arXiv:2603.18254v1 [cs.DS] 18 Mar 2026

贝叶斯估计中的计算—效用—隐私权衡

Sitan Chen*
哈佛大学

Jingqiu Ding †
苏黎世联邦理工学院

Mahbod Majid ‡
MIT

Walter McKelvie §
哈佛大学

2026年3月18日

摘要

贝叶斯方法是现代数据科学的核心，为数据受限设定下的估计，以及以原则性方式量化和传播不确定性，提供了强有力的框架。然而，在许多实际应用中，部署这些方法时，自然需要保护数据所涉及个人的隐私。尽管已有不少研究尝试通过分析后验分布固有的隐私属性，或对现成的贝叶斯方法进行隐私化，来解决差分隐私贝叶斯估计 问题，但总体而言，这些研究在低维设定之外并未提供严格的效用保证。事实上，即便对于高斯均值估计和线性回归这类典型任务，即使在未知参数来自高斯先验这一最简单的情形下，隐私算法能够达到多接近贝叶斯最优误差，此前也尚无定论。

在本文中，我们为这两个问题首次给出了多项式时间算法，使均方误差达到 $(1 + o(1))\mathrm{OPT}$，并进一步表明，这两项任务都呈现出一种引人关注的计算-统计差距。对于贝叶斯均值估计，我们证明，在低次框架内，我们的方法所达到的超额风险在所有高效算法中都是最优的，但可证明地劣于指数时间算法所能达到的水平。对于线性回归，我们也证明了定性上类似的下界。我们的算法借鉴了 [HKMN23], 提出的隐私到鲁棒性框架，但其中有一个奇特之处：为了实现私有的贝叶斯最优估计，我们需要针对经验均值和普通最小二乘估计量等本质上不具鲁棒性的对象，设计基于平方和的鲁棒估计量。在此过程中，我们还为平方和工具箱增添了一类基于短平分解的新约束。

*电子邮件：sitan@seas.harvard.edu，部分经费由美国国家科学基金会 CAREER 奖项 CCF-2441635 资助

†电子邮箱：jingqiu.ding@inf.ethz.ch，由欧盟地平线 2020研究与创新计划下的欧洲研究理事会（ERC）资助（资助协议编号 815464）

‡电子邮箱：mahbod@mit.edu§电子邮箱：wmckelvie@fas.harvard.edu，部分由美国国家科学基金会研究生研究奖学金计划（GRFP）奖项 2140743 和 CAREER 奖项 CCF-2441635 资助

目录

1 引言 41.1 相关工作 .9

2 技术概述 **10**

2.1 数据依赖统计的隐私与鲁棒性.102.2 贝叶斯均值估计 .112.3 贝叶斯线性回归 .12

3 展望 **15**

4 预备知识 **16**

4.1 鲁棒性与隐私条件下的贝叶斯估计 .164.2 低次困难性 .174.3 平方和背景知识.194.3.1 平方和证明与算法 .194.3.2 有用的平方和不等式 .204.3.3 近似满足线性算子 .214.4 鲁棒性与隐私之间的联系 .224.5 矩阵的稀疏谱范数.23

5 稳健与私有贝叶斯均值估计 23

5.1 重新表述为私有经验均值估计.245.2 私有各向同性经验均值估计的统计速率 .255.3 私有各向同性经验均值估计的计算速率 .275.3.1 高效稳健经验均值估计 .275.3.2 从鲁棒性到隐私的归约 .285.4 贝叶斯均值估计中的信息—计算差距.305.4.1 稳健经验均值估计的困难性.305.4.2 私有经验均值估计的困难性 .335.5 持续观测下私有高斯均值估计的应用 .34

6 稳健且隐私贝叶斯线性回归 36

6.1 具有临界先验的隐私贝叶斯线性回归的计算速率 . . . . . 376.1.1 具有临界先验的高效稳健回归 .376.1.2 从鲁棒性到隐私的归约 .406.2 具有弱先验的隐私贝叶斯线性回归的计算速率 42

6.2.1 多阶段算法概述 .426.2.2 第1阶段：粗略估计 .436.2.3 第 2 阶段：细化 .446.2.4 汇总全部内容 .49

6.3 稳健贝叶斯回归的困难性 .496.3.1 引理 6.23 的证明 .506.3.2 将假设检验约化为稳健后验均值估计。 . . . . . 526.3.3 引理 6.25 的证明 .546.4 隐私贝叶斯回归的困难性 .57

参考文献 **58**

A 第5节的延后证明 **64**
A.1 从经验均值到后验均值 .64A.2 通过分桶从各向同性推广到各向异性 .65A.3 引理 5.8 的证明：可行性的韧性界 .66A.4 各向同性估计引理 .68

B 频率学派推论 **69**
B.1 频率学派均值估计 .70B.2 频率学派线性回归 .70

C 第6节的延后证明 **72**
C.1 可行性的集中不等式 .72C.2 定理 6.2 的证明：稳健且隐私保护的贝叶斯回归的低效算法 .75C.2.1 鲁棒回归的低效算法 .75C.2.2 从鲁棒性到隐私的归约 .80C.3 引理 6.17 的证明：具有常数误差率的鲁棒回归 .81C.4 弱先验下鲁棒回归算法的隐私化 .82C.4.1 对第1阶段的粗略估计器进行隐私化 .82C.4.2 对第 2 阶段的精化估计量进行隐私化 .83

1 引言

给定数据 x 由一个具有未知参数 θ的过程生成，我们该如何更新关于 θ的先验信念？这正是贝叶斯推断的目标。为实现最优推断，人们可以利用丰富的算法和统计工具集，从数据中尽可能多地提取关于 θ 的信息 x。然而，在许多实际应用场景中，这套工具可能会——而且确实会——被采用，而最优推断与另一项正交的要求之间存在内在张力： 隐私。流行病学家经常使用贝叶斯方法来估计大流行病的传播速度 [AHS+25，FMG+20]；美国人口普查局利用这些方法推断小亚群体内的统计数据 [FIH79，CD20]，；遗传学家则用这些方法预测个体罹患复杂疾病的风险 [VYF+15]。有没有办法在保留这些技术统计效能的同时，保护其所分析数据对应个体的隐私？

尽管差分隐私（DP） [DMNS06a] 为研究这一问题提供了理想的形式化框架，但有关差分隐私统计的大部分工作都集中于频率学派设定，在这一设定下，人们无法获得关于 θ的先验信息。Dimitrakakis 等人 [DNZ+17] 和 Wang 等人 [WFS15], 的早期工作是一个例外，他们提出了一个颇具吸引力的设想。他们认为，贝叶斯推断所针对的对象，即后验分布 $\mathrm{p}(\theta \mid \mathrm{x})$，本质上是一个概率对象，因此，来自该后验分布的样本或许包含足够的噪声，能够“免费地”满足差分隐私。这一设想奠定了单后验样本（OPS）机制 [Zhe15, DNZ+17, WFS15, FGWC16, Mir13, BST14, ZRD16, GSC17] 这一富有成果的研究方向；该方向旨在从后验分布的某个退火版本中输出一个样本，其依据是：在对先验分布和对数似然 $\ln \mathrm{p}(\mathrm{x} \mid \theta)$作出（非常强的）假设时，这样做已经满足差分隐私。

遗憾的是，在许多自然场景中，真实后验分布并不具备隐私性，即便是在均值估计和线性回归这类典型测试场景中也是如此。尽管人们已尝试通过替代算法来解决这一问题，例如对充分统计量进行适当加噪处理 [FGWC16、BS18、BS19、FS24]，或者运行标准贝叶斯方法的差分隐私类比方法 [WFS15、FGWC16、HJDH19、ZZ23、Jäl17、HDN+18、PFCW20],但这些方法的效用保证要么尚不明确，要么只有在低维情况下才具有定量意义。

单后验样本范式的另一个严重局限是，在贝叶斯推断的典型应用中，从后验分布中获得单个样本并没有特别大的用处。如果后续步骤需要将后验分布作为截至当前时刻不确定性的完整摘要继续使用，那么真正需要的是对后验分布进行私密估计，获得其描述；或者，获得一个后验抽样预言机，以便随后从后验分布中生成任意数量的样本。例如，如果我们能够对数据的充分统计量进行私密估计，就可以做到这一点。

因此，我们提出以下问题:

我们能否严谨地 为高维情形下的贝叶斯后验采样确立隐私-效用权衡?

在这项工作中，我们对高斯均值估计和高斯设计下的线性回归这两项基本任务给出了肯定的回答。不同于以往差分隐私贝叶斯推断中的方法，我们利用了强大的平方和（*SoS*）层级。该层级近期被用于在频率学派差分隐私设定下，为上述任务 [CCAd+23、ABMT25、HKMN23]等均值估计、回归及相关问题取得了最先进的理论保证。尽管我们将以这些技术为基础，但正如我们将看到的，贝叶斯设定会带来许多频率学派设定中不存在的微妙现象和技术难题。

贝叶斯均值估计。 在这个问题最基本的形式中，未知参数 θ对应于 $\mathbb{R}^d$ 中的未知均值，该均值从某个先验分布 π中采样；观测到的数据 x 由独立同分布样本 x_1、、. . .、x_n组成，这些样本来自 $\mathcal{N}(\theta, \mathrm{Id})$。在整个均值估计的讨论中，我们将使用记号 μ 而不是 θ 来表示均值向量。目标是设计一个（随机化的）

估计量 $\hat{\mu}(x)$，使得均方误差（MSE）

$$\mathrm{MSE}(\hat{\mu}) := \mathop{\mathbb{E}}_{\mu\sim\pi} \mathop{\mathbb{E}}_{x\sim\mathcal{N}(\mu,\mathrm{Id})^{\otimes n}} \|\mu - \hat{\mu}(x)\|_2^2 \tag{1}$$

在确保$\hat{\mu}$ 满足差分隐私的同时，使其尽可能小；具体而言，对于最坏情况下的相邻数据集x,x'，其输出分布$\hat{\mu}(x)$ 和$\hat{\mu}(x')$ 应当足够接近。若不考虑隐私约束，使均方误差最小的贝叶斯最优估计量 是后验均值 $\mathbb{E}[\mu \mid x]$。请注意，要使后验均值（甚至均方误差）有明确定义，必须事先知道先验分布 π。本文将考虑协方差 Σ已知的高斯先验。

对于这项任务，我们给出了首个既能保护隐私，又能在先验分布 π 为高斯分布时，经证明可与贝叶斯最优估计量竞争（事实上，近乎达到最优）的估计量。首先，通过简单的方差分解，任意估计量$\hat{\mu}$ 所达到的均方误差 可写成两部分之和：由 $\mathbb{E}[\mu \mid x]$, 达到的最小均方误差（MMSE），以及$\hat{\mu}(x)$ 与 $\mathbb{E}[\mu \mid x]$ 之间的期望平方差（式（14））。因此，我们的目标是尽可能准确地估计 $\mathbb{E}[\mu \mid x]$。还要注意，由于先验分布是高斯分布，后验分布也是高斯分布，且其协方差已知[1]，，因此一旦我们得到了 $\mathbb{E}[\mu \mid x]$, 的估计值，就能对完整的后验分布 做出统计上足够接近的描述，随后便可从中生成任意数量的样本，而无需额外付出隐私成本。

定理 1.1（私有贝叶斯均值估计的上界）。给定从先验分布 $\pi = \mathcal{N}(0,\Sigma)$中抽取的 μ，若数据 x 由来自 $\mathcal{N}(\mu,\mathrm{Id})$ 的 n 个 $\geqslant \tilde{\Omega}(\frac{d}{\varepsilon}\log(\sqrt{\mathrm{Tr}(\Sigma)/\alpha}))$ 独立同分布样本组成，则存在

1. 一个计算上低效的 *ε-DP* 估计量 $\hat{\mu}(x)$，使得其以至少 $1-\beta$ 的概率满足，

$$\|\hat{\mu}(x) - \mathbb{E}[\mu \mid x]\|_2 \leqslant \tilde{O}\left(\frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}}\log(1/\beta)}{\varepsilon n}\right) := \tilde{O}(\alpha_{\mathrm{stat}}) \tag{2}$$

2. 一个多项式时间 *ε-DP* 估计量 $\hat{\mu}(x)$，使得其以至少 $1-\beta$ 的概率满足，

$$\|\hat{\mu}(x) - \mathbb{E}[\mu \mid x]\|_2 \leqslant \tilde{O}\left(\frac{\|\Lambda\|_{4/3} + \|\Lambda\|_{\mathrm{op}}\log^{3/4}(1/\beta)}{\varepsilon^{1/2} n^{3/4}} + \alpha_{\mathrm{stat}}\right) := \tilde{O}(\alpha_{\mathrm{comp}} + \alpha_{\mathrm{stat}}). \tag{3}$$

其中 $\Lambda := (\mathrm{Id} + n\Sigma^{-1})^{-1}$，而 $\|\Lambda\|_{4/3}$ 表示 Λ 的 *Schatten* 4/3-范数。这里，$\tilde{O}$ 隐去了$\log 1/\alpha$、$\log\log 1/\beta$ 和 $\log d$ 的因子。当 $\Sigma = \mathrm{Id}$ 时，我们不会损失 $\log d$ 因子。

首先请注意，与先前频率学派的研究不同，这里没有“非隐私”项：取 $\varepsilon \to \infty$ 就能以任意高的精度估计后验均值。这是合理的，因为在非隐私地访问数据的情况下，可以精确计算后验均值。

该结果的大部分直觉体现在各向同性情形中，此时 $\Sigma = \sigma^2\mathrm{Id}$ 对于 $\sigma \asymp 1/\sqrt{n}$，此时

$$\alpha_{\mathrm{stat}} = \frac{d + \log(1/\beta)}{\varepsilon n} \qquad \text{and} \qquad \alpha_{\mathrm{comp}} = \frac{d^{3/4} + \log^{3/4}(1/\beta)}{\varepsilon^{1/2} n^{3/4}}. \tag{4}$$

在这种情况下，可以计算得出

$$\mathbb{E}[\mu \mid x] = \frac{1}{n + 1/\sigma^2}\sum_{i=1}^{n} x_i. \tag{5}$$

换言之，贝叶斯最优估计量是通过计算数据的经验均值，并将其向原点收缩 来得到的，收缩因子取决于样本量和先验方差。在这种设定下，

[1]我们采用交换邻接，因此n不具备隐私性

可以将我们的估计量达到的误差与真实均值 μ达到的误差进行比较，并按$n\ n+1/\sigma^2 = \Theta(1)$缩放，其中

$$\left\| \frac{n}{n+1/\sigma^2} \cdot \mu - \mathbb{E}[\mu \mid x] \right\|_2 \leqslant \sqrt{\frac{d + 2\log(1/\beta)}{n}} \tag{6}$$

以至少 $1-\delta$ 的概率。请注意，在 $\varepsilon = \Theta(1)$ 的情形下，此界优于 α_{stat} 和 α_{comp}。事实上，我们通过构造用于估计经验均值的隐私估计器来证明 定理 1.1；对于任何 μ，只要它位于半径为 R 的某个球内，这些估计器就能以高概率成功。由于我们的经验均值估计器无论 μ 的取值如何都能成功（且无需了解其取值），因此即使在 频率学派框架下也可应用，从而得到一个具有以下性质的 μ 估计器：

定理 1.2（具有最优常数的频率学派均值估计）。设 Λ 为半正定矩阵。对于任意满足$\|\mu\| \leqslant R$ 的 $\mu \in \mathbb{R}^d$，若数据 x 由 n 个 $\geqslant \tilde{\Omega}\left(d \text{ 个 } \varepsilon \log\left(\frac{R+\|\Lambda\|_F}{\alpha}\right)\right)$个来自 $\mathcal{N}(\mu, \Lambda^2)$ 的独立同分布样本组成，则存在

1. 一种计算上效率低下的 ε-*DP* 估计量 $\mu(x)$，使得其以至少 $1-\beta$ 的概率，

$$\|\mu(x) - \mu\|_2 \leqslant (1 + o(1)) \sqrt{\frac{\|\Lambda\|_F^2 + \|\Lambda\|_{\text{op}}^2 \log \frac{1}{\beta}}{n}} + \tilde{O}(\alpha_{\text{stat}})$$

2. 一种多项式时间 ε-*DP* 估计量 $\mu(x)$，使得其以至少 $1-\beta$ 的概率，

$$\|\mu(x) - \mu\|_2 \leqslant (1 + o(1)) \sqrt{\frac{\|\Lambda\|_F^2 + \|\Lambda\|_{\text{op}}^2 \log \frac{1}{\beta}}{n}} + \tilde{O}(\alpha_{\text{comp}} + \alpha_{\text{stat}}).$$

此处的 $\tilde{O}$ 隐去了$\log 1/\alpha$、$\operatorname{loglog} 1/\beta$ 和$\log d$ 因子。当 $\Sigma = \mathrm{Id}$ 时，我们不会损失$\log d$ 因子。主导项中的 $o(1)$ 在 $\|\Lambda\|_F/\|\Lambda\|_{\text{op}} \to \infty$ 或 $\beta \to 0$ 的情形下成立。

这与 [HKMN23定理 15],所达到的均值估计渐近最优速率相匹配，但前导非隐私项具有最优常数。因此，它可以被视为 [KV17定理 12] 在更高维度上的推广（证明见附录 B）。

这一切凸显了我们工作的第一个概念性转折：在贝叶斯设定中，隐私均值估计关注的是对数据的经验均值进行隐私估计，而非真实均值。事实上，尽管前者在无限样本极限下会收敛到后者，式（6）中的有限样本误差仍然显著——其量级与 MMSE 相同；而我们的定理 1.1表明，只需付出低阶额外误差，就有可能以隐私方式达到 MMSE。聚焦于经验均值尤其违反直觉，因为我们最终将证明定理 1.1 时，会运用 [HKMN23] 的方法，将问题化归为一种不同寻常的稳健统计问题：给定一个受污染数据集，估计其经验均值。经验均值虽是众所周知的非稳健统计量，但稳健地估计它对于实现差分隐私贝叶斯估计至关重要。我们用于稳健经验均值估计的算法，基于对 KMZ 提出的稳健参数均值估计平方和分析的谨慎改编 [KMZ22]。正式定义见第 4.1 节，稳健问题的保证见第 5 节。

有人可能会想，计算复杂度定理 1.1中的额外项是否只是我们所用技术造成的假象。我们的研究设定中，下一个令人意外的发现是，这一差距实际上是固有的：差分隐私贝叶斯均值估计天然存在计算-统计差距！这或许出人意料，因为该问题的表述并不具备此类差距通常具有的特征，例如稀疏性和低秩结构。

定理 1.3（非正式表述，参见 定理 5.2）。 根据低次启发式 *[KWB19, Wei25]*, 不存在多项式时间估计器 $\hat{\mu}$ 能够达到 $\|\hat{\mu}(x)-\mathbb{E}[\mu\,|\,x]\|_2 \ll \alpha_{\text{comp}}+\alpha_{\text{stat}}$ 当 $\beta=\exp(-\Theta(d))$ 且 $\pi=\mathcal{N}(0,\sigma^2\mathrm{Id})$对于 $\sigma \asymp 1/\sqrt{n}$。

我们通过仔细地将最近由 [DHPT25] 提出并证明为低次难解的判别问题，归约到差分隐私贝叶斯均值估计问题，从而证明这一点。

跳出各向同性情形来看，定理 1.1 的另一个引人注目的特点是，α_{stat} 和 α_{comp} 所体现的隐私成本并不显式依赖维度（至多相差对数因子），而是取决于先验协方差 Σ：换言之，先验分布会告知我们的算法应将隐私预算用于重点估计哪个维度。这是差分隐私贝叶斯估计相较于差分隐私频率学估计的一大优势：先验信息使我们能够通过对数据的不同方向审慎地施加不同程度的噪声，从而最优地分配隐私预算。这在数据受限设定中尤为重要，因为仅通过添加各向同性噪声来保护隐私，可能会严重损害效用。

请注意，我们的设定与关于“协方差感知”均值估计的文献 [BGS+21][BHS23] [KDH23] 有两点不同：首先，在我们的设定中，协方差是先验地（按字面意义）已知的；其次，我们关注的是 ℓ_2 而非马氏误差。相比之下，我们的结果最接近参数化对应物 [ALNP23] 和 [DJY+24]，它们研究的是在协方差已知的各向异性噪声下进行隐私均值估计，以 ℓ_2 范数衡量。

贝叶斯视角的另一个优势在于，它能够自然地实现模块化的不确定性传播。贝叶斯统计建立在对后验分布估计值进行后处理的基础上，而差分隐私（DP）不受后处理影响，这是一个标准事实。在第 5.5 节中，我们用一个简单的例子说明这一点：我们使用自己的算法，在流式设定下，随着分布的样本逐步提供给我们，私密地估计其均值。

贝叶斯线性回归。接下来，我们转向线性回归的保证。在这里，未知参数 θ 对应于 $\mathbb{R}^d$ 中的一个未知回归器，该回归器同样从某个先验分布中抽取。讨论线性回归时，我们始终用字母w 代替 θ。给定w 后，观测到的数据x现在以独立同分布样本对的形式给出 $(x_1,y_1),\ldots,(x_n,y_n)$，其中$x_i \sim \mathcal{N}(0,\mathrm{Id})$，且 $y_i = \langle w, x_i\rangle + \xi_i$，其中$\xi_i \sim \mathcal{N}(0,\varsigma^2)$。目标仍然是最小化均方误差（MSE），其定义如下：

$$\mathrm{MSE}(\hat{w}) := \mathop{\mathbb{E}}_{w\sim\pi}\mathop{\mathbb{E}}_{\{(x_i,y_i)\}|w}\|w-\hat{w}(\{x_i,y_i\})\|^2, \tag{7}$$

同时确保 $\hat{w}$ 满足差分隐私。在不受隐私约束时，最优估计量仍然是后验均值 $\hat{w}(x)=\mathbb{E}[w \mid \{(x_i,y_i)\}]$。令 $X \in \mathbb{R}^{n\times d}$ 表示设计矩阵，其各行由 $x_1,\ldots,x_n$组成；令 $y \in \mathbb{R}^n$ 表示标签向量 $y_1,\ldots,y_n$，则后验均值为

$$w_{\text{post}} := \mathbb{E}[w \mid \{(x_i,y_i)\}] = (\sigma^{-2}\mathrm{Id}+XX^\intercal)^{-1}Xy\,. \tag{8}$$

针对这一任务，我们给出了首个同时保护隐私、且经证明能够在先验 π 为各向同性高斯分布时与贝叶斯最优估计量相竞争的估计量。

定理 1.4（非正式表述，参见定理 *6.2* 和定理 *6.3*）。令 $\varepsilon \leqslant O(1)$。给定从先验分布$\pi = \mathcal{N}(0,\sigma^2 Id)$中抽取的 w，如果数据由 $n \geqslant \Omega(\frac{d}{\varepsilon}log(\sigma\sqrt{d}))$ 个独立同分布样本组成，形式为 (x_i,y_i)，其中 $x_i \sim \mathcal{N}(0,Id)$，且 $y_i = \langle w, x_i\rangle + \xi_i$，则存在：

- 一种计算 *nally* 低效的 *ε-DP* 估计量 $\hat{w}(\{(x_i,y_i)\})$ 使得概率至少为 $1-\beta$ ，

$$\|\hat{w}(\{(x_i,y_i)\})-w_{\text{post}}\|_2 \leqslant \tilde{O}\left(\frac{d+\log(1/\beta)}{\varepsilon n}\right) \tag{9}$$

• 一个多项式时间的 ε-DP 估计量 $\hat{w}(\{(xi, yi)\})$ 使得概率至少为 $1-\beta$,

$$\|\hat{w}(\{(x_i, y_i)\}) - w_{\text{post}}\|_2 \leqslant \tilde{O}\left(\frac{(d+\log(1/\beta))^{3/4}}{\varepsilon^{1/2} n^{3/4}} + \frac{d+\log(1/\beta)}{\varepsilon n}\right). \tag{10}$$

均值估计设定中的相同特殊之处在这里同样适用。在指出后验均值与普通最小二乘（*OLS*）估计量之间的联系后，我们将为隐私贝叶斯线性回归给出一种算法，其方法是为一个本质上不具鲁棒性的对象——OLS 估计量——构造稳健估计量。与定理 1.1的证明一样，我们采用 [HKMN23], 的鲁棒性到隐私框架；这需要设计一种高效算法，在数据中有 η比例的部分被篡改时，近似计算 OLS 估计量（正式定义和保证见第 6.2 节）。

与均值估计一样，我们的线性回归结果也立即适用于频率学派统计学。我们的隐私估计器能够近似最小二乘估计器$w_{\text{OLS}} := (XX^\intercal)^{-1}Xy$，无论$w$的分布如何，因此可以应用于频率学派设定，得到以下保证：

定理 1.5 （具有最优常数的频率学派线性回归）。 对于任意 $w \in \mathbb{R}^d$ 满足 $\|w\| \leqslant R$ 的情形，若数据由 $n \geqslant \tilde{\Omega}\left(\frac{d}{\varepsilon}\log(R/\alpha)\right)$ 个如下形式的独立同分布样本组成：(x_1, y_1)、……、(x_n, y_n)，其中 $x_i \sim \mathcal{N}(0, \mathrm{Id})$，且 $y_i = \langle w, x_i\rangle + \xi_i$，则存在

1. 一种计算上 效率低下的 ε-DP 估计量 $\hat{w}(\{(x_i, y_i)\})$，使得其概率至少为 $1-\beta$ ，

$$\|\hat{w}(\{(x_i, y_i)\}) - w\|_2 \leqslant (1+o(1))\sqrt{\frac{d+\log\frac{1}{\beta}}{n}} + \tilde{O}\left(\frac{d+\log(1/\beta)}{\varepsilon n}\right).$$

2. A 多项式时间 ε-DP 估计量 $\hat{w}(\{(x_i, y_i)\})$，使得其概率至少为 $1-\beta$ ，

$$\|\hat{w}(\{(x_i, y_i)\}) - w\|_2 \leqslant (1+o(1))\sqrt{\frac{d+\log\frac{1}{\beta}}{n}} + \tilde{O}\left(\frac{(d+\log(1/\beta))^{3/4}}{\varepsilon^{1/2} n^{3/4}} + \frac{d+\log(1/\beta)}{\varepsilon n}\right).$$

~此处的 O 隐去了 $\log(1/\alpha)$、$\log\log(1/\beta)$ 和 $\log d$ 的因子。

该结果通过将针对 w_{OLS} 的相应隐私近似保证与标准高概率界 $\|w_{\text{OLS}} - w\|_2 \leqslant (1+o(1))\sqrt{\frac{d+\log(1/\beta)}{n}}$ 相结合而得。

请注意，与现有关于稳健或隐私保护高斯线性回归的 SoS 研究不同，定理 1.4的样本复杂度关于维度是线性的，而非二次的。这得益于我们对模型采取了较宽松的选择：现有 SoS 文献主要关注未知协方差情形，在该情形下，统计查询下界表明， $\Omega(d^2)$ 个样本在计算上是必要的（见 [ABMT25，备注 1。5] 然而，已知各向同性协方差情形下已有高效的非 SoS 估计量 [DKS19a]。我们构造了首个针对各向同性协方差高斯线性回归的 SoS 稳健估计量，其样本复杂度关于维度是线性的，从而绕过了此前在认证四阶矩超收缩性方面的障碍。正如我们在 第 2 节中所概述的，我们的贝叶斯线性回归结果需要新的思路，例如约束 SoS 程序仅在数据的某些子集上进行搜索，使得这些子集所对应的 y 值可以分解为一个短向量和一个平坦向量。

请注意，与现有关于稳健或隐私高斯线性回归的 SoS 研究不同，定理 1.4的样本复杂度关于维度呈线性而非二次增长。现有 SoS 文献主要关注未知协方差情形，有证据表明，在该情形下， $\Omega(d^2)$ 个样本可能是

任何计算高效算法所必需的（参见 [ABMT25，备注 1。5]）。另一方面，对于已知各向同性协方差情形，已有高效的非 SoS 估计量 [DKS19a]。我们构造了首个用于各向同性协方差下高斯线性回归的基于 SoS 的估计器，其样本复杂度关于维度呈线性增长。正如我们在第 2 节中所概述的，我们的贝叶斯线性回归结果需要新的思路，例如约束 SoS 程序仅在数据的某些子集上进行搜索，使得对应的y 值能够分解为一个短向量和一个平坦向量。

最后，与均值估计的情形类似，我们证明指数时间算法与计算高效算法之间确实存在差距：一旦同时要求隐私和贝叶斯最优性，就会再次出现计算-统计差距：

定理 1.6 (非正式表述，参见 定理 6.4)。 在低次启发式 *[KWB19* 下，*Wei25]*, 不存在多项式时间估计器 $\hat{w}$ 能够在 $O(d^{1/3}\log(1/\beta)/\varepsilon^{2/3}\alpha^{4/3})$ 个样本内达到 $\|\hat{w}(\{(x_i,y_i)\})-\mathbb{E}[w\mid\{(x_i,y_i)\}]\|_2\ll\alpha$。
当 $\beta=\exp(-\Theta(n))$ 时，这表明定理 1.4 中计算高效算法达到的速率在常数因子范围内是最优的。

1.1 相关工作

差分隐私贝叶斯估计。相关文献十分丰富，全面综述可参见例如[Sar24, Section 7.2.2] 。这方面的研究大致可归为以下几类主题：（A）“单个后验样本”方法，分析单个后验样本或其温度化版本所固有的隐私性；（B）对标准贝叶斯方法进行隐私化实现；（C）基于发布带噪充分统计量的算法。（A）类工作包括 [Mir13, DNZ+17, WFS15, GSC17, ZRD16, Zhe15],，其常见的证明策略是论证对数似然 $\ln p(x\mid\theta)$ 关于数据集距离在 x 上满足 Lipschitz 条件——但在我们的设定中，这一条件并不成立。（B）类工作包括 MCMC 的隐私化 [WFS15, FGWC16, HJDH19, ZZ23], 、变分推断 [Jäl17, HDN+18, PFCW20], ，以及图模型推断中的带噪贝叶斯更新。除缺乏具体的效用保证外，这些方法的另一缺点是每次迭代都会产生隐私成本，这会使针对纯差分隐私的朴素方法难以为继。我们还注意到，许多研究探讨了朗之万蒙特卡洛的收敛隐私性，参见例如 [BST14, CYS21, ABT24, AT22, BSA24], ；这些研究的出发点是优化而非推断。（C）类工作包括 [FGWC16] 以及[BS18, BS19, FS24] 关于指数族和贝叶斯线性回归推断的研究；这些研究基于带噪充分统计量，但没有提供可证明的效用保证。这些研究也与基于隐私化数据进行贝叶斯推断的相关工作密切相关，其中隐私机制被纳入似然 $p(x\mid\theta)$，参见例如 [XX12, WM10, JAGR22]。最后，我们指出，还有一些研究探讨了贝叶斯统计与隐私定义之间的联系，例如[KS14, TF20],，但这些工作与我们的研究无关。

用于鲁棒性和隐私的平方和。 关于使用平方和进行高维统计估计的文献过于丰富，无法在此一一详述。相反，我们聚焦于使用平方和将鲁棒估计与差分隐私联系起来的相关工作。首个利用这种联系的平方和工作是 [HKM22], , 随后 [HKMN23] 对其进行了扩展，提出了我们的工作所依据的形式化框架。这些工作利用这种联系，首次获得了具有最优样本复杂度的纯差分隐私均值估计的计算高效算法。其他利用这一联系的工作包括 [ABMT25、 CCAd+23]。我们还提及使用平方和进行鲁棒均值估计的众多工作 [HL18、 KS17、 KMZ22] 以及稳健回归 [KKM18、BP21、CKMY22b、 ABMT25] 。

在鲁棒均值估计方面，早期的平方和研究，例如 [HL18、KS17] 针对可认证次高斯的超收缩族，提出了用于鲁棒矩估计的一般松弛方法 /,

但将这些方法专用于高斯分布时，所得分析并未恢复对污染比例的最优依赖关系。 [KMZ22] 的贡献在于弥合了这一差距：他们表明，对于高斯分布，~本质上相同的典范松弛已能通过多项式时间算法达到最优的 O(η) 误差；其新分析能够利用高斯矩下界，而无需匹配的平方和证书。我们在第 5.3.1 节中用于稳健经验均值估计的估计量，是对这一分析的审慎改编；A.4 节则给出了我们使用的关键引理。在稳健回归方面，与我们最接近的平方和研究是 [KKM18、BP21]。前者在可认证的 4 阶超收缩性假设下给出保证，后者则将其推广到可认证的 k 阶超收缩性，并在噪声相互独立或具有适当负相关的条件下，得到阶为 $\tilde{O}(\eta^{2-2/k})$的超额预测误差；在各向同性高斯情形下，这对应于阶为 $\tilde{O}(\eta^{1-1/k})$的参数误差。然而，这些保证并不能直接满足我们的设定。首先，我们需要对未污染样本所对应、依赖于数据的普通最小二乘/ 后验均值进行鲁棒近似。其次，常数次矩假设会导致我们所需的 η依赖关系次优：例如，4 次矩只能带来 $\eta^{3/4}$类型的参数保证；而增大 k 虽能改善 η依赖关系，却会使维度依赖关系变差，因为必须根据样本认证相应的更高阶矩。具体而言，仅对 4 次矩进行认证，在高斯设计下就至少需要维度的二次样本复杂度。

有关现代稳健统计学发展的更全面综述，请读者参阅 [DK23]。

稳健推断。 相比之下，关于稳健贝叶斯统计学的文献要稀少得多。值得一提的是，近期 [HL24] 研究了树上的广播问题中的稳健贝叶斯推断，以及 [CKMY22a] 关于稳健卡尔曼滤波的研究。另有一些研究在鲁棒性背景下考察了与贝叶斯相关的问题，例如 [LM22] 关于实现稳健社区检测极小极大速率的研究，以及 [IS24、IS25] 关于稳健自旋玻璃优化的研究。

与先前研究相比的频率学派应用。 先前关于隐私估计的研究大致分为两派。一派 [KLN+08、KV17、KLSU19、HKMN23] 采用学习理论视角：从某个分布中抽取样本，估计量旨在学习该分布的某种性质。另一派 [DMNS06a、BLR08、HT10] 旨在对任意样本的某个数据相关统计量进行隐私估计。 定理 1.2表明，前一派更强的分布假设甚至可以有效地用于实现后一派的目标：如果对数据不作任何分布假设，我们所能做的最好选择就是采用拉普拉斯机制，得到$n = \Theta(\frac{dR}{\varepsilon n})$；当 Λ较小或R较大时，这远不如高斯性这一分布假设所能达到的效果。

—

隐私估计中的既有计算下界。 已有研究 [DNR+09、UV20]利用强密码学假设，得到了差分隐私的下界。然而，对密码学问题进行编码所需的结构远多于我们这里具备的结构：这些困难性结果通常仅适用于合成数据生成。 [KMUW17] 的适用范围超出了合成数据，但需要一个规模呈指数级增长的查询类。相比之下，高斯噪声下的均值估计是一项非常简单、具体且结构极少的统计任务。

2 技术概述

2.1 数据依赖统计中的隐私与鲁棒性

我们的首个概念性贡献很简单，却颇具新意，或许适用于稳健与隐私统计中的许多不同问题。传统上，我们这一研究领域的论文一直聚焦于参数估计：某个参数对算法而言是未知的，数据是从以该参数为索引的

分布中独立同分布地生成的，而算法必须根据数据估计该参数。这是统计学中最常见的设定，因为它概括了统计任务的核心：根据噪声测量进行猜测。

我们不再估计模型的某个属性，而仅仅着眼于估计样本的某个属性。从统计估计的角度来看，这使得问题变得完全简单。然而，在施加鲁棒性或隐私要求后，问题反而比参数估计设置更加困难：我们不再能够将分析上的松弛隐藏在一个占主导地位的统计项之后，而必须刻画并正视鲁棒性或隐私本身的代价。

我们的主要观察是，许多为参数估计设置而发展出的技术，都可以直接应用于数据估计设置；通过更加细致的分析，还能为这些技术提供更严格的保证。特别是，[HKMN23] 提出的“鲁棒性蕴含隐私”结果，即使当统计量是数据的属性而非模型的属性时仍然成立，前提是我们的鲁棒估计量是在对未污染样本的数据依赖属性进行“鲁棒”估计，而不是对模型的底层参数进行估计。我们在第 5.3.2 节中针对均值估计详细阐述了这一观点，并在第 6.1.2 节和C.4 节中针对线性回归进行了讨论。

2.2 贝叶斯均值估计

归约到稳健经验均值估计。 在本概述中，我们聚焦于各向同性先验的情形，因为该情形包含了大部分技术直觉。在这一设定下，正如引言中所述，后验均值只不过是对经验均值进行重新缩放。为了对这一对象进行隐私估计，我们将借助 [HKM22、HKMN23]所发展出的隐私与稳健统计之间的强大联系。我们在第 5.3.2 节 中针对我们的设定对这一工具进行了仔细调整，将问题的本质提炼为稳健统计中一个颇为独特的问题：给定一个高斯分布数据集的 η-污染，且其参数均值未知，估计原始未污染数据集的经验均值。乍看之下，这似乎是一个悖论，因为经验均值是参数均值的极不稳健估计量。之所以不存在矛盾，是因为我们关注的是未污染数据的经验均值，而根据测度集中性，该均值无论如何都接近参数均值。

然而，关键在于我们不希望为这种集中误差付出代价，因为它可能与原始贝叶斯估计问题中的贝叶斯最优误差处于同一量级！相反，在我们的设定中，必须专门针对经验均值，以获得相对于贝叶斯最优误差而言阶数更低的超额风险界。

鲁棒经验均值估计算法。事实证明，为了通过套用定理 1.4中的稳健性到隐私的机制 [HKMN23],来实现该保证，我们必须将经验均值稳健地估计到误差$O(\eta\sqrt{\log 1/\eta}+\sqrt{\eta}\sqrt{d/n})$。这与先前研究中鲁棒均值估计的收敛速率一致；先前研究表明，参数均值可以在 $\tilde{O}(\eta)$的$n = O(d/\eta^2)$ 个样本下估计到相应误差，但在我们的设定中，我们希望得到一种界：当样本数低于$O(d/\eta^2)$ 个样本时，该界会平滑地变差；尤其是当 $\eta \to 0$时，它仍趋于零。

值得庆幸的是，对于均值估计，现有的频率学派高斯均值估计研究可以加以改造，从而给出这样的误差界。在 第 5.3.1 节中，我们将介绍如何借鉴 [KMZ22] 近期研究中的思路，在平方和框架下构造一个多项式时间算法，以稳健地估计高斯分布的均值。从整体上看，该平方和程序会寻找一个数据的大子集，使其协方差有界，而非高阶矩有界。其策略是在平方和框架下重现 [DKK+19] 基于韧性的方法，同时避开韧性可能无法被高效认证这一问题。我们将说明如何仔细处理 [KMZ22] 分析中出现的误差项，并由此推导出，同一分析实际上也能给出一种稳健的经验均值估计算法。由于这些技术工具已在先前研究中出现，我们在此不再赘述

，并将正式论述留待 第 5.3.1 节 和 A.4 节。

均值估计中的信息—计算差距。令人惊讶的是，我们用于稳健经验均值估计的平方和算法所达到的速率劣于统计最优速率。事实上，在第5.2节中，我们给出了一种计算上低效的算法，它能以误差 $O(\eta\sqrt{\log 1/\eta}+\sqrt{\eta}\sqrt{d/n})$ 稳健地估计经验均值。粗略地说，后一个算法会通过暴力搜索，寻找一个点集 $(y_1,\ldots,y_n)$；该点集与受污染数据集之间的差异至多涉及比例为 η 的点，而且关键的是，该点集满足如下性质：任意 η 比例的 y_i 的均值都足够接近所有 y_i 的经验均值 $\bar{y}$。也就是说，对于所有大小为 ηn 的 $S\subseteq[n]$，我们要求 $\frac{1}{\eta n}\sum_{i\in S}y_i$ 接近 $\bar{y}$。所有此类 S 中的最大距离称为矩阵 M 的最大 η-稀疏奇异值，其中 M 的列由 $\frac{1}{\eta n}(y_i-\bar{y})$ 组成。M 的最大奇异值为 $\sup_{v\in\mathbb{S}^{n-1}}\|Mv\|$，而最大 η-稀疏奇异值定义为 $\sup_{v\in\mathbb{S}^{n-1}:\|v\|_0=\eta n}\|Mv\|$。

简单应用并集界可知，如果我们将 y_i 视为真实且未受污染的高斯样本，那么最大 η 稀疏奇异值至多为 $O(\sqrt{\eta}\sqrt{d/n})$，最终便可得到统计最优速率。不幸的是，要计算高效地搜索具有这一性质的 y_i 并非易事。这并非没有原因：[DHPT25] 证明了，要对最大稀疏奇异值进行最优认证，在计算上是困难的。

尽管稳健经验均值估计在形式上并不等同于稀疏奇异值认证，我们仍能采用与 [DHPT25] 中类似的下界构造，论证前一个问题要达到最优速率在计算上同样困难。在第5.4.1节中，我们在低次框架下给出证据表明，任何计算高效算法若要在污染比例为 α 的情况下将经验均值的估计误差控制在 η，就需要 $n\geqslant\Omega\left(\frac{d\eta^2}{\alpha^4}\right)$ 个样本。

从整体上看，困难性结果的证明过程如下。我们构造两个高斯分布组成的不平衡混合分布，其中一个成分的 η 质量（定义 5.12）很小，另一个成分的均值则沿着一个未知的随机方向偏离原点 $\Theta(\alpha)$。通过标准计算，我们证明，任何低次数算法都难以使用 $o(d\eta^2/\alpha^4)$ 个样本将该系综与标准高斯分布区分开来。重要的是，这意味着此类算法无法判断数据究竟是未受污染、独立同分布地从标准高斯分布中抽取的，还是受到 η 污染并从非零均值高斯分布中抽取的。接下来，我们证明这个困难问题可以归约为稳健经验均值估计。事实上，要区分这两个系综，只需比较经验均值与稳健估计量输出之间的距离。事实证明，这种归约足以证明稳健经验均值估计所需的紧确的信息—计算权衡。

要在我们原始的贝叶斯设定中推导出下界，还需要做一些额外工作，因为上述归约证明其难度的均值参数先验分布，是 0 与长度为 $\Theta(\alpha)$ 的向量均匀分布的混合；具体细节推迟到定理 5.2 的证明中，见第 5.4.2 节末尾。关于私有贝叶斯均值估计的困难性，则可由 [GH22] 的隐私到鲁棒性归约得出。

2.3 贝叶斯线性回归

在贝叶斯线性回归中，我们希望估计的后验均值为 $w_{\text{post}}=(\frac{1}{\sigma^2}\mathrm{Id}+XX^\intercal)^{-1}Xy$，它与频率学派对应的普通最小二乘（OLS）估计量 $w_{\text{OLS}}=(XX^\intercal)^{-1}Xy$ 密切相关。正如获得贝叶斯均值估计的隐私算法时，大部分难点归结为一个非传统的问题：对数据的一个简单经验统计量（经验均值）进行鲁棒估计；这里分析的核心同样围绕着对 w_{OLS} 进行鲁棒估计。也就是说，给定一个 η 受污染数据集 $(\tilde{X},\tilde{y})$，我们希望尽可能估计 w_{OLS}，针对

原始未损坏数据集 $(X、y)$。此外，由于$XX^\intercal \approx n\mathrm{Id}$在算子范数意义下，论证的本质实际上围绕着估计向量 $\frac{1}{n}X\bar{y}$展开。

尽管稳健统计文献研究过此类问题的无数变体，据我们所知，尚无方法能达到我们所追求的精细界。事实上，为了获得所声称的隐私率，对于相应的稳健问题，必须在存在 η 被破坏数据的情况下，将 $\mathrm{w_{OLS}}$ 估计到误差 $\eta\sqrt{\log 1/\eta} + \sqrt{\eta}\sqrt{\mathrm{d}/\mathrm{n}}$，这与我们此前为稳健经验均值估计所设定的目标率直接类似。请注意，这大大限制了我们的选择：我们需要寻找在样本数仅与维度呈线性关系时也能奏效的算法。众所周知，使用非 SoS 方法，只需 $\mathrm{O}(\mathrm{d})$ 个样本即可达到误差 $\tilde{\mathrm{O}}(\eta)$；但即使早在 [DKS19b]，的工作中，我们就发现，在我们的设定下，必须使用平方和，才能利用 [HKMN23] 的稳健性到隐私的机制。遗憾的是，所有已知能达到误差 $\tilde{\mathrm{O}}(\eta)$ 的稳健回归平方和结果，都至少需要认证协变量的四阶矩，而这要求 $\mathrm{O}(\mathrm{d}^2)$ 的样本复杂度。这正是核心技术挑战，下文将介绍一种新颖的解决方法。

使用短平分解的计算高效算法。注意，如果 $\sigma \ll 1/\sqrt{n}$，那么先验分布足够强，只需忽略数据并从先验分布中采样，就能达到贝叶斯最优误差。相比之下，当 $\sigma \gg 1/\sqrt{n}$ （“大方差”）时，先验分布足够弱，可以放心忽略，隐私贝叶斯问题便简化为隐私普通最小二乘。大方差情形实际上是线性回归结果中技术性最强的部分，需要依次求解三个相互构建的平方和程序；我们将推迟到本文后文再讨论这一点（第 6.2.1 节）。

本概述聚焦于“方差成比例”情形，即 $\sigma \asymp 1/\sqrt{n}$，在这种情形下，先验分布和数据都不能忽略，必须结合二者才能达到贝叶斯最优误差率。如上所述，关键挑战是为 $\frac{1}{n}X\bar{y}$设计一个稳健的估计量。

为此，我们将采用一种平方和算法，它在以往方法的基础上加入了一项关键改动。首先，与该领域的常见做法一样，我们寻找数据中满足真实未污染数据集所具有结构特征的 $(1-\eta)$ 比例部分。在我们的设定中，其中一项结构特征是标准的，即要求所选点的经验协方差接近总体协方差。更准确地说，我们寻找一个 $n \times d$ 矩阵 $\hat{X}$，它在 $(1-\eta)$ 比例的行上与受污染数据集的 $\tilde{X}$ 一致，并且满足 $\frac{1}{n}\hat{X}\hat{X}^\intercal \approx \mathrm{Id}$。

我们工作的关键创新在于，还对所选 y 值施加了一种非标准的韧性要求。更准确地说，我们寻找一个向量 $\hat{y}$，它在 $(1-\eta)$ 比例的部分上与受污染数据集的 $\tilde{y}$ 值一致，并且还具有加性分解

$$\hat{y} = y' + y''$$

转化为:

- 一个范数为y' 的向量，其 $L2$ 范数为 $O(\sqrt{\eta n \log(1/\eta)})$
- 一个范数为y'' 的向量，其 $L\infty$ 范数为$O(\sqrt{\log 1/\eta})$。

我们首先解释为何预期这样的 $\hat{y}$ 及其分解存在，然后说明这对我们的算法有何用处。

如果 $\hat{y}$ 实际上由未污染数据集中的所有标签构成，那么每个值都是一个独立的标量高斯变量，其$O(1)$方差。根据标准次序统计量界，幅度超过 η 的“离群值”比例至多为$O(\sqrt{\log 1/\eta})$ (定理 C.3)。因此，剩余条目的 $(1-\eta)$比例可以视为y''。此外，可以证明，离群值中占η 比例的部分，其平方和有界于$O(\eta n \log(1/\eta))$ (定理 C.2)。我们注意到，这种分解

$\hat{y} = y' + y''$ 与短平分解这一概念有松散的关联。该概念由 [KLL+23] 在一个截然不同的背景下（用于快速的半随机矩阵补全）提出，但据我们所知，我们的工作是首次在平方和框架中运用这一概念。

最后，让我们看看这种短平分解能带来什么。最终目标是论证$\frac{1}{n}\hat{X}\hat{y}$接近$\frac{1}{n}Xy$，也就是说，通过平方和程序搜索得到的对象 $\hat{X}$ 和 $\hat{y}$ 所计算出的普通最小二乘估计量，接近使用真实未污染数据计算出的普通最小二乘估计量。我们选取 $\hat{X}$ 和 $\hat{y}$ 时，已经使它们在除$O(\eta n)$个数据点以外的所有数据点上，都与真实的X 和y 一致，因此，$\frac{1}{n}\hat{X}\hat{y}$ 与 $\frac{1}{n}Xy$ 之间的差异，实际上只来自那一小部分$O(\eta n)$ 个数据点；这些数据点要么被破坏了，要么被我们的平方和程序识别为已破坏。

Deno用S表示这组小小的“坏点”，并考察它对 $\frac{1}{n}\hat{X}\hat{y}$的贡献，即

$$\frac{1}{n}\sum_{i\in S}\hat{y}_i\hat{X}_i\,. \tag{11}$$

我们要证明这一贡献很小，也就是说，它与任意单位方向u 的内积都很小。根据柯西-施瓦茨不等式，可通过以下方式对其进行界定：

$$\left|\frac{1}{n}\sum_{i\in S}\hat{y}_i\langle\hat{X}_i,u\rangle\right| \leqslant \left(\frac{1}{n}\sum_{i\in S}\hat{y}_i^2\right)^{1/2}\cdot\left(\frac{1}{n}\sum_{i\in S}\langle\hat{X}_i,u\rangle^2\right)^{1/2}. \tag{12}$$

右侧第二项可以利用标准工具（引理 4.16）来界定，这些工具最近由 [ABMT25] 引入，使我们能够控制任意一个大小为 $O(\eta n)$ 的 $\hat{X}_i$ 子集的协方差。相反，我们证明中的主要创新在于如何处理右侧第一项。这正是我们利用 $\hat{y}$ 具有短平分解这一事实的地方。事实上，我们有

$$\frac{1}{n}\sum_{i\in S}\hat{y}_i^2 = \frac{1}{n}\sum_{i\in S}(y_i'+y_i'')^2 \leqslant \frac{2}{n}\sum_{i\in S}y_i'^2 + \frac{2}{n}\sum_{i\in S}y_i''^2 \leqslant \frac{2}{n}\|y'\|_2^2 + 2\eta\|y''\|_\infty^2 \leqslant \tilde{O}(\eta)\,, \tag{13}$$

结果表明，这对于我们的目的而言足够小。

线性回归中的信息—计算差距。 与均值估计的情形一样，对于线性回归，也存在一种计算上低效的算法，其达到的速率严格优于我们的平方和估计量（C.2节）。该算法与均值估计算法类似，涉及暴力搜索一个最大稀疏奇异值较小的特定矩阵（算法 C.7）。

随后在 6.4 节 中，我们给出低次证据，表明我们的多项式时间估计器与统计最优估计器之间的差距是固有的。与均值估计的情形相比，主要挑战在于，如何设计一对彼此难以区分的系综，要困难得多。具体而言，虽然其中一个系综有一个自然的选择——只需令 (x, y) 组成的数据集服从独立标准高斯分布——但另一系综该如何设计，才能满足以下条件，尚不明确：(1) 在计算上与前一个系综不可区分；(2) 在稳健线性回归预言机下表现得与前一个系综不同。对于第一个条件，与均值估计情形中的下界实例类似，我们希望 (x, y) 的联合分布看起来与完全高斯分布不可区分，唯独在一个隐藏方向上有所不同。事实证明，正确的选择是采用线性回归的不平衡混合，其中回归向量是某个未知随机方向的缩放形式（定义 6.22）。

通过比均值估计情形更复杂的计算，可以证明，对于任意低次数估计器，这两个系综在 $n = o(d\,\eta^2/\alpha^4)$ 个样本下难以区分（引理 6.25）。与均值估计情形中的类似论证一样，随后可通过比较稳健估计量的输出与非稳健后验均值的输出，将这一难以区分的任务归约为鲁棒贝叶斯线性回归（引理 6.24）。借助 [GH22]，的隐私到鲁棒性归约，我们便可得到私有贝叶斯回归的相应困难性结果。（定理 6.4）。

3 展望

在这项工作中，我们迈出了第一步，开始阐明高维情形下私有贝叶斯估计的计算与统计复杂性。我们在低次框架下提出了高效算法并证明了相匹配的计算困难性；我们首次表明，即使在最简单的均值估计和具有高斯先验的线性回归情形中，私有贝叶斯估计也呈现出引人关注的信息—计算差距。我们的工作开启了众多值得探索的未来方向：

超越高斯先验。 关于一般先验下高维贝叶斯推断的算法及计算-统计差距，已有丰富的研究文献（例如，参见 [ZK16、KZ24、MS+24]，其中有综述）。绘制相图，厘清何时能够进行计算高效的估计、何时能够进行任何非平凡估计，这需要运用源自统计物理学的一套精深工具；而刻画施加隐私或鲁棒性约束后相图如何变化，则是一个令人着迷的未来研究方向。

相关地，对于高斯分布这类简单的指数族先验，后验采样可直接通过后验均值估计实现，但一般情况下并非如此。对于其他先验，能否在这些更具挑战性的情形下开发隐私保护或鲁棒的后验采样算法?

其他推断任务。 还有许多其他值得研究的贝叶斯问题，可以沿着本文的思路加以探讨，例如线性二次估计、社区检测、矩阵与张量估计、群组测试等。正如第1.1节所讨论的，针对其中一些任务，已有部分关于鲁棒性的研究，但其中的保证似乎尚无法用于获得隐私保证。即使在均值估计和线性回归问题中，本研究也揭示了引人入胜的现象和技术细节，这表明在这些其他任务中还有许多有趣的见解有待发掘。

近似差分隐私。 虽然我们的两种鲁棒估计量都能通过 [HKMN23],得到近似差分隐私估计量，但这些近似差分隐私估计量与最优统计速率之间仍存在同样的准确度差距。然而，对于高效的近似差分隐私估计量，要证明相匹配的计算下界就困难得多，因为 [GH22] 的隐私到鲁棒性归约不再适用。一个自然的后续方向是，紧密刻画近似差分隐私下贝叶斯估计的计算成本，以及其他隐私概念。

紧确的计算下界。 我们针对隐私经验均值估计得到的低次数下界，仅在情形 $\beta = \exp(-d)$ 下是紧确的，因此原则上并未排除存在失败概率为常数（或仅以多项式速度随 d 缩小）的更优估计量。值得探究的是，在这些情形之一中是否存在更优的高效隐私估计器，或者我们的下界能否得到加强，从而排除这类改进。

超越分桶。 我们简要指出，维度 依赖隐藏在 $\tilde{O}(\cdot)$ 中，而 定理 5.3 中的这种依赖并非显然不可避免。一个尚待解决的问题是，对于具有各向异性（已知）协方差的数据，能否在不受分桶带来的 log维度 依赖影响的情况下，以最优方式进行经验均值估计——甚至传统的参数均值估计——即便从信息论角度来看也是如此。 [DJY+24]在 近似-DP 设定下，通过采用我们的记号，实现了参数均值估计的维度无关性：

$$n = \tilde{O}\left(\frac{\mathrm{tr}(\Lambda^2) + \|\Lambda^2\|_{\mathrm{op}} \log \frac{1}{\beta}}{\alpha^2} + \frac{(\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}} \log \frac{1}{\beta}) \sqrt{\log \frac{1}{\delta}}}{\alpha\varepsilon} + \frac{\log \frac{1}{\delta}}{\varepsilon} \right)$$

样本；事实上，可以证明，他们的算法对经验均值的估计具有由上述最后两项给出的样本复杂度。然而，$\mathrm{tr}(\Lambda)$ and 之间存在次优的耦合

$\sqrt{}$对数 $\frac{1}{\delta}$ 出现在第二项中，而且似乎很难使他们的基于修剪的方法适用于纯差分隐私设定。在各向同性情形下，通过采用 [HKMN23] 的从鲁棒性到隐私的归约，这些问题是可以解决的。不幸的是，该方法需要对体积比进行推理，而这似乎与各向异性设定不兼容（根本问题在于，缩放后的 Σ-球的体积将按秩的幂次(Σ) 而非 $\mathrm{tr}(\Sigma)$ 进行缩放，从而引入对 d 的依赖）。

4 预备知识

4.1 鲁棒性与隐私保护下的贝叶斯估计

我们首先回顾标准的（非鲁棒、非隐私）贝叶斯估计设定。给定一个从某个先验分布 $\mathscr{P}$ 中抽取的参数 θ，以及n 个从由参数 θ 参数化的分布 $\mathscr{D}_\theta$ 中独立同分布抽取的样本，贝叶斯估计旨在输出一个估计值 θ，使贝叶斯风险最小，其定义为（对于某个损失函数 ℓ）

$$\mathop{\mathbb{E}}_{\theta\sim\mathscr{P}}\mathop{\mathbb{E}}_{x_1,\dots,x_n\sim\mathscr{D}_\theta}[\ell(\hat{\theta}(x_1,\dots,x_n),\theta)]$$

在均值估计的情形下（例如 $\mathscr{D}_\theta := \mathcal{N}(\theta,\mathrm{Id}_d)$），通常将损失函数设为平方误差

$$\ell_{\mathrm{mse}} := \|\hat{\theta}-\theta\|_2^2,$$

ˆ 此时，贝叶斯风险称为均方误差，记作 $\mathrm{MSE}(\theta)$。记 $\mathbb{E}[\theta|x]$ 为给定样本后的 θ 后验均值，可得

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \mathop{\mathbb{E}}_{\theta\sim\mathscr{P}}\mathop{\mathbb{E}}_{x_1,\dots,x_n\sim\mathscr{D}_\theta}[\|\hat{\theta}-\mathbb{E}[\theta|x]+\mathbb{E}[\theta|x]-\theta\|_2^2]\\ &= \mathbb{E}[\|\hat{\theta}(x)-\mathbb{E}[\theta|x]\|_2^2]+\mathbb{E}[\|\mathbb{E}[\theta|x]-\theta\|^2]+2\mathbb{E}[(\hat{\theta}(x)-\mathbb{E}[\theta|x])(\mathbb{E}[\theta|x]-\theta)]\\ &= \mathbb{E}[\|\hat{\theta}(x)-\mathbb{E}[\theta|x]\|_2^2]+\mathbb{E}[\|\mathbb{E}[\theta|x]-\theta\|^2] \qquad (14)\end{aligned}$$

ˆ 其中，最后一个等式成立是因为 $\mathbb{E}[(\mathbb{E}[\theta|x]-\theta)\,|x]=0$；因此，只需令 θ 等于根据已观测样本得到的后验分布的均值，即可使均方误差最小。将均方误差选作损失函数的一个理由是

$$\mathop{\mathbb{E}}_{z\sim\mathscr{D}_\theta}\|\hat{\theta}-z\|_2^2 = \mathop{\mathbb{E}}_{z\sim\mathscr{D}_\theta}\|\hat{\theta}-\theta\|_2^2+\mathrm{Var}_\theta(z)$$

因此，使均方误差最小的估计量，也会使新样本上的期望平方误差最小。贝叶斯风险的“新样本视角”促使我们在线性回归设定中选择这种损失：如果$\mathscr{D}_\theta$是 $(x、y)$，其中 $x\sim\mathcal{N}(0,\mathrm{Id})$ 且 $y\sim\mathcal{N}(\theta^\intercal x,\ \varsigma^2)$，那么我们有

$$\begin{aligned}\mathbb{E}_{(x,y)\sim\mathscr{D}_\theta}\|\hat{\theta}^\intercal X-y\|_2^2 &= \varsigma^2+\mathbb{E}_{(x,y)\sim\mathscr{D}_\theta}\|\hat{\theta}^\intercal x-\theta^\intercal x\|_2^2\\ &= \varsigma^2+\mathbb{E}_{(x,y)\sim\mathscr{D}_\theta}\|\hat{\theta}-\theta\|_2^2\end{aligned}$$

因此，如果我们希望线性模型在给定新数据时具有近乎最优的预测能力，就会关注以 ℓ_2-范数估计回归向量。

需要特别指出的是，这些准确性保证本质上属于平均情形。不过，在本文中，我们将考虑隐私和鲁棒性等其他约束；前者完全是最坏情形保证，而后者则兼具平均情形和最坏情形的特征。接下来，我们将对这些目标作形式化定义。

差分隐私。我们将在 ε-差分隐私的标准"纯"设置下开展研究，该设置最初由[DMNS06b]提出。

定义 4.1 (纯差分隐私)。 如果 X，$X' \in \mathbb{R}^{n\times m}$ 相邻，即 X' 可通过更改 X 中的一行得到，则称这两个可能的数据集为相邻。若对于所有相邻的 X，X' 以及所有可测事件 E，随机机制 $\mathscr{M}$ 满足以下条件，则称其为 ε-差分隐私：

$$\mathbb{P}[\mathscr{M}(X) \in E] \leqslant e^{\varepsilon} \cdot \mathbb{P}[\mathscr{M}(X) \in E].$$

ˆ 在隐私统计估计中，机制 $\mathscr{M}$ 是估计量 $\theta(x_1, \ldots\ldots, x_n)$，而数据集 X 的各行由样本 $x_1, \ldots\ldots, x_n$ 组成。重要的是，学习算法必须对所有数据集 X 遵守隐私约束，而不仅仅是从 $\mathscr{D}_\theta$ 中抽取的"分布内"数据集。

对自适应污染的鲁棒性。本文中凡提及鲁棒性，均指在强污染模型 [DKK+19，LRV16]：

定义 4.2 (强污染模型)。 设 $0 \leqslant \eta < 1/2$。设 $\mathscr{D}_\theta$ 为某个参数化分布，并令 $x_1, \ldots\ldots, \tilde{x}_n$ 表示从 $\mathscr{D}_\theta$ 中抽取的独立同分布样本。在强污染模型中，对手可以查看所有这些样本，从中任意选取 ηn 个，并将其替换为任意点。令 $(\tilde{x}_1, \ldots\ldots, \tilde{x}_n)$ 表示受污染数据集。

先前研究该模型的工作着重于在能够访问被破坏数据的情况下，估计定义生成未污染数据的分布的参数 θ。而在本文中，我们转而考虑估计样本的依赖于数据的统计量（即后验均值）。

定义 4.3 （数据依赖统计量的鲁棒估计）。 设 $0 \leqslant \eta < 1/2$。设 $\mathscr{D}_\theta$ 是 $\mathbb{R}^d$ 上的某个参数化分布；设 $(x_1, \ldots, x_n)$ 表示从 $\mathscr{D}_\theta$ 中独立同分布抽样得到的数据集，且设 $(\tilde{x}_1, \ldots, \tilde{x}_n)$ 表示在强污染模型下对其进行 η-污染后得到的数据。设 $f\colon (\mathbb{R}^d)^n \to \mathbb{R}^m$ 表示某个依赖于数据的统计量。若 $\mathbb{P}[\|\mathscr{M}(\tilde{x}_1, \ldots, \tilde{x}_n) - f(x_1, \ldots, x_n)\|_2 \leqslant \alpha] \geqslant 1-\beta$，则称学习算法 $\mathscr{M}$ 以概率 $1-\beta$ 将 f 鲁棒地估计到误差 α 以内。

请注意，鲁棒性是一种平均情形概念，因为未污染数据集 $(x_1, \ldots, x_n)$ 被假设来自真实分布 $\mathscr{D}_\theta$；但它也是一种最坏情形概念，因为对数据集 η 比例部分的污染可以是任意的。

4.2 低次困难性

在本文中，我们将证明以下低次框架下的计算困难性结果 [BHK+19、HS17、HKP+17]所引入的框架。这里我们将简要介绍这一框架，关于详细论述请参见 [KWB19、Wei25]的综述。粗略地说，对于"性质良好"的统计估计任务，该框架预测：计算困难性与无法使用低次多项式实现的估计量这一现象相一致。对于假设检验问题，这一点可以形式化如下：

定义 4.4 （D 次-D 优势）。 设 $(\mathscr{P}_N)_{N\in\mathbb{N}}$ 和 $(\mathscr{Q}_N)_{N\in\mathbb{N}}$ 是两个分布序列。考虑一系列假设检验问题，旨在区分数据集 X 是从 $\mathscr{P}_N$ 还是从 $\mathscr{Q}_N$ 中抽样得到。对于给定的 N，该问题的D 次-D 优势 为

$$\mathrm{Adv}_{\leqslant D,N} := \sup_{f\in\mathbb{R}[X]_{\leqslant D}} \frac{\mathbb{E}_{X\sim\mathscr{P}_N}[f(X)]}{\sqrt{\mathbb{E}_{X\sim\mathscr{Q}_N}[f(X)^2]}},$$

其中 $\mathbb{R}[X]_{\leqslant D}$ 表示 X 中总次数至多为 D 的多项式环。标准计算（参见*[KWB19，命题 1.15]*）表明，若 $L_{\leqslant D,N}$ 表示通过将似然比 $L_N := \mathrm{d}\mathscr{P}_N \mathrm{d}\mathscr{Q}_N$ 投影到关于 $L_2(\mathscr{Q}_N)$的 D 次多项式空间而得到的低次似然比，则

$$\mathrm{Adv}_{\leqslant D,N} = \|L_{\leqslant D,N}\|_{L_2(\mathscr{Q}_N)}. \tag{15}$$

如果 $D = \omega(\log N)$，我们就称这一系列假设检验问题为低次难解。

$$Adv_{\leqslant D,N} = O_N(1)$$

当上下文中 N 的含义明确时，我们省略上述下标中的 N。

尽管目前尚无正式猜想表明，在对分布施加一组具体条件后，低次困难性就意味着计算困难性，$\mathscr{P}_N$，$\mathscr{Q}_N$，[2]但对于缺乏代数结构的平均情形问题，它在预测计算阈值方面表现出惊人的准确性，并且通常与大量支持困难性的互补证据相一致（例如统计查询下界 [BBH+20]、景观分析 [BEAH+22]、平方和下界 [BS14、HKP+17]、消息传递失效 [MW25]）。在低次框架错误地预测计算困难性的情况下，问题通常具有代数 [HW20、BHJK25] 或格 [DK22、ZSWB22] 结构；鉴于我们所考虑的问题中并未出现这种结构，我们认为，在我们的设定下，低次困难性是计算困难性的合适替代指标。

更准确地说，我们将通过展示从低次难解问题到隐私贝叶斯估计的计算高效（但不一定是低次）归约，来证明隐私贝叶斯估计在计算上是困难的。

下面，我们汇总一些用于证明低次下界的标准工具。

定义 4.5（非高斯成分分析）。设 A 是 $\mathbb{R}$ 上的一个分布，且 $d(N)$ 和 $n(N)$ 是参数 N 的递增函数，分别对应维度和样本复杂度。非高斯成分分析（NGCA）是一个判别问题，用于区分来自 $\mathscr{P}_N$ 和 $\mathscr{Q}_N$ 的样本，定义如下。令 $\mathscr{Q}_N = \mathcal{N}(0, \mathrm{Id}_{d(N)})$。为从 $\mathscr{P}_N$ 中抽样，先从 $\mathbb{S}^{d(N)-1}$ 中均匀分布地抽取一个单位向量 v，然后从其在 v 上的正交投影为 A、在 v 的正交补上的正交投影为 $\mathcal{N}(0,\mathrm{Id}_{d(N)-1})$ 的乘积分布中独立同分布地抽取 $y_1,\ldots,y^{n(N)}$。

引理 4.6 ([DHPT25]，《次数D 优势与厄米特系数的增长》中定理 7.2；在 [MW21]中隐含提及**)**。假设单变量分布 A 满足以下条件，其中 j 为偶数，$_\star \in \mathbb{N}$，参数为 κ，$\tau \in \mathbb{R}_+$：

- $\mathbb{E}_{X\sim A}\, h_j(X) = 0 \quad for\ j \in [j_\star - 1]$.
- *For* $j \geqslant j_\star$, $\left|\mathbb{E}_{X\sim A} h_j(X)\right|^2 \leqslant j^j\, \kappa\, \tau^{-j/j_\star}$.

那么，如果 $n\kappa \geqslant 1$ 且

$$n \leqslant \frac{1}{\mathrm{poly}(D)} \frac{d^{j_\star/2}\tau}{\kappa},$$

隐藏分布为 A 的 *NGCA* 问题的 D 次优势满足 优势$_{\leqslant D,\ n} \leqslant O(1)$。

事实 4.7 ([DHPT25]，《厄米特多项式的性质》中的事实 3.10**)**。对于标准化的概率论厄米特多项式 $h_j(X)$，我们有以下性质：

- *For any* x, $\mathbb{E}_{X\sim\mathcal{N}(x,1)}[h_j(X)] = x^j/\sqrt{j!}$.

[2]请参见 [HW20、BHJK25]，其中给出了此前尝试表述此类猜想的反例。

- 对于 $\sigma^2 \leqslant 1$，令 $\rho^2 := 1-\sigma^2$。则 $\mathbb{E}_{X\sim\mathcal{N}(x,\ \sigma^2)}[h_j(X)] = \rho^j h_j(x/\rho)$。
- 存在一个普适常数 C，使得对于任意 $|x|^2 \leqslant 2i$，$hi(x)^2 \leqslant Cex^{2}/2/\sqrt{i}$。特别地，对于所有 $|x| \leqslant 1$，$hi(x)^2 \leqslant 3C/\sqrt{i}$。

4.3 平方和背景

在本文中，我们使用平方和（SoS）半定规划层级 [BS14、RSS18]来进行算法设计与分析。平方和证明到算法框架已被证明有助于在算法统计学中取得许多最优或最先进的结果。我们在此简要介绍伪分布、平方和证明和平方和算法。

4.3.1 平方和证明与算法

伪分布。我们可以用其概率质量函数 μ 表示定义在 $\mathbb{R}^n$ 上的有限支撑概率分布：$\mathbb{R}^n \to \mathbb{R}$，使得 $\mu \geqslant 0$ 且 $\sum_{z\in\mathrm{supp}(\mu)} \mu(z) = 1$。我们将伪分布定义为此类概率质量分布的推广，放宽约束 $\mu \geqslant 0$，只要求 μ通过某些低次数非负性检验。

定义 4.8（伪分布）。在 $\mathbb{R}^n$ 上的层级为 ℓ伪分布 μ 是一个有限支撑函数 $\mu: \mathbb{R}^n \to \mathbb{R}$，满足 $\sum_{z\in\mathrm{supp}(\mu)} \mu(z) = 1$ 且 $\sum_{z\in\mathrm{supp}(\mu)} \mu(z) f(z)^2 \geqslant 0$，对每个次数至多为 $\ell/2$ 的多项式 f 均成立。

我们可以像定义有限支撑概率分布的期望那样，定义伪分布的期望。

定义 4.9（伪期望）。给定 $\mathbb{R}^n$ 上的伪分布 μ，我们将函数 f 的伪期望 定义为 $f: \mathbb{R}^n \to \mathbb{R}$，即

$$\tilde{\mathbb{E}}_\mu f := \sum_{z\in\mathrm{supp}(\mu)} \mu(z) f(z). \tag{16}$$

下面的定义将伪分布满足多项式约束系统这一概念形式化。

定义 4.10（受约束伪分布）。设 $\mu: \mathbb{R}^n \to \mathbb{R}$ 是定义在 $\mathbb{R}^n$ 上的层级为 ℓ 的伪分布。设 $\mathcal{A} = \{f_1 \geqslant 0, \ldots, f_m \geqslant 0\}$ 是一个多项式约束系统。若对每个多重集 $S \subseteq [m]$ 以及每个满足 $\deg(h) + \sum_{i\in S} \max\{\deg(f_i), r\} \leqslant \ell$ 的平方和多项式 h，μ 满足 $\mathcal{A}$，则称其在层级 r 上满足 μ r $\mathcal{A}$，

$$\tilde{\mathbb{E}}_\mu h \cdot \prod_{i\in S} f_i \geqslant 0. \tag{17}$$

如果 μ $_0$ $\mathcal{A}$，我们就称 μ 满足 $\models\mathcal{A}$，并记作 μ $\models$ $\mathcal{A}$（不再具体说明次数）。

我们指出，如果 μ 是一个实际的有限支撑概率分布，那么当且仅当 μ 支撑在满足 $\mathcal{A}$的解上时，μ $\models$ $\mathcal{A}$ 成立。

平方和证明。我们将平方和证明引入为伪分布的对偶对象，可用于推理伪分布的性质。若存在多项式 (q_i)，使得 $p = \sum_i q_i^2$，则称多项式 p 为平方和多项式。

定义 4.11（平方和证明）。一个关于多项式约束系统 $\mathcal{A} = \{f_1 \geqslant 0, \dots, f_m \geqslant 0\}$ 蕴含 $q \geqslant 0$ 的平方和证明由平方和多项式 $(p_S)_{S \subseteq [m]}$ 组成，满足[3]

$$q = \sum_{\text{multiset } S \subseteq [m]} p_S \cdot \prod_{i \in S} f_i .$$

如果存在这样的证明，我们就说 $\mathcal{A}$（由平方和）证明 $q \geqslant 0$ 的次数不超过 ℓ，记作 $\mathcal{A} \vdash_\ell q \geqslant 0$。为明确证明中量化的变量，我们通常写作 $\mathcal{A}(z) \vdash_\ell z\, q(z) \geqslant 0$。

下面的引理表明，平方和证明使我们能够推导出满足某些约束的伪分布的性质。

引理 4.12。设 μ 为一个伪分布，$\mathcal{A}, \mathcal{B}$ 为多项式约束系统。假设存在一个平方和证明 $\mathcal{A} \vdash_r \mathcal{B}$。若 $\mu \models_r \mathcal{A}$，则 $\mu \models_{r \cdot r' + r'} \mathcal{B}$。

平方和算法。给定一个多项式约束系统，平方和算法使用半定规划在满足该多项式系统的伪分布空间中进行搜索。

由于半定规划只能近似求解，我们只能找到近似满足给定多项式系统的伪分布。我们称一个层级为-ℓ 的伪分布近似满足一个多项式系统，是指公式 (17) 中的不等式允许加性误差 $2^{-n^\ell} \cdot \|h\| \cdot \prod_{i \in S} \|f_i\|$，其中 $\|\cdot\|$ 表示单项式基下多项式系数的欧几里得范数[4]。

定理 4.13 (平方和算法)。存在一个运行时间为 $(n+m)^{O(\ell)}$ 的算法，对于任意一个显式有界且可满足的、包含 n 个变量和 m 个多项式约束的系统，输出一个层级为-ℓ的伪分布，使其近似满足 $\mathcal{A}$。

注记 4.14 (近似误差与位复杂度)。对于仅近似满足多项式系统的伪分布，我们仍可以像处理引理 *4.12*那样，利用平方和证明进行推理。为避免近似误差在推理过程中不断放大，我们需要确保平方和证明中系数的位复杂度有多项式界。

4.3.2 有用的平方和不等式

本节将介绍一些对本文有用的基本 SoS 证明。

引理 4.15. 给定常数 $C \geqslant 0$，我们有

$$\left\{-C \leqslant z \leqslant C\right\} \Big|_{\frac{z}{2}} z^2 \leqslant C^2 .$$

[3]这里我们遵循以下约定：$\prod_{i \in S} f_i = 1$ 对于 $S = \emptyset$。[4]此处范数的选择并不重要，因为因子 2^{-n^ℓ} 会压过选择其他范数所带来的影响。[5]如果一个多项式约束系统包含形如 $\|z\|^2 \leqslant M$ 的约束，则称其为显式有界。[6]这里我们假设 $\mathcal{A}$ 中约束的位复杂度为 $(n+m)^{O(1)}$。

引理 4.16 (平方和选择器引理)**.** 设 $a_1,\ldots,a_n \in \mathbb{R}$ 和 $B \in \mathbb{R}$。假设对于任意子集 $S \subseteq [n]$，满足$|S| \leqslant k$，我们有 $|\sum_{i\in S} a_i| \leqslant B$。那么

$$\left\{0 \leqslant x_1,\ldots,x_n \leqslant 1,\ \sum_i x_i \leqslant k\right\} \vdash_1^{x_1,\ldots,x_n} \left|\sum_{i=1}^n a_i x_i\right| \leqslant B\,.$$

证明。 我们首先证明 $\sum_{i=1}^n a_i x_i \leqslant B$。不失一般性，假设 $a_1 \geqslant \ldots \geqslant a_n$。**情形 1：** $a_k \geqslant 0$。显然可知 $\{0 \leqslant x_1,\ldots,x_n \leqslant 1,\ \sum_i x_i \leqslant k\} \vdash_1$

$$\begin{aligned}
B - \sum_{i=1}^n a_i x_i &\geqslant \sum_{i=1}^k a_i - \sum_{i=1}^n a_i x_i && (\textstyle\sum_{i=1}^k a_i \leqslant B)\\
&= \sum_{i=1}^k a_i(1-x_i) - \sum_{i=k+1}^n a_i x_i\\
&\geqslant \sum_{i=1}^k a_k(1-x_i) - \sum_{i=k+1}^n a_k x_i && (0 \leqslant x_i \leqslant 1)\\
&= a_k \cdot \left(k - \sum_{i=1}^n x_i\right)\\
&\geqslant 0\,. && (\textstyle\sum_i x_i \leqslant k)
\end{aligned}$$

情形 2： $a_k < 0$。令 ℓ 为满足 $a_\ell \geqslant 0$的最大下标。（注意 $\ell \in \{0,1,\ldots,k-1\}$。）于是

$$\begin{aligned}
B - \sum_{i=1}^n a_i x_i &\geqslant \sum_{i=1}^\ell a_i - \sum_{i=1}^n a_i x_i && (\textstyle\sum_{i=1}^\ell a_i \leqslant B)\\
&= \sum_{i=1}^\ell a_i(1-x_i) + \sum_{i=\ell+1}^n (-a_i)x_i\\
&\geqslant 0\,,
\end{aligned}$$

其中，在最后一个不等式中，我们用到了$a_1,\ldots,a_\ell \geqslant 0$和 $a_{\ell+1},\ldots,a_n \leqslant 0$，以及对所有$i$，$0 \leqslant x_i \leqslant 1$。

注意，我们可以将上述相同的论证应用于 $-a_1,\ldots,-a_n$，并得出 $\sum_{i=1}^n (-a_i)x_i \leqslant B$，或者等价地，$\sum_{i=1}^n a_i x_i \geqslant -B$。

因此，

$$\left\{0 \leqslant x_1,\ldots,x_n \leqslant 1,\ \sum_i x_i \leqslant k\right\} \vdash_1^{x_1,\ldots,x_n} \left\{-B \leqslant \sum_{i=1}^n a_i x_i \leqslant B\right\}. \qquad \square$$

4.3.3 近似满足线性算子

类似于 [HKMN23], ，我们给出近似满足线性算子的定义，以处理平方和指数机制中的数值问题。

定义 4.17 (τ-近似满足线性算子)。 令 $n \in \mathbb{N}$ 和 $T \in [n]$。令 $\mathscr{A}$ 为变量 $w = (w_1,\ldots,w_n)$ 以及 z（可能有很多个）上的多项式不等式系统，使得

$$\mathscr{A}_T = \{q_1(w,z) \geqslant 0,\ldots,q_m(w,z) \geqslant 0\} \cup \left\{\sum_{i=1}^n w_i \geqslant n - T\right\}.$$

令 $\tau > 0$。我们称 th 在线性算子 $\mathscr{L}$ τ 下，若满足以下条件，则在次数 D 上近似满足 $\mathscr{A}_T$:

1. $\mathscr{L}1 = 1$。

2. 对所有满足$\deg(p^2) \leqslant D$且 $\|\mathscr{R}(p)\|_2 \leqslant 1$ 的多项式 p（其中 $\mathscr{R}(p)$ 是 p 的系数的向量表示），均有 $\mathscr{L}p^2 \geqslant -\tau T$。

3. 对所有 $i = 1, \ldots, m$ 以及满足$\deg(p^2 \cdot q_i) \leqslant D$且 $\|\mathscr{R}(p)\|_2 \leqslant 1$ 的多项式 p（其中 $\mathscr{R}(p)$ 是 p 的系数的向量表示），均有 $\mathscr{L}p^2 q_i \geqslant -\tau T$。

4. 对每个满足$\deg\left(p^2 \cdot \sum_{i=1}^{n} w_i - n + T\right) \leqslant D$且 $\|\mathscr{R}(p)\|_2 \leqslant 1$ 的多项式 p（其中 $\mathscr{R}(p)$ 是 p 的系数的向量表示），均有

$$\mathscr{L}\left(p^2 \sum_{i=1}^{n} w_i - n + T\right) \geqslant -5\tau T n.$$

4.4 鲁棒性到隐私的联系

我们将使用鲁棒性到隐私框架 [HKMN23] 将鲁棒算法转换为隐私算法。我们将同时使用这一归约的低效和高效实现。

引理 *4.18*（低效鲁棒性到隐私（ [*HKMN*23]中的引理 *2.1* 和 *2.2*））。设 $\eta_0 < \eta^\star \in [0, 1]$，其中 $\eta^\star$ n 是一个充分大的常数。对于每个 ε、$\delta > 0$，存在一个 $(O(\varepsilon)$、$O(e^{2\varepsilon}\delta))$-差分隐私机制，该机制对于任意 $\theta^* \in \mathbb{R}^d$，接收 $x_1, \ldots, x_n \sim p^{\theta^*}$，并以概率 $1 - \beta$输出 θ，使得$\|\theta - \theta^*\| \leqslant 2\alpha(\eta_0)$，如果

$$n \geqslant O\left(\max_{\eta_0 \leqslant \eta \leqslant \eta^*} \frac{d \cdot \log\left(\frac{2\alpha(\eta)}{\alpha(\eta_0)}\right) + \log(1/\beta) + \log(\eta n)}{\eta\varepsilon} + \frac{\log(1/\delta)}{\eta^*\varepsilon}\right).$$

此外，如果满足以下条件，则存在一个纯 ε-差分隐私机制，可以获得相同的保证

$$n \geqslant O\left(\max_{\eta_0 \leqslant \eta \leqslant 1} \frac{d \cdot \log\left(\frac{2\alpha(\eta)}{\alpha(\eta_0)}\right) + \log(1/\beta)}{\eta\varepsilon}\right).$$

定理 4.19 (纯差分隐私的高效鲁棒性（ [HKMN23]中的定理 4.1））。 设$0 < \eta,\ r < 1 < R$ 为固定参数。假设我们有一个评分函数 $S(\theta, \mathscr{Y}) \in [0, n]$，它以数据集 $\mathscr{Y} = \{y_1, \ldots, y_n\}$和参数 $\theta \in \Theta \subset \mathbb{B}(\mathbb{R})^d$（其中 Θ是凸的，且包含在半径为 R 的球内），并具有以下性质：

- （有界敏感性 ）对于任意两个相邻数据集 $\mathscr{Y}$，$\mathscr{Y}'$以及任意 $\theta \in \Theta$, $|S(\theta, \mathscr{Y}) - S(\theta, \mathscr{Y}')| \leqslant 1$.
- *(Q*拟凸性）对于任意固定数据集 $\mathscr{Y}$、任意 θ，$\theta' \in \Theta$ 以及任意 $0 \leqslant \lambda \leqslant 1$，我们有

$$S\left(\lambda\theta + (1-\lambda)\theta', \mathscr{Y}\right) \leqslant \max\left(S(\theta, \mathscr{Y}), S(\theta', \mathscr{Y})\right).$$

- （可高效计算）对于任意给定的 $\theta \in \Theta$ 和数据集 $\mathscr{Y}$，我们都可以在 $\mathrm{poly}\left(n,\ d,\ \log R\, r,\ \log \gamma^{-1}\right)$ 时间内，将 $S(\theta, \mathscr{Y})$ 计算到误差γ 以内，其中 $\gamma > 0$ 任意。
- （鲁棒算法找到低分点）对于给定的数据集 $\mathscr{Y}$，令 $T = \min_{\theta_0 \in \Theta} S(\theta_0, \mathscr{Y})$。那么，我们可以在$\mathrm{poly}\left(n,\ d,\ \log R\, r\right)$时间内，找到某个点 θ，使得对于距离 θ 不超过 r 的所有 θ'，$S(\theta',\ \mathscr{Y}) \leqslant T + 1$。
- （体积）对于任意给定的数据集 $\mathscr{Y}$和 $\eta' \geqslant \eta$，令 $V_{\eta'}(\mathscr{Y})$ 表示分数不超过 $\eta' n$ 的点 $\theta \in \Theta \subset \mathbb{R}$ d 的 d 维勒贝格体积。（注意，$V_1(\mathscr{Y})$ 是 Θ 的完整体积。）

那么，对于大小为 n 的数据集，我们有一个纯 ε-DP 算法 $\mathscr{M}$，其运行时间为 多项式$\left(n、d、\log \frac{R}{r}\right)$，并具有以下性质。对于任意数据集 $\mathscr{Y}$，如果存在 θ，满足 $S(\theta,\mathscr{Y}) \leqslant \eta n$，且

$$n \geqslant \Omega\left(\max_{\eta':\eta\leqslant\eta'\leqslant 1} \frac{\log\left(\frac{V_{\eta'}(\mathscr{Y})}{V_{\eta}(\mathscr{Y})}\right)+\log\left(\frac{1}{\beta\cdot\eta}\right)}{\varepsilon\cdot\eta'}\right),$$

然后 $\mathscr{M}(\mathscr{Y})$ 以概率 $1-\beta$输出一些得分至多为 $2\eta n$ 的 $\theta\in\Theta$。

注释 4.20 （将鲁棒性应用于贝叶斯估计中的隐私）。 我们注意到，*[HKMN23]* 中的鲁棒性到隐私的引理和定理，针对的是与数据无关的固定 θ^*。然而，仔细考察后可以发现，对于输入变量的任意固定函数的估计，同样的引理和定理也成立。更确切地说，设 f 表示任意固定函数，并设 $\{x_1^*, \ldots, x^*n\}$ 表示未污染样本。那么，如果将 θ^* 替换为输入变量的任意固定函数 $f(x_1^*, \ldots, x^*n)$，同样的引理和定理仍然成立。

为了证明差分隐私的计算下界，我们还需要以下反方向的定理归约，即从隐私到鲁棒性。

定理 4.21 （隐私到鲁棒性归约， [GH22]中的定理 3.1）。 设 $\mathscr{M}: \mathscr{X}^* \to \mathscr{O}$ 是一个ε-隐私映射，将数据集 $\mathscr{X}^*$映射为输出 $\mathscr{O}$。对于每个数据集 $X_1,\ldots, X_n$，令 $G_{X_1,\ \ldots,\ X_n}\subseteq\mathscr{O}$为一组良好输出。假设 $\mathscr{M}(X_1,\ldots, X_n)\in G_{X_1,\ \ldots,\ X_n}$ 的概率至少为 $1-\beta$，其中 $\beta=\beta(n)$。那么，对于每个 $n\in\mathbb{N}$，在 n 元数据集上， $\mathscr{M}$ 对于任意 $\eta(n)$-比例输入的对抗性污染均具有鲁棒性，其中

$$\eta(n)=O\left(\frac{\log(1/\beta)}{\varepsilon n}\right),$$

这意味着，对于每个 $X_1,\ldots, X_n$ 以及仅在 ηn 个元素上存在差异的 $X_1', \ldots, X'n$， $\mathscr{M}(X_1', \ldots, X'n)\in G_{X_1,\ \ldots,\ X_n}$的概率至少为 $1-\beta^{\Omega(1)}$。

4.5 矩阵的稀疏谱范数

在分析贝叶斯线性回归算法时，我们使用矩阵的稀疏谱范数。

定义 4.22 (矩阵的稀疏谱范数)。 设 $X\in\mathbb{R}^{d\times n}$。对于指标集 $S\subseteq[n]$，，令 $X_S\in\mathbb{R}^{d\times|S|}$表示由指标属于 S 的列构成的子矩阵。对于 $k\in[n]$，定义 k-稀疏谱范数

$$\|X\|_{\mathrm{sp}}(k) := \max_{\substack{S\subseteq[n]\\ |S|\leqslant k}} \|X_S\|_{\mathrm{op}}.$$

5 鲁棒且私密的贝叶斯均值估计

本节将证明贝叶斯高斯均值估计的主要结果。模型如下:

定义 5.1 (高斯先验下的贝叶斯均值估计)。 从先验分布$\mathcal{N}(0,\Sigma)$中抽取一个参数 μ，然后从 $\mathcal{N}(\mu,\mathrm{Id}_d)$中独立同分布地抽取 n 个样本 $x=(x_1,\ldots,x_n)$。给定这些样本，我们的目标是私密地估计后验均值 $\mu_{\mathrm{post}}:=\mathbb{E}[\mu|x]$，使其达到 ℓ_2 范数。

我们针对这一问题的主要正面结果是定理 1.1，为方便起见，在此重述:

定理 （定理 1.1 的重新表述）。给定 μ 从先验分布 $\pi=\mathcal{N}(0,\Sigma)$中抽取，若数据 x 由 n 个 $\geqslant\tilde{\Omega}(\frac{d}{\varepsilon}\log(\sqrt{\mathrm{Tr}(\Sigma)/\alpha}))$个来自 $\mathcal{N}(\mu,\mathrm{Id})$的独立同分布样本组成，则存在

1. 一个计算上低效的 ε-*DP* 估计量 $\hat{\mu}(x)$，使得其以至少 $1-\beta$ 的概率满足，

$$\|\hat{\mu}(x)-\mathbb{E}[\mu \mid x]\|_2 \leqslant \tilde{O}\left(\frac{\operatorname{tr}(\Lambda)+\|\Lambda\|_{\mathrm{op}} \log(1/\beta)}{\varepsilon n}\right) := \tilde{O}(\alpha_{\mathrm{stat}}) \tag{18}$$

2. 一个多项式时间 ε-*DP* 估计量 $\hat{\mu}(x)$，满足其概率至少为 $1-\beta$，

$$\|\hat{\mu}(x)-\mathbb{E}[\mu \mid x]\|_2 \leqslant \tilde{O}\left(\frac{\|\Lambda\|_{4/3}+\|\Lambda\|_{\mathrm{op}} \log^{3/4}(1/\beta)}{\varepsilon^{1/2} n^{3/4}}+\alpha_{\mathrm{stat}}\right) := \tilde{O}(\alpha_{\mathrm{comp}}+\alpha_{\mathrm{stat}}). \tag{19}$$

其中 $\Lambda := (\text{单位矩阵} + \frac{1}{n}\Sigma^{-1})^{-1}$，$\|\Lambda\|_{4/3}$ 表示 Λ 的 *Schatten* 4/3-范数。此处的 $\tilde{O}$ 隐去了对数 $1/\alpha$、双重对数 $1/\beta$ 和对数 d 因子。当 $\Sigma = \text{单位矩阵}$ 时，我们不会损失对数 d 因子。

我们还将证明，在低次困难性框架下，统计速率与计算速率之间的差距是根本性的：

定理 5.2（贝叶斯后验均值估计的信息—计算差距）。考虑各向同性先验 $\mu \sim \mathcal{N}(0, \sigma^2 \text{单位矩阵}_d)$ 的情形，并假设 $\beta < \frac{1}{12}$、$d > 2$ 和 $\alpha \leqslant \frac{\sqrt{2\pi}}{63}\sigma$。存在一个判别问题 H，它是次数为 n 的低次难解问题；对于在先验单位矩阵 d 下进行私有后验均值估计的多项式时间算法，

$$n = o\left(\frac{\min(d, \log \frac{1}{\beta})}{\alpha^{4/3}\varepsilon^{2/3}}\right)$$

样本，就会存在一个针对 Π 的多项式时间区分算法。

5.1 重新表述为私有经验均值估计

为了证明这些结果，我们首先将其直接归约为私有经验均值估计任务。

不失一般性，假设 Σ 可逆（否则，我们只需通过换基排除其核）。通过标准计算可知，后验均值与经验均值 $\bar{x} := \frac{1}{n}\sum_i x_i$ 之间存在简单的线性对应关系：

$$\mu_{\mathrm{post}} = \left(\mathrm{Id} + \frac{1}{n}\Sigma^{-1}\right)^{-1} \bar{x} =: \Lambda \bar{x}.$$

因此，在 L_2 范数下获得后验均值的一个估计值 $\hat{\mu}$，对应于在某个马氏范数下获得经验均值的一个估计值 $\hat{x}$：

$$\|\hat{\mu} - \mu_{\mathrm{post}}\|_2 = \|\Lambda(\hat{x} - \bar{x})\| = \|\hat{x} - \bar{x}\|_{\Lambda^{-2}}$$

此外，请注意，对数据集应用（已知的）线性变换 Λ 后，这等价于在 L_2 范数下估计一个由来自 $\mathcal{N}(0, \Lambda^2)$ 的样本构成的数据集的经验均值。

因此，本节余下部分将重点讨论协方差为 Λ^2 的高斯分布的经验均值估计这一后一个问题。定理 1.1 可直接由以下上界推出（详见 附录 A.1）：

定理 5.3（私有经验均值估计的上界）。固定一个公开参数 $R > 1$。对于所有 $\mu \in \mathbb{R}^d$，给定来自 $\mathcal{N}(\mu, \Lambda^2)$、经验均值为 $\|\bar{x}\|_2$ 的 n 个独立同分布样本，若 $\|\bar{x}\|_2 \leqslant R$，

1. 存在一个（计算上低效的） ε-隐私估计量，其至多使用

$$n = \tilde{O}\left(\frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}} \log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(R/\alpha)}{\varepsilon}\right),$$

然后，估计量将以至少 $1-\beta$ 的概率输出 $\hat{x}$，使得 $\|\hat{x} - \bar{x}\|_2 \leqslant \alpha$。

2. 存在一个多项式时间的 ε-差分隐私估计量，其使用量至多为

$$n = \tilde{O}\left(\frac{\mathrm{tr}(\Lambda^{4/3}) + \|\Lambda^{4/3}\|_{\mathrm{op}} \log(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}} \log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(R/\alpha)}{\varepsilon}\right),$$

那么同样的结论仍然成立。

~这里，O 隐去了 α, n, d 的对数因子。如果 Λ 与 Id_d 成比例，那么它仅隐去了 α, n 的对数因子。

对 Λ 的特征值进行简单的分桶论证可知，要证明一般 Λ 情形下的这些界，只需证明特殊情形 Λ 与 Id 成比例$_d$，代价是增加一个对数因子 d (详情请参见附录 A.2)。

定理 5.3 中的统计速率在对数 $1/\alpha$ 因子范围内是最优的，因为改进该速率将违反现有的差分隐私参数均值估计下界。更准确地说，该速率中的最后一项由标准打包论证可知是必要的，$\mathrm{tr}(\Lambda)$ 项由 [DJY+24定理 1 可知是必要的，3], 而 $\|\Lambda\|_{\mathrm{op}}\log(1/\beta)$ 项则由 [KV17定理 6 中的单变量下界可知是必要的，2] 该下界应用于 Λ。

此外，我们 提供低次证据，表明定理 5.3 中的计算速率是最优的 l.

定理 5.4 （私有经验均值估计中的信息—计算差距）。设 $\beta < \frac{1}{3}$ 且 $d > 1$。存在一个判别问题 Π（见 *定义 5.12*），它是次数为 d 的 $o(1)$ 困难问题；因此，如果存在一个多项式时间 ε-*DP* 算法，能够对各向同性高斯分布数据进行经验均值估计，并达到误差 α，且其概率为 $1-\beta$，使用

$$n = o\left(\frac{\min(d, \log\frac{1}{\beta})}{\alpha^{4/3}\varepsilon^{2/3}}\right)$$

个样本，那么就会存在一个针对 Π 的高效判别算法。

尽管上述结果本身很有意义，但它只是一个最坏情况下界，不足以证明定理 5.2，后者即使在 $\mu \sim \mathcal{N}(0, \mathrm{Id}_d)$ 上的平均情况下也成立。我们证明了这一最坏情况下界的一个平均情况变体，该变体在 α 和 β足够小时成立。完整证明见附录 A.1。

5.2 私有各向同性经验均值估计的统计速率

在本节中，我们证明定理 5.3 中 Λ 为各向同性时的统计界。下面，我们首先证明各向同性高斯分布的鲁棒经验均值估计的统计上界，然后应用 [HKMN23]提出的隐私到鲁棒性归约。

定理 5.5 （稳健经验均值估计的统计速率）。 假设 $0 \leqslant \eta < 1/2$ 与 $1/2$保持正距离。存在一个（计算上低效的） η-稳健估计量，对于所有 $\mu \in \mathbb{R}^d$，给定来自 $\mathcal{N}(\mu, \mathrm{Id}_d)$ 的 n 个独立同分布样本，其经验均值为 $\bar{x}$，该估计量输出 $\hat{x}$，使得 $\|\hat{x} - \bar{x}\|_2 \leqslant \alpha$，且其概率至少为 $1-\beta$，对于

$$\alpha = O\left(\eta\sqrt{\log(1/\eta)} + \sqrt{\eta} \cdot \sqrt{\frac{d + \log(1/\beta)}{n}}\right). \tag{20}$$

请注意，这意味着，要在稳健估计经验均值时达到误差 $\widetilde{O}(\eta\sqrt{\log(1/\eta)})$，只需使用 $\widetilde{O}(d/\eta)$ 个样本，而不是通常的“参数”速率 $\widetilde{O}(d/\eta^2)$。此外，正如预期，随着 $\eta \to 0$，误差趋于零。

定理 5.5 的证明是一个基本的韧性型论证 [SCV17]:

证明。 我们按如下方式构造估计量：令$x_1,\ldots,x_n$ 为均值为 $\widetilde{x}$的未污染输入样本，并令 $\widetilde{x}_1,\ldots,\widetilde{x}_n$ 为污染输入样本。我们的估计量将在满足以下约束的向量值变量 $y_1,\ldots,y_n$ 上进行搜索：

- $|\{i : \widetilde{x}i \neq yi\}| \leqslant \eta n$
- 对于所有大小至多为 $2\eta S \subseteq [n]$ 的子集：

$$\left\| \frac{1}{|S|} \sum_{i \in S} (y_i - \bar{y}) \right\|_2 \leqslant C \sqrt{\frac{d + \log \frac{1}{\beta}}{\eta n} + \log \frac{1}{\eta}}$$

其中 $C > 0$为充分大的绝对常数，且 $\mathrm{y} = {}^{1}\mathrm{n}\sum \mathrm{i}\,\mathrm{y}$ $\underline{i}_{\circ}$ 如果上述条件不可行，估计量将输出 $\perp$。然而，通过标准的集中性以及对所有大小至多为 $2\eta n$ 的 $O(1/\eta)^{\widetilde{\eta}n}$个子集 S 使用并集界，约束由真实值 $(y_1,\ldots,y_n) = (x_1,\ldots,x_n)$满足。接下来，令 $(y_1,\ldots,y_n)$和 $(y'_1,\ldots,y'n)$为两个都满足上述约束的赋值，并令 $S = \{i : y_i \neq y'i\}$（注意 $|S| \leqslant 2\eta n$）。定义 $y' = {}^{1}\mathrm{n}\sum \mathrm{i}\,y'i$。我们将证明，对于 α 中由式（20）给出的 $\|\widetilde{y} - \widetilde{y}'\|_2 \leqslant O(\alpha)$，这将意味着满足上述约束集合的任何赋值都将达到定理中所声称的界。然后

$$\begin{aligned}
\|\bar{y} - \bar{y}'\|_2 &= \frac{1}{n} \left\| \sum_{i \in S} (y_i - y'_i) \right\|_2 \\
&= \frac{1}{n} \left\| \sum_{i \in S} (y_i - \bar{y}) - \sum_{i \in S} (y'_i - \bar{y}') + 2\eta n (\bar{y} - \bar{y}') \right\|_2 \\
&\leqslant 2C\eta \sqrt{\frac{d + \log \frac{1}{\beta}}{\eta n} + \log \frac{1}{\eta}} + 2\eta \cdot \|\bar{y} - \bar{y}'\|_2 .
\end{aligned}$$

重新整理并利用 η 与 $1/2$ 相距有界的假设后，我们得到 $\|\widetilde{y} - \widetilde{y}'\|_2 \leqslant O(\alpha)$，如所需。

□

现在，我们通过结合定理 5.3在 $\Lambda = \mathrm{Id}$ 时的第一部分，定理 5.5 和引理 4.18，证明私有各向同性经验均值估计的统计速率。

定理 5.3 第一部分的证明。令 $\eta^* := 1/3$，并令 $\alpha(\eta)$ 为稳健经验均值估计中定理 5.5 的误差项；对于更大的 η，令 $\alpha(\eta) = \mathrm{R}$。令 α 为隐私算法的目标准确度，我们假设它小于一个足够小的常数；再令 η_0使得 $\alpha(\eta_0) = \alpha$。现在根据引理 4.18，存在一个计算上低效的 ε-DP 算法，它从 $\mathcal{N}(\mu, \mathrm{Id})$中获取 n 个样本，并输出 $\hat{\mathrm{x}}$，使得 $\|\hat{\mathrm{x}} - \widetilde{\mathrm{x}}\|_2 \leqslant 2\alpha(\eta_0) = 2\alpha$的概率为$1 - 2\beta$，只要

$$n \geqslant \max_{\eta_0 \leqslant \eta' \leqslant 1} \frac{d \cdot \log \frac{2\alpha(\eta')}{\alpha(\eta_0)} + \log(1/\beta)}{\eta' \varepsilon} . \tag{21}$$

这个最大值不超过 $\eta_0 \leqslant \eta' \leqslant \eta^*$ 和 $\eta^* \leqslant \eta' < 1$上的最大值之和，即

$$\frac{d \log \frac{2\alpha(\eta^*)}{\alpha} + \log(1/\beta)}{\eta_0 \varepsilon} + \frac{d \log \frac{2R}{\alpha} + \log(1/\beta)}{\varepsilon}.$$

注意到，由于 η_0满足 $\alpha = O(\eta_0 \sqrt{\log 1/\eta_0 + \sqrt{\eta_0 / dn}})$，我们必有 $\eta_0 = \tilde{\Omega}(\min\{\alpha, \alpha^2 n/d\})$。将其代入式（21）即可得到所述界。 □

5.3 隐私各向同性经验均值估计的计算速率

这里我们将证明定理 5.3 在 Λ 为各向同性时的计算界。我们将首先提出一种用于稳健经验均值估计的高效平方和算法，然后应用隐私到鲁棒性的归约。

5.3.1 高效稳健经验均值估计

定理 5.6。 设 $\eta > 0$ 小于一个充分小的常数。存在一个高效的 η-稳健估计量，使得对于所有 $\mu \in \mathbb{R}^d$，给定来自 $\mathcal{N}(\mu, \mathrm{Id}_d)$ 的 n 个独立同分布样本，且未污染样本的经验均值为 $\bar{x}$ 时，该估计量将输出一个 $\hat{x}$，使得 $\|\hat{x} - \bar{x}\|_2 \leqslant \alpha$，除非发生概率至多为 β的事件

$$\alpha \leqslant O\left(\eta \sqrt{\log 1/\eta} + \sqrt{\eta \sqrt{\frac{d + \log \frac{1}{\beta}}{n}}}\right).$$

证明。 证明源于对 [KMZ22] 中算法的仔细分析。令 $\alpha_0, \alpha_1, \alpha_2$ 为稍后确定的值。令 $x_1, \ldots, x_n$ 表示观测到的样本，其中有 η 比例已被污染。令 $x^*_1, \ldots, x^*_n$ 表示原始的、未污染的样本；我们无法访问这些样本，其经验均值为 $\bar{x}$。我们考虑以下关于不定元 $w_1, \ldots, w_n$ 的多项式约束（它们表示我们对哪些点未受污染的猜测），以及 $x'_1, \ldots, x'_n$：

$$\mathcal{A} = \left\{ \begin{array}{ll} \forall i \in [n]: & w_i = w_i^2 \\ \forall i \in [n]: & w_i x_i = w_i x'_i \\ & \sum_{i=1}^n w_i \geqslant (1-\eta) n \\ \forall v \in \mathbb{S}^{d-1}: & \mathbb{E}_i \langle x'_i - \mu', v \rangle^2 \leqslant (1 + \alpha_0) \end{array} \right\}$$

其中 $\mu' = \mathbb{E}_i x'_i$。那么算法如下：

算法 5.7 (稳健经验均值估计算法)。输入： η-受污染样本 $x_1, \ldots x_n \in \mathbb{R}^d$输出： 经验均值的一个多项式时间估计值 $\hat{x} \in \mathbb{R}^d$

1. 求解约束系统 $\mathcal{A}(x', w; x)$ 的次数为 $O(1)$ 的 *SoS* 松弛，并得到一个伪-$\tilde{\mathbb{E}}$期望算子 $\mathbb{E}$。
2. 令 $\hat{x} \leftarrow \tilde{\mathbb{E}}[\mu']$。
3. 返回$\hat{x}$.

假设我们知道，对于所有满足 $\mathbb{E}_i b_i \geqslant 1-\eta$ 的向量$v \in \mathbb{S}^{d-1}$ 和$b \in [0,1]^n$，界限

$$\left|\mathop{\mathbb{E}}_i b_i \langle x_i^\circ - \bar{x}, v\rangle\right| \leqslant \alpha_1$$
$$\left|\mathop{\mathbb{E}}_i b_i \langle x_i^\circ - \bar{x}, v\rangle^2 - 1\right| \leqslant \alpha_2$$

保持不变。接着，追踪 [KMZ22 引理 4 的证明中的误差项1] 可知，若 $\tilde{\mathbb{E}}$ 是满足 $O(1)$ 次数的伪期望，且 $\mathscr{A}$，那么

$$\|\tilde{\mathbb{E}}\mu' - \bar{x}\|_2 \leqslant O(\alpha_1 + \sqrt{\eta\cdot(\alpha_0+\alpha_2)+\eta^2}).$$

我们在 A.4 节 （见 引理 A.1）中给出该结论的证明。现在只需令 $\{\alpha_0,\ \alpha_1,\ \alpha_2\}$。应用 引理 5.8，由于 $n \geqslant d\ +\log(1/\beta)$，我们可以取

$$\alpha_0 = \alpha_2 = O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \eta\log(1/\eta)\Big),$$
$$\text{and } \alpha_1 = O\Big(\sqrt{\frac{\eta(d+\log(1/\beta))}{n}} + \eta\sqrt{\log(1/\eta)}\Big).$$

代入这些对 α_0、α_1、α_2 的取值，即完成证明。对于可行性，注意伪期望推广了期望，而期望又推广了单点上的质量。为了验证可行性，只需令 $\tilde{\mathbb{E}}$ 为这样一个伪期望：它将每个 x'_i 映射为 x°_i，并将每个 w_i 设为 1（若该样本未被污染），否则设为 0。

□

引理 5.8 （稳定性）。 设 $x_1,\ldots,x_n$ 是来自 $\mathcal{N}(\mu, \mathrm{Id}_d)$ 的独立同分布样本， $\bar{x}$ 是它们的经验均值。那么，以 $1-\beta$的概率，对于所有满足 $\mathbb{E}_i b_i \geqslant 1-\eta$的向量 $v \in \mathbb{S}^{d-1}$ 和 $b \in [0,1]^n$，都有

$$\left|\mathop{\mathbb{E}}_i b_i \langle x_i - \bar{x}, v\rangle\right| \leqslant O\Big(\sqrt{\frac{\eta(d+\log(1/\beta))}{n}} + \eta\sqrt{\log(1/\eta)}\Big)$$
$$\left|\mathop{\mathbb{E}}_i b_i \langle x_i - \bar{x}, v\rangle^2 - 1\right| \leqslant O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \frac{d+\log(1/\beta)}{n} + \eta\log(1/\eta)\Big)$$

我们在 A.3 节中证明该界。

5.3.2 从鲁棒性到隐私的归约

本节中，我们应用从鲁棒性到隐私的归约，并证明定理 5.3的高效各向同性情形。我们的目标是应用 定理 4.19。我们紧密遵循 [HKMN23]第 5 节中的归约。我们在定理5.6 的证明中使用的程序，与 [HKMN23],所用的程序相同，唯一的区别在于我们希望从该程序获得的准确性保证。他们关注的是估计分布的参数均值，而我们关注的是估计未污染样本的经验均值。尽管存在这一区别，我们仍可通过使用相同的评分函数来实例化同一归约。定理 4.19 所需的大多数性质——有界敏感度、拟凸性、高效可计算性、高效可计算性以及高效寻找低评分点——都会自动成立。我们只需验证准确性保证和体积比。

现在我们定义评分函数。首先，需要定义候选点作为可认证均值的含义：

定义 5.9（(α,τ,T)-可认证均值）。设 x 为一组输入。若存在一个线性泛函 $\mathscr{L}$，其输入为关于不定元 $(x'、w、M)$ 的次数为 $O(1)$ 的多项式，并满足以下性质，则称候选点 $\tilde{x}$ 是这组输入的 (α,τ,T)-可认证均值：$\mathscr{L}$ 满足 $\|\mathscr{L}\mu'-\tilde{x}\|_2\leqslant\alpha$ 和 $\|\mathscr{L}\mu'\|_2\leqslant 2R+\tau T$，其中 $\mu'=\mathbb{E}_i x'i$、$\|\mathscr{R}(\mathscr{L})\|_{\mathrm{F}}\leqslant R'+n\tau$，对于 $R'=poly(n、d、R)$ 足够大时成立，其中 $\mathscr{R}$ 将 $\mathscr{L}$ 映射为其矩阵表示。此外，$\mathscr{L}$ 必须近似满足（如定义 4.17 所述）以下多项式不等式系统，次数至多为 6。

$$\left\{\begin{array}{ll}\forall i\in[n]: & w_i=w_i^2\\ \forall i\in[n]: & w_ix_i=w_ix_i'\\ & \sum_{i=1}^n w_i\geqslant n-T\\ & \mathbb{E}_i(x_i'-\mu')(x_i'-\mu')^{\mathsf{T}}+MM^{\mathsf{T}}=(1+\alpha_0)I\end{array}\right\}$$

我们将最后一个等式视为 k^2 个相等式。此外，我们选择 η 使得 $\alpha=\alpha(\eta)=\Theta(\eta\sqrt{\log 1/\eta}+\sqrt{\eta}\sqrt{\frac{d+\log(1/\beta)}{n}}$（如定理 5.6中所示），并按照定理 5.6的证明来选择 $\alpha_0=\alpha_0(\eta)$，使其取决于 η，因而也取决于 α, $\alpha_0(\eta)=\Theta(\eta\log(1/\eta)+\sqrt{\frac{d+\log(1/\beta)}{n}})$。

根据上述定义，评分函数定义如下：

定义 5.10 (评分函数)。设 $\alpha,\tau>0$。设 $\mathbb{B}_2^d(2R+n\tau+\alpha)$表示 d 维空间中半径为$2R+n\tau+\alpha$的 ℓ_2球。给定一个输入集合 x，且 $\tilde{x}\in\mathbb{R}^d$，我们定义在 $\mathbb{B}_2^d(2R+n\tau+\alpha)$ 上的评分函数 $\mathscr{S}$为

$$\mathscr{S}(\tilde{x};x,\alpha,\tau)=\min\{T\geqslant 0:\tilde{x}\ \textit{is an}\ (\alpha,\tau,T)\textit{-certifiable mean with respect to}\ x\}.$$

接下来我们界定体积比。

引理 5.11 (高分与低分点的体积)。假设$R=\Omega(\alpha)$ 足够大。设 $\eta\geqslant 0$ 至多为一个足够小的常数 η^*。假设输入点集合 x 满足引理 5.8的性质。那么：

1. 在以 $\tilde{x}$ 为中心、半径为 $\alpha(\eta)$ 的球内，每个点的分数至多为 ηn
2. 每个分数至多为 $\eta^* n$ 的点都包含在半径为 $O(\alpha(\eta^*))$ 的球内。

证明。注意，由于样本集合满足引理 5.8的性质，原始程序是可行的：考虑由点质量给出的解，它将每个x'_i 赋值为未受污染的值。因此，只需令 $\mathscr{L}$ 为对应于该赋值的伪期望。于是，$\mathscr{L}$上的接近性约束，即 $\|\mathscr{L}\tilde{\mu}-\mu\|_2\leqslant\alpha$，将得到满足；并且由于$R\geqslant\alpha$，约束$\|\mathscr{L}\mu\|_2\leqslant 2R+\tau T$也得到满足。因此，半径为 $\alpha(\eta)$ 的球内每个点的分数都必须至多为 ηn。

对于第二部分，注意，由于 $\mathscr{L}$ 满足约束系统，根据鲁棒算法的保证，我们必有 $\|\mathscr{L}\mu'-\tilde{x}\|_2\leqslant O(\alpha(\eta^*))$。此外，根据 $\mathscr{L}$ 上的接近性约束，我们知道$\|\mathscr{L}\mu'-\tilde{\mu}\|_2\leqslant\alpha(\eta)\leqslant\alpha(\eta^*)$。由三角不等式即可得到所需结果。□

如前所述，定理 4.19的其余条件均已满足，因为得分函数与 [HKMN23]中的相同。现在，我们可以证明在先验分布为各向同性的情形下，定理 5.3的计算高效版本。

定理 *5.3*第二部分的证明。设 η 满足 $\alpha = \alpha(\eta)$，其中 $\alpha(\eta)$ 如定理 5.6所述。根据定理 4.19可知，只需正态分布 满足

$$n \geqslant \Omega\Big(\max_{\eta' : \eta \leqslant \eta' \leqslant 1} \frac{\log V'_\eta / V_\eta + \log(1/\beta\eta)}{\varepsilon \eta'}\Big).$$

我们将上面的内容分为两部分，$[\eta, \eta^*] \cup [\eta^*, 1]$。现在应用引理 5.11。由第一部分，只需满足

$$n \geqslant \Omega\Big(\frac{d \log(\alpha(\eta^*)/\alpha(\eta)) + \log(1/\beta\eta)}{\varepsilon\eta}\Big).$$

由第二部分，只需满足

$$n \geqslant \Omega\Big(\frac{d \log(R/\alpha(\eta^*)) + \log(1/\beta\eta)}{\varepsilon}\Big).$$

代入 η 和 $\alpha(\eta^*)$ 关于 α、d、n、β的定义，只需取满足以下条件的n：

$$n \geqslant \tilde{\Omega}\Big(\frac{d + \log(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{d + \log(1/\beta)}{\alpha\varepsilon} + \frac{d \log(R/\alpha) + \log(1/\beta)}{\varepsilon}\Big),$$

其中对数因子被隐藏在 α中。 □

5.4 贝叶斯均值估计的信息—计算差距

我们可能希望存在一种计算高效的平方和算法，甚至能够达到定理 5.5 中稳健经验均值估计的统计最优速率，从而可以高效地应用 [HKMN23] 中的归约，在私有经验均值估计中实现统计最优速率。本节将证明这是不可能的。我们首先使用低次框架证明稳健问题的计算下界，然后通过 [GH22]提出的隐私到鲁棒性归约，证明私有问题的计算下界。

5.4.1 稳健经验均值估计的困难性

为简单起见，假设 ηn 是整数。为证明稳健问题中的信息—计算差距，我们将展示从以下问题出发的一种计算高效的归约：

定义 5.12 ((η, δ)-高斯混合分布区分问题)。 给定 η， $\delta > 0$和 n， $d \in \mathbb{N}$，考虑以下两个分布：

- 零假设分布： $\mathcal{N}(0, \mathrm{Id}_d)$
- 植入分布：从单位球面上均匀分布地采样 v。于是，植入分布为$(1-\eta)\mathcal{N}(-\eta\delta v, \mathrm{Id}_d) + \eta\mathcal{N}((1-\eta)\delta v, \mathrm{Id}_d)$。

给定 n 个样本，问题是以常数优势区分这两个分布。

这与 [DHPT25], 中考虑的一个判别问题密切相关；并且在本节末尾的引理 5.16 中，我们将验证，在低次框架下，该问题在计算上是困难的，适用于 $\delta = \Theta(\alpha/\eta)$。

我们现在继续 将该问题归约为经验均值估计问题，以获得困难性结果 .

定理 5.13 （经验均值估计的最坏情况困难性）。 设 $\eta \in [0, 1/3)$。考虑任意一个多项式时间估计量，该估计量使用来自 $\mathcal{N}(\mu, \mathrm{Id}_d)$ 的 $n \leqslant o(d\eta^2/\alpha^4)$ 个样本，以执行误差为 α 的 η-稳健经验均值估计。如果对于所有 $\|\mu\| \leqslant 21\alpha$，任意此类估计量以最坏情况概率 2 3 成功，那么就存在一个多项式时间算法，可以解决 $(\eta, \Theta(\alpha/\eta))$-高斯混合分布区分问题。

同样的归约将给出以下平均情形下界；由于私有贝叶斯均值估计本质上是关于从先验分布中抽取 μ 的平均情形问题，这一结果对于证明其困难性至关重要：

引理 5.14 （稳健经验均值估计的平均情况困难性）。 令 $\eta \in [0, \frac{1}{3})$ 和 $\mu \sim 21\alpha$ 服从 $Bern(1/2)$，其中 $v \sim \mathbb{S}^{d-1}$ 均匀分布。考虑任意一个多项式时间估计器，它使用来自 $\mathcal{N}(\mu, \mathrm{Id}_d)$ 的 $n = o\left(d\frac{\eta^2}{\alpha^4}\right)$ 个样本，以执行 η-稳健经验均值估计，使误差达到 α。如果任意此类估计器在如上所述对 μ 的抽样下，以平均情形概率 5 6 成功，那么就存在一个用于 $(\eta, \Theta(\alpha/\eta))$-高斯混合分布区分问题的多项式时间算法。

这两个结果都可由以下归约直接得到：

引理 5.15 (估计蕴含可区分性)。 假设存在一个多项式时间的 η-鲁棒经验均值估计器，它接收来自 $\mathcal{N}(\mu, \mathrm{Id})$ 的 n 个样本，其中 μ 生成方式为以下两者之一

- $\mu = 0$
- $\mu = 21\alpha v$，其中 $v \sim \mathbb{S}d-1$ 均匀分布。

并且在两种情况下都以 ℓ_2 准确度 α 和概率 $1 - \beta > 2/3$输出经验均值的一个估计值。那么，根据 定义 5.12 ，存在一个多项式时间算法，使用 n 个样本以常数优势解决 (η, δ)-高斯混合可区分性问题，且 $\delta = 21\alpha/\eta$。

证明。 我们的算法如下：在这 n 个样本上运行鲁棒经验均值估计器，得到估计值 $\hat{}\mu$，并令 $\tilde{}x$ 为输入样本的均值。如果 $\|\hat{}\mu - \tilde{}x\| \leqslant \alpha$，我们就猜测为“零假设”，否则猜测为“植入分布”。其思想是，如果我们从零假设分布中抽样，显然

$$\|\hat{\mu} - \tilde{x}\| \leqslant \alpha$$

以概率 $1 - \beta > 2/3$ （根据 $\eta = 0$ 情形下的准确性保证）。然而，如果我们从植入分布中抽样，就可以将“植入”过程视为对从 $\mathcal{N}(-\eta\delta v, \mathrm{Id}_d)$ 中抽取的干净样本进行 η 腐化；用 $\bar{}x$ 表示这些“干净样本”的经验均值。此外，令 $X_i \sim \mathcal{N}(-\eta\delta v, \mathrm{Id}_d)$、$Y_i \sim \mathcal{N}((1-\eta)\delta v, \mathrm{Id}_d)$，以及$Z_i \sim \mathcal{N}(\delta v, 2\mathrm{Id}_d)$，我们看到 $\tilde{}x - \bar{}x$服从如下分布

$$\frac{1}{n}\sum_{i=1}^{\eta n}[Y_i - X_i] = \frac{1}{n}\sum_{i=1}^{\eta n} Z_i \sim \mathcal{N}(\eta\delta v, \frac{2\eta}{n}\mathrm{Id}_d)$$

不难发现

$$\sup_{\sigma>0} \mathbb{P}_{X\sim\mathcal{N}(v,\sigma^2\mathrm{Id}_d)}\left[\|X\|_2 \leqslant \frac{1}{10}\right] \leqslant \frac{1}{9^d\sqrt{\pi d}} < \frac{1}{10}$$

（通过对球上高斯分布的概率密度函数和球的体积取上界，代入一阶最优性条件得到的 σ，再使用斯特林近似），因此我们尤其可以通过重新缩放得到

$$\mathbb{P}\left[\|\tilde{x} - \bar{x}\|_2 \leqslant \frac{1}{10}\eta\delta\right] < \frac{1}{10}$$

因此，在第二种情况下，上述结果与算法的准确性保证之间的反三角不等式足以推出

$$\begin{aligned}\|\hat{\mu}-\tilde{x}\| &\geqslant \|\bar{x}-\tilde{x}\|-\|\hat{\mu}-\bar{x}\| \\ &\geqslant \frac{1}{10}\eta\delta-\alpha \\ &> \alpha\end{aligned}$$

除了概率为 $\frac{1}{3}+\frac{1}{10}<\frac{1}{2}$ 的情况外，根据并集界，这使我们的区分器具有常数优势。 □

请注意，任何优于定理 5.13 中所述下界的经验均值估计器，都显然满足引理 5.15中所用估计量的条件。此外，任何优于引理 5.14 中平均情形下界的估计量，都必须在两种情况下分别以概率 2/3 成功，因此也可以通过引理 5.15归约为检验问题。

仍需验证 (η,δ)-高斯分布混合分布区分问题的低次困难性：

引理 5.16. </style}给定 n 个样本时，O上的 -高斯分布混合可区分性问题的 D 次优势至多为 O，前提是 n且 -n d多项式(D)**。**

证明。 我们将应用 引理 4.6，令 $\kappa:=\eta^2$， $\tau:=\frac{1}{\delta^4}$，并令 $j_*:=2$ 。引理 4.6 中的第一个条件引理 4.6 由构造可知成立；对于第二个条件，我们需要证明对所有 $j\geqslant j_*$，

$$\left|\mathop{\mathbb{E}}_{X\sim A} h_j(X)\right|^2 \leqslant j^j\kappa\tau^{-j/j^*}$$

利用 定理 4.7，

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{X\sim A} h_j(X)\right| &= \left|(1-\eta)\mathop{\mathbb{E}}_{X\sim\mathcal{N}(-\eta\delta,1)}[h_j(X)]+\eta\mathop{\mathbb{E}}_{X\sim\mathcal{N}(\delta-\eta\delta,1)}[h_j(X)]\right| \\ &\leqslant \left|(1-\eta)\cdot(-\eta\delta)^j/\sqrt{j!}\right|+\left|\eta\cdot(\delta-\eta\delta)^j/\sqrt{j!}\right|\end{aligned}$$

两边平方后得到

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{X\sim A} h_j(X)\right|^2 &\leqslant 2\left|(1-\eta)\cdot(-\eta\delta)^j/\sqrt{j!}\right|^2+2\left|\eta\cdot(\delta-\eta\delta)^j/\sqrt{j!}\right|^2 \\ &\leqslant 2\left|(-\kappa^{1/2}\tau^{-1/4})^j/\sqrt{j!}\right|^2+2\kappa\left|\tau^{-j/4}/\sqrt{j!}\right|^2 \\ &\leqslant 2\kappa^j\tau^{-j/2}/j!+2\kappa\tau^{-j/2}/j! \\ &\leqslant 4\kappa\tau^{-j/j^*}/j! \\ &\leqslant j^j\kappa\tau^{-j/j^*}\end{aligned}$$

如所需。这给出了样本复杂度的一个次数为D 的下界

$$n\geqslant d^{j^*/2}\frac{\tau}{\kappa}\cdot\frac{1}{\mathrm{poly}(D)}=\Omega\left(\frac{d\eta^2}{\alpha^4}\right)\cdot\frac{1}{\mathrm{poly}(D)}$$

如所述。 □

5.4.2 私有经验均值估计的计算困难性

我们可以利用 [GH22] 的隐私到鲁棒性归约，结合 定理 5.13 来证明 定理 5.4 的计算下界：

定理 *5.4* 的证明。设 p、$q > 0$ 稍后确定，并假设我们能够使用 n_{priv} 个各向同性样本进行私有经验均值估计，其中

$$n_{\text{priv}} = o\left(\frac{\min(d, \log\frac{1}{\beta})}{\alpha^p \varepsilon^q}\right)$$

不失一般性，设 $\log 1/\beta = d$（否则我们可以降低维度或增大 β，这应当会使问题更容易）。根据 [GH22], 上文重述为定理 4.21中的隐私到鲁棒性归约，存在一个 η-鲁棒经验均值估计器，其样本复杂度为 $n_{\text{rob}} = o\left(\frac{d}{\alpha^p \varepsilon^q}\right)$，对应的污染比例为 $\eta = d\,\varepsilon n_{\text{rob}}$。

解出 ε 并重新整理变量，可得一个 η-鲁棒经验均值估计器，其样本复杂度为

$$n_{\text{rob}} = o\left(\frac{d}{\alpha^p} \cdot \left(\frac{\eta n_{\text{rob}}}{d}\right)^q\right) = o\left(\frac{d^{1-q}\eta^q}{\alpha^p} n_{\text{rob}}^q\right) = o\left(\frac{d\eta^{q/(1-q)}}{\alpha^{p/(1-q)}}\right)$$

对于$p = 4/3$ 和$q = 2/3$，这将与定理 5.13关于鲁棒经验均值估计的计算困难性结果相矛盾。

□

我们可以将相同的归约应用于引理 5.14 中所述的平均情形鲁棒下界，从而得到平均情形隐私下界：

推论 5.17。 令 $\beta < 1/6$，并假设 $\mu = 21\alpha v$ ·伯努利分布$(1/2)$，其中 $v \sim \mathbb{S}^{d-1}$。如果存在一个多项式时间 ε-*DP* 经验均值估计量，对于来自$\mathcal{N}(\mu, \mathrm{Id}_d)$的 n 个样本，能够以误差 α 和置信度 β 成功，并且其样本复杂度为

$$o\left(\frac{\min\left(d, \log\frac{1}{\beta}\right)}{\alpha^{4/3}\varepsilon^{2/3}}\right),$$

那么，对于$\eta = \Theta(\alpha^{4/3}/\varepsilon^{1/3})$ 和 $\delta = \Theta(\varepsilon^{1/3}/\alpha^{1/3})$，存在一个使用 $o(d_{\alpha^{4/3}\varepsilon^{2/3}})$ 个样本的多项式时间算法，用于解决 (η, δ)-高斯混合分布区分问题；~~该问题~~是低次困难的。

最后，我们完成定理 5.2的证明。该定理为原始的隐私贝叶斯均值估计问题建立了信息—计算差距。证明需要调整推论 5.17 中的平均情形下界，使得关于 μ 的分布与定理 5.2所设定的高斯先验相匹配。

定理 定理 *5.2*的证明。 我们构造 (X, Y)的一个耦合，使其具有以下性质：

- X的边缘分布为 $\mathcal{D}_1 := \mathcal{N}(0, \sigma^2 I)$。
- Y 的边缘分布 $\mathcal{D}_2$ 与先验分布的全变差距离至多为 $1/6$。
- (X, Y)的联合分布如下：首先，按照 $\mathcal{D}_1$ 抽取X。然后，将Y设为$X + 21\alpha v$，其中$v \sim \mathbb{S}^{d-1}$ 均匀且相互独立地取值，并适当选择 α。

根据高斯分布的总变差距离与参数距离之间的标准界，在此构造下，存在一个绝对常数$c > 0$，使得只要满足 $\alpha \leqslant c\sigma$，第二个性质便成立。

第三个性质结合推论 5.17表明，不存在使用 ε个样本的隐私估计器，能够估计经验均值

$$o\left(\frac{\min(d, \log \frac{1}{\beta})}{\alpha^{4/3}\varepsilon^{2/3}}\right) \tag{22}$$

达到 α精度的概率不会超过 5/6，其中 μ 取自 $\frac{1}{2}(\mathscr{D}_1 + \mathscr{D}_2)$。事实上，反设某个高效估计量的成功概率为 5/6。那么，根据平均原理，必定存在某个x，使得在条件$X = x$下，它在 5/6 上的成功概率为 $\mu = x + 21\alpha\nu$·伯努利分布(1/2)。推论 5.17中的保证显然具有平移不变性，因此，这个估计量的存在意味着存在一种高效算法，可以解决 (η, δ)-高斯混合分布区分问题，这与低次困难性矛盾。

与 $\frac{1}{2}(\mathscr{D}_1 + \mathscr{D}_2)$ $\mathscr{D}_1$ 之间的全变差距离至多为 1/12，因此根据三角不等式，样本复杂度由式（22）给出的隐私估计器，不可能以超过 α的概率达到11 12 精度，其中 μ 取自先验分布。因此我们要求 $\beta < \frac{1}{12}$，而第 5.1 节中阐明的贝叶斯均值估计与经验均值估计之间的联系，给出了前者所需的下界。

□

5.5 持续观测下私有高斯分布均值估计的应用

作为我们针对高斯分布先验的私有贝叶斯均值估计算法的一项应用，我们考虑持续观测下的在线均值估计模型：

定义 5.18 (私有在线均值估计)。 设 $\mu \in \mathbb{R}^d$ 是一个满足 $|\mu| \leqslant R$ 的未知参数。假设我们观察到由 k 个批次组成的数据流，每个批次包含 n 个样本，且这些样本是从 $\mathcal{N}(\mu, \mathrm{Id})$ 独立抽取的。每个批次到来后，我们都必须根据迄今为止观察到的数据报告 μ 的一个估计值，并且整个机制必须满足关于数据流中所有 nk 个样本的差分隐私。

如果流式算法能够安全地将非隐私状态对公众隐藏起来，解决方案就很简单：令

$$\varepsilon_i = \frac{\varepsilon}{i \log k}$$

并在每次迭代后发布迄今为止所观察到样本的一个 ε_i-私有经验均值估计值。通过简单计算可知，如果 $n \geqslant d\log_{R/\varepsilon} \alpha$，那么以 $1-\beta$ 的概率，对于所有 $i \in [k]$

$$\|\hat{\mu}_i - \mu\|_2 \leqslant \tilde{O}\left(\sqrt{\frac{d + \log \frac{k}{\beta}}{n}} + \frac{d + \log \frac{k}{\beta}}{\varepsilon n} \cdot \log k\right)$$

然而，更强的全景隐私概念通常在流式设定中进行研究。在这种模型中，对手除了能看到所有输出外，还可以在其选择的m 个时刻（处理连续批次之间）查看算法保留的隐藏状态。常见的研究问题是在批次大小为 1，且x_i 从某个有界域中任意选取时估计经验均值；在这种设定下，可以证明，抵御m 次入侵的全景隐私要求准确度损失随 $\sqrt{m}$ [DNPR10]而增大。相比之下，我们将证明，在随机设定下——当x_i 独立同分布地取自高斯分布，而非任意选取，且批次大小足够大时——无论入侵次数有多少，都能保证全景隐私，因为批次之间完全不保留隐藏状态。

尽管由于 μ没有先验分布，这项任务属于频率学派范畴，但我们将采用以下自然策略，使用私有贝叶斯均值估计算法。我们将从 $\mathbb{R}^d$ 上的均匀非正常先验开始，因此，每个批次结束后，在给定迄今为止看到的所有样本的条件下，后验均值就是最大似然估计（MLE）。每个新批次开始时，我们都有一个基于截至上一批次（包括上一批次）的全部数据得到的后验分布估计，并将此后验分布作为先验分布，利用新批次中的样本进行私有贝叶斯更新。将第5节中的隐私-效用保证组合应用于各批次，可得以下私有在线均值估计保证：

定理5.19。 设 $\mu \in \mathbb{R}^d$ 为算法未知的均值参数，满足 $\|\mu\| \leqslant R$。给定由 k 个批次组成的数据流，每个批次包含 n 个独立同分布样本 $x_i \sim \mathcal{N}(\mu, \mathrm{Id}_d)$，且 $n > d \log^{\alpha} \frac{R}{\varepsilon}$，则存在一种计算上低效（分别地，计算高效）的 ε-差分隐私机制，该机制在每个批次后输出一个 μ 的估计值，使得以概率 $1-\beta$，对所有 $t \in [k]$，，第 t 个估计值的误差至多为

$$\tilde{O}\left(\sqrt{\frac{d+\log\frac{k}{\beta}}{nt}} + \frac{d+\log\frac{k}{\beta}}{\varepsilon n}\right), \quad \text{resp.} \quad \tilde{O}\left(\sqrt{\frac{d+\log\frac{k}{\beta}}{nt}} + \left(\frac{d+\log\frac{k}{\beta}}{\varepsilon^{2/3} n}\right)^{3/4} + \frac{d+\log\frac{k}{\beta}}{\varepsilon n}\right).$$

证明。 为了获得频率学派保证，我们从均匀非正常先验开始（因此，每次迭代后，我们要估计的真实后验均值就是最大似然估计量）。每个批次之后，我们将先前的后验分布用作先验分布，并利用该批次的数据计算后验均值估计值。一如既往，假设 $n \geqslant \tilde{\Omega}\left(d \log^{\alpha} \frac{R}{\varepsilon}\right)$。我们注意到，第 t 步的后验方差与样本无关，其定义为

$$\sigma_{t+1}^2 = \left(n + \frac{1}{\sigma_t^2}\right)^{-1}$$

其（满足 $\sigma_0^{-2} = 0$）的解为

$$\sigma_t^{-2} = nt$$

后验均值更新可写为

$$\begin{aligned}\mu_{t+1} &= \mu_t + \frac{1}{1+\frac{1}{\sigma_t^2 n}}(\bar{x}_t - \mu_t) \\ &= \frac{1}{1+\sigma_t^2 n}\mu_t + \frac{1}{1+\frac{1}{\sigma_t^2 n}}\bar{x}_t\end{aligned}$$

设 E_t 为时刻 t 后验均值（或在此情形下等价的最大似然）估计值的误差上界，并假设我们可以在 ε-隐私和 β-置信度下，以误差 C 估计 n 个样本在 d 维度中的经验均值。那么我们注意到

$$\begin{aligned}E_{t+1} &\leqslant \frac{1}{1+\sigma_t^2 n}E_t + \frac{1}{1+\frac{1}{\sigma_t^2 n}} \cdot C \\ &= \frac{t}{1+t}E_t + \frac{1}{1+t} \cdot C\end{aligned}$$

这意味着 $E_t \leqslant C$ 对所有 $t \geqslant 1$。因此，利用定理 5.3，我们可以低效地（分别在计算上高效地）估计最大似然估计的误差

$$\alpha \leqslant \tilde{O}\left(\frac{d+\log\frac{k}{\beta}}{\varepsilon n}\right), \quad \text{resp.} \quad \alpha \leqslant \tilde{O}\left(\frac{d+\log\frac{k}{\beta}}{\varepsilon n} + \left(\frac{d+\log\frac{k}{\beta}}{\varepsilon^{2/3} n}\right)^{3/4}\right).$$

根据最大似然估计收敛到高斯模型中的参数均值，即可得到该定理。

□

6 稳健且保护隐私的贝叶斯线性回归

在本节中，　　我们将证明关于贝叶斯线性回归的主要结果。模型如下　　s:

定义 6.1（高斯先验下的贝叶斯回归）。 参数 $w^* \in \mathbb{R}^d$ 从先验分布$\mathcal{N}(0,\Sigma)$中抽取，然后按如下方式生成数据。设计矩阵 $X^\circ \in \mathbb{R}^{d\times n}$ 的各列独立同分布地采样自分布 $\mathcal{N}(0,\mathrm{Id}_d)$，响应向量 $y^\circ \sim \mathcal{N}$(为 $X^\circ{}^\intercal w^*, \mathrm{Id}_n$)。每个样本表示为$(X^\circ(\cdot,i), y^\circ(i))$，其中 $X^\circ(\cdot,i)$ 表示设计矩阵 X°的第 i 列。给定这些样本后，我们的目标是以 ℓ_2 范数私密地估计后验均值 $w^{\text{post}} := \mathbb{E}[w^* \mid X^\circ, y^\circ]$。

我们还将考虑这样一种情形：有一部分样本（比例为 η）被任意污染。在这种情况下，隐私是针对受污染样本定义的，但我们希望估计的后验均值是针对未污染样本而言的，尽管算法只能访问受污染的样本。

对于这个问题，我们的主要正面结果是针对 $\Sigma = \sigma^2 \mathrm{Id}_d$ 这一特殊情形的算法。具体来说，我们将证明以下对 定理 1.4 的略微加强版。

定理 *6.2*（隐私保护的鲁棒贝叶斯线性回归的统计速率）。设 $\Sigma = \sigma^2 Id_d$，其中已知参数为 $\sigma > 0$。设 $\beta \in [0,1]$, $\varepsilon \leqslant O(1)$。假设 $n \geqslant max\{d\, log(\sigma\sqrt{d})/\varepsilon, (d + log(1/\beta))/\eta\}$。那么，存在一种计算上低效的算法，对于按定义 *6.1* 生成的 η-受污染样本 $(X^{\text{input}}, y_{\text{input}})$，输出一个 ε-*DP* 估计量 $\hat{w}$，使得其概率至少为 $1-\beta$,

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|_2 \leqslant \tilde{O}(\eta + \frac{d + \log(1/\beta)}{n\varepsilon}),$$

其中，$(X^\circ、y^\circ)$ 是未污染样本。

定理 *6.3*（隐私保护的鲁棒贝叶斯线性回归的计算速率）。设 $\Sigma = \sigma^2 Id_d$，其中参数 $\sigma > 0$已知。设 $\beta \in [0,1]$, $\varepsilon \leqslant O(1)$。假设 $n \geqslant max\{d\, log(\sigma\sqrt{d})/\varepsilon, (d + log(1/\beta))/\eta\}$。存在一个多项式时间算法，对于按照定义 *6.1* 生成的 η污染样本 (X, y)，输出一个 ε-*DP* 估计量 $\hat{w}$，使得其以至少 $1-\beta$的概率满足：

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|_2 \leqslant \tilde{O}\left(\sqrt{\eta\sqrt{\frac{d+\log(1/\beta)}{n}}} + \eta + \frac{\left(d+\log(1/\beta)\right)^{3/4}}{n^{3/4}\varepsilon^{1/2}} + \frac{d+\log(1/\beta)}{n\varepsilon}\right),$$

其中，$(X^\circ、y^\circ)$ 是未污染样本。

我们注意到，贝叶斯最优估计量的均方误差 $\mathrm{MSE}(\hat{w})^{1/2}$（定义见式（8））为$(1 + {}^{1}{}_{n\sigma^2})^{-1}\sqrt{}d/n$</style}的量级。由此可得推论：在满足条件 $d/n \ll \eta + d/n\varepsilon$ 且$n \geqslant \Omega(d\log(\sigma\sqrt{d})/\varepsilon)$时，可以渐近地达到贝叶斯最优误差率。

σ有三种不同的情形：

- 过强先验：当 $\sigma \ll 1/\sqrt{n}$ 时，在这种情形下进行隐私保护的鲁棒估计很简单。先验分布足够强，因此可以忽略样本，仅通过输出先验分布的均值，便能渐近地达到贝叶斯最优均方误差。

- 过弱先验：当 $\sigma \gg 1/\sqrt{n}$ 时，私有且鲁棒的贝叶斯估计会退化为私有且鲁棒的频率学派估计。先验分布足够弱，因此普通最小二乘渐近地即可达到贝叶斯最优均方误差，不过，要证明在隐私和鲁棒性约束下也能最优地做到这一点，仍需付出一些努力。

- 临界先验：最有趣的情形是 $\sigma = \Theta(1/\sqrt{n})$，此时需要结合先验分布和样本中的信息，才能与贝叶斯估计量竞争。

本节将重点讨论计算高效算法，并将 定理 6.2 中关于统计速率的证明推迟到 C.2.1 节。在 6.1 节中，我们考虑临界先验情形；在 第 6.2 节中，我们考虑弱先验情形。

此外，我们提供了低次证据，表明计算速率是最优的。

定理 6.4 （隐私贝叶斯线性回归的信息—计算差距）。 假设 $\varepsilon \leqslant O(1)$且 $\alpha \leqslant \min\{\eta d^{1/4}/D,\ \eta^2\}$。存在一个判别问题 Π （见 定义 *6.22*），该问题是 D 次困难的；如果存在一个多项式时间 ε-*DP* 算法，用于在标准高斯设计和 $\mathcal{N}(0, \sigma^2 \mathrm{Id})$ 先验下进行贝叶斯线性回归，并以 α 的误差、以 $1-\beta$的概率

$$n = o\left(\frac{d^{1/3}\log^{2/3}(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}\cdot D^{1/3}}\right)$$

使用样本，那么就会存在一个针对 Π 的高效判别算法。

注意，当失败概率 β 在 d 上呈指数级小的时候，这与定理 6.3达到的速率一致（相差对数因子）。事实上，即使对于一般的 β，我们也得到了一个真正的信息—计算差距（定理 6.30）。

上述下界的证明同样借助了非高斯成分分析，与贝叶斯均值估计中的做法类似；不过前者涉及更多，因为将非高斯成分分析嵌入贝叶斯线性回归中更为微妙。相关细节见6.3 节和6.4 节。

6.1 Com隐私贝叶斯线性回归在临界先验下的计算速率 r

我们在临界先验情形下的主要保证如下：

定理 6.5 (临界先验下的计算速率)。设 $\Sigma = \sigma^2 \mathrm{Id}_d$，已知参数 $\sigma = \Theta(1/\sqrt{n})$。存在一个多项式时间算法，给定 η-受污染样本 (X, y)，其生成方式如定义 *6.1*所述。该算法输出一个 ε-差分隐私估计量 $\hat{w}$，使得以至少 $1-\beta$的概率，

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|_2 \leqslant O(\alpha),$$

每当 $\alpha \geqslant \eta$对数$(1/\eta)$, $\varepsilon \leqslant O(1)$ 和

$$n \geqslant \tilde{\Omega}\Big(\frac{d+\log(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{d+\log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(\sigma\sqrt{d})}{\varepsilon} + \frac{(d+\log(1/\beta))\eta^2}{\alpha^4}\Big).$$

我们首先给出用于估计后验均值的稳健非隐私算法，然后应用隐私到鲁棒性归约。

6.1.1 使用临界先验分布的高效稳健回归

在此，我们给出一种用于估计后验均值的稳健非隐私算法，其中 $\sigma = \Theta(\sqrt{1/n})$。

我们将设计一个平方和估计量，用于 $(\Sigma^{-1} + n\mathrm{Id}_d)^{-1} X \circ y^\circ$，然后通过集中性证明它接近 w_{out}。正如第 2 节所讨论的，与我们的贝叶斯均值估计算法相比，关键区别在于要对估计值 y 强制要求其存在一个针对真实响应变量 y°的短平分解。

令X_{input}、y_{input}表示观测数据，其中有 η 的数据遭到破坏。令X°、y°表示原始的未污染数据，这是我们无法获取的。我们考虑以下关于不定元 $\xi=(\xi(1),\ldots,\xi(n))$（表示我们对哪些点遭到破坏的猜测）以及$X$、$y$的多项式约束。首先，我们有如下约束，表示 $\xi(i)$ 试图选出遭到破坏的点：

$$\mathscr{A}_{\text{corr}}(X,y,\xi;X_{\text{input}},y_{\text{input}})=\left\{\begin{array}{l}\xi\odot\xi=\xi,\quad \|\xi\|_1\leqslant\eta n,\\ (1-\xi(i))X(i,\cdot)=(1-\xi(i))X_{\text{input}}(i,\cdot),\ \forall i\in[n],\\ (1-\xi(i))\,y(i)=(1-\xi(i))\,y_{\text{input}}(i),\ \forall i\in[n]\end{array}\right\}. \tag{23}$$

接下来，我们给出韧性约束，以确保对未被篡改的协变量的估计值 X 具有有界协方差，并且对未被篡改的响应变量的估计值 y 可以分解为一个不定元向量 z_1 （其 ℓ_2 范数较小）和一个向量 z_2 （其 ℓ_∞范数较小）的短平分解：

$$\mathscr{A}_{\text{res}}(X,y)=\left\{\begin{array}{l}\left\|\frac{1}{n}XX^{\intercal}-\mathrm{Id}\right\|_{\text{op}}\leqslant\chi\sqrt{\frac{d+\log(1/\beta)}{n}},\\ y=z_1+z_2\\ \|z_1\|_2^2\leqslant\chi\eta n\log(1/\eta)+\chi\log(1/\beta),\\ \eta n\|z_2\|_\infty^2\leqslant\chi\eta n\log(1/\eta)+\chi\log(1/\beta)\end{array}\right\}. \tag{24}$$

其中，$\chi>0$ 是某个固定的普适常数。

最后，我们给出一个约束，用于定义一个与后验均值估计值相对应的变量：

$$\mathscr{A}_{\text{post}}(X,y,w_{\text{out}})=\left\{w_{\text{out}}=(1/\sigma^2+n)^{-1}Xy\right\}. \tag{25}$$

在满足这些约束的基础上，我们在下方给出算法 算法 6.6 。

算法 6.6 （用于近似后验均值估计量的鲁棒算法 ）.
输入： η-篡改的设计矩阵 $X_{\text{input}}\in\mathbb{R}^{n\times m}$ 和响应变量 $y_{\text{input}}\in\mathbb{R}^n$输出：
一个用于后验均值的多项式时间估计值 $\hat{w}\in\mathbb{R}^d$

1. 求解约束的次数为 $O(1)$ 的 *SoS* 松弛

$$\mathscr{A}(X,\ y\xi,w;X_{\text{input}},y_{\text{input}})=\mathscr{A}_{\text{corr}}(X,y,\xi;X_{\text{input}},y_{\text{input}})\cup\mathscr{A}_{\text{res}}(X,y,w)\cup\mathscr{A}_{\text{post}}(X,y,w_{\text{out}})$$

得到一个伪期望算子 $\tilde{\mathbb{E}}$。

2. 设置 $\hat{w}\leftarrow\tilde{\mathbb{E}}[w_{\text{out}}]$.

3. 返回$\hat{w}$。

首先，我们验证其 可行性，程序 $\mathscr{A}$ 的证明要点见 C.1 节 .

引理 6.7。 干净样本以至少 $1-\beta$ 的概率满足程序约束 $\mathscr{A}(X、y、\xi、w;\ X_{\text{input}}、y_{\text{input}})$。

证明。 根据定义，满足污染约束 $\mathscr{A}_{\text{corr}}$。对于韧性约束 $\mathscr{A}_{\text{res}}$，由定理 C.1可知，以至少 $1-\beta$ 的概率，$\left\|\frac{1}{n}X\circ X^{\circ\intercal}-\mathrm{Id}\right\|_{\text{op}}$上的界成立。另一方面，

向量 y 服从方差不超过 2 的独立同分布高斯分布。我们可以令z_1 表示 ηn 个 y 中绝对值最大的元素，而 z_2 表示其余元素。由定理 C.3可得

$$\|z_2\|_\infty^2 \leqslant \chi \log(1/\eta) + \frac{\chi \log(1/\beta)}{\eta n}.$$

根据定理 C.2，我们有

$$\|z_1\|_2^2 \leqslant \chi \eta n \log(1/\eta) + \chi \log(1/\beta).$$

因此，该韧性约束以至少为 $1-\beta$ 的概率得到满足。 □

Ne接下来，我们证明我们的算法输出了 $(1/\sigma^2+n)^{-1}X^\circ y^\circ$ 的精确估计值：

引理 6.8。 考虑定理 *6.5*的情形。则以至少为 $1-\beta$ 的概率，算法 *6.6*输出 $\hat{}w \in \mathbb{R}^d$，使得

$$\|\hat{w} - (1/\sigma^2+n)^{-1}X^\circ y^\circ\|_2^2 \leqslant O\Big(\eta^2 \log^2(1/\eta) + \eta \log(1/\eta)\sqrt{\frac{d+\log(1/\beta)}{n}}\Big).$$

证明。 令 $r \in \{0,1\}^n$ 为受污染样本的真实指示变量。于是，在 $(1-\xi)\odot(1-r)$的支撑集上，程序变量 X、y 与真实样本相同，即

$$X((1-\xi)\odot(1-r)) = X^\circ((1-\xi)\odot(1-r))$$
$$(1-\xi)\odot(1-r)\odot y = (1-\xi)\odot(1-r)\odot y^\circ.$$

我们有

$$Xy - X^\circ y^\circ = X(y\odot v) - X^\circ(y^\circ \odot v).$$

其中 $v = 1-(1-r)\odot(1-\xi)$ 表示真实遭到污染的点，/或被 ξ标记为遭到污染的点。

现在我们分别界定右侧的两项。对于任意单位向量 u，利用柯西-施瓦茨不等式，我们有

$$\langle X(y\odot v), u\rangle^2 \leqslant \|Xu\odot v\|^2 \|y\odot v\|^2.$$

现在根据约束 $\mathscr{A}_{\text{res}}(X、y、\xi)$，我们有

$$\|y\odot v\|^2 \lesssim \|z_1\odot v\|^2 + \|z_2\odot v\|^2 \leqslant \|z_1\|^2 + \eta n\|z_2\|_\infty^2 \lesssim \eta n\log(1/\eta) + \log(1/\beta).$$

此外，与我们在经验均值估计中使用各向同性估计引理（引理 A.1）的情形类似，我们有

$$\|Xu\odot v\|^2 = \sum_i v_i \langle X(\cdot,i),u\rangle^2 \leqslant \eta n\log(1/\eta) + \sqrt{\frac{d+\log(1/\beta)}{n}}n.$$

我们可以类似地界定 $\langle X。(y^\circ\odot v)、u\rangle^2$，因为真实值是一个可行解。因此，对于每个方向 u，我们都有一个次数为 $O(1)$ 的平方和证明，表明

$$\langle Xy - X^\circ y^\circ, u\rangle^2 \lesssim (\eta n\log(1/\eta) + \log(1/\beta))\cdot\Big(\eta n\log(1/\eta) + \sqrt{\frac{d+\log(1/\beta)}{n}}n\Big).$$

对$w_{\text{out}} = (1/\sigma^2+n)^{-1}Xy$取伪期望，我们得到一个具有所述误差率的估计量：

$$\alpha \leqslant O\Big(\eta\log(1/\eta) + \sqrt{\eta\log(1/\eta)\sqrt{\frac{d+\log(1/\beta)}{n}}}\Big).$$ □

最后，我们通过集中性证明，我们的估计量接近后验均值。

Lem引理 6.9。 考虑 定理 6.5中的设定。以至少 $1-\beta$ 的概率，我们有

$$\|\mathbb{E}[w^* \mid X^\circ, y^\circ] - (\Sigma^{-1} + n\mathrm{Id}_n)^{-1} X^\circ y^\circ\| \lesssim \frac{d + \log(1/\beta)}{n}.$$

证明。 我们的回归模型的后验均值为

$$\mathbb{E}[w^* \mid X^\circ, y^\circ] = \left(X^\circ X^{\circ\intercal} + \Sigma^{-1}\right)^{-1} X^\circ y^\circ.$$

根据 定理 C.1，以 $1-\beta$ 的概率，我们有

$$\left\| \frac{1}{n} X^\circ X^{\circ\intercal} - \mathrm{Id}_n \right\|_{\mathrm{op}} \leqslant \sqrt{\frac{d + \log(1/\beta)}{n}}$$

因此，我们得到

$$\left\| n(\Sigma^{-1} + n\mathrm{Id}_n)^{-1} - n(\Sigma^{-1} + X^\circ X^{\circ\intercal})^{-1} \right\|_{\mathrm{op}} \lesssim \sqrt{\frac{d + \log(1/\beta)}{n}}.$$

由于以至少 $1-\beta$ 的概率，我们有 $\left\| {}^{1}{}_{n} X \circ y^\circ \right\| \leqslant O(\sqrt{d + \log(1/\beta)n})$，故结论成立。 □

定理 定理 6.5的证明。 结合 引理 6.9 和 引理 6.8 以及三角不等式，我们知道，以至少 $1-\beta$ 的概率，有

$$\|\hat{w} - (1/\sigma^2 + X^\circ X^{\circ\intercal})^{-1} X^\circ y^\circ\|^2 \leqslant O\left(\eta^2 \log^2(1/\eta) + \eta \log(1/\eta) \sqrt{\frac{d + \log(1/\beta)}{n}}\right).$$

转换为样本复杂度界后，即得所述结论。 □

6.1.2 从鲁棒性到隐私的归约

本节中，我们将应用从鲁棒性到隐私的归约来证明 定理 6.5。我们将对我们的算法 6.6进行隐私化。我们对回归的可认证后验均值作如下定义。

定义 6.10 ((α, τ, T)-回归的可认证后验均值)。 固定 $R > 0$ 和 $\alpha, \tau > 0$。对于输入 $(X_{\mathrm{in}}, y_{\mathrm{in}})$，若存在一个定义在不定元 $(X, y, \xi, w, w_{\mathrm{out}}, M)$ 上的次数为 $O(1)$ 的线性泛函 L（伪期望），使得向量 $\tilde{w} \in \mathbb{R}^d$ 满足 (α, τ, T)-可认证，则称其具有该性质：

1. ***Closeness/boundedness:*** $\|L[w_{\mathrm{out}}] - \tilde{w}\|_2 \leqslant \alpha$ *and* $\|L[w_{\mathrm{out}}]\|_2 \leqslant 2R + \tau T$.
2. 近似可行性（次数为 $\boldsymbol{6}$）：L τ-近似满足约束族的并集

$$\mathscr{A} = \underbrace{\mathscr{A}_{\mathrm{corr}}}_{\textit{corruption}} \cup \underbrace{\mathscr{A}_{\mathrm{res}}}_{\textit{design/residual stability}} \cup \underbrace{\mathscr{A}_{\mathrm{post}}}_{w_{\mathrm{out}} = (1/\sigma^2 + n)^{-1} X y},$$

具体而言：

- $\mathscr{A}_{\text{相关性}}$： $\xi \odot \xi = \xi, \|\xi\|_1 \leqslant \eta n$ 和 $(1 - \xi(i))X(i, \cdot) = (1 - \xi(i))X_{\text{在}}(i, \cdot), (1 - \xi(i))y(i) = (1 - \xi(i))y_{\text{在}}(i$对所有 i。

- $\mathscr{A}_{\text{结果}}$: $\left\| \frac{1}{n} X X^\top - I \right\|_{\text{运算}} \lesssim \sqrt{(d + \text{对数}(1/\beta))/n}$ 以及分解 $y = z_1 + z_2$ 与 $\|z_1\|_2^2 \lesssim \eta n$ 对数 $(1/\eta) + \text{对数}(1/\beta)$ 和 $\|z_2\|_\infty^2 \lesssim \text{对数}(1/\eta) + \text{对数}(1/\beta)/(\eta n)$。
- $\mathscr{A}_{\text{后}}$: $w_{\text{超出}} = (1/\sigma^2 + n)^{-1} X y$.

我们选择 $R = \Theta(\sigma\sqrt{d})$，并令 $\mathscr{A}_{\text{res}}$ 内的松弛量为

$$\alpha_0(\eta) \asymp \sqrt{\frac{d + \log(1/\beta)}{n}} + \eta \log 1/\eta\,.$$

定义 6.11 (回归评分)。对于 $\widetilde{w} \in \mathbb{B}_2^d(2R + n\tau + \alpha)$，定义

$$\mathscr{S}(\widetilde{w}; X_{\text{in}}, y_{\text{in}}, \alpha, \tau) := \min\left\{T \geqslant 0 : \widetilde{w} \text{ is } (\alpha, \tau, T)\text{-certifiable for } (X_{\text{in}}, y_{\text{in}})\right\}.$$

现在我们针对不同的评分值界定体积比。

引理 6.12 (低分集合局部化于w_{out}附近)。存在一个通用常数 $\eta_\star > 0$，使得在由韧性约束定义的稳定性事件上（该事件以概率 $1 - \beta$ 成立），以下结论同时成立。令

$$\alpha(\eta)^2 \lesssim \eta^2 \log^2 1/\eta + \eta \log(1/\eta) \cdot \sqrt{\frac{d + \log(1/\beta)}{n}}\,.$$

1. 完备性（内球）。若 $\|\widetilde{w} - w_{\text{out}}\|_2 \leqslant \alpha(\eta)$ 则 $\mathscr{S}(\widetilde{w}; \cdot) \leqslant \eta n$。
2. 可靠性（外球）。若 $\mathscr{S}(\widetilde{w}; \cdot) \leqslant \eta_\star n$，则 $\|\widetilde{w} - w_{\text{out}}\|_2 \lesssim \alpha(\eta_\star)$。

证明。 对于 (1)，使用一个见证，该见证将点质量赋予与$(X_{\text{in}}, y_{\text{in}})$一致的未污染赋值；这些约束是可行的，且$L[w_{\text{out}}] = w_{\text{out}}$。对于 (2)，结合 $\mathscr{A}_{\text{res}}$ 回归的 SoS 可识别性与定义 6.10中的接近性约束。

□

我们需要证明，对于所有候选点 $\widetilde{w}$，评分函数的值都有上界n。

引理 6.13 （评分函数的良定义性）。对于任意 $\widetilde{w} \in \mathbb{R}^d$，评分函数的值不超过 n。

证明。 关于可行性，我们注意到可以取$y = \mathbf{1}$，证明的其余部分与 [ABMT25，引理 6.15] 类似。具体而言，当 $\widetilde{w}$ 的范数不超过 $2R + \tau T$时，我们可以设$X_i = \left(\frac{1}{n\sigma^2} + 1\right) \widetilde{w}$。否则，我们将<style id="30">$w$ 投影到半径为 $2R + \tau T - \phi$ 的球上，其中 ϕ，并从分布正态分布 $(\widetilde{w}, \mathrm{Id})$ 中抽样，其中 $\widetilde{w}$ 为投影结果。

□

引理 6.14。 在定义 *6.11* 中定义的评分函数是拟凸的。

证明。 大多数约束与经验均值估计程序中的约束相似，但对y的韧性约束除外。然而，根据伪分布的性质，很容易看出这些凸约束满足拟凸性。

□

现在我们通过效用分析完成对定理 6.5 的证明。

定理 *6.5*的证明。 将从鲁棒性到隐私的归约应用于定义域 $\mathscr{S}$上的评分函数 $\mathbb{B}_2^d(2R + n\tau + \alpha)$。令 V_η 表示该定义域内 $\{\widetilde{w} : \mathscr{S}(\widetilde{w}) \leqslant \eta n\}$ 的体积。根据 引理 6.12：

- 局部区域 $\eta \leqslant \eta' \leqslant \eta_\star$：低分集合夹在以$w$out为中心、半径为 ℓ_2-球的$\alpha(\eta)$和 $\alpha(\eta')$之间，因此

$$\log \frac{V_{\eta'}}{V_\eta} \lesssim d \cdot \log \frac{\alpha(\eta')}{\alpha(\eta)}.$$

- 全局情形 $\eta_\star \leqslant \eta' \leqslant 1$：得分-$\leqslant \eta_\star$ 集合位于一个$O(\alpha(\eta_\star))$-球内，而环境半径为$O(R+n\tau)$，因此

$$\log \frac{V_{\eta'}}{V_\eta} \lesssim d \cdot \log \frac{R+n\tau}{\alpha(\eta_\star)}.$$

最后，很容易验证该评分函数是拟凸的，且可以高效计算。于是，定理 4.19 中的归约给出了一个充分条件

$$n \gtrsim \max_{\eta \leqslant \eta' \leqslant \eta_\star} \frac{d \log(\alpha(\eta')/\alpha(\eta)) + \log(1/\beta_\eta)}{\varepsilon \eta'} + \max_{\eta_\star \leqslant \eta' \leqslant 1} \frac{d \log\big((R+n\tau)/\alpha(\eta_\star)\big) + \log(1/\beta_\eta)}{\varepsilon}.$$

使用 $\alpha(\eta)^2 \asymp \eta \cdot {}^{d+\log(1/\beta)}{}_n$ 对 η 进行优化，在小-η 情形下得到两个“局部”项$d+\log^{(1/\beta)}_{\alpha^{4/3}\varepsilon^{2/3}}$ 和$d+\log^{(1/\beta)}_{\alpha\varepsilon}$。对于全局项，取$R = \Theta(\sigma\sqrt{d})$，这样其贡献为$d\log(\sigma\sqrt{d}/\varepsilon)$，相差对数因子。

□

注 6.15 (关于 σ 以及从w_{out}进行转换)。 $d\log(\sigma\sqrt{d}/\varepsilon)$ 项是回归中对应于均值估计里的全局体积项的项，完全源于环境球与局部化评分集之间的比值。在 $\sigma = \Omega(n^{-1/2})$的情形下，$R = \Theta(\sigma\sqrt{d})$ 与 $\|w_{\text{out}}\|_2$的典型尺度相符。标准集中性结果表明，$w_{\text{out}} = (\frac{1}{\sigma^2} + n)^{-1} X^\top y$与真实贝叶斯后验均值相差不超过 $\widetilde{O}(\sqrt{(d + \log(1/\beta))/n})$，在所述范围内，该误差可被吸收到 α 中。

6.2 弱先验下私有贝叶斯线性回归的计算速率

我们在弱先验区间的主要保证如下：

定理 6.16. 设 $\Sigma = \sigma^2 \mathrm{Id}_d$，已知参数 $\sigma = \omega(1/\sqrt{n})$。设 $\beta \in [0,1]$，$\overline{\varepsilon} \leqslant O(1)$。存在一个多项式时间算法，给定 η个受污染样本 (X, y)，这些样本按照定义 *6.1*生成，算法会输出一个 ε-*DP* 估计量 $\hat{w}$，使得其成立的概率至少为 $1-\beta$，

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|^2 \leqslant \alpha^2,$$

其中 (X°, y°) 是未污染样本，且满足条件 $\alpha \geqslant \eta \log(1/\eta)$ 以及

$$n \geqslant \tilde{\Omega}\Big(\frac{d+\log(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{d+\log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(\sigma\sqrt{d})}{\varepsilon} + \frac{d\eta^2}{\alpha^4}\Big).$$

与临界先验情形相比，证明过程更加复杂，因此我们将部分细节推迟至附录 C。

在 $\sigma = \omega(1/\sqrt{n})$ 的情形下，先验分布足够弱，因此普通最小二乘估计量渐近达到贝叶斯最优误差。我们设计了一种多阶段算法。在第一阶段，我们获得一个误差率为$O(1)$的粗略估计器。在第二和第三阶段，我们将估计量逐步改进至目标准确度。

6.2.1 多阶段算法概述

这里我们概述弱先验区间中的方法。首先，如果没有对真实回归向量 先验方差的上界 w，就无法直接应用临界先验情形中的策略，因为我们无法控制标签 y的大小。因此，第一步是通过学习 粗略估计值 w (第 6.2.2 节)，对标签进行“重新中心化”。这同样可以通过短平分解约束实现，但这次约束并非施加于标签向量 $\hat{y}$，而是施加于残差向量 $\hat{y} - \hat{X}^\top \hat{w}$。一个

可以用与上述类似的论证证明，这将得到真实 w_{init} 的一个估计值，其精度达到 w，误差在 $O(1)$ 内（按 L_2 度量）。

经过粗估计阶段后，直观而言，我们现在可以从 w 中“减去”粗略估计值 w_{init}，将问题简化为类似比例情形的问题。但这会遇到若干障碍。

首先，由于 $\tilde{X}$ 中存在污染，我们不想直接从 $\tilde{y}$ 中减去 $\tilde{X}w_{\text{init}}$，因此，我们转而将粗略估计值间接纳入一个新的平方和程序，并寻找 $\hat{y} - \hat{X}^\top \hat{w}$ 的短平分解，其中 $\hat{w}$ 是回归器的 SoS 变量；同时也寻找 $X^\top(\hat{w} - w_{\text{init}})$ 的短平分解。

其次，事实证明，即使如此，也只能将 w_{init} 达到的 $O(1)$ 误差细化为估计后验均值时 L_2 中的 $O(\sqrt{d/n})$ 误差。这是因为 $\frac{1}{n} X X^\top - \mathrm{Id}$ 的谱范数以高概率仅能界定为 $O(\sqrt{d/n})$。因此，一直接推导只能表明 $\frac{1}{n} X y$ 与最小二乘估计量 $(X X^\top)^{-1} X y$ 在 L_2 中的距离为 $\sqrt{d/n}$。于是，给定达到后一误差界的向量 w_1，上述寻找标签和残差短平分解的程序需要重新运行，并将 w_{init} 替换为 w_1。尤其是，得益于 w_1，的误差界得到细化，现在可以证明向量 $v_1 := (n + \frac{1}{\sigma^2})^{-1} X (y - X^\top w_1)$ 与 $v_2 := (X X^\top + \frac{1}{\sigma^2})^{-1} X (y - X^\top w_1)$ 很接近，具体而言，以高概率有

$$\|v_1 - v_2\| \leqslant O\Big(\sqrt{\eta \log 1/\eta} + \sqrt{d/n}\Big) \cdot \Big(\sqrt{d/n}\Big).$$

这将使我们能够证明所需的误差界 $\tilde{O}(\eta + \sqrt{\eta}\sqrt{d/n})$。关于后两个 SoS 程序的详细内容，我们将在第 6.2.3 节中讨论。

最后，我们指出，由于我们是在贝叶斯设定下工作，因此依次运行三个 SoS 程序以获得最终后验均值估计值时，必然都要使用同一数据集。这可能看起来有些问题，因为后两个 SoS 程序是自适应构造的：其中某些约束取决于前面的 SoS 程序计算出的估计值w_{init}、w_1，而这种数据重用可能会破坏可行性。正如我们将看到的（引理 6.20），事实证明这并不是问题，原因基本上在于，我们可以利用设计矩阵X的“稀疏奇异值”界来说明，无论w_{init}、w_1，为何，只要它们与$O(1)$-接近，w，向量$X^\top(w - w_{\text{init}})$和$X^\top(w - w_1)$都存在短平分解。

6.2.2 第1阶段：粗略估计

在第1阶段，我们将使用一个与临界先验情形中所用程序相似的平方和程序，但存在以下关键差异。

首先，污染约束 $\mathscr{A}_{\text{corr}}$ 与之前相同（式（23））。韧性约束与之前类似，但我们不再要求标签 y 具有短平分解，而是要求残差 $y - X^\top w$ 具有短平分解：

$$\mathscr{A}'_{\text{res}}(X, y, w) = \left\{ \begin{array}{l} \left\| \frac{1}{n} X X^\top - \mathrm{Id} \right\|_{\text{op}} \leqslant \chi \sqrt{\frac{d + \log(1/\beta)}{n}}, \\ y - X^\top w = z_1 + z_2, \|z_1 + z_2\|^2 \leqslant 2n \\ \|z_1\|^2 \leqslant \eta n \log(1/\eta) + \log(1/\beta), \\ \|z_2\|_\infty^2 \leqslant \chi \log(1/\eta) + \log(1/\beta)/\eta n \end{array} \right\}. \tag{26}$$

形式化地，我们有以下保证，表明真实参数 w^* 可以在常数误差范围内恢复：

引理 *6.17*（弱先验下的粗略估计）。设 $\Sigma = \sigma^2 Id_d$，其中参数 $\sigma = \omega(1/\sqrt{n})$ 已知。存在一个多项式时间算法，给定按定义 *6.1* 生成的 η 受污染样本 (X, y)，输出 $\hat{w}$，使得以至少 $1-\beta$ 的概率，有 $\|\hat{w} - w^*\|_2 \leqslant O(1)$，前提是 $n \geqslant \Omega((d + log(1/\beta))/\eta)$。

该算法将求解程序的次数为 $O(1)$ SoS 松弛 $\mathscr{A}_{\text{corr}} \cup \mathscr{A}'_{\text{res}}$，并对得到的伪分布进行朴素舍入（见 算法 C.10）。SoS 可识别性证明的细节将在C.3 节中给出。

6.2.3 第 2 阶段：精细化

在接下来的阶段中，我们将依次运行两个 SoS 程序。第一个程序将把前一阶段得到的估计值作为参数，记为 w_{init}。这将得到一个新的估计值 $\hat{w}_1$，其对真实参数 w^* 的估计误差为 $O(\sqrt{\eta \log 1/\eta} + \sqrt{\frac{d+\log(1/\beta)}{n}})$。我们将把 $\hat{w}_1$ 输入最终的 SoS 程序，进一步精细化，得到估计值 $\hat{w}_2$，使其在逼近后验均值时达到所需的误差水平。

与第1阶段一样，我们使用相同的污染约束 $\mathscr{A}_{\text{corr}}$，来自式（23）。对于韧性约束，在第一个程序中，除了要求残差$y - X^\intercal w$具有短平分解外，我们还要求初始化得到的标签$X^\intercal w_{\text{init}}$ 与w_{init} 在标签$X^\intercal w$下由SoS变量w预测得到的标签之间的差值也具有短平分解：

$$\mathscr{A}''_{\text{res}}(X, y; w_{\text{init}}) = \left\{ \begin{array}{l} \left\| \frac{1}{n} XX^\intercal - \text{Id} \right\|_{\text{op}} \leqslant \chi \sqrt{\frac{d+\log(1/\beta)}{n}}, \\ y = X^\intercal w + z, z = z_1 + z_2, \|z_1 + z_2\|^2 \leqslant 2n \\ \|z_1\|^2 \leqslant \eta n \log(1/\eta) + \log(1/\beta), \\ \|z_2\|_\infty^2 \leqslant \chi \log(1/\eta) + \log(1/\beta)/\eta n \\ X^\intercal (w - w_{\text{init}}) = b_1 + b_2, \\ \|b_1\|^2 \leqslant \eta n \log(1/\eta), \|b_2\|_\infty^2 \leqslant \chi \log(1/\eta) \end{array} \right\}. \tag{27}$$

类似地，对于第二个程序，我们将w_{init} 替换为 $\hat{w}_1$，得到 $\mathscr{A}''\text{res}(X、y; \hat{w}_1)$。

对于第一个程序，我们添加一个关于最小二乘估计值定义的约束 r:

$$\mathscr{A}_{\text{LS}}(X, y, w, w_{\text{LS}}) = \left\{ w_{\text{LS}} = \frac{1}{n} Xy, \|w - w_{\text{LS}}\|^2 \leqslant \frac{\chi(d + \log(1/\beta))}{n} \right\}, \tag{28}$$

对于第二个程序，我们将使用式（25）中关于（简化的）后验均值定义的约束。

形式化地，我们有以下保证，表明后验均值可以在所需误差范围内得到估计：

引理 6.18。设 $\Sigma = \sigma^2 \text{Id}_d$ 对应已知参数 $\sigma = \omega(1/\sqrt{n})$。存在一个多项式时间算法，给定 η受污染样本 $(X、y)$，其生成方式如 定义 *6.1*所述，并给定粗略估计值 w_{init}，满足 $\|w_{init} - w^*\| \leqslant O(1)$，输出 $\hat{w}$，使得以至少 $1-\beta$的概率，有

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|^2 \leqslant O\Big(\eta \log(1/\eta) \sqrt{\frac{d+\log(1/\beta)}{n}} + \eta^2 \log^2(1/\eta)\Big)$$

给定 $n \geqslant (d + \log(1/\beta))/\eta$。

结合阶段 1 和阶段 2 的完整算法如下：算法 6.19：

算法 6.19（用于近似后验均值估计量的鲁棒算法）。输入：η 受污染矩阵 $X_{\text{input}} \in \mathbb{R}^{n\times m}$、$y_{\text{input}} \in \mathbb{R}^n$输出：一个多项式时间估计器 $\hat{w} \in \mathbb{R}^d$

1. 运行算法 *C.10*，以获得估计量 w_{init}。 *2.* 求解约束 $\mathscr{A}(X, y, \xi, w; X_{\text{input}}, y_{\text{input}}, w_{init}) = \mathscr{A}_{\text{corr}}(X, y, \xi; X_{\text{input}}, y_{\text{input}}) \cup \mathscr{A}''\, res(X, y, w; w_{init}) \cup \mathscr{A}_{LS}(X, y, w_{LS})$的次数为 $O(1)$ 的 *SoS* 松弛，其中 $t = O(1)$，从而得到伪期望算子 $\tilde{\mathbb{E}}$。 *3.* 对伪分布进行舍入，得到 $\hat{w}_1 = \tilde{\mathbb{E}}[w]$。 *4.* 求解约束 $\mathscr{A}(X, y, \xi, w; X_{\text{input}}, y_{\text{input}}, \hat{w}_1) = \mathscr{A}_{\text{corr}}(X, y, \xi; X_{\text{input}}, y_{\text{input}}) \cup \mathscr{A}''\, res(X, y, w; \hat{w}_1)$ 的次数为 $O(1)$ 的 *SoS* 松弛，其中 $t = O(1)$，从而得到伪期望 $\tilde{}$算子 $\mathbb{E}$。 *5.* 对伪分布进行舍入，得到 $\hat{w}_2 = \tilde{\mathbb{E}}[(n + \frac{1}{\sigma^2})^{-1} X (y - X\hat{w}_1)]$，并返回 $(n + \frac{1}{\sigma^2})^{-1}\hat{w}_1 + \hat{w}_2$。

首先，我们验证相关程序的可行性。

引理 6.20。 算法*2* 和 *4* 中的程序约束 算法 *6.19* 以至少 $1 - 3\beta$ 的概率满足真实值。

证明。 污染约束 $\mathscr{A}_{\text{corr}}$ 根据定义得到满足。对于韧性约束，根据定理 C.1，以至少 $1 - \beta$ 的概率，关于 $\left\| \frac{1}{n} X \circ X \circ^{\intercal} - \mathrm{Id} \right\|_{\text{op}}$ 的界成立。另一方面，向量 $y^\circ - (X^\circ)^{\intercal} w$ 是方差不超过 1 的独立同分布高斯分布。令 z_1 表示 ηn 个 y 中绝对值最大的元素。根据 定理 C.3，我们有

$$\|z_2\|_\infty^2 \leqslant \chi \log(1/\eta) + \frac{\chi \log(1/\beta)}{\eta n}.$$

根据 定理 C.2，我们有

$$\|z_1\|^2 \leqslant \chi\, \eta n \log(1/\eta) + \chi \log(1/\beta).$$

固定 $u := w^* - w_{\text{init}}$，并令 $a := X^{\circ\intercal}(w^* - w_{\text{init}}) \in \mathbb{R}^n$。令 $S \subseteq [n]$ 为 $|a_i|$ 中最大 $k = \eta n$ 个值对应的索引（并列时任意打破），并定义

$$b_1 := a \odot \mathbf{1}_S, \qquad b_2 := a \odot \mathbf{1}_{S^c}.$$

那么根据构造，$a = b_1 + b_2$。

根据引理 C.4，存在一个普适常数 $C > 0$，使得对于 X，以至少 $1 - \beta$ 的概率，有 $\|X\|_{\text{sp}}(\eta n) \leqslant L$，其中 $L = C(\sqrt{d} + \sqrt{\eta n \log \frac{e}{\eta}} + \sqrt{\log \frac{1}{\beta}})$。在条件 $n \geqslant \overline{d + \log \frac{1}{\eta}(1/\beta)}$ 下，有 $L \lesssim \sqrt{\eta n \log(1/\eta)}$。

对于 b_1，注

$$\|b_1\|_2 = \|a_S\|_2 = \|X_S^{\intercal} u\|_2 \leqslant \|X_S\|_{\text{op}} \|u\|_2 \leqslant \|X\|_{\text{sp}}(k) \|u\|_2 \leqslant L\|u\|_2,$$

这意味着 $\|b_1\|_2^2 \leqslant L^2 \|u\|_2^2$。

对于b_2，的上界，令 $|a|_{(1)} \geqslant \cdots \geqslant |a|_{(n)}$为$\{|a_i|\}^n{}_{i=1}$的次序统计量。根据$S$的定义，对每个$i \notin S$，都有 $|b_2(i)| = |a_i| \leqslant |a|_{(k+1)} \leqslant |a|_{(k)}$。此外，

$$|a|_{(k)}^2 \leqslant \frac{1}{k}\sum_{j=1}^{k}|a|_{(j)}^2 = \frac{1}{k}\|a_S\|_2^2 = \frac{1}{\eta n}\|b_1\|_2^2.$$

因此，

$$\|b_2\|_\infty^2 = \max_{i\notin S}|a_i|^2 \leqslant |a|_{(k)}^2 \leqslant \frac{1}{k}\|b_1\|_2^2 \leqslant \frac{L^2}{\eta n}\|u\|_2^2,$$

其中最后一步用到了b_1 的界。因此，程序约束得以满足。 □

接下来，我们证明第 2 阶段中第一个程序生成的 $\hat{}w_1$ 达到误差$O(\sqrt{\eta \log\ 1/\eta} + \sqrt{d+\log 1/\beta\, n})$:

引理 6.21。 考虑第 *3* 步中的算法 *6.19*。存在一个次数为 $O(1)$ 的平方和证明，使得

$$\mathscr{A}(X^{(1)}, y^{(1)}, \xi^{(1)}, w^{(1)}; X_{\text{input}}, y_{\text{input}}),\ \mathscr{A}(X^{(2)}, y^{(2)}, \xi^{(2)}, w^{(2)}; X_{\text{input}}, y_{\text{input}})$$
$$\vdash \|w^{(1)} - w^{(2)}\|^2 \leqslant O\Big(\eta\log(1/\eta) + \frac{d+\log(1/\beta)}{n}\Big).$$

由标准舍入可知，

$$\|\hat{w}_1 - w^*\|_2 \leqslant O\Big(\sqrt{\eta\log 1/\eta} + \sqrt{\frac{d+\log(1/\beta)}{n}}\Big). \tag{29}$$

证明。 对于$j \in \{1,2\}$，定义$y_1^{(j)} = y^{(j)} - (X^{(j)})^\intercal w_{\text{init}}$和$w_1^{(j)} = w^{(j)} - w_{\text{init}}$。于是有 $\|w_1^{(j)}\|^2 \leqslant O(1)$。我们的目标是给出一个平方和证明，证明

$$\|w_{\text{LS}}^{(1)} - w_{\text{LS}}^{(2)}\|^2 \leqslant O\Big(\eta\log(1/\eta) + \frac{d+\log(1/\beta)}{n}\Big).$$

令 $v = 1 - (1-\xi^{(1)}) \odot (1-\xi^{(2)})$，注意到

$$X^{(1)}(1-v) = X^{(2)}(1-v)$$
$$(1-v)\odot y_1^{(1)} = (1-v)\odot y_1^{(2)}.$$

我们进行如下拆分

$$X^{(1)}y_1^{(1)} - X^{(2)}y_1^{(2)} = X^{(1)}(y_1^{(1)}\odot v) - X^{(2)}(y_1^{(2)}\odot v)$$

并利用 X 的谱范数界分别估计这两项。具体而言，我们有

$$\|X^{(1)}(y_1^{(1)}\odot v)\|^2 = \|X^{(1)}\|_{\text{op}}^2 \cdot \|y_1^{(1)}\odot v\|^2 \leqslant \chi n \cdot \|y_1^{(1)}\odot v\|^2.$$

根据 $y^{(1)}$ 的短平分解，

$$\|y_1^{(1)}\odot v\|^2 \leqslant \chi^2\eta n\log(1/\eta) + \log(1/\beta).$$

因此，我们有

$$\|X^{(j)}(y_1^{(j)}\odot v)\|^2 \leqslant O(\eta n^2\log(1/\eta) + \log(1/\beta)n)$$

对于 $j = 1$、2，因此有 $\|w^{(1)}\text{LS} - w^{(2)}\text{LS}\|^2 \leqslant \eta\log(1/\eta) + \log(1/\beta)/n$。另一方面，根据程序约束，我们有 $\|w^{(j)} - w^{(j)}\text{LS}\|^2 \leqslant O(^{d+\log(1/\beta)}{}_n)$，由三角不等式可得 $\|w^{(1)} - w^{(2)}\|$ 的界。

□

最后，我们证明 算法 6.19 中的最后一个程序能够对后验分布均值给出足够精确的近似。

引理 6.18 的证明。令 $y' = y - X^\intercal \hat{w}_1$, and $w' = w - \hat{w}_1$。令 $y^{\circ\prime} = y^\circ - X^\circ{}^\intercal \hat{w}_1$ and $w^{*\prime} = w^* - \hat{w}_1$。根据引理 6.21 对 $\hat{w}_1$ 的保证，我们有 $\|w'\|^2 \leqslant O\Big(\eta \log(1/\eta) + d + \log(1/\beta) n\Big)$。

现在我们给出平方和证明，说明对于任意单位向量 u

$$\langle X y' - X^\circ y^{\circ\prime}, u\rangle^2 \leqslant O\Big(\eta^2 \log(1/\eta) + \eta\sqrt{\frac{d+\log(1/\beta)}{n}}\Big).$$

令 r 为受损节点集合的指示函数。令 $v = 1 - (1-\xi)\odot(1-r)$，注意到

$$X((1-v)\odot y) = X^\circ((1-v)\odot y^{\circ\prime})$$
$$(1-v)\odot y = (1-v)\odot y^{\circ\prime}.$$

和之前一样，我们进行如下拆分

$$X y' - X^\circ y^{\circ\prime} = X(y'\odot v) - X^\circ(y^{\circ\prime}\odot v)$$

并利用 X 的谱范数界分别估计这两项。特别地，对于任意单位向量 u,

$$\langle X(y'\odot v), u\rangle^2 \leqslant \|Xu\odot v\|^2 \|y'\odot v\|^2.$$

现在类似于 我们关于经验均值估计的证明，我们有一个平方和证明，表明 t

$$\frac{1}{n}\|Xu\odot v\|^2 \leqslant O\Big(\eta\log(1/\eta) + \sqrt{\frac{d+\log(1/\beta)}{n}}\Big)$$
$$\|y'\odot v\|^2 \leqslant (\eta n \log(1/\eta) + \log(1/\beta)).$$

结合以上结果，我们有一个平方和证明，表明

$$\frac{1}{n}\langle X(y'\odot v), u\rangle^2 \leqslant O\Big(\eta\log(1/\eta) + \sqrt{\frac{d+\log(1/\beta)}{n}}\Big)\cdot(\eta n\log(1/\eta) + \log(1/\beta)), \tag{30}$$

以及对 ${}^1{}_n\langle X_\circ(y^{\circ\prime}\odot v), u\rangle_-{}^2$ 的相同界。因此，根据 第 6.2.3 节，对于任意满足约束的伪分布 $\tilde{\mathbb{E}}$ 以及任意单位向量 u，有 $\tilde{\mathbb{E}}\Big[{}^1{}_n\langle Xy' - X_\circ y^{\circ\prime}, u\rangle^2\Big]$ 的上界为式（30）右侧的值。因此，通过对伪分布进行舍入，我们得出

$$\Big\|\frac{1}{n}\Big(\tilde{\mathbb{E}}[Xy']\Big) - X^\circ y^{\circ\prime}\Big\|^2 \leqslant O\Big(\eta\log(1/\eta) + \sqrt{\frac{d+\log(1/\beta)}{n}}\Big)\cdot(\eta\log(1/\eta) + \log(1/\beta)/n). \tag{31}$$

现在我们证明，（正态分布 $+\frac{1}{\sigma^2}$）$^{-1}X_\circ y^{\circ\prime} + \Big(1 + {}^1{}_{\text{正态分布}\sigma^2}\Big)^{-1}\hat{w}_1$ 接近真实后验均值 $(X_\circ X^{\circ\intercal} + \frac{1}{\sigma^2}\mathrm{Id})^{-1}X^\circ y^\circ$。根据定义，我们有 $y^{\circ\prime} = y^\circ - X_\circ{}^\intercal \hat{w}_1$。现在我们有

$$\begin{aligned}
&(n+\frac{1}{\sigma^2})^{-1}X^\circ y^{\circ\prime}\\
&\quad= \Big(X^\circ X^{\circ\intercal} + \frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}X^\circ(y^\circ - X^{\circ\intercal}\hat{w}_1) + \Big((n+\frac{1}{\sigma^2})^{-1} - \Big(X^\circ X^{\circ\intercal} + \frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\Big)X^\circ(y^\circ - X^{\circ\intercal}\hat{w}_1)\\
&\quad= \Big(X^\circ X^{\circ\intercal} + \frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}X^\circ y^\circ - \hat{w}_1 + \Big(X^\circ X^{\circ\intercal} + \frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\frac{\hat{w}_1}{\sigma^2}
\end{aligned}$$

$$+\Big((n+\frac{1}{\sigma^2})^{-1}-\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\Big)X^{\circ}(y^{\circ}-X^{\circ\intercal}\hat{w}_1).$$

首先，

$$\left\|\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\frac{\hat{w}_1}{\sigma^2}-\frac{\hat{w}_1}{n\sigma^2+1}\right\|\leqslant\left\|\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}-\frac{\sigma^2}{n\sigma^2+1}\right\|_{\mathrm{op}}\frac{\|\hat{w}_1\|}{\sigma^2}$$

此外，我们有

$$\left\|\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}-\frac{\sigma^2}{n\sigma^2+1}\mathrm{Id}\right\|_{\mathrm{op}}\leqslant\frac{1}{n}\left\|\frac{1}{n}X^{\circ}X^{\circ\intercal}-\mathrm{Id}\right\|_{\mathrm{op}}\leqslant\frac{1}{n}\sqrt{\frac{d+\log(1/\beta)}{n}}.$$

综合起来，我们得到

$$\left\|\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\frac{\hat{w}_1}{\sigma^2}-\frac{\hat{w}_1}{n\sigma^2+1}\right\|\leqslant\sqrt{\frac{d+\log(1/\beta)}{n}}\cdot\frac{\|\hat{w}_1\|}{n\sigma^2}\lesssim\sqrt{\frac{d+\log(1/\beta)}{n}}\cdot\frac{1}{n\sigma}\lesssim\frac{d+\log(1/\beta)}{n}.$$

最后，我们希望界定 $\Big((n+\frac{1}{\sigma^2})^{-1}-\Big(\underline{X}\circ\underline{X}^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\Big)\underline{X}^{\circ}(y^{\circ}-\underline{X}^{\circ\intercal}\hat{w}_1)$ 的范数。特别地，以至少 $1-\beta$ 的概率，对于某个普适常数 C，我们有

$$\left\|(n+\frac{1}{\sigma^2})^{-1}\mathrm{Id}-\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\right\|_{\mathrm{op}}\leqslant\frac{1}{n}\left\|\frac{1}{n}X^{\circ}X^{\circ\intercal}-\mathrm{Id}\right\|_{\mathrm{op}}\leqslant\frac{1}{n}\sqrt{\frac{d+\log(1/\beta)}{n}}.$$

另一方面，我们有

$$\|\frac{1}{n}X^{\circ}(y^{\circ}-X^{\circ\intercal}\hat{w}_1)\|=\|\frac{1}{n}X^{\circ}X^{\circ\intercal}(w^*-\hat{w}_1)\|+\left\|\frac{1}{n}X^{\circ}z_0\right\|\leqslant O\Big(\sqrt{\eta\log 1/\eta}+\sqrt{\frac{d+\log(1/\beta)}{n}}\Big).$$

因此，我们有

$$\left\|\Big((n+\frac{1}{\sigma^2})^{-1}\mathrm{Id}-\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\Big)X^{\circ}(y^{\circ}-X^{\circ\intercal}\hat{w}_1)\right\|$$
$$\leqslant O\Big(\sqrt{\eta\log 1/\eta}+\sqrt{\frac{d+\log(1/\beta)}{n}}\Big)\cdot\Big(\sqrt{\frac{d+\log(1/\beta)}{n}}\Big).$$

综合所有结果，我们得到

$$\begin{aligned}&\left\|(n+\frac{1}{\sigma^2})^{-1}X^{\circ}y^{\circ\prime}+\Big(1+\frac{1}{n\sigma^2}\Big)^{-1}\hat{w}_1-(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id})^{-1}X^{\circ}y^{\circ}\right\|\\&\leqslant\left\|\Big((n+\frac{1}{\sigma^2})^{-1}\mathrm{Id}-\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\Big)X^{\circ}(y^{\circ}-X^{\circ\intercal}\hat{w}_1)\right\|+\left\|\Big(X^{\circ}X^{\circ\intercal}+\frac{1}{\sigma^2}\mathrm{Id}\Big)^{-1}\frac{\hat{w}_1}{\sigma^2}-\frac{\hat{w}_1}{n\sigma^2+1}\right\|\\&\leqslant O\Big(\sqrt{\eta\log 1/\eta}+\sqrt{\frac{d+\log(1/\beta)}{n}}\Big)\cdot\Big(\sqrt{\frac{d+\log(1/\beta)}{n}}\Big).\end{aligned}$$

结合公式（31），我们即可得出证明。 □

6.2.4 综合上述结果

要得到一个隐私算法，还需要将上述鲁棒性保证代入鲁棒性到隐私的归约中。我们在C.4 节中实现这一点，以证明定理 6.16。最后，将其与临界先验情形下的保证结合起来，我们便可立即推出定理 6.3。

定理的证明 定理 6.3。 当 $\sigma = o(1/\sqrt{n})$ 时，通过从先验分布中抽样即可达到所声称的误差率。当 $\sigma = \Omega(1/\sqrt{n})$ 时，我们可以结合定理 6.5、定理 6.16 和式（14）。具体而言，根据定理 6.5和定理 6.16，当$n \geqslant d\log(\sigma\sqrt{d})/\varepsilon$ 时，以概率 $1-\beta$，我们可以达到误差率 $\alpha = O(\bar{\eta}' + \sqrt{\eta'}\sqrt{d+\log(1/\beta)}{n})$，其中 $\eta' = \eta + d+\log(1/\beta)\, n\varepsilon$。 □

6.3 稳健贝叶斯回归的计算困难性

现在，我们给出稳健后验均值估计的计算下界，适用于 $\eta\log(1/\eta) \lesssim \alpha \leqslant \eta d^{1/4-o(1)}$ 这一情形。与本小节其余部分一样，首先使用简化的后验统计量进行分析会更方便。

$$w_{\mathrm{simp}}(X, y) := \left(n + \frac{1}{\sigma^2}\right)^{-1} Xy.$$

在临界先验区域 $\sigma^2 = \Theta(1/n)$，引理 6.9 推出 $w_{\mathrm{simp}}(X^\circ、 y^\circ)$ 与精确后验均值相差 $O((d + \log(1/\beta))/n)$。

定义 6.22 （区分线性回归混合分布）。 固定 $\eta \in (0, 1/2)$。令 K 为某个足够大的常数。固定一个信号参数 $\delta > 0$，并定义

$$s := \sqrt{1 + K^2\alpha^2}, \qquad a := -\frac{K\alpha}{\eta s^2}.$$

我们考虑关于 $(x_i、 y_i) \in \mathbb{R}^d \times \mathbb{R}$ 的如下单样本假设检验问题。

- 零假设分布 Q： $xi = gi$， $yi = szi$，其中 $gi \sim \mathcal{N}(0, \mathrm{Id}d)$，$zi \sim \mathcal{N}(0, 1)$，且 gi、zi 相互独立。

- 植入分布 P： 首先抽取 $u \sim$均匀分布$(\mathbb{S}d^{-1})$，$B \sim$伯努利分布(q)。 给定 (u, B)，独立地抽取 $gi \sim \mathcal{N}(0, \mathrm{Id}d)$，$zi \sim \mathcal{N}(0, 1)$，并令 $yi := K\alpha \langle gi, u\rangle + zi$。 若 $B = 0$，令 $xi := gi$。 若 $B = 1$，令 $xi := gi + a\, yiu$。

我们记录判别问题的一些基本性质：

引理 6.23。 假设 $\alpha^2 \lesssim \eta$。 则有以下结论。

1. 在 P 下，给定 u，观测到的样本是一个 q-污染的干净高斯设计回归样本，其回归器为 $w^\star = K\alpha u$。

2. 在 Q 和 P 下，y_i 的边缘分布恰好为 $\mathcal{N}(0, s^2)$。

3. 在 Q 下， $\mathbb{E}_Q[x_i y_i] = 0$，而在 P 下，给定 u， $\mathbb{E}_P[x_i y_i \mid u] = 0$。

4. 零假设分布 Q 是一个 η-污染的干净高斯设计回归样本，其回归器为 $w^\star = 0$。

我们证明，精确后验均值的稳健估计量可用于解决判别问题。

引理 6.24（从区分到稳健后验均值估计的归约）。 假设 $\sigma^2 = \Theta(1/n)$，并定义

$$w_{\mathrm{simp}}(X, y) := \left(n + \frac{1}{\sigma^2}\right)^{-1} X y.$$

考虑定义 *6.22*中的检验问题。假设

$$n \geqslant C_1 \frac{d + \log(1/\beta)}{\alpha} + C_2 \frac{\eta(d + \log(1/\beta))}{\alpha^2} + C_3 \frac{\log(1/\beta)}{\eta} \tag{32}$$

对于充分大的通用常数 C_1, C_2, C_3。

假设存在一个多项式时间估计量 $\mathscr{A}$，满足以下保证：对于任意 η污染(X, y)，其中 (X°, y°)来自定义 *6.1* 中的 $\|w^*\|_2 \leqslant \alpha$，其概率至少为 $9/10$，

$$\|\mathscr{A}(X, y) - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|_2 \leqslant \frac{\alpha}{20}.$$

那么存在一个多项式时间检验，能够以至少定义 *6.22* 中给出的概率区分 Q 与 $P1 - O(\beta) - \exp(-\Omega(\eta n))$。

具体而言，我们给出如下检验：计算 $w_{\mathrm{obs}} := w_{\mathrm{simp}}(X, y)$, $\widehat{w} := \mathscr{A}(X, y)$, 并在 $\|\widehat{w}$植入分布 $w - w_{\mathrm{obs}}\|_2 \geqslant 2\alpha$ 时输出，否则输出零假设 。

最后，我们给出一个引理，说明在我们所关注的情形下，该假设检验问题是低次难解的。

引理 6.25。 假设 $\alpha \leqslant$ 最小值$\{\eta d^{1/4}/D,\ \eta^2\}$ 且 $D \leqslant d^{10^{-4}}$。考虑定义 *6.22*中的检验问题，并令优势 $\leqslant_D$ 表示 D 次优势。则存在一个普适常数 $c > 0$，使得当 $n \leqslant c\, d\eta^2 D^2\alpha^4$时，有优势$\leqslant_D O(1)$。

由这两个结果，我们可以得出回归中稳健后验均值估计的信息—计算差距。

定理 6.26（回归中稳健后验均值估计的信息—计算差距）。存在绝对常数 $c_0, C_0 > 0$，使得对所有充分大的 d，以下结论成立。假设$\alpha \leqslant$ 最小值$\{\eta d^{1/4}/D,\ \eta^2\}$。设

$$C_0 \frac{d + \log(1/\beta)}{\alpha} + C_0 \frac{\eta(d + \log(1/\beta))}{\alpha^2} + C_0 \frac{\log(1/\beta)}{\eta} \leqslant n \leqslant c_0 \frac{d\eta^2}{D^2\alpha^4}.$$

假设假设检验问题定义 *6.22*中的低次猜想成立。

那么，任何能稳健地将精确后验均值估计到误差 $\alpha/20$ 以内，且其在每个 $9/10$ 上以至少 η的概率，对来自定义 *6.1* 且带有 $\|w^*\|_2 \leqslant \alpha$ 腐化的干净样本进行估计的多项式时间估计器，都能给出该检验问题的多项式时间区分器。

证明。这可由结合引理 6.24与引理 6.25得出。 □

6.3.1 引理 6.23的证明

引理 *6.23*的证明。令 $\delta = K\alpha$。对于 (i)，定义干净样本

$$x_i^\circ := g_i, \qquad y_i^\circ := \delta\langle g_i, u\rangle + z_i.$$

那么 (X°, Y°)恰好是来自定义 6.1 的一个干净样本，其回归器为$w^\star = \delta u$。在P下，若$B = 0$，我们保持样本不变；若$B = 1$，则将i 处的x°替换为i 处的$x^\circ + a y^\circ{}_i\, u$，同时保持$i$ 处的y°不变。这就证明了 (i)。

对于 (ii)，在Q 下，结论显然成立。在P下，给定u，在两个分支中我们都有

$$y_i = \delta\langle g_i, u\rangle + z_i.$$

由于 $\langle g_i, u\rangle \sim \mathcal{N}(0,1)$ 且与$z_i \sim \mathcal{N}(0,1)$相互独立，因此$y_i \sim \mathcal{N}(0, 1+\alpha^2) = \mathcal{N}(0, s^2)$。对于 (iii)，在$Q$ 下，变量相互独立且中心化，因此 $\mathbb{E}_Q[x_i y_i] = 0$。在$P$下，给定$u$，在干净分支$B = 0$中，

$$\mathbb{E}[x_i y_i \mid u, B=0] = \mathbb{E}\big[x_i(\delta\langle G, u\rangle + Z) \mid u\big] = \delta u.$$

在异常值分支$B = 1$中，

$$\mathbb{E}[x_i y_i \mid u, B=1] = \mathbb{E}[(g_i + a u y_i) y_i \mid u] = \mathbb{E}[g_i y_i \mid u] + a\,\mathbb{E}[y_i^2 \mid u] u = \delta u + a s^2 u.$$

对B ∼服从伯努利分布(q)取平均，

$$\mathop{\mathbb{E}}_{P}[x_i y_i \mid u] = (1-q)\delta u + q(\delta + a s^2)u = \big(\delta + q a s^2\big)u = 0,$$

因为$a = -\delta/(q s^2)$。

对于 (iv)，令 φ_σ 表示 $\mathcal{N}(0, \sigma^2)$的密度，并定义

$$r_0(y) := \frac{\varphi_s(y) - (1-\eta)\varphi_1(y)}{\eta}.$$

由于

$$\frac{\varphi_s(y)}{\varphi_1(y)} = \frac{1}{s}\exp\Big(\frac{\alpha^2}{2(1+\alpha^2)} y^2\Big),$$

该比值在$y = 0$处达到最小，因此

$$\inf_y \frac{\varphi_s(y)}{\varphi_1(y)} = \frac{1}{s}.$$

此外，

$$1 - \frac{1}{\sqrt{1+t}} \leqslant \frac{t}{2} \qquad (t \geqslant 0),$$

因此

$$1 - \frac{1}{s} \leqslant \frac{\alpha^2}{2} \leqslant \eta.$$

因此

$$\varphi_s(y) \geqslant (1-\eta)\varphi_1(y) \qquad \text{for all } y,$$

所以$r_0(y) \geqslant 0$。此外，

$$\int_{\mathbb{R}} r_0(y)\,dy = \frac{1 - (1-\eta)}{\eta} = 1,$$

所以r_0 是一个有效的密度。

现在从一个干净的$w^\star = 0$样本$x^\circ{}_i = g_i$和$y^\circ{}_i = z_i$开始，其中$g_i \sim \mathcal{N}(0, \mathrm{Id}_d)$，且$z_i \sim \mathcal{N}(0,1)$相互独立。保持$x^\circ{}_i$不变。以 $1-\eta$的概率，保留$y^\circ{}_i$；以 η的概率，将其替换为独立抽样$R \sim r_0$。那么观测响应的密度为

$$(1-\eta)\varphi_1 + \eta r_0 = \varphi_s,$$

so $y_i \sim \mathcal{N}(0, s^2)$，因此得到的观测对 (x_i, y_i) 恰好服从 Q. □

6.3.2 将假设检验化简为稳健后验均值估计。

本节中，我们给出引理 6.24的证明

引理 引理 *6.24* 的证明。设

$$\lambda := \Big(n + \frac{1}{\sigma^2}\Big)^{-1}, \qquad \nu := n\lambda = \frac{n\sigma^2}{1+n\sigma^2}.$$

由于 $\sigma^2 = \Theta(1/n)$，我们有 $\nu = \Theta(1)$。

我们首先考虑零假设分布下的情形。根据引理 6.23，Q 可以表示为一个干净的$w^\star = 0$样本 $(X^\circ、y^\circ)$的 η-污染，其中$X^\circ_i = G_i$，且$y^\circ_i = Z_i$。令$C_i \sim$伯努利分布(η)表示被污染响应变量的指示函数。于是，由切尔诺夫界可得

$$\mathbb{P}\Big[\sum_{i=1}^n C_i \leqslant \eta n\Big] \geqslant 1 - \exp(-\Omega(\eta n)).$$

在此事件下， 适用于 $\mathscr{A}$ 的稳健保证。此外，根据引理 6.9以及式（32）中的第一项 ，

$$\big\|\mathbb{E}[w^* \mid X^\circ, y^\circ] - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \leqslant \frac{\alpha}{20}$$

除以概率至多为 β 的情况外。因此，除以概率 $O(\beta)$ 的情况外，

$$\big\|\widehat{w} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \leqslant \frac{\alpha}{10}. \tag{33}$$

现在比较观测到的简化统计量与干净简化统计量。由于只有响应变量被修改，

$$w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ) = \lambda \sum_{i=1}^n X_i^\circ \Delta_i, \qquad \Delta_i := C_i(R_i - Z_i).$$

替换密度 r_0 的方差为

$$\mathrm{Var}(R_i) = \frac{s^2 - (1-\eta)}{\eta} = 1 + \frac{K^2\alpha^2}{\eta} \leqslant O(1),$$

因此 Δ_i 以 0 为中心，且为具有方差代理 $O(\eta)$ 的次高斯变量。于是

$$\sum_{i=1}^n \Delta_i^2 \leqslant C\eta n$$

以至少 $1 - \exp(-\Omega(\eta n))$ 的概率成立。在给定 Δ_i 的条件下，向量 $\sum_i X^\circ{}_i \Delta_i$ 服从协方差为 $(\sum_i \Delta^2{}_i)$ Id_d 的高斯分布。因此，以至少 $1 - \beta - \exp(-\Omega(\eta n))$ 的概率，

$$\big\|w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \leqslant C\lambda\sqrt{\eta n(d + \log(1/\beta))} \leqslant C\sqrt{\frac{\eta(d+\log(1/\beta))}{n}} \leqslant \alpha,$$

其中，最后一步使用了（32）中的第二项。结合（33），

$$\|\widehat{w} - w_{\mathrm{obs}}\|_2 \leqslant \frac{11}{10}\alpha < 2\alpha$$

除了概率为 $O(\beta) + \exp(-\Omega(\eta n))$ 的情况外。因此，该检验输出 零假设。

接下来我们考虑植入分布。根据构造，植入分布是通过先抽取一个干净样本来实现的

$$X_i^\circ = G_i, \qquad y_i^\circ = \alpha\langle G_i, u\rangle + Z_i$$

with 回归器 $w^\star = \alpha u$，然后抽取独立同分布指示变量 $B_i \sim$ 伯努利分布(q)，最后令

$$(X_i, y_i) = \begin{cases} (X_i^\circ, y_i^\circ), & B_i = 0, \\ (X_i^\circ + a y_i^\circ u,\ y_i^\circ), & B_i = 1. \end{cases}$$

再次

$$\mathbb{P}\Big[\sum_{i=1}^n B_i \leqslant \eta n\Big] \geqslant 1 - \exp(-\Omega(\eta n)).$$

在此事件下， $\mathscr{A}$ 的稳健保证适用，并由引理 6.9 结合（32）中的第一项可得，

$$\big\|\widehat{w} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \leqslant \frac{\alpha}{10} \tag{34}$$

除概率为$O(\beta)$的情形外。

现在比较观测到的简化统计量与干净简化统计量。由于只有X 发生变化，

$$w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ) = \lambda \sum_{i=1}^n B_i\, a\, (y_i^\circ)^2\, u.$$

在给定u的条件下取期望，

$$\mathbb{E}[w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ) \mid u] = \lambda n \eta a s^2 u = -\nu K \alpha\, u.$$

此外，$y^\circ{}_i \sim \mathcal{N}(0, s^2)$，因此由伯恩斯坦不等式可得

$$\left|\sum_{i=1}^n B_i\big((y_i^\circ)^2 - s^2\big)\right| \leqslant C s^2 \sqrt{\eta n \log(1/\beta)}$$

其成立的概率至少为 $1 - \beta - \exp(-\Omega(\eta n))$。因此，

$$\left\|\Big(w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ)\Big) + \nu\alpha u\right\|_2 \leqslant C\lambda|a|s^2\sqrt{\eta n \log(1/\beta)} = C\,\nu K \alpha \sqrt{\frac{\log(1/\beta)}{\eta n}}.$$

由（32）中的第三项可知，右端至多为 $\nu K\alpha/4$。因此，除概率为 $\beta + \exp(-\Omega(\eta n))$ 的情况外，

$$\big\|w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \geqslant \frac{3\nu K}{4}\alpha.$$

结合（34），

$$\|\widehat{w} - w_{\mathrm{obs}}\|_2 \geqslant \big\|w_{\mathrm{obs}} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 - \big\|\widehat{w} - w_{\mathrm{simp}}(X^\circ, y^\circ)\big\|_2 \geqslant \frac{3\nu K}{4}\alpha - \frac{\alpha}{10}.$$

由于 $\nu = \Theta(1)$，充分增大 K 可得

$$\frac{3\nu K}{4}\alpha - \frac{\alpha}{10} \geqslant 2\alpha.$$

因此，该检验输出植入分布，除非发生概率为$O(\beta) + \exp(-\Omega(\eta n))$的情况。 □

6.3.3 引理 6.25 的证明

当零假设分布是标准高斯分布时，更容易证明低次数下界。我们定义固定线性映射

$$T:\mathbb{R}^{d+1}\to\mathbb{R}^{d+1},\qquad T(x,y):=(x,y/s).$$

由于T 是可逆线性映射，且保持多项式总次数不变，因此逐样本应用T 不会改变次数为D 时的优势。下文将使用经T变换后的数据，此时似然比可展开如下。

引理 6.27。令 $\theta:=K^{\alpha}s_{\circ}$ 在 T 下，零假设变为 $\mathbb{R}^{d+1}$: $T_{\#}Q=\mathcal{N}(0,\mathrm{Id}_{d+1})$。在植入分布下，给定 u，若记 $t:=\langle x,u\rangle$,$y:=Y/s$, 则相对于白化后的零假设分布，单样本似然比可展开为

$$\ell_u(x,y)=1+\sum_{k\geqslant 1}h_k(\langle x,u\rangle)\,\psi_k(y), \tag{35}$$

其中

$$\psi_k(y)=\theta^k\Big((1-q)h_k(y)+\eta\,h_k\Big(\Big(1-\frac{1}{\eta}\Big)y\Big)\Big). \tag{36}$$

特别地， $\psi_1(y)\equiv 0$。

证明。在零假设下，$X\sim\mathcal{N}(0,\mathrm{Id}_d)$和$Y/s\sim\mathcal{N}(0,1)$相互独立，因此$T_{\#}Q=\mathcal{N}(0,\mathrm{Id}_{d+1})$。令$\delta=K\alpha$。

现在以植入方向 u 为条件，并令 $g:=\langle G,u\rangle$ 。那么

$$y=\frac{Y}{s}=\frac{\delta g+Z}{s}=\theta g+\sqrt{1-\theta^2}\,\xi,$$

其中 $\xi\sim\mathcal{N}(0,1)$。因此 $y\sim\mathcal{N}(0,1)$。

在干净分支 $B=0$ 中，我们有 $x=G$，因此 $t=g$。于是

$$t\mid y,B=0\sim\mathcal{N}(\theta y,\,1-\theta^2).$$

在异常值分支 $B=1$ 中，我们有

$$x=G+aYu=G+as\,y\,u,$$

so

$$t=g+by,\qquad b:=as=-\frac{\delta}{\eta s}=-\frac{\theta}{\eta}.$$

因此

$$t\mid y,B=1\sim\mathcal{N}\Big(\theta\Big(1-\frac{1}{\eta}\Big)y,\,1-\theta^2\Big).$$

根据事实 4.7，若 $T\sim\mathcal{N}(\rho x,\ 1-\rho^2)$，则

$$\mathbb{E}[h_k(T)]=\rho^k h_k(x).$$

将此结果应用于两个分支中的 $\rho=\theta$，可得

$$\mathbb{E}[h_k(t)\mid y,B=0]=\theta^k h_k(y),\qquad \mathbb{E}[h_k(t)\mid y,B=1]=\theta^k h_k\Big(\Big(1-\frac{1}{\eta}\Big)y\Big).$$

对 B ~伯努利分布(η) 取平均可得（36），由此立即得到

$$\psi_1(y)=\theta\Big((1-\eta)h_1(y)+\eta h_1\Big(\Big(1-\frac{1}{\eta}\Big)y\Big)\Big)=0\,.$$ □

为证明 引理 6.25，我们先证明以下两个辅助引理。

Le引理 6.28。 设 $Y \sim \mathcal{N}(0,1)$，设 $|c| \geqslant 1$，并设 $k \geqslant 0$。则 $\mathbb{E}[h_k(cY)^2] \leqslant 2(4c^2)^k$。

证明。 由 Mehler 恒等式，对每个 $|r| < 1$，

$$\sum_{k\geqslant 0} r^k h_k(x)^2 = \frac{1}{\sqrt{1-r^2}} \exp\Big(\frac{rx^2}{1+r}\Big).$$

代入$x = cY$并对$Y \sim \mathcal{N}(0,1)$取期望，得到

$$\sum_{k\geqslant 0} r^k \,\mathbb{E}[h_k(cY)^2] = \frac{1}{\sqrt{1-r^2}} \cdot \frac{1}{\sqrt{1-\frac{2rc^2}{1+r}}}.$$

Set

$$r = \frac{1}{4c^2}.$$

由于 $|c| \geqslant 1$，我们有 $0 < r < 1$，且

$$1 + r - 2rc^2 = 1 + \frac{1}{4c^2} - \frac{1}{2} \geqslant \frac{1}{2}, \qquad 1 - r \geqslant \frac{3}{4}.$$

因此，右侧至多为 2。由于所有系数均非负，

$$r^k \,\mathbb{E}[h_k(cY)^2] \leqslant 2,$$

因此

$$\mathbb{E}[h_k(cY)^2] \leqslant 2r^{-k} = 2(4c^2)^k.$$ □

引理 6.29。 设 $a_k := \|\psi_k\|^2_{L_2(\mathcal{N}(0,1))}$。 则 $a_1 = 0$， 且对每个 $k \geqslant 2$，$a_k \leqslant \eta^2 \Big(\frac{C\theta^2}{\eta^2}\Big)^k$，其中 $C > 0$为某个普适常数。

证明。 恒等式 $a_1 = 0$ 由 $\psi_1 \equiv 0$得出。对于 $k \geqslant 2$，由（36），

$$a_k \leqslant 2\theta^{2k}\Big((1-\eta)^2 \|h_k\|_2^2 + \eta^2 \,\mathbb{E}\Big[h_k\Big(\Big(1-\frac{1}{\eta}\Big)Y\Big)^2\Big]\Big).$$

由于 $\|h_k\|_2 = 1$ 且 $|1 - 1/\eta| \geqslant 1$， 引理 6.28 给出

$$a_k \leqslant 2\theta^{2k}\Big(1 + 2\eta^2\Big(4\Big(1-\frac{1}{\eta}\Big)^2\Big)^k\Big).$$

由于我们有 $|1 - 1/\eta| \leqslant 1/\eta$，因此 $a_k \leqslant \eta^2\Big(\frac{C\theta^2}{\eta^2}\Big)^k$，调整常数 C 后即得。 □

引理 引理 *6.25*的证明。 根据 引理 6.27，只需分析经 T缩放后的问题。对于 $k \geqslant 2$，定义

$$\phi_{u,k}(x,y) := h_k(\langle x,u\rangle)\psi_k(y).$$

于是，单样本似然比为

$$\ell_u(x,y) = 1 + \sum_{k\geqslant 2} \phi_{u,k}(x,y).$$

令u、v $\in \mathbb{S}^{d-1}$，并令 $\gamma := \langle u、v\rangle$ 。在白化后的零假设分布下，x 和y 是相互独立的标准高斯变量，因此根据正交性，

$$\langle \phi_{u,k}, \phi_{v,\ell}\rangle_{L_2(Q)} = \mathbf{1}_{\{k=\ell\}}\gamma^k a_k. \tag{37}$$

此外，对于相互独立且服从$u,v \sim$均匀分布($\mathbb{S}^{d-1}$)的变量，

$$\mathbb{E}[|\gamma|^m] \leqslant \Big(\frac{Cm}{d}\Big)^{m/2} \tag{38}$$

对所有整数 $m \geqslant 1$。令 $K := \lfloor D/2 \rfloor$。由于每个 $\phi_{u,k}$ 的总次数为 $2k$，n-样本 D-投影恰好保留所选隐藏次数之和不超过 K 的项。展开n-样本似然比，并利用（37），得到

$$\mathrm{Adv}^2_{\leqslant D} \leqslant 1 + \sum_{\ell \geqslant 1}\binom{n}{\ell} \sum_{\substack{k_1,\dots,k_\ell \geqslant 2 \\ k_1+\cdots+k_\ell \leqslant K}} \Big(\prod_{j=1}^{\ell} a_{k_j}\Big)\mathbb{E}\big[|\gamma|^{k_1+\cdots+k_\ell}\big]. \tag{39}$$

令 $S := n\alpha^4/d\eta^2$。由于 $\alpha^2 \leqslant \eta$，我们有 $s = \Theta(1)$，因此 $\theta = \Theta(\alpha)$。故由引理 6.29可得

$$a_k \leqslant \eta^2\Big(\frac{C\delta^2}{\eta^2}\Big)^k \qquad (k \geqslant 2). \tag{40}$$

我们将（39）中的求和拆分为 $\ell = 1$ 部分和 $\ell \geqslant 2$ 部分。

我们先对满足 $\ell = 1$ 的项的贡献进行上界估计。对于 $m \geqslant 2$， $\ell = 1$ 项为 $na_m\,\mathbb{E}[|\gamma|^m]$。对于 $m = 2$，利用（40）和（38），$na_2\,\mathbb{E}[\gamma^2] \leqslant C\,n\alpha^4 \cancel{d\eta^2} = CS$。

现在固定$m \geqslant 4$。利用（40）和（38），

$$\begin{aligned} na_m\,\mathbb{E}[|\gamma|^m] &\leqslant n\eta^2\Big(\frac{C\delta^2}{\eta^2}\Big)^m\Big(\frac{Cm}{d}\Big)^{m/2} \\ &= S\cdot\Big(\frac{\delta}{\eta}\Big)^{2m-4}\cdot(Cm)^{m/2}\,d^{1-m/2}. \end{aligned}$$

当$m \leqslant K \leqslant d^{10^{-4}}$时，有 $(\delta/\eta)^4 \cdot m/d \leqslant O(1)$。因此

$$na_m\,\mathbb{E}[|\gamma|^m] \leqslant S\cdot\Big(\Big(\frac{\delta}{q}\Big)^4\cdot\frac{m}{d}\Big)^{m/2-1} \leqslant S\cdot d^{-0.001}.$$

对于每个$m \geqslant 4$。苏 对$m = 4$，……，K求和，我们得出整个 $\ell = 1$ 贡献为$O(S)$。

对于满足 $\ell \geqslant 2$ 的各项所产生的贡献。固定$m \in \{4, \dots\dots, K\}$。将$m$拆分为 ℓ 个部分且每个部分至少为 2 的拆分数为

$$\binom{m-\ell-1}{\ell-1} \leqslant 2^m.$$

利用这一点，结合（40）和（38），所有总隐藏次数为 m 且$\ell \geqslant 2$ 的项的贡献至多为

$$\begin{aligned} &\mathbb{E}[|\gamma|^m]\sum_{\ell=2}^{\lfloor m/2\rfloor}\binom{n}{\ell}\binom{m-\ell-1}{\ell-1}\eta^{2\ell}\Big(\frac{C\delta^2}{\eta^2}\Big)^m \\ &\leqslant \Big(\frac{Cm}{d}\Big)^{m/2}\cdot 2^m\sum_{\ell=2}^{\lfloor m/2\rfloor}(n\eta^2)^\ell\Big(\frac{C\delta^2}{\eta^2}\Big)^m. \end{aligned}$$

由于 $n\eta^2 \geqslant 1$ 处于所关注的情形中，且 $\ell \leqslant m/2$, $\sum_{\ell=2}^{\lfloor m/2 \rfloor} (n\eta^2)^\ell \leqslant m(n\eta^2)^{m/2}$。因此，总的 $\ell \geqslant 2$贡献有界于

$$m(Cm)^m \Big(\frac{n\delta^4}{d\eta^2}\Big)^{m/2} = m(Cm)^m S^{m/2}.$$

若 $S \leqslant c/K^2$ 其中普适常数 $c > 0$足够小，则

$$m(Cm)^m S^{m/2} \leqslant 2^{-m} \qquad \text{for all } m \leqslant K.$$

因此，对所有 $m \geqslant 4$ 和 $\ell \geqslant 2$ 求和后得到 $O(1)$。结合 $\ell = 1$ 和 $\ell \geqslant 2$ 的估计值，我们得到 $\mathrm{Adv}_{\leqslant D}^2 \leqslant 1 + O(S) + O(1)$。若 $n \leqslant c\, d\eta^2 D^2 \delta^4$, ~~,~~ 则 $S \leqslant c/D^2$, , 因此 $\mathrm{Adv}_{\leqslant D} = O(1)$。 □

6.4 私有贝叶斯回归的困难性

现在，我们可以应用 [GH22] 的隐私到鲁棒性归约，从而得到私有贝叶斯回归中所声称的信息—计算差距，定理 6.4。

定理 (定理 6.4的重述)。设 $\varepsilon \leqslant O(1)$ 且 $\alpha \leqslant \min\{\eta d^{1/4}/D, \eta^2\}$。存在一个判别问题 Π *(参见定义 6.22)*，它是 D 次困难的；若存在一个多项式时间的 ε-*DP* 算法，用于具有标准高斯设计和 $\mathcal{N}(0, \sigma^2 \mathrm{Id})$ 先验分布的贝叶斯线性回归，并能估计后验均值，误差不超过 α，且概率至少为 $1-\beta$，使用

$$n = o\Big(\frac{d^{1/3}\log^{2/3}(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3} \cdot D^{1/3}}\Big)$$

样本，那么将存在一个针对 Π 的高效判别算法。

注释 6.30。 在低次框架下，我们给出了误差率的计算阈值 $\alpha_{\mathrm{comp}} = \Big(\frac{d\log^2(1/\beta)}{3\varepsilon^2}\Big)^{1/4} n$。相比之下，通过 ~~*[GH22]*~~ 的黑盒隐私归约，再结合我们的指数时间鲁棒算法，我们知道指数时间算法可以达到误差率 $\alpha \leqslant \tilde{O}\Big(\frac{d+\log(1/\beta)}{\varepsilon}\Big) n$。因此，指数时间算法可达到的误差为 $\alpha_{\mathrm{IT}} = \tilde{O}(d/n\varepsilon)$。结果是，~~即使对于~~ 恒定的失败概率 0.001，当 $\varepsilon^2 \gtrsim d^3/n$ 时，α_{comp} 与 α_{IT} 之间仍存在差距。

定理 *6.4* 的证明。 我们通过反证法，并应用定理 4.21中的隐私到鲁棒性归约。

令 $\mu_{\mathrm{post}} := \mathbb{E}[w^* \mid X^\circ, y^\circ]$ 表示给定模型下贝叶斯线性回归的后验均值。假设存在一个多项式时间 ε-DP 算法 $\mathscr{A}$，该算法在给定n 个独立同分布样本后，输出一个满足以下条件的估计值 θ

$$\mathbb{P}\big[\, \|\widehat{\theta} - \mu_{\mathrm{post}}\| \leqslant O(\alpha) \,\big] \geqslant 1 - \beta.$$

将定理 4.21 应用于目标估计量 μ_{post}，可得一个多项式时间 η-稳健算法 $\mathscr{A}_{\mathrm{rob}}$，它针对比例为 η 的污染，能够以至少 $1-\beta$ 的概率达到误差 α，其中$\eta = O(\log(1/\beta)/n\varepsilon)$。根据引理 6.24，如果存在一个高效的 η-稳健回归算法，能够使用$n = \tilde{O}\Big(\frac{d\,\eta^2}{\alpha^4} D\Big)$个样本达到误差 α，那么就能用同样多的样本高效地解决定义 6.22 中的判别问题 Π。然而，引理 6.25 表明，任何次数为D 的估计量都无法用这么多样本做到这一点。因此，对于任何此类高效的 ε-DP 算法 $\mathscr{A}$，必有$n = \tilde{\Omega}\Big(\frac{d\,\eta^2}{\alpha^4} D\Big)$。

代入 $\eta = O(\log(1/\beta)/(n\varepsilon))$ 并整理可得 $n = \tilde{\Omega}\big(\frac{d}{\alpha^4 D} \cdot \log^2(1/\beta)\, \frac{}{n^2\varepsilon^2}\big)$, 由此即可得出所述下界。

□

致谢

作者感谢 Cynthia Dwork、Pranay Tankala 和 Linjun Zhang 在本项目早期进行的有益交流，感谢 Sam Hopkins、Gautam Kamath、Adam Smith 和 Salil Vadhan 富有见地的讨论，以及他们对本稿先前版本提出的慷慨反馈。

参考文献

[ABMT25] Prashanti Anderson、Ainesh Bakshi、Mahbod Majid 和 Stefan Tiegel，《多项式时间内样本最优的隐私回归》，《第57届 ACM 计算理论年会论文集》，2025年，第2341–2349页。[ABT24] Jason M Altschuler、Jinho Bok 和 Kunal Talwar，《凸优化中噪声随机梯度下降的隐私》，《SIAM计算杂志》53（2024年），第4期，969–1001。[AHS+25] Sam Abbott、Joel Hellewell、Katharine Sherratt、Katelyn Gostic、Joe Hickson、Hamada S. Badr、Michael DeWitt、James M. Azam、EpiForecasts 和 Sebastian Funk，《EpiNow2：估计实时病例数和时变流行病学参数》，2025年，R 软件包版本 1.7.1，https://epiforecasts.io/EpiNow2/dev/，https://github.com/epiforecasts/EpiNow2。[ALNP23] Martin Aumuller、Christian Janos Lebeda、Boel Nelson 和 R. Pagh，《Plan：方差感知隐私均值估计》，arXiv abs/**2306**.08745（2023年）。[AT22] Jason Altschuler 和 Kunal Talwar，《噪声随机梯度下降的隐私：迭代次数增加而隐私损失不变》，《神经信息处理系统进展》35（2022年），3788–3800。[AUZ23] Hilal Asi、Jonathan Ullman 和 Lydia Zakynthinou，《从鲁棒性到隐私，再回到鲁棒性》，《国际机器学习会议》，PMLR，2023年，第1121–1146页。[BBH+20] Matthew Brennan、Guy Bresler、Samuel B Hopkins、Jerry Li 和 Tselil Schramm，《统计查询算法与低次检验几乎等价》，arXiv 预印本 arXiv:2009.06107（2020年）。[BEAH+22] Afonso S Bandeira、Ahmed El Alaoui、Samuel Hopkins、Tselil Schramm、Alexander S Wein 和 Ilias Zadik，《高维统计中的 Franz-Parisi 判据与计算权衡》，《神经信息处理系统进展》35（2022年），33831–33844。[BGS+21] Gavin Brown、Marco Gaboardi、Adam Smith、Jonathan Ullman 和 Lydia Zakynthinou，《无需私有协方差估计的协方差感知隐私均值估计》，《神经信息处理系统进展》（M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan 编），第34卷，Curran Associates, Inc.，2021年，第7950–7964页。[BHJK25] Rares-Darius Buhai、Jun-Ting Hsieh、Aayush Jain 和 Pravesh K Kothari，《拟多项式低次猜想不成立》，arXiv 预印本 arXiv:2505.17360（2025年）。[BHK+19] Boaz Barak、Samuel Hopkins、Jonathan Kelner、Pravesh K Kothari、Ankur Moitra 和 Aaron Potechin，《植入团问题的近乎紧致平方和下界》，《SIAM计算杂志》48（2019年），第2期，687–735。

[BHS23] Gavin Brown、Samuel Hopkins 和 Adam Smith，《面向次高斯分布的快速、样本高效、仿射不变的私有均值与协方差估计》，《第三十六届学习理论会议论文集》（Gergely Neu 和 Lorenzo Rosasco 编），《机器学习研究论文集》，第195卷，PMLR，2023年7月12日至15日，第5578–5579页。[BLR08] Avrim Blum、Katrina Ligett 和 Aaron Roth，《非交互式数据库隐私的学习理论方法》，《第四十届 ACM 计算理论年度研讨会论文集》（美国纽约州纽约市），STOC '08，计算机协会，2008年，第609–618页。[BP21] Ainesh Bakshi 和 Adarsh Prasad，《稳健线性回归：多项式时间内的最优速率》，《第53届 ACM SIGACT 计算理论年度研讨会论文集》，2021年，第102–115页。[BS14] Boaz Barak 和 David Steurer，《平方和证明与寻求最优算法》，arXiv 预印本 arXiv:1404.5236（2014年）。[BS18] Garrett Bernstein 和 Daniel R Sheldon，《指数族的差分隐私贝叶斯推断》，《神经信息处理系统进展》31（2018年）。[BS19] ，《差分隐私贝叶斯线性回归》，《神经信息处理系统进展》32（2019年）。[BSA24] Jinho Bok、Weijie J Su 和 Jason Altschuler，《差分隐私的移位插值》，《第41届国际机器学习会议》，2024年，第4230–4266页。[BST14] Raef Bassily、Adam Smith 和 Abhradeep Thakurta，《重新审视私有经验风险最小化》，06/2014 2014 (英文)。[CCAd+23] Hongjie Chen、Vincent Cohen-Addad、Tommaso d'Orsi、Alessandro Epasto、Jacob Imola、David Steurer 和 Stefan Tiegel，《随机块模型与混合模型的隐私估计算法》，2023年。[CD20] Hee Cheol Chung 和 Gauri Sankar Datta，《用于小区域估计的贝叶斯层级空间模型》，《Statistics》（2020年），07。[CKMY22a] Sitan Chen、Frederic Koehler、Ankur Moitra 和 Morris Yau，《具有对抗性污染的卡尔曼滤波》，《第54届 ACM SIGACT 计算理论年度研讨会论文集》，2022年，第832–845页。[CKMY22b] ，《在线且分布无关的鲁棒性：具有 Huber 污染的回归与上下文老虎机》，《2021 IEEE 第62届计算机科学基础年度研讨会》（FOCS），IEEE，2022年，第684–695页。[CYS21] Rishav Chourasia、Jiayuan Ye 和 Reza Shokri，《朗之万扩散与噪声梯度下降的差分隐私动态》，《神经信息处理系统进展》34（2021年），14771–14781。[DHPT25] Ilias Diakonikolas、Samuel B Hopkins、Ankit Pensia 和 Stefan Tiegel，《稀疏奇异值及其应用的 SoS 证书：稳健统计、子空间畸变等》，《第57届 ACM 计算理论年会论文集》，2025年，第1701–1709页。

[DJY+24] Yuval Dagan、Michael Jordan、Xuelin Yang、Lydia Zakynthinou 和 Nikita Zhivotovskiy，《各向异性分布的无维度私有均值估计》，《神经信息处理系统进展》37（2024年），120667–120698。[DK22] Ilias Diakonikolas 和 Daniel Kane，《通过格基约化进行非高斯成分分析》，学习理论会议，PMLR，2022年，第4535–4547页。[DK23] Ilias Diakonikolas 和 Daniel M. Kane，《算法高维稳健统计》，剑桥大学出版社，2023年。[DKK+19] Ilias Diakonikolas、Gautam Kamath、Daniel Kane、Jerry Li、Ankur Moitra 和 Alistair Stewart，《高维稳健估计量：无需面对计算难题》，SIAM计算杂志48（2019年），第2期，742–864。[DKS19a] Ilias Diakonikolas、Weihao Kong 和 Alistair Stewart，《稳健线性回归的高效算法与下界》，第三十届ACM-SIAM离散算法研讨会论文集（美国），SODA '19，工业与应用数学学会，2019年，第2745–2754页。[DKS19b] ，《稳健线性回归的高效算法与下界》，第三十届ACM-SIAM离散算法研讨会论文集，SIAM，2019年，第2745–2754页。[DMNS06a] Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith，《在隐私数据分析中根据敏感度校准噪声》，密码学理论会议，施普林格，2006年，第265–284页。[DMNS06b] ，《在隐私数据分析中根据敏感度校准噪声》，第三届密码学理论会议论文集（柏林，海德堡），TCC'06，施普林格出版社，2006年，第265–284页。[DNPR10] Cynthia Dwork、Moni Naor、Toniann Pitassi 和 Guy N Rothblum，《持续观测下的差分隐私》，第~~四十二~~届ACM计算理论研讨会论文集，2010年，第715–724页。[DNR+09] Cynthia Dwork、Moni Naor、Omer Reingold、Guy N. Rothblum 和 Salil Vadhan，《差分隐私数据发布的复杂性：高效算法与困难性结果》，第四十一届ACM计算理论年度研讨会论文集（美国纽约州纽约市），STOC '09，计算机协会，2009年，第381–390页。[DNZ+17] Christos Dimitrakakis、Blaine Nelson、Zuhe Zhang、Aikaterini Mitrokotsa 和 Benjamin IP Rubinstein，《通过后验采样实现贝叶斯推断的差分隐私》，《机器学习研究杂志》18（2017年），第11期，1–39。[FGWC16] James Foulds、Joseph Geumlek、Max Welling 和 Kamalika Chaudhuri，《保护隐私的贝叶斯数据分析的理论与实践》，第三十二届人工智能不确定性会议论文集，2016年，第192–201页。[FIH79] Robert E Fay III 和 Roger A Herriot，《小区域收入估计：詹姆斯-斯坦程序在人口普查数据中的应用》，《美国统计学会杂志》74（1979年），第366a期，269–277页。

[FMG+20] Seth Flaxman、Swapnil Mishra、Axel Gandy、H Juliette T Unwin、Thomas A Mellan、Helen Coupland、Charles Whittaker、HarrisonZhu、Tresnia Berah、Jeffrey W Eaton 等，《估计非药物干预措施对欧洲新冠肺炎的影响》，《自然》584（2020年），第7820期，257–261页。[FS24] Cecilia Ferrando 和 Daniel Sheldon，《通过数据依赖的充分统计量扰动实现隐私回归》，arXiv预印本 arXiv:2405.15002（2024年）。[GH22] Kristian Georgiev 和 Samuel B. Hopkins，《隐私带来鲁棒性：信息与计算差距及稀疏均值估计》，ArXiv abs/**2211**.00724（2022年）。[GSC17] Joseph Geumlek、Shuang Song 和 Kamalika Chaudhuri，《用于后验采样的Rényi差分隐私机制》，《神经信息处理系统进展》30（2017年）。[HDN+18] Antti Honkela、Mrinal Das、Arttu Nieminen、Onur Dikmen 和 Samuel Kaski，《高效的差分隐私学习提升药物敏感性预测》，《生物学直通》13（2018年），第1期，1。[HJDH19] Mikko Heikkilä、Joonas Jälkö、Onur Dikmen 和 Antti Honkela，《差分隐私马尔可夫链蒙特卡罗》，《神经信息处理系统进展》32（2019年）。[HKM22] Samuel B Hopkins、Gautam Kamath 和 Mahbod Majid，《通过平方和指数机制实现纯差分隐私下的高效均值估计》，第54届 ACM SIGACT 计算理论年度研讨会论文集，2022年，第1406–1417页。[HKMN23] Samuel B Hopkins、Gautam Kamath、Mahbod Majid 和 Shyam Narayanan，《统计估计中的鲁棒性蕴含隐私》，第55届ACM计算理论研讨会论文集，2023年，第497–506页。[HKP+17] Samuel B Hopkins、Pravesh K Kothari、Aaron Potechin、Prasad Raghavendra、Tselil Schramm 和 David Steurer，《平方和方法在检测隐结构方面的威力》，2017年第58届IEEE计算机科学基础年会（FOCS），IEEE，2017年，第720–731页。[HL18] Samuel B Hopkins 和 Jerry Li，《混合模型、鲁棒性与平方和证明》，第50届 ACM SIGACT 计算理论研讨会论文集，2018年，第1021–1034页。[HL24] Samuel B Hopkins 和 Anqi Li，《通过置信传播实现树上的对抗鲁棒推断》，第三十七届学习理论年会，PMLR，2024年，第2389–2417页。[HS17] Samuel B Hopkins 和 David Steurer，《基于少量样本的高效贝叶斯估计：社区检测及相关问题》，2017年第58届IEEE计算机科学基础年会（FOCS），IEEE，2017年，第379–390页。[HT10] Moritz Hardt 和 Kunal Talwar，《差分隐私的几何学》，第四十二届ACM计算理论研讨会论文集（美国纽约州纽约市），STOC '10，计算机协会，2010年，第705–714页。[HW20] Justin Holmgren 和 Alexander S. Wein，《低次猜想的反例》，信息技术融合与服务，2020年。

[IS24] Misha Ivkov 和 Tselil Schramm，半定规划能够稳健地模拟近似消息传递，《第56届 ACM 计算理论年会论文集》，2024年，第348–357页。[IS25] , 快速、稳健的近似消息传递，《第57届 ACM 计算理论年会论文集》，2025年，第2237–2248页。[JAGR22] Nianqiao Ju、Jordan Awan、Ruobin Gong 和 Vinayak Rao，基于隐私化数据进行贝叶斯推断的数据增强马尔可夫链蒙特卡洛，《神经信息处理系统进展》第35卷（2022年），第12732–12743页。[Jäl17] Joonas Jälkö，非共轭模型的差分隐私变分推断，《人工智能》第11卷（2017年），第08页。[KDH23] Rohith Kuditipudi、John Duchi 和 Saminul Haque，一种用于自适应隐私均值估计的相当快速算法，《第三十六届学习理论会议论文集》（Gergely Neu 和 Lorenzo Rosasco 编），机器学习研究论文集，第195卷，PMLR，2023年7月12日至15日，第2511–2551页。[KKM18] Adam Klivans、Pravesh K Kothari 和 Raghu Meka，离群值鲁棒回归的高效算法，《学习理论会议》，PMLR，2018年，第1420–1430页。[KLL+23] Jonathan Kelner、Jerry Li、Allen X Liu、Aaron Sidford 和 Kevin Tian，近线性时间的半随机稀疏恢复，《第三十六届学习理论年度会议》，PMLR，2023年，第2352–2398页。[KLN+08] Shiva Prasad Kasiviswanathan、Homin K. Lee、Kobbi Nissim、Sofya Raskhodnikova 和 Adam Smith，我们能够私密地学习什么？《2008年第49届 IEEE 计算机科学基础年会》，2008年，第531–540页。[KLSU19] Gautam Kamath、Jerry Li、Vikrant Singhal 和 Jonathan Ullman，私密学习高维分布，《第三十二届学习理论会议论文集》（Alina Beygelzimer 和 Daniel Hsu 编），机器学习研究论文集，第99卷，PMLR，2019年6月25日至28日，第1853–1902页。[KMUW17] Lucas Kowalczyk、Tal Malkin、Jonathan Ullman 和 Daniel Wichs，基于单向函数的非交互式差分隐私的困难性，密码学电子印本文库，论文2017/1107，2017年。[KMZ22] Pravesh K Kothari、Peter Manohar 和 Brian Hu Zhang，多项式时间平方和能够以最优方式稳健估计高斯分布的均值和协方差，《国际算法学习理论会议》，PMLR，2022年，第638–667页。[KS14] Shiva P Kasiviswanathan 和 Adam Smith，论差分隐私的“语义”：一种贝叶斯表述，《隐私与保密期刊》第6卷（2014年），第1期。[KS17] Pravesh K Kothari 和 David Steurer，通过平方和进行离群值鲁棒的矩估计，arXiv电子预印本（2017年），arXiv–1711。[KV17] Vishesh Karwa 和 Salil P. Vadhan，有限样本差分隐私置信区间，《信息技术融合与服务》，2017年。

[KWB19] Dmitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira，关于假设检验计算困难性的笔记：利用低次似然比进行预测，ISAAC Congress（国际分析及其应用与计算学会），施普林格，2019年，第1–50页。[KZ24] Florent Krzakala 和 Lenka Zdeborová，优化与机器学习中的统计物理方法，课程讲义（2024年）。[LM22] Allen Liu 和 Ankur Moitra，稳健社区检测的极小极大速率，《2022年 IEEE 第63届计算机科学基础年会》（FOCS），2022年，第823–831页。[LRV16] Kevin A Lai、Anup B Rao 和 Santosh Vempala，无先验的均值和协方差估计，《2016年 IEEE 第57届计算机科学基础年会》（FOCS），IEEE，2016年，第665–674页。[Mir13] Darakhshan J Mir，差分隐私：隐私-效用权衡格局探索，罗格斯新泽西州立大学研究生院，2013年。[MS+24] Andrea Montanari、Subhabrata Sen 等，面向非物理学家的平均场自旋玻璃技术友好教程，《机器学习基础与趋势》第17卷（2024年），第1期，第1–173页。[MW21] Cheng Mao 和 Alexander S Wein，子空间中植入向量的最优谱恢复，arXiv预印本 arXiv:2105.15081（2021年）。[MW25] Andrea Montanari 和 Alexander S Wein，秩一矩阵估计中近似消息传递与低次多项式的等价性，《概率论及相关领域》第191卷（2025年），第1期，第181–233页。[PFCW20] Mijung Park、James Foulds、Kamalika Chaudhuri 和 Max Welling，隐私环境中的变分贝叶斯（VIPS），《人工智能研究杂志》第68卷（2020年），第109–157页。[RSS18] Prasad Raghavendra、Tselil Schramm 和 David Steurer，通过平方和证明进行高维估计，《国际数学家大会论文集：2018 年里约热内卢》，世界科技出版公司，2018年，第3389–3423页。[Sar24] Anand D Sarwate，差分隐私机器学习，《机密数据共享手册》，Chapman and Hall/CRC，2024年，第91–114页。[SCV17] Jacob Steinhardt、Moses Charikar 和 Gregory Valiant，韧性：存在任意离群值时学习的判据，arXiv预印本 arXiv:1703.04940（2017年）。[TF20] Aleksei Triastcyn 和 Boi Faltings，面向机器学习的贝叶斯差分隐私，《国际机器学习会议》，PMLR，2020年，第9583–9592页。[UV20] Jonathan Ullman 和 Salil Vadhan，概率可检验证明与生成合成数据的困难性，《密码学杂志》第33卷（2020年），第4期，第2078–2112页。[VYF+15] Bjarni J Vilhjálmsson、Jian Yang、Hilary K Finucane、Alexander Gusev、Sara Lindström、Stephan Ripke、Giulio Genovese、Po-Ru Loh、Gaurav Bhatia、Ron Do 等，建模连锁不平衡可提高多基因风险评分的准确度，《美国人类遗传学杂志》第97卷（2015年），第4期，第576–592页。

[Wei25] Alexander S Wein, 统计学的计算复杂性：低次多项式带来的新见解，arXiv 预印本 arXiv:2506.10748（2025年）。[WFS15] Yu-Xiang Wang、Stephen Fienberg 和 Alex Smola，免费的隐私保护：后验采样与随机梯度蒙特卡洛，国际机器学习会议，PMLR，2015年，第 2493–2502 页。[WM10] Oliver Williams 和 Frank McSherry，概率推断与差分隐私，《神经信息处理系统进展》**23**（2010年）。[XX12] Yonghui Xiao 和 Li Xiong，差分隐私下的贝叶斯推断，arXiv 预印本 arXiv:1203.0617（2012年）。[Zhe15] Shijie Zheng，贝叶斯推断的差分隐私，学士学位论文，哈佛学院，2015年。[ZK16] Lenka Zdeborová 和 Florent Krzakala，推断的统计物理：阈值与算法，《物理学进展》**65**（2016年），第 5 期，453–552。[ZRD16] Zuhe Zhang、Benjamin Rubinstein 和 Christos Dimitrakakis，论贝叶斯推断的差分隐私，AAAI人工智能会议论文集，第 30 卷，2016年。[ZSWB22] Ilias Zadik、Min Jae Song、Alexander S Wein 和 Joan Bruna，基于格的方法在聚类中优于平方和方法，学习理论会议，PMLR，2022年，第 1247–1248 页。[ZZ23] Wanrong Zhang 和 Ruqi Zhang，*DP-fast MH*：面向大规模贝叶斯推断的私有、快速且准确的梅特*ropolis*–黑斯廷斯算法，国际机器学习会议，PMLR，2023年，第 41847–41860 页。

附录 A 延后证明：来自 第 5 节

A.1 从经验均值到后验均值

本节将补充细节，说明如何从 定理 1.1 （私有后验均值估计）推导出 定理 5.3 （私有经验均值估计）。

定理 *1.1* 中低效上界的证明。 我们将应用 定理 5.3 中的低效算法，其中

$$R = \sqrt{\left(\|\Sigma\|_{\mathrm{op}} + \frac{1}{n}\right)\left(d + \log\frac{1}{\beta}\right)}$$

so，使得前提 $\|\bar{x}\|_2 \leqslant R$ 以概率 $1 - O(\beta)$ 成立。然后，如果我们要求 e

$$n \geqslant \frac{d\log\frac{R}{\alpha}}{\varepsilon} = \tilde{\Theta}\left(\frac{d\log\frac{\|\Sigma\|_{\mathrm{op}}}{\alpha}}{\varepsilon}\right)$$

并应用重缩放归约，我们得到所需的误差

$$\alpha \leqslant \tilde{O}\left(\frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}}\log\frac{1}{\beta}}{\varepsilon n}\right) = \tilde{O}\left(\frac{\mathrm{tr}\left(I + \frac{1}{n}\Sigma^{-1}\right)^{-1} + \left\|\left(I + \frac{1}{n}\Sigma^{-1}\right)^{-1}\right\|_{\mathrm{op}}\log\frac{1}{\beta}}{\varepsilon n}\right) \qquad \square$$

定理 *1.1* 中高效上界的证明定理 *1.1*。 沿用前一个证明的思路，我们使用定理 5.3 中的高效算法，并取相同的R。然后，假设$n \geqslant \tilde{\Theta}\left(d \log_{\varepsilon}^{\|\Sigma\|_{\mathrm{op}}}\right)$，我们有

$$\alpha \leqslant \tilde{O}\left(\left(\frac{\operatorname{tr}(\Lambda^{4/3}) + \|\Lambda^{4/3}\|_{\mathrm{op}} \log \frac{1}{\beta}}{n\varepsilon^{2/3}}\right)^{3/4} + \frac{\operatorname{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}} \log \frac{1}{\beta}}{\varepsilon n}\right)$$

如所需。 □

A.2 通过分桶从各向同性到各向异性

这里我们将从定理 5.3 在第5节中证明的各向同性先验这一特殊情形推导出来。

低效版本的定理 *5.3*的证明。我们的方案是按特征值将 Λ^2 拆分成多个桶，每个桶对应 $\sigma^2{}_i = 2^{-i} \cdot \|\Lambda^2\|_{\mathrm{op}}$，并在前$M+1$ 个桶上运行隐私参数为 ε_{M+1} 的定理 5.3 中的各向同性估计量，然后对其余所有桶输出 ~~0~~。形式化地，令 α_i 表示桶 $\sigma^2{}_i$上的误差，令 $\Lambda^2{}_i$ 表示限制在桶i上的协方差矩阵，令d_i 表示桶i的大小。

$$\alpha \leqslant \sum_{i=0}^{\infty} \alpha_i \leqslant \sum_{i=1}^{M} \alpha_i + 2^{-M/2} \cdot d \cdot \left\|\Lambda^2\right\|_{\mathrm{op}} \cdot \frac{d + \log \frac{1}{\beta}}{\varepsilon n}$$

现在先忽略第二项（稍后将M 设得足够大，使其可以忽略）。利用均方-算术平均不等式，第一项至多为（相差常数因子）

$$\begin{aligned}
\sum_{i=0}^{M} \alpha_i &\leqslant \sum_{i=1}^{M} \sigma_i (M+1) \cdot \frac{d_i + \log \frac{1}{\beta}}{\varepsilon n} \\
&\leqslant (M+1) \sum_{i=1}^{M} \sigma_i \cdot \frac{d_i + \log \frac{1}{\beta}}{\varepsilon n} \\
&\leqslant (M+1) \cdot \frac{\operatorname{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}} \log \frac{1}{\beta}}{\varepsilon n}
\end{aligned}$$

其中我们使用了不等式 $\sum^{M}{}_{i=0} \sigma^2{}_i \leqslant 2\|\Lambda^2\|_{\mathrm{op}}$ 和 $\sum^{M}{}_{i=0} \sigma_i \leqslant 2\|\Lambda\|_{\mathrm{op}}$。最后，为了界定剩余的桶，只需令 $M = \Omega(\mathrm{polylog}(n、d、\frac{1}{\varepsilon}、\log \frac{1}{\beta}))$，即可完成论证。 □

定理 *5.3* 的高效变体的证明定理 *5.3*。 我们采用了与上文高效情形证明相似的方法。与之前类似，

$$\alpha \leqslant \sum_{i=0}^{\infty} \alpha_i \leqslant \sum_{i=1}^{M} \alpha_i + 2^{-M/2} \cdot d \cdot \left\|\Lambda^2\right\|_{\mathrm{op}} \cdot \left(\left(\frac{d + \log \frac{1}{\beta}}{n\varepsilon^{2/3}}\right)^{3/4} + \frac{d + \log \frac{1}{\beta}}{\varepsilon n}\right).$$

我们暂时忽略第二项（稍后将 M 设得足够大，使其可以忽略）。第一项利用 $f(x) = x^{3/4}$ 的凹性，至多为（差一个常数因子）

$$\sum_{i=0}^{M} \alpha_i \leqslant \sum_{i=1}^{M} \left[\sigma_i \left((M+1)^{2/3} \frac{d_i + \log \frac{1}{\beta}}{n\varepsilon^{2/3}}\right)^{3/4} + \sigma_i (M+1) \cdot \frac{d_i + \log \frac{1}{\beta}}{\varepsilon n}\right]$$

$$\leqslant M^{1/4}\sqrt{M+1}\cdot\left(\sum_{i=1}^{M}\sigma_i^{4/3}\cdot\frac{d_i+\log\frac{1}{\beta}}{n\varepsilon^{2/3}}\right)^{3/4}+(M+1)\sum_{i=1}^{M}\sigma_i\cdot\frac{d_i+\log\frac{1}{\beta}}{\varepsilon n}$$
$$\leqslant M^{1/4}\sqrt{M+1}\cdot\left(\frac{\mathrm{tr}(\Lambda^{4/3})+\|\Lambda^{4/3}\|_{\mathrm{op}}\log\frac{1}{\beta}}{n\varepsilon^{2/3}}\right)^{3/4}+(M+1)\cdot\frac{\mathrm{tr}(\Lambda)+\|\Lambda\|_{\mathrm{op}}\log\frac{1}{\beta}}{\varepsilon n}$$

其中我们使用了不等式 $\sum^{M}{}_{i=0}\sigma^{4/3}{}_{i}\leqslant 2\|\Lambda^{4/3}\|_{\mathrm{op}}$ 以及 $\sum^{M}{}_{i=0}\sigma_i\leqslant 2\|\Lambda\|_{\mathrm{op}}$。最后，要界定剩余的桶，只需令 $M=\Omega(\mathrm{polylog}(n、d、\frac{1}{\varepsilon}、\log\frac{1}{\beta}))$，即可完成论证。 □

A.3 引理 5.8 的证明：可行性的韧性界

本节将证明所需的韧性界，以表明真实值对于主程序 $\mathscr{A}$ 第 5.3.1 节中的可行性：

引理 （引理 5.8 的重述）。 设 $x_1,\ldots,x_n$ 是来自 $\mathcal{N}(\mu,\mathrm{Id}_d)$ 的独立同分布样本，且 $\bar{x}$ 是它们的经验均值。则以 $1-\beta$ 的概率，对于所有向量 $v\in\mathbb{S}^{d-1}$ 和 $b\in[0,1]^n$ 满足 $\mathbb{E}_i b_i\geqslant 1-\eta$，都有

$$\left|\mathop{\mathbb{E}}_i b_i\langle x_i-\bar{x},v\rangle\right|\leqslant O\Big(\sqrt{\frac{\eta(d+\log(1/\beta))}{n}}+\eta\sqrt{\log(1/\eta)}\Big)$$
$$\left|\mathop{\mathbb{E}}_i b_i\langle x_i-\bar{x},v\rangle^2-1\right|\leqslant O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}}+\frac{d+\log(1/\beta)}{n}+\eta\log(1/\eta)\Big)$$

证明。固定 $b\in[0,1]^n$ 满足 $\mathbb{E}_i b_i\geqslant 1-\eta$。令 $v\in\mathbb{S}^{d-1}$ 为任意向量。利用 $\bar{x}=\mathbb{E}_i x_i$，我们有

$$\begin{aligned}\mathop{\mathbb{E}}_i b_i\langle x_i-\bar{x},v\rangle&=\mathop{\mathbb{E}}_i b_i\langle x_i,v\rangle-(\mathop{\mathbb{E}}_i b_i)\mathop{\mathbb{E}}_j\langle x_j,v\rangle\\&=\mathop{\mathbb{E}}_i\Big(b_i-(\mathop{\mathbb{E}}_j b_j)\Big)\langle x_i,v\rangle.\end{aligned}$$

将 $x_i=\mu+(x_i-\mu)$ 分解。由于 $\mathbb{E}_i\left(b_i-(\mathbb{E}_j b_j)\right)=0$，$\mu$ 项抵消，因此

$$\mathop{\mathbb{E}}_i b_i\langle x_i-\bar{x},v\rangle=\mathop{\mathbb{E}}_i\Big(b_i-(\mathop{\mathbb{E}}_j b_j)\Big)\langle x_i-\mu,v\rangle.$$

现在注意，$u=\mathbb{E}_i(b_i-(\mathbb{E}_j b_j))(x_i-\mu)$ 服从均值为零、方差为 $\mathbb{E}_i(b_i-(\mathbb{E}_j b_j))^2 Id$ 的高斯分布（按洛温纳序）。现在注意，我们可以写成

$$\begin{aligned}\mathop{\mathbb{E}}_i(b_i-\mathop{\mathbb{E}}_j b_j)^2&=\mathop{\mathbb{E}}_i b_i^2-(\mathop{\mathbb{E}}_i b_i)^2\\&\leqslant 1-(1-\eta)^2\\&=2\eta-\eta^2\\&\leqslant 2\eta\,.\end{aligned}$$

因此，根据方差至多为$O(\eta)Id$、均值为 0 的高斯分布的 ℓ_2 范数的集中性，即u，我们可以得出，以概率至少为 $1-\beta$，对于所有$v\in\mathbb{S}^{d-1}$，都有

$$|\mathop{\mathbb{E}}_i b_i\langle x_i-\bar{x},v\rangle|\leqslant O\Big(\sqrt{\frac{\eta(d+\log(1/\beta))}{n}}\Big).$$

现在，为了将这一结论推广到所有b，只需对所有满足 $\mathbb{E}_i b_i \geqslant 1-\eta$ 的整点$b \in \{0,1\}^n$ 取并集界，因为它们可能是极点。对数量为$\binom{n}{\eta n} \leqslant (e/\eta)^{\eta n} \leqslant \exp(O(\eta \log(1/\eta) n))$ 的这些b取并集界，可得以概率 $1-\beta$，对于所有向量$v \in \mathbb{S}^{d-1}$和$b \in [0,1]^n$，只要满足 $\mathbb{E}_i b_i \geqslant 1-\eta$，就有

$$|\mathop{\mathbb{E}}_i b_i \langle x_i - \bar{x}, v\rangle| \leqslant O\Big(\sqrt{\frac{\eta(d+\log(1/\beta))}{n}} + \eta\sqrt{\log(1/\eta)}\Big).$$

如所需。

现在我们证明引理的第二部分。固定 $b \in \{0,1\}^n$ 满足 $\mathbb{E}_i b_i \geqslant 1-\eta$。我们可以写成

$$\Big|\mathop{\mathbb{E}}_i b_i \langle x_i - \bar{x}, v\rangle^2 - 1\Big| \leqslant \Big|\mathop{\mathbb{E}}_i \langle x_i - \bar{x}, v\rangle^2 - 1\Big| + \mathop{\mathbb{E}}_i (1-b_i)\langle x_i - \bar{x}, v\rangle^2.$$

我们分别对每一项进行界定。对于第一项，我们可以写成

$$\begin{aligned}\Big|\mathop{\mathbb{E}}_i \langle x_i - \bar{x}, v\rangle^2 - 1\Big| &= \Big|\mathop{\mathbb{E}}_i \langle x_i - \mu, v\rangle^2 - \langle \bar{x} - \mu, v\rangle^2 - 1\Big| \\ &\leqslant \Big|\mathop{\mathbb{E}}_i \langle x_i - \mu, v\rangle^2 - 1\Big| + \langle \bar{x} - \mu, v\rangle^2 \\ &\leqslant \Big\|\mathop{\mathbb{E}}_i (x_i-\mu)(x_i-\mu)^\mathsf{T} - I_d\Big\|_{\mathrm{op}} + \|\bar{x}-\mu\|_2^2.\end{aligned}$$

根据标准高斯分布的标准集中界，我们知道以概率 $1-\beta$,

$$\Big\|\mathop{\mathbb{E}}_i (x_i-\mu)(x_i-\mu)^\mathsf{T} - I_d\Big\|_{\mathrm{op}} \leqslant O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \frac{d+\log(1/\beta)}{n}\Big), \quad \|\bar{x}-\mu\|_2^2 \leqslant O\Big(\frac{d+\log(1/\beta)}{n}\Big).$$

因此，以概率 $1-\beta$，对于所有 $v \in \mathbb{S}^{d-1}$，我们得到

$$\Big|\mathop{\mathbb{E}}_i \langle x_i - \bar{x}, v\rangle^2 - 1\Big| \leqslant O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \frac{d+\log(1/\beta)}{n}\Big).$$

现在我们来界定第二项。加上并减去 μ，我们可以写成

$$\mathop{\mathbb{E}}_i (1-b_i)\langle x_i - \bar{x}, v\rangle^2 \leqslant 2\mathop{\mathbb{E}}_i (1-b_i)\langle x_i - \mu, v\rangle^2 + 2\mathop{\mathbb{E}}_i (1-b_i)\langle \mu - \bar{x}, v\rangle^2.$$

我们可以证明，利用高斯分布集中性的标准界以及假设 $\mathbb{E}_i b_i \geqslant 1-\eta$，第二项以概率 $1-\beta$ 至多为 $O\Big(\frac{\eta(d+\log(1/\beta))}{n}\Big)$。接下来需要界定第一项。再次应用经验协方差的标准集中性界，我们可以写出，以概率 $1-\beta$

$$\mathop{\mathbb{E}}_i (1-b_i)\langle x_i - \mu, v\rangle^2 \leqslant \eta \cdot \left[1 + O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \frac{d+\log(1/\beta)}{n}\Big)\right].$$

总之，对于固定的 b，我们可以得到，以概率 $1-\beta$,

$$\Big|\mathop{\mathbb{E}}_i b_i \langle x_i - \bar{x}, v\rangle^2 - 1\Big| \leqslant O\Big(\sqrt{\frac{d+\log(1/\beta)}{n}} + \frac{d+\log(1/\beta)}{n} + \eta\Big).$$

对 β 的选择应用并集界，即可得到所需的界。

□

A.4 各向同性估计引理

这里我们将陈述并证明以下引理，它是 [KMZ22] 中引理 4.1 的一种改编，也是 第 5.3.1 节证明中的关键组成部分。由于我们只需要在 $V(\mu_0, v) = 1$ 的情况下使用该引理，因此这里也在这一设定下陈述并证明。我们将完全沿用他们的策略，唯一的区别是，我们将其改编到这样一种情形：我们不一定能获得带有 $O(\eta)$ 误差的韧性保证，而可能只有较弱的保证，例如在高斯均值估计的 $n \ll d/\eta^2$情形中。

引理 A.1。 令 $x_1, \ldots, x_n \in \mathbb{R}^d$，并固定一个参考均值 $\mu_0 \in \mathbb{R}^d$。假设以下鲁棒矩界成立：对于每个 $v \in S$ 以及每个权重向量 $a = (a_1, \ldots, a_n) \in [0,1]^n$，其中 $\sum_{i=1}^n a_i \geqslant (1-\eta)n$，

$$\left| \frac{1}{n} \sum_{i=1}^n a_i \langle x_i - \mu_0, v \rangle \right| \leqslant \alpha_1, \qquad \left| \frac{1}{n} \sum_{i=1}^n a_i \big(\langle x_i - \mu_0, v \rangle^2 - 1 \big) \right| \leqslant \alpha_2. \tag{41}$$

令 $y_1, \ldots, y_n$ 是 $x_1, \ldots, x_n$的 η-污染版本。令 $\tilde{\mathbb{E}}$ 为形式变量 $x'_1, \ldots, x'_n \in \mathbb{R}^d$ 和 $w_1, \ldots, w_n$上的一个次数足够高的伪期望，并令 $\mu' := \frac{1}{n} \sum_{i=1}^n x'_i$。假设：

1. 对所有 $i \in [n]$，$\tilde{\mathbb{E}}[w_i^2] = w_i$；
2. $\tilde{\mathbb{E}}\left[\sum_{i=1}^n w_i\right] = (1-\eta)n$；
3. 对所有 $i \in [n]$，$\tilde{\mathbb{E}}[w_i x'_i = w_i y_i]$
4. *for every* $v \in S$, $\tilde{\mathbb{E}}\left[\frac{1}{n} \sum_{i=1}^n \langle x'_i - \mu', v \rangle^2 \right] \leqslant 1 + \alpha_0$.

定义 $w'_i := w_i \mathbf{1}\{x_i = y_i\}$，并令 $a_i := \tilde{\mathbb{E}}[w'_i]$，其中 $i \in [n]$。于是 $a_i \in [0,1]$ 以及 $\frac{1}{n} \sum_i a_i \geqslant 1 - 2\eta$。对于每个 $v \in S$，记 $\widehat{\mu} := \tilde{\mathbb{E}}[\mu']$，则有：

$$\tilde{\mathbb{E}}\big[\langle \mu' - \mu_0, v \rangle^2 \big] \leqslant \frac{4\eta(\alpha_0 + \alpha_2) + 2\alpha_1^2}{1 - 4\eta} \leqslant 8\eta(\alpha_0 + \alpha_2) + 4\alpha_1^2 \quad (\eta \leqslant \tfrac{1}{8}), \tag{42}$$

$$\left| \langle \widehat{\mu} - \mu_0, v \rangle \right| \leqslant \alpha_1 + \sqrt{2\eta \Big(\alpha_0 + \alpha_2 + \tilde{\mathbb{E}}\big[\langle \mu' - \mu_0, v \rangle^2 \big] \Big)}. \tag{43}$$

特别地，结合(42)和(43)，可得界

$$\left| \langle \widehat{\mu} - \mu_0, v \rangle \right| \leqslant \alpha_1 + \sqrt{C_1 \eta(\alpha_0 + \alpha_2) + C_2 \eta \alpha_1^2 + C_3 \eta^2 (\alpha_0 + \alpha_2)} \leqslant C\Big(\alpha_1 + \sqrt{\eta(\alpha_0 + \alpha_2)} + \eta \Big), \tag{44}$$

对于通用常数 $C_1, C_2, C_3, C > 0$。因此，该速率

$$\alpha_1 + \sqrt{\eta(\alpha_0 + \alpha_2) + \eta^2}$$

在通用常数因子以内成立。

引理 A.1 的证明。定义 $w'_i := w_i \mathbf{1}\{x_i = y_i\}$。则 $\tilde{\mathbb{E}}[w'^2_i = w'_i]$，$\tilde{\mathbb{E}}[w'_i x'_i = w'_i x_i]$，且 $\frac{1}{n} \sum_{i=1}^n \tilde{\mathbb{E}}[1 - w'_i] \leqslant 2\eta$。令 $a_i := \tilde{\mathbb{E}}[w'_i]$；则 $a_i \in [0,1]$ 以及 $\frac{1}{n} \sum_i a_i \geqslant 1 - 2\eta$。

步骤 *1*：分解估计误差。对于任意$v \in S$,

$$\langle \widehat{\mu} - \mu_0, v \rangle = \tilde{\mathbb{E}}\left[\frac{1}{n} \sum_{i=1}^n \langle x'_i - \mu_0, v \rangle \right]$$

$$= \tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle\right] + \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbb{E}}[w_i']\,\langle x_i-\mu_0,\ v\rangle.$$

因此，

$$\left|\langle\widehat{\mu}-\mu_0,\ v\rangle\right| \leqslant \underbrace{\left|\tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle\right]\right|}_{\text{(A)}} + \underbrace{\left|\frac{1}{n}\sum_{i=1}^{n}a_i\langle x_i-\mu_0,\ v\rangle\right|}_{\leqslant\ \alpha_1\ \text{ by (41)}}. \tag{45}$$

对 (A) 和 $w'_i{}^2 = w'_i$ 应用柯西-施瓦茨不等式，得到

$$\begin{aligned}(\text{A})^2 &\leqslant \tilde{\mathbb{E}}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle\right)^2\right] \leqslant \tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\right]\cdot\tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle^2\right]\\ &\leqslant 2\eta\cdot\tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle^2\right].\end{aligned} \tag{46}$$

代入 $w'_i x'_i = w'_i x_i$，

$$\begin{aligned}\tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle^2\right] &= \tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}\langle x_i'-\mu_0,\ v\rangle^2\right] - \frac{1}{n}\sum_{i=1}^{n}a_i\langle x_i-\mu_0,\ v\rangle^2\\ &\leqslant \underbrace{\tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^{n}\langle x_i'-\mu',\ v\rangle^2\right]}_{\leqslant\ 1+\alpha_0} + \tilde{\mathbb{E}}\left[\langle\mu'-\mu_0,\ v\rangle^2\right] - \underbrace{\frac{1}{n}\sum_{i=1}^{n}a_i\langle x_i-\mu_0,\ v\rangle^2}_{\geqslant\ 1-\alpha_2}\\ &\leqslant \alpha_0+\alpha_2 + \tilde{\mathbb{E}}\left[\langle\mu'-\mu_0,\ v\rangle^2\right].\end{aligned} \tag{47}$$

结合 (46) 和 (47)，

$$(\text{A})^2 \leqslant 2\eta\Big(\alpha_0+\alpha_2+\tilde{\mathbb{E}}\left[\langle\mu'-\mu_0,\ v\rangle^2\right]\Big). \tag{48}$$

将 (48) 代入 (45)，得到 (43)。

现在我们对 $\tilde{\mathbb{E}}[\langle\mu'-\mu_0, v\rangle^2]$ 作出界。分解

$$\langle\mu'-\mu_0,\ v\rangle = \frac{1}{n}\sum_{i=1}^{n}(1-w_i')\langle x_i'-\mu_0,\ v\rangle + \frac{1}{n}\sum_{i=1}^{n}a_i\langle x_i-\mu_0,\ v\rangle.$$

根据 $(x+y)^2 \leqslant 2x^2+2y^2$, (48) 和 (41)，

$$\begin{aligned}\tilde{\mathbb{E}}\left[\langle\mu'-\mu_0,\ v\rangle^2\right] &\leqslant 2(\text{A})^2 + 2\left(\frac{1}{n}\sum_{i=1}^{n}a_i\langle x_i-\mu_0,\ v\rangle\right)^2\\ &\leqslant 4\eta\Big(\alpha_0+\alpha_2+\tilde{\mathbb{E}}\left[\langle\mu'-\mu_0,\ v\rangle^2\right]\Big)+2\alpha_1^2.\end{aligned}$$

重新整理得（42）。将（42）代入（43）并化简，得到（44 ）. □

B 频率学派推论

本节证明主要结果的频率学派推论，具体表述见 定理 1.2 和定理 1.5。

B.1 频率学派均值估计

定理 *1.2* 的证明。 除概率为 $\frac{\beta}{3}$ 的事件外，样本均值 $\bar{x}$ 满足

$$\|\bar{x}\|_2 \leqslant O\Big(R + \sqrt{\mathrm{Tr}(\Lambda^2) + \log 1/\beta}\Big).$$

在上述条件下，根据定理 5.3，只要

$$n \geqslant \tilde{\Omega}\Big(\frac{d}{\varepsilon}\log\Big(\frac{R + \|\Lambda\|_F}{\alpha}\Big)\Big),$$

其中 $\tilde{\Omega}$ 隐藏了 $\log\log \frac{1}{\beta}$ 的因子，除去概率为 $\frac{\beta}{3}$ 的情形，我们的低效估计量输出 $\hat{\mu}$，满足

$$\|\bar{x} - \hat{\mu}\| \leqslant \tilde{O}\Big(\frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}}\log(1/\beta)}{\varepsilon n}\Big),$$

而我们的高效估计量输出 $\hat{\mu}$，满足

$$\|\bar{x} - \hat{\mu}\| \leqslant \tilde{O}\Big(\frac{\|\Lambda\|_{4/3} + \|\Lambda\|_{\mathrm{op}}\log^{3/4}(1/\beta)}{\varepsilon^{1/2}n^{3/4}} + \frac{\mathrm{tr}(\Lambda) + \|\Lambda\|_{\mathrm{op}}\log(1/\beta)}{\varepsilon n}\Big)$$

最后，除去概率为 $\frac{\beta}{3}$ 的情形，

$$\|\bar{x} - \mu\|_2 \leqslant \sqrt{\frac{\|\Lambda\|_F^2 + \|\Lambda\|_{\mathrm{op}}^2\log\frac{3}{\beta}}{n}} = (1 + o(1))\sqrt{\frac{\|\Lambda\|_F^2 + \|\Lambda\|_{\mathrm{op}}^2\log\frac{1}{\beta}}{n}}.$$

其中，当有效维数 $\|\Lambda\|_F/\|\Lambda\|_{\mathrm{op}} \to \infty$ 或 $\beta \to 0$ 时，$o(1)$ 成立。对所有三个事件应用并集界和三角不等式，即可得出该定理。 □

B.2 频率学派线性回归

在本部分，我们给出定理 1.5的证明。首先，我们给出一个关于普通最小二乘估计量所达到误差率的引理。

引理 B.1。 设 $X := [x_1, \ldots, x_n] \in \mathbb{R}^{d\times n}, y = X^\intercal w + \xi, \xi \sim \mathcal{N}(0, \mathrm{Id}_n)$。，并设

$$w_{\mathrm{OLS}} := (XX^\intercal)^{-1}Xy.$$

对于任意 $\beta \in (0, 1/2)$，定义

$$r_\beta := C\sqrt{\frac{d + \log(4/\beta)}{n}},$$

其中 $C > 0$ 是一个足够大的绝对常数。如果 $r_\beta \leqslant 1/2$，那么以至少 $1 - \beta$ 的概率，

$$\|w_{\mathrm{OLS}} - w\|_2^2 \leqslant \frac{1}{(1 - r_\beta)n}\left(d + 2\sqrt{d\log\frac{4}{\beta}} + 2\log\frac{4}{\beta}\right).$$

因此，如果 $r^\beta = o(1)$，那么对于某个普适常数 c，

$$\|w_{\mathrm{OLS}} - w\|_2 \leqslant (1 + o(1))\sqrt{\frac{d + c\log(1/\beta)}{n}}.$$

证明。 考虑随机事件

$$\mathscr{E}_{\mathrm{cov}} := \left\{ \left\| \frac{1}{n} XX^\intercal - \mathrm{Id} \right\|_{\mathsf{op}} \leqslant r_\beta \right\}.$$

根据定理 C.1，如有必要可增大C $\mathbb{P}(\mathscr{E}_{\mathrm{cov}}) \geqslant 1-\beta/2$。在 $\mathscr{E}_{\mathrm{cov}}$ 上，特别地，$XX^\intercal$可逆，且 $(1-r_\beta)$ $n\mathrm{Id} \preceq XX^\intercal \preceq (1+r_\beta)n\mathrm{Id}$。

此外，$w_{\mathrm{OLS}} - w = (XX^\intercal)^{-1}X\,\xi$。给定$X$，这是一个中心高斯分布，其协方差为 $\mathbb{R}^d \Sigma_X := (XX^\intercal)^{-1}$。因此，根据标准高斯尾界，给定$X$，以至少 $1-\beta/2$ 的概率，

$$\|w_{\mathrm{OLS}} - w\|_2^2 \leqslant \mathrm{tr}(\Sigma_X) + 2\sqrt{\mathrm{tr}(\Sigma_X^2)\log\frac{2}{\beta}} + 2\|\Sigma_X\|_{\mathsf{op}}\log\frac{2}{\beta}.$$

在 $\mathscr{E}_{\mathrm{cov}}$ 上，我们有

$$\mathrm{tr}(\Sigma_X) \leqslant \frac{d}{(1-r_\beta)n}, \qquad \mathrm{tr}(\Sigma_X^2) \leqslant \frac{d}{(1-r_\beta)^2 n^2}, \qquad \|\Sigma_X\|_{\mathsf{op}} \leqslant \frac{1}{(1-r_\beta)n}.$$

代入　　这些界得到：在 $\mathscr{E}_{\mathrm{cov}}$ 上，条件概率至少为 $1-\delta/2$,

$$\|w_{\mathrm{OLS}} - w\|_2^2 \leqslant \frac{1}{(1-r_\beta)n}\left(d + 2\sqrt{d\log\frac{2}{\beta}} + 2\log\frac{2}{\beta} \right).$$

由并集界可得第一个结论，而第二个结论则直接由$r^\beta = o(1)$推出。 □

定理 *1.5* 的证明。 按照定理 6.16 和定理 6.2的类似证明，在 $\eta = 0$ 和半径界R的条件下，只要定理 1.5 的假设成立，则以至少 $1-\frac{\beta}{3}$ 的概率，我们的低效估计量输出 $\hat{w}$，使得 −

$$\|\hat{w} - w_{\mathrm{OLS}}\|_2 \leqslant \widetilde{O}\left(\frac{d+\log(3/\beta)}{\varepsilon n} \right) = \widetilde{O}\left(\frac{d+\log(1/\beta)}{\varepsilon n} \right),$$

除概率 $\frac{\beta}{3}$ 外，我们的高效估计量输出 $\hat{w}$ ，使得

$$\|\hat{w} - w_{\mathrm{OLS}}\|_2 \leqslant \widetilde{O}\left(\frac{(d+\log(3/\beta))^{3/4}}{\varepsilon^{1/2} n^{3/4}} + \frac{d+\log(3/\beta)}{\varepsilon n} \right) = \widetilde{O}\left(\frac{(d+\log(1/\beta))^{3/4}}{\varepsilon^{1/2} n^{3/4}} + \frac{d+\log(1/\beta)}{\varepsilon n} \right).$$

最后，应用引理 B.1 并取 $\delta = \beta/3$，除概率 $\beta/3$外，

$$\|w_{\mathrm{OLS}} - w\|_2^2 \leqslant \frac{1}{(1-r_\beta)n}\left(d + 2\sqrt{d\log\frac{12}{\beta}} + 2\log\frac{12}{\beta} \right),$$

其中

$$r_\beta := C\sqrt{\frac{d+\log(12/\beta)}{n}} = o(1).$$

特别地，如果 $\log(1/\beta) = o(d)$，则

$$\|w_{\mathrm{OLS}} - w\|_2 \leqslant (1+o(1))\sqrt{\frac{d+\log(1/\beta)}{n}}.$$

该定理可由对这三个事件应用并集界，以及w、w_{OLS} 和 $\hat{w}$ 之间的三角不等式得出。

□

C 延后证明自 第 6 节

C.1 可行性的集中不等式

这里我们证明在证明真实值对于程序可行时所用到的集中不等式 $\mathscr{A}$，见算法 6.6。

定理 C.1。 令 $x_1, x_2, \ldots, x_n$ 为从高斯分布 $N(0,\Sigma)$中独立同分布抽取的样本。则以至少 $1-\beta$ 的概率，有

$$\left\| \frac{1}{n} \sum_{i\in[n]} x_i x_i^{\intercal} - \Sigma \right\|_{\mathrm{op}} \leqslant \sqrt{\frac{\mathrm{Tr}(\Sigma) + \log(1/\beta)\|\Sigma\|_{\mathrm{op}}}{n}}.$$

定理 C.2 （高斯坐标的最大子集）。 令 $v = (v_1, \ldots, v_d) \sim$服从 $N(0, \mathrm{Id}_d)$，令 $\eta \in (0,1)$，并令 $\beta \in (0,1)$。设 $m := \eta d$。则以至少 $1-\beta$ 的概率（关于 v 的抽样），

$$\max_{\substack{S\subseteq[d] \\ |S|=m}} \sum_{i\in S} v_i^2 \leqslant C\Big(\eta d \log\frac{e}{\eta} + \log\frac{1}{\beta}\Big), \tag{49}$$

其中 $C > 0$ 是绝对常数。换言之：对于每个大小为 ηd 的坐标子集 S，这些坐标的平方和同时至多为右侧的值。

证明。 固定任意大小为 $S \subseteq [d]$的子集 $m = \eta d$。由于 $v \sim N(0, I_d)$，随机变量

$$X_S := \sum_{i\in S} v_i^2$$

服从具有 m 个自由度的 χ^2 分布，即 $X_S \sim \chi^2{}_m$。我们将使用卡方分布的标准 Laurent–Massart 尾界：若 $X \sim \chi^2{}_m$，则对每个 $x > 0$，

$$\mathbb{P}\big(X \geqslant m + 2\sqrt{mx} + 2x\big) \leqslant e^{-x}. \tag{50}$$

将 (50) 应用于 X_S，取参数 $x > 0$，其值稍后确定：

$$\mathbb{P}\big(X_S \geqslant m + 2\sqrt{mx} + 2x\big) \leqslant e^{-x}. \tag{51}$$

我们希望上界同时 对所有大小为 m 的子集 S 都成立。这类子集的总数为

$$\binom{d}{m} \leqslant \Big(\frac{ed}{m}\Big)^m = \Big(\frac{e}{\eta}\Big)^{\eta d}. \tag{52}$$

由并集界，结合 (51) 和 (52)，可得

$$\mathbb{P}\Big(\exists S \subseteq [d], |S| = m : X_S \geqslant m + 2\sqrt{mx} + 2x\Big) \leqslant \binom{d}{m} e^{-x} \leqslant \Big(\frac{e}{\eta}\Big)^{\eta d} e^{-x}.$$

选择

$$x := \eta d \log\frac{e}{\eta} + \log\frac{1}{\beta}. \tag{53}$$

Then

$$\Big(\frac{e}{\eta}\Big)^{\eta d} e^{-x} = \exp\big(\eta d \log(e/\eta)\big)\, \exp\big(-\eta d \log(e/\eta) - \log(1/\beta)\big) = \beta.$$

因此，概率至少为 $1-\beta$,

$$\forall S \subseteq [d], |S| = m: \quad \sum_{i \in S} v_i^2 \leqslant m + 2\sqrt{mx} + 2x, \tag{54}$$

其中 x 由式（53）给出。

接下来只需化简右侧。回想一下，$m = \eta d$ 且

$$x = \eta d \log \frac{e}{\eta} + \log \frac{1}{\beta}.$$

Then

$$m + 2x \leqslant 3\eta d \log \frac{e}{\eta} + 2 \log \frac{1}{\beta},$$

and

$$2\sqrt{mx} = 2\sqrt{\eta d\big(\eta d \log(e/\eta) + \log(1/\beta)\big)} \leqslant \eta d \log \frac{e}{\eta} + \log \frac{1}{\beta},$$

放大常数后（这是形如 $2\sqrt{ab} \leqslant a + b$ 的标准不等式）。因此，对于某个普适常数 $C > 0$,

$$m + 2\sqrt{mx} + 2x \leqslant C\Big(\eta d \log \frac{e}{\eta} + \log \frac{1}{\beta}\Big).$$

将其代回式（54）便恰好得到式（49）。 □

定理 C.3（高斯向量的次序统计量及其尖锐的 β-依赖性）。设 $v = (v_1, \ldots, v_d) \sim N(0, \mathrm{Id}_d)$，且令 η, $\beta \in (0,1)$。将坐标平方按非递增顺序排列：

$$v_{(1)}^2 \geqslant v_{(2)}^2 \geqslant \ldots \geqslant v_{(d)}^2.$$

那么，以至少 $1-\beta$ 的概率，

$$v_{(\lceil \eta d \rceil)}^2 \leqslant 2 \log \frac{e}{\eta} + \frac{2}{\eta d} \log \frac{1}{\beta}. \tag{55}$$

等价地：以至少 $1-\beta$ 的概率，至多 η 个坐标的平方值超过 (55) 右侧的值。

证明。固定 $\eta \in (0,1)$ 并设 $m := \lceil \eta d \rceil$。令 $T > 0$ 为待选定的阈值。我们要界定存在m 个v 的坐标且它们的平方值均至少为T的概率。

对于单个坐标$v_i \sim$ 正态分布 $(0,1)$，我们有

$$\mathbb{P}(v_i^2 \geqslant T) = \mathbb{P}(|v_i| \geqslant \sqrt{T}) \leqslant e^{-T/2}.$$

现在固定一个$S \subseteq [d]$ 的子集，其中 $|S| = m$。根据独立性，

$$\mathbb{P}\big(v_i^2 \geqslant T \text{ for all } i \in S\big) \leqslant \big(e^{-T/2}\big)^m = e^{-(T/2)m}.$$

这样的子集S共有 $\binom{d}{m}$ 个。根据并集界，

$$\mathbb{P}\Big(\exists S \subseteq [d], |S| = m : v_i^2 \geqslant T \ \forall i \in S\Big) \leqslant \binom{d}{m} e^{-(T/2)m}. \tag{56}$$

我们将用标准的类熵上界来估计二项式系数

$$\binom{d}{m} \leqslant \Big(\frac{ed}{m}\Big)^m.$$

由于 $m = \lceil \eta d \rceil \geqslant \eta d$，我们有

$$\binom{d}{m} \leqslant \Big(\frac{ed}{\eta d}\Big)^m = \Big(\frac{e}{\eta}\Big)^m.$$

代入式（56）得

$$\mathbb{P}\Big(\exists S : |S| = m,\ v_i^2 \geqslant T\ \forall i \in S\Big) \leqslant \Big(\frac{e}{\eta}\Big)^m e^{-(T/2)m} = \exp\Big(m\big(\log \tfrac{e}{\eta} - \tfrac{T}{2}\big)\Big).$$

我们希望右侧至多为 β。由于$m \geqslant \eta d$，只需满足

$$\exp\Big(\eta d\big(\log \tfrac{e}{\eta} - \tfrac{T}{2}\big)\Big) \leqslant \beta.$$

两边取对数，

$$\eta d\big(\log \tfrac{e}{\eta} - \tfrac{T}{2}\big) \leqslant \log \beta \quad \Longleftrightarrow \quad -\tfrac{T}{2} \leqslant \frac{1}{\eta d}\log\beta - \log\tfrac{e}{\eta}.$$

乘以 -1 （并反转不等号）：

$$\frac{T}{2} \geqslant \log \tfrac{e}{\eta} + \frac{1}{\eta d}\log\frac{1}{\beta}.$$

因此，只需选择

$$T := 2\log\frac{e}{\eta} + \frac{2}{\eta d}\log\frac{1}{\beta}.$$

取此T值后，式（56）中的概率至多为 β。

但（56）中的事件恰好是：存在 m 个坐标都大于 $\geqslant T$。因此，其补事件表示至多有 $m-1 \leqslant \eta d$ 个坐标超过 T。等价地，在该补事件中，我们有

$$v_{(m)}^2 \leqslant T,$$

i.e.

$$v_{(\lceil \eta d \rceil)}^2 \leqslant 2\log\frac{e}{\eta} + \frac{2}{\eta d}\log\frac{1}{\beta},$$

其概率至少为 $1-\beta$，这恰好就是（55）。 □

最后，我们使用高斯子矩阵的最大稀疏奇异值。

引理 C.4 （高斯子矩阵的稀疏谱范数）。 设 $X \in \mathbb{R}^{d\times n}$ 具有独立同分布列 $x_1, \ldots, x_n \sim \mathcal{N}(0, I_d)$。对于指标集 $S \subseteq [n]$, , 令 $X_S \in \mathbb{R}^{d\times|S|}$ 表示由 S 中指标对应的列组成的子矩阵。对于 $k \in [n]$, , 回顾 k- 稀疏谱范数的定义（定义 *4.22*）

$$\|X\|_{\mathrm{sp}}(k) := \max_{\substack{S\subseteq[n]\\ |S|\leqslant k}} \|X_S\|_{\mathrm{op}}.$$

那么，存在普适常数 $C > 0$, 使得对于每个 $k \in [n]$ 和 $\beta \in (0,1)$, 其概率至少为 $1-\beta$,

$$\|X\|_{\mathrm{sp}}(k) \leqslant C\left(\sqrt{d} + \sqrt{k\log\frac{en}{k}} + \sqrt{\log\frac{1}{\beta}}\right).$$

等价地，在同一事件上，对所有满足 $\subseteq [n]$ 且 $|S| \leqslant k$ 的 S,

$$\|X_S\|_{\mathrm{op}}^2 \leqslant C^2\left(d + k\log\frac{en}{k} + \log\frac{1}{\beta}\right).$$

证明。 固定$s \in \{1, \ldots, k\}$，并取一个集合$S \subseteq [n]$，满足 $|S| = s$。根据$d \times s$ 高斯矩阵的标准算子范数尾界，对于所有$t \geqslant 0$,

$$\mathbb{P}\Big[\|X_S\|_{\mathsf{op}} \geqslant \sqrt{d} + \sqrt{s} + t\Big] \leqslant 2e^{-t^2/2}. \tag{57}$$

至多有$\binom{n}{s} \leqslant (en/s)^s$ 个这样的子集。令

$$t_s := \sqrt{2\left(s \log \frac{en}{s} + \log \frac{2k}{\beta}\right)}.$$

根据并集界和（57），

$$\mathbb{P}\left[\max_{|S|=s} \|X_S\|_{\mathsf{op}} \geqslant \sqrt{d} + \sqrt{s} + t_s\right] \leqslant \binom{n}{s} \cdot 2e^{-t_s^2/2} \leqslant \frac{\beta}{k}.$$

取并集界_久久爱 $s = 1$、…、k进行并集界处理可得：以至少 $1-\beta$ 的概率，对所有满足 $|S| \leqslant k$ 的 S， ，

$$\|X_S\|_{\mathsf{op}} \leqslant \sqrt{d} + \sqrt{|S|} + t_{|S|} \leqslant \sqrt{d} + \sqrt{k} + \sqrt{2k \log \frac{en}{k}} + \sqrt{2 \log \frac{2k}{\beta}}.$$

将 $\sqrt{k}$ 和 $\sqrt{\log(2k/\beta)}$ 吸收到所显示的表达式中（通过增大普适常数 C），即可得到所述界。

□

C.2 定理 6.2的证明：用于稳健且私有贝叶斯回归的低效算法

在本节中，我们首先给出一种用于贝叶斯回归的低效鲁棒算法。为简单起见，我们基于先前提出的鲁棒且高效的线性回归估计量来构造算法。与经验均值估计类似，我们可以利用更优的韧性界，鲁棒地近似后验均值估计量（或等价地，岭回归估计量）。最后，我们使用黑盒鲁棒性到隐私的归约[HKMN23、AUZ23] 来开发一种具有最优误差保证的差分隐私算法。

C.2.1 用于稳健回归的低效算法

定理 C.5。 设 $\Sigma = \sigma^2 \mathrm{Id}_d$ 为已知参数 $\sigma > 0$。设 $\beta \in [0, 1]$。那么存在一种计算上低效的算法，该算法在给定 η-污染样本 $(X^{\text{input}}, y_{\text{input}})$（按 定义 6.1生成）后，输出一个估计量 $\hat{w}$，使得其概率至少为 $1-\beta$,

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|_2 \leqslant O\Big(\eta \log(1/\eta)\Big),$$

其中 (X°, y°) 是未污染样本，且满足条件 $n \geqslant (d + \log(1/\beta))/\eta$..

全文中，(X°, y°) 表示未被污染的样本，而 $(X^{\text{input}}, y_{\text{input}})$ 表示被 η污染的样本。设$I^\star \subseteq [n]$为（未知的）未被污染样本索引集合，因此 $|I^\star| \geqslant (1-\eta)n$。

假设 $\Sigma = \sigma^2 \mathrm{Id}_d$，并定义

$$\lambda := \frac{1}{n\sigma^2}, \qquad A^\star := \frac{1}{n} X^\circ X^{\circ\intercal}, \qquad b^\star := \frac{1}{n} X^\circ y^\circ.$$

那么后验均值为

$$w_{\text{post}} := \mathbb{E}[w \mid X^\circ, y^\circ] = (A^\star + \lambda \mathrm{Id}_d)^{-1} b^\star.$$

对于任意$u \in \mathbb{R}^d$，定义（干净的）残差向量

$$r^\star(u) := y^\circ - X^{\circ\intercal} u \in \mathbb{R}^n, \qquad g^\star(u) := \frac{1}{n} X^\circ r^\star(u) \in \mathbb{R}^d.$$

注意$g^\star(u) = b^\star - A^\star u$，因此以下恒等式对每个$u$都成立：

$$w_{\text{post}} = u + (A^\star + \lambda \mathrm{Id}_d)^{-1}\big(g^\star(u) - \lambda u\big). \tag{58}$$

定义失败概率 $\beta_1 = \beta_2 = \beta_3 = \beta/3$。

对于序列$\{z_i\}_{i=1}^n \subset \mathbb{R}^d$ 和 $\eta \in (0,1)$，定义韧性

$$\mathrm{Res}_\eta(z) := \sup_{\substack{S \subseteq [n] \\ |S| \leqslant \eta n}} \left\| \frac{1}{n} \sum_{i \in S} z_i \right\|_2.$$

我们将算法定义为 算法 C.6。

算法 C.6 （两阶段鲁棒后验均值估计）。输入：η污染的设计矩阵 $X_{\text{input}} \in \mathbb{R}^{d \times n}$（列 x_i）、响应变量 $y_{\text{input}} \in \mathbb{R}^n$、污染比例 $\eta \in (0, 1/2)$、失败概率 $\beta \in (0,1)$、先验方差 $\sigma > 0$（因此 $\Sigma = \sigma^2 I_d$）。输出：后验均值 $w_{\text{post}} = \mathbb{E}[w \mid X^\circ, y^\circ]$的一个估计量$\hat{w} \in \mathbb{R}^d$。

1. *Assume* $n \gtrsim (d + \log(1/\beta))/\eta$. *Set* $\beta_1 = \beta_2 = \beta_3 = \beta/3$ *and* $\lambda \leftarrow (n\sigma^2)^{-1}$.
2. （阶段 ***1***：粗略鲁棒回归）运行定理 *6.3* 中的多项式时间鲁棒回归算法，以获得估计量 $\tilde{w}$。
3. （阶段 ***2***：残差化）计算残差 $r \leftarrow y_{\text{input}} - X_{\text{输入}}^\intercal \tilde{w}$ 和向量 $v \leftarrow X_{\text{input}} r$。
4. 设置

$$\delta_0 \leftarrow C_1\Big(\sqrt{\tfrac{d + \log(1/\beta_1)}{n}} + \eta\Big),$$

and

$$\tau \leftarrow 2\eta\,\delta_0 + C_2(1 + \delta_0)\Big(\sqrt{\eta \log \tfrac{1}{\eta}}\sqrt{\tfrac{d + \log(1/\beta_2)}{n}} + \eta \log \tfrac{1}{\eta}\Big).$$

5. 在 $(\{v_i\}_{i=1}^n$、η、$\tau)$ 上运行算法 *C.7*，以获得$\hat{g}$。
6. 返回

$$\hat{w} \leftarrow \tilde{w} + \frac{1}{1+\lambda}\big(\hat{g} - \lambda \tilde{w}\big).$$

算法 C.7（替换式稳健均值估计）。
输入：向量 $v_1,\dots,v_n\in\mathbb{R}^d$，污染比例 $\eta\in(0,1/2)$，韧性界 $\tau>0$。输出：干净均值 $\frac{1}{n}\sum_{i=1}^{n}v^\star_i$ 的稳健估计 $\hat{g}$。

1. 找到向量 $v'_1,\dots,v'_n\in\mathbb{R}^d$ 以及指标集 $J\subseteq[n]$，使得 $|J|\leqslant\eta n$：

   (a) 对所有 $i\notin J$，都有 $v'_i=v_i$；(b)

$$\sup_{\substack{S\subseteq[n]\\|S|\leqslant\eta n}}\left\|\frac{1}{n}\sum_{i\in S}v'_i\right\|_2\leqslant\tau.$$

2. 输出

$$\hat{g}\leftarrow\frac{1}{n}\sum_{i=1}^{n}v'_i.$$

引理 C.8（证书 $\Rightarrow$ 与干净均值的接近程度）。设 $\{z_i\}$ 是 η污染的未知干净序列 $\{z^\star_i\}$，即它们至多在 η个指标上不同。假设存在 τ 使得 $\mathrm{Res}_{2\eta}(z)\leqslant\tau$，且算法 C.7 输出替换后的 z'，满足 $\mathrm{Res}_{2\eta}(z')\leqslant\tau$。则

$$\left\|\frac{1}{n}\sum_{i=1}^{n}z'_i-\frac{1}{n}\sum_{i=1}^{n}z^\star_i\right\|_2\leqslant2\tau.$$

证明。令 $\Delta:=\{i:z'_i\neq z^\star_i\}$。由于 z 与 $z^\star$ 至多在 ηn 个指标上不同，且 z' 与 z 至多在 ηn 个指标上不同，因此有 $|\Delta|\leqslant2\eta n$。于是

$$\frac{1}{n}\sum_{i=1}^{n}(z'_i-z^\star_i)=\frac{1}{n}\sum_{i\in\Delta}(z'_i-z^\star_i),$$

因此，根据三角不等式，

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(z'_i-z^\star_i)\right\|_2\leqslant\left\|\frac{1}{n}\sum_{i\in\Delta}z'_i\right\|_2+\left\|\frac{1}{n}\sum_{i\in\Delta}z^\star_i\right\|_2\leqslant\mathrm{Res}_{2\eta}(z')+\mathrm{Res}_{2\eta}(z^\star)\leqslant2\tau.$$

□

现在我们给出韧性界。定义 $\tilde{w}$处的（干净）残差：

$$r^\star_i:=y^\circ(i)-\langle X^\circ(\cdot,i),\tilde{w}\rangle=\langle X^\circ(\cdot,i),w-\tilde{w}\rangle+\xi_i,\qquad\xi_i\sim\mathcal{N}(0,1),$$

以及干净向量

$$v^\star_i:=X^\circ(\cdot,i)\,r^\star_i\in\mathbb{R}^d.$$

记 $\Delta:=w-\tilde{w}$，并注意 $\mathbb{E}[v^\star_i\,|\,\Delta]=\Delta$。

定义 $\tilde{w}$处的干净残差为

$$r^\star_i:=y^\circ(i)-\langle X^\circ(\cdot,i),\tilde{w}\rangle=\langle X^\circ(\cdot,i),w-\tilde{w}\rangle+\xi_i,\qquad\xi_i\sim\mathcal{N}(0,1),$$

以及干净向量

$$v^\star_i:=X^\circ(\cdot,i)\,r^\star_i\in\mathbb{R}^d.$$

记 $\Delta:=w-\tilde{w}$。注意，对于每个i,

$$v^\star_i=X^\circ(\cdot,i)X^\circ(\cdot,i)^\intercal\Delta+X^\circ(\cdot,i)\,\xi_i.$$

引理 C.9 （{vi } 的小集合鲁棒性界 $\{v^\star\ i\}$）。 假设 $\|\Delta\|_2 \leqslant \delta$，并令 $k\colon = \lceil 2\eta n\rceil$。在 $n \gtrsim d_{+\log_\eta^{(1/\beta_2)}}$ 的假设下，存在一个~~绝对常数~~ C，使得在干净抽样 X 上，y°) 以至少 $1-\beta_2$ 的概率成立，

$$\mathrm{Res}_{2\eta}(v^\star) \leqslant C(1+\delta)\left(\sqrt{\eta\log\frac{e}{\eta}}\sqrt{\frac{d+\log(1/\beta_2)}{n}} + \eta\log\frac{e}{\eta}\right).$$

证明。 令 $x_i \coloneqq X^\circ(\cdot,\ i)$，对于 $S \subseteq [n]$，令 $X^\circ{}_S$ 为由索引属于 S 的列组成的子矩阵。另令 $\xi_S \in \mathbb{R}^{|S|}$ 表示 $(\xi_1,\ldots,\xi_n)$ 在 S 上的限制。

对于每个 $S \subseteq [n]$，满足 $|S| \leqslant k$，利用

$$v_i^\star = x_i x_i^\intercal \Delta + x_i \xi_i,$$

我们有

$$\frac{1}{n}\sum_{i\in S} v_i^\star = \frac{1}{n} X^\circ{}_S X^{\circ\intercal}_S \Delta + \frac{1}{n} X^\circ{}_S \xi_S.$$

因此，

$$\left\|\frac{1}{n}\sum_{i\in S} v_i^\star\right\|_2 \leqslant \frac{1}{n}\|X^\circ{}_S\|_{\mathrm{op}}^2\|\Delta\|_2 + \frac{1}{n}\|X^\circ{}_S\|_{\mathrm{op}}\|\xi_S\|_2.$$

对所有 $|S| \leqslant k$ 取上确界，得到

$$\mathrm{Res}_{2\eta}(v^\star) \leqslant \frac{\delta}{n}\|X^\circ\|_{\mathrm{sp}}(k)^2 + \frac{1}{n}\|X^\circ\|_{\mathrm{sp}}(k)\max_{|S|\leqslant k}\|\xi_S\|_2. \tag{59}$$

现在我们对右侧的两个因子进行界定。

首先，根据引理 C.4，以至少 $1-\beta_2/2$ 的概率，

$$\|X^\circ\|_{\mathrm{sp}}(k) \leqslant C_1\left(\sqrt{d} + \sqrt{k\log\frac{en}{k}} + \sqrt{\log\frac{2}{\beta_2}}\right).$$

因此，在该事件上，

$$\frac{\delta}{n}\|X^\circ\|_{\mathrm{sp}}(k)^2 \leqslant C\delta\left(\frac{d+\log(1/\beta_2)}{n} + \frac{k}{n}\log\frac{en}{k}\right). \tag{60}$$

其次，由于 $(\xi_1,\ldots,\ \xi_n) \sim$ 正态分布 $(0, I_n)$，且平方后各坐标均为非负，

$$\max_{|S|\leqslant k}\|\xi_S\|_2^2 = \max_{|S|=k}\sum_{i\in S}\xi_i^2.$$

应用定理 C.2 在 n 维空间中，子集大小为 k，我们得到以至少 $1-\beta_2/2$ ，

$$\max_{|S|\leqslant k}\|\xi_S\|_2^2 \leqslant C_2\left(k\log\frac{en}{k} + \log\frac{2}{\beta_2}\right). \tag{61}$$

在事件 (60) 和 (61) 的交集上，其概率至少为 $1-\beta_2$，(59) 中的第二项有界为

$$\begin{aligned}\frac{1}{n}\|X^\circ\|_{\mathrm{sp}}(k)\max_{|S|\leqslant k}\|\xi_S\|_2 &\leqslant \frac{C}{n}\left(\sqrt{d} + \sqrt{k\log\frac{en}{k}} + \sqrt{\log\frac{1}{\beta_2}}\right)\left(\sqrt{k\log\frac{en}{k}} + \sqrt{\log\frac{1}{\beta_2}}\right)\\ &\leqslant C\left(\sqrt{\frac{d+\log(1/\beta_2)}{n}}\sqrt{\frac{k\log(en/k)+\log(1/\beta_2)}{n}} + \frac{k\log(en/k)+\log(1/\beta_2)}{n}\right).\end{aligned}$$

由于 $k = \lceil 2\eta n \rceil$，因此有 $k/n \asymp \eta$ 且

$$\log \frac{en}{k} \asymp \log \frac{e}{\eta}.$$

将这些结果代入上述两个界，并结合（59）即可证明

$$\mathrm{Res}_{2\eta}(v^\star) \leqslant C\delta\left(\frac{d+\log(1/\beta_2)}{n} + \eta \log \frac{e}{\eta}\right) + C\left(\sqrt{\eta \log \frac{e}{\eta}}\sqrt{\frac{d+\log(1/\beta_2)}{n}} + \eta \log \frac{e}{\eta} + \frac{\log(1/\beta_2)}{n}\right).$$

最后，在以下假设成立的条件下

$$n \gtrsim \frac{d+\log(1/\beta_2)}{\eta},$$

项 $\delta(d+\log(1/\beta_2))/n$ 和 $\log(1/\beta_2)/n$ 均可吸收到

$$C(1+\delta)\,\eta \log \frac{e}{\eta},$$

由此得到最终的误差界。 □

定理 *C.5* 的证明。在给定 $(X^\circ、y^\circ)$的条件下，后验分布为高斯分布，其协方差为

$$\mathrm{Cov}(w^* \mid X^\circ, y^\circ) = (X^\circ X^{\circ\intercal} + \sigma^{-2}\mathrm{Id}_d)^{-1} = \frac{1}{n}(A^\star + \lambda \mathrm{Id}_d)^{-1}.$$

在事件 $\lambda_{\min}(A^\star) \geqslant 1/2$ 上，我们有 $(A^\star + \lambda \mathrm{Id}_d)^{-1} \preceq 2\mathrm{Id}_d$，因此 $\mathrm{Cov}(w \mid X\circ、y^\circ) \preceq {}^{2}{}_{n}\mathrm{Id}_d$。因此，以至少 $1-\beta_3$ 的概率（可能需要调整常数），

$$\|w^* - w_{\mathrm{post}}\|_2 \leqslant C_5\sqrt{\frac{d+\log(1/\beta_3)}{n}}. \tag{62}$$

应用引理 C.9，取 $\delta := \|\Delta\|_2$(给定粗略估计器，我们有 $\delta \leqslant C_1(\sqrt{(d+\log(1/\beta_1))/n} + \eta)$)。令 τ 为引理 C.9右侧的表达式。观测向量 $v_i = x_i r_i$ 受到 η-污染，来自 $v^\star{}_i$，因此证书补全过程（定义 C.7）生成 $\hat{g}$，满足

$$\|\hat{g} - g^\star(\tilde{w})\|_2 \leqslant 2\tau \tag{63}$$

根据引理 C.8，其成立的概率至少为 $1-\beta_2$。

现在，我们将 $\hat{w}$ 与 w^{post} 联系起来，并控制 $A^\star - \mathrm{Id}_d$ 带来的误差。利用（58），取 $u = \tilde{w}$，

$$w_{\mathrm{post}} = \tilde{w} + (A^\star + \lambda \mathrm{Id}_d)^{-1}\big(g^\star(\tilde{w}) - \lambda \tilde{w}\big).$$

回忆一下，$\hat{w} = \tilde{w} + \frac{1}{1+\lambda}(\hat{g} - \lambda \tilde{w})$。相减可得

$$\hat{w} - w_{\mathrm{post}} = \frac{1}{1+\lambda}(\hat{g} - g^\star(\tilde{w})) + \Big(\frac{1}{1+\lambda}\mathrm{Id}_d - (A^\star + \lambda \mathrm{Id}_d)^{-1}\Big)\big(g^\star(\tilde{w}) - \lambda \tilde{w}\big).$$

利用预解式恒等式

$$(A^\star + \lambda \mathrm{Id}_d)^{-1} - \frac{1}{1+\lambda}\mathrm{Id}_d = (A^\star + \lambda \mathrm{Id}_d)^{-1}(\mathrm{Id}_d - A^\star)\frac{1}{1+\lambda}\mathrm{Id}_d$$

以及 $\frac{1}{1+\lambda} \leqslant 1$，我们得到

$$\|\hat{w} - w_{\mathrm{post}}\|_2 \leqslant \|\hat{g} - g^\star(\tilde{w})\|_2 + \|A^\star - \mathrm{Id}_d\|_{\mathrm{op}} \cdot \|w_{\mathrm{post}} - \tilde{w}\|_2, \tag{64}$$

其中我们用到了 $(A^\star+\lambda \mathrm{Id}_d)^{-1}(g^\star(\tilde{w})-\lambda\tilde{w})=w_{\text{post}}-\tilde{w}$ 根据（58）。

接下来,

$$\|w_{\text{post}}-\tilde{w}\|_2 \leqslant \|w_{\text{post}}-w\|_2+\|w-\tilde{w}\|_2.$$

在粗略估计误差保证事件与（62）的交集上,

$$\|w_{\text{post}}-\tilde{w}\|_2 \leqslant C_6\Big(\sqrt{\frac{d+\log(1/\beta)}{n}}+\eta\Big). \tag{65}$$

根据 Wishart矩阵的集中性，我们有

$$\|A^\star-\mathrm{Id}_d\|_{\text{op}} \leqslant C_4\sqrt{\frac{d+\log(1/\beta)}{n}}. \tag{66}$$

结合（65）和（66），

$$\|A^\star-\mathrm{Id}_d\|_{\text{op}}\cdot\|w_{\text{post}}-\tilde{w}\|_2 \leqslant C_7\left(\frac{d+\log(1/\beta)}{n}+\eta\sqrt{\frac{d+\log(1/\beta)}{n}}\right).$$

在所假设的情形 $n \gtrsim (d+\log(1/\beta))/\eta$ 下，我们有 $\frac{d+\log(1/\beta)}{n} \lesssim \eta$ 以及 $\eta\sqrt{\frac{d+\log(1/\beta)}{n}} \lesssim \eta$。因此,

$$\|A^\star-\mathrm{Id}_d\|_{\text{op}}\cdot\|w_{\text{post}}-\tilde{w}\|_2 \lesssim \eta. \tag{67}$$

最后，将（63）和（67）代入（64），

$$\|\hat{w}-w_{\text{post}}\|_2 \leqslant 2\tau+O(\eta).$$

最后，使用 $\delta=\|w^*-\tilde{w}\|_2 \lesssim \sqrt{\frac{d+\log(1/\beta)}{n}}+\eta$ 引理 C.9 中的结果，并将低阶 $O(\eta)$项吸收到 $\eta\log(1/\eta)$ 中（当 $\eta \leqslant 1/2$时），我们得到

$$\|\hat{w}-w_{\text{post}}\|_2 \leqslant O\left(\sqrt{\eta\log\frac{1}{\eta}}\sqrt{\frac{d+\log(1/\beta)}{n}}+\eta\log\frac{1}{\eta}\right)$$

在对上述使用的事件应用并集界后，以至少 $1-\beta$ 的概率成立。 □

C.2.2 从鲁棒性到隐私的归约

现在，利用针对指数时间算法的隐私到鲁棒性归约 [HKMN23]，我们得到 定理 6.2，为方便起见，在此重述:

定理（定理 *6.2* 的重述）。设 $\Sigma=\sigma^2 Id_d$，其中已知参数为 $\sigma>0$。设 $\beta\in[0,1]$, $\varepsilon \leqslant O(1)$。假设 $n \geqslant max\{d\,log(\sigma\sqrt{d})/\varepsilon, (d+log(1/\beta))/\eta\}$。那么存在一个计算上低效的算法，它在输入 η-受污染样本（$X^{\text{input}}, y_{\text{input}}$）（按定义 *6.1* 生成）时，输出一个 ε-*DP* 估计量 $\hat{w}$，使得以至少 $1-\beta$ 的概率,

$$\|\hat{w}-\mathbb{E}[w^*\mid X^\circ, y^\circ]\|_2 \leqslant \tilde{O}(\eta+\frac{d+\log(1/\beta)}{n\varepsilon}),$$

其中 (X°, y°) 是未污染样本。

证明。根据定理 C.5，我们得到一个低效算法，其误差率为 $\eta\log(1/\eta)$ 当$n \geqslant \tilde{\Omega}(d/\eta)$。使用黑盒隐私-鲁棒性归约（[HKMN23，引理 2。1]），我们得到一个隐私算法，其误差为 α 当

$$n \geqslant \tilde{\Omega}\Big(\frac{d+\log(1/\beta)}{\alpha\varepsilon}+\frac{d\log(\sigma\sqrt{d})}{\varepsilon}+\frac{(d+\log(1/\beta))\eta}{\alpha^2}\Big).$$

在满足条件 $\alpha \geqslant \eta\log(1/\eta)$ 的情况下。现在取 $\alpha=\eta\log(1/\eta)+\frac{d+\log(1/\beta)}{n\varepsilon}\log\Big(\frac{d+\log(1/\beta)}{\varepsilon}\Big)n$，即可得到定理 6.2的结论。

□

C.3 引理 6.17的证明：常数误差率下的稳健回归

这里我们给出弱先验区间内稳健回归在常数误差下的粗略估计过程的详细内容。为方便起见，结果陈述如下：

引理（引理 *6.17* 的重述）。设 $\Sigma = \sigma^2 Id_d$，其中 $\sigma = \omega(1/\sqrt{n})$ 是已知参数。存在一个多项式时间算法，对于按定义 *6.1* 生成的 η-受污染样本 (X, y)，输出 $\hat{w}$，使得以至少 $1-\beta$ 的概率，我们有 $\|\hat{w} - w^*\|_2 \leqslant O(1)$，前提是 $n \geqslant \Omega((d + log(1/\beta))/\eta)$。
回顾程序 $\mathscr{A} = \mathscr{A}_{\text{corr}} \cup \mathscr{A}'_{\text{res}}$ 的定义，其中 $\mathscr{A}_{\text{corr}}$ 定义于 式（23），而 $\mathscr{A}'_{\text{res}}$ 定义于 式（26）。形式上，我们的算法描述见 算法 C.10。

算法 C.10（恒定误差率下的稳健回归算法）。输入：η 受污染矩阵 $X_{input} \in \mathbb{R}^{n\times m}$，$y_{input} \in \mathbb{R}^n$输出：一个多项式时间估计器 $\hat{w} \in \mathbb{R}^d$

1. 对约束 $\mathscr{A}(X、y、\xi、w;\ X_{\text{input}}、y_{\text{input}}) = \mathscr{A}_{\text{corr}}(X、y、\xi) \cup \mathscr{A}'_{\text{res}}(X、y、w)$求解次数为 $O(1)$ 的 *SoS* 松弛，其中 $t = O(1)$，得到一个伪期望算子。
2. 使用 $\tilde{\mathbb{E}}[w]$ 对伪分布进行舍入。

引理 *6.17* 的证明引理 *6.17*。首先，正如定理 C.3所证明的，平方和程序中的约束以至少 $1-\beta$ 的概率得到满足。接下来，我们给出平方和可辨识性证明。令$r \in \{0,1\}^n$ 为污染集合。令 $v = 1-(1-\xi)\odot(1-r)$ 表示被污染的点以及/或被 ξ 标记为受污染的点所对应的指示函数。于是我们有

$$(y - X^\intercal w)\odot(1-v) = (y^\circ - X^{\circ\intercal} w)\odot(1-v),$$

so

$$(y^\circ - X^{\circ\intercal} w) - (y - X^\intercal w) = (y^\circ - X^{\circ\intercal} w)\odot v - (y - X^\intercal w)\odot v.$$

我们可以对 右侧的第二项使用$y - X^\intercal w$的短平分解来进行上界估计：

$$\|(y - X^\intercal w)\odot v\|^2 \leqslant \|z_1\|^2 + 2\eta n\|z_2\|_\infty^2 \leqslant \eta n\log(1/\eta) + \log(1/\beta).$$

另一方面，我们有

$$y^\circ - X^{\circ\intercal} w = y^\circ - X^{\circ\intercal} w^* + X^{\circ\intercal}(w^* - w)$$

As 真实值是可行的，$y^\circ - X^{\circ\intercal} w^*$ 也存在短平分解，因此

$$\|(y^\circ - X^{\circ\intercal} w^*)\odot v\|^2 \leqslant \eta n\log(1/\eta) + \log(1/\beta).$$

因此，接下来只需对 $\|(X^{\circ\intercal}(w - w^*))\odot v\|^2$进行上界估计。我们有如下平方和上界

$$\|(X^{\circ\intercal}(w - w^*))\odot v\|^2 = \sum_{i\in[n]} v_i \langle X^\circ(\cdot, i), w - w^*\rangle^2 \leqslant \|w - w^*\|^2,$$

其中第二步由选择器引理（引理 4.16）得出。由三角不等式可得

$$\left\|y^\circ - X^{\circ\intercal} w\right\|^2 \leqslant 2\eta n\log(1/\eta)\|w - w^*\|^2 + 2n.$$

于是我们有

$$\left\|X^{\circ\intercal} w - X^{\circ\intercal} w^*\right\|^2 \leqslant 2\eta n\log(1/\eta)\|w - w^*\|^2 + 4n,$$

由于我们有 $\left\|\frac{1}{n} X^\circ X^{\circ\intercal} - \mathrm{Id}\right\|_{\text{op}} \leqslant \sqrt{\frac{d + \log(1/\beta)}{n}}$，该伪分布满足

$$\|w - w^*\|^2 \leqslant O(1).$$

By 标准舍入后，我们得到一个估计量 $\tilde{\mathbb{E}}[w]$，其误差率为常数 te. □

C.4 弱先验下私有化鲁棒回归算法

C.4.1 私有化第1阶段的粗略估计器

在本节中， 我们对以下算法进行隐私化处理：算法 C.10。我们证明以下定理。

定理 C.11。 存在一个多项式时间的 ε-*DP* 算法，对于给定的 η-污染样本 $(X_{\text{in}}, y_{\text{in}})$，输出 $\widehat{w}$，使得

$$\mathbb{E}\big[\|\widehat{w}-w^*\|_2^2\big] \leqslant O(1)$$

只要 $n \geqslant \widetilde{\Omega}\Big(\frac{d\log(\sigma\surd)}{\varepsilon}\Big)d$。 此外，通过 w_{out} 集中于真实贝叶斯后验均值$\mathbb{E}[w^*|X\circ, y^\circ]$，对于 $\mathbb{E}\|\widehat{w}-\mathbb{E}[w^*|X\circ, y^\circ]\|_2^2$，同样的界（除低阶项外）也成立。

定义 C.12 ((α, τ, T)-可认证回归器)。 固定 $R>0$和 $\alpha, \tau>0$。对于输入 $(X_{\text{in}}, y_{\text{in}})$，向量$\widetilde{w}\in\mathbb{R}^d$是 (α,τ,T)-可认证的，当且仅当存在一个关于不定元 (X, y, ξ, w)的次数为 $O(1)$的线性泛函 L（一个伪期望），使得：

1. 接近/有界性： $\|L[w]-\widetilde{w}\|_2\leqslant\alpha$ 和 $\|L[w]\|_2\leqslant 2R+\tau T$。
2. 近似可行性（六阶）： L τ-近似满足约束族的并集

$$\mathscr{A} = \underbrace{\mathscr{A}_{\text{corr}}}_{\textit{corruption}} \cup \underbrace{\mathscr{A}_{\text{res}}}_{\textit{design/residual stability}}$$

具体而言：

- $\mathscr{A}_{\text{相关}}$： $\xi\odot\xi=\xi, \|\xi\|_1\leqslant\eta n$ 和 $(1-\xi(i))X(i,\cdot)=(1-\xi(i))X_{\text{属于}}(i,\cdot), (1-\xi(i))y(i)=(1-\xi(i))y_{\text{属于}}(i)$ 对所有 i 均成立。
- $\mathscr{A}_{\text{剩余}}$： $\Big\|\frac{1}{n}XX^\intercal-I\Big\|_{\text{运算}}\lesssim\surd(d+\text{对数}(1/\beta))/n$ 的六阶~~SoS替代函数~~，以及分解 $y-X^\intercal w=z_1+z_2$，其中 $\|z_1\|_2^2\lesssim\eta n$ 对数$(1/\eta)+$对数$(1/\beta)$且 $\|z_2\|_\infty\lesssim$对数$(1/\eta)$。

我们取 $R=\Theta(\sigma\surd d)$。

引理 C.13 （低分集合局部化于w附近）。 设 $\eta_\star=R$。设

$$\alpha(\eta)^2 \lesssim \eta^2\log\tfrac{1}{\eta} + \eta\cdot\frac{d+\log(1/\beta)}{n}.$$

那么以下各项同时成立。

1. 完备性（内球）。若 $\|\widetilde{w}-w_{\text{out}}\|_2\leqslant\alpha(\eta)$则 $\mathscr{S}(\widetilde{w};\cdot)\leqslant\eta n$。
2. 可靠性（外球）。若 $\mathscr{S}(\widetilde{w};\cdot)\leqslant\eta_\star n$，则 $\|\widetilde{w}-w_{\text{out}}\|_2\lesssim\alpha(\eta_\star)$。

定理 *C.11* 的证明定理 *C.11*。 容易验证评分函数的拟凸性。为证明评分函数对 $\widetilde{w}\in\mathbb{B}_2^d(2R+n\tau+\alpha)$是良定义的，我们注意到可以令 $y=0$且 $X=0$。于是可以令 w为 $\widetilde{w}$在半径为 $2R+\tau T$的球上的投影。现在我们应用定理 4.19中的归约。特别地，这得到样本复杂度

$$n \gtrsim \max_{\eta\leqslant\eta'\leqslant 1}\frac{d\log\big((R+n\tau)/\alpha(\eta)\big)+\log(1/\beta_\eta)}{\varepsilon}.$$

证明至此完成。 □

C.4.2 对第 2 阶段的精细估计器进行隐私化

最后，我们 证明以下关于以隐私方式近似最小二乘估计量的定理 .

定理 C.14. 存在一个多项式时间的 ε-*DP* 算法，对于一个 η-污染样本 $(X_{\text{in}}, y_{\text{in}})$，输出 $\hat{w}$，满足

$$\mathbb{E}\left[\|\hat{w} - w^*\|_2^2\right] \leqslant O(\alpha^2)$$

只要 $n \geqslant \tilde{\Omega}\left(\frac{d \log(\sigma\sqrt{\ })}{\varepsilon}\right) d$. 此外，由于 w_{out} 集中于真实贝叶斯后验均值$\mathbb{E}[w^* \mid X_\circ, y^\circ]$附近，对于 $\mathbb{E}\|\hat{w} - \mathbb{E}[w^* \mid X_\circ, y^\circ]\|_2^2$，同样的界（忽略低阶项）也成立。

在本节中，我们将 算法 6.6隐私化。我们对回归的可认证后验均值作如下定义。

定义 C.15 ((α, τ, T)-回归的可认证后验均值)**.** 固定 $R > 0$ 和 $\alpha, \tau > 0$。对于输入 $(X_{\text{in}}, y_{\text{in}})$，若存在一个次数为 $O(1)$ 的线性泛函 L（一个伪期望），定义在不定元 $(X, y, \xi, w, w_{\text{out}}, M)$ 上，使得向量 $\tilde{w} \in \mathbb{R}^d$ 是 (α, τ, T)-可认证的：

1. 接近度/有界性：$\|L[w_{\text{out}}] - \tilde{w}\|_2 \leqslant \alpha$ 和 $\|L[w_{\text{out}}]\|_2 \leqslant 2R + \tau T$。
2. 近似 可行性（六阶）：L τ-近似满足约束族的并集

$$\mathscr{A} = \underbrace{\mathscr{A}_{\text{corr}}}_{corruption} \cup \underbrace{\mathscr{A}''_{\text{res}}}_{design/residual\ stability} \cup \underbrace{\mathscr{A}_{\text{LS}}}_{w_{LS} = \frac{1}{n} X y},$$

具体而言：

- $\mathscr{A}_{\text{相关}}$：$\xi \odot \xi = \xi$, $\|\xi\|_1 \leqslant \eta n$ 和 $(1 - \xi(i))X(i, \cdot) = (1 - \xi(i))X_{\text{中}}(i, \cdot)$, $(1 - \xi(i))y(i) = (1 - \xi(i))y_{\text{中}}(i)$ 对所有 i 均成立。
- $\mathscr{A}''res$：$\left\|\frac{1}{n} X X^\intercal - I\right\|_{op} \lesssim \sqrt{(d + log(1/\beta))/n}$ 的六阶SoS替代函数，以及分解 $y - X^\intercal w = z_1 + z2$，其中 $\|z_1\|_2^2 \lesssim \eta n\, log(1/\eta) + log(1/\beta)$ 且 $\|z_2\|_\infty^2 \lesssim log(1/\eta) + log(1/\beta)/\eta n$；此外，我们有 $X^\intercal(w^* - winit) = b_1 + b2$，使得 $\|b_1\|_2^2 \lesssim \eta n\, log(1/\eta) + log(1/\beta)$ 且 $\|b_2\|_\infty^2 \lesssim log(1/\eta) + log(1/\beta)/\eta n$。
- $\mathscr{A}_{\text{LS}}$：$wLS = \frac{1}{n} X y$。

我们选择 $R = \Theta(\sigma\sqrt{d})$。

定理（重述定理 6.16）。设 $\Sigma = \sigma^2 \mathrm{Id}_d$ 为已知参数 $\sigma = \omega(1/\sqrt{n})$。设$\beta \in [0, 1]$，$\varepsilon \leqslant O(1)$。存在一个多项式时间算法，给定 η受污染样本 (X, y)，其生成方式如定义 6.1所述，输出一个 ε-*DP* 估计量$\hat{w}$，使得其概率至少为 $1 - \beta$,

$$\|\hat{w} - \mathbb{E}[w^* \mid X^\circ, y^\circ]\|^2 \leqslant \alpha^2,$$

其中 (X°, y°) 是未污染样本，条件为 $\alpha \geqslant \eta$ 对数$(1/\eta)$ 且

$$n \geqslant \tilde{\Omega}\left(\frac{d + \log(1/\beta)}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{d + \log(1/\beta)}{\alpha\varepsilon} + \frac{d \log(\sigma\sqrt{d})}{\varepsilon} + \frac{d\eta^2}{\alpha^4}\right).$$

证明。 将从鲁棒性到隐私的归约应用于定义域 $\mathscr{S}$上的分数 $\mathbb{B}_2^d(2R + n\tau + \alpha)$。令 V_η表示 $\{\tilde{w}: \mathscr{S}(\tilde{w}) \leqslant \eta n\}$ 在该定义域内的体积。根据 引理 6.12：

• 局部区域 $\eta \leqslant \eta' \leqslant \eta_\star$：低分集合被夹在以 $w\mathrm{out}$ 为中心、半径分别为 $\alpha(\eta)$ 和 $\alpha(\eta')$ 的 ℓ_2-球之间，因此

$$\log \frac{V_{\eta'}}{V_\eta} \lesssim d \cdot \log \frac{\alpha(\eta')}{\alpha(\eta)}.$$

• 全局情形 $\eta_\star \leqslant \eta' \leqslant 1$：评分函数-$\leqslant \eta_\star n$ 集合位于一个$O(\alpha(\eta_\star))$-球内，而环境半径为$O(R + n\tau)$，因此

$$\log \frac{V_{\eta'}}{V_\eta} \lesssim d \cdot \log \frac{R + n\tau}{\alpha(\eta_\star)}.$$

最后，容易验证评分函数是拟凸的，且可以高效计算。定理 4.19 随后给出一个充分条件

$$n \gtrsim \max_{\eta \leqslant \eta' \leqslant \eta_\star} \frac{d \log(\alpha(\eta')/\alpha(\eta)) + \log(1/\beta_\eta)}{\varepsilon \eta'} + \max_{\eta_\star \leqslant \eta' \leqslant 1} \frac{d \log\big((R + n\tau)/\alpha(\eta_\star)\big) + \log(1/\beta_\eta)}{\varepsilon}.$$

在小-η 情形下，使用 $\alpha(\eta)^2 \asymp \eta \cdot d + \log^{(1/\beta)} n$ 对 η 进行优化，可得到两个局部项$d + \log^{(1/\beta)} n$ 和$d + \log^{(1/\beta)}_{\alpha\varepsilon}$。对于全局项，取$R = \Theta(\sigma\sqrt{d})$，使得其贡献为$d \log(\sigma\sqrt{d}/\varepsilon^{4/3}\varepsilon^{2/3})$（省略多项式对数因子）。

□