_1 = 0$，因为 $\boldsymbol{\beta}$ 是 k-稀疏的。综合以上各项，

$$C_{t+1} = \left\|\boldsymbol{n}^{(t+1)}\right\|_1 \leq \frac{\rho(1+\rho)}{1-\nu}\left(1 + \frac{1}{k+1-\widetilde{k}}\right)\left\|\boldsymbol{n}^{(t)}\right\|_1 = \mu C_t.$$

由于根据假设有 $\mu < 1$，可推出 $C_{t+1} \leq R$，因此条件 $C_t \leq R$ 对所有 $t \geq t_0$ 都成立。因此，对于所有 $t \geq t_0$，都有 $C_{t+1} \leq \mu C_t$，证明完毕。 ■

注意，如果 ρ 足够小——具体而言，如果 $\rho(1+\rho) < 2/3$——那么对于任意 $k \leq k-1$，都存在某个 $\nu > 0$ 使得 $\mu < 1$。因此，只要 $\boldsymbol{\beta}^*$ 是 $(k-1)$-稀疏的，$\boldsymbol{\beta}^{(t)}$ 就会线性收敛到 $\widehat{\boldsymbol{\beta}}$。

10.2.3 一阶增广拉格朗日算法

与罚函数法（算法 12）类似，ℓ_1-最小化问题（P_1）可以通过其拉格朗日形式高效求解（参见例如 Boyd 和 Vandenberghe (2004年)；Figueiredo等 (2008年)；Aybat 和 Iyengar (2012)）：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{\sigma\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2\right\}, \tag{10.47}$$

其中，惩罚参数 $\sigma > 0$ 逐渐减小至零。

用于求解该问题的增广拉格朗日函数（*ALF*）定义如下

$$L_\sigma(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sigma\|\boldsymbol{\beta}\|_1 - \sigma\boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2, \tag{10.48}$$

其中，$\boldsymbol{\lambda}$是与等式约束条件 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$对应的拉格朗日乘子向量。换言之，增广拉格朗日函数通过添加一个二次罚项，对标准拉格朗日函数$L(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \|\boldsymbol{\beta}\|_1 - \boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})$进行了修改，因此得名“增广”拉格朗日函数（Lu，2025年）。

更一般地，对于目标函数为 $f(\boldsymbol{\beta})$、等式约束条件为 $\boldsymbol{c}(\boldsymbol{\beta}) = [c_i(\boldsymbol{\beta})]_{i\in\mathcal{E}}$ 的优化问题，增广拉格朗日方法的核心思想概述于算法25中，其中 $\mathcal{E}$ 表示等式约束的指标集。我们在每次迭代 t中，迭代更新参数 $\boldsymbol{\beta}^{(t)}$ 和乘子 $\boldsymbol{\lambda}^{(t)}$，同时逐步减小惩罚参数 σ_t。在初期迭代中，通常固定乘子（例如，设 $\boldsymbol{\lambda} = \boldsymbol{0}$），并让惩罚参数 σ 占主导地位，从而推动迭代点趋于可行，并使我们更接近真正的解 $\boldsymbol{\beta}_*$。

作为针对带等式约束的ℓ_1最小化问题而设计的增广拉格朗日方法的一种改进，算法 26 给出了用于求解问题 $(\boldsymbol{P}_1)$ 的一阶增广拉格朗日（FAL）算法。子程序 PGM在步骤 (FAL_7) 中使用，其定义见算法 27 (参见第 9.2.2 节)（Aybat 和 Iyengar，2012）。[7]

FAL 算法通过近似求解以下形式的一系列子问题来解决 ℓ_1-最小化问题

$$\begin{aligned}&\min_{\boldsymbol{\beta}\in\mathbb{R}^p:\|\boldsymbol{\beta}\|_1\leq\eta_t}\left\{\sigma_t\|\boldsymbol{\beta}\|_1 - \sigma_t\boldsymbol{\lambda}^{(t)\top}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2\right\}\\ \equiv &\min_{\boldsymbol{\beta}\in\mathbb{R}^p:\|\boldsymbol{\beta}\|_1\leq\eta_t} P_t(\boldsymbol{\beta}) \triangleq \sigma_t\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2\end{aligned} \tag{10.49}$$

7. 原论文 Aybat 和 Iyengar (2012) 使用加速近端梯度法求解该子问题；为简便起见，此处我们使用标准近端梯度法。

算法 25 增广拉格朗日方法：一般情形（Lu，2025年 ）

要求：函数 $f(\boldsymbol{\beta})$ 和一组等式约束 $\{c_i(\boldsymbol{\beta})\}, i \in \mathcal{E}$; 1: 输入：选择初始点 $\boldsymbol{\beta}^{(1)}$、乘子 $\boldsymbol{\lambda}^{(1)}$、罚因子 $\sigma_1 > 0$、罚因子更新常数 $\rho \in (0, 1)$、约束违反容差 $\varepsilon > 0$ 和精度要求 $\gamma_t > 0$。2: 定义： $L_\sigma(\boldsymbol{\beta}, \boldsymbol{\lambda}) = f(\boldsymbol{\beta}) + \boldsymbol{\lambda}^\top \boldsymbol{c}(\boldsymbol{\beta}) + \frac{1}{2}\sigma \boldsymbol{c}(\boldsymbol{\beta})^\top \boldsymbol{c}(\boldsymbol{\beta})$; 3: 对于 $t = 1, 2, \ldots,$ 执行 4: 以 $\boldsymbol{\beta}^{(t)}$ 作为初始点，求解 $\min_{\boldsymbol{\beta}} L_{\sigma_t}(\boldsymbol{\beta}, \boldsymbol{\lambda}^{(t)})$，_得到满足精度条件 $\left\|\nabla_{\boldsymbol{\beta}} L_{\sigma_t}(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\lambda}^{(t)})\right\|_2 \leq \gamma_t$ 的解 $\boldsymbol{\beta}^{(t+1)}$。 5: 如果 $\left\|\boldsymbol{c}(\boldsymbol{\beta}^{(t+1)})\right\|_2 \leq \varepsilon$ ， 则 6: 返回近似解 $\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\lambda}^{(t)}$，终止迭代。 7: 结束 if 8: 更新乘子： $\boldsymbol{\lambda}^{(t+1)} \leftarrow \boldsymbol{\lambda}^{(t)} + \boldsymbol{c}(\boldsymbol{\beta}^{(t+1)})/\sigma_t$。 9: 更新罚因子： $\sigma_{t+1} \leftarrow \rho\sigma_t$。 10: 结束 for 11: 返回最终的 $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^{(t)}$;

对于适当选取的序列 $\{\sigma_t, \boldsymbol{\lambda}^{(t)}, \eta_t\}$、 $t = 0$、1、.....FAL 停止准则记为 FAL停止准则，其定义如下

$$\texttt{FALstop} = \left\|\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t-1)}\right\|_\infty \leq \gamma, \tag{10.50}$$

其中 γ 是预先给定的容差。最后，PGM子程序在第 ϵ_t 次迭代计算子问题(10.49)的一个 t-最优解。

算法 26 FAL$\{(\sigma_t, \epsilon_t, \tau_t)^{t\in\mathbb{Z}_+}$， $\widetilde{\eta}\}$ Aybat 和 Iyengar (2012)

Ensure: Find $\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \left\{\sigma_t \|\boldsymbol{\beta}\|_1 - \sigma_t \boldsymbol{\lambda}^{(t)\top}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2\right\}$;

input: $L_b = \sigma_{\max}(\boldsymbol{X}\boldsymbol{X}^\top)$, $\rho_1, \rho_2, \rho_3 \in (0, 1)$;
initialization: $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\lambda}^{(0)} \leftarrow \boldsymbol{0}$, $t \leftarrow 0$;
while (FALstop not true) **do**
  $g_t(\boldsymbol{\beta}) \leftarrow \sigma_t \|\boldsymbol{\beta}\|_1$, $f_t(\boldsymbol{\beta}) \leftarrow \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t \boldsymbol{\lambda}^{(t)}\right\|_2^2$; ▷ (FAL_1)
  $P_t(\boldsymbol{\beta}) \leftarrow f_t(\boldsymbol{\beta}) + g_t(\boldsymbol{\beta})$; ▷ (FAL_2)
  $\eta_t \leftarrow \widetilde{\eta} + (\sigma_t/2)\left\|\boldsymbol{\lambda}^{(t)}\right\|_2$; ▷ (FAL_3)
  $\mathbb{S}^{(t)} \leftarrow \{\boldsymbol{\beta} \in \mathbb{R}^p \mid \|\boldsymbol{\beta}\|_1 \leq \eta_t\}$; ▷ (FAL_4)
  $\ell_{t,\max} \leftarrow \sigma_{\max}(\boldsymbol{X})(\eta_t + \|\boldsymbol{\beta}^{(t-1)}\|_2)\sqrt{2/\epsilon_t}$; ▷ (FAL_5)
  PGMstop $= \{\ell \geq \ell_{t,\max}\}$ or $\{\exists \boldsymbol{g} \in \partial P_t(\boldsymbol{\zeta}^{(\ell)})$, with $\|\boldsymbol{g}\|_2 \leq \tau_t\}$; ▷ (FAL_6)
  $\boldsymbol{\beta}^{(t)} \leftarrow \texttt{PGM}(g_t, f_t, L_b, \mathbb{S}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \texttt{PGMstop})$, see Algorithm 27; ▷ (FAL_7)
  $\boldsymbol{\lambda}^{(t+1)} \leftarrow \boldsymbol{\lambda}^{(t)} - (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})/\sigma_t$; ▷ (FAL_8)
  $\sigma_{t+1} \leftarrow \rho_1\sigma_t$, $\epsilon_{t+1} \leftarrow \rho_2\epsilon_t$, $\tau_{t+1} \leftarrow \rho_3\tau_t$; ▷ (FAL_9)
  $t \leftarrow t + 1$;
end while
output: $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}^{(t)}$

更具体地说，FAL 算法的输入包括一个序列 $\{(\sigma_t, \epsilon_t, \tau_t)\}_{t\in\mathbb{Z}_+}$、一个初始点 $\boldsymbol{\beta}^{(0)}$，以及最优解 $\boldsymbol{\beta}_*$ 的 ℓ_1-范数上界 $\widetilde{\eta}$，用于 ℓ_1-最小化

算法 27 近端梯度法：PGM(g_t, f_t, L_b, $\mathbb{S}^{(t)}$, $\boldsymbol{\beta}^{(0)}$, PGMstop)

要求： 函数 $f(\boldsymbol{\beta}) = f_t(\boldsymbol{\beta})$，以及满足以下条件的闭凸函数 $g(\boldsymbol{\beta}) = g_t(\boldsymbol{\beta})$
定理 9.12 中的 (A1) 和 (A2)。 $f(\boldsymbol{\beta})$ 是 L_b-光滑的（见例 2.72）；

初始化： $\boldsymbol{\zeta}^{(0)} \leftarrow \boldsymbol{\beta}^{(0)}$, $\ell \leftarrow 0$;
当(PGM停止条件不成立)**时**
  选择步长 $\eta_\ell = \frac{1}{L_\ell} = \frac{1}{L_b}$; ▷ ($\mathrm{PGM}_1$)
  $\boldsymbol{\alpha}^{(\ell+1)} \leftarrow \boldsymbol{\zeta}^{(\ell)} - \eta_\ell \nabla f(\boldsymbol{\zeta}^{(t)})$; ▷ ($\mathrm{PGM}_2$)
  $\boldsymbol{\zeta}^{(\ell+1)} \leftarrow \mathcal{P}\mathrm{rox}_{\eta_\ell g}(\boldsymbol{\alpha}^{(\ell+1)}) \triangleq \mathcal{T}^{f,\ g}_{L_\ell}(\boldsymbol{\zeta}^{(t)})$; ▷ ($\mathrm{PGM}_3$)
结束 while
返回 $\boldsymbol{\zeta}^{(t)}$;

问题。一个这样的上界 $\widetilde{\eta}$ 可以按如下方式计算。令

$$\widetilde{\boldsymbol{\beta}} \triangleq \arg\min\{\|\boldsymbol{\beta}\|_2 \mid \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}\} = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{y} \tag{10.51}$$

表示相应 ℓ_2-最小化问题的解。根据定义，我们有

$$\|\boldsymbol{\beta}_*\|_1 \le \widetilde{\eta} \triangleq \left\|\widetilde{\boldsymbol{\beta}}\right\|_1. \tag{10.52}$$

现在我们详细介绍该算法的各个组成部分。如上所述，对于 ℓ_1-最小化问题（P_1），其增广拉格朗日函数可写为

$$P(\boldsymbol{\beta}) \triangleq L_\sigma(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sigma\|\boldsymbol{\beta}\|_1 - \sigma\boldsymbol{\lambda}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2, \tag{10.53}$$

其中，σ 是惩罚参数，$\boldsymbol{\lambda}$ 是与约束$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$相关的对偶变量。根据算法$_2$（$\sim$）$_5$（26）的步骤（FAL），FAL 在将 $\boldsymbol{\lambda}$固定为当前值 $\boldsymbol{\lambda}^{(t)}$的同时，近似最小化该增广拉格朗日函数。具体而言，它在函数

$$\begin{aligned} P_t(\boldsymbol{\beta}) &= \sigma_t\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2 \\ &= \sigma_t\|\boldsymbol{\beta}\|_1 - \sigma_t(\boldsymbol{\lambda}^{(t)})^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \frac{1}{2}\left\|\sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2, \end{aligned} \tag{10.54}$$

上使用 PGM 算法（或 PGD；见算法3）对约束集 $\mathbb{S}^{(t)} = \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta}\|_1 \le \eta_t\}$ 进行最小化。

▶ η_t的选择。应用 PGM 时，我们必须确保 $P_t(\boldsymbol{\beta})$ 在 $\mathbb{R}^p$ 上的极小化解位于可行集 $\mathbb{S}^{(t)}$ 内，即 $\mathbb{S}^{(t)} \cap \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{P_t(\boldsymbol{\beta})\} \neq \varnothing$。令 $\boldsymbol{\beta}_*^{(t)} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} P_t(\boldsymbol{\beta})$。由于 $P_t(\boldsymbol{\beta}_*^{(t)}) \le P_t(\boldsymbol{\beta}_*)$，且 $\boldsymbol{X}\boldsymbol{\beta}_* = \boldsymbol{y}$ 和 $\|\boldsymbol{\beta}_*\|_1 \le \widetilde{\eta}$ 由式（10.52）可得$\left\|\boldsymbol{\beta}_*^{(t)}\right\|_1 \le \eta_t$。因此，$\boldsymbol{\beta}_*^{(t)} \in \mathbb{S}^{(t)}$，这就解释了为何选择

$$\eta_t = \widetilde{\eta} + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \tag{10.55}$$

在 FAL 算法中，该算法保证 $\boldsymbol{\beta}_*^{(t)} \in \mathbb{S}^{(t)}$。

▶ **停止准则**。现在讨论步骤（FAL_7）中的停止条件。近端梯度法的收敛速率（参见定理 9.12）确保

$$P_t(\boldsymbol{\zeta}^{(\ell)}) \leq P_t(\boldsymbol{\beta}_*^{(t)}) + \epsilon_t, \quad \text{for} \quad \ell \geq \frac{L_b \left\| \boldsymbol{\beta}^{(t-1)} - \boldsymbol{\beta}_*^{(t)} \right\|_2^2}{2\epsilon_t} + 1,$$

其中 $\{\boldsymbol{\zeta}^{(\ell)}\}_{\ell \in \mathbb{Z}_+}$ 是 $\boldsymbol{\zeta}$在 t子问题上应用 PGM 算法时的迭代点序列，而 L_b 表示梯度 $\nabla f_t(\boldsymbol{\beta})$ 的 Lipschitz 常数（即 L_b-光滑；定义 2.78）。由于 $\left\|\boldsymbol{\beta}_*^{(t)}\right\|_1 \leq \eta_t$ 且 $\|\cdot\|_1 \geq \|\cdot\|_2$（习题 1.26），由三角不等式可得

$$\left\|\boldsymbol{\beta}_*^{(t)} - \boldsymbol{\beta}^{(t-1)}\right\|_2 \leq \left\|\boldsymbol{\beta}_*^{(t)}\right\|_1 + \left\|\boldsymbol{\beta}^{(t-1)}\right\|_2 \leq \eta_t + \left\|\boldsymbol{\beta}^{(t-1)}\right\|_2. \tag{10.56}$$

由于 $f_t(\boldsymbol{\beta}) \triangleq \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t \boldsymbol{\lambda}^{(t)}\right\|_2^2$ 且 $\nabla f_t(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t \boldsymbol{\lambda}^{(t)})$，可得 $L_b = \sigma_{\max}^2(\boldsymbol{X})$，其中 $\sigma_{\max}(\boldsymbol{X})$ 表示 $\boldsymbol{X}$ 的最大奇异值。因此，可得

$$\frac{L_b \left\| \boldsymbol{\beta}^{(t-1)} - \boldsymbol{\beta}_*^{(t)} \right\|_2^2}{2\epsilon_t} + 1 \leq \frac{\sigma_{\max}^2(\boldsymbol{X})}{2\epsilon_t} \left(\eta_t + \left\|\boldsymbol{\beta}^{(t-1)}\right\|_2\right)^2 \triangleq \ell_{t,\max}.$$

因此，运行 PGM 至少 $\ell_{t,\ \max}$ 次可保证得到第 t 个子问题的 ϵ_t-最优解。

因此可知，算法 FAL 的步骤 (FAL_7) 中的停止准则 PGMstop 确保迭代点 $\boldsymbol{\beta}^{(t)}$ 满足以下两个条件之一：

$$P_t(\boldsymbol{\beta}^{(t)}) \leq \min_{\boldsymbol{\beta} \in \mathbb{R}^p} P_t(\boldsymbol{\beta}_*^{(t)}) + \epsilon_t; \tag{10.57a}$$

$$\exists \boldsymbol{g} \in \partial P_t(\boldsymbol{\beta}^{(t)}) \text{ with } \|\boldsymbol{g}\|_2 \leq \tau_t, \tag{10.57b}$$

其中，$\partial P_t(\boldsymbol{\beta}^{(t)})$ 表示函数 P_t 在 $\boldsymbol{\beta}^{(t)}$ 处的次梯度集合，且 $\boldsymbol{\beta}_*^{(t)} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} P_t(\boldsymbol{\beta})$。

最后，在步骤 (FAL_8) 中，使用增广拉格朗日方法所采用的标准规则更新对偶变量 $\boldsymbol{\lambda}^{(t)}$（参见算法 25）。

FAL 的收敛性

现在给出 FAL 算法的主要收敛结果。以下定理证明，由 $\{\boldsymbol{\beta}^{(t)}\}_{t \in \mathbb{Z}_+}$ FAL 生成的迭代序列中的每个极限点都是 ℓ_1-最小化问题 (P_1) 的最优解。

为证明这一结果，我们需要以下引理。

引理 10.16：（ϵ-**最优梯度引理**）设 $h: \mathbb{R}^p \to \mathbb{R}$ 是一个凸函数，并假设 f 是 L_b-光滑的（即，其梯度 ∇h 是 Lipschitz 连续的，常数为L_b，如定理 2.80所述）。固定 $\epsilon > 0$，并令 $\boldsymbol{\beta} \in \mathbb{R}^p$ 满足

$$\sigma \left\|\overline{\boldsymbol{\beta}}\right\|_1 + h(\overline{\boldsymbol{\beta}}) - \left(\sigma \left\|\widehat{\boldsymbol{\beta}}\right\|_1 + h(\widehat{\boldsymbol{\beta}})\right) \leq \epsilon,$$

其中 $\widehat{\boldsymbol{\beta}} \in$ 最小值点集合$\{\sigma \|\boldsymbol{\beta}\|_1 + h(\boldsymbol{\beta})\colon \ \boldsymbol{\beta} \in \mathbb{R}^p\}$ （即，$\overline{\boldsymbol{\beta}}$ 对于该最小化问题是 ϵ-最优的）。那么

$$\frac{1}{2L_b} \sum_{i:|\nabla h_i(\overline{\boldsymbol{\beta}})|>\sigma} \big(\big|\nabla h_i(\overline{\boldsymbol{\beta}})\big| - \sigma\big)^2 \leq \epsilon. \tag{10.58}$$

上界 (10.58) 意味着 $\big\|\nabla h(\overline{\boldsymbol{\beta}})\big\|_\infty \leq \sqrt{2L_b\epsilon} + \sigma$。

证明 [引理 10.16]关于 $\|\cdot\|_1$ 和 L_b-光滑性的三角不等式（定义 2.78）意味着，对所有 $\boldsymbol{\alpha} \in \mathbb{R}^p$

$$\sigma \|\boldsymbol{\alpha}\|_1 + h(\boldsymbol{\alpha}) \leq \sigma \big\|\overline{\boldsymbol{\beta}}\big\|_1 + h(\overline{\boldsymbol{\beta}}) + \nabla h(\overline{\boldsymbol{\beta}})^\top(\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}) + \frac{L_b}{2}\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_2^2 + \sigma\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_1.$$

对两边关于 $\boldsymbol{\alpha} \in \mathbb{R}^p$ 取最小值，得到

$$\sigma\big\|\widehat{\boldsymbol{\beta}}\big\|_1 + h(\widehat{\boldsymbol{\beta}}) \leq \sigma \big\|\overline{\boldsymbol{\beta}}\big\|_1 + h(\overline{\boldsymbol{\beta}}) + \min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\left\{\nabla h(\overline{\boldsymbol{\beta}})^\top(\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}) + \frac{L_b}{2}\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_2^2 + \sigma\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_1\right\}. \tag{10.59}$$

令 $\boldsymbol{g} \triangleq \nabla h(\overline{\boldsymbol{\beta}})$。右侧的最小化问题可改写为

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\left\{\overline{\boldsymbol{g}}^\top(\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}) + \frac{L_b}{2}\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_2^2 + \sigma\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_1\right\} \\
&= \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\left\{\frac{1}{2}\left\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}} - \frac{-\overline{\boldsymbol{g}}}{L_b}\right\|_2^2 + \frac{\sigma}{L_b}\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_1\right\} \\
&\overset{\dagger}{=} \overline{\boldsymbol{\beta}} + \operatorname{sign}\left(\frac{-\overline{\boldsymbol{g}}}{L_b}\right) \circ \max\left\{\left|\frac{-\overline{\boldsymbol{g}}}{L_b}\right| - \frac{\sigma}{L_b}\mathbf{1}, \mathbf{0}\right\} = \overline{\boldsymbol{\beta}} - \frac{\operatorname{sign}(\overline{\boldsymbol{g}})}{L_b} \circ \max\{|\overline{\boldsymbol{g}}| - \sigma\mathbf{1}, \mathbf{0}\},
\end{aligned}$$

其中，等式 (†) 由例 2.104（ℓ_1范数的近端算子）得出，所有向量算子，例如 $|\cdot|$、符号$(\cdot)$ 和 $\max\{\cdot,\cdot\}$，均按分量定义。将 $\widehat{\boldsymbol{\alpha}}$代入 (10.59)，得到

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}\in\mathbb{R}^p}\left\{\overline{\boldsymbol{g}}^\top(\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}) + \frac{L_b}{2}\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_2^2 + \sigma\big\|\boldsymbol{\alpha} - \overline{\boldsymbol{\beta}}\big\|_1\right\} \\
&= -\sum_i \frac{|\overline{g}_i|}{L_b}\max\{|\overline{g}_i| - \sigma, 0\} + \frac{1}{2L_b}\sum_i \max\{|\overline{g}_i| - \sigma, 0\}^2 + \frac{\sigma}{L_b}\sum_i \max\{|\overline{g}_i| - \sigma, 0\} \\
&= \frac{1}{L_b}\sum_{i:|\overline{g}_i|>\sigma}\left(-|\overline{g}_i| + \frac{1}{2}(|\overline{g}_i| - \sigma) + \sigma\right)(|\overline{g}_i| - \sigma) = -\frac{1}{2L_b}\sum_{i:|\overline{g}_i|>\sigma}(|\overline{g}_i| - \sigma)^2.
\end{aligned}$$

界（10.58）由以下假设可得：$\sigma\big\|\overline{\boldsymbol{\beta}}\big\|_1 + h(\overline{\boldsymbol{\beta}}) - (\sigma\big\|\widehat{\boldsymbol{\beta}}\big\|_1 + h(\widehat{\boldsymbol{\beta}})) \leq \epsilon$。界（10.58）显然意味着，对于$|g_i| \leq \sqrt{2L_b\epsilon} + \sigma$ 所有 i，由向量范数的标准界（习题1.26）可得，即 $\big\|\nabla h(\overline{\boldsymbol{\beta}})\big\|_\infty \leq \sqrt{2L_b\epsilon} + \sigma$。 ■

该引理给出了 h 在增广拉格朗日子问题（10.54）的任意 ϵ-最优解处梯度的定量界。下面给出一个直接推论。

引理 10.17 设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足 $n \leq p$ 且满秩，并令 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 为一个固定向量。定义函数

$$P(\boldsymbol{\beta}) \triangleq \sigma\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma\boldsymbol{\lambda}\|_2^2.$$

假设 $\overline{\boldsymbol{\beta}}$是 $\min_{\boldsymbol{\beta}\in\mathbb{R}^p} P(\boldsymbol{\beta})$ 的一个 ϵ-最优解，即 $0 \le P(\overline{\boldsymbol{\beta}}) - \min_{\boldsymbol{\beta}\in\mathbb{R}^p} P(\boldsymbol{\beta}) \le \epsilon$。则有以下界：

$$\left\| \boldsymbol{X}^\top (\boldsymbol{X}\overline{\boldsymbol{\beta}} - \boldsymbol{y} - \sigma\boldsymbol{\lambda}) \right\|_\infty \le \sqrt{2\epsilon}\,\sigma_{\max}(\boldsymbol{X}) + \sigma; \tag{10.60a}$$

$$\left\| \boldsymbol{X}\overline{\boldsymbol{\beta}} - \boldsymbol{y} - \sigma\boldsymbol{\lambda} \right\|_2 \le \frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})} \left(\sqrt{2\epsilon}\,\sigma_{\max}(\boldsymbol{X}) + \sigma \right), \tag{10.60b}$$

其中，$\sigma_{\max}(\boldsymbol{X})$ 和 $\sigma_{\min}(\boldsymbol{X})$ 分别表示 $\boldsymbol{X}$ 的最大和最小非零奇异值。

证明 [引理 10。17] 设 $h(\boldsymbol{\beta}) \triangleq \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma\boldsymbol{\lambda}\|_2^2$。则 $\nabla h(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma\boldsymbol{\lambda})$。对于任意 $\boldsymbol{\beta}$, $\boldsymbol{\alpha} \in \mathbb{R}^p$我们有

$$\|\nabla h(\boldsymbol{\beta}) - \nabla h(\boldsymbol{\alpha})\|_2 = \left\| \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\alpha}) \right\|_2 \le \sigma_{\max}^2(\boldsymbol{X})\,\|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2,$$

参见奇异值不等式（1.29）。因此，$h: \mathbb{R}^p \to \mathbb{R}$ 是一个凸函数，并且 h 是L_b-光滑函数，其常数为 $L_b = \sigma_{\max}^2(\boldsymbol{X})$。

由于$\overline{\boldsymbol{\beta}}$是一个 ϵ-最优解，满足 $\min_{\boldsymbol{\beta}\in\mathbb{R}^p} P(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{\sigma\|\boldsymbol{\beta}\|_1 + h(\boldsymbol{\beta})\}$，引理 10.16立即推出（10.60a）。再次利用奇异值不等式（1.29）和向量范数的标准界（习题 1.26），第二个界由以下事实得出：

$$\|\boldsymbol{X}\overline{\boldsymbol{\beta}} - \boldsymbol{y} - \sigma\boldsymbol{\lambda}\|_2 \le \frac{\|\boldsymbol{X}^\top(\boldsymbol{X}\overline{\boldsymbol{\beta}} - \boldsymbol{y} - \sigma\boldsymbol{\lambda})\|_2}{\sigma_{\min}(\boldsymbol{X})} \le \frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})} \left\| \boldsymbol{X}^\top(\boldsymbol{X}\overline{\boldsymbol{\beta}} - \boldsymbol{y} - \sigma\boldsymbol{\lambda}) \right\|_\infty.$$

证明至此完成。 ■

利用上述引理，我们现在证明 FAL 算法的收敛性。

定理10.18：（FAL 的收敛性（Aybat 和 Iyengar，2012））假设$\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满行秩。固定一个初始点 $\boldsymbol{\beta}^{(0)} \in \mathbb{R}^p$、一个常数 $\widetilde{\eta} > 0$ 使得 $\widetilde{\eta} \ge \|\boldsymbol{\beta}_*\|_1$，以及一个参数序列 $\{(\sigma_t,\ \epsilon_t,\ \tau_t)\}_{t\in\mathbb{Z}_+}$ 满足以下条件：

（i）惩罚参数，$\sigma_t \searrow 0$；（ii）近似最优性参数，$\epsilon_t \searrow 0$，使得对于所有 $t \ge 1$，$\epsilon_t/\sigma_t^2 \le B_1$；（iii）近似可行性参数，$\tau_t \searrow 0$，使得对于所有 $t \ge 1$，$\tau_t/\sigma_t \le B_2$；并且当 $t \to \infty$ 时，$\tau_t/\sigma_t \to 0$。

令 $\mathcal{X} = \{\boldsymbol{\beta}^{(t)}\}_{t\in\mathbb{Z}_+}$ 表示使用这些参数通过 FAL 算法（算^法 26）生成的迭代序列。那么，$\mathcal{X}$ 是一个有界序列，且其任意极限点 $\boldsymbol{\beta}$都是 $\mathcal{X}$-最小化问题 ℓ_1（P_1）的最优解。

定理 [的证明 10.18] 由于 $\sigma_{\max}(\boldsymbol{X})$ 对于所有 $t \ge 1$均为有限值，因此序列$\mathcal{X}$ 存在。

有界性。为证明 $\mathcal{X}$ 有界，我们首先建立对偶乘子序列 $\{\boldsymbol{\lambda}^{(t)}\}_{t\in\mathbb{Z}_+}$的一个均匀上界，然后利用（10.55）为迭代点 $\{\boldsymbol{\beta}^{(t)}\}$给出上界。在第 t次迭代中，假设 PGM 算法以满足条件（10.57a）的迭代点 $\boldsymbol{\beta}^{(t)}$终止。将引理 10.17应用于 $P_t(\boldsymbol{\beta}) = \sigma_t\|\boldsymbol{\beta}\|_1 + \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}\|_2^2$ 和（10.54），可得

$$\left\| \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}) \right\|_\infty \le \sqrt{2\epsilon_t}\,\sigma_{\max}(\boldsymbol{X}) + \sigma_t. \tag{10.61}$$

或者，如果迭代点 $\boldsymbol{\beta}^{(t)}$ 满足条件（10.57b），则存在 $\boldsymbol{q}^{(t)} \in \partial\left\|\boldsymbol{\beta}^{(t)}\right\|_1$使得 $\partial P_t(\boldsymbol{\beta}^{(t)}) \ni \boldsymbol{g}^{(t)} \triangleq \left\|\sigma_t \boldsymbol{q}^{(t)} + \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t \boldsymbol{\lambda}^{(t)})\right\|_2 \le \tau_t$，因此根据三角不等式和向量范数的标准界（习题 1.26），有

$$\begin{aligned}\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)})\right\|_\infty &\le \left\|\sigma_t\boldsymbol{q}^{(t)} + \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)})\right\|_\infty + \sigma_t\left\|\boldsymbol{q}^{(t)}\right\|_\infty \\ &\le \tau_t + \sigma_t, \end{aligned} \tag{10.62}$$

其中最后一个不等式由以下事实得出：$\left\|\boldsymbol{q}^{(t)}\right\|_\infty \le 1$ 对所有 $\boldsymbol{q}^{(t)} \in \partial\left\|\boldsymbol{\beta}^{(t)}\right\|_1$；参见习题 2.58。

由于 $\boldsymbol{\lambda}^{(0)} = \boldsymbol{0}$，$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} - (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})/\sigma_t$ 对所有 $t \ge 1$ 且 $\boldsymbol{X}$ 具有满行秩，由 (1.29)、(10.61)、(10.62以及向量范数的标准界（习题1.26）可得

$$\begin{aligned}\left\|\boldsymbol{\lambda}^{(t+1)}\right\|_2 &\le \frac{1}{\sigma_t\sigma_{\min}(\boldsymbol{X})}\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)})\right\|_2 \\ &\le \frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})}\left(\max\left\{\sigma_{\max}(\boldsymbol{X})\sqrt{\frac{2\epsilon_t}{\sigma_t^2}},\ \frac{\tau_t}{\sigma_t}\right\} + 1\right), \quad \forall\, t \ge 1. \end{aligned} \tag{10.63}$$

结合界 $\frac{\epsilon_t}{\sigma_t^2} \le B_1$，$\frac{\tau_t}{\sigma_t} \le B_2$（由假设 (ii) 和 (iii) 得出）以及 (10.63)，可得

$$\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \le B_{\boldsymbol{\lambda}} \triangleq \frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})}\left(\max\left\{\sqrt{2B_1}\,\sigma_{\max}(\boldsymbol{X}),\ B_2\right\} + 1\right), \quad \forall\, t > 1, \tag{10.64}$$

它将 $\left\|\boldsymbol{\beta}^{(t)}\right\|_1$ 上界为

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \eta_t = \widetilde{\eta} + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t)}\right\|_2^2 \le B_b \triangleq \widetilde{\eta} + \frac{1}{2}\sigma_1 B_{\boldsymbol{\lambda}}^2. \tag{10.65}$$

因此，序列 $\mathcal{X}$ 有界，因而至少存在一个极限点。令 $\boldsymbol{\beta}$ 为任意一个这样的极限点，并令 $\mathbb{T} \subset \mathbb{Z}_+$ 为一个无限子序列，使得 $\lim_{t\in\mathbb{T}} \boldsymbol{\beta}^{(t)} = \widehat{\boldsymbol{\beta}}$。

满足条件 (10.57a)时的收敛。 假设存在一个进一步的子序列 $\mathbb{T}_a \subset \mathbb{T}$，使得对所有 $t \in \mathbb{T}_a$，迭代点 $\boldsymbol{\beta}^{(t)}$ 满足条件 (10.57a)。那么，对于 $t \in \mathbb{T}_a$，有

$$\begin{aligned}\left\|\boldsymbol{\beta}^{(t)}\right\|_1 &\le \frac{P_t(\boldsymbol{\beta}^{(t)})}{\sigma_t} \le \frac{P_t(\boldsymbol{\beta}_*^{(t)}) + \epsilon_t}{\sigma_t} \le \frac{P_t(\boldsymbol{\beta}_*) + \epsilon_t}{\sigma_t} = \frac{P_t(\boldsymbol{\beta}_*)}{\sigma_t} + \frac{\epsilon_t}{\sigma_t} \\ &= \|\boldsymbol{\beta}_*\|_1 + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t)}\right\|_2^2 + \frac{\epsilon_t}{\sigma_t} \le \|\boldsymbol{\beta}_*\|_1 + \frac{1}{2}\sigma_t B_{\boldsymbol{\lambda}}^2 + \sigma_t B_1, \end{aligned} \tag{10.66}$$

其中，第一个不等式由以下事实得出：$f_t(\boldsymbol{\beta}) \ge 0$；第二个不等式由停止条件得出；第三个不等式由以下事实得出：$P_t(\boldsymbol{\beta}_*^{(t)}) \le P_t(\boldsymbol{\beta}_*)$（$\boldsymbol{\beta}_*$ 表示 ℓ_1-最小化问题（P_1）的最优解；第二个等式由以下事实得出：$\boldsymbol{X}\boldsymbol{\beta}_* = \boldsymbol{y}$；最后一个不等式由界$\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \le B_{\boldsymbol{\lambda}}$ 和 $\frac{\epsilon_t}{\sigma_t^2} \le B_1$得出。由于 $\sigma_t \searrow 0$，令 $t \in \mathbb{T}_a \to \infty$趋于极限，我们得到

$$\left\|\widehat{\boldsymbol{\beta}}\right\|_1 = \lim_{t\in\mathbb{T}_a}\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \|\boldsymbol{\beta}_*\|_1 + \lim_{t\in\mathbb{T}_a}\left\{\sigma_t\left(\frac{1}{2}B_{\boldsymbol{\lambda}}^2 + B_1\right)\right\} = \|\boldsymbol{\beta}_*\|_1. \tag{10.67}$$

^接下来，考虑极限点 $\boldsymbol{\beta}$的可行性。根据 P_t 的定义和停止条件（10.57a），

$$\begin{aligned}\frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}-\sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2 &\leq P_t(\boldsymbol{\beta}^{(t)}) \leq P_t(\boldsymbol{\beta}_*^{(t)})+\epsilon_t \leq P_t(\boldsymbol{\beta}_*)+\epsilon_t \\ &\leq \sigma_t\|\boldsymbol{\beta}_*\|_1+\frac{1}{2}\left\|\sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2+\epsilon_t \leq \sigma_t\|\boldsymbol{\beta}_*\|_1+\frac{1}{2}(\sigma_t B_{\boldsymbol{\lambda}})^2+\epsilon_t,\end{aligned}$$

其中，第一个不等式由事实 $\sigma_t\left\|\boldsymbol{\beta}^{(t)}\right\|_2 \geq 0$ 得出，第三个不等式由事实 $P_t(\boldsymbol{\beta}_*^{(t)}) \leq P_t(\boldsymbol{\beta}_*)$ 得出（其中 $\mathbb{S}^{(t)} \cap \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{P_t(\boldsymbol{\beta})\}$），第四个不等式由事实 $\boldsymbol{X}\boldsymbol{\beta}_* = \boldsymbol{y}$ 得出，最后一个不等式由上界 $\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \leq B_{\boldsymbol{\lambda}}$ 得出。令$t \in \mathbb{T}_a \to \infty$取极限，得到

$$\frac{1}{2}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{y}\right\|_2^2 \leq 0,$$

即 $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}$， $\widehat{\boldsymbol{\beta}}$的可行性。由于 $\widehat{\boldsymbol{\beta}}$ 是可行的，且 $\|\widehat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}_*\|_1$，因此 $\widehat{\boldsymbol{\beta}}$是 ℓ_1-最小化问题（P_1）的最优解。

在条件 (10.57b) 下收敛。 现在，考虑补集情形，即存在 $T \in \mathbb{T}$，使得对所有 $t \in \mathbb{T}_b \triangleq \mathbb{T} \cap \{t \geq T\}$，迭代点 $\boldsymbol{\beta}^{(t)}$ 满足条件 (10.57b)。对所有 $t \in \mathbb{T}_b$，存在 $\boldsymbol{q}^{(t)} \in \partial\left\|\boldsymbol{\beta}^{(t)}\right\|_1$ 使得

$$\partial P_t(\boldsymbol{\beta}^{(t)}) \ni \left\|\sigma_t\boldsymbol{q}^{(t)}+\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}-\sigma_t\boldsymbol{\lambda}^{(t)})\right\|_2 \leq \tau_t. \tag{10.68}$$

那么，对所有 $t \in \mathbb{T}_b$，再次利用 (1.29) 和三角不等式：

$$\begin{aligned}&\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}\right\|_2 \leq \frac{1}{\sigma_{\min}(\boldsymbol{X})}\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y})\right\|_2 \\ &\leq \frac{1}{\sigma_{\min}(\boldsymbol{X})}\left(\left\|\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}-\sigma_t\boldsymbol{\lambda}^{(t)})+\sigma_t\boldsymbol{q}^{(t)}\right\|_2+\left\|\sigma_t\boldsymbol{q}^{(t)}\right\|_2+\left\|\boldsymbol{X}^\top(\sigma_t\boldsymbol{\lambda}^{(t)})\right\|_2\right), \\ &\leq \frac{1}{\sigma_{\min}(\boldsymbol{X})}\left(\tau_t+\sigma_t\sqrt{p}+\sigma_{\max}(\boldsymbol{X})\sigma_t B_{\boldsymbol{\lambda}}\right),\end{aligned}$$

其中，最后一个不等式由 (10.68)、上界$\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \leq B_{\boldsymbol{\lambda}}$以及以下事实得出：$\|\boldsymbol{q}\|_2 \leq \sqrt{p}$ 对任意 $\boldsymbol{q} \in \partial\left\|\boldsymbol{\beta}^{(t)}\right\|_1$（见习题 2.58）。由于 $\tau_t \searrow 0$ 且 $\sigma_t \searrow 0$，令 $t \in \mathbb{T}_b \to \infty$ 时取极限，得到 $\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{y}\right\|_2 \leq 0$，这意味着 $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}$，从而保证了 ^极限 $\boldsymbol{\beta}$的可行性。

对所有 $t \in \mathbb{T}_b$， $\boldsymbol{q}^{(t)} \in \partial\left\|\boldsymbol{\beta}^{(t)}\right\|_1$，因此， $\left\|\boldsymbol{q}^{(t)}\right\|_\infty \leq 1$。于是，存在一个子序列 $\mathbb{T}_b' \subset \mathbb{T}_b$，使得 $\lim_{t\in\mathbb{T}_b'}\boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}$ 存在。容易证明 $\widehat{\boldsymbol{q}} \in \partial\left\|\widehat{\boldsymbol{\beta}}\right\|_1$。将 (10.68) 两边除以 σ_t，并对 $\boldsymbol{\lambda}^{(t)}$使用更新规则 ($\mathrm{FAL}_8$)，可得

$$\left\|\boldsymbol{q}^{(t)}-\boldsymbol{X}^\top\boldsymbol{\lambda}^{(t+1)}\right\|_2 \leq \frac{\tau_t}{\sigma_t}, \tag{10.69}$$

对所有 $t \in \mathbb{T}_b'$。由于 $\lim_{t\in\mathbb{T}_b'}\boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}$，$\lim_{t\in\mathbb{T}_b'}\frac{\tau_t}{\sigma_t} = 0$（假设（iii）），且 $\boldsymbol{X}$ 具有满行秩，因此 $\lim_{t\in\mathbb{T}_b'}\boldsymbol{\lambda}^{(t+1)} = \widehat{\boldsymbol{\lambda}}$ 存在。沿着 $\mathbb{T}_b'$对 (10.69) 两边取极限，可得

$$\widehat{\boldsymbol{q}} = \boldsymbol{X}^\top\widehat{\boldsymbol{\lambda}}. \tag{10.70}$$

结合（10.70）以及以下事实： $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}$ 和 $\widehat{\boldsymbol{q}} \in \partial\left\|\widehat{\boldsymbol{\beta}}\right\|_1$，可知^ ^最优性的 KKT 条件在 $\boldsymbol{\beta}$处成立；因此， $\boldsymbol{\beta}$ 是 ℓ_1-最小化问题的最优解

问题（P_1），由引理3.18。证明完毕。 ■

在压缩感知中，只有当优化问题 $\min \|\boldsymbol{\beta}\|_1$满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ 具有唯一解时，才能实现精确恢复。在定理 10.18中所述相同条件下，FAL 算法收敛于该唯一解。

FAL 的收敛速度

现在我们分析 FAL 的有限迭代性能，从而得到定理10.20中给出的显式收敛速度。

引理 10.19：（FAL 的迭代性质）固定一个初始点 $\boldsymbol{\beta}^{(0)} \in \mathbb{R}^p$、一个常数$\widetilde{\eta} > 0$ 使得 $\widetilde{\eta} \geq \|\boldsymbol{\beta}_*\|_1$，以及一个参数序列 $\{(\sigma_t,\ \epsilon_t,\ \tau_t)\}_{t\in\mathbb{Z}_+}$，该序列满足定理 10.18中的所有条件。此外，假设对于所有 $t \geq 1$，$\tau_t \leq c\,\epsilon_t$ 均存在某个$0 < c < 1$。令 $\{\boldsymbol{\beta}^{(t)}\}_{t\in\mathbb{Z}_+}$ 表示使用这些参数由 FAL 算法（算法 26）生成的迭代序列。那么，对于所有 $t \geq 1$，

(i) $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \leq 2B_{\boldsymbol{\lambda}}\sigma_t$。

(ii) $\left\|\boldsymbol{\beta}^{(t)}\right\|_1 - \|\boldsymbol{\beta}_*\|_1 \leq \max\left\{B_{\boldsymbol{\lambda}}^2/2 + B_1 \cdot \max\{1,\, 2cB_b\},\ \frac{1}{2}\left(\frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})} + B_{\boldsymbol{\lambda}}\right)^2\right\}\sigma_t$

注意，$B_{\boldsymbol{\lambda}} = \frac{\sqrt{2B_1\,\sigma_{\max}(\boldsymbol{X})\ B_2\}+1)}}{\sigma_{\min}(\boldsymbol{X})}$（最大值 $\{\sqrt{2B_1\,\sigma_{\max}(\boldsymbol{X})}\ B_2\}+1)$ 以及 $B_b = \widetilde{\eta} + \frac{1}{2}\sigma_1 B_{\boldsymbol{\lambda}}^2$，其定义见式 (10.64) 和式 (10.65)。

证明 [引理 10.19] **(i) 可行性误差界。** 根据定理 10.18的证明，我们已经得到以下均匀界 $\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \leq B_{\boldsymbol{\lambda}}$ 以及 $\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \leq B_b$，分别由式 (10.64) 和式 (10.65) 得出。根据界以及 FAL 算法步骤（FAL_8）（即 $\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} - (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})/\sigma_t$）中的对偶更新，可得

$$\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \leq \left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2 + \sigma_t\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 = \sigma_t\left\|\boldsymbol{\lambda}^{(t+1)}\right\|_2 + \sigma_t\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \leq 2B_{\boldsymbol{\lambda}}\sigma_t,$$

由此证明了 (i) 部分。

(ii) 目标函数次优性界。 FAL 算法的步骤 (FAL_8) 中的对偶更新还表明，增广拉格朗日函数满足

$$P_t(\boldsymbol{\beta}^{(t)}) = \sigma_t\|\boldsymbol{\beta}^{(t)}\|_1 + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y} - \sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2 = \sigma_t\left(\left\|\boldsymbol{\beta}^{(t)}\right\|_1 + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t+1)}\right\|_2^2\right).$$

对于任意 $\boldsymbol{\beta}_*^{(t)} \in$ 最小值点集合$_{\boldsymbol{\beta}\in\mathbb{R}^p} P_t(\boldsymbol{\beta})$，以及所有 $t \geq 1$，上述单位矩阵表明

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \geq \frac{P_t(\boldsymbol{\beta}_*^{(t)})}{\sigma_t} - \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t+1)}\right\|_2^2. \tag{10.71}$$

为了给出 $P_t(\boldsymbol{\beta}_*^{(t)})$的下界，考虑以下 ℓ_1-最小化问题的原始-对偶对（见命题3.17）：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}; \tag{10.72a}$$

$$\max_{\boldsymbol{w}\in\mathbb{R}^n} \boldsymbol{y}^\top\boldsymbol{w} \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{w}\right\|_\infty \leq 1. \tag{10.72b}$$

令 $\boldsymbol{w}_* \in \mathbb{R}^n$ 表示式（10.72b）中最大化问题的最优解。接下来，考虑与 ℓ_1-最小化问题的增广拉格朗日函数形式相对应的问题的原始-对偶对：

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sigma\|\boldsymbol{\beta}\|_1 + \tfrac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}-\sigma\boldsymbol{\lambda}\|_2^2; \tag{10.73a}$$

$$\max_{\boldsymbol{w}\in\mathbb{R}^n} \sigma(\boldsymbol{y}+\sigma\boldsymbol{\lambda})^\top\boldsymbol{w} - \tfrac{\sigma^2}{2}\|\boldsymbol{w}\|_2^2 \quad \text{s.t.} \quad \left\|\boldsymbol{X}^\top\boldsymbol{w}\right\|_\infty \le 1. \tag{10.73b}$$

由于 $\boldsymbol{w}_*$ 对于（10.73b）中的最大化问题也可行，因此可得

$$P_t(\boldsymbol{\beta}_*^{(t)}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left(\sigma_t\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}-\sigma_t\boldsymbol{\lambda}^{(t)}\right\|_2^2\right) \overset{\dagger}{\ge} \sigma_t\left(\boldsymbol{y}^\top\boldsymbol{w}_* + \sigma_t(\boldsymbol{\lambda}^{(t)})^\top\boldsymbol{w}_* - \frac{\sigma_t}{2}\|\boldsymbol{w}_*\|_2^2\right),$$
$$\overset{\ddagger}{\ge} \sigma_t\left(\|\boldsymbol{\beta}_*\|_1 - \sigma_t\left\|\boldsymbol{\lambda}^{(t)}\right\|_2\|\boldsymbol{w}_*\|_2 - \frac{\sigma_t}{2}\|\boldsymbol{w}_*\|_2^2\right),$$

其中，不等式（†）由原始-对偶问题（10.73）的弱对偶性得到，而不等式（‡）由原始-对偶问题（10.72）的强对偶性得到——即 $\boldsymbol{y}^\top\boldsymbol{w}_* = \|\boldsymbol{\beta}_*\|_1$（$\boldsymbol{\beta}_*$ 表示 ℓ_1-最小化问题（10.72a）的最优解）——以及柯西–施瓦茨不等式。因此，上述不等式（‡）结合（10.71）可推出

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \ge \|\boldsymbol{\beta}_*\|_1 - \frac{\sigma_t}{2}\left(2\left\|\boldsymbol{\lambda}^{(t)}\right\|_2\|\boldsymbol{w}_*\|_2 + \|\boldsymbol{w}_*\|_2^2 + \left\|\boldsymbol{\lambda}^{(t+1)}\right\|_2^2\right) \ge \|\boldsymbol{\beta}_*\|_1 - \frac{\sigma_t}{2}\left(\frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})} + B_{\boldsymbol{\lambda}}\right)^2,$$

其中，第二个不等式由以下事实得出：$\sigma_{\min}(\boldsymbol{X})\|\boldsymbol{w}_*\|_2 \le \left\|\boldsymbol{X}^\top\boldsymbol{w}_*\right\|_2 \le \sqrt{p}\left\|\boldsymbol{X}^\top\boldsymbol{w}_*\right\|_\infty \le \sqrt{p}$（参见习题1.26 和（1.29））。

现在，我们在这两个停止条件下推导相应的匹配上界。

条件 (10.57a) 下的迭代性质。假设迭代点 $\boldsymbol{\beta}^{(t)}$ 满足停止条件 (10.57a)。根据定理 10.18 的证明中的不等式 (10.66)，可得

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \|\boldsymbol{\beta}_*\|_1 + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t)}\right\|_2^2 + \frac{\epsilon_t}{\sigma_t} \le \|\boldsymbol{\beta}_*\|_1 + \sigma_t\left(\frac{1}{2}B_{\boldsymbol{\lambda}}^2 + B_1\right), \tag{10.74}$$

因为 $\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \le B_{\boldsymbol{\lambda}}$ 和 $\epsilon_t/\sigma_t^2 \le B_1$。

条件 (10.57b) 下的迭代性质。假设迭代点 $\boldsymbol{\beta}^{(t)}$ 满足停止条件 (10.57b)。令 $\boldsymbol{g}^{(t)} \in \partial P_t(\boldsymbol{\beta}^{(t)})$，使得 $\left\|\boldsymbol{g}^{(t)}\right\|_2 \le \tau_t$。那么，$P_t$ 的凸性意味着

$$P_t(\boldsymbol{\beta}^{(t)}) - P_t(\boldsymbol{\beta}_*) \le -(\boldsymbol{g}^{(t)})^\top(\boldsymbol{\beta}_* - \boldsymbol{\beta}^{(t)}) \le \left\|\boldsymbol{g}^{(t)}\right\|_2\left\|\boldsymbol{\beta}_* - \boldsymbol{\beta}^{(t)}\right\|_2 \le \tau_t\left\|\boldsymbol{\beta}_* - \boldsymbol{\beta}^{(t)}\right\|_2.$$

现在，采用与推导上界 (10.66) 类似的论证，可得

$$\left\|\boldsymbol{\beta}^{(t)}\right\|_1 \le \|\boldsymbol{\beta}_*\|_1 + \frac{\sigma_t}{2}\left\|\boldsymbol{\lambda}^{(t)}\right\|_2^2 + \frac{\tau_t}{\sigma_t}\left\|\boldsymbol{\beta}_* - \boldsymbol{\beta}^{(t)}\right\|_2 \le \|\boldsymbol{\beta}_*\|_1 + \sigma_t\left(\frac{1}{2}B_{\boldsymbol{\lambda}}^2 + 2cB_bB_1\right), \tag{10.75}$$

其中用到了以下事实：$\left\|\boldsymbol{\lambda}^{(t)}\right\|_2 \le B_{\boldsymbol{\lambda}}$，$\tau_t \le c\epsilon_t$，以及 $\epsilon_t/\sigma_t^2 \le B_1$。由 (10.74) 和 (10.75)，即可得到 (ii) 中的上界。 ■

最后，我们利用上述引理中得到的界，推导出 FAL 算法的收敛速率。

定理 10.20：（FAL 的收敛速度（Aybat 和 Iyengar，2012）） 固定一个 $0 < \alpha < 1$。那么，存在一个参数序列——并且可以显式构造该序列—— $\{\sigma_t, \epsilon_t, \tau_t\}_{t\in\mathbb{Z}_+}$，使得由 FAL 算法（算法26）生成的迭代点 ϵ是 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \leq \epsilon$可行的，即 ϵ，并且 $\left\|\boldsymbol{\beta}^{(t)}\right\|_1 - \|\boldsymbol{\beta}_*\|_1 \leq \epsilon$是$t \geq T_{\mathrm{FAL}}(\epsilon) = \mathcal{O}\left(\log_{\frac{1}{\alpha}}\left(\frac{1}{\epsilon}\right)\right)$。

定理 [的证明 10.20] 我们先对问题进行缩放。定义 $(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{y}}) \triangleq \frac{1}{\sigma_{\max}(\boldsymbol{X})}(\boldsymbol{X}, \boldsymbol{y})$。那么，对于重新缩放的问题，光滑度参数为 $L_b = \sigma_{\max}(\boldsymbol{X}) = 1$，而条件数保持不变：$\kappa(\boldsymbol{X}) = \kappa(\boldsymbol{X}) = \sigma_{\max}(\boldsymbol{X})/\sigma_{\min}(\boldsymbol{X})$。我们将 FAL 算法应用于重新缩放后的实例 $(\boldsymbol{X}, \widetilde{\boldsymbol{y}})$。

将初始参数设为 $\sigma_1 \triangleq 1$, $\epsilon_1 \triangleq 2$，并递归定义$t \geq 1$ 的序列如下：

$$\sigma_{t+1} = \alpha \cdot \sigma_t, \qquad \epsilon_{t+1} = \alpha^2 \cdot \epsilon_t, \qquad \tau_t = \frac{\epsilon_t}{2\max\left\{1, \widetilde{\eta} + \frac{9p}{2}\kappa(\boldsymbol{X})^2\right\}}.$$

采用此选择后，引理 10.19 中出现的常数可简化如下

$$B_1 = \max_{t\geq 1}\left\{\frac{\epsilon_t}{\sigma_t^2}\right\} = \frac{\epsilon_1}{\sigma_1^2} = 2, \qquad B_2 = \max_{t\geq 1}\left\{\frac{\tau_t}{\sigma_t}\right\} = \frac{\alpha^{t-1}}{\max\left\{1, \widetilde{\eta} + \frac{9p}{2}\kappa(\boldsymbol{X})^2\right\}} \leq 1.$$

因此，$\left\|\boldsymbol{\lambda}^{(t)}\right\|_2$ 和 $\left\|\boldsymbol{\beta}^{(t)}\right\|_1$ 的均匀上界为

$$B_{\boldsymbol{\lambda}} = \frac{\sqrt{p}}{\sigma_{\min}(\boldsymbol{X})}\left(\max\left\{\sqrt{2B_1}\,\sigma_{\max}(\boldsymbol{X}),\ B_2\right\} + 1\right) \leq 3\kappa(\boldsymbol{X})\sqrt{p};$$
$$B_b = \widetilde{\eta} + \frac{1}{2}\sigma_1 B_{\boldsymbol{\lambda}}^2 \leq \widetilde{\eta} + \frac{9p}{2}\kappa(\boldsymbol{X})^2.$$

对于重新缩放的问题（$\boldsymbol{X}$，$\widetilde{\boldsymbol{y}}$），引理 10.19 保证对所有 $t \geq 1$,

$$\left|\left\|\boldsymbol{\beta}^{(t)}\right\|_1 - \|\boldsymbol{\beta}_*\|_1\right| \leq \max\left\{\left(\frac{B_{\boldsymbol{\lambda}}^2}{2} + B_1\right), \frac{(\sqrt{p}\kappa(\boldsymbol{X}) + B_{\boldsymbol{\lambda}})^2}{2}\right\}\sigma_1\alpha^{t-1} \leq 8p\kappa(\boldsymbol{X})^2\alpha^{t-1},$$

这里我们利用了这样一个事实：$\kappa(\boldsymbol{X}) \geq 1$。因此，对于所有满足以下条件的 $t \in \mathbb{Z}_+$，有 $\left\|\boldsymbol{\beta}^{(t)}\right\|_1 - \|\boldsymbol{\beta}_*\|_1 \leq \epsilon$

$$t \geq \log_{\frac{1}{\alpha}}\left(\frac{8p\kappa(\boldsymbol{X})^2}{\epsilon}\right) + 1. \tag{10.76}$$

根据引理 10.19，我们还得到对所有 $t \geq 1$,

$$\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \leq 2B_{\boldsymbol{\lambda}}\sigma_t = 2B_{\boldsymbol{\lambda}}\sigma_1\alpha^{t-1} \leq 6\sqrt{p}\kappa(\boldsymbol{X})\alpha^{t-1}.$$

因此，对所有 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \leq \epsilon$

$$t \geq \log_{\frac{1}{\alpha}}\left(\frac{6\sqrt{p}\kappa(\boldsymbol{X})}{\epsilon}\right) + 1. \tag{10.77}$$

结合（10.76）和（10.77）可知，对所有 $\epsilon > 0$，$T_{\mathrm{FAL}}(\epsilon)$，计算一个 ϵ-可行且 ϵ-最优解所需的 FAL 迭代次数至多为

$$T_{\mathrm{FAL}}(\epsilon) \leq \left\lceil \log_{\frac{1}{\alpha}}\left(\frac{8p\kappa(\boldsymbol{X})^2}{\epsilon}\right)\right\rceil + 1. \tag{10.78}$$

至此，证明完成。 ■

10.2.4 布雷格曼迭代算法

我们将使用第8章介绍的原始-对偶框架，提出一种统一算法，用于求解本书讨论的四个优化问题（即（P_1）、（$P1_\epsilon$）、（L_L）和（L_C）。在这些模型中，ℓ_1-最小化形式是 ℓ_0-范数最小化问题的凸松弛，而约束LASSO模型则提供了一种与基追踪去噪模型等价的线性预测表示。

现在考虑这四个优化问题的一般形式（参见第8.3节）：

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{J(\boldsymbol{\beta}) + \sigma F(\boldsymbol{\beta})\}, \tag{10.79}$$

其中 $J: \mathbb{S} \to \mathbb{R}$ 和 $H: \mathbb{S} \to \mathbb{R}$ 是定义在闭凸集 $\mathbb{S} \subseteq \mathbb{R}^p$上的非负凸函数。函数 J 通常不可微，而 F 可微。

求解问题（10.79）的一种著名迭代方法是*Bregman*迭代，它基于 Bregman距离（定义2.87）：

$$D_\phi^{\boldsymbol{g}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \phi(\boldsymbol{\alpha}) - \phi(\boldsymbol{\beta}) - \langle \boldsymbol{g}, \boldsymbol{\alpha} - \boldsymbol{\beta} \rangle, \tag{10.80}$$

其中，$\phi(\boldsymbol{\alpha})$: $\mathbb{S} \subseteq \mathbb{R}^p \to \mathbb{R}$ 是一个（可能非光滑的）凸函数，而 $\boldsymbol{g} \in \partial\phi(\boldsymbol{\beta})$是 ϕ 在点 $\boldsymbol{\beta} \in \mathbb{R}^p$处的次梯度向量。注意，Bregman 距离并非经典意义上的真正度量，因为它通常不对称：$D_\phi^{\boldsymbol{g}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq D_\phi^{\boldsymbol{h}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$。

考虑非光滑函数 $J(\boldsymbol{\beta})$在第t个迭代点 $\boldsymbol{\beta}^{(t)}$处的一阶泰勒近似，记为 $J(\boldsymbol{\beta}) = J(\boldsymbol{\beta}^{(t)}) + \langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \rangle$，其中 $\boldsymbol{g}^{(t)} \in \partial J(\boldsymbol{\beta}^{(t)})$。近似误差可由 Bregman 距离精确度量：

$$D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) = J(\boldsymbol{\beta}) - J(\boldsymbol{\beta}^{(t)}) - \left\langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle. \tag{10.81}$$

▶ **Bregman 迭代算法**。早在 1965年，Bregman 就提出将无约束问题（10.79）改写为以下迭代方案（Bregman，1965年）：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}} \left\{ D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + \sigma F(\boldsymbol{\beta}) \right\}, \tag{10.82}$$

$$\equiv \arg\min_{\boldsymbol{\beta}} \left\{ J(\boldsymbol{\beta}) - \left\langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle + \sigma F(\boldsymbol{\beta}) \right\}. \tag{10.83}$$

这一过程称为*Bregman* 迭代算法 或 *Bregman* 迭代正则化。其核心思想类似于近端梯度法，而后者本身建立在投影梯度法的基础上（参见第 9.2.2节）。下文将讨论 Bregman 迭代算法及其推广。

At 第 t 次迭代，将 Bregman 子问题的目标函数定义为

$$L(\boldsymbol{\beta}) = J(\boldsymbol{\beta}) - \langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \rangle + \sigma F(\boldsymbol{\beta}).$$

一阶最优性条件 $\mathbf{0} \in \partial L(\boldsymbol{\beta})$ 蕴含 $\mathbf{0} \in \partial J(\boldsymbol{\beta}) - \boldsymbol{g}^{(t)} + \sigma \nabla F(\boldsymbol{\beta})$。因此，在第 $(t+1)$ 次迭代点 $\boldsymbol{\beta}^{(t+1)}$ 处，我们得到次梯度的更新规则：

$$\boldsymbol{g}^{(t+1)} \leftarrow \boldsymbol{g}^{(t)} - \sigma \nabla F(\boldsymbol{\beta}^{(t+1)}) \qquad \text{and} \qquad \boldsymbol{g}^{(t+1)} \in \partial J(\boldsymbol{\beta}^{(t+1)}). \tag{10.84}$$

方程 (10.83) 和 (10.84) 共同构成了*Bregman* 迭代算法，最初由Osher 等人 (2005年) 在图像处理背景下提出。完整过程总结于算法28中。

现在我们回顾 Bregman 迭代算法的收敛性质，陈述如下，不再证明。

算法 28 Bregman 迭代算法（Osher 等人，2005年）

Ensure: Approximate the solution of $\min\{J(\boldsymbol{\beta})+\sigma F(\boldsymbol{\beta})\}$;

1: **initialize:** $\boldsymbol{\beta}^{(0)}=\mathbf{0}$, $\boldsymbol{g}^{(0)}=\mathbf{0}$;
2: **for** $t=0,1,2,\ldots$ **do**
3: $\quad\boldsymbol{\beta}^{(t+1)}\leftarrow\arg\min_{\boldsymbol{\beta}}\left\{D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)})+\sigma F(\boldsymbol{\beta})\right\}$; ▷ (BI_1)
4: $\quad\boldsymbol{g}^{(t+1)}\leftarrow\boldsymbol{g}^{(t)}-\sigma\nabla F(\boldsymbol{\beta}^{(t+1)})$; ▷ (BI_2)
5: $\quad$Exit if $\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}\right\|_2\le\varepsilon$;
6: **end for**
7: **return** $\boldsymbol{\beta}\leftarrow\boldsymbol{\beta}^{(t)}$.

定理 10.21：（Bregman迭代算法的收敛性（Osher 等人，2005年）） 假设 J 和 F 是凸函数，并且 F 可微。如果在每次迭代中方程（10.83）都存在解，则以下结论成立：

(i) 单调性。序列 $\{F(\boldsymbol{\beta}^{(t)})\}$ 单调不增。具体而言，$\sigma F(\boldsymbol{\beta}^{(t+1)})\le\sigma F(\boldsymbol{\beta}^{(t+1)})+D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}^{(t+1)},\boldsymbol{\beta}^{(t)})\le\sigma F(\boldsymbol{\beta}^{(t)})$。(ii) 函数 F将收敛到最优解 $F(\boldsymbol{\beta}^*)$，其速率估计为 $F(\boldsymbol{\beta}^{(t)})\le F(\boldsymbol{\beta}^*)+J(\boldsymbol{\beta}^*)/t$。

最小二乘目标函数的 Bregman 迭代算法

当 $\sigma F(\boldsymbol{\beta})=\frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2$ 时，Bregman 迭代算法有两种等价表述：

(1) 初始化 $\boldsymbol{\beta}^{(0)}=\mathbf{0}$、$\boldsymbol{g}^{(0)}=\mathbf{0}$；迭代过程如下

$$\boldsymbol{\beta}^{(t+1)}\leftarrow\arg\min_{\boldsymbol{\beta}}\left\{D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta},\boldsymbol{\beta}^{(t)})+\frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2\right\};\tag{10.85a}$$

$$\boldsymbol{g}^{(t+1)}\leftarrow\boldsymbol{g}^{(t)}-\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}-\boldsymbol{y}).\tag{10.85b}$$

(2) 初始化 $\boldsymbol{y}^{(0)}=\mathbf{0}$、$\boldsymbol{\beta}^{(0)}=\mathbf{0}$；迭代过程如下

$$\boldsymbol{y}^{(t+1)}\leftarrow\boldsymbol{y}+(\boldsymbol{y}^{(t)}-\boldsymbol{X}\boldsymbol{\beta}^{(t)});\tag{10.86a}$$

$$\boldsymbol{\beta}^{(t+1)}\leftarrow\arg\min_{\boldsymbol{\beta}}\left\{J(\boldsymbol{\beta})+\frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}^{(t+1)}\right\|_2^2\right\}.\tag{10.86b}$$

Yin 等人已证明Yin 等人 (2008年) 这两个版本在数学上是等价的。现在我们将说明，对最小二乘目标应用 Bregman 迭代算法，可以有效地近似 ℓ_1-最小化问题（P_1）：

$$(\mathrm{P}_1)\qquad\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\mu\|\boldsymbol{\beta}\|_1\quad\text{s.t.}\quad\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{y},\tag{10.87}$$

其中 μ 是一个正数，我们此前使用过 $\mu=1$。我们证明，Bregman迭代（10.85）（或等价地，（10.86））会生成一个解序列 $\{\boldsymbol{\beta}^{(t)}\}$，该序列在有限步内收敛到 $\boldsymbol{\beta}^*$ -最小化问题（ ℓ_1 10.87）的一个最优解；也就是说，存在一个 T，使得每个 $\boldsymbol{\beta}^{(t)}$ 在 $t>T$ 时，都是（10.87）的解。

定理 10.22：（近似 ℓ_1-最小化问题（Yin 等人，2008年）考虑 Bregman 迭代算法中（10.85）或（10.86）中的迭代，其中 $J(\boldsymbol{\beta}) = \mu\|\boldsymbol{\beta}\|_1$ 且 $\sigma F(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$。那么：

(a) 假设一个迭代点 $\boldsymbol{\beta}^{(t)}$ 来自（10.85a）且满足 $\boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{y}$；那么 $\boldsymbol{\beta}^{(t)}$ 是该 ℓ_1-最小化问题（10.87）的一个解。(b) 存在有限整数 $T < \infty$，使得每个满足 $t \geq T$ 的迭代点 $\boldsymbol{\beta}^{(t)}$ 都是该 ℓ_1-最小化问题（10.87）的最优解。

证明 [定理 10.22] **(a)**。由于 $\boldsymbol{g}^{(t)} \in \partial J(\boldsymbol{\beta}^{(t)})$，对于任意 $\boldsymbol{\beta}$，由 Bregman 距离的非负性（注2.88），可得

$$J(\boldsymbol{\beta}^{(t)}) \leq J(\boldsymbol{\beta}) - \left\langle \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}, \boldsymbol{g}^{(t)} \right\rangle = J(\boldsymbol{\beta}) - \left\langle \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}, \boldsymbol{X}^\top(\boldsymbol{y}^{(t)} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}) \right\rangle \tag{10.88}$$

$$= J(\boldsymbol{\beta}) - \left\langle \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}^{(t)}, \boldsymbol{y}^{(t)} - \boldsymbol{X}\boldsymbol{\beta}^{(t)} \right\rangle = J(\boldsymbol{\beta}) - \left\langle \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}, \boldsymbol{y}^{(t)} - \boldsymbol{y} \right\rangle, \tag{10.89}$$

其中，第二个等式由表述（10.85）与（10.86）的等价性得出；参见问题 10.6和 Yin 等人（2008年）。因此，$\boldsymbol{\beta}^{(t)}$满足 $J(\boldsymbol{\beta}^{(t)}) \leq J(\boldsymbol{\beta})$，对任意满足 $\boldsymbol{\beta}$ 的 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$；因此，$\boldsymbol{\beta}^{(t)}$ 是该 ℓ_1-最小化问题（10.87）的最优解。

(b). 设 $(\mathbb{I}_+^{(j)}, \mathbb{I}_-^{(j)}, \mathbb{E}^{(j)})$ 是指标集 $\{1, 2, \ldots, p\}$ 的一个划分，并定义 e

$$\mathbb{U}^{(j)} \triangleq \mathbb{U}(\mathbb{I}_+^{(j)}, \mathbb{I}_-^{(j)}, \mathbb{E}^{(j)}) = \{\boldsymbol{\beta} \mid \beta_i \geq 0,\ i \in \mathbb{I}_+^{(j)};\ \beta_i \leq 0,\ i \in \mathbb{I}_-^{(j)};\ \beta_i = 0,\ i \in \mathbb{E}^{(j)}\}; \tag{10.90}$$

$$\mathbb{H}^{(j)} \triangleq \min_{\boldsymbol{\beta}} \left\{ \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 : \boldsymbol{\beta} \in \mathbb{U}^{(j)} \right\}. \tag{10.91}$$

这样的划分只有有限多个，$(\mathbb{I}_+^{(j)}, \mathbb{I}_-^{(j)}, \mathbb{E}^{(j)})$，且所有可能的 $\mathbb{U}^{(j)}$ 的并集为 $\mathcal{H}$，即 $\boldsymbol{\beta} \in \mathbb{R}^p$ 的整个空间。

在第 t 次迭代时，按如下方式根据 $\boldsymbol{g}^{(t)}$ 定义 $(\mathbb{I}_+^{(t)}, \mathbb{I}_-^{(t)}, \mathbb{E}^{(t)})$:

$$\mathbb{I}_+^{(t)} = \{i \mid g_i^{(t)} = \mu\},\ \mathbb{I}_-^{(t)} = \{i \mid g_i^{(t)} = -\mu\},\ \mathbb{E}^{(t)} = \{i \mid g_i^{(t)} \in (-\mu, \mu)\}. \tag{10.92}$$

根据定义 (10.90)，以及由习题 2.58 可知 $\boldsymbol{g}^{(t)} \in \partial J(\boldsymbol{\beta}^{(t)}) = \partial(\mu\|\boldsymbol{\beta}^{(t)}\|_1)$ 且 μ符号函数$(\boldsymbol{\beta}^{(t)}) \in \partial(\mu\|\boldsymbol{\beta}^{(t)}\|_1)$ 这一事实，我们得到 $\boldsymbol{\beta}^{(t)} \in \mathbb{U}^{(t)}$。为应用定理10.21，令 $\boldsymbol{\beta}^*$ 满足 $F(\boldsymbol{\beta}^*) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta}^* - \boldsymbol{y}\|_2^2 = 0$。在定理10.21 (ii) 中使用这个 $\boldsymbol{\beta}^*$，可知对每个满足 $\mathbb{H}^{(j)} > 0$ 的 j，都存在充分大的 T_j，使得 $\boldsymbol{\beta}^{(t)}$ 不属于 $\mathbb{U}^{(j)}$，其中 $t \geq T_j$。因此，令 $T \triangleq \max_j\{T_j : \mathbb{H}^{(j)} > 0\}$，我们有 $F(\boldsymbol{\beta}^{(t)}) = 0$，其中 $t \geq T$。也就是说，$\boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{y}$，其中 $t \geq T$。

因此，由（10.85b），次梯度趋于稳定：$\boldsymbol{g}^{(T-1)} = \boldsymbol{g}^{(T)} = \ldots$，而（10.86a）意味着 $\boldsymbol{y}^{(T)} = \boldsymbol{y}^{(T+1)} = \ldots$。由于 Bregman迭代（10.85a）和（10.86b）的极小化点不一定唯一，迭代点 $\boldsymbol{\beta}^{(t)}$ 对于 $t \geq T$ 不一定相同。然而，由（a）可知，所有 $\boldsymbol{\beta}^{(t)}$ 对于 $t \geq T$ 都是该 ℓ_1-最小化问题（10.87）的最优解。 ■

线性化 Bregman 迭代算法

Bregman 迭代算法为求解结构化优化问题提供了一个有效的框架。然而，在每次迭代中，它都需要最小化目标函数

$D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + \sigma F(\boldsymbol{\beta})$，当 F是非线性或非二次函数时，计算开销可能很大。为提高计算效率，Yin 等人（2008年）提出了线性化 *Bregman* 迭代算法。

关键思路是利用当前迭代点 $F(\boldsymbol{\beta})$ 处的一阶泰勒展开来近似（可能非线性的）函数 $\boldsymbol{\beta}^{(t)}$ (定理2.13):

$$F(\boldsymbol{\beta}) \approx F(\boldsymbol{\beta}^{(t)}) + \left\langle \nabla F(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle.$$

于是，优化问题（10.79）在 $\sigma = 1$ 下变为

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}} \left\{ D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + F(\boldsymbol{\beta}^{(t)}) + \left\langle \nabla F(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle \right\}. \tag{10.93}$$

注意，这种线性近似仅在 $\boldsymbol{\beta}^{(t)}$附近准确，而常数项$F(\boldsymbol{\beta}^{(t)})$不会影响关于 $\boldsymbol{\beta}$的极小化解。不过，为确保稳定性和收敛性——尤其当 F 不是强凸函数时，需要添加二次邻近项。这便得到以下正则化子问题:

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta}} \left\{ D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + \left\langle \nabla F(\boldsymbol{\beta}^{(t)}), \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle + \frac{1}{2\eta} \left\| \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\|_2^2 \right\} \tag{10.94a}$$

$$\equiv \arg\min_{\boldsymbol{\beta}} \left\{ D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + \frac{1}{2\eta} \left\| \boldsymbol{\beta} - (\boldsymbol{\beta}^{(t)} - \eta \nabla F(\boldsymbol{\beta}^{(t)})) \right\|_2^2 \right\}, \tag{10.94b}$$

其中 η 为正数，作为步长。等价性成立是因为（10.94a）和（10.94b）仅相差一个与 $\boldsymbol{\beta}$无关的常数项。Darbon 和 Osher 在Darbon 和 Osher (2007年) 中使用了这一更新规则来求解图像去模糊/反卷积问题 $\min_{\boldsymbol{\beta}}\{TV(\boldsymbol{\beta}) : \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}\}$。

▶ **与近端梯度更新和 Bregman 迭代更新的联系**。更新（10.94b）类似于近端梯度更新（9.31b），但也因采用正则化而有所不同。近端梯度更新（9.31b）最小化 J，而（10.94b）最小化基于 J的 Bregman 距离 $D_J^{\boldsymbol{g}^{(t)}}$ 。另一方面，（10.94b）通过将保真项 $\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2/2$ 替换为其在 $\boldsymbol{\beta}^{(t)}$ 处的一阶近似与在 $\boldsymbol{\beta}^{(t)}$ 处的 ℓ_2-近端项之和，从而区别于（$\sigma F(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$。

将线性化 Bregman 迭代应用于最小二乘目标函数

特别地，当 $\sigma F(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$时，我们有 $\nabla F(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})$ 。将其代入（10.94b）可得

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \text{最小值点集合}_{\boldsymbol{\beta}} \left\{ D_J^{\boldsymbol{g}^{(t)}}(\boldsymbol{\beta}, \boldsymbol{\beta}^{(t)}) + \frac{1}{2\eta} \left\| \boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}) \right) \right\|_2^2 \right\}. \tag{10.95}$$

考虑式（10.95）中的目标函数，它可以写为

$$L(\boldsymbol{\beta}) = J(\boldsymbol{\beta}) - J(\boldsymbol{\beta}^{(t)}) - \left\langle \boldsymbol{g}^{(t)}, \boldsymbol{\beta} - \boldsymbol{\beta}^{(t)} \right\rangle + \frac{1}{2\eta} \left\| \boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}) \right) \right\|_2^2.$$

根据次微分最优性条件 $\mathbf{0} \in \partial L(\boldsymbol{\beta})$，可得

$$\mathbf{0} \in \partial J(\boldsymbol{\beta}) - \boldsymbol{g}^{(t)} + \frac{1}{\eta} \left(\boldsymbol{\beta} - \left(\boldsymbol{\beta}^{(t)} - \eta \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}) \right) \right) \tag{10.96}$$

$$\Longrightarrow \boldsymbol{g}^{(t+1)} - \boldsymbol{g}^{(t)} = -\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}) - \frac{\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}}{\eta}, \tag{10.97}$$

其中 $\boldsymbol{g}^{(t+1)} \in \partial J\left(\boldsymbol{\beta}^{(t+1)}\right)$。给定初始化 $\boldsymbol{\beta}^{(0)} = \boldsymbol{g}^{(0)} = \mathbf{0}$，对迭代 0 至 t中的（10.97）进行叠加，可得次梯度的累积更新：

$$\boldsymbol{g}^{(t+1)} = \sum_{i=1}^{t} \boldsymbol{X}^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(i)}) - \frac{\boldsymbol{\beta}^{(t+1)}}{\eta} \triangleq \boldsymbol{u}^{(t)} - \frac{\boldsymbol{\beta}^{(t+1)}}{\eta} \tag{10.98}$$

其中 $\boldsymbol{u}^{(t)} \triangleq \sum_{i=1}^{t} \boldsymbol{X}^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(i)})$。

由此，我们推导出两个关键的迭代公式。首先，重新整理（10.98）得到 $\boldsymbol{\beta}$ 的更新公式，而 $\boldsymbol{u}^{(t+1)}$ 的定义给出了 $\boldsymbol{u}$ 的更新公式：

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \eta(\boldsymbol{u}^{(t)} - \boldsymbol{g}^{(t+1)}); \tag{10.99a}$$

$$\boldsymbol{u}^{(t+1)} \leftarrow \boldsymbol{u}^{(t)} + \boldsymbol{X}^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(t+1)}). \tag{10.99b}$$

该方案构成了用于求解复合优化问题的线性化 *Bregman* 迭代算法

$$\min_{\boldsymbol{\beta}} \left\{ J(\boldsymbol{\beta}) + \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2^2 \right\}. \tag{10.100}$$

该算法极其简单且高效：每次迭代只涉及矩阵-向量乘法以及一个（可能很简单的）次梯度或收缩操作，因此非常适合大规模问题。

根据定理 10.22，（线性化）Bregman 迭代能否成功近似 ℓ_1-最小化问题，取决于能否使残差趋于零 $\boldsymbol{X}\boldsymbol{\beta}^{(t)} = \boldsymbol{y}$，即最小二乘目标函数 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2 \searrow 0$。下面的结果给出了线性化变体的收敛保证。

定理 10.23：（线性 Bregman 迭代的收敛速度，适用于 ℓ_1-最小化（Darbon 和 Osher，2007年；Yin 等人，2008年）） 考虑将线性 Bregman 迭代算法（10.99）应用于最小二乘目标函数 $\sigma F(\boldsymbol{\beta}) = \frac{1}{2}\left\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right\|_2^2$。

(i) 若 $\eta \boldsymbol{X}^{\top}\boldsymbol{X} \prec \boldsymbol{I}$（即，$\boldsymbol{I} - \eta \boldsymbol{X}^{\top}\boldsymbol{X}$ 为严格正定），则残差范数 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2$ 单调递减：

$$\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} - \boldsymbol{y}\right\|_2^2 + \left(\frac{1}{\eta} - \left\|\boldsymbol{X}^{\top}\boldsymbol{X}\right\|_2\right)\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right\|_2^2 \leq \left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2^2.$$

(ii) 进一步假设 J 严格凸，并令 $\boldsymbol{\beta}^*$ 为 $\min\{J(\boldsymbol{\beta}) : \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}\}$ 的一个最优解。若 $\boldsymbol{\beta}^{(t)} \to \boldsymbol{w}$，则

$$J(\boldsymbol{w}) \leq J(\boldsymbol{\beta}^*) + \frac{1}{\eta}\left\langle \boldsymbol{\beta}^* - \boldsymbol{w}, \boldsymbol{w} \right\rangle. \tag{10.101}$$

(iii) 若 J严格凸且 $\eta \boldsymbol{X}\boldsymbol{X}^{\top} \prec 2\boldsymbol{I}$（即，$\boldsymbol{I} - \frac{\eta}{2}\boldsymbol{X}\boldsymbol{X}^{\top}$ 为严格正定），则残差范数 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2$ 随 t呈指数衰减。

证明 [定理 10。23] **(i)**。由于（10.95）中使用的 Bregman 距离非负，因此我们有

$$\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)} + \eta \boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\right\|_2^2 \leq \left\|\eta \boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\right\|_2^2.$$

展开等式两边可得

$$\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right\|_2^2 + 2\eta\left\langle \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}, \boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y})\right\rangle \leq 0.$$

重新排列后，可得

$$\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}\right\|_2^2+\eta\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}-\boldsymbol{y}\right\|_2^2-\eta\left\langle\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)},\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})\right\rangle\leq\eta\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}\right\|_2^2.$$

利用 (1.29)，可得

$$\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}-\boldsymbol{y}\right\|_2^2+\left(\frac{1}{\eta}\boldsymbol{I}-\left\|\boldsymbol{X}^\top\boldsymbol{X}\right\|_2\right)\left\|\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}\right\|_2^2\leq\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}\right\|_2^2.$$

(ii). 再次令 $\boldsymbol{g}^{(t)}\in\partial J(\boldsymbol{\beta}^{(t)})$。我们有

$$\boldsymbol{\beta}^{(t+1)}=\arg\min_{\boldsymbol{\beta}}\left\{J(\boldsymbol{\beta})-J(\boldsymbol{\beta}^{(t)})-\left\langle\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)},\boldsymbol{g}^{(t)}\right\rangle+\frac{1}{2\eta}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}+\eta\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y})\right\|_2^2\right\},$$

由 (10.99)，可得

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)}=\arg\min_{\boldsymbol{\beta}}\Big\{&J(\boldsymbol{\beta})-J(\boldsymbol{\beta}^{(t)})-\left\langle\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)},\boldsymbol{u}^{(t-1)}\right\rangle\\&+\frac{1}{\eta}\left\langle\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)},\boldsymbol{\beta}^{(t)}\right\rangle+\frac{1}{2\eta}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}+\eta\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y})\right\|_2^2\Big\}.\end{aligned}$$

由 Bregman 距离的非负性及 $\boldsymbol{u}^{(t-1)}\triangleq\boldsymbol{X}^\top\sum_{i=1}^{t-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(i)})$ 的定义，可得

$$\begin{aligned}J(\boldsymbol{\beta}^{(t)})&\leq J(\boldsymbol{\beta}^*)-\left\langle\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)},\boldsymbol{g}^{(t)}\right\rangle=J(\boldsymbol{\beta}^*)-\left\langle\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)},\boldsymbol{u}^{(t-1)}\right\rangle+\frac{1}{\eta}\left\langle\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)},\boldsymbol{\beta}^{(t)}\right\rangle\\&=J(\boldsymbol{\beta}^*)-\left\langle\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(t)},\sum_{i=0}^{t-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(i)})\right\rangle+\frac{1}{\eta}\left\langle\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(t)},\boldsymbol{\beta}^{(t)}\right\rangle.\end{aligned}$$

我们将证明 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(i)}-\boldsymbol{y}\right\|_2$ 随 i 指数衰减；那么上式右端的中间项将趋于零，从而得到所需结果。

(iii). 由 J的严格凸性，存在一个对称正定算子 $\boldsymbol{Q}^{(t)}$（习题2.35），使得 $\boldsymbol{g}^{(t+1)}-\boldsymbol{g}^{(t)}=\boldsymbol{Q}^{(t)}(\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)})$，且（10.97）变为

$$\boldsymbol{\beta}^{(t+1)}-\boldsymbol{\beta}^{(t)}=-\left(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I}\right)^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}),$$

这意味着

$$\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}-\boldsymbol{y}=\left(\boldsymbol{I}-\boldsymbol{X}\left(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I}\right)^{-1}\boldsymbol{X}^\top\right)(\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}).$$

要使 t 中 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)}-\boldsymbol{y}\right\|_2$呈指数衰减，我们要求矩阵 $\left(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I})^{-1}\boldsymbol{X}^\top\right)$ 的谱半径严格小于 1。当$\boldsymbol{X}(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I})^{-1}\boldsymbol{X}^\top$ 的所有特征值都位于 (0，2) 内时，这一条件成立。下界由 $\boldsymbol{X}(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I})^{-1}\boldsymbol{X}^\top$的严格正定性得出。要得到上界，我们注意到 $\boldsymbol{X}(\boldsymbol{Q}^{(t)}+\frac{1}{\eta}\boldsymbol{I})^{-1}\boldsymbol{X}^\top\prec\eta\boldsymbol{X}\boldsymbol{X}^\top$，这由 Sherman–Morrison–Woodbury公式得出；因此，只需满足 $\eta\boldsymbol{X}\boldsymbol{X}^\top\prec 2\boldsymbol{I}$。证明完毕。

■

最后，假设对于某个基函数 j，有 $J(\boldsymbol{\beta}) \triangleq \mu \cdot j(\boldsymbol{\beta})$。请注意，即使 η 满足条件 $\boldsymbol{I} - \frac{\eta}{2}\boldsymbol{X}\boldsymbol{X}^\top \succ 0$，$t$ 中 $\left\|\boldsymbol{X}\boldsymbol{\beta}^{(t)} - \boldsymbol{y}\right\|_2$ 的指数衰减速率仍取决于 μ 的值。由于 $\boldsymbol{Q}^{(t)}$ 关于 μ 是线性的（假设 $\boldsymbol{\beta}^{(t)}$ 和 $\boldsymbol{\beta}^{(t+1)}$ 固定），我们令 $\boldsymbol{Q}^{(t)} \triangleq \mu\boldsymbol{Q}^{(t)}$。当 $\mu \gg \frac{1}{\eta}$ 时，$(\boldsymbol{Q}^{(t)} + \frac{1}{\eta}\boldsymbol{I})^{-1}$ 由 $\frac{1}{\mu}(\boldsymbol{Q}^{(t)})^{-1}$ 主导，因此 $\boldsymbol{X}(\boldsymbol{Q}^{(t)} + \frac{1}{\eta}\boldsymbol{I})^{-1}\boldsymbol{X}^\top$ 的最小特征值在 μ 中线性衰减至 0。因此，尽管残差仍呈指数收敛，但随着 μ 增大，收敛速率会变慢。在实践中，这意味着我们可以通过调节参数 μ 来控制收敛速度。

当 $J(\boldsymbol{\beta}) = \mu\|\boldsymbol{\beta}\|_1$，取 μ 足够大的值会促使极限点 $\boldsymbol{w} = \lim_{t\to\infty}\boldsymbol{\beta}^{(t)}$ 趋近于以下约束 ℓ_1-最小化问题的一个解：$\min\|\boldsymbol{\beta}\|_1$ 满足 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$。尽管 $J(\boldsymbol{\beta}) = \mu\|\boldsymbol{\beta}\|_1$ 不是严格凸的，但可以通过严格凸扰动（例如，添加一个小的 ℓ_2 项）任意精确地近似，详见 Yin 等人 (2008年)。

总而言之，对于求解 ℓ_1-最小化问题（10.87），该问题传统上被表述并求解为一个约束线性规划，我们证明，将简单的 Bregman 迭代方案应用于其无约束拉格朗日松弛 $\mu\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$，可在有限次迭代内得到精确解。在实践中，采用适中的惩罚参数 μ 通常能使大多数问题实例仅经过几次迭代便收敛，因此该方法既高效又易于实现。

第10章 问题

1. **RIP下IHT的收敛性（Foucart 和 Rauhut，2013年）**。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 $3k$、常数为

$$\delta_{3k} < \frac{1}{2}.$$

对于每个 k-稀疏向量 $\boldsymbol{\beta}^* \in \mathbb{R}^p$，证明由IHT（算法22）根据测量值 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^*$ 定义的序列 $\{\boldsymbol{\beta}^{(t)}\}_{t\geq 0}$ 收敛到 $\boldsymbol{\beta}^*$。

2. **RIP下IHT和HTP的稳定恢复（Foucart 和 Rauhut，2013年）**。设 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ 满足阶为 $3k$、常数为

$$\delta_{3k} < \frac{1}{\sqrt{3}}.$$

对于 $\boldsymbol{\beta}^* \in \mathbb{R}^p$，$\boldsymbol{\epsilon} \in \mathbb{R}^n$ 以及 $\mathbb{S} \subseteq \{1, 2, \ldots, p\}$ 满足 $|\mathbb{S}| = k$ 的 $\{\boldsymbol{\beta}^{(t)}\}_{t\geq 0}$ 由IHT（算法22）或HTP（算法23）生成的序列 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$ 满足

$$\left\|\boldsymbol{\beta}_{\mathbb{S}}^{(t)} - \boldsymbol{\beta}_{\mathbb{S}}^*\right\|_2 \leq \mu^t\left\|\boldsymbol{\beta}_{\mathbb{S}}^{(0)} - \boldsymbol{\beta}_{\mathbb{S}}^*\right\|_2 + \nu\|\boldsymbol{X}_{\mathbb{S}^c}\boldsymbol{\beta}_{\mathbb{S}^c} + \boldsymbol{\epsilon}\|_2, \quad \forall t \geq 0,$$

其中，对于迭代硬阈值，$\mu = \sqrt{3}\,\delta_{3k} < 1$，$\nu \leq 2.18/(1-\mu)$；对于硬阈值追踪，$\mu = \sqrt{2\delta_{3k}^2/(1-\delta_{3k}^2)} < 1$，$\nu \leq 5.15/(1-\mu)$。

3. **收敛**。设 $F_t : \mathbb{S} \subset \mathbb{R}^p \to \mathbb{R}$ 是一个连续函数递减序列，逐点收敛到函数 $F : \mathbb{S} \to \mathbb{R}$。也就是说，对于每个 $\boldsymbol{\beta} \in \mathbb{S}$，$F_{t+1}(\boldsymbol{\beta}) \leq F_t(\boldsymbol{\beta})$，且 $\lim_{t\to\infty} F_t(\boldsymbol{\beta}) = F(\boldsymbol{\beta})$。假设 F 在 $\mathbb{S}$ 上连续且强制。令 $\boldsymbol{\beta}^{(t)}$ 表示 F_t 在 $\mathbb{S}$ 上的一个极小化解。证明序列 $\{\boldsymbol{\beta}^{(t)}\}_{t\geq 0}$ 的每个聚点都是 F 的极小化解。此外，若 F 有唯一极小化解，则整个序列 $\{\boldsymbol{\beta}^{(t)}\}$ 收敛到该极小化解，当 $t \to \infty$ 时。

4. 证明函数 $F_\varepsilon(\boldsymbol{\beta})$ 在 10.39 中是严格凸的，当 $\varepsilon > 0$ 时。
5. 设 $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \in \mathbb{R}^p$，并将它们的非递增重排记为 $[\boldsymbol{\alpha}]^{\downarrow}$ 和 $[\boldsymbol{\beta}]^{\downarrow}$；参见（10.34）。证明

(i) $\left\|[\boldsymbol{\alpha}]^{\downarrow} - [\boldsymbol{\beta}]^{\downarrow}\right\|_\infty \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_\infty$。(ii) $|\sigma_k(\boldsymbol{\alpha}) - \sigma_k(\boldsymbol{\beta})| \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_1$ 对于 $k \in \{1, 2, \ldots, p\}$。(iii) $[\boldsymbol{\alpha}]_t^{\downarrow} \leq \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_1 + \sigma_k(\boldsymbol{\beta})_1$ 对于任意 $t > k$。
6. 证明，将 Bregman迭代应用于最小二乘损失可得到（10.85）和（10.86）中给出的两个等价表述。
7. 讨论如何调整线性化 Bregman 迭代算法，以求解约束 LASSO 问题：$\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$，满足 $\|\boldsymbol{\beta}\|_1 \leq \Sigma$。

参考文献

Alekh Agarwal、Sahand Negahban 和 Martin J Wainwright。用于高维统计恢复的梯度方法的快速全局收敛速率。神经信息处理系统进展，23，2010年。Kenneth Joseph Arrow、Leonid Hurwicz、Hirofumi Uzawa、Hollis Burnley Chenery、Selmer Johnson 和 Samuel Karlin。线性与非线性规划研究，第2卷。斯坦福大学出版社，斯坦福，1958年。Necdet S Aybat 和 Garud Iyengar。用于压缩感知的一阶增广拉格朗日方法。*SIAM* 优化杂志，22(2):429–459，2012。Richard Baraniuk、Mark Davenport、Ronald DeVore 和 Michael Wakin。随机矩阵受限等距性质的一个简单证明。《构造性逼近》，28:253–263，2008年。Richard Baraniuk、Mark A Davenport、Marco F Duarte、Chinmay Hegde 等。压缩感知导论。*Connexions* 电子教材，第24–76页，2011年。Jonathan Barzilai 和 Jonathan M Borwein。两点步长梯度方法。*IMA* 数值分析杂志，8(1):141–148，1988年。Thomas Bayes。试图解决概率学说中的一个问题。*Biometrika*，45(3-4):296–315，1958年。Amir Beck。非线性优化导论：理论、算法及 *MATLAB* 应用。SIAM，2014年。Amir Beck。优化中的一阶方法。SIAM，2017年。Amir Beck 和 Marc Teboulle。用于约束总变差图像去噪和去模糊问题的快速基于梯度的算法。*IEEE* 图像处理汇刊，18(11):2419–2434，2009年。Dennis S. Bernstein。矩阵数学：理论、事实与公式。2008年。Serguei Bernstein。关于切比雪夫不等式的一种修正。《乌克兰拯救科学研究所数学部年刊》，第38–49页，1924年。

Dimitri Bertsekas。凸优化理论，第1卷。Athena Scientific，2009年。Dimitri P Bertsekas。非线性规划。《运筹学会杂志》，48(3):334–334，1997 年。Dimitri P Bertsekas、A Nedi´c 和 AE Ozdaglar。凸分析与优化，Athena Scientific，马萨诸塞州贝尔蒙特，2003。与 *Angelia Nedic* 和 *Asuman E. Ozdaglar* 合著。*MR2184037 (2006j: 90001)*。Arne Beurling。论傅里叶绝对收敛积分及其在函数变换中的应用。载于第九届斯堪的纳维亚数学大会，第1938卷，第345–366页，1938年。˚Ake Bj¨orck。最小二乘问题的数值方法。SIAM，2024。Jean Bourgain、Stephen Dilworth、Kevin Ford、Sergei Konyagin 和 Denka Kutzarova。RIP 矩阵的显式构造及相关问题。2011年。Stephen Boyd、Neal Parikh、Eric Chu、Borja Peleato、Jonathan Eckstein 等。通过交替方向乘子法进行分布式优化与统计学习。基础与趋势® 机器学习，3(1):1–122，2011年。Stephen P Boyd 和 Lieven Vandenberghe。凸优化。剑桥大学出版社，2004年。Lev M Bregman。通过连续投影法求凸集的公共点。载于《苏联数学文献》，第6卷，第688–692页，1965年。Alfred M Bruckstein、David L Donoho 和 Michael Elad。从方程组的稀疏解到信号与图像的稀疏建模。《*SIAM*评论》，51(1):34–81，2009年。Valeriĭ Vladimirovich Buldygin 和 IU V Kozachenko。随机变量与随机过程的度量刻画，第188卷。美国数学会，2000年。Martin Burger 和 Stanley Osher。凸变分正则化的收敛速率。逆问题，20(5):1411，2004年。T Tony Cai 和 Tiefeng Jiang。随机矩阵相干性的极限定律及其在协方差结构检验和压缩感知矩阵构造中的应用。2011年。T Tony Cai、Lie Wang 和 Guangwu Xu。平移不等式与稀疏信号恢复。《*IEEE*信号处理汇刊》，58(3):1300–1308，2009年。EJ Cand。通过 L1 最小化实现近乎理想的模型选择。《统计年刊》，37(5A):2145–2177，2008年。Emmanuel Candes 和 Terence Tao。Dantzig选择器：当 p 远大于 n 时的统计估计。2007年。Emmanuel J Candes。受限等距性质及其对压缩感知的意义。《科学院院报：数学》，346(9-10):589–592，2008年。E mmanuel J Candes 和 Terence Tao。通过线性规划解码。《*IEEE*信息论汇刊》，51(12):4203–4215，2005年。

Emmanuel J Candes 和 Terence Tao。由随机投影实现近乎最优的信号恢复：通用编码策略？《*IEEE*信息论汇刊》，52(12): 5406–5425，2006年。Emmanuel J Candes、Yonina C Eldar、Deanna Needell 和 Paige Randall。利用相干冗余字典进行压缩感知。《应用与计算调和分析》，31(1):59–73，2011年。Antonin Chambolle 和 Thomas Pock。用于凸问题的一阶原始-对偶算法及其在成像中的应用。《数学成像与视觉杂志》，40 (1):120–145，2011年。Scott Shaobing Chen、David L Donoho 和 Michael A Saunders。基于基追踪的原子分解。《*SIAM*评论》，43(1):129–159，2001年。Shaobing Chen 和 David Donoho。基追踪。载于《*1994*年第二十八届阿西洛马信号、系统与计算会议论文集》，第1卷，第41–44页。IEEE，1994年。Albert Cohen、Wolfgang Dahmen 和 Ronald DeVore。压缩感知与最佳 k 项近似。《美国数学会杂志》，22(1):211–231，2009年。Harald Cram´er。论概率论中的一个新极限定理。载于《文集》第二卷，第895–913页。施普林格，1994年。Wei Dai 和 Olgica Milenkovic。用于压缩感知信号重建的子空间追踪。《*IEEE*信息论汇刊》，55(5):2230–2249，2009年。Jerome Darbon 和 Stanley Osher。稀疏逼近与反卷积的快速离散优化。预印本，2007年。Sanjoy Dasgupta 和 Anupam Gupta。Johnson–Lindenstrauss 定理的一个初等证明。《随机结构与算法》，22(1):60–65，2003。Ingrid Daubechies、Ronald DeVore、Massimo Fornasier 和 C Sinan G¨unt¨urk。用于稀疏恢复的迭代重加权最小二乘最小化。《纯粹与应用数学通讯：库朗数学科学研究所主办的期刊》，63(1):1–38，2010年。Mark A Davenport、Jason N Laska、Petros T Boufounos 和 Richard G Baraniuk。随机矩阵具有民主性的一个简单证明。*arXiv* 预印本 *arXiv:0911.0736*，2009年。Mark A Davenport、Marco F Duarte、Yonina C Eldar 和 Gitta Kutyniok。压缩感知导论，2012。Geoff Davis、Stephane Mallat 和 Marco Avellaneda。自适应贪心逼近。《构造性逼近》，13:57–98，1997 年。Ronald A DeVore。压缩感知矩阵的确定性构造。《复杂性杂志》，23(4-6):918–925，2007年。David L Donoho。通过软阈值处理去噪。《*IEEE*信息论汇刊》，41(3):613–627，2002年。

David L Donoho。对于大多数大型欠定线性方程组，其最小 l1 范数解也是最稀疏解。《纯粹与应用数学通讯：库朗数学科学研究所主办的期刊》，59 (6):797–829，2006年。David L Donoho 和 Michael Elad。通过 l1 最小化，在一般（非正交）字典中实现最优稀疏表示。《美国国家科学院院刊》，100(5):2197–2202，2003。David L Donoho 和 Benjamin F Logan。信号恢复与大筛法。《*SIAM*应用数学杂志》，52(2):577–591，1992年。David L Donoho、Xiaoming Huo 等。 不确定性原理与理想原子分解。《*IEEE*信息论汇刊》，47(7):2845–2862，2001年。DL Donoho、Y Tsaig、I Drori 和 JL Starck。通过分阶段正交匹配追踪（StOMP）求解欠定线性方程组的稀疏解。投稿中，2007年。Petros Drineas、Ravi Kannan 和 Michael W Mahoney。用于矩阵的快速蒙特卡洛算法 II：计算矩阵的低秩近似。《*SIAM*计算杂志》，36(1):158–183，2006年。John Duchi、Shai Shalev-Shwartz、Yoram Singer 和 Tushar Chandra。用于高维机器学习的 l1 球高效投影。载于《第*25*届国际机器学习会议论文集》，第272–279页，2008年。Bradley Efron、Trevor Hastie、Iain Johnstone 和 Robert Tibshirani。最小角回归。2004年。Michael Elad、Peyman Milanfar 和 Ron Rubinstein。信号先验中的分析与合成。《逆问题》，23(3):947，2007年。M´ario AT Figueiredo、Robert D Nowak 和 Stephen J Wright。用于稀疏重建的梯度投影：在压缩感知及其他逆问题中的应用。《*IEEE*信号处理精选主题期刊》，1(4):586–597，2008年。Ronald A Fisher。多重测量在分类学问题中的应用。《优生学年鉴》，7(2):179–188，1936年。Massimo Fornasier、Holger Rauhut 和 Rachel Ward。通过迭代重加权最小二乘最小化进行低秩矩阵恢复。*SIAM* 优化杂志，21(4): 1614–1640，2011年。Simon Foucart。关于通过 l 1-最小化保证稀疏恢复的注记。《应用与计算调和分析》，29(1):97–103，2010年。Simon Foucart 和 Ming-Jun Lai。通过 Lq 最小化求解欠定线性系统的最稀疏解，其中 0¡ q¡ 1。《应用与计算调和分析》，26(3): 395–407，2009年。Simon Foucart 和 Holger Rauhut。压缩感知数学导论。施普林格，2013年。

Andrei Yur'evich Garnaev 和 Efim Davydovich Gluskin。欧几里得球的宽度。载于《科学院通报》，第277卷，1048–1052页。俄罗斯科学院，1984年。Semen Aronovich Gershgorin。关于矩阵特征值的界限。《俄罗斯科学院通报·数学辑》，(6):749–754，1931年。Laurent El Ghaoui、Vivian Viallon 和 Tarek Rabbani。用于 LASSO 和稀疏监督学习问题的安全特征消除。*arXiv*预印本 *arXiv:1009.4219*，2010年。Bastian Goldluecke 和 Daniel Cremers。乘积标签空间上多标签问题的凸松弛。载于欧洲计算机视觉会议，225–238页。施普林格，2010年。Irina F Gorodnitsky 和 Bhaskar D Rao。使用 FOCUSS 从有限数据中重构稀疏信号：一种重加权最小范数算法。《*IEEE*信号处理汇刊》，45(3):600–616，1997 年。Markus Grasmair、Otmar Scherzer 和 Markus Haltmeier。l_1 正则化线性收敛的必要且充分条件。《纯粹与应用数学通讯》，64(2):161–182，2011年。Thomas Nall Eden Greville。关于矩阵乘积广义逆的注记。《*SIAM*评论》，8(4):518–521，1966年。Nathan Halko、Per-Gunnar Martinsson 和 Joel A Tropp。借助随机性发现结构：构造近似矩阵分解的概率算法。《*SIAM*评论》，53(2):217–288，2011年。Zhu Han、Husheng Li 和 Wotao Yin。无线网络的压缩感知。剑桥大学出版社，2013年。Trevor Hastie、Robert Tibshirani 和 Martin Wainwright。稀疏性统计学习：*LASSO*及其推广。CRC出版社，2015年。Jarvis Haupt 和 Robert Nowak。从含噪随机投影中重构信号。《*IEEE*信息论汇刊》，52(9):4036–4048，2006年。Ishay Haviv 和 Oded Regev。子采样傅里叶矩阵的受限等距性质。载于《泛函分析的几何方面：以色列研讨会（*GAFA*）*2014– 2016*》，163–179页。施普林格，2017年。Matthew A Herman 和 Thomas Strohmer。通过压缩感知实现高分辨率雷达。《*IEEE*信号处理汇刊》，57(6):2275–2284，2009年。Wassily Hoeffding。有界随机变量之和的概率不等式。*Journal of the American Statistical Association*，58(301):13–30，1963年。Piotr Indyk。稀疏信号压缩感知的显式构造。载于*SODA*，第8卷，30–33页。Citeseer，2008年。Prateek Jain、Ambuj Tewari 和 Purushottam Kar。高维 M 估计的迭代硬阈值方法。神经信息处理系统进展，27，2014年。

Prateek Jain、Purushottam Kar 等。机器学习的非凸优化。*Foundations and Trends® in Machine Learning*，10(3-4):142–336，2017年。William B Johnson、Joram Lindenstrauss 等。将利普希茨映射扩展到希尔伯特空间。*Contemporary mathematics*，26(189-206):1，1984年。Jeena Joy、Salice Peter 和 Neetha John。使用软阈值进行去噪。*InternationalJournalof AdvancedResearchin Electrical, Electronics and Instrumentation Engineering*，2(3):1027–1032，2013年。Boris Sergeevich Kashin。八面体的直径。*Uspekhi Matematicheskikh Nauk*，30 (4):251–252，1975年。Seung-Jean Kim、Kwangmoo Koh、Stephen Boyd 和 Dimitry Gorinevsky。 \ell 1 趋势滤波。《*SIAM*评论》，51(2):339–360，2009年。Yongdai Kim、Hosik Choi 和 Hee-Seok Oh。高维下的平滑剪切绝对偏差。*Journal of the American Statistical Association*，103(484):1665–1673，2008年。Felix Krahmer 和 Rachel Ward。利用受限等距性质改进 Johnson–Lindenstrauss 嵌入。*SIAM Journal on Mathematical Analysis*，43(3):1269–1281，2011年。Jason N Laska、Petros T Boufounos、Mark A Davenport 和 Richard G Baraniuk。民主付诸实践：量化、饱和与压缩感知。*Applied and Compu-tational Harmonic Analysis*，31(3):429–443，2011年。Michel Ledoux。测度集中现象。第89号。美国数学会，2001年。Yulong Liu、Tiebin Mi 和 Shidong Li。利用基于最优对偶的 l1 分析，通过一般框架实现压缩感知。《*IEEE*信息论汇刊》，58(7):4201– 4214，2012。Benjamin Franklin Logan。高通信号的性质。博士论文，哥伦比亚大学，1965年。Jun Lu。数值矩阵分解。*arXiv*预印本 *arXiv:2107.02579*，2021a。Jun Lu。线性模型严谨导论。*Elive Press*，*arXiv*预印本 *arXiv:2105.04240*，2021b。Jun Lu。矩阵分解及其应用。*arXiv*预印本 *arXiv:2201.00145*，2022年。Jun Lu。贝叶斯矩阵分解及其应用。*arXiv*预印本 *arXiv:2302.11337*，2023年。Jun Lu。优化中的实用主题。*arXiv*预印本 *arXiv:2503.05882*，2025年。M Lustig、DL Donoho 和 JM Pauly。利用压缩感知和随机欠采样三维傅里叶变换轨迹快速进行磁共振成像。载于第*14*届*ISMRM*年会论文集。Citeseer，2006年。

Michael Lustig、David Donoho 和 John M Pauly。稀疏磁共振成像：压缩感知在快速磁共振成像中的应用。《医学磁共振》：国际磁共振医学学会官方期刊，58(6):1182–1195，2007年。St´ephane Mallat。《信号处理小波之旅》。爱思唯尔，1999。St´ephane G Mallat 和 Zhifeng Zhang。基于时频字典的匹配追踪。《*IEEE*信号处理汇刊》，41(12):3397–3415，1993 年。Gonzalo Mateos、Juan Andr´es Bazerque 和 Georgios B Giannakis。分布式稀疏线性回归。《*IEEE*信号处理汇刊》，58(10):5262–5276，2010年。Shahar Mendelson、Alain Pajor 和 Nicole Tomczak-Jaegermann。伯努利系综与次高斯系综的一致不确定性原理。《构造性逼近》，28:277–289，2008年。Qun Mo 和 Song Li。受限等距常数的新界 δ_{2k}。《应用与计算调和分析》，31(3):460–468，2011年。Sangnam Nam、Mike E Davies、Michael Elad 和 R´emi Gribonval。余稀疏分析模型与算法。《应用与计算调和分析》，34(1):30–56，2013年。Balas Kausik Natarajan。线性方程组的稀疏近似解。《*SIAM*计算杂志》，24(2):227–234，1995 年。Deanna Needell 和 Joel A Tropp。CoSaMP：从不完整且不准确的样本中迭代恢复信号。《应用与计算调和分析》，26(3):301–321，2009年。Deanna Needell 和 Roman Vershynin。基于正则化正交匹配追踪的一致不确定性原理与信号恢复。《计算数学基础》，9(3):317–334，2009年。Deanna Needell 和 Roman Vershynin。通过正则化正交匹配追踪从不完整且不准确的测量值中恢复信号。《*IEEE*信号处理精选主题期刊》，4(2):310–316，2010年。Harry Nyquist。电报传输理论中的若干主题。*IEEE* 会议录，90(2):280–305，2002。Henrik Ohlsson、Lennart Ljung 和 Stephen Boyd。使用范数和正则化对 ARX 模型进行分段。《自动化》，46(6):1107–1111，2010年。Bruno A Olshausen 和 David J Field。通过学习自然图像的稀疏编码，涌现出简单细胞感受野特性。《自然》，381(6583):607–609，1996年。Michael R Osborne、Brett Presnell 和 Berwin A Turlach。最小二乘问题中变量选择的一种新方法。*IMA* 数值分析杂志，20(3):389–403，2000年。Stanley Osher、Martin Burger、Donald Goldfarb、Jinjun Xu 和 Wotao Yin。一种基于全变分的图像复原迭代正则化方法。《多尺度建模与仿真》，4(2):460–489，2005年。

Yagyensh Chandra Pati、Ramin Rezaiifar 和 Perinkulam Sambamurthy Krishnaprasad。正交匹配追踪：用于小波分解等应用的递归函数逼近。载于第*27*届阿西洛马信号、系统与计算会议论文集，第40–44页。IEEE，1993。William B Pennebaker 和 Joan L Mitchell。《JPEG：静止图像数据压缩标准》。施普林格科学与商业媒体，1992年。Roger Penrose。矩阵的广义逆。载于《剑桥哲学学会数学会议录》，第51卷，第406–413页。剑桥大学出版社，1955年。Arnau Tibau Puig、Ami Wiesel、Gilles Fleury 和 Alfred O Hero。多维收缩阈值算子与组套索惩罚。《*IEEE*信号处理快报》，18(6):363–366，2011年。Majid Rabbani 和 Rajan Joshi。JPEG 2000 静止图像压缩标准概述。《信号处理：图像通信》，17(1):3–48，2002。Garvesh Raskutti、Martin J Wainwright 和 Bin Yu。相关高斯设计的受限特征值性质。《机器学习研究杂志》，11:2241–2259，2010年。Ralph Tyrell Rockafellar。《凸分析：(pms-28)》。1970年。Moshe Rosenfeld。赞颂格拉姆矩阵。《保罗·埃尔德什数学文集 ¨第二卷》，第318–323页，1997 年。Volker Roth。广义 LASSO。《*IEEE*神经网络汇刊》，15(1):16–28，2004年。Leonid I Rudin、Stanley Osher 和 Emad Fatemi。非线性全变差噪声去除算法。《物理学*D*辑：非线性现象》，60(1-4):259–268，1992年。Claude E Shannon。噪声环境下的通信。《*IRE*会刊》，37 (1):10–21，2006年。Thomas Strohmer 和 Robert W Heath Jr。格拉斯曼框架及其在编码与通信中的应用。《应用与计算调和分析》，14(3):257–275，2003。David S Taubman、Michael W Marcellin 和 Majid Rabbani。JPEG 2000：图像压缩基础、标准与实践。《电子成像杂志》，11(2):286–287，2002。Robert Tibshirani。通过 LASSO 实现回归收缩与变量选择。英国皇家统计学会杂志 *B*辑：统计方法学，58(1):267–288，1996年。Ryan J Tibshirani。LASSO 问题与唯一性。2013年。Andrei N Tikhonov。错误表述问题的求解与正则化方法。苏联科学院院刊，4:1035–1038，1963年。Joel A Tropp。随机子字典的条件数。《应用与计算调和分析》，25(1):1–24，2008年。

Joel A Tropp 和 Stephen J Wright。线性逆问题稀疏解的计算方法。*IEEE* 会议录，98(6):948–958，2010年。Vladimir Vapnik。《统计学习理论的本质》。施普林格科学与商业媒体，2013年。Roman Vershynin。随机矩阵非渐近分析导论。*arXiv* 预印本 *arXiv:1011.3027*，2010年。Jie Wang、Jiayu Zhou、Peter Wonka 和 Jieping Ye。通过对偶多面体投影实现 Lasso 筛选规则。神经信息处理系统进展，26，2013年。Rachel Ward。利用交叉验证进行压缩感知。《*IEEE*信息论汇刊》，55(12):5773–5782，2009年。Sanford Weisberg。《应用线性回归》，第528卷。约翰·威立父子公司，2005年。Lloyd Welch。信号最大互相关的下界（通信）。《*IEEE*信息论汇刊》，20(3):397–399，2003。Christopher KI Williams 和 Carl Edward Rasmussen。《机器学习中的高斯过程》，第2卷。麻省理工学院出版社，马萨诸塞州剑桥，2006年。Przemyslaw Wojtaszczyk。压缩感知中高斯测量的稳定性与实例最优性。《计算数学基础》，10(1):1–13，2010年。David P Woodruff 等人。草图化：数值线性代数的工具。理论计算机科学基础与趋势® 理论计算机科学，10(1–2):1–157，2014年。Hui-Hua Wu 和 Shanhe Wu。柯西-施瓦茨不等式的多种证明。《*Octogon* 数学杂志》，17(1):221–229，2009年。Huan Xu、Constantine Caramanis 和 Shie Mannor。稳健回归与 Lasso。神经信息处理系统进展，21，2008年。Wotao Yin、Stanley Osher、Donald Goldfarb 和 Jerome Darbon。用于 \ell 1 最小化的 Bregman 迭代算法及其在压缩感知中的应用。*SIAM* 成像科学杂志，1(1):143–168，2008年。Ming Yuan 和 Yi Lin。含分组变量的回归模型选择与估计。英国皇家统计学会杂志 *B*辑：统计方法学，68(1):49–67，2006年。Cun-Hui Zhang。极小极大凹惩罚下近乎无偏的变量选择。2010年。Hui Zhang、Wotao Yin 和 Lizhi Cheng。l1 范数最小化解唯一性的必要充分条件。优化理论与应用杂志，164(1):109–122，2015年。Hui Zhang、Ming Yan 和 Wotao Yin。关于 l1-合成与 l1-分析最小化解唯一性及稳健性的一个条件。计算数学进展，42(6):1381–1399，2016。