arXiv:2603.02682v1 [math.OC] 3 Mar 2026

# $\ell_{1\text{-}2}$ 稀疏优化的正则化：一致性与全局收敛

YaohuaHu*、Hao Wang†、Xiaoqi Yang‡

摘要： $\ell_{1\text{-}2}$ 正则化方法具有很强的稀疏促进能力，可用于求解线性逆问题中的稀疏解，并已成功应用于数学及应用科学的多个领域。本文旨在研究 $\ell_{1\text{-}2}$ 正则化问题的一致性理论和全局收敛算法。在理论方面，我们针对 $\ell_{1\text{-}2}$惩罚项引入受限特征值条件，并利用该条件为 $\ell_{1\text{-}2}$ 正则化问题的全局解建立神谕性质和恢复界。在算法方面，我们提出两类迭代阈值算法，分别采用截断技术和延拓技术来求解 $\ell_{1\text{-}2}$ 正则化问题。此外，在广为人知的限制等距性质这一假设下，我们证明所提出的算法能够收敛至真实稀疏解，且误差在与噪声水平及恢复界相关的容差范围内。初步数值结果表明，与文献中常用的稀疏优化算法相比，我们提出的算法能够逼近真实稀疏解，并显著增强稀疏恢复能力。

关键词：稀疏优化， $\ell_{1\text{-}2}$ 正则化，一致性理论，迭代阈值算法，全局收敛，真实解。

## 1 引言

稀疏优化旨在求解欠定线性系统的一个稀疏解：

$$b = Ax + \varepsilon, \tag{1.1}$$

其中带有 $m \ll n$ 的感知矩阵 $A \in \mathbb{R}^{m\times n}$ 已知，观测 $b \in \mathbb{R}^m$ 是在未知噪声 $\varepsilon \in \mathbb{R}^m$下采集的，变量 $x \in \mathbb{R}^n$ 有待估计。变量 $x$ 的稀疏性等于其非零分量的个数，记为 $\|x\|_0$。

*深圳大学数学科学学院，中国人民共和国深圳 518060（may-hhu@szu.edu.cn）。

†上海科技大学信息科学与技术学院，中国人民共和国上海 201210（wanghao1@shanghaitech.edu.cn）。

‡香港理工大学应用数学系，中国香港九龙（mayangxq@polyu.edu.hk）。

逼近(1.1)的稀疏解的一种流行且实用的技术是如下正则化形式:

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2+\lambda\|x\|_0, \tag{1.2}$$

其中 $\lambda>0$ 是一个正则化参数，用于在数据保真度与稀疏性之间进行权衡。遗憾的是，由于 $\ell_0$ 范数具有非凸性和组合性，计算问题（1.2）的全局/局部解是NP困难的；参见 [26]。

为克服设计数值算法时的这一困难，如下（凸） $\ell_1$正则化问题已得到广泛研究:

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2+\lambda\|x\|_1,$$

其中 $\|x\|_1:=\sum_{i=1}^n|x_i|$ 是一种稀疏促进范数，作为 $\ell_0$ 范数最紧的凸近似。得益于凸性，大量研究致力于探索$\ell_1$ 正则化问题的理论性质并发展其数值算法。其良好的统计性质和一致性理论，包括神谕性质和恢复界，已在矩阵 $A$的若干类正则性条件下得到充分研究，例如互不相干性条件（MIP） [5], 限制等距性质（RIP） [7], 零空间性质（NSP） [14] 以及限制特征值条件（REC） [2]。许多独特而高效的算法已在收敛保证下被提出并得到发展，例如迭代软阈值算法（ISTA） [9], 近端梯度算法（PGA） [1], 交替方向乘子法（ADMM） [35] 以及块坐标下降法（BCD） [32]。然而， $\ell_1$ 正则化方法在实际应用中存在若干不足，包括对变量 [13] 的较大绝对值分量过度惩罚，以及应用于压缩感知时需要大量测量 [34]。

为克服应用 $\ell_1$ 正则化时的这些缺点，人们提出并发展了多种非凸正则化方法，例如平滑截断绝对偏差（SCAD） [13], 极小极大凹惩罚（MCP） [38]、 $\ell_p$ 惩罚（参数为 $p\in(0,1)$ [8]）、 $\ell_{1\text{-}2}$ 惩罚 [36] 和 $\ell_1/\ell_2$ 惩罚[12], 。文献 [21, 36] 指出， $\ell_{1\text{-}2}$ 范数的等高线比 $\ell_1$ 范数和 $\ell_p$ 范数的等高线更接近 $\ell_0$ 范数的等高线，因此， $\ell_{1\text{-}2}$ 正则化问题

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2+\lambda(\|x\|_1-\|x\|_2) \tag{1.3}$$

在感知矩阵 $A$ 高度相干时，与 $\ell_1$ 和 $\ell_p$ 正则化相比，它可能表现更优：其稀疏促进能力显著更强，并且只需更少的线性测量就能获得更稀疏的解。 $\ell_{1\text{-}2}$ 正则化方法已成功应用于众多领域，例如点源超分辨率 [22], 、图像复原 [23] 和 MRI重建 [24]。受这些重要应用的推动，许多研究致力于发展 $\ell_{1\text{-}2}$ 正则化问题（1.3）的理论性质和优化算法。尽管

$\ell_{1\text{-}2}$ 惩罚项的非凸性和非可分性，可能会给 $\ell_{1\text{-}2}$ 正则化问题（1.3）的理论和算法研究带来一些困难，但在 RIP [36] 和 NSP [15], 的假设下，分别已建立带有观测测量约束的 $\ell_{1\text{-}2}$ 最小化问题的精确恢复性质。据我们所知，目前仍没有研究致力于探讨 $\ell_{1\text{-}2}$ 正则化问题（1.3）的神谕性质和一致性理论。人们已开发出许多实用算法来求解 $\ell_{1\text{-}2}$ 正则化问题（1.3），例如 PGA 和交替方向乘子法（ADMM） [20], 以及差分凸函数算法（DCA） [36]。然而，受限于 $\ell_{1\text{-}2}$ 正则化的非凸性，现有文献仅证明了问题（1.3）收敛到驻点，尚无理论证据能够保证其“全局收敛”，即收敛到全局最小值或真实解。这是非凸优化算法的典型缺陷。

最近，在迭代有限收缩阈值算法框架下，研究者已探讨了对于某些非凸稀疏正则化问题, PGA 收敛到近似全局最小值或真实解的情形。这些问题包括在 RIP 假设下的 SCAD、MCP 和正则化。然而， [17] 中的框架仅关注可分惩罚项，无法用于处理非可分惩罚问题（1.3）。

本文旨在探讨 $\ell_{1\text{-}2}$ 正则化问题 (1.3) 中若干优化算法的一致性理论与全局收敛。在一致性理论方面，我们首先引入相对于 $\ell_{1\text{-}2}$ 正则化的受限特征值条件（REC），并依据 MIP 和 RIP 等经典正则性条件研究其充分条件；见命题 2.1。在 REC 假设下，对于 $\ell_{1\text{-}2}$ 正则化问题 (1.3) 中真实解处水平集内的任意点（包括问题 (1.3) 的全局最优解），我们得到一个预言机不等式和一个阶为 $\mathcal{O}\left(\lambda^2(\sqrt{s}+1)^2\right)$ 的恢复界；见定理2.1。在算法方面，我们通过简化近端映射的解析表达式 [20], ，为 $\ell_{1\text{-}2}$ 正则化引入一种阈值算子。该算子能够克服软阈值算子对绝对值较大分量的过度惩罚现象；详情见注释 3.1。借助 $\ell_{1\text{-}2}$ 阈值算子，我们提出一种带截断技术的迭代阈值算法 [4] (ITAT)，以及一种带延拓技术的迭代阈值算法 [16] (ITAC)；分别见算法 3.1 和 3.2。所提出的 ITAT 和 ITAC 形式简单、存储需求低，因此在求解大规模稀疏优化问题时效率极高。更重要的是，在 RIP 假设下，我们证明 ITAT 生成的序列以几何速率收敛到 (1.1) 的近似真实稀疏解，且误差容差由噪声水平和恢复界共同决定；ITAC 的输出则以与噪声水平成正比的容差逼近 (1.1) 的近似真实稀疏解；分别见定理 3.1 和 3.2。初步数值实验表明，ITAT 和 ITAC 具有很强的稀疏促进能力，在精度和鲁棒性方面均优于标准 PGA。

本文结构安排如下。第2节建立 $\ell_{1\text{-}2}$ 正则化问题 (1.3) 的一致性理论。第3节提出用于求解问题 (1.3) 的 ITAT 和 ITAC，并在 RIP 假设下证明它们收敛到 (1.1) 的近似真实稀疏解。第4节给出初步数值结果。

本文采用的记号在相关文献中均为标准记法。按照惯例，小写字母 $x$、$y$ 和 $z$ 表示向量，花体字母 $\mathcal{I}$、$\mathcal{J}$ 和 $\mathcal{S}$ 表示指标集。对于 $x \in \mathbb{R}^n$ 和 $A \in \mathbb{R}^{m\times n}$，我们分别用 $x_i$ 和 $A_i$ 表示 $x$ 的第 $i$ 个分量和 $A$ 的第 $i$ 列。对于 $\mathcal{I} \subseteq [n]$，我们用 $\mathcal{I}^c$ 表示 $\mathcal{I}$ 的补集，并用 $x_{\mathcal{I}}$ 和 $A_{\mathcal{I}}$ 分别表示由 $x$ 中分量 $i \in \mathcal{I}$ 组成的子向量，以及由 $A$ 中列 $i \in \mathcal{I}$ 组成的子矩阵。按照惯例，我们记 $[n] := \{1, \cdots, n\}$ 和 $x_+ := \max\{x, 0\}$，并用 $x \odot y := (x_i y_i)_{i=1}^n$ 表示向量 $x$ 和 $y$ 的 Hadamard 积。以下 $\ell_1$、$\ell_2$ 和 $\ell_\infty$ 范数之间的关系是众所周知的

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty. \tag{1.4}$$

## 2 一致性理论

在本节中，我们假设观测 $b$ 来自于对一个具有支撑集 $\mathcal{S}$ 的真实 $s$-稀疏解 $\bar{x}$ 施加的线性变换 $A$：

$$b = A\bar{x} \text{ with } \mathcal{S} := \{i : \bar{x}_i \neq 0\} \text{ and } s := |\mathcal{S}|. \tag{2.1}$$

为了求得稀疏解 $\bar{x}$，我们研究 $\ell_{1\text{-}2}$ 正则化问题：

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda(\|x\|_1 - \|x\|_2). \tag{2.2}$$

本节旨在探讨 $\ell_{1\text{-}2}$ 正则化问题（2.2）的一致性理论，包括神谕性质和恢复界。为此，我们首先介绍正则性条件的概念，并讨论其性质。

### 2.1 正则性条件

在稀疏优化场景中，感知矩阵 $A$ 需要满足某些正则性条件，以保证稀疏正则化问题的一致性理论；参见 [2, 5, 18, 25] 及其中的参考文献。例如，限制特征值条件（REC）由 [2] 引入，用于研究 $\ell_1$ 正则化问题的一致性理论。

受 [2]，中 REC 思想的启发，我们引入一个相对于 $\ell_{1\text{-}2}$ 惩罚项的 REC 概念，以保证 $\ell_{1\text{-}2}$ 正则化问题 (2.2) 的一致性理论。为此，令 $\bar{x}$ 和 $x^*$ 分别表示由 (2.1) 定义的真实 $s$-稀疏解，以及问题 (2.2) 的全局最优解。注意，残差 $\hat{x} := x^* - \bar{x}$ 始终满足

$$\|\hat{x}_{\mathcal{S}^c}\|_1 \le \|\hat{x}_{\mathcal{S}}\|_1 + \|\hat{x}\|_2;$$

参见定理2.1的证明。与此相关，我们引入如下相对于 $\ell_{1\text{-}2}$ 惩罚项的 REC 概念。值得一提的是，相对于 $\ell_{1\text{-}2}$ 惩罚项的 REC 类似于 [39] 中提出的 $\ell_{1\text{-}2}$ 零空间性质。给定 $1 \le s \le t \ll n$、 $x \in \mathbb{R}^n$ 和 $\mathcal{I} \subseteq [n]$，我们用 $\mathcal{I}(x;t)$ 表示 $\mathcal{I}^c$ 中 $x$ 绝对值最大的前 $t$ 个坐标所对应的索引集合。

定义 2.1. 若矩阵 $A \in \mathbb{R}^{m\times n}$ 满足相对于 $\ell_{1\text{-}2}$ 惩罚项和 $(s,t)$ 的受限特征值条件（$\ell_{1\text{-}2}$-REC$(s,t)$ $inshort$），则称其满足该条件：

$$\phi(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_{\mathcal{J}}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_1 \le \|x_{\mathcal{I}}\|_1 + \|x\|_2, \mathcal{J} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

为研究稀疏正则化问题的一致性理论及其良好的统计性质，人们提出了几种正则性条件，现回顾如下。

定义 2.2（RIC [7]）。（i） $A$的 $s$-限制等距常数记为 $\delta_s$，定义为满足以下条件的最小值：对于任意 $x \in \mathbb{R}^n$ 和 $J \subseteq [n]$，若 $|J| \le s$，则

$$(1-\delta_s)\|x_J\|_2^2 \le \|Ax_J\|_2^2 \le (1+\delta_s)\|x_J\|_2^2. \tag{2.3}$$

（ii） $A$ 的 $(s,t)$-限制正交常数记为 $\theta_{s,\ t}$，定义为满足以下条件的最小值：对于任意 $x \in \mathbb{R}^n$ 和 $J$，若 $T \subseteq [n]$、 $|J| \le s$ 和 $|T| \le t$，且 $J \cap T = \emptyset$，则

$$|\langle Ax_J, Ax_T\rangle| \le \theta_{s,t}\|x_J\|_2\|x_T\|_2. \tag{2.4}$$

定义 2.3（MIC [11]）。 $A$ 的互不相干常数定义为

$$\mu := \sup\{|A_j^\top A_i| : \forall 1 \le i \ne j \le n\}. \tag{2.5}$$

定义 2.4（SEC [10]）。 $A$的 $s$-稀疏最小特征值和 $s$-稀疏最大特征值分别定义为

$$\sigma_{\min}(s) := \min_{\|x\|_0\le s}\sqrt{\frac{x^\top A^\top Ax}{x^\top x}}, \quad \sigma_{\max}(s) := \max_{\|x\|_0\le s}\sqrt{\frac{x^\top A^\top Ax}{x^\top x}}. \tag{2.6}$$

研究上述 $\ell_{1\text{-}2}$-REC与其他类型正则性条件之间的关系是很自然的。下面的命题分别根据 SEC 、 RIC 和 MIC 给出了 $\ell_{1\text{-}2}$-REC 成立的一些充分条件。证明见附录 A。

命题 2.1。若满足以下条件之一，则 $A \in \mathbb{R}^{m\times n}$ 满足 $\ell_{1\text{-}2}$-REC$(s,t)$：

(i) $(\sqrt{t}-1)\sigma_{\min}(s+t) > (\sqrt{s}+1)\sigma_{\max}(t)$.

(ii) $(\sqrt{s}+1)\theta_{t,\ s+t} < (\sqrt{t}-1)(1-\delta_{s+t})$。

(iii) $(\sqrt{t}-1)\sigma_{\min}(s+t) > 2\theta_{s+t,1}\sqrt{t}(\sqrt{s}+1)$.

(iv) $(\sqrt{t}-1)\sigma_{\min}(s+t) > 2\mu\sqrt{t(s+t)}(\sqrt{s}+1)$.

(v) $\|A_i\|_2 = 1$ *for each* $i \in [n]$, *and* $\mu < \frac{\sqrt{t}-1}{(s+t)(\sqrt{t}-1)+2\sqrt{t(s+t)}(\sqrt{s}+1)}$.

下面的引理引自 [18]，在证明定理2.1中问题（2.2）的恢复界时将会用到。令 $\text{rank}_i(x)$ 表示 $|x_i|$ 在 $\mathcal{I}^c$ 中的秩（按降序排列），并将 $\mathcal{I}^c$ 划分为一系列互不相交的指标集 $\{\mathcal{I}_k\}$，其中每个集合包含 $t$ 个索引，且按 $\text{rank}_i(x)$递增排列。也就是说，$r := \left\lceil \frac{n-|\mathcal{I}|}{t} \right\rceil$（其中 $\lceil u \rceil$ 表示不大于 $u$的最大整数），并且

$$\mathcal{I}_k(x;t) := \{i \in \mathcal{I}^c : \text{rank}_i(x) \in \{kt+1, \ldots, (k+1)t\}\} \quad \text{for} \quad k \in [r]. \tag{2.7}$$

引理 2.1（[18, 引理 7]）。设 $x \in \mathbb{R}^n$、$\mathcal{I} \subset [n]$ 与 $p \geq 1$，并记 $\mathcal{J} := \mathcal{I} \cup \mathcal{I}(x;t)$ 与 $\mathcal{I}_k := \mathcal{I}_k(x;t)$。则有

$$\|x_{\mathcal{J}^c}\|_p \leq \sum_{k=1}^{r} \|x_{\mathcal{I}_k}\|_p \leq t^{\frac{1}{p}-1}\|x_{\mathcal{I}^c}\|_1.$$

## 2.2 神谕性质与恢复界

在本小节中，我们给出 $\ell_{1\text{-}2}$ 正则化问题 (2.2) 的一致性理论，包括其在 $\ell_{1\text{-}2}$-受限特征值条件下的神谕性质与恢复界。为简便起见，令 $h(x) := \|Ax-b\|_2^2 + \lambda(\|x\|_1 - \|x\|_2)$ 表示问题 (2.2) 的目标函数，并将其水平集定义为

$$\text{lev}_h(\bar{x}) := \{x \in \mathbb{R}^n : h(x) \leq h(\bar{x})\}. \tag{2.8}$$

定理2.1。设 $(\bar{x}, \mathcal{S}, s)$ 由 (2.1) 定义。设 $x^* \in \text{lev}_h(\bar{x})$，并假设 $A$ 满足 $\ell_{1\text{-}2}$-受限特征值条件$(s, s)$。则下列陈述成立。

(i) 预言机不等式成立

$$\frac{1}{2}\|Ax^*-b\|_2^2 + \lambda(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2) \leq \frac{2\lambda^2(\sqrt{s}+1)^2}{\phi^2(s,s)}. \tag{2.9}$$

(ii) 恢复界成立

$$\|x^*-\bar{x}\|_2^2 \leq \left(4 + \max\left\{\frac{\sqrt{s}+1}{\sqrt{s}}, \frac{4}{\sqrt{s}-2}\right\}^2\right)\frac{\lambda^2(\sqrt{s}+1)^2}{\phi^4(s,s)}. \tag{2.10}$$

证明。(i) 令 $x^* \in \text{lev}_h(\bar{x})$。由 (2.8) 和 $\bar{x}_{\mathcal{S}^c} = 0$（参见 (2.1)）可得

$$\frac{1}{2}\|Ax^*-b\|_2^2 + \lambda(\|x^*\|_1 - \|x^*\|_2) \leq \lambda(\|\bar{x}\|_1 - \|\bar{x}\|_2) = \lambda(\|\bar{x}_{\mathcal{S}}\|_1 - \|\bar{x}_{\mathcal{S}}\|_2). \tag{2.11}$$

注意，

$$\|x^*\|_1 = \|x^*_{\mathcal{S}}\|_1 + \|x^*_{\mathcal{S}^c}\|_1 \quad \text{and} \quad \|x^*\|_2 \le \|x^*_{\mathcal{S}}\|_2 + \|x^*_{\mathcal{S}^c}\|_2 \tag{2.12}$$

（根据 $\ell_2$ 范数的次可加性）。于是（2.11）化简为

$$\frac{1}{2}\|Ax^* - b\|_2^2 + \lambda(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2) \le \lambda(\|\bar{x}_{\mathcal{S}}\|_1 - \|x^*_{\mathcal{S}}\|_1) + \lambda(\|x^*_{\mathcal{S}}\|_2 - \|\bar{x}_{\mathcal{S}}\|_2). \tag{2.13}$$

注意，根据 $\ell_p$ 范数的次可加性（$p = 1,\ 2$），可得

$$\|x^*_{\mathcal{S}}\|_2 - \|\bar{x}_{\mathcal{S}}\|_2 \le \|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2 \quad \text{and} \quad \|\bar{x}_{\mathcal{S}}\|_1 - \|x^*_{\mathcal{S}}\|_1 \le \|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_1 \le \sqrt{s}\|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2 \tag{2.14}$$

（由（1.4）可知）。于是（2.13）化简为

$$\frac{1}{2}\|Ax^* - A\bar{x}\|_2^2 + \lambda(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2) \le \lambda(\sqrt{s}+1)\|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2. \tag{2.15}$$

另一方面，由（2.1）可知 $\bar{x}_{\mathcal{S}^c} = 0$，结合（2.14）以及 $\ell_2$ 范数的次可加性，可得

$$\begin{aligned}\|x^*_{\mathcal{S}^c} - \bar{x}_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_1 - \|x^* - \bar{x}\|_2 &\le \|x^*_{\mathcal{S}^c}\|_1 - (\|\bar{x}_{\mathcal{S}}\|_1 - \|x^*_{\mathcal{S}}\|_1) - (\|x^*\|_2 - \|\bar{x}\|_2)\\ &= \|x^*\|_1 - \|x^*\|_2 - (\|\bar{x}\|_1 - \|\bar{x}\|_2)\\ &\le 0\end{aligned}$$

（由（2.12）和（2.11）可知）。这表明 $x^* - \bar{x}$ 属于 $\ell_{1\text{-}2}$-REC$(s,s)$ 的约束集。因此，根据定义 2.1 可得

$$\|x^*_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2 \le \frac{1}{\phi(s,s)}\|Ax^* - A\bar{x}\|_2.$$

因此（2.15）化简为

$$\frac{1}{2}\|Ax^* - A\bar{x}\|_2^2 + \lambda(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2) \le \frac{\lambda(\sqrt{s}+1)}{\phi(s,s)}\|Ax^* - A\bar{x}\|_2; \tag{2.16}$$

因此，

$$\|Ax^* - A\bar{x}\|_2 \le \frac{2\lambda(\sqrt{s}+1)}{\phi(s,s)}. \tag{2.17}$$

结合式 (2.16)，可得预言机不等式 (2.9)。

(ii) 定义 $\mathcal{N} := \mathcal{S} \cup \mathcal{S}(x^*; s)$。 $\ell_{1\text{-}2}$-受限特征值条件$(s,s)$意味着

$$\|x^*_{\mathcal{N}} - \bar{x}_{\mathcal{N}}\|_2 \le \frac{\|Ax^* - A\bar{x}\|_2}{\phi(s,s)} \le \frac{2\lambda(\sqrt{s}+1)}{\phi^2(s,s)} \tag{2.18}$$

（由（2.17））。根据引理 2.1（以 2、 $s$、 $\mathcal{N}$、 $\mathcal{S}$ 替代 $p$、 $t$、 $\mathcal{J}$、 $\mathcal{I}$），可得

$$\|x^*_{\mathcal{N}^c}\|_2 \le \frac{1}{\sqrt{s}}\|x^*_{\mathcal{S}^c}\|_1; \tag{2.19}$$

因此,

$$\|x^*_{\mathcal{N}^c}\|_2 - \frac{1}{\sqrt{s}}\|x^*_{\mathcal{S}^c}\|_2 \le \frac{1}{\sqrt{s}}(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2). \tag{2.20}$$

由 (2.16) 和柯西不等式可知 $a^2 + b^2 \ge 2ab$

$$\|Ax^* - A\bar{x}\|_2\sqrt{2\lambda(\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2)} \le \frac{\lambda(\sqrt{s}+1)}{\phi(s,s)}\|Ax^* - A\bar{x}\|_2;$$

因此

$$\|x^*_{\mathcal{S}^c}\|_1 - \|x^*_{\mathcal{S}^c}\|_2 \le \frac{\lambda(\sqrt{s}+1)^2}{2\phi^2(s,s)}.$$

结合 (2.20)，可得

$$\|x^*_{\mathcal{N}^c}\|_2 - \frac{1}{\sqrt{s}}\|x^*_{\mathcal{S}^c}\|_2 \le \frac{\lambda(\sqrt{s}+1)^2}{2\sqrt{s}\,\phi^2(s,s)}. \tag{2.21}$$

下面考虑以下两种情形。
情形 1：假设 $\|x^*_{\mathcal{N}^c}\|_2 \ge \frac{2}{\sqrt{s}}\|x^*_{\mathcal{S}^c}\|_2$。则 (2.21) 化简为

$$\|x^*_{\mathcal{N}^c}\|_2 \le \frac{\lambda(\sqrt{s}+1)^2}{\sqrt{s}\,\phi^2(s,s)}. \tag{2.22}$$

情形 2：假设 $\|x^*_{\mathcal{N}^c}\|_2 < \frac{2}{\sqrt{s}}\|x^*_{\mathcal{S}^c}\|_2$。注意到 $\mathcal{N} = \mathcal{S} \cup \mathcal{S}(x^*;\, s)$ 是一个不交分解，且 $\bar{x}_{\mathcal{S}(x^*;s)} = 0$ 当 $\mathcal{S}(x^*; s) \subseteq \mathcal{S}^c$ 时，我们得到

$$\|x^*_{\mathcal{S}(x^*;s)}\|_2 = \|x^*_{\mathcal{S}(x^*;s)} - \bar{x}_{\mathcal{S}(x^*;s)}\|_2 \le \|x^*_{\mathcal{N}} - \bar{x}_{\mathcal{N}}\|_2 \le \frac{2\lambda(\sqrt{s}+1)}{\phi^2(s,s)}$$

（由（2.18））, 然后根据 $\ell_2$ 范数的次可加性可得

$$\|x^*_{\mathcal{S}^c}\|_2 \le \|x^*_{\mathcal{N}^c}\|_2 + \|x^*_{\mathcal{S}(x^*;s)}\|_2 \le \|x^*_{\mathcal{N}^c}\|_2 + \frac{2\lambda(\sqrt{s}+1)}{\phi^2(s,s)}.$$

这与情形 2 的假设一起推出

$$\|x^*_{\mathcal{N}^c}\|_2 \le \frac{4}{\sqrt{s}-2}\frac{\lambda(\sqrt{s}+1)}{\phi^2(s,s)}. \tag{2.23}$$

结合（2.22）和（2.23）可得

$$\|x^*_{\mathcal{N}^c}\|_2 \le \max\left\{\frac{\sqrt{s}+1}{\sqrt{s}}, \frac{4}{\sqrt{s}-2}\right\}\frac{\lambda(\sqrt{s}+1)}{\phi^2(s,s)}.$$

结合（2.18）和 $\bar{x}_{\mathcal{N}^c} = 0$，可得

$$\|x^* - \bar{x}\|_2^2 = \|x^*_{\mathcal{N}} - \bar{x}_{\mathcal{N}}\|_2^2 + \|x^*_{\mathcal{N}^c}\|_2^2 \le \left(4 + \max\left\{\frac{\sqrt{s}+1}{\sqrt{s}}, \frac{4}{\sqrt{s}-2}\right\}^2\right)\frac{\lambda^2(\sqrt{s}+1)^2}{\phi^4(s,s)},$$

由此得到式 (2.10)，证明完成。 ■

定理2.1是一项重要的理论结果，因为它为 $\ell_{1\text{-}2}$ 正则化问题 (2.2) 提供了预言机不等式 (2.9) 和恢复界 (2.10)。具体而言，预言机不等式 (2.9) 给出了线性系统的平方误差加上 $x^*$ 非零分量违反量的上界，而恢复界 (2.10) 则给出了 $x^*$ 与真实解 $\bar{x}$之间距离的上界。

注记2.1。从（2.10）可以看出， $\ell_{1\text{-}2}$正则化问题（2.2）的全局最优解 $x^*(\lambda)$ 有如下恢复界：

$$\|x^*(\lambda) - \bar{x}\|_2^2 \leq \mathcal{O}\left(\lambda^2(\sqrt{s}+1)^2\right).$$

特别地，当 $s \geq 16$ 时，由（2.10）可得

$$\|x^*(\lambda) - \bar{x}\|_2^2 \leq \frac{8}{\phi^4(s,s)}\lambda^2(\sqrt{s}+1)^2.$$

# 3 迭代阈值算法

迭代阈值算法（ITA）是一类最受欢迎且高效的数值算法之一，其形式简单、计算复杂度低，可用于求解（凸或非凸）稀疏优化问题；参见 [4, 9, 14, 18, 34] 及其中的参考文献。特别地，用于求解 $\ell_{1\text{-}2}$ 正则化问题（2.2）的 ITA 已在 [20]中提出。然而，受限于 $\ell_{1\text{-}2}$ 正则化的非凸性，其 ITA 的收敛理论仍远不能令人满意： [20, 定理 1]仅证明了其收敛到问题（2.2）的一个驻点；而目前仍无理论依据能够保证其收敛到全局最小值或真实稀疏解。

为弥补非凸优化算法中这一典型缺口，本节旨在提出两种新颖的 ITA 型算法，以求解 $\ell_{1\text{-}2}$ 正则化问题（2.2），并研究它们收敛到真实稀疏解的性质。

## 3.1 $\ell_{1\text{-}2}$ 阈值算子

近端梯度算法（PGA） [1, 18] 是求解复合优化问题最流行且实用的算法之一

$$\min_{x\in\mathbb{R}^n} f(x) + \varphi(x), \tag{3.1}$$

其中， $f:\ \mathbb{R}^n \to \mathbb{R}$ 是可微函数， $\varphi:\ \mathbb{R}^n \to \mathbb{R}$ 具有某些特殊结构。用于求解（3.1）的 PGA 迭代公式为

$$x^{k+1} := \mathrm{Prox}_{\varphi}(x^k - v\nabla f(x^k)). \tag{3.2}$$

其中 $\varphi$ 的近端映射$\mathrm{Prox}_{\varphi}: \mathbb{R}^n \to \mathbb{R}^n$ 定义为

$$\mathrm{Prox}_{\varphi}(y) := \arg\min_{x\in\mathbb{R}^n}\ \varphi(x) + \frac{1}{2}\|x-y\|_2^2 \quad \text{for each } y \in \mathbb{R}^n.$$

这类迭代阈值算法可以理解为将近端梯度算法应用于求解带有特定正则项的复合优化问题(3.1)。例如，迭代软（相应地，硬、半）阈值算法可以理解为用于求解 $\ell_1$（相应地， $\ell_0$、$\ell_{1/2}$）正则化问题的PGA；分别参见 [9, 4, 34] 。

特别地，当应用于问题 (2.2)（在 (3.1) 中用 $\frac{1}{2}\|Ax-b\|^2$ 和 $\lambda(\|x\|_1-\|x\|_2)$ 替换 $f(x)$ 和 $\varphi(x)$）时，用于求解 $\ell_{1\text{-}2}$ 正则化问题 (2.2) 的 ITA 具有如下迭代公式

$$x^{k+1} := \mathrm{Prox}_{v\lambda(\|x\|_1-\|x\|_2)}(x^k - vA^\top(Ax^k-b)). \tag{3.3}$$

在 [20, 引理 1] 中给出，$\mathrm{Prox}_{\lambda(\|x\|_1-\|x\|_2)}$ 具有如下解析公式。

引理 3.1. 解 $x^* \in \mathrm{Prox}_{\lambda(\|x\|_1-\|x\|_2)}(y)$ 具有如下公式。

(i) 若 $\|y\|_\infty > \lambda$，则 $x^* = \left(1+\frac{\lambda}{\|z\|_2}\right)z$，其中 $z := (|y|-\lambda)_+ \odot \mathrm{sign}(y)$。(ii) 若 $\|y\|_\infty = \lambda$，则 $\|x^*\|_2 = \lambda$ ，其中 $x_i^* y_i \geq 0$ 对每个 $i \in [n]$ 均成立，并且当$|y_i| < \lambda$时 $x_i^* = 0$ 。

(iii) 若 $\|y\|_\infty < \lambda$，则 $x^*$ 是一个满足 $\|x^*\|_2 = \|y\|_\infty$ 的 1-稀疏向量，其中对每个 $i \in [n]$ 有 $x_i^* y_i \geq 0$ ，且当 $|y_i| < \|y\|_\infty$ 时有 $x_i^* = 0$ 。

结合一致性理论（参见定理2.1），正则化参数 $\lambda$ 在数值实验中应始终设置得相当小，以保证精确恢复和数值性能。在 $\lambda$ 被设置得很小的情况下，引理 3.1 (i) 中的情形总会在数值计算过程中出现。因此，为简化 $\ell_{1\text{-}2}$ 正则化的近端映射计算，我们定义一个 $\ell_{1\text{-}2}$ 阈值算子如下

$$\mathbb{T}_\lambda(x) := \mathbb{E}_\lambda \circ \mathbb{S}_\lambda(x), \tag{3.4}$$

其中 $\mathbb{S}_\lambda$ 和 $\mathbb{E}_\lambda$ 分别为软阈值算子和扩张算子，定义如下

$$\mathbb{S}_\lambda(x) := (|x|-\lambda)_+ \odot \mathrm{sign}(x) \quad \text{and} \quad \mathbb{E}_\lambda(x) := \left(1+\frac{\lambda}{\|x\|_2}\right)x. \tag{3.5}$$

由引理 3.1(i) 和 (3.4) 可得

$$\mathbb{T}_\lambda(y) = \mathrm{Prox}_{\lambda(\|x\|_1-\|x\|_2)}(y) \quad \text{whenever } \|y\|_\infty > \lambda. \tag{3.6}$$

下面我们讨论 $\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$的一些有用性质。

引理 3.2。 $\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$： $\mathbb{R}^n \to \mathbb{R}^n$ 满足以下性质：

(i) 阈值性质：

$$(\mathbb{T}_\lambda(x))_i = 0 \quad \textit{whenever } |x_i| \leq \lambda. \tag{3.7}$$

(ii) 收缩性质：

$$\|\mathbb{T}_\lambda(x) - x\|_\infty \leq \lambda. \tag{3.8}$$

证明。(i) 由 (3.5) 和 (3.4) 可直接得到

$$|x_i| \leq \lambda \quad \Rightarrow \quad (\mathbb{S}_\lambda(x))_i = 0 \quad \Rightarrow \quad (\mathbb{T}_\lambda(x))_i = 0.$$

(ii) 从 (3.7) 可以看出

$$|(\mathbb{T}_\lambda(x))_i - x_i| \leq \lambda \quad \text{whenever } |x_i| \leq \lambda. \tag{3.9}$$

接下来讨论 $|x_i| > \lambda$ 的情形。在这种情况下，由 (3.5) 和 (3.4) 可得

$$(\mathbb{S}_\lambda(x))_i = (|x_i| - \lambda)\,\mathrm{sign}(x_i) = x_i - \lambda\,\mathrm{sign}(x_i), \tag{3.10}$$

and

$$(\mathbb{T}_\lambda(x))_i = (\mathbb{E}_\lambda \circ \mathbb{S}_\lambda(x))_i = \left(1 + \frac{\lambda}{\|\mathbb{S}_\lambda(x)\|_2}\right)(\mathbb{S}_\lambda(x))_i.$$

结合上述两个不等式，可得

$$\begin{aligned}(\mathbb{T}_\lambda(x))_i - x_i &= \left(1 + \frac{\lambda}{\|\mathbb{S}_\lambda(x)\|_2}\right)(\mathbb{S}_\lambda(x))_i - ((\mathbb{S}_\lambda(x))_i + \lambda\,\mathrm{sign}(x_i)) \\ &= \frac{\lambda}{\|\mathbb{S}_\lambda(x)\|_2}(\mathbb{S}_\lambda(x))_i - \lambda\,\mathrm{sign}(x_i) \\ &= \lambda\,\mathrm{sign}(x_i)\left(\frac{|(\mathbb{S}_\lambda(x))_i|}{\|\mathbb{S}_\lambda(x)\|_2} - 1\right)\end{aligned} \tag{3.11}$$

（因为 $(\mathbb{S}_\lambda(x))_i$ 和 $x_i$ 的符号相同）。注意到 $1 - \frac{|(\mathbb{S}_\lambda x \ \ i|\,())()\|_2}{\|\mathbb{S}_\lambda\ x} \in [0, 1)$，可知 $|(\mathbb{T}_\lambda(x))_i - x_i| < \lambda$。结合 (3.9)，即可得到 (3.8)。 ■

注 3.1。(i) 引理 3.2 表明，$\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$ 具有阈值和收缩性质，这些性质对于求得稀疏解至关重要。

(ii) 扩张算子 $\mathbb{E}_\lambda$ 增大非零分量的绝对值，因此 $\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$ 能够克服软阈值算子 $\mathbb{S}_\lambda$ 对大幅值分量的过度惩罚现象。特别地，由(3.10)和(3.11)可得

$$0 < \frac{x_i - (\mathbb{T}_\lambda(x))_i}{x_i - (\mathbb{S}_\lambda(x))_i} = 1 - \frac{|(\mathbb{S}_\lambda(x))_i|}{\|\mathbb{S}_\lambda(x)\|_2} \leq 1,$$

从而

$$|(\mathbb{S}_\lambda(x))_i| \leq |(\mathbb{T}_\lambda(x))_i| \leq |x_i|.$$

与 $\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$相关联，用于求解 $\ell_{1\text{-}2}$正则化问题 (2.2) 的标准 ITA 具有如下迭代形式：

$$x^{k+1} := \mathbb{T}_\lambda(x^k - vA^\top(Ax^k - b)). \tag{3.12}$$

注 3.2. 引理 3.1(i) 以及式 (3.6) 指出，当 $\lambda$ 较小时，阈值算子$\mathbb{T}_\lambda$ 等于 $\ell_{1\text{-}2}$ 正则化的近端算子。因此，当 $\lambda$ 设为较小时，ITA (3.12) 等价于用于求解 $\ell_{1\text{-}2}$正则化问题 (2.2) 的 PGA (3.3)。

$\ell_{1\text{-}2}$ 正则化问题(2.2)的理论研究与算法研究之间存在尚未弥合的差距。具体而言，定理2.1中的一致性理论为问题(2.2)的全局最小值到真实稀疏解提供了恢复界；而 [20, 定理 1] 中的收敛理论仅表明标准ITA (3.12)收敛到问题(2.2)的驻点。然而，仍没有理论证据保证其收敛到全局最小值或真实稀疏解。这也是非凸优化算法的一个典型差距。

本文始终假设，观测 $b$ 由线性变换 $A$ 作用于支撑集为 $\mathcal{S}$ 的真实 $s$-稀疏解 $\bar{x}$ ，并叠加未知噪声 $\varepsilon$而得到:

$$b = A\bar{x} + \varepsilon \ \text{ with } \ \mathcal{S} := \{i : \bar{x}_i \neq 0\} \ \text{ and } \ s := |\mathcal{S}|. \tag{3.13}$$

在本节余下部分，我们旨在通过为求解 $\ell_{1\text{-}2}$正则化问题(2.2)的ITA型算法的稀疏恢复能力提供肯定的理论保证，来填补这一差距。具体而言，我们将分别结合截断技术 [4] 与延拓技术 [37], 提出两类迭代阈值算法，并在限制等距性质（RIP） [7]的假设下，研究它们向(1.1)的真实稀疏解 $\bar{x}$ 的收敛性。

注 3.3. RIP *[7]* 已广泛用于稀疏优化算法的收敛性分析 *[4, 14, 27, 33]*。已有研究表明，包括高斯、伯努利和部分傅里叶矩阵在内的多种随机矩阵，以指数级高的概率满足 RIP *[3]*。

本小节最后给出引理中的一些有用的 RIP 性质，这些性质摘自 [17, 引理 1], 以及 [27, 命题 3.1] 和 [14, 引理 6.16 和 6.20], ，并有助于分析我们所提出算法的收敛性。

引理 3.3. 假设 $A$ 满足 RIP，且 $\delta_s < 1$。令 $x \in \mathbb{R}^n$、 $\varepsilon \in \mathbb{R}^m$、 $\mathcal{I}$、 $\mathcal{J} \subseteq [n]$和 $v \geq 0$。则以下结论成立。

(i) 若 $|\mathcal{I} \cup \text{supp}(x)| \leq s$，则 $\|((\mathbb{I} - vA^\top A)x)_\mathcal{I}\|_2 \leq (|1 - v| + v\delta_s)\|x\|_2$。 (ii) 若 $|\mathcal{I}| \leq s$，则 $\|A_\mathcal{I}^\top \varepsilon\|_2 \leq \sqrt{1+\delta_s}\|\varepsilon\|_2$。 (iii) 若 $\mathcal{I} \cap \mathcal{J} = \emptyset$ 且 $|\mathcal{I} \cup \mathcal{J}| \leq s$，则 $\|A_\mathcal{J}^\top A_\mathcal{I} x_\mathcal{I}\|_2 \leq \delta_s \|x_\mathcal{I}\|_2$。

## 3.2 带截断技术的迭代阈值算法

在稀疏优化算法生成的序列中，幅值较大的分量被视为对真实稀疏解的良好近似，而

幅值较小的分量则只能归因于观测中存在噪声。为避免小分量造成噪声，稀疏优化算法中广泛应用截断算子 $\mathbb{H}_s$，该算子将向量中除幅值最大的 $s$ 个元素外的所有元素置为零，以确保解的稀疏结构；参见 [4, 14, 40] 及其中的参考文献。

借助标准的 $\ell_{1\text{-}2}$ 阈值算子 (3.4) 和截断算子，我们提出了一种采用截断技术的迭代阈值算法 (ITAT)，用于求解 $\ell_{1\text{-}2}$ 正则化问题 (2.2)，并逼近 (1.1) 的真实稀疏解。

算法 3.1 (ITAT)。选择一个正则化参数 $\lambda > 0$、一个截断参数$s \in \mathbb{N}$，并设定一个随机初始点 $x^0 \in \mathbb{R}^n$ 和步长 $v > 0$。对于每个 $k \in \mathbb{N}$，已知$x^k$，我们通过以下方式确定 $x^{k+1}$

$$x^{k+1} := \mathbb{H}_s \circ \mathbb{T}_{v\lambda}\left(x^k - vA^\top(Ax^k - b)\right). \tag{3.14}$$

注 3.4。(i) ITAT 采用截断算子 $\mathbb{H}_s$ 来维持序列 $\{x^k\}$的稀疏度$s$ ，这有助于保证收敛到具有 (1.1) 所需稀疏度的真实解；见定理 3.1。

(ii) 由于截断算子 $\mathbb{H}_s$ 和 $\ell_{1\text{-}2}$ 阈值算子 $\mathbb{T}_\lambda$ 的计算非常简单，ITAT 继承了公式简单、计算复杂度低的显著优势，因此在大规模问题中极为高效。

下面的引理回顾了 [28, 定理 1] 中关于截断算子$\mathbb{H}_s(\cdot)$ 的一个紧确界，该界在收敛性分析中很有用。

引理 3.4. 设 $x \in \mathbb{R}^n$ 为一个 $s$-稀疏向量， $z \in \mathbb{R}^n$ 为任意向量，则

$$\|\mathbb{H}_s(z) - x\|_2 \le \frac{\sqrt{5}+1}{2}\|z - x\|_2.$$

本小节的主要结果如下：在 RIP 假设下，我们证明 ITAT 收敛到 (1.1) 的近似真实稀疏解。

定理 3.1. 假设 $A$ 满足 $3s$-RIP，且其 $\delta_{3s} < \frac{\sqrt{5}-1}{2} \approx 0.618$为；令 $\{x^k\}$为由算法 3.1 生成、步长满足下式的序列：

$$\frac{0.382}{1-\delta_{3s}} \approx \frac{3-\sqrt{5}}{2(1-\delta_{3s})} < v < \frac{\sqrt{5}+1}{2(1-\delta_{3s})} \approx \frac{1.618}{1-\delta_{3s}}. \tag{3.15}$$

则 $\{x^k\}$ 以几何速率近似收敛到 $\bar{x}$；特别地，

$$\|x^k - \bar{x}\|_2 \le \rho^k\|x^0 - \bar{x}\|_2 + \frac{v(\sqrt{5}+1)}{2(1-\rho)}(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2 + \sqrt{2s}\lambda), \tag{3.16}$$

其中 $\rho := \frac{\sqrt{5}+1}{2}(|1-v| + v\delta_{3s}) \in (0, 1)$。

证明。为进行收敛性分析，我们将算法 3.1 的过程改写为以下三个步骤:

$$y^k := x^k - vA^\top(Ax^k - b), \quad z^k := \mathbb{T}_{v\lambda}(y^k), \quad x^{k+1} := \mathbb{H}_s(z^k). \tag{3.17}$$

此外，为简便起见，我们记

$$r^k := x^k - \bar{x}, \quad \mathcal{S}_k := \mathrm{supp}(x^k) \quad \text{and} \quad \mathcal{I}_k = \mathcal{S} \cup \mathcal{S}_k; \tag{3.18}$$

于是，由 $\bar{x}$ 的 $s$-稀疏性以及 (3.17) 的第三个等式可知

$$|\mathcal{S}| = s, \quad |\mathcal{S}_k| \le s \quad \text{and} \quad |\mathcal{I}_k| \le 2s \quad \text{for each } k \in \mathbb{N}. \tag{3.19}$$

由 (3.18) 可知

$$\mathrm{supp}(\bar{x}) = \mathcal{S} \subseteq \mathcal{I}_{k+1} \quad \text{and} \quad \mathrm{supp}(x^{k+1}) = \mathcal{S}_{k+1} \subseteq \mathcal{I}_{k+1}, \tag{3.20}$$

由 (3.17) 以及引理 3.4（用 $z^k_{\mathcal{I}_{k+1}}$ 和 $\bar{x}_{\mathcal{I}_{k+1}}$ 分别代替 $z$ 和 $x$），可得

$$\begin{aligned} \|r^{k+1}\|_2 &= \|x^{k+1}_{\mathcal{I}_{k+1}} - \bar{x}_{\mathcal{I}_{k+1}}\|_2 \\ &\le \frac{\sqrt{5}+1}{2}\|z^k_{\mathcal{I}_{k+1}} - \bar{x}_{\mathcal{I}_{k+1}}\|_2 \\ &\le \frac{\sqrt{5}+1}{2}(\|\bar{x}_{\mathcal{I}_{k+1}} - y^k_{\mathcal{I}_{k+1}}\|_2 + \|y^k_{\mathcal{I}_{k+1}} - z^k_{\mathcal{I}_{k+1}}\|_2). \end{aligned} \tag{3.21}$$

由 (3.17) 可知 $z^k = \mathbb{T}_{v\lambda}(y^k)$，根据引理 3.2 可得 $\|z^k - y^k\|_\infty \le v\lambda$。结合 (3.19)，由 (1.4) 可得

$$\|y^k_{\mathcal{I}_{k+1}} - z^k_{\mathcal{I}_{k+1}}\|_2 \le \sqrt{|\mathcal{I}_{k+1}|}\|y^k - z^k\|_\infty \le \sqrt{2s}v\lambda. \tag{3.22}$$

另一方面，由 (3.17) 的第一个等式和 (3.13) 可得

$$y^k = x^k - vA^\top(Ax^k - b) = x^k - vA^\top Ar^k + vA^\top\varepsilon$$

（由 (3.18)）；因此可得

$$\|y^k_{\mathcal{I}_{k+1}} - \bar{x}_{\mathcal{I}_{k+1}}\|_2 \le \|((I - vA^\top A)r^k)_{\mathcal{I}_{k+1}}\|_2 + v\|A^\top_{\mathcal{I}_{k+1}}\varepsilon\|_2. \tag{3.23}$$

注意，由 (3.20) 可知 $|\mathcal{I}_{k+1}| \le 2s$ 和 $|\mathcal{I}_{k+1} \cup \mathrm{supp}(r^k)| = |\mathcal{I}_{k+1} \cup \mathcal{S}_k| \le 3s$。然后，根据 $A$ 的 $3s$-限制等距性质（RIP）假设及 (3.15)，由引理 3.3(i) 和 (ii) 可得

$$\|((I - vA^\top A)r^k)_{\mathcal{I}_{k+1}}\|_2 \le (|1-v| + v\delta_{3s})\|r^k\|_2 \quad \text{and} \quad \|A^\top_{\mathcal{I}_{k+1}}\varepsilon\|_2 \le \sqrt{1+\delta_{2s}}\|\varepsilon\|_2,$$

分别。由上述两个不等式，(3.23) 可化简为

$$\|y^k_{\mathcal{I}_{k+1}} - \bar{x}_{\mathcal{I}_{k+1}}\|_2 \le (|1-v| + v\delta_{3s})\|r^k\|_2 + v\sqrt{1+\delta_{2s}}\|\varepsilon\|_2.$$

结合 (3.21) 和 (3.22)，可得

$$\|r^{k+1}\|_2 \leq \frac{\sqrt{5}+1}{2}(|1-v|+v\delta_{3s})\|r^k\|_2 + \frac{\sqrt{5}+1}{2}v(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda) \tag{3.24}$$

令 $\rho := \frac{\sqrt{5}+1}{2}(|1-v|+v\delta_{3s})$。根据限制等距性质假设 $\delta_{3s} < \frac{\sqrt{5}-1}{2}$，我们可以验证，当 $\frac{3-\sqrt{5}}{2}\frac{}{1-\delta_{3s}}() < v \leq 1$ 或 $1 \leq v < \frac{\sqrt{5}+1}{2}\frac{}{1-\delta_{3s}}()$ 时，$\rho \in (0,1)$ 成立；也就是说，假设 (3.15) 成立。然后由 (3.24) 归纳可得

$$\begin{aligned}\|x^{k+1}-\bar{x}\|_2 &\leq \rho\|x^k-\bar{x}\|_2 + \frac{\sqrt{5}+1}{2}v(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda)\\ &\leq \cdots\\ &\leq \rho^{k+1}\|x^0-\bar{x}\|_2 + \frac{v(\sqrt{5}+1)}{2(1-\rho)}(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda).\end{aligned}$$

证明完毕。 ■

注 3.5。(i) 定理 3.1 表明，ITAT 以几何收敛速率收敛到 (1.1) 的近似真实稀疏解，且误差在容差范围内。(3.16) 中的容差由两项相加构成：一项与噪声水平 $\mathcal{O}(\|\varepsilon\|_2)$ 有关，另一项与正则化参数 $\mathcal{O}(\sqrt{s}\lambda)$ 有关，其阶与 (2.10) 中的恢复界相同。当选取较小的正则化参数 $\lambda$ 时，项 $\mathcal{O}(\sqrt{s}\lambda)$ 可能较小；第 4 节的数值实验将说明，最佳正则化参数约为 $\lambda = 10^{-3}$。

(ii) 定理 3.1 改进了 *[17, 定理 2]* 中的收敛结果，因为它所要求的限制等距性质和步长假设更弱。特别是，假设$\delta_{3s} < 0.618$ 和 (3.15) 比 $\delta_{3s} < \frac{1}{2}$ 和 $1\ \frac{}{2(1-\delta_{3s})} < v \leq \frac{1}{1-\delta_{3s}}$在 *[17, 定理 2]*中采用的假设更弱。

(iii) 根据 (3.16)，我们得到 ITAT 的复杂度为

$$\|x^{k^*}-\bar{x}\|_2 \leq \frac{\sqrt{5}+3-2\rho}{1-\rho}v(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda), \tag{3.25}$$

至多经过 $k^* := \lceil \log_{\rho^{-1}}{}_{v}\ \frac{\|x^0-\bar{x}\|}{(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda)}\rceil$ 次迭代即可达到。事实上，根据 $k^*$ 的定义，我们有

$$\rho^{k^*}\|x^0-\bar{x}\|_2 \leq v(\sqrt{1+\delta_{2s}}\|\varepsilon\|_2+\sqrt{2s}\lambda).$$

由此结合 (3.16)，可得 (3.25)。

### 3.3 采用延拓技术的迭代阈值算法

请注意，正则化参数 $\lambda$ 对稀疏优化算法的数值性能起着重要作用。根据恢复界理论，正则化参数 $\lambda$ 应当较小，以保证更好的恢复效果；然而，计算数学理论和大量数值研究表明，参数过小会导致子问题不适定，而适当增大参数则能加快收敛速度。

为了兼顾理论和数值方面的优势，延拓技术是稀疏优化算法中一种常用的加速策略：它采用从较大值开始、按几何级数递减的正则化参数序列$\{\lambda_k\}$，取代固定但较小的正则化参数；例如，参见 [16, 31, 33, 37]。

受 $\ell_{1\text{-}2}$ 阈值算子 (3.4) 和延拓技术的启发，我们提出了采用延拓技术的迭代阈值算法（ITAC），以求解 $\ell_{1\text{-}2}$ 正则化问题 (2.2)。

算法 3.2（ITAC）。选取正则化参数 $\lambda_0 > 0$, $\lambda > 0$ 和延拓参数 $\gamma \in (0, 1)$，并设置初始点 $x^0 := 0$ 和步长 $v > 0$。对于每个 $k \in \mathbb{N}$，已知 $x^k$ 和 $\lambda_k$，若 $\lambda_k < \lambda$，则停止并输出 $x^* := x^k$；否则，更新正则化参数 $\lambda_{k+1} := \gamma\lambda_k$，并通过以下方式确定 $x^{k+1}$

$$x_i^{k+1} := \mathbb{T}_{v\lambda_k}\left(x^k - vA^\top(Ax^k - b)\right). \tag{3.26}$$

备注 3.6。ITAC 与 ITA 的唯一区别在于使用了正则化参数的动态序列。ITAC 继承了公式简单、计算复杂度低等显著优点，因此对于大规模稀疏优化问题极为高效。

本小节的主要定理如下。该定理给出了 ITAC 中某些参数的设置（与 RIP 相关），以保证其在容差与噪声水平成正比的条件下，收敛到式 (1.1) 的近似真实稀疏解。此外，ITAC 输出的支撑集不存在错误预测，并且恰好是真实稀疏解支撑集的子集。

定理 3.2. 假设 $A$满足限制等距性质，且

$$(\sqrt{s}+1)\delta_{s+1} < 1. \tag{3.27}$$

*Let*

$$\eta \in \left(0, 1-(\sqrt{s}+1)\delta_{s+1}\right), \tag{3.28}$$

并且步长 $v \in (0, 1]$ 以及

$$\lambda_0 \geq \frac{\|\bar{x}\|_2}{\sqrt{s}+1}, \quad \lambda := \frac{\sqrt{1+\delta_s}}{\eta}\|\varepsilon\|_2 \quad \textit{and} \quad \gamma \in \left[\frac{(\sqrt{s}+1)v\delta_{s+1}}{1-\eta} + 1 - v, 1\right). \tag{3.29}$$

令算法 3.2 使用这些参数输出 $x^*$。则有

$$\mathrm{supp}(x^*) \subseteq \mathcal{S} \quad \textit{and} \quad \|x^* - \bar{x}\|_2 \leq \frac{(1-\eta)\sqrt{1+\delta_s}}{\eta\delta_{s+1}}\|\varepsilon\|_2. \tag{3.30}$$

证明。根据假设 (3.27)，可验证 $1-(\sqrt{s}+1)\delta_{s+1} > 0$，因此 (3.28) 中的 $\eta$ 有定义。由此可得 $(\sqrt{s}+1)\frac{v\delta_{s+1}}{1-\eta} + 1 - v < 1$。因此，~~(3.29) 中的~~ $\gamma$ 有定义，采用这些参数的算法 3.2 也同样有定义。

为完善证明，令算法 3.2 生成有限序列 $\{x^k\}_{k=0}^K$ 并输出 $x^* = x^K$。记

$$\rho := \frac{1-\eta}{\delta_{s+1}}, \tag{3.31}$$

and

$$y^k := x^k - vA^\top(Ax^k - b), \quad \mathcal{S}_k := \mathrm{supp}(x^k) \quad \text{and} \quad r^k := x^k - \bar{x} \tag{3.32}$$

对每个 $k \in [K]$。我们将用归纳法证明，对每个 $k \in [K]$，以下包含关系和估计成立:

$$\mathcal{S}_k \subseteq \mathcal{S} \quad \text{and} \quad \|r^k\|_2 \le \rho\lambda_k. \tag{3.33}$$

根据初始选择 $x^0 := 0$，可得 $\mathcal{S}_0 = \emptyset \subseteq \mathcal{S}$。由 (3.29) 中的 $\lambda_0$ 假设以及 (3.31) 中的 $\rho$ 定义，再由 (3.28) 可得

$$\rho\lambda_0 \ge \frac{1-\eta}{\delta_{s+1}} \frac{\|\bar{x}\|_2}{\sqrt{s}+1} > \|\bar{x}\|_2 = \|r^0\|_2.$$

可以证明，(3.33) 对 $k = 0$ 成立。

现在假设 (3.33) 对迭代点 $k$ ($< K$) 成立。由 (3.32) 和 (3.13)，可得

$$y^k = x^k - vA^\top(Ax^k - A\bar{x} - \varepsilon) = x^k - vA^\top A_{\mathcal{S}} r^k_{\mathcal{S}} + vA^\top\varepsilon, \tag{3.34}$$

第二个等号由 (3.33) 中的假设 $\mathcal{S}_k \subseteq \mathcal{S}$ 得到。固定 $i \in S^c$。由 (3.33) 中的假设 $\mathcal{S}_k \subseteq \mathcal{S}$ 可得 $x_i^k = 0$，于是 (3.34) 化为

$$|y_i^k| \le v|A_i^\top A_{\mathcal{S}} r^k_{\mathcal{S}}| + v|A_i^\top\varepsilon|. \tag{3.35}$$

由于 $\{i\} \cap \mathcal{S} = \emptyset$，由引理 3.3(iii) 和 (ii) 可得

$$|A_i^\top A_{\mathcal{S}} r^k_{\mathcal{S}}| \le \delta_{s+1}\|(x^k - \bar{x})_{\mathcal{S}}\|_2 = \delta_{s+1}\|r^k\|_2 \quad \text{and} \quad |A_i^\top\varepsilon| \le \sqrt{1+\delta_1}\|\varepsilon\|_2 \le \sqrt{1+\delta_s}\|\varepsilon\|_2$$

（由于 $\mathcal{S}_k \subseteq \mathcal{S}$ 以及 $\delta_1 \le \delta_s$的单调不减性质）。结合 (3.35)，可得

$$|y_i^k| \le v\delta_{s+1}\|r^k\|_2 + v\sqrt{1+\delta_s}\|\varepsilon\|_2. \tag{3.36}$$

根据停止准则 $\lambda_k < \lambda$ 以及 (3.29) 中 $\lambda$ 的定义，可知 $\lambda_k \ge \lambda = \frac{\sqrt{1+\delta_s}}{\eta}\|\varepsilon\|_2$，即 $\|\varepsilon\|_2 \le \frac{\eta\lambda_k}{\sqrt{1+\delta_s}}$。结合假设 (3.33)，由此可将 (3.36) 推得

$$|y_i^k| \le v\delta_{s+1}\|r^k\|_2 + v\sqrt{1+\delta_s}\|\varepsilon\|_2 \le v\delta_{s+1}\rho\lambda_k + v\eta\lambda_k = v\lambda_k, \tag{3.37}$$

其中，等式由 (3.31) 中 $\rho$ 的定义成立。因此，由 (3.7) 可得 $x_i^{k+1} = 0$；进而，$i \in \mathcal{S}_{k+1}^c$。这对任意 $i \in \mathcal{S}^c$均成立。于是得到 $\mathcal{S}^c \subseteq \mathcal{S}_{k+1}^c$，等价地，$\mathcal{S}_{k+1} \subseteq \mathcal{S}$。

另一方面，由包含关系 $\mathcal{S}_{k+1} \subseteq \mathcal{S}$ 可得

$$\|x^{k+1} - \bar{x}\|_2 = \|x^{k+1}_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2 \le \|x^{k+1}_{\mathcal{S}} - y^k_{\mathcal{S}}\|_2 + \|y^k_{\mathcal{S}} - \bar{x}_{\mathcal{S}}\|_2. \tag{3.38}$$

由 (1.4) 并结合算法 3.2 可知 $x^{k+1}=\mathbb{T}_{v\lambda_k}(y^k)$，因此我们得到

$$\|x^{k+1}_{\mathcal{S}}-y^k_{\mathcal{S}}\|_2\le\sqrt{s}\|x^{k+1}-y^k\|_\infty=\sqrt{s}\|\mathbb{T}_{v\lambda_k}(y^k)-y^k\|_\infty\le v\sqrt{s}\lambda_k, \tag{3.39}$$

（由引理 3.2），并由 (3.34) 可知

$$\|y^k_{\mathcal{S}}-\bar{x}_{\mathcal{S}}\|_2=\|((I-vA^\top A)r^k)_{\mathcal{S}}+vA^\top_{\mathcal{S}}\varepsilon\|_2\le\|((I-vA^\top A)r^k)_{\mathcal{S}}\|_2+v\|A^\top_{\mathcal{S}}\varepsilon\|_2. \tag{3.40}$$

注意到 $v\le 1$，由引理 3.3(i) 和 (ii) 可得

$$\|((I-vA^\top A)r^k)_{\mathcal{S}}\|_2\le(1-v+v\delta_s)\|r^k\|_2\quad\text{and}\quad\|A^\top_{\mathcal{S}}\varepsilon\|_2\le\sqrt{1+\delta_s}\|\varepsilon\|_2,$$

分别成立。这些结果连同 (3.38)-(3.40) 表明

$$\|x^{k+1}-\bar{x}\|_2\le v\sqrt{s}\lambda_k+(1-v+v\delta_s)\|r^k\|_2+v\sqrt{1+\delta_s}\|\varepsilon\|_2 \tag{3.41}$$

由 $\delta_s\le\delta_{s+1}$ 这一事实以及 (3.37)，可得

$$v\delta_s\|r^k\|_2+v\sqrt{1+\delta_s}\|\varepsilon\|_2\le v\delta_{s+1}\|r^k\|_2+v\sqrt{1+\delta_s}\|\varepsilon\|_2\le v\lambda_k.$$

因此，由 (3.33)，(3.41) 可化为

$$\|x^{k+1}-\bar{x}\|_2\le\left(\frac{v(\sqrt{s}+1)}{\rho}+1-v\right)\rho\lambda_k. \tag{3.42}$$

注意到，由 (3.31) 中 $\rho$ 的定义可知

$$\frac{v(\sqrt{s}+1)}{\rho}+1-v=\frac{(\sqrt{s}+1)v\delta_{s+1}}{1-\eta}+1-v\le\gamma$$

（由（3.29）中 $\gamma$ 的定义），（3.42）可化简为 $\|x^{k+1}-\bar{x}\|\le\rho\gamma\lambda_k=\rho\lambda_{k+1}$。结合 $\mathcal{S}_{k+1}\subseteq\mathcal{S}$，可知（3.33）对每个迭代点 $k\in[K]$ 均成立。于是，由（3.33）可得 supp($x^*$) $\subseteq\mathcal{S}$ 且

$$\|x^*-\bar{x}\|_2\le\rho\lambda_K<\rho\lambda=\frac{(1-\eta)\sqrt{1+\delta_s}}{\delta_{s+1}\eta}\|\varepsilon\|_2$$

由（3.29）和（3.31）中 $\lambda$ 和 $\rho$ 的定义可得。证明完毕。 ■

注 3.7. *[19]*证明了在 MIP 假设下，采用延拓技术的 ISTA（IS-TAC）求解 $\ell_1$ 正则化问题时，会收敛到（1.1）的一个近似真实稀疏解。定理 3.2 在以下几个方面扩展并改进了 *[19,* 定理 *2]*：

- 定理 3.2 从以下意义上扩展了 *[19,* 定理 *2]*：（a）$\ell_{1\text{-}2}$ 正则化问题（2.2）是 *[19]* 中所考虑的 $\ell_1$ 正则化问题的扩展；（b）定理 3.2 考虑了 $A$ 为一般感知矩阵且步长为 $v\in(0,1]$）的情形，这是对 *[19]* 的扩展；后者假设 $A$ 列归一化且 $v=1$。

- 定理 3.2 改进了 *[19,* 定理 *2]*，因为它只要求满足 RIP 假设，而该假设弱于 *[19,* 定理 *2]* 中的 MIP 假设。

# 4 数值实验

本节旨在通过将 ITAT 和 ITAC 与若干常用的稀疏优化算法进行比较，对它们求解 $\ell_{1\text{-}2}$ 正则化问题 (2.2) 的表现开展数值实验。所有数值实验均在个人台式机（Intel Core Duo i7-8700，3.20GHz，32.00 GB 内存）上使用 R (3.5.2) 和 MATLAB R2018a 完成。

模拟数据生成如下。感知矩阵 $A \in \mathbb{R}^{m\times n}$ 由两种随机矩阵生成：

(i) 随机高斯矩阵，其定义为

$$A_i \backsim \mathcal{N}(0, \mathbb{I}_m/m).$$

(ii) 随机部分离散余弦变换（PDCT）矩阵，即

$$A_i := \frac{1}{\sqrt{m}}\cos(2i\pi\xi),$$

其中， $\xi \in \mathbb{R}^m \backsim \mathcal{U}([0,1]^m)$从 $[0,1]$中均匀且独立地抽取。

随机高斯矩阵和 PDCT 矩阵适用于压缩感知，因为它们具有非相干性，并且以高概率具有较小的RIP常数 [3, 6]。

真实稀疏解 $\bar{x} \in \mathbb{R}^n$ 通过随机选取其 $s$ 个分量作为非零分量生成，这些分量的取值也随机生成为独立同分布高斯变量，其余分量均设为零。观测矩阵 $b$ 由 $b = A\bar{x} + \sigma\,\epsilon$ 生成，其中 $\epsilon \in \mathbb{R}^m$ 为独立同分布高斯噪声。在数值实验中，问题规模设为 $(m, n)$ =(256, 1024)，加性高斯噪声的标准差设为 $\sigma = 0.1\%$。

在 ITAT 和 ITAC 的实现中，步长和初始点默认分别设为 $v_k \equiv 0.5$ 和 $X^0 = 0$ ，停止准则设为 $\frac{\|x^k - x^{k-1}\|_2}{\|x^{k-1}\|_2} \le 10^{-6}$ ，或迭代次数超过 500。算法的性能通过以下指标进行评估：

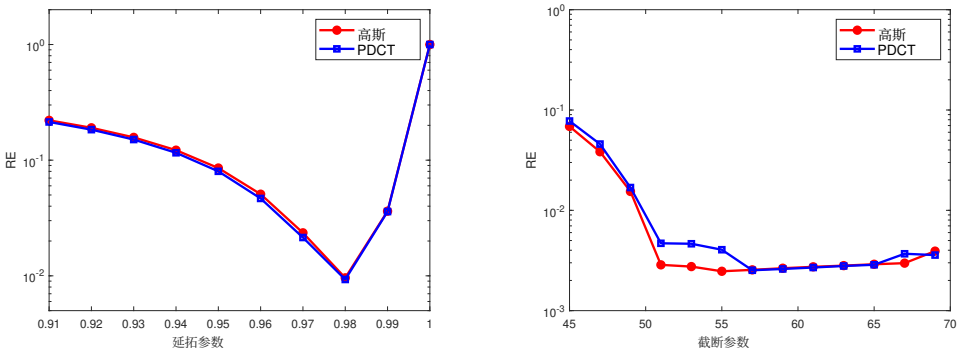
-（精度）与真实解的相对误差（相对误差）：

$$\mathrm{RE} := \frac{\|x - \bar{x}\|_2}{\|\bar{x}\|_2}.$$

-（鲁棒性）在 500 次随机试验中，相对误差 $< 0.01$ 的成功恢复比例 s.

第一个实验旨在展示 ITAC 在不同延拓参数 $\gamma$ 下，以及 ITAT 在不同截断参数 $s$ 下的数值表现。在该实验中，真解的稀疏性设为 $\bar{s} = 51$。图 1(a) 和 (b) 分别绘出了 500 次随机试验中，ITAC 在 $\gamma$ 从 (0.9, 1) 变化时，以及 ITAT 在 $s$ 从 (45,70) 变化时所生成解的平均相对误差。结果表明

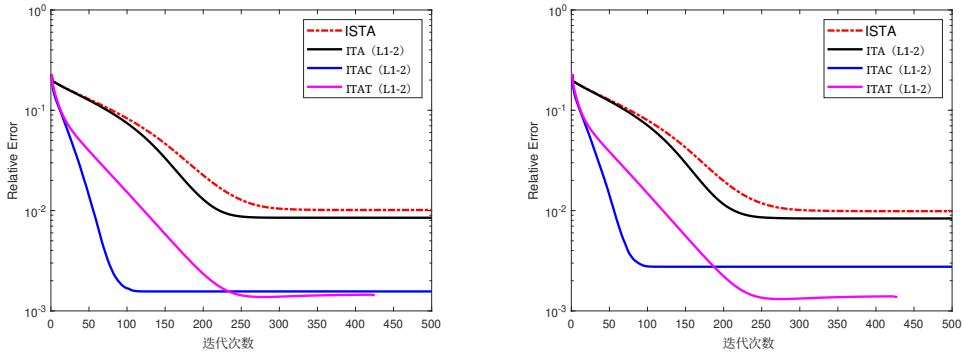
从图 1(a) 可以看出，当 $\gamma = 0.98$ 时，ITAC 在获得精确估计方面表现最佳。图 1(b) 表明，当$s < \bar{s}$ 时，ITAT 无法获得精确估计；当 $s \geq \bar{s}$ 略微变化（在 20% 以内）时，则接近于获得精确估计。因此，在接下来的数值实验中，我们分别为 ITAC 和 ITAT 设置延拓参数 $\gamma = 0.98$ 和截断参数 $s = \bar{s}$ 。




(a) $\gamma$ 在 ITAC 中。 (b) $s$ 在 ITAT 中。

图 1：不同参数下 ITAC 和 ITAT 的数值结果。

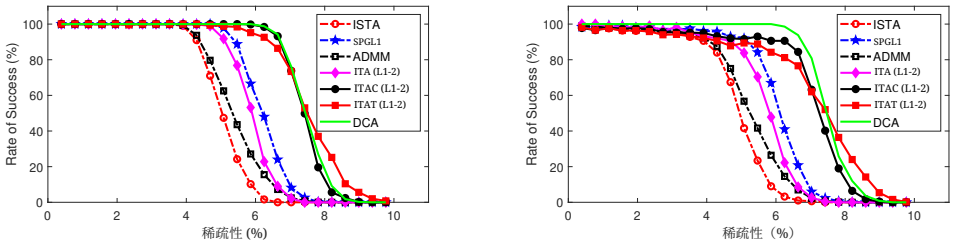
第二个实验旨在比较 ITAC 和 ITAT 与标准 ITA [20] 在 $\ell_{1\text{-}2}$ 正则化下，以及 ISTA [9] 在 Lasso 下的收敛表现。在该实验中，真解的稀疏比设为 $s/n := 5\%$。图2绘出了这些算法在高斯矩阵和 PDCT 矩阵上的 500 次随机试验中，相对误差随迭代次数变化的平均值。从图2可以看出，与标准 ITA 和 ISTA 相比，ITAC 和 ITAT 收敛更快，且得到的解更准确。这验证了 PGA 中用于稀疏优化的延拓技术和截断技术具有加速能力，并能收敛到近似真解；同时也体现了 $\ell_{1\text{-}2}$ 正则化相较于 Lasso 的优势。




（a）高斯 （b）PDCT

图2：稀疏优化中迭代阈值算法的收敛行为。

第三个实验旨在比较 ITAC 和 ITAT 与标准 ITA [20] 以及 DCA [36] 在 $\ell_{1\text{-}2}$ 正则化下的稳定性，并与 Lasso 的常用算法进行比较：ISAT [9], 、ADMM [35] 和谱投影梯度法（SPGL1） [30]。图3绘出了成功恢复率，表1记录了这些算法在高斯矩阵和 PDCT 矩阵上针对每个稀疏度进行的 500 次随机试验中的平均 CPU 时间（秒）。从图3可以看出，与标准 ITA 以及 Lasso 的常用算法相比，ITAC 和 ITAT 的成功恢复率更高。尽管 DCA 的成功恢复率略高于 ITAC 和 ITAT，但表1显示，DCA 消耗的 CPU 时间远超其他算法，包括 ITAC 和 ITAT，这是因为 DCA 使用 CVX 软件求解凸子问题时耗时更多。




(a) 高斯　　(b) PDCT

图3：稀疏优化中迭代阈值算法及其他算法的成功恢复率。

表1：稀疏优化中迭代阈值算法及其他算法的 CPU 时间（秒）。

| 问题 | Lasso | | | $\ell_{1\text{-}2}$ 正则化 | | | |
|---|---|---|---|---|---|---|---|
| 算法 | ITA | SGPL1 | AMDD | ITA | ITAC | ITAT | DCA |
| CPU 时间 | 0.34 | 0.81 | 0.07 | 0.35 | 0.28 | 0.34 | 2.34 |

总之，初步数值结果表明，基于$\ell_{1\text{-}2}$ 正则化的 ITAC 和 ITAT 具有很强的稀疏促进能力，在精度和鲁棒性方面均优于基于 $\ell_{1\text{-}2}$ 正则化的标准 ITA 以及用于 Lasso 的常用算法，这得益于 $\ell_{1\text{-}2}$ 惩罚项以及延拓或截断技术。

## 附录A 命题 2.1 的证明

以下　引理对于证明命题 2.1 很有用，该引理取自 [29, 引理 3.1　　].

引理 A.1. 设 $x$、 $y \in \mathbb{R}^n$和 $\gamma \in (0, 1)$满足 $-\langle x$、 $y\rangle \leq \gamma\|x\|_2^2$。则 $(1-\gamma)\ \|x\|_2 \leq \|x+y\|_2$。

命题 2.1 的证明。对于 $\ell_{1\text{-}2}$-受限特征值条件$(s, t)$，我们定义可行集

$$C(s) := \{x \in R^n : \|x_{\mathcal{I}^c}\|_1 \leq \|x_{\mathcal{I}}\|_1 + \|x\|_2 \text{ for some } |\mathcal{I}| \leq s\}.$$

固定 $x \in C(s)$。那么存在 $\mathcal{I} \subseteq [n]$ 使得

$$|\mathcal{I}| \le s \quad \text{and} \quad \|x_{\mathcal{I}^c}\|_1 \le \|x_{\mathcal{I}}\|_1 + \|x\|_2. \tag{A.1}$$

按 (2.7) 记 $\mathcal{I}_k := \mathcal{I}_k(x;t)$，为简便起见，记 $\mathcal{J} := \mathcal{I} \cup \mathcal{I}_0$。因此

$$\{\mathcal{J}, \mathcal{I}_1, \ldots, \mathcal{I}_r\} \text{ is a disjoint decomposition of } [n] \text{ with } |\mathcal{I}_k| \le t \text{ and } |\mathcal{J}| \le s+t. \tag{A.2}$$

(i) 由引理 2.1 和 (A.1) 可得

$$\sum_{k=1}^{r} \|x_{\mathcal{I}_k}\|_2 \le \frac{1}{\sqrt{t}} \|x_{\mathcal{I}^c}\|_1 \le \frac{1}{\sqrt{t}} \|x_{\mathcal{I}}\|_1 + \frac{1}{\sqrt{t}} \|x\|_2. \tag{A.3}$$

注意，由 (1.4)（以 $x_{\mathcal{I}}$ 代替 $x$）可得

$$\|x_{\mathcal{I}}\|_1 \le \sqrt{s} \|x_{\mathcal{I}}\|_2 \le \sqrt{s} \|x_{\mathcal{J}}\|_2, \tag{A.4}$$

再由 $\ell_2$ 范数的次可加性和 (A.2) 可得 $\|x\|_2 \le \sum_{k=1}^{r} \|x_{\mathcal{I}_k}\|_2 + \|x_{\mathcal{J}}\|_2$。于是 (A.3) 可化为

$$(\sqrt{t}-1) \sum_{k=1}^{r} \|x_{\mathcal{I}_k}\|_2 \le (\sqrt{s}+1) \|x_{\mathcal{J}}\|_2. \tag{A.5}$$

注意，由 (A.2) 可知，对每个 $k \in \mathbb{N}$，$|\mathcal{J}| \le s+t$ 且 $|\mathcal{I}_k| \le t$，因此由 (2.6) 可得

$$\sigma_{\min}(s+t) \|x_{\mathcal{J}}\|_2 \le \|Ax_{\mathcal{J}}\|_2 \quad \text{and} \quad \|Ax_{\mathcal{I}_k}\|_2 \le \sigma_{\max}(t) \|x_{\mathcal{I}_k}\|_2.$$

由于 $\ell_2$ 范数具有次可加性，由此并结合 (A.5) 可得

$$\|Ax\|_2 \ge \|Ax_{\mathcal{J}}\|_2 - \sum_{k=1}^{r} \|Ax_{\mathcal{I}_k}\|_2 \ge \left( \sigma_{\min}(s+t) - \frac{\sqrt{s}+1}{\sqrt{t}-1} \sigma_{\max}(t) \right) \|x_{\mathcal{J}}\|_2.$$

由于满足 (A.1) 的 $x$ 和 $\mathcal{I}$ 是任意的，因此可以证明

$$\phi(s,t) \ge \sigma_{\min}(s+t) - \frac{\sqrt{s}+1}{\sqrt{t}-1} \sigma_{\max}(t) > 0$$

根据假设 (i)。

(ii) 由定义 2.2(ii)（参见 (2.4)）及 (A.2)，可得

$$|\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c} \rangle| \le \sum_{k=1}^{r} |\langle Ax_{\mathcal{J}}, Ax_{\mathcal{I}_k} \rangle| \le \theta_{t,s+t} \|x_{\mathcal{J}}\|_2 \sum_{k=1}^{r} \|x_{\mathcal{I}_k}\|_2.$$

由（A.5）可得

$$|\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c} \rangle| \le \theta_{t,s+t} \frac{\sqrt{s}+1}{\sqrt{t}-1} \|x_{\mathcal{J}}\|_2^2 \le \frac{\theta_{t,s+t}(\sqrt{s}+1)}{(\sqrt{t}-1)(1-\delta_{s+t})} \|Ax_{\mathcal{J}}\|_2^2 \tag{A.6}$$

（根据（2.3））。根据本命题的假设（ii）可验证

$$0 < \frac{\theta_{t,s+t}(\sqrt{s}+1)}{(\sqrt{t}-1)(1-\delta_{s+t})} < 1. \tag{A.7}$$

结合 (A.6)，可知引理 A.1 适用（以 $Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c}$，$\theta_{t,\ s+t}(\sqrt{s}+1)\,(\sqrt{t}-1)(1-\delta_{s+t})$代替 ~~$x, y, \gamma$~~），从而得出

$$\|Ax\|_2 \geq \left(1 - \frac{\theta_{t,s+t}(\sqrt{s}+1)}{(\sqrt{t}-1)(1-\delta_{s+t})}\right)\|Ax_{\mathcal{J}}\|_2 \geq \sqrt{1-\delta_{s+t}}\left(1 - \frac{\theta_{t,s+t}(\sqrt{s}+1)}{(\sqrt{t}-1)(1-\delta_{s+t})}\right)\|x_{\mathcal{J}}\|_2$$

（由 (2.3) 得）。由于满足 (A.1) 的 $x$ 和 $I$ 是任意的，根据定义 2.1 和 (A.7)，有

$$\phi(s,t) \geq \sqrt{1-\delta_{s+t}}\left(1 - \frac{\theta_{t,s+t}(\sqrt{s}+1)}{(\sqrt{t}-1)(1-\delta_{s+t})}\right) > 0.$$

（iii）注意到

$$\|Ax\|_2^2 = \|Ax_{\mathcal{J}}\|_2^2 + 2\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c}\rangle + \|Ax_{\mathcal{J}^c}\|_2^2 \geq \|Ax_{\mathcal{J}}\|_2^2 - 2|\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c}\rangle|. \tag{A.8}$$

特别地，由 $\mathcal{I} \subseteq \mathcal{J}$、(A.1) 和 (A.4) 可得

$$\|x_{\mathcal{J}^c}\|_1 \leq \|x_{\mathcal{I}^c}\|_1 \leq \|x_{\mathcal{I}}\|_1 + \|x\|_2 \leq (\sqrt{s}+1)\|x_{\mathcal{J}}\|_2 + \|x_{\mathcal{J}^c}\|_2. \tag{A.9}$$

注意到由 (A.5) 可知，$\|x_{\mathcal{J}^c}\|_2 \leq \sum_{k=1}^{r}\|x_{\mathcal{I}_k}\|_2 \leq \frac{\sqrt{s}+1}{\sqrt{t}-1}\|x_{\mathcal{J}}\|_2$、(A.9) 可化简为

$$\|x_{\mathcal{J}^c}\|_1 \leq \frac{\sqrt{t}(\sqrt{s}+1)}{\sqrt{t}-1}\|x_{\mathcal{J}}\|_2. \tag{A.10}$$

此外，由 (2.4) 可得 $\|A_{\mathcal{J}^c}^{\top}Ax_{\mathcal{J}}\|_\infty \leq \theta_{s+t,\ 1}\|x_{\mathcal{J}}\|_2$。结合以上两个不等式，可得

$$|\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c}\rangle| \leq \|x_{\mathcal{J}^c}\|_1\|A_{\mathcal{J}^c}^{\top}Ax_{\mathcal{J}}\|_\infty \leq \theta_{s+t,1}\frac{\sqrt{t}(\sqrt{s}+1)}{\sqrt{t}-1}\|x_{\mathcal{J}}\|_2^2. \tag{A.11}$$

另一方面，由式 (2.6) 可得

$$\|Ax_{\mathcal{J}}\|_2^2 \geq \sigma_{\min}(s+t)\|x_{\mathcal{J}}\|_2^2. \tag{A.12}$$

由式 (A.11) 和 (A.12)，式 (A.8) 可推出

$$\|Ax\|_2^2 \geq \left(\sigma_{\min}(s+t) - 2\theta_{s+t,1}\frac{\sqrt{t}(\sqrt{s}+1)}{\sqrt{t}-1}\right)\|x_{\mathcal{J}}\|_2^2.$$

由于满足 (A.1) 的 $x$ 和 $\mathcal{I}$ 是任意的，由假设 (iii) 可知

$$\phi(s,t) \geq \left(\sigma_{\min}(s+t) - 2\theta_{s+t,1}\frac{\sqrt{t}(\sqrt{s}+1)}{\sqrt{t}-1}\right)^{\frac{1}{2}} > 0.$$

(iv) 注意，根据互相不相干性 (2.5),

$$|\langle Ax_{\mathcal{J}}, Ax_{\mathcal{J}^c}\rangle| \le \mu \sum_{i\in\mathcal{J}^c}\sum_{j\in\mathcal{J}} |x_i x_j| \le \mu \|x_{\mathcal{J}}\|_1 \|x_{\mathcal{J}^c}\|_1 \le \mu \frac{\sqrt{t(s+t)}(\sqrt{s}+1)}{\sqrt{t}-1}\|x_{\mathcal{J}}\|_2^2. \quad \text{(A.13)}$$

（由 (1.4) 和 (A.10)）。将其与 (A.8) 和 (A.12) 结合，可得

$$\|Ax\|_2^2 \ge \left(\sigma_{\min}(s+t) - 2\mu\frac{\sqrt{t(s+t)}(\sqrt{s}+1)}{\sqrt{t}-1}\right)\|x_{\mathcal{J}}\|_2^2.$$

由于满足 (A.1) 的 $x$ 和 $\mathcal{I}$ 是任意的，根据假设 (iv) 可知

$$\phi(s,t) \ge \left(\sigma_{\min}(s+t) - 2\mu\frac{\sqrt{t(s+t)}(\sqrt{s}+1)}{\sqrt{t}-1}\right)^{\frac{1}{2}} > 0.$$

（v）将 $x_{\mathcal{J}}^\top A^\top A x_{\mathcal{J}}$ 的对角项与非对角项分开，可得

$$\|Ax_{\mathcal{J}}\|_2^2 = \sum_{i\in\mathcal{J}}\langle A_i x_i, A_i x_i\rangle + \sum_{i,j\in\mathcal{J}:i\neq j}\langle A_i x_i, A_j x_j\rangle. \quad \text{(A.14)}$$

根据对每个 $i\in[n]$ 均有 $\|A_i\|=1$ 的假设，可得 $\sum_{i\in\mathcal{J}}\langle A_i x_i, A_i x_i\rangle = \|x_{\mathcal{J}}\|_2^2$。由互相不相干性 (2.5)，可得

$$\sum_{i,j\in\mathcal{J}:i\neq j}\langle A_i x_i, A_j x_j\rangle \ge -\mu\sum_{i,j\in\mathcal{J}:i\neq j}|x_i x_j| \ge -\mu\|x_{\mathcal{J}}\|_1^2 \ge -(s+t)\mu\|x_{\mathcal{J}}\|_2^2.$$

因此，(A.14) 可化简为 $\|Ax_{\mathcal{J}}\|_2^2 \ge (1-(s+t)\mu)\|x_{\mathcal{J}}\|_2^2$。结合 (A.8) 和 (A.13)，可得

$$\|Ax\|_2^2 \ge \left(1-(s+t)\mu - 2\mu\frac{\sqrt{t(s+t)}(\sqrt{s}+1)}{\sqrt{t}-1}\right)\|x_{\mathcal{J}}\|_2^2.$$

由于满足 (A.1) 的 $x$ 和 $\mathcal{I}$ 是任意的，因此有

$$\phi(s,t) \ge \left(1-(s+t)\mu - 2\mu\frac{\sqrt{t(s+t)}(\sqrt{s}+1)}{\sqrt{t}-1}\right)^{\frac{1}{2}} > 0$$

由假设 (v) 可得。证明完毕。 ■

## 参考文献

[1] A. Beck 和 M. Teboulle。用于线性逆问题的快速迭代收缩阈值算法。SIAM 成像科学期刊，2(1):183–202，2009。

[2] P. J. Bickel、Y. Ritov 和 A. B. Tsybakov。Lasso 与 Dantzig 选择器的同时分析。统计年刊，37:1705–1732，2009。[3] J. D. Blanchard、C. Cartis 和 J. Tanner。压缩感知：限制等距性质有多精确？《SIAM评论》，53(1):105–125，2011。[4] T. Blumensath 和 M. E. Davies。压缩感知的迭代硬阈值。《应用与计算谐波分析》，27(3):265–274，2009。[5] F. Bunea、A. Tsybakov 和 M. Wegkamp。Lasso 的稀疏性预言机不等式。《电子统计学期刊》，1:169–194，2007。[6] E. Cand`es、J. K. Romberg 和 T. Tao。利用不完整且不准确的测量实现稳定信号恢复。《纯粹与应用数学通讯》，59(8):1207–1223，2006。[7] E. Cand`es 和 T. Tao。线性规划解码。《IEEE信息论汇刊》，51:4203–4215，2005。[8] X. Chen、F. Xu 和 Y. Ye。 $\ell_2$-$\ell_p$最小化解中非零元素的下界理论。SIAM 科学计算期刊，32(5):2832–2852，2010。[9] I. Daubechies、M. Defrise 和 C. De Mol。带稀疏性约束的线性逆问题的迭代阈值算法。《纯粹与应用数学通讯》，57:1413–1457，2004。[10] D. L. Donoho。对于大多数大型欠定线性方程组，最小$\ell_1$-范数解也是最稀疏解。《纯粹与应用数学通讯》，59(6):797–829，2006。[11] D. L. Donoho 和 X. M. Huo。不确定性原理与理想原子分解。《IEEE信息论汇刊》，47(7):2845–2862，2001。[12] E. Esser、Y. Lou 和 J. Xin。一种求解非负最小二乘问题结构化稀疏解的方法及其应用。SIAM 成像科学期刊，6(4):2010–2046，2013。[13] J. Fan 和 R. Li。通过非凹惩罚似然进行变量选择及其预言机性质。《美国统计学会杂志》，96(456):1348–1360，2001。[14]S. Foucart 和 H. Rauhut。《压缩感知数学导论》。施普林格，纽约，2013。[15]H. Ge、J. Wen 和 W. Chen。截断 $\ell_{1-2}$-最小化的零空间性质。IEEE 信号处理快报，25(8):1261–1265，2018。

[16]E. T. Hale、W. Yin 和 Y. Zhang。不动点延拓用于 $\ell_1$-最小化：方法论与收敛性。SIAM 优化期刊，19(3)：1107–1130，2008。[17] Y. Hu、X. Hu 和 X. Yang。关于迭代阈值算法收敛至复合非凸优化近似稀疏解的问题。数学规划，211：181–206，2025。[18] Y. Hu、C. Li、K. Meng、J. Qin 和 X. Yang。通过 $\ell_{p,\ q}$ 正则化进行组稀疏优化。机器学习研究期刊，18(30)：1–52，2017。[19] Y. Jiao、B. Jin 和 X. Lu。结合同伦延拓的迭代软/硬阈值法用于稀疏恢复。IEEE 信号处理快报，24(6)：784–788，2017。[20] Y. Lou 和 M. Yan。通过近端算子快速求解 L1-L2 最小化。科学计算期刊，74：767–785，2018。[21] Y. Lou、P. Yin、Q. He 和 J. Xin。基于 $l_1$ 与 $l_2$ 之差，在高度相干字典中计算稀疏表示。科学计算期刊，64：178–196，2015。[22] Y. Lou、P. Yin 和 J. Xin。基于非凸 $l_1$ 方法的点源超分辨率。科学计算期刊，68：1082–1100，2016。[23]Y. Lou、T. Zeng、S. Osher 和 J. Xin。用于图像处理的各向异性与各向同性全变分加权差分模型。SIAM 成像科学期刊，8(3)：1798–1823，2015。[24] T.-H. Ma、Y. Lou 和 T.-Z. Huang。用于稀疏恢复和秩最小化的截断 $l_{1-2}$ 模型。SIAM 成像科学期刊，10(3)：1346–1380，2017。[25] N. Meinshausen 和 B. Yu。高维数据中稀疏表示的 Lasso 型恢复。统计年刊，37：246–270，2009。[26] B. Natarajan。线性系统的稀疏近似解。SIAM 计算期刊，24(2)：227–234，1995。[27] D. Needell 和 J. Tropp。CoSaMP：从不完整且不准确的样本中迭代恢复信号。《应用与计算谐波分析》，26(3)：301–321，2009。[28] J. Shen 和 P. Li。硬阈值的紧界。机器学习研究期刊，18：1–42，2018。[29] S. A. van de Geer 和 P. B¨uhlmann。关于用于证明 Lasso 预言机结果的条件。《电子统计学期刊》，3：1360–1392，2009。[30] E. van den Berg 和 M. P. Friedlander。探索基追踪解的帕累托前沿。SIAM 科学计算期刊，31(2)：890–912，2008。

[31] Z. Wen、W. Yin、D. Goldfarb 和 Y. Zhang。一种基于收缩、子空间优化和延拓的快速稀疏重构算法。SIAM 科学计算期刊，32(4)：1832–1857，2010。[32] S. J. Wright。坐标下降算法。数学规划，151(1)：3–34，2015。[33] L. Xiao 和 T. Zhang。用于稀疏最小二乘问题的近端梯度同伦方法。SIAM 优化期刊，23(2)：1062–1091，2013。[34] Z. Xu、X. Chang、F. Xu 和 H. Zhang。 $L_{1/2}$ 正则化：阈值表示理论与快速求解器。IEEE神经网络与学习系统汇刊，23：1013–1027，2012。[35] J. Yang 和 Y. Zhang。用于压缩感知中 $\ell_1$ 问题的交替方向算法。SIAM 科学计算期刊，33(1)：250–278，2011。[36] P. Yin、Y. Lou、Q. He 和 J. Xin。压缩感知中的 $\ell_{1-2}$ 最小化。SIAM 科学计算期刊，37(1)：A536–A563，2015。[37] W. Yin、S. Osher 和 D. Goldfarb。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 成像科学期刊，1(1)：143–168，2008。[38] C.-H. Zhang。在极小极大凹惩罚下近乎无偏的变量选择。统计年刊，38(2)：894–942，2010。[39] J. Zhang 和 S. Zhang。具有先验支撑信息的 $\ell_{1-2}$ 最小化的零空间性质。IEEE 信号处理快报，28：1779–1783，2021。[40] Y.-B. Zhao。用于稀疏优化问题的最优 $k$-阈值算法。SIAM 优化期刊，30(1)：31–55，2020。