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

Gavin Brown[*] GRBROWN@BU.EDU
Samuel B. Hopkins[†] SAMHOP@MIT.EDU
Adam Smith[*] ADS22@BU.EDU

arXiv:2301.12250v2 [cs.LG] 25 Apr 2023

摘要

我们提出了一种快速的、满足差分隐私的高维协方差感知均值估计算法，其样本复杂度近乎最优。此前只有指数时间估计器能够达到这一保证。给定来自均值 μ 未知、协方差 Σ 未知的（次）高斯分布的 n 个样本，我们的（ε，δ）差分隐私估计量生成 $\tilde{\mu}$，使得 $\|\mu-\tilde{\mu}\|_\Sigma \le \alpha$，只要 $n \gtrsim \frac{d}{\alpha^2} + \frac{d\sqrt{\log 1/\delta}}{\alpha\varepsilon} + \frac{d\,\text{对数}\,1/\delta}{\varepsilon}$。马氏误差度量 $\|\mu-\hat{\mu}\|_\Sigma$衡量 $\hat{\mu}$ 与 μ 之间相对于 Σ 的距离；它刻画了样本均值的误差。我们的算法运行时间为 $\tilde{O}(nd^{\omega-1}+nd/\varepsilon)$，其中 $\omega < 2.38$ 是矩阵乘法指数。

我们采用 Brown、Gaboardi、Smith、Ullman 和 Zakynthinou（2021）的指数时间方法，提出高效的稳定均值和协方差估计子程序变体，并将样本复杂度进一步降低至上述近乎最优的界。

我们的稳定协方差估计器可用于对无约束次高斯分布进行隐私协方差估计。使用 $n \gtrsim d^{3/2}$ 个样本时，我们的估计在谱范数下是准确的。这是首个使用 $n = o(d^2)$个样本的此类算法，回答了 Alabi 等人（2022）提出的一个开放问题。使用 $n \gtrsim d^2$ 个样本时，我们的估计在弗罗贝尼乌斯范数下是准确的。这带来了一个快速且近乎最优的算法，可在全变差距离下对无约束高斯分布进行隐私学习。

Duchi、Haque 和 Kuditipudi（2023）独立且同期得到了类似结果 .

目录

1 引言 1.1 Our Tech 法.1.2 与同期结果的比较.1.3 相关工作. **2** 4 8 9

2 主要结果与分析。 2 1.2.2 准确性分析.1 2.3 运行时间.1 1 隐私分析 **10** 1 4 7

3 稳定协方差估计 3.1 Families 3.2 权重良好（引理24的证明）.23 3.3 权重稳定（引理25的证明）.24 3.4 SCORE 具有低敏感度.24 **19** 最大 Goodness

4 StableMean 均值估计 26 4.1 最大核心族.29 4.2 权重良好（引理 37 的证明）.30 4.3 权重是稳定的（引理 38 的证明）.30 4.4 分数具有低敏感度.31

5 隐私协方差估计与高斯分布的快速学习 5.1 讨论 ion of Private Covariance Estimation . . . . . 5.2 定理 42 的证明. **32** 33 34

A 预备知识 38 A.1 记号与初等事实.38 A.2 次高斯随机变量与集中不等式 . . . 38 A.3 差分隐私.39

B 延后证明：可识别性与敏感度 39

[*] 波士顿大学计算机科学系。获得美国国家科学基金会 CNS-2120667 和 CCF-1763786 号奖项以及 Apple Faculty Award 的支持。
[†] 麻省理工学院电气工程与计算机科学系。部分获得 MLA@CSAIL 计划资金以及美国国家科学基金会 CCF-2238080 号奖项的支持。

© G. Brown、S.B. Hopkins 和 A. Smith。

1. 引言

我们考虑这样一项统计任务：在差分隐私约束下，根据独立样本估计高维子高斯分布的均值。差分隐私使算法设计者能够控制并分析统计学和机器学习中的隐私损失，并且是保护个人数据隐私的标准方法，目前已被学术界、工业界和美国政府采用。我们聚焦于以下均值估计问题的协方差感知版本。

问题1（协方差感知均值估计）给定来自均值 μ 未知且协方差为 Σ的次高斯分布 D 在 $\mathbb{R}^d$ 上的 n 样本，输出$\hat{}\mu$，使得 $\|\hat{}\mu-\mu\|_\Sigma \le \alpha$，其中$\hat{}\mu-\mu\|_\Sigma$ 表示 $\|\Sigma^{-1/2}\left(\hat{}\mu-\mu\right)\|_2$。

马氏距离 $\|\hat{}\mu-\mu\|_\Sigma$是衡量均值估计量统计精度的自然仿射不变方式。马氏误差较小的估计量会自动适应位置 μ中不确定性较低的方向——在任意特定方向上，估计量的误差都会按照该方向上的方差进行缩放。次高斯分布族是对高斯分布的推广；定义暂且留待后文。

在不考虑隐私约束的情况下，经验均值已经以 $n \gtrsim d/\alpha^2$ 个样本达到了问题1所能获得的最佳保证。（此处 $\gtrsim$ 隐去了常数。）然而，它通常无法保护其输入数据——例如，一个对该分布有粗略了解的对手，仅通过观察经验均值，就能轻易判断数据集中是否在某个特定方向上包含极端异常值。因此，我们旨在构造具有差分隐私的估计量：

定义*2*（差分隐私（*DP*））设 $\mathcal{X}$ 和 $\mathcal{Y}$ 为集合。随机算法 $A:\mathcal{X}^n \to \mathcal{Y}$ 满足 (ε,δ)-差分隐私（*DP*），如果对于任意 $x=(x_1,\dots,x_n)\in\mathcal{X}^n$ 和 $x'=(x'_1,\dots,x'_n)\in\mathcal{X}^n$，只要 x 和 x' 除一个坐标外均相同，$A(x)$ 和 $A(x')$ 就是 (ε,δ)-不可区分的（记作 $A(x)\approx_{(\varepsilon,\delta)}A(x')$）；也就是说，对于任意事件 $Y\subseteq\mathcal{Y}$，

$$\Pr[A(x)\in Y] \le e^{\varepsilon}\Pr[A(x')\in Y]+\delta\,.$$

差分隐私估计量提供了一项强有力的保证：无论外部对手事先掌握有关数据集的哪些信息，他们从任何特定个体 Alice 身上了解到的信息大致相同，无论 *Alice* 的数据是否用于计算（例如，参见 Kasiviswanathan 和 Smith（2014））。我们将只在一条记录上存在差异的数据集称为相邻数据集。

直到最近，许多研究人员都认为，受差分隐私约束的协方差感知均值估计至少需要与私有估计协方差一样多的样本，而后者在非隐私条件下也需要 $O(d)$ 个样本；对于差分隐私估计量，则需要 $\Omega(d^{3/2})$ 个样本（Kamath 等人（2022））。

Brown、Gaboardi、Smith、Ullman 和 Zakynthinou（2021）（下文简称“BGSUZ”）通过提出对指数机制的改进，推翻了这一假设；只要满足以下条件，该机制就能解决问题1（针对高斯数据）：

$$n \gtrsim \underbrace{\frac{d}{\alpha^2}}_{\substack{\text{nonprivate}\\ \text{sample}\\ \text{complexity}}} + \frac{d}{\alpha\varepsilon} + \frac{\log 1/\delta}{\varepsilon} \tag{1}$$

这个表达式省略了一个绝对常数和一个加性项 log(1/εα)/(εα)。即使协方差已知，它也几乎是紧的：公式（1）中的第一项是非私有估计所必需的，第二项几乎匹配 Kamath 等人（2019）给出的 Ω(d/(αε)log(d)) 下界，而第三项即使在估计均值为无约束均值、方差为单位方差的一维高斯分布时也是必需的。后续研究中，Liu、Kong 和 Oh（2022）将基于指数机制的方法扩展到了许多统计任务；具体而言，他们以相同的近最优样本复杂度解决了针对一般次高斯分布的问题1。不幸的是，所有这些估计量似乎都至少需要指数时间才能计算。

我们的主要结果是一个用于问题1的多项式时间算法，其样本复杂度还线性依赖于 d：

The定理3（非正式表述，参见定理8）算法 *1* 满足(ε, δ)-差分隐私。给定

$$n \gtrsim \frac{d}{\alpha^2} + \frac{d\sqrt{\log 1/\delta}}{\alpha\varepsilon} + \frac{d \log 1/\delta}{\varepsilon}$$

来自定义在 $\mathbb{R}^d$ 上、均值为 μ 且协方差为 Σ的次高斯分布的样本，以高概率，它会输出 $\tilde{\mu}$ 使得 $\|\tilde{\mu} - \mu\|_\Sigma \leq \alpha$。其运行时间为 $\tilde{O}\big(nd^{\omega-1} + nd/\varepsilon\big)$，其中 $\omega < 2.38$ 是矩阵乘法指数。

当 d 相对于 $1/\varepsilon$较大时，我们的算法的运行时间主要由计算协方差矩阵 $nd^{\omega-1}$所需的时间决定。我们的算法的样本复杂度中的项 $d\sqrt{\log\ 1/\delta}/(\alpha\varepsilon)$ 和 $d\log(1/\delta)/\varepsilon$ 略微次优；然而，在适中的隐私参数下，非私有样本复杂度仍占主导。

如下面将要讨论的，我们的主要算法依赖于一种新的非隐私协方差估计器 StableCovariance。该估计器与 Alabi 等人（2022）的“高斯采样机制”结合后，可得到一种快速的隐私协方差估计算法，并在谱范数和弗罗贝尼乌斯范数下均提供强误差保证。对于无约束次高斯分布，该算法是首个以 $n = o(d^2)$ 个样本实现低谱范数误差的隐私算法。它还以 $O(d^2)$ 个样本恢复了与 Ashtiani 和 Liaw（2021）类似的弗罗贝尼乌斯范数估计保证；这种误差保证结合我们的均值估计保证，已知足以在低全变差距离下学习无约束高斯分布。

定理 4（非正式表述，参见定理 42）算法 *7* 满足(ε, δ)-差分隐私，并返回一个~向量 $\tilde{\mu}$ 和矩阵 Σ。假设它接收从一个定义在 $\mathbb{R}^d$ 上、均值为 μ 、协方差为 Σ 的次高斯分布中独立同分布抽取的 n 个样本。

- 若 $n \gtrsim \frac{d}{\alpha^2} + \frac{d^{3/2}\sqrt{\text{对数}}}{\alpha\varepsilon}\ \frac{1/\delta}{} + \frac{d\text{对数}}{\varepsilon}^{1/\delta}$，则以高概率有 $\left\|\Sigma^{-1/2}\tilde{\Sigma}\Sigma^{-1/2} - \mathbb{I}\right\|_2 \leq \alpha$。
- 若 $n \gtrsim \frac{d^2}{\alpha^2} + \frac{d^2\sqrt{\log 1/\delta}}{\alpha\varepsilon} + \frac{d\text{对数}\ 1/\delta}{\varepsilon}$, *then,* 以高概率 , $\left\|\Sigma^{-1/2}\tilde{\Sigma}\Sigma^{-1/2} - \mathbb{I}\right\|_F \leq \alpha$;

~此外，若分布为高斯分布，则 总变差距离$(\mathcal{N}(\mu, \Sigma)\ \mathcal{N}(\tilde{\mu}, \Sigma)) = O(\alpha)$。其运行时间为 $\tilde{O}(nd^{\omega-1} + nd/\varepsilon)$，其中 $\omega < 2.38$ 是矩阵乘法指数。

在全文中，我们的准确性分析均假设底层分布具有满秩协方差矩阵。对于不满足此假设的分布，存在样本效率和时间效率兼具的预处理算法（Singhal and Steinke（2021年）；Kamath 等人（2021年）；Ashtiani 和 Liaw（2021年））。我们的隐私分析不对数据作任何假设。

在本引言的其余部分，我们将概述所采用的技术，并综述相关工作。

1.1. 我们的技术

背景：经验协方差方法 除了对指数机制进行修改之外，BGSUZ 还给出了另一种均值估计算法，该算法同样需要指数时间，适用于次高斯分布，但对隐私参数的依赖更差。这一算法是我们构造高效算法的基础。

该算法的基本思路是，首先解决“良好”数据集 y 上的问题；这里的良好是指满足一组次高斯数据通常具备的条件。例如，良好数据集 y 不应包含显著的异常值，这可通过 $\|(\cdot) - \mu_y\|_{\Sigma_y}$ 来衡量，其中 μ_y 和 Σ_y 分别是 y的经验均值和协方差矩阵。BGSUZ 表明，如果能将输入限制为良好数据集，那么发布从 $\mathcal{N}(\mu_y, \sigma^2\Sigma_y)$ 中抽取的单个样本的机制（对于 $\sigma \approx \frac{\sqrt{d}}{\varepsilon n}$）实际上满足差分隐私。依赖于数据选择协方差 Σ_y 对于实现马氏距离保证至关重要，因为它确保隐私保护所添加的噪声在 y 自身方差较小的方向上仍然较小。

给定一个数据集 x，BGSUZ 的算法首先将 x 投影到最近的“良好”数据集 y，然后检查 y与 x 的距离是否过大；最后，如果该检验通过，则发布 $\tilde{\mu} \sim \mathcal{N}(\mu_y, \sigma^2\Sigma_y)$。投影步骤确保隐私对所有数据集都成立。

尽管添加噪声在计算上是高效的，但 BGSUZ 算法在投影步骤中需要对数据集进行穷举搜索。此外，它还会带来额外的样本复杂度开销——大致需要 $n \gtrsim \frac{d}{\alpha^2} + \frac{d}{\alpha\varepsilon^{2\circ}}$ 与公式（1）相比，$\frac{1}{\varepsilon}$ 的额外因子源于投影并不稳定：两个相邻的数据集 x和 x' 可能被投影到彼此相距很远的“良好”数据集 y 和 y'。

我们方法概述 在本文中，我们克服了经验协方差方法的关键计算瓶颈；与此同时，我们恢复了对 ε的最优依赖关系。

我们的主要技术工具是新型高效的非私有估计量 StableCovariance(x) 和 StableMean(x)。我们还设计了算法 SCORE(x)，用于近似从 x 到最近“良好”数据集的汉明距离。我们证明，对于相邻的 x, x'，|SCORE(x) −SCORE(x')| ≤ 2，并且如果 SCORE(x) 很小——对于次高斯 x，这一点以高概率成立——那么 StableCovariance(x) $\approx$ StableCovariance(x') 且 StableMean(x) $\approx$ StableMean(x')。

总体而言，对于输入 x，我们的算法首先私下测试 SCORE(x) 是否小于某个阈值，约为 $1/\varepsilon$。如果测试失败，我们就停止并不返回任何内容（如同 Dwork 和 Lei（2009）的“PTR”框架）。如果测试通过，我们就从 $\mathcal{N}$(StableMean(x)$_{-}$ σ^2 ·StableCovariance(x)) 中输出一个样本，其中 $\sigma \approx \frac{\sqrt{d}}{\varepsilon n}$ 是之前所述的尺度参数。 —

StableMean(x) 和 StableCovariance(x) 分别取代 BGSUZ 算法中投影 y 的经验参数 μ_y 和 Σ_y。它们的优势在于计算可处理性和稳定性：StableMean 和 StableCovariance 比它们在 BGSUZ 中的对应项更稳定，因此我们对 ε 的依赖达到了最优。我们证明了 StableCovariance 的以下关键性质（StableMean 的思路类似）：

- 准确性：当 x_i 独立同分布地取自次高斯分布时，以高概率有 StableCovariance(x) $= \Sigma_x$，其中 Σ_x 是 x的经验协方差。[1]

1. 实际上，“配对”的经验协方差 $\frac{2}{n}\sum_{i \leq n/2}(x_i - x_{i+n/2})(x_i - x_{i+n/2})^T$。

• 稳定性：如果 x 是一个满足 SCORE(x) $\lesssim 1/\varepsilon$的数据集，那么对于任意相邻的 x'，矩阵$\Sigma_1 =$ StableCovariance(x) 和 $\Sigma_2 =$ StableCovariance(x') 很接近。具体而言，

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}_d\right\|_{\mathrm{tr}}, \left\|\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - \mathbb{I}_d\right\|_{\mathrm{tr}} \le O(d/n),$$

其中，$\|\cdot\|_{\mathrm{tr}}$ 表示特征值绝对值之和。

• 效率：StableCovariance 可在多项式时间内运行。

稳定协方差 现在我们概述 StableCovariance。暂且假设我们有一个数据集 x，并认为它来自某个未知 Σ 对应的 $\mathcal{N}(0, \Sigma)$。（我们通过标准的样本配对技巧将问题转化为零均值情形，见方程 (3)。）最终，StableCovariance(x) 会生成一个权重向量 $w \in [0, 1/n]^n$，并输出 $\Sigma_w = \sum_{i\in[n]} w_i \cdot x_i x_i^T$。我们将证明这些权重满足以下条件。

(a) 对次高斯输入均匀：当 x_i 独立同分布地来自次高斯分布时，以高概率有 $w = (1/n, \ldots, 1/n)$。

(b) λ-良好加权：对于任何 $i \in [n]$ 使得 $w_i > 0$的情况，我们有 $\left\|\Sigma_w^{-1/2} x_i\right\|_2^2 \le \lambda$。

(c) 稳定性：如果 x 是一个满足 SCORE(x) $\lesssim 1/\varepsilon$的数据集，那么对于任意相邻的 x'，相应的权重向量 w，w' 具有 $\|w - w'\|_1 \le O(1/n)$。

这里 $\lambda > 0$ 是一个超参数，我们最终会将其设置为大约 d。实现性质 (a) 意味着我们关于 StableCovariance 的准确性声明成立，因为将所有权重设置为 $1/n$ 后，得到的正好是 x的二阶矩矩阵。由于我们假设 x 的均值为零，因此该矩阵可代替经验协方差。

性质 (b) 和 (c) 共同保证了 StableCovariance 的参数估计具有稳定性。直观地说，考虑 w 和 w' 恰好只有一个取值不同的情形：假设 $w'_i = w_i + \eta_i$ 对于 $\eta_i > 0$。进一步假设两个权重均非零，且 x 和 x' 在索引 i 上相同。因此，我们有 $\Sigma_{w'} = \Sigma_w + \eta_i \cdot x_i x_i^T$。

$$\begin{aligned}\Sigma_w^{-1/2}\Sigma_{w'}\Sigma_w^{-1/2} - \mathbb{I} &= \Sigma_w^{-1/2}\left(\Sigma_w + \eta_i \cdot x_i x_i^T\right)\Sigma_w^{-1/2} - \mathbb{I} \\ &= \eta_i \cdot \Sigma_w^{-1/2} x_i x_i^T \Sigma_w^{-1/2}.\end{aligned}$$

这是一个秩一矩阵，其唯一的非零特征值恰好为 $\eta_i \|x_i\|_{\Sigma_w}^2$，根据性质 (b)，该值至多为 $\eta_i\lambda$。

稳定性证明的其余部分相对直接：一般而言，$\Sigma_w - \Sigma_{w'}$ 是秩一矩阵之和，其系数 η_i 满足$\sum_i |\eta_i| = \|w - w'\|_1$。迹范数等于特征值绝对值之和，因此我们得到大致为 $\lambda\|w - w'\|_1 \approx d/n$的上界。证明还需要进一步处理一个索引 i^*，其中$x_{i^*} \neq x'_{i^*}$，同时还要处理 w_i 为零而 w'_i 不为零的索引（在这种情况下，根据假设，$\|x_i\|_{\Sigma_w}^2$ 没有界）。

Good 子集 现在我们讨论 StableCovariance 如何计算权重向量 w。

定义 5 λ-良好加权的一个特例是一个 λ-良好子集。若子集 $S \subseteq [n]$ 所关联的权重向量 w_S 是 λ-良好的，则该子集 $S \subseteq [n]$ 是 λ-良好的，其中 $(w_S)_i = 1/n$ 对于 $i \in S$ 成立，而在其他情况下为0 。

StableCovariance 计算一个良好子集的嵌套序列 $S_0 \subseteq S_1 \subseteq \ldots \subseteq S_{2k}$，其中$S_\ell$对于某个异常值阈值序列 $\lambda_1 \leq \ldots \leq \lambda_{2k}$是 λ_ℓ良好的，而 k是一个稍后设定的参数。随后，它对相关权重向量中的一个子集取平均，以生成 $w = \frac{1}{k}\sum_{\ell=k+1}^{2k} w_{S_\ell}$。它选择进行平均的子集由以下引理描述：

引理 6（非正式表述）设 x 是一个数据集， $\lambda > 0$ 是一个异常值阈值。存在唯一最大的 λ-良好子集 $S \subseteq [n]$ ，它对应于 x ，并包含所有其他 λ-良好子集。此外，可以使用贪心算法 `LargestGoodSubset`找到这个最大的良好子集。

StableCovariance 将 S_ℓ 取为离群值阈值 λ_ℓ 对应的 LargestGoodSubset。根据引理 6，可立即推出 $S_0 \subseteq \ldots \subseteq S_{2k}$，因为 S_ℓ 对所有 ℓ 均为 $\lambda_{\ell+1}$-良好的。

出于我们将在下文看到的原因，我们根据 $\lambda_{\ell+1} = e^\varepsilon \lambda_\ell$ 选择异常值阈值，并设定 $k \approx 1/\varepsilon$，从而使 $\lambda_{2k} \leq O(\lambda_0)$。现在，我们已经几乎完整地描述了 StableCovariance，也掌握了足够的信息，可以看出它计算出的权重为何满足上文的性质 (a) 和 (b)。我们将确保，如果 $[n]$ 本身是一个良好子集，那么 $S_0 = \ldots = S_{2k} = [n]$，由此得到上文的性质 (a)。由于子集 S_ℓ 全都是 $O(\lambda_0)$-良好的，因此不难证明 w 本身也是$O(\lambda_0)$-良好的，由此得到上文的性质 (b)。

现在我们转向性质（c），它取决于 x， x'是否具有 SCORE(x)，SCORE(x') $\lesssim 1/\varepsilon$。接下来，我们需要定义 SCORE(x)。

评分函数 我们已经计算出 S_ℓ，即 x 的最大 λ_ℓ-良好子集，适用于 $\ell = 0$、1、…、$2k$。StableCovariance(x) 对 $\ell = k+1$、…、$2k$ 中的项目取平均值，而 SCORE(x) 使用序列的前半部分：

$$\texttt{SCORE}(x) \stackrel{\text{def}}{=} \min_{0 \leq \ell \leq k} \left(n - |S_\ell| + \ell\right).$$

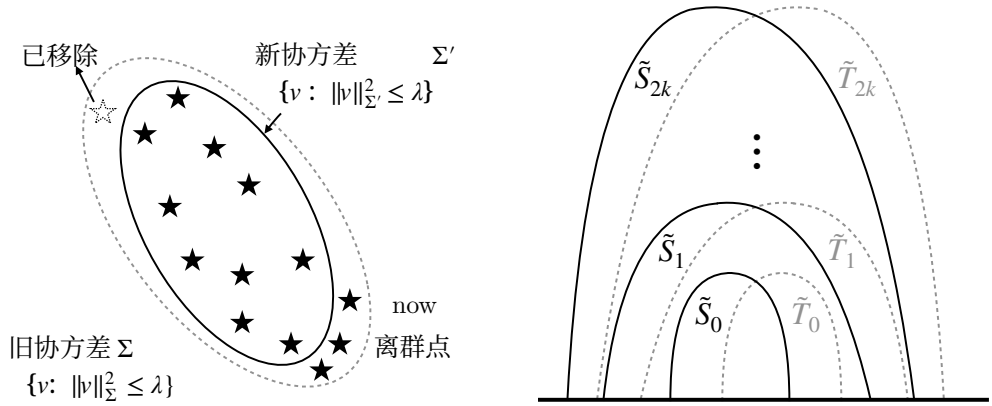
正如我们将在下文讨论的，该函数的敏感度为 2：相邻数据集具有大小和离群值阈值相近的良好子集。我们的算法最终使用 SCORE′(x) $= \min\{k, \text{SCORE}(x)\}$，这使得一些表述更简洁。

直观地说，如果 $[n]$ 相对于离群值阈值 λ_0、…、λ_k 是良好的，那么 SCORE(x) $= 0$。结合 SCORE(x) 的稳定性，这意味着 SCORE(x) 大致反映了 x 与某个数据集之间的汉明距离，而在该数据集中， $[n]$ 是一个良好子集。

相邻数据集上的良好子集 为了为我们刚才描述的算法建立性质（c），我们首先考察如果从数据集 x 中移除单个样本， λ-良好子集 S 会发生什么。

如果我们从 x 中移除样本 x_i，其中 $i \notin S$，那么当然有 $S \setminus \{i\} = S$，它对于 $x \setminus \{x_i\}$ 仍然是 λ-良好的。另一方面，如果我们移除 x_i（其中 $i \in S$），所得的 $S \setminus \{i\}$ 对于 $x \setminus \{x_i\}$ 可能不再是 λ-良好的；其中可能包含许多异常值！图 1（a）对此进行了说明。不过，对于我们所考虑的参数范围，可以证明 $S \setminus \{i\}$ 对于 $x \setminus \{x_i\}$ 仍然是 λ'-良好的，其中 $\lambda' = e^\varepsilon \lambda$。

这使我们能够推理在移除和添加的情况下的数据集对，这正是为了处理相邻 x， x'而必须考虑的概念。如果 S 是 x 的一个 λ-良好子集，且数据集 x' 与其仅在索引 i^*处不同，那么 $S \setminus \{i^*\}$ 就是 x'的一个 $e^\varepsilon\lambda$-良好子集。



(*a*) 从良好子集中移除一个点　　(*b*) 交织：　　$\tilde{S}_\ell \stackrel{\text{def}}{=} S_\ell \setminus \{i^*\}$

图 1：(*a*) 一个良好子集：相对于经验协方差，没有点是异常值。移除一个点可能会改变协方差（从 Σ 变为 Σ'），并导致一些点成为（轻微）异常值。(*b*) 令 S_ℓ 表示 x 的最大 λ_ℓ-良好子集，令 T_ℓ 表示 ~相邻 x'的对应集合。令 S_ℓ 表示 $S_\ell \setminus \{i^*\}$，其中 x, x' 在索引 i^* 处不同（ T_ℓ也同样如此）。~ ~ ~ ~这些集合相互交织：对于所有 ℓ，都有 $S_\ell \cup T_\ell \subseteq S_{\ell+1} \cap T_{\ell+1}$。

相邻数据集上的最大良好子集族现在让我们回到针对输入 x、使用离群值阈值 λ_0，……，λ_{2k} 计算得到的最大良好子集族 S_0，……，S_{2k}，以及针对与 x相邻的输入 x'得到的对应子集族 T_0，……，T_{2k}。我们将证明 S_0，……，S_{2k}与 T_0，……，T_{2k}之间存在一种关键的“交织性”关系；其示意图见图1(*b*)。

引理 7（交织性，非正式）假设 x、x' 在索引 i^* 处不同。对于所有 ℓ，定义 $\tilde{S}_\ell \stackrel{\text{定义}}{=} S_\ell \setminus \{i^*\}$ 和 $\tilde{T}_\ell \stackrel{\text{定义}}{=} T_\ell \setminus \{i^*\}$，我们知道 $\tilde{S}_\ell \cup \tilde{T}_\ell$ 是 $\tilde{S}_{\ell+1} \cap \tilde{T}_{\ell+1}$ 的子集。

证明 [概要] 根据上文的讨论，$S_\ell \setminus \{i^*\}$ 是 x 和 x' 的一个 $\lambda_{\ell+1}$-良好子集（假设$\lambda_{\ell+1} \le e^\varepsilon \lambda_\ell$，这将由构造保证）。此外，根据引理 6，$T_{\ell+1}$（ x'的最大 $\lambda_{\ell+1}$-良好子集）包含 x'的所有 $\lambda_{\ell+1}$-良好子集；对 $S_{\ell+1}$ 也同理。因此，$T_{\ell+1}$，$S_{\ell+1} \supseteq$~ ~ ~ ~ ~S_ℓ。而且，根据定义，S_ℓ 不包含 $\{i^*\}$，所以 $T_{\ell+1}$，$S_{\ell+1} \supseteq S_\ell$。对称的论证~同样适用于$T_\ell$。

■

现在我们概述证明交织性蕴含性质（c）的主要思路，即 $\|w - w'\|_1 \le O(1/n)$，其中 w 是与输入 x 相关联的权重，而 w' 是与相邻x'相关联的权重。由于 SCORE(x),SCORE(x') $< k \approx 1/\varepsilon$，存在 $\ell < k$ 使得 $|S_\ell| \ge n - k$；又由于$S_\ell \subseteq S_{k+1}$，也有 $|S_{k+1}| \ge n - k$。根据嵌套性质 $S_\ell \subseteq S_{\ell+1}$，对于 T_ℓ 也同样如此，我们有

$$\left| \bigcap_{\ell=k+1}^{2k} S_\ell \cap T_\ell \right| \ge n - O(k)$$

根据定义，

$$|w_i - w'_i| = \frac{1}{kn} \left| \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\} - \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in T_\ell\} \right|. \tag{2}$$

对于 $i \in \bigcap_{\ell=k+1}^{2k} S_\ell \cap T_\ell$，公式（2）中的差值为 0。因此，只剩下 $O(k)$ 个索引 i 其中 $|w_i - w_i'|$ 可能非零。我们将其分为两种情况。

如果 $i = i^*$，公式（2）中的两项求和最多相差 k，因此 $|w_{i^*} - w_{i^*}'| \le 1/n$。

对于 $i \neq i^*$，根据交织性质（引理 7），我们知道，如果 $w_i \neq w_i'$，那么只存在一个 ℓ 使得 $\mathbb{1}\{i \in S_\ell\} \neq \mathbb{1}\{i \in T_\ell\}$。因此，对于这样的 i，我们有 $|w_i - w_i'| \le 1/$（kn），而这些项的总贡献为 O（$1/n$）。至此，我们对 StableCovariance 的证明思路概述完毕。

从 StableCovariance 到均值估计 现在回到均值估计任务：给定 StableCovariance(x) 的输出 Σ，估计均值 $\hat{}$。一种自然的策略是抽取 n 个新样本x_i，并令 $\tilde{x}_i = \hat{\Sigma}^{-1/2} x_i$，然后将这些样本交给一个非协方差感知型的私有均值估计算法。如果 Σ 确实是隐私的，那么根据组合性质，该算法将满足差分隐私。然而，StableCovariance 的稳定性保证（而非隐私保证）无法转化为该方案整体的隐私保证。

相反，我们的算法在第二阶段计算稳定均值，并向其中添加噪声。我们的估计量采用了 StableCovariance 中的相同思想：对不包含离群点的大型数据子集进行平均；但它采用了不同的“异常值”概念，该概念取决于点之间的两两距离（相对于估计协方差 Σ）。其思路类似于近期隐私估计方法：寻找由彼此都足够接近的大量点（或“核心”）组成的集合（Tsfadia 等人（2022）；Ashtiani 和 Liaw（2021））。

超越均值估计：隐私协方差估计与高斯分布学习 我们的隐私证明的一个关键部分是 StableCovariance 的稳定性：对于相邻输入 x 和 x'，我们要么失败，要么生成协方差估计 $\Sigma_1 \leftarrow$ StableCovariance(x) 和$\Sigma_2 \leftarrow$StableCovariance(x')，使得 $\mathcal{N}(0, \Sigma_1) \approx_{(\varepsilon,\delta)} \mathcal{N}(0, \Sigma_2)$。我们的主要算法将这一点与稳定均值估计过程相结合，但仅凭这一点，它就能产生一个差分隐私算法：计算 $\hat{\Sigma} \leftarrow$ StableCovariance(x)，如果没有失败，则返回 $z \sim \mathcal{N}(0, \hat{\Sigma})$。该算法的实用性值得怀疑：$z$ 不包含关于 x 均值的信息，并且只包含很少的协方差信息。然而，正如 Alabi 等人（2022）最近观察到的，多个此类样本很有用：给定私有样本 $z_1, \ldots, z_N \overset{\text{iid}}{\sim} \mathcal{N}(0, \hat{\Sigma})$，可以对其进行后处理，以生成私有协方差估计 $\tilde{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} z_i z_i^T$。我们的主要分析表明，当 $n \gtrsim d$ 时，我们可以抽取 $N = 1$样本；通过高级组合，当 $n \gtrsim d\sqrt{N}$ 时，我们可以抽取 N 样本。Alabi 等人（2022）将其称为高斯采样机制；他们将其与基于平方和框架的稳定估计量相结合，从而在多项式时间内对高斯分布的协方差进行隐私且鲁棒的学习。将其与 StableCovariance 相结合，可以得到一种快速的隐私协方差估计算法，并在谱范数和弗罗贝尼乌斯范数下都具有强误差保证，尤其实现了两者最优的维度依赖。

1.2. 与同期结果的比较

Duchi 等人（2023年）（下文简称“DHK”）独立地、同期地提出了一种私有均值估计算法，其运行时间和准确性保证与我们的相似。两篇论文都通过设计稳定均值和协方差估计器，补充完善了 BGSUZ 的基本框架。DHK 的稳定协方差估计器与我们的思路相似：迭代移除异常值并更新经验协方差。不过，两者的分析和算法细节大不相同。DHK 的稳定均值估计器和我们的方法都考虑点与点之间的距离（先按稳定协方差进行重缩放）。但是，DHK 的算法考察的是随机生成的小型

数据子集内部的距离。他们的算法和分析巧妙地利用了子集分配中的随机性，与我们的方法有很大不同。

在数据独立同分布地取自满秩次高斯分布这一基本设定下，我们的准确性保证略强一些。粗略地说，我们为实现隐私而添加的噪声造成的误差量级为 $\frac{d\sqrt{\log 1/\delta}}{\varepsilon n}$，而他们分析中的相应项为 $\frac{d}{\log \varepsilon n}\sqrt{1/\delta}$。此外，他们的算法需要 $n \gtrsim \frac{d}{\log \varepsilon^2}\sqrt{1/\delta}$ 个样本才能保证隐私，而我们的算法所需样本数在渐近意义下更少：$n \gtrsim \frac{d}{\log \varepsilon}\sqrt{1/\delta}$。

DHK 扩展了他们的算法和分析，以处理重尾分布或秩亏分布。类似的扩展很可能也适用于我们的算法。我们将稳定协方差估计器用于学习次高斯分布的协方差这一任务；DHK 的类似估计器也能发挥相同作用，并提供相近的保证。

1.3. 相关工作

本节概述了与在差分隐私约束下学习高维、无限制高斯分布和次高斯分布的均值和／或协方差最相关的研究。这些结果针对的是近似差分隐私；若没有参数的先验界，满足更强隐私概念（如纯差分隐私或集中式差分隐私）的算法无法达到准确估计。

在这些界之外，关于差分隐私统计的研究也相当丰富。Kamath 和 Ullman（2020）的入门读物介绍了差分隐私统计，并附有一篇综述。其中一项值得注意的工作是 Karwa 和 Vadhan（2017）的研究，他们建立了学习一元高斯分布的方法，其界与数据的位置和尺度无关（使用了 Dwork 和 Lei（2009）的工具）。另一项有影响力的工作是 Kamath、Li、Singhal 和 Ullman（2019）的研究，他们提出了用于学习高维高斯分布均值和协方差的算法，并满足集中式差分隐私。他们要求均值和协方差具有先验界，但其误差上界仅以多对数方式依赖于这些量。这两篇论文都证明了适用于我们所研究设定的下界。

我们利用非私有参数估计稳定性的方式并不常见。类似稳定性的性质，例如 Lipschitz 常数的界或“全局敏感度”，经常出现在差分隐私算法的设计中。通常的做法是论证某个函数具有稳定性，从而可以向该函数的输出添加少量噪声（如拉普拉斯机制和通常的高斯机制）。相比之下，我们利用 StableCovariance 的稳定性来论证，噪声本身（添加到 StableMean(x) 中）在相邻输入上不可区分。这个概念近期也以各种形式出现在其他研究中（Kothari 等人，2021年；Ashtiani 和 Liaw，2021年；Tsfadia 等人，2022年；Alabi 等人，2022年），不过我们应用这一概念的具体方式与之截然不同。

高斯均值估计 对于无限制高斯分布的私有协方差感知均值估计，其样本复杂度已得到充分理解，几乎匹配的上界（Brown 等人，2021）和下界（Karwa 和 Vadhan，2017；Kamath 等人，2019）均已确立。对于次高斯分布，也存在样本复杂度近乎最优的算法（Liu 等人，2022）；对于某些相关的分布族，已知的下界更强（Cai 等人，2019）。此前，已知的解决这一任务的方法要么需要指数时间，要么需要 $\Omega(d^2)$ 个样本，这对应于在谱范数下进行私有协方差学习的已知最佳样本复杂度。一

可以将这种近似用作预条件器：重缩放后，该分布将近似各向同性，从而可以逐维应用 Karwa 和 Vadhan（2017）的技术。

高斯协方差估计差分隐私高斯协方差估计近年来备受关注，尤其是因为它是隐私学习高斯分布并将全变差距离控制在较小范围内的关键一步。对于这一任务，样本复杂度已得到充分理解：在隐私参数适中时，下界（Vadhan，2017；Kamath 等人，2022年）与简洁的信息论上界（Aden-Ali 等人，2021年）均与非隐私情形下的样本复杂度相匹配。一系列不断改进的上界已经建立了几乎达到这些结果的多项式时间算法（Liu 等人，2022；Kothari 等人，2021年；Kamath 等人，2021年；Ashtiani 和 Liaw，2021年；Alabi 等人（2022）；Hopkins 等人，2022年）。更全面的概述（也讨论了鲁棒性和更强的隐私概念）请参阅 Alabi 等人（2022）和 Hopkins 等人（2022年）。

私有生成谱范数协方差近似（通常可用作预条件器）的任务进展较为缓慢。非私有情况下，该任务仅需 $n \approx d$ 个样本；我们的算法则需要 $d^{3/2}$ 个样本。最近已证明，在 $\alpha = O(1/\sqrt{d})$ 的情形下，这种对维度的依赖是最优的（Kamath 等人，2022年）。据我们所知，我们是首个针对无约束高斯分布、能够实现此类保证的差分隐私多项式时间算法，从而解决了 Alabi 等人（2022）提出的一个未解决问题。在假设 $\mathbb{I} \preceq \Sigma \preceq \kappa\mathbb{I}$ 的情况下，标准高斯机制需要 $\Omega(\mathrm{poly}(\kappa) \cdot d^{3/2})$ 个样本，而 Kamath 等人（2019）提出的私有预条件器需要 $\Omega(\mathrm{polylog}(\kappa) \cdot d^{3/2})$ 个样本（忽略对其他参数的依赖）。

2. 主要结果与分析

定理8（主定理） 固定 $\varepsilon, \beta \in (0,1)$，$\delta \in (0, \varepsilon/10]$，以及 $n, d \in \mathbb{N}$。算法 *1* 接收一个包含 n 个点的数据集，其中每个点属于 $\mathbb{R}^d$，隐私参数为 ε, δ，离群值阈值为 λ_0。

- 对于任意 $\lambda_0 \geq 1$，算法 *1* 均满足 (ε, δ)-差分隐私。
- 对于 $\mu \in \mathbb{R}^d$ 和正定 $\Sigma \in \mathbb{R}^{d \times d}$，令 $x = x_1, \ldots, x_n$ 从参数为 $K_{\mathcal{D}}$、均值为 μ 且协方差为 Σ 的次高斯分布 $\mathcal{D}$ 中独立同分布地抽取。存在绝对常数 K_1, K_2 和 K_3，使得当 $\lambda_0 = K_1 K_{\mathcal{D}}^2 (d + \log\, n/\beta)$ 且

$$n \geq K_2 K_{\mathcal{D}}^2 \cdot \frac{\log 1/\delta}{\varepsilon}\left(d + \log\left(\frac{K_{\mathcal{D}} \log 1/\delta}{\varepsilon\beta}\right)\right),$$

则以至少 $1 - \beta$ 的概率，算法 *1* 返回 $\hat{\mu}$，使得

$$\|\hat{\mu} - \mu\|_{\Sigma} \leq K_3 K_{\mathcal{D}}\left(\sqrt{\frac{d + \log 1/\beta}{n}} + \frac{d\sqrt{\log 1/\delta}}{\varepsilon n} + \frac{\log n/\beta \sqrt{\log 1/\delta}}{\varepsilon n}\right).$$

- 算法 *1* 可以实现为需要：一个形式为 $A^T A$ 的乘积，其中 $A \in \mathbb{R}^{n \times d}$；两个形式为 AB 的乘积，其中分别为 $A \in \mathbb{R}^{n \times d}$ 和 $B \in \mathbb{R}^{d \times d}$；对矩阵 $\mathbb{R}^{d \times d}$ 求逆，且其比特复杂度为 $\tilde{}$对数级；以及额外的计算开销为 $O(nd/\varepsilon)$。

我们通过分别证明以下各项子主张来证明定理8：引理15中的隐私性、引理19中的准确性，以及引理 20中的运行时间。

算法 1： 私有均值估计，$\mathcal{A}^{\varepsilon,\ \delta,\ \lambda_0}_{\text{主}}(x)$

输入： 数据集 $x \in \mathbb{R}^{n\times d}$；隐私参数 $\varepsilon,\ \delta$；离群值阈值 λ_0

要求： $n \geq \frac{192e^2\lambda_0 \text{对数}\, 6/\delta}{\varepsilon} + 160e^2\lambda_0$

/* 初始化 */

$k \leftarrow \left\lceil \frac{6\text{对数}\, 6/\delta}{\text{对数}\, \varepsilon} \right\rceil + 4$; $M \leftarrow 6k + \lceil 18\text{对数}\ 16n/\delta \rceil$; $c^2 \leftarrow \frac{720e^2\lambda_0 \text{对数}\, 12/\delta}{\varepsilon^2 n^2}$;

$R \sim$ 均匀分布$(\{R' \subseteq [n] :\ |R'| = M\})$;

/* 计算非私有参数估计 */

$\hat{\Sigma}$,SCORE$_1 \leftarrow$ StableCovariance(x, λ_0, k); /* $\hat{\Sigma} \in \mathbb{R}^{d\times d}$, SCORE$_1 \in \mathbb{N}$ */

$\hat{}\,\hat{\mu}$,SCORE$_2 \leftarrow$ StableMean$(x, \Sigma, \lambda_0, k, R)$; /* $\hat{\mu} \in \mathbb{R}^d$,SCORE$_2 \in \mathbb{N}$ */

/* 测试分数并发布 */

如果 $\mathcal{M}^{\varepsilon/3,\delta/6}_{\text{PTR}}(\max\{\text{SCORE}_1, \text{SCORE}_2\}) = $通过 **那么**

  返回 $\tilde{\mu} \sim \mathcal{N}(\hat{}\,\mu, c^2\hat{\Sigma})$;

else

  返回 失败

end

2.1. 隐私分析

设 x 和 x' 为相邻数据集。我们的主要算法，即算法 1，要么失败，要么计算非隐私估计（$\hat{}\,\mu$、$\hat{}\,\Sigma$），并针对某个实数 c 输出一个样本 $\tilde{\mu} \sim \mathcal{N}(\hat{}\,\mu,\ c^2\,\hat{}\,\Sigma)$。根据标准的提出-测试-发布式分析，只需证明：(i) 在 x 和 x' 下失败的概率是（$\varepsilon,\ \delta$）-不可区分的；(ii) 对于算法 1 在 x 上计算得到的任意一对（$\hat{}\,\mu$、$\hat{}\,\Sigma$）以及在 x' 上计算得到的（$\hat{}\,\mu'$、$\hat{}\,\Sigma'$），如果不失败，则

$$\mathcal{N}(\hat{\mu}, c^2\hat{\Sigma}) \approx_{(\varepsilon,\delta)} \mathcal{N}(\hat{\mu}', c^2\hat{\Sigma}').$$

失败概率的不可区分性源于我们的分数函数具有较低的敏感度。与 BGSUZ 中一样，为了证明高斯分布的不可区分性，我们分别“改变均值”和“改变协方差”，并利用组合来论证结果是不可区分的。我们在本小节末尾的引理15中证明整体隐私结论。

我们对两个分数（一个针对协方差，另一个针对均值）中的最大值调用提出-测试-发布函数。我们首先证明这个最大值具有较低的敏感度。

引理 9 固定离群值阈值 $\lambda_0 > 0$、参考集 $R \subseteq [n]$和隐私参数 $\varepsilon > 0$ 和$0 < \delta < 1$；设 $k = \left\lceil \frac{6}{\log_\varepsilon 6/\delta} \right\rceil + 4$如算法 *1* 中所示。假设 $n \geq 32e^2\lambda_0 k$。令 x和 x'为大小为 n的相邻数据集。令

$$\hat{\Sigma}, \text{SCORE}_1 \leftarrow \texttt{StableCovariance}(x, \lambda_0, k)$$
$$\hat{\Sigma}', \text{SCORE}_1' \leftarrow \texttt{StableCovariance}(x', \lambda_0, k)$$
$$\hat{\mu}, \text{SCORE}_2 \leftarrow \texttt{StableMean}(x, \hat{\Sigma}, \lambda_0, k, R)$$
$$\hat{\mu}', \text{SCORE}_2' \leftarrow \texttt{StableMean}(x', \hat{\Sigma}', \lambda_0, k, R).$$

然后 $|$最大 $\{\text{SCORE}_1, \text{SCORE}_2\} -$ 最大 $\{\text{SCORE}_1', \text{SCORE}_2'\}| \leq 2$。

引理31的证明 在第 3.4 节中给出，该引理表明 $|\mathtt{SCORE}_1 - \mathtt{SCORE}'_1| \le 2$，前提是 $k \le \frac{n}{4e^2\lambda_0}$（回顾一下，配对数据集的大小为 m，即 $n/2$）。根据假设，这一条件成立。

第 41, 44, $|\ _2 - \ '_2| \le 2$ $1-\gamma\ \Sigma \preceq \Sigma' \preceq$节证明的引理指出，$\frac{1}{1-\gamma}\Sigma$ $\gamma \le \frac{1}{2}$ $k \le \frac{1}{2\gamma}$ 11 $_1,\ '_1 < k$ SCORE SCORE 在（）ˆ ˆ ˆ 对于 和 的条件下成立。根据下文的引理，如果 SCORE SCORE，那么只要 $n \ge 32e^2\lambda_0 k$，这些不等式对 $\gamma = \frac{16e^2\lambda_0}{n}$就成立，正如我们所假设的那样。如果 $\mathtt{SCORE}_1$ 或 $\mathtt{SCORE}'_1$之一等于 k，那么我们仍然具有低敏感性：不失一般性地假设 $\mathtt{SCORE}_1 = k$，我们知道 $\mathtt{SCORE}'_1 \ge k-2$，因此

$$\max\{\mathtt{SCORE}_1, \mathtt{SCORE}_2\} = k$$
$$\max\left\{\mathtt{SCORE}'_1, \mathtt{SCORE}'_2\right\} \ge \mathtt{SCORE}'_1 \ge k-2.$$

我们用到了这样一个事实：四个分数都不超过 k。 ■

算法 1 将这两个分数中的最大值传递给 $\mathcal{M}^{\varepsilon,\delta}{}_{,\ \mathrm{PTR}}$，这是一种私有提出-测试-发布式检查，用于检验分数是否较低以及能否安全地继续。教科书中的方法使用拉普拉斯噪声来完成这项任务；我们的机制与之类似，但对于零输入总是通过，对于较大的输入则总是失败。这一改动简化了我们的论证，将参数稳定性与私有检查的（随机）结果分离开来。我们在附录B中给出了命题10的（初等）证明。

命题10 固定 $0 < \varepsilon \le 1$ 和 $0 < \delta \le \frac{\varepsilon}{10}$。存在一个算法 $\mathcal{M}^{\varepsilon,\delta}{}_{,\ \mathrm{PTR}} \colon \mathbb{R} \to \{$通过，失败$\}$，满足以下条件：

(1) 设 $\mathcal{U}$为一个集合，$g \colon \mathcal{U}^n \to \mathbb{R}_{\ge 0}$为一个函数。若对于所有相差一个条目的 x, $x' \in \mathcal{U}^n$，$|g(x) - g(x')| \le 2$，则 $\mathcal{M}^{\varepsilon,\delta}{}_{,\ \mathrm{PTR}}(g(\cdot))$ 满足(ε, δ)-差分隐私。

(2) $\mathcal{M}^{\varepsilon,\ \delta}{}_{,\ \mathrm{PTR}}(0)$ =通过。 (3) 对于所有 $z \ge {}^{2}\log_{\varepsilon}{}^{1/\delta} + 4$, $\mathcal{M}^{\varepsilon,\ \delta}{}_{,\ \mathrm{PTR}}(z)$ =失败。

现在我们给出非私有参数估计的稳定性保证。回顾一下，StableCovariance 是确定性的，并输出一个估计值 $\hat{\Sigma} \in \mathbb{R}^{d\times d}$，以及一个整数 SCORE。SCORE 的值对相邻输入具有 2- 敏感性；当 SCORE 不太大时，ˆ 参数估计 Σ 本身也具有我们所需的隐私稳定性。

引理 11 固定数据集大小 n、离群值阈值 $\lambda_0 > 0$ 和离散化参数 $k \in \mathbb{N}$。假设 $k \le \frac{n}{4e^2\lambda_0}$，并令 $\gamma = \frac{16e^2\lambda_0}{n}$。设 x、x' 为相邻数据集，且

$$\Sigma_1, \mathtt{SCORE} \leftarrow \mathtt{StableCovariance}(x, \lambda_0, k)$$
$$\Sigma_2, \mathtt{SCORE}' \leftarrow \mathtt{StableCovariance}(x', \lambda_0, k).$$

假设 $\mathtt{SCORE}, \mathtt{SCORE}' < k$。那么 Σ_1、$\Sigma_2 \succ 0$ 和 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$。如果 $\gamma \le \frac{1}{2}$，那么

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}}, \left\|\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \le (1+2\gamma)\gamma.$$

第 3 节的大部分内容都用于证明引理 11。结合以下论断，它立即意味着零均值重缩放高斯分布的不可区分性。

命题 12 *Fi* x $\varepsilon \in (0,1)$ 且 $\delta \in (0, 1/10]$ 并设 $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d\times d}$ 为正定矩阵。如果

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}}, \left\|\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \le \frac{\varepsilon}{3\log 2/\delta},$$

则 $\mathcal{N}(0,\ \Sigma_1) \approx_{(\varepsilon,\delta)} \mathcal{N}(0,\ \Sigma_2)$。

命题 12 由一个直接的高斯集中论证得出。参见 BGSUZ 中的引理 4.15，并注意到其证明经过适当调整后仍然成立，且常数更优（用 3 替代 5）。另请参见 Alabi 等人（2022），我们在第 5 节中使用了其中类似的定理 5.1。

第 4 节的主要结果与此类似：StableMean 是确定性的，并返回一个估计值 $\hat{}\mu \in \mathbb{R}^d$ 和一个整数 SCORE。后者具有 2-敏感性，并且当其不太大时，均值估计具有稳定性。我们还需考虑一些额外问题：对于输入 x 和 x'，我们可能此前已经计算出了略有不同的协方差 Σ_1 和 Σ_2。因此，我们分析 StableMean 在同时从 $(x,\ \Sigma_1)$ 移动到 $(x',\ \Sigma_2)$ 时的稳定性。此外，为提高计算效率，StableMean 接受一个“参考点”集合 $R \subseteq [n]$，并据此估计哪些点属于离群点。为了满足我们的稳定性条件，该集合需要足够大且具有充分的代表性（严格定义见定义 32）。

引理 13 固定数据集大小为 n、维度为 d、离群值阈值为 $\lambda_0 \ge 1$、离散化参数为$k \in \mathbb{N}$。使用满足 $|R| > 6k$的参考集 $R \subseteq [n]$。令 x 和 x' 为相邻的 d维数据集，且大小为 n。令 Σ_1、$\Sigma_2 \in \mathbb{R}^{d\times d}$为满足$(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$ $(\gamma = 16e^2\frac{\lambda_0}{n})$ 的正定矩阵。假设 $k \le \frac{1}{2\gamma}$。令

$$\hat{\mu}, \mathtt{SCORE} \leftarrow \mathtt{StableMean}(x, \Sigma_1, \lambda_0, k, R)$$
$$\hat{\mu}', \mathtt{SCORE}' \leftarrow \mathtt{StableMean}(x', \Sigma_2, \lambda_0, k, R).$$

若SCORE、SCORE′ $< k$ 和 R 对于 x 和 x' （参见定义 *32*）均具有度代表性，则$\|\hat{}\mu - \hat{}\mu'\|_{\Sigma_1}^2 \le (1+2\gamma)\frac{38e^2\lambda_0}{n^2}$。

非隐私均值的接近程度恰好符合我们建立具有相同协方差的高斯分布不可区分性所需的条件。回想一下，$\|\hat{}\mu - \hat{}\mu'\|_{\Sigma_1}^2 = \|\Sigma_1^{-1/2}\hat{}\mu - \Sigma_1^{-1/2}\hat{}\mu'\|_2^2$。

主张14（标准高斯机制）令向量 u 和 v 满足 $\|u-v\|_2^2 \le \Delta^2$。对于任意$\varepsilon, \delta \in (0, 1)$，令 $c^2 = \Delta^2 \cdot {}^{2}\log_{\varepsilon^2}{}^{2/\delta}$。则 $\mathcal{N}(u, c^2\mathbb{I}) \approx_{(\varepsilon,\delta)} \mathcal{N}(0, c^2\mathbb{I})$。

我们可以借助标准工具证明算法 1 的隐私保证。

引理15（主要隐私主张）固定 $0 < \varepsilon < 1$ 和 $0 < \delta < \frac{\varepsilon}{10}$。对于任意 $\lambda_0 \ge 1$，算法 *1* 是(ε, δ)-差分隐私。

证明 令 $\varepsilon' = \varepsilon/3$ 和 $\delta' = \delta/6$。我们将证明，算法 1 是 $(3\varepsilon',\ 6\delta')$-差分隐私的。

该算法需要 $n \ge {}^{192e^2\lambda_0}\log_{\varepsilon}\frac{6/\delta}{+160e^2\lambda_0}$ 才能继续执行。设定这一条件是为了使 $n \ge 32e^2\lambda_0 k$ 对于 $k = \left\lceil {}^{6}\log_{\varepsilon}{}^{6/\delta}\right\rceil + 4$ 成立，这是我们稍后将要用到的事实。假设该下界成立，因为否则算法会立即失败（因此具有隐私性）。

固定相邻数据集 x和 x'。由于 $M = |R| > 18\log 16n/\delta = 18\log 4n/\delta'$，主张50以及对 x 和 x' 中所有数据点应用并集界可证明，R 对于 x 和 x'是度代表性的

其概率至少为 $1-\delta'$。我们将证明，在 R对于Y和 Y'是度代表性的条件下，该机制是 $(3\varepsilon', 5\delta')$-差分隐私的。这将证明整个算法满足参数为 $(3\varepsilon', 6\delta') = (\varepsilon, \delta)$ 的隐私性。

我们希望 $\mathcal{M}^{\varepsilon', \delta'}$ PTR 满足 (ε', δ')-差分隐私。对于任意固定的参考集 $R = r \subseteq [n]$，引理 9 告诉我们，当 $k = \left\lceil \frac{2 \log 1/\delta'}{\varepsilon'} \right\rceil + 4 = \left\lceil \frac{6 \log}{\varepsilon}^{6/\delta} \right\rceil + 4$且 $n \geq 32e^2\lambda_0 k$时，这一条件成立。我们对 n 所作的下界假设正是为保证这一点而设定的。

当我们没有失败时，根据 $\mathcal{M}^{\varepsilon, \delta}$ PTR 的保证，我们知道两个分数都严格小于k。令 (μ_1, Σ_1) 表示在 x 上计算得到的非隐私参数，令 (μ_2, Σ_2) 表示在 x'上计算得到的非隐私参数。引理 11 和命题 12 共同推出 $\mathcal{N}(0, \Sigma_1) \approx_{(\varepsilon', \delta')} \mathcal{N}(0, \Sigma_2)$。要应用引理 11，我们需要 $n \geq 4e^2\lambda_0 k$和 $n \geq 32e^2\lambda_0$，而根据假设它们均得到满足。该引理为 $\gamma = \frac{16e^2\lambda_0}{n}$建立了 $\Delta = (1+2\gamma)\gamma$的迹范数界。要应用命题 12，我们需要 $\Delta \leq 3^{\varepsilon'} \log 2/\delta'$。（这些约束连同 $\varepsilon \leq 1$和 $\delta \leq 1/10$一起推出 $\gamma \leq 0.025$，因此 $(1+2\gamma)\gamma \leq 1.05 \cdot \gamma$和 $\Delta \leq 17e^2\lambda_0/n$。）重新整理可知，$\mathcal{N}(0, \Sigma_1)$ 与 $\mathcal{N}(0, \Sigma_2)$ 的 (ε', δ')-不可区分性要求 $n \geq \frac{153e^2\lambda_0}{\varepsilon} \log^{12/\delta}$，这比我们最初假设的界更弱。特别地，这意味着 $\mathcal{N}(\mu_2, c^2\Sigma_1) \approx_{(\varepsilon', \delta')} \mathcal{N}(\mu_2, c^2\Sigma_2)$，因为加上固定向量或乘以固定值不会影响不可区分性。

引理 13，结合上述观察，即 $2\gamma \leq 0.05$对于我们考虑的参数，结论为 $\|\mu_1 - \mu_2\|^2_{\Sigma_1} \leq \frac{40e^2\lambda_0}{n^2}$。结合主张14，这意味着 $\mathcal{N}(\mu_1, c^2\Sigma_1) \approx_{(\varepsilon', \delta')} \mathcal{N}(\mu_2, c^2\Sigma_1)$，其中

$$c^2 = \left(\frac{40e^2\lambda_0}{n^2}\right)\left(\frac{2\log 2/\delta'}{(\varepsilon')^2}\right) = \frac{720e^2\lambda_0 \log 12/\delta}{\varepsilon^2 n^2},$$

与算法 1 中相同。要应用引理 13，我们需要满足几个条件。首先，$|R| > 6k$ 按构造即满足。对于 $\gamma = \frac{16e^2\lambda_0}{n}$，条件 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$得到满足；这是引理 11 的推论。我们还要求 $k \leq \frac{1}{2\gamma}$，这等价于 $n \geq 32e^2\lambda_0 k$；我们已假设该条件成立。最后，分数严格小于 k（因为我们没有失败），并且 R 在假设下是 x 和 x' 的度代表性。

现在应用事实 52，将 $\mathcal{N}(\mu_1, c^2\Sigma_1) \approx_{(\varepsilon', \delta')} \mathcal{N}(\mu_2, c^2\Sigma_1)$ 和 $\mathcal{N}(\mu_2, c^2\Sigma_1) \approx_{(\varepsilon', \delta')} \mathcal{N}(\mu_2, c^2\Sigma_2)$ 的陈述结合起来。我们得到 $\mathcal{N}(\mu_1, c^2\Sigma_1) \approx_{(2\varepsilon', (1+e^{\varepsilon'})\delta')} \mathcal{N}(\mu_2, c^2\Sigma_2)$。由于 $e^{\varepsilon'} \leq e \leq 3$，对于值 $(2\varepsilon', 4\delta')$，我们也有相同的陈述。基本组合（事实 51）允许我们将这些高斯分布的保证与 $\mathcal{M}^{\varepsilon, \delta}$ PTR 的保证结合起来，从而证明：在 R 是 x 和 x' 的度代表性这一事实的条件下，该算法是 $(3\varepsilon', 5\delta')$-差分隐私的。由于 R 不是 x 和 x' 的度代表性的概率至多为δ'，证明完成。 ■

2.2. 准确性分析

设 x 是一个大小为 n 的 d 维数据集。我们定义经验均值和配对经验协方差：

$$\mu_x \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \Sigma_x \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}\lfloor n/2 \rfloor}\sum_{i=1}^{\lfloor n/2 \rfloor}\left(x_i - x_{i+\lfloor n/2 \rfloor}\right)\left(x_i - x_{i+\lfloor n/2 \rfloor}\right)^T. \tag{3}$$

这种“配对”是一种标准技巧，它通过将样本量减半来使数据居中：给定输入 $x = (x_1, \dots, x_n)$，我们通过设定 $y_i = (1/\sqrt{2})(x_i - x_{i+m})$ 来构造 $y = (y_1, \dots, y_m)$，其中$m = \lfloor n/2 \rfloor$。如果我们有 x_i, x_{i+m} 独立同分布$\sim \mathcal{N}(\mu, \Sigma)$，那么 $y_i \sim \mathcal{N}(0, \Sigma)$。这种变换保持相邻关系：如果 x 和 x' 相邻，那么 y 和 y' 也相邻。

现在我们引入良好数据集的概念。在这些数据集上，StableCovariance 和 StableMean 会“让数据通过”，分别始终准确返回配对经验协方差和经验均值。此外，对于此类输入，算法 1 从不返回失败。我们在观察 17 中将这些要点形式化。该观察虽然简单，却是准确性分析的关键。

定义 16（良好集中） 固定 $\lambda_0 > 0$，令 x 为一个大小为 n 个样本的 d 维数据集。令 Σ_x 为配对经验协方差。如果 Σ_x 可逆，并且对于所有 i、$j \in [n]$、$\|x_i - x_j\|_{\Sigma_x}^2 \le \lambda_0$ 均成立，则称 x 是 λ_0-良好集中的。

观察 17（在良好集中输入上始终通过） 固定 $\lambda_0 > 0$。如果数据集 $x \in \mathbb{R}^{n \times d}$ 是 λ_0-良好集中的，并且 $n \ge {}^{192e^2\lambda_0 \log_{\varepsilon}} {}^{6/\delta} + 160e^2\lambda_0$（因此我们不会立即失败），则算法 *1* 计算 $\mathtt{SCORE}_1 = \mathtt{SCORE}_2 = 0$，确定性地通过，并返回来自 $\mathcal{N}(\mu_x, c^2\Sigma_x)$ 的一个样本，其中 μ_x 是经验均值，Σ_x 是配对经验协方差，c 是算法 *1* 中设定的噪声尺度。

我们关于次高斯输入分布所使用的关键事实是：它们恰好以这种方式集中。有关次高斯分布的定义和集中不等式，请参见 A.2 节。命题 18 的证明略去，但根据这些陈述可以直接得到。回忆一下，高斯分布 $\mathcal{N}(\mu, \Sigma)$ 是参数为 $K = O(1)$ 的次高斯分布。

命题 18（次高斯数据具有良好集中性） 设 $x \in \mathbb{R}^{n \times d}$ 从参数为 $K_{\mathcal{D}}$、均值任意且具有满秩协方差的次高斯分布 $\mathcal{D}$ 中独立同分布地抽取。当 $\lambda_0 \ge K_1 K_{\mathcal{D}}^2 (d + \log n/\beta)$ 且 $n \ge K_2 K_{\mathcal{D}}^2 (d + \log 1/\beta)$ 时，存在常数 K_1 和 K_2，使得 x 以至少 $1 - \beta$ 的概率是 λ_0-良好集中的。

现在我们证明主要准确性结论。计算是标准的：以高概率，真实均值接近经验均值，而经验均值又接近我们的私有估计。

引理 19（主要准确性结论） 固定 $\varepsilon, \delta \in (0,1)$、$\delta \in (0, \varepsilon/10]$ 以及 $n, d \in \mathbb{N}$。对于 $\mu \in \mathbb{R}^d$ 和正定 $\Sigma \in \mathbb{R}^{d \times d}$，令 $x = x_1, \dots, x_n$ 独立同分布地来自参数为 $K_{\mathcal{D}}$、均值为 μ 且协方差为 Σ 的次高斯分布 $\mathcal{D}$。存在绝对常数 K_1、K_2 和 K_3，使得如果 $\lambda_0 = K_1 K_{\mathcal{D}}^2 (d + \log n/\beta)$ 且

$$n \ge K_2 K_{\mathcal{D}}^2 \cdot \frac{\log 1/\delta}{\varepsilon} \left(d + \log \left(\frac{K_{\mathcal{D}} \log 1/\delta}{\varepsilon \beta} \right) \right),$$

那么，以至少 $1 - \beta$ 的概率，算法 *1* 返回 $\hat{}\mu$，使得

$$\|\hat{\mu} - \mu\|_{\Sigma} \le K_3 K_{\mathcal{D}} \left(\sqrt{\frac{d + \log 1/\beta}{n}} + \frac{d\sqrt{\log 1/\delta}}{\varepsilon n} + \frac{\log n/\beta \sqrt{\log 1/\delta}}{\varepsilon n} \right).$$

证明 为了不立即失败，我们要求

$$n \ge \frac{192e^2\lambda_0 \log 6/\delta}{\varepsilon} + 160e^2\lambda_0 \gtrsim K_{\mathcal{D}}^2 \cdot \frac{(d + \log n/\beta) \log 1/\delta}{\varepsilon}.$$

此表达式的两边都有 n；利用 n 对数 $n \gtrsim \eta$ 意味着 $n \gtrsim \eta$ 对数 η 这一事实，我们可以将其改写为

$$n \gtrsim K_{\mathcal{D}}^2 \cdot \frac{\log 1/\delta}{\varepsilon}\left(d + \log\left(\frac{K_{\mathcal{D}} \log 1/\delta}{\varepsilon\beta}\right)\right),$$

正如我们所要求的。特别是，这意味着 $n \gtrsim K_{\mathcal{D}}^2(d + \log 1/\beta)$。

除此之外，我们还担心四个坏事件：（i）数据集并非 λ_0-良好集中；（ii）经验均值与真实均值相差甚远；（iii）经验协方差与真实协方差差异很大；（iv）为实现隐私而添加的噪声很大。我们将证明这些事件各自发生的概率至多为 $\beta/4$，在这些事件均未发生的假设下分析准确性，并使用并集界完成证明。

根据观察 17 和命题 18，以至少 $1 - \beta/4$ 的概率，数据集 x 是 λ_0-良好集中，且算法 1 发布 $\tilde{\mu} \sim \mathcal{N}(\mu_x,\ c^2\Sigma_x)$，其中 μ_x 是经验均值，Σ_x 是配对经验协方差，并且

$$c^2 = \frac{72e^2\lambda_0 \log 2/\delta}{\varepsilon^2 n^2} \lesssim K_{\mathcal{D}}^2(d + \log n/\beta) \cdot \frac{\log 1/\delta}{\varepsilon^2 n^2}. \tag{4}$$

我们加上再减去经验均值，并应用三角不等式：

$$\begin{aligned}\|\mu - \tilde{\mu}\|_\Sigma &= \|\mu - \mu_x + \mu_x - \tilde{\mu}\|_\Sigma \\ &\leq \|\mu - \mu_x\|_\Sigma + \|\mu_x - \tilde{\mu}\|_\Sigma .\end{aligned}$$

μ_x 是经验均值，因此根据命题 48，以至少 $1 - \beta/4$ 的概率，我们有

$$\|\mu - \mu_x\|_\Sigma \lesssim K_{\mathcal{D}}\sqrt{\frac{d + \log 1/\beta}{n}}. \tag{5}$$

由于 $n \gtrsim K_{\mathcal{D}}^2(d + \log 1/\beta)$，Claim 49 意味着，以至少 $1 - \beta/4$ 的概率，对于所有向量 v，我们有$\|v\|_\Sigma \lesssim \|v\|_{\Sigma_x}$。因此$\|\mu_x - \tilde{\mu}\|_\Sigma \lesssim \|\mu_x - \tilde{\mu}\|_{\Sigma_x}$这是控制 $\mu_x - \tilde{\mu}$ 时“正确”的范数，因为 $\tilde{\mu} \sim \mathcal{N}(\mu_x, c^2\Sigma_x)$。滥用记号并将分布与随机变量混为一谈，我们有

$$\|\mu_x - \tilde{\mu}\|_{\Sigma_x} = \left\|\Sigma_x^{-1/2}\big(\mathcal{N}(0, c^2\Sigma_x)\big)\right\|_2 = c \cdot \|\mathcal{N}(0, \mathbb{I})\|_2 .$$

再次应用命题 48，以至少 $1-\beta/4$ 的概率，我们有 $\|\mu_x - \tilde{\mu}\|_{\Sigma_x} \lesssim c\sqrt{d + \log 1/\beta}$。将方程 (4) 中 c 的值代入，并利用 $d + \log n/\beta \geq d + \log\ 1/\beta$，我们有

$$\begin{aligned}\|\mu_x - \tilde{\mu}\|_{\Sigma_x} &\lesssim K_{\mathcal{D}}\sqrt{d + \log n/\beta} \cdot \frac{\sqrt{\log 1/\delta}}{\varepsilon n} \cdot \sqrt{d + \log 1/\beta} \\ &\lesssim K_{\mathcal{D}} \cdot (d + \log n/\beta) \cdot \frac{\sqrt{\log 1/\delta}}{\varepsilon n}.\end{aligned}$$

将此结果与公式 (5) 中关于 $\|\mu - \mu_x\|_\Sigma$ 的上界结合，即可完成证明。 ■

2.3. 运行时间

在隐私参数适中的情况下，我们的运行时间主要取决于少数几项矩阵运算，这些运算与计算经验协方差和重缩放数据有关。

引理 20 算法 *1* 的实现至多需要

(1) 一个形如 A^TA *for* $A \in \mathbb{R}^{n\times d}$ 的乘积；*(2)* 两个形如 AB *for* $A \in \mathbb{R}^{n\times d}$ *and* $B \in \mathbb{R}^{d\times d}$ 的乘积；*(3)* 在 $\mathbb{R}^{d\times d}$ 中将一个正定矩阵求逆至多项式精度（对数位复杂度）；以及 ˜*(4)* 额外的计算开销 $O(nd/\varepsilon)$。

我们的所有算法都可以在字 RAM 模型中使用有限精度算术实现，并将比特复杂度设为相对于其他参数的对数级。非正式地说，(1) 或 (2) 中形式的矩阵乘积可以在时间 $\tilde{O}($ $nd^{\omega-1})$ 内执行，而求逆可以在时间 ˜(O) 内完成，因此总体运行时间为

$$\tilde{O}\big(nd^{\omega-1} + nd/\varepsilon\big).$$

我们使用了 $n = \Omega(kd) = \Omega(d)$，其中 $k = \Theta(\log(1/\delta)/\varepsilon)$ 是我们的离散化参数。这些关系是我们的隐私论证所要求的。在下文中，我们忽略运算的精度，因为对运行时间的精确理解并非本文的重点。还需注意，渐近意义下最快的矩阵运算算法并不实用；“教科书式”矩阵运算可以用数字 3 代替 ω 来执行。

协方差估计 我们计算协方差 $\hat{\Sigma} = \frac{1}{n}y^Ty$ 及其逆矩阵 $\hat{\Sigma}^{-1}$。对于每个（成对的）点 $y_i \in [m]$，我们对其进行重新缩放：$\tilde{y}_i \leftarrow \hat{\Sigma}^{-1}y_i$。该操作可以写成 $\tilde{y} \leftarrow y\hat{\Sigma}^{-1}$，其中 $y \in \mathbb{R}^{n\times d}$ 和 $\hat{\Sigma}^{-1} \in \mathbb{R}^{d\times d}$。（或者，也可以直接求解方程组 $\hat{\Sigma}\tilde{y}^T = y^T$。）每个马氏范数的平方（从而识别出任何离群点）都可以根据 y 和 $\tilde{y}$（通过逐元素乘法和求和）计算˜，耗时 $O(nd)$。

LargestGoodSet 的朴素实现会在发现异常值后，从头开始重复上述过程。此外，StableCovariance 会调用 LargestGoodSet $2k+1$ 次来生成 $\{S_\ell\}_{\ell=0}^{2k}$，即满足 S_ℓ 是针对 y 的最大 $e^{\ell/k}\lambda_0$-良好子集的集合族。为了改进这种朴素方法，我们注意到以下三点。

(1) 找到离群点后，需要对马氏范数进行秩一更新（以检查新的离群点）。这类更新有众所周知的高效技术，其耗时为 $\tilde{O}(nd + d^2) = \tilde{O}(nd)$。详细信息如下。

(2) 对于任何 $\ell \le \ell'$，我们根据引理 27 知道 $S_\ell \subseteq S_{\ell'}$。因此，按递减顺序依次处理异常值阈值后，我们找到的每个离群点在后续过程中都可以忽略。

（3）一旦我们找到一个 ℓ^* 使得 $|S_{\ell^*}| \le m-k$，就可以立即停止，并返回失败或输出协方差估计。回顾一下，我们计算的是

$$\texttt{SCORE} = \min\left\{k, \min_{0\le\ell\le k}\{m - |S_\ell| + \ell\}\right\}.$$

如果 $\ell^* \geq k+1$，那么我们知道 SCORE $= k$，因为对于所有 $\ell \leq k$，我们都有 $|S_\ell| \leq |S_{k+1}| \leq n-k$。这意味着调用 $\mathcal{M}^{\varepsilon,\ \delta}{}_{\text{PTR}}$ 必然会失败。如果 $\ell^* < k+1$，我们仍然可以继续下一步：对于所有 $\ell \leq \ell^*$，我们都有 $m - |S_\ell| + \ell \geq k$，因此较小的 S_ℓ 的大小不会影响我们的得分计算。

综合这些事实可知，我们至多需要执行 k 次异常值移除，并在每次移除后更新所有 n 个马哈拉诺比斯距离。跟踪移除过程中的 Σ，我们还可以计算最终的加权协方差。

下面更详细地讨论移除异常值后的更新过程。假设我们想在从协方差矩阵中移除 u 后，重新计算点 v 的范数。设 Σ_1 为初始协方差矩阵，$\Sigma_2 = \Sigma_1 - \frac{1}{m}uu^T$ 为移除 u 后的矩阵。我们希望为数据集中的每个点计算 $v^T(\Sigma_2)^{-1}v$。Sherman-Morrison公式指出：

$$\Sigma_2^{-1} = \left(\Sigma_1 - \frac{1}{m}uu^T\right)^{-1} = \Sigma_1^{-1} + \frac{\Sigma_1^{-1}uu^T\Sigma_1^{-1}}{m - u^T\Sigma_1^{-1}u} \tag{6}$$

因此，

$$\begin{aligned} v^T\Sigma_2^{-1}v &= v^T\Sigma_1^{-1}v + v^T\left(\frac{\Sigma_1^{-1}uu^T\Sigma_1^{-1}}{m - u^T\Sigma_1^{-1}u}\right)v \\ &= v^T\Sigma_1^{-1}v + \frac{\left(v^T\Sigma_1^{-1}u\right)^2}{m - u^T\Sigma_1^{-1}u}. \end{aligned}$$

我们已经计算出 $v^T\Sigma_1^{-1}v$ 和 $u^T\Sigma_1^{-1}u$。我们可以计算一次 $\Sigma_1^{-1}u$（耗时 $\tilde{O}(d^2)$），从而能够在 $\tilde{O}(d)$ 时间内为任意向量 v 计算 $\left(v^T\Sigma_1^{-1}u\right)^2$。此外，注意到方程 (6) 使我们能够直接在 $\tilde{O}(d^2)$ 时间内计算 Σ_2^{-1}。

我们已经证明，移除一个异常值后的每次更新都可以在时间$\tilde{O}(d^2 + nd) = \tilde{O}(nd)$内完成。由于我们最多执行 k 次更新，因此总耗时为 $\tilde{O}(knd)$。

请注意，我们可以同时计算最终的加权协方差，额外耗时为 $\tilde{O}(kd^2) = \tilde{O}(knd)$。此外，我们可以返回加权逆矩阵；这同样只需对我们计算出的原始 $\left(\frac{1}{m}y^Ty\right)^{-1}$ 执行至多 k 次秩一更新。

因此，我们可以通过一次计算$\hat{\Sigma} = y^Ty$、一次矩阵求逆以得到$\hat{\Sigma}^{-1}$、一次乘积 $y\hat{\Sigma}^{-1}$，以及额外的 $\tilde{O}(knd) = \tilde{O}(nd/\varepsilon)$开销来执行StableCovariance。

均值估计 与协方差估计一样，朴素地将LargestCore重复执行 $2k+1$次会造成不必要的开销。相反，假设我们有一个矩阵 $D \in \mathbb{R}^{n\times M}$，其中 $M = |R| = \Theta(k + \log n/\delta)$，且 D 的元素为$D_{i,j} = \|x_i - R_j\|^2_{\hat{\Sigma}}$。请注意，有了这个矩阵 D，我们便可以计算核心集合族$\{S_\ell\}_{\ell=0}^{2k}$，进而在时间 $\tilde{O}(nM) = \tilde{O}(n/\varepsilon)$内计算分数和权重。给定这些权重后，计算均值需要时间 $O(nd)$。

我们必须计算这个由平方距离构成的矩阵 D。正如我们所指出的，StableCovariance 可以返回 $\hat{\Sigma}^{-1}$，即最终加权协方差的逆矩阵。我们对数据进行重新缩放：$\tilde{x} \leftarrow x\hat{\Sigma}^{-1}$。有了$\tilde{x}$ 和 x，D 中的每个条目都可以在时间 $O(d)$内计算出来。

因此，我们可以在 $\tilde{O}(nd/\varepsilon)$ 时间内执行 StableMean，并进行一次 $x\hat{\Sigma}^{-1}$ 乘积。

主算法 我们可以通过抽取 $z \sim \mathcal{N}(0, \mathbb{I}_m)$ 并返回 $\hat{}\mu + y^T W^{1/2} z$，在 $\tilde{O}(nd)$ 时间内从 $\mathcal{N}(\hat{}\mu, c^2 \hat{}\Sigma)$ 中抽样，其中 $y \in \mathbb{R}^{m \times d}$ 是成对样本组成的矩阵，W 是对角矩阵，其对角元素为 StableCovariance 计算得到的权重。（数字计算机无法从高斯分布中精确抽样，但我们预计可以进行适当的离散化处理，正如 Canonne 等人（2020）所做的那样。）

要确认这是正确的分布，请注意，根据定义 $\hat{}\Sigma = y^T W y$，并展开 $\left(y^T W^{1/2} z\right)\left(y^T W^{1/2} z\right)^T$ 的期望。

3. 稳定协方差估计

本节介绍 StableCovariance，这是我们用于非私有协方差估计的确定性算法。关键在于，当它计算出的分数较小时，其参数估计是稳定的。

引理 21（重述引理 11） 设数据集大小为 n，离群值阈值为 $\lambda_0 > 0$，离散化参数为 $k \in \mathbb{N}$。假设 $k \leq \frac{n}{4e^2\lambda_0}$，并令 $\gamma = \frac{16e^2\lambda_0}{n}$。设 x、x' 为相邻数据集，并且

$$\Sigma_1, \texttt{SCORE} \leftarrow \texttt{StableCovariance}(x, \lambda_0, k)$$
$$\Sigma_2, \texttt{SCORE}' \leftarrow \texttt{StableCovariance}(x', \lambda_0, k).$$

假设 $\texttt{SCORE}$、$\texttt{SCORE}' < k$。那么 Σ_1、$\Sigma_2 \succ 0$ 和 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$。若 $\gamma \leq \frac{1}{2}$，则

$$\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}}, \left\|\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \leq (1+2\gamma)\gamma.$$

一旦给出以下定义和引理，我们就可以迅速证明引理 21。算法 2 的第一步是使用一个标准技巧将数据中心化：如果 $x_i,\ x_j \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\ \Sigma)$，那么 $\frac{1}{\sqrt{2}}(x_i - x_j) \sim \mathcal{N}(0,\ \Sigma)$。（对于次高斯分布，也有类似的结论。）鉴于此，本节中的引理将涉及大小为 $m = \lfloor n/2 \rfloor$ 的数据集 y，其中 $y_i = \frac{1}{\sqrt{2}}\left(x_i - x_{i+\lfloor n/2 \rfloor}\right)$。我们强调，本节属于隐私分析，并不假设数据实际上是零均值的。我们将在第 2.2 节分析准确性

定义 22（良好加权/子集） 令 $\lambda > 0$。对于数据集 $y = y_1, \ldots, y_m$，向量 $w \in [0,1]^m$ 是 λ-良好加权的 y，如果对于 $\Sigma_w = \sum_{j=1}^m w_j \cdot y_i y_i^T$，

$$\forall i \in \mathrm{supp}(w), \quad \left\|\Sigma_w^{-1/2} y_i\right\|_2^2 \leq \lambda.$$

如果权重在集合 $S \subseteq [m]$ 上取值 $\frac{1}{m}$，并在其他地方为零，我们将 S 称为 λ-良好子集 y。

权重可以看作数据点上的一个分布；如果不要求它们的总和为一，我们的分析会更简单。注意，这一定义要求协方差矩阵具有可逆性。

我们的引理 21 证明需要以下引理。该引理表明，如果相邻数据集y，y'分别具有“相似”的良好权重分配 w，v那么相应的加权协方差在谱距离和迹距离意义下都很接近。BGSUZ 中曾出现过几乎相同的计算；为完整起见，我们在附录B中证明我们的版本（重述为引理53）。

引理 23 设 $\lambda > 0$，且 y 和 y' 是大小为 m 的相邻数据集。设 w 是 y 的 λ-良好权重分配，设 v 是 y' 的 λ-良好权重分配，且满足 $\|w - v\|_1 \leq \rho$ 和 $\|v\|_\infty, \|w\|_\infty \leq \eta$。设 Σ_w和 Σ_v 是相应的加权协方差，如定义 *22* 中所述。设 $\gamma = \lambda(\rho + 2\eta)$。

那么 Σ_w、$\Sigma_v \succ 0$ 和 $(1-\gamma)\,\Sigma_w \preceq \Sigma_v \preceq \frac{1}{1-\gamma}\Sigma_w$。此外，如果 $\gamma \leq \frac{1}{2}$，那么

$$\left\|\Sigma_v^{-1/2}\Sigma_w\Sigma_v^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}}, \left\|\Sigma_w^{-1/2}\Sigma_v\Sigma_w^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \leq (1+2\gamma)\gamma.$$

这个引理及其均值估计对应引理 36，都是关于“可识别性”的论述。粗略地说，稳健统计学通过一种标准框架运用可识别性：如果 y 是一个“良好”的数据集（按照某种“良好”的定义），而 y' 是对 y的污染，那么找到一个与 y' 高度重叠的“良好”数据集 z 意味着 z 也会与 y高度重叠；而“良好性”将确保基于 z 得到的经验参数估计因此对于 y是准确的。

遵循 BGSUZ 的做法，我们利用可识别性来确立隐私：对于相邻数据集，我们将生成非常相似的“良好”权重向量。这最终会产生无法区分的高斯分布，并据此得到最终的均值估计。

要应用引理 23，我们必须论证 StableCovariance 计算出了良好权重。

引理 24 固定数据集 $y \in \mathbb{R}^{m\times d}$、离群值阈值 $\lambda_0 > 0$和离散化参数 $k \in \mathbb{N}$。假设 $k \leq \frac{m}{2e^2\lambda_0}$。如果 `StableCovariance` 计算出 SCORE $< k$，那么它生成的权重 $\{w_i\}_{i\in[m]}$ 是 y的 $2e^2\lambda_0$-良好权重。

最后，为了证明引理 21，我们必须证明在相邻数据集上，StableCovariance 计算出的权重分配在 ℓ_1 方面接近，并且在 ℓ_∞方面不会过大。

引理 25 固定数据集大小 m、离群值阈值 $\lambda_0 > 0$ 和离散化参数 $k \in \mathbb{N}$。假设 $k \leq \frac{m}{2e^2\lambda_0}$。令 y 和 y' 为大小为 m 的相邻数据集。令 w 为算法 *2* 在 y 上计算得到的权重，并令 w' 为在 y' 上计算得到的权重。假设在这两种情况下，我们都计算SCORE $< k$。则 $\|w - w'\|_1 \leq 2/m$。此外，$\|w\|_\infty$、$\|w'\|_\infty \leq \frac{1}{m}$。

算法 2： StableCovariance(x, λ_0, k)，用于非私有协方差估计

输入： 数据集 $x = (x_1, \ldots, x_n) \in \mathbb{R}^{n\times d}$、离群值阈值 λ_0、离散化参数 $k \in \mathbb{N}$

$m \leftarrow \lfloor n/2 \rfloor$;

$\forall i \in [m], y_i \leftarrow \frac{1}{\sqrt{2}}(x_i - x_{i+m})$; /* 配对并重新缩放 */

对于 $\ell = 0, 1, \ldots, 2k$ **执行**

  $S_\ell \leftarrow$ LargestGoodSubset$(y, e^{\ell/k}\lambda_0)$; /* 算法 3*/

end

SCORE $\leftarrow$ 最小$\{k,$ 最小$_{0\leq\ell\leq k}\{m - |S_\ell| + \ell\}\}$;

对于 $i = 1, \ldots, m$ **do**

  $w_i \leftarrow \frac{1}{km}\sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$;

end

$\Sigma \leftarrow \sum_{i\in[m]} w_i \cdot y_i y_i^T$;

返回 $\hat{\Sigma}$, SCORE;

现在我们可以证明本节的主要结果了。

证明 [引理 21] 的证明：对于相邻的 x 和 x'，我们进行配对并重新缩放，得到 y 和 y'（也相邻）。引理 25 表明，我们可以得到向量 w 和 w'，使得 $\|w-w'\|_1 \le 2/m$，且 $\|w\|_\infty$ 和 $\|w'\|_\infty$ 均至多为 $\frac{1}{m}$。根据引理 24，w 和 w' 分别是其对应数据集的 $2e^2\lambda_0$-良好权重分配。令 $\rho = 2/m$ 和 $\lambda = 2e^2\lambda_0$，引理 23 告诉我们，对于

$$\gamma = \lambda(\rho + \|w\|_\infty + \|v\|_\infty) \le \frac{8e^2\lambda_0}{m},$$

我们有所需的不等式。最后，使用 $m = n/2$。 ■

在本节余下部分，我们将证明引理 23、引理 24 和引理 25。最后，在第 3.4 节中，我们证明 StableCovariance 返回的分数具有低敏感性。

算法 3： LargestGoodSubset(y, λ)，用于协方差估计的子程序

输入： 数据集 $y = (y_1, \ldots, y_m) \in \mathbb{R}^{m\times d}$，离群值阈值 λ/* 对于向量 v 和奇异矩阵 A，定义 $\|A^{-1}v\|_2 = +\infty$ */**重复**OUT $\leftarrow \left\{ i \in S : \left\| \left(\frac{1}{m}\sum_{j\in S} y_j y_j^T \right)^{-1/2} y_i \right\|_2 > \lambda \right\}$;

$S \leftarrow [m]$;

$S \leftarrow S \setminus$OUT**直到**OUT $= \emptyset$; **返回** S;

3.1. 最大良好子集族

在开始证明引理 24 和 25 之前，我们先确立一些关于算法 2 操作的事实。

算法 2 StableCovariance 接受一个离散化参数 $k \in \mathbb{N}$作为输入。对于每个$\ell = 0$，1、.....、$2k$，它调用 LargestGoodSubset 来寻找集合 S_ℓ。将这一子集族记为 $\{S_\ell\}_{\ell=0}^{2k}$。在本小节中，我们证明每个 S_ℓ 都是唯一的最大良好子集，并讨论此类集合族的一些性质。

引理26 对于任意数据集 y 和 $\lambda > 0$，存在唯一的最大 λ-良好子集 S^*，并且满足以下性质：如果 S_0 是一个 λ-良好子集，则 $S_0 \subseteq S^*$。此外，算法 *3* 返回 S^*。

证明 注意，算法 3 要么返回一个 λ-良好子集，要么返回空集。我们将证明，如果 y 含有一个 λ-良好集合 S_0，那么算法 3 返回的集合包含 S_0。这就完成了证明：对于任意两个不同的 λ-良好集合，输出集合包含它们的并集，因此必然严格大于二者。

假设 y 有一个 λ-良好集合 S_0。我们证明 $S_0 \subseteq S$ 是一个算法不变量，其中 S表示算法 3 中被操作的集合。初始化时，我们有 $S_0 \subseteq S = [m]$。取一个

索引 $j \in S_0$，并考虑算法 3 的**重复直到**代码块开始时的一个 S 值。我们将证明 j 不会被添加到 OUT。令 $\Sigma_S = \frac{1}{m}\sum_{i\in S} y_i y_i^T$，并类似地定义 Σ_{S_0}。

移除点不会使矩阵在半正定序下增大，因此

$$\Sigma_S = \frac{1}{m}\sum_{i\in S} y_i y_i^T \succeq \frac{1}{m}\sum_{i\in S_0} y_i y_i^T = \Sigma_{S_0}.$$

由于 Σ_{S_0} 可逆（根据良性假设），Σ_S 也可逆。因此我们还知道 $\Sigma_S^{-1} \preceq \Sigma_{S_0}^{-1}$，进而

$$\|y_j\|_{\Sigma_S}^2 \leq \|y_j\|_{\Sigma_{S_0}}^2.$$

右侧一项至多为 λ，根据 S_0 上的 λ-良性假设，j 不会被作为异常值移除。最后请注意，算法 3 仅在找到一个 λ-良好子集时才会终止。 ■

下面的引理表明，在相邻数据集上计算得到的良好子集集合族紧密交织。

引理 27 设 $\lambda_0 > 0$ 和 $k \in \mathbb{N}$。设 y 和 y' 是大小为 m、仅在索引 i^* 处不同的相邻数据集。设 $\{S_\ell\}_{\ell=0}^{2k}$ 是这样的集合族：其中 S_ℓ 是 y 的最大 $e^{\ell/k}\lambda_0$-良好子集。设 $\{T_\ell\}_{\ell=0}^{2k}$ 是在 y' 上计算得到的最大 $e^{\ell/k}\lambda_0$- 良好集合所对应的集合族。假设 $k \leq \frac{m}{2e^2\lambda_0}$。我们有以下性质：

(i) For any $0 \leq \ell \leq \ell' \leq 2k$*, we have* $S_\ell \subseteq S_{\ell'}$.

(ii) 对任意 $0 \leq \ell < 2k$，我们都有 $S_\ell \setminus \{i^*\} \subseteq T_{\ell+1}$。

在证明引理 27 之前，我们将证明一个关于在单个数据集 y 上移除一个索引时良好子集的稳定性的命题。如果从一个 λ-良子集S中移除索引 j，协方差可能会减小，所得子集可能不再是 λ-良子集。不过，我们可以证明所得集合是 λ'-良子集，其中 λ' 比 λ 大一个小的乘法因子，该因子取决于 λ 和 m。

断言 28 设 $\lambda > 0$ 和 $m \geq 2\lambda$。设 y 是一个大小为 m 的数据集，并设 S 是y的一个 λ-良好子集。对于任意 $j \in [m]$，$S \setminus \{j\}$ 是 y的一个$e^{2\lambda/m}\lambda$-良好子集。

证明 令 $S' = S \setminus \{j\}$。假设 $j \in S$，否则结论显然成立。根据构造，我们有 $\Sigma_{S'} = \Sigma_S - \frac{1}{m} y_j y_j^T$。我们声称 $\Sigma_{S'} \succeq \Sigma_S - (\lambda/m)\Sigma_S$。这等价于 $\frac{1}{m} y_j y_j^T \preceq (\lambda/m)\Sigma_S$，或者

$$\Sigma_S^{-1/2} y_j y_j^T \Sigma_S^{-1/2} \preceq \lambda \mathbb{I}.$$

这是因为 $\left\|\Sigma_S^{-1/2} y_j\right\|_2^2 \leq \lambda$（因为 $j \in S$）。根据假设，$\frac{\lambda}{m}$ 至多为 $\frac{1}{2}$，因此我们有

$$\Sigma_{S'}^{-1} \preceq \frac{1}{1-\lambda/m} \cdot \Sigma_S^{-1} \preceq e^{2\lambda/m} \cdot \Sigma_S^{-1}.$$

对于任意 $i \in S' \subseteq S$，我们有 $\|y_i\|_{\Sigma_S}^2 \leq \lambda$，因此也有 $\|y_i\|_{\Sigma_{S'}}^2 \leq e^{2\lambda/m}\lambda$。 ■

证明 [引理 27] 的证明为证明命题（i），选取两个索引 $\ell \le \ell'$。我们知道，S_ℓ 是 y的一个 $e^{\ell'/k}\lambda_0$-良好子集，因为提高离群值阈值不会导致 S_ℓ中的任何点变成离群点。根据构造，$S_{\ell'}$ 是最大的 $e^{\ell'/k}\lambda_0$-良好子集，适用于 y。此外，根据引理 26，$S_{\ell'}$ 包含所有 $e^{\ell'/k}\lambda_0$-良好子集。特别地，$S_\ell \subseteq S_{\ell'}$。

为证明陈述 (ii)，考虑 $S_\ell \setminus \{i^*\}$，其中数据集 y 和 y' 仅在索引 i^*处不同。我们知道，S_ℓ 是 y的 $e^{\ell/k}\lambda_0$-良好子集，因此根据断言 28，我们知道 $S_\ell \setminus \{i^*\}$ 对于

$$\begin{aligned}\lambda' &= \exp\left\{\frac{2}{m}\cdot e^{\ell/k}\lambda_0\right\}\cdot e^{\ell/k}\lambda_0 \\ &\le \exp\left\{\frac{2}{m}\cdot e^{2}\lambda_0\right\}\cdot e^{\ell/k}\lambda_0 \\ &\le e^{(\ell+1)/k}\lambda_0,\end{aligned}$$

其中，第一个不等式使用了 $\ell \le 2k$，第二个使用了 $k \le \frac{m}{2e^2\lambda_0}$。

下一个关键事实是，对于 $S_\ell \setminus \{i^*\}$中的索引 i，我们有 $y_i = y'_i$。因此，$S_\ell \setminus \{i^*\}$ 是 $e^{(\ell+1)/k}\lambda_0$-良好子集，对于 y'而言。由于根据引理 26，$T_{\ell+1}$ 包含所有此类集合，因此 $S_\ell \setminus \{i^*\} \subseteq T_{\ell+1}$。 ■

3.2. 权重良好（引理 24 的证明）

引理 24 表明，当 StableCovariance 计算 SCORE $< k$ 时，它生成的权重实际上构成了良好的加权。

证明 [引理 24] 的证明对于 $\ell \in \{0,\dots,2k\}$，令 S_ℓ 表示 y 的最大 $e^{\ell/k}\lambda_0$-良好子集。令 $\Sigma_\ell = \frac{1}{m}\sum_{i\in S_\ell} y_i y_i^T$。算法 2 计算权重 $w_i = \frac{1}{km}\sum_{\ell=k+1}^{2k}\mathbb{1}\{i \in S_\ell\}$，这意味着我们可以将发布的 Σ 重写为

$$\hat{\Sigma} = \sum_{i=1}^{m} w_i \cdot y_i y_i^T = \frac{1}{k}\sum_{\ell=k+1}^{2k}\Sigma_\ell.$$

这个1　　较大的集合包含较小的集合，因此我们可以得到关于 Σ 的下界，其形式为 Σ_{k+1}:

$$\hat{\Sigma} \succeq \frac{1}{k}\sum_{\ell=0}^{2k}\Sigma_{k+1} = \Sigma_{k+1}.$$

现在，任意两个值 ℓ 和 ℓ'，只要 $k+1 \le \ell \le \ell' \le 2k$，就分别对应于集合 S_ℓ，$S_{\ell'}$，且这两个集合至多相差 k 个元素：S_ℓ 是 $S_{\ell'}$ 的子集，$S_{\ell'}$ 的大小至多为 m，而 S_ℓ 的大小至少为 $m-k+1 > m-k$。（要理解最后一个事实，注意到 SCORE $< k$ 意味着存在某个 $\ell^* \le k$ 使得 $n - |S_{\ell^*}| + \ell^* < k$，由此可推出 $|S_{\ell^*}| > n-k$。然后注意到 $S_{\ell^*} \subseteq S_\ell$，因为 $\ell^* \le k \le \ell$。）根据良好集合的可识别性引理（引理 23），对于 $\gamma = e^2\lambda_0 k/m$，我们有 $\Sigma_\ell \succeq (1-\gamma)\Sigma_{\ell'}$。（两个集合都是 $e^2\lambda_0$-良好的，因为 $\ell,\ \ell' \le 2k$。）这进一步给出了关于 Σ_ℓ的平方马氏距离的上界。

任取一点 $i \in \bigcup_{\ell=k+1}^{2k} S_\ell$。我们知道 $i \in S_{2k}$，因此 $\|y_i\|^2_{\Sigma_{2k}} \le e^2\lambda_0$。利用上述计算，我们有

$$\begin{aligned}\|y_i\|^2_{\hat{\Sigma}} &\le \|y_i\|^2_{\Sigma_{k+1}} \\ &\le (1 + 2e^2\lambda_0 k/m)\,\|y_i\|^2_{\Sigma_{2k}} \\ &\le (1 + 2e^2\lambda_0 k/m)\cdot e^2\lambda_0.\end{aligned}$$

由于我们假设了 $k \leq \frac{m}{2e^2\lambda_0}$，因此这至多为 $2e^2\lambda_0$。 ■

3.3. 权重是稳定的（引理 25 的证明）

引理 25 指出，对于相邻输入，StableCovariance 会生成几乎相同的权重向量（假设分数不大）。

证明 [引理 25] 的证明：对于 $\ell \in \{0, \ldots, 2k\}$，令 S_ℓ 表示 y 的最大 $e^{\ell/k}\lambda_0$-良好子集，并令 T_ℓ 表示 y' 的最大 $e^{\ell/k}\lambda_0$-良好子集。回顾算法 2 中的记号：对于每个 i，我们计算计数 $c_i = \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$，并令 $c_i + \Delta_i = \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in T_\ell\}$。假设 y 和 y' 仅在索引 i^* 处不同。由于计数介于 0 和 k 之间，因此 $|\Delta_i^*| \leq k$。根据引理 27，对于所有 $i \neq i^*$，我们知道 $\Delta_i \in \{-1, 0, +1\}$。

我们可以界定具有非零 Δ_i 的索引 $i \neq i^*$ 的数量。由于 y 上的 SCORE 小于 k，集合 S_k 的大小至少为 $m-k+1$。根据引理 27，对于每个 $\ell > k$，我们都有 $S_k \setminus \{i^*\} \subseteq S_\ell$ 和 $S_k \setminus \{i^*\} \subseteq T_\ell$，因此对于所有索引 $i \in S_k \setminus \{i^*\}$，我们都可以精确计算权重 $\frac{1}{m}$。所以具有 $\Delta_i \neq 0$ 的索引 $i \neq i^*$ 的数量至多为 k。因此

$$\begin{aligned}
\left\|w - w'\right\|_1 &= \left|w_{i^*} - w'_{i^*}\right| + \sum_{i \in [m] \setminus \{i^*\}} \left|w_i - w'_i\right| \\
&= \left|\frac{c_{i^*}}{km} - \frac{c_{i^*} + \Delta_{i^*}}{km}\right| + \sum_{i \in [m] \setminus \{i^*\}} \left|\frac{c_i}{km} - \frac{c_i + \Delta_j}{km}\right| \\
&= \left|\frac{\Delta_{i^*}}{km}\right| + \sum_{i \in [m] \setminus \{i^*\}} \left|\frac{\Delta_i}{km}\right| \\
&\leq \frac{1}{m} + k \cdot \frac{1}{km}.
\end{aligned}$$

最后，注意到根据构造，所有权重都至多为 $1/m$。 ■

3.4. SCORE 具有低敏感度

在本小节中，我们介绍函数 $g(\cdot)$，StableCovariance 通过该函数确定 SCORE 的值并将其返回。我们证明 SCORE 具有低敏感度，这使我们能够将其用作 $\mathcal{M}^{\varepsilon,\ \delta}{}_{\text{PTR}}$（我们的提出-测试-发布子程序）的输入。

定义 29 固定数据集 y、离群值阈值 $\lambda_0 > 0$ 和离散化参数 $k \in \mathbb{N}$。对于 $\ell \in \{0, 1, \ldots, k\}$，令 S_ℓ 表示 y 的最大 $e^{\ell/k}\lambda_0$-良好子集；如果不存在良好子集，则令其为 $S_\ell = \emptyset$。定义

- $f(y) \stackrel{def}{=} \min_{\ell \in \{0,\ldots,k\}} \{m - |S_\ell| + \ell\}$ *and*
- $g(y) \stackrel{def}{=} \min\{f(y), k\}$.

当存在 λ 不比 λ_0 大太多且能产生大集合S_λ时，函数 $f(y)$ 的值较小。特别地，当数据集 y 具有 $S = [m]$ 作为 λ_0- 良好子集时，则 $f(y) = g(y) = 0$。

为获得更多直观理解，请回顾 BGSUZ 中的一个关键子程序：计算与不含异常值的数据集空间之间的汉明距离。我们的评分函数近似于这一量。

我们在下面的论断中对其进行精确定义；该论断旨在帮助读者理解，并不会直接用于其他部分。

断言 30 设 $\mathfrak{S}_\lambda$ 为由大小为 m 的数据集组成的空间，其中 $[m]$ 是一个 λ-良好子集。令 $d(y, \mathfrak{S}_\lambda) = \min_{z\in\mathfrak{S}_\lambda} d(y,z)$ 表示到该空间的汉明距离。如果我们计算 $g(y) < k$，那么我们知道 $d(y, \mathfrak{S}_{e\lambda_0}) \le g(y) \le 2\cdot d(y, \mathfrak{S}_{\lambda_0})$。

证明 对于上界，假设 $d(y, \mathfrak{S}_{\lambda_0}) = r$，因此存在一个 z，使得 $d(y,z) = r$ 且 $r \in \mathfrak{S}_{\lambda_0}$。从 z 中移除 r 个点，我们得到 y 的一个 $e^{r/k}\lambda_0$- 良好子集，其大小为 $m-r$，因此 $g(y) \le m - m + r + r = 2\cdot d(y, \mathfrak{S}_{\lambda_0})$。

现在假设 $g(y) = r < k$，因此存在某个 ℓ，使得 $m - |S_\ell| + \ell = r$。我们知道 $S_\ell \subseteq S_k$，所以向 S_ℓ 中至多添加 r 个 $\mathbf{0}$ 的副本，可以得到集合 $z \in \mathfrak{S}_{e\lambda_0}$，使得 $d(y,z) = r$。因此 $d(y, \mathfrak{S}_{e\lambda_0}) \le r = g(y)$。 ■

本节的主要结果表明，我们的评分函数具有较低的敏感度。

引理31 固定数据集大小 m、离群值阈值 $\lambda_0 > 0$和离散化参数 $k \in \mathbb{N}$。假设 $k \le \frac{m}{2e^2\lambda_0}$。对于所有相邻的 y 和 y'，我们有 $|g(y) - g(y')| \le 2$。

我们在引理 27 中论证，对于某些参数，如果数据集 y 具有较大的良好集合，那么相邻的 y' 必定具有一个（稍微小一些的）良好集合（对应稍微大一些的离群值阈值）。因此，在这些条件下，我们有 $f(y) \approx f(y')$。当 y在适度的 λ下没有良好集合，导致 $f(y)$ 较大时，这些条件可能不成立。为了解决这个问题，我们引入 $g(y)$，其值不会大于 k。

证明 [引理 31] 的证明。不失一般性，假设 $g(y) \le g(y')$。我们将通过分析两种情况来证明$g(y') \le g(y) + 2$。

情况 1： 假设 $g(y) = k$。那么，根据构造，$g(y') \le g(y) \le g(y) + 2$，因为 $g(\cdot)$ 的上限为 k。

情况 2： 假设 $g(y) < k$。只有当 $f(y) < k$ 时才会出现这种情况，因此存在一个 $\ell^* \in \{0,\dots,k\}$ 和子集 S_{ℓ^*}，满足 (i) 对于 y 是 $e^{\ell^*/k}\lambda_0$-良好的，且 (ii) 满足 $m - |S_{\ell^*}| + \ell^* < k$。此外，我们知道 $\ell^* < k$，因为 $m - |S_{\ell^*}|$ 非负。

现在可以应用引理 27，因为我们已假设 $k \le \frac{m}{2e^2\lambda_0}$。对于~~$\ell$~~ $\in \{0,\dots,k\}$，令 T_ℓ 为 y' 的最大 $e^{\ell/k}\lambda_0$-良好子集。引理 27 表明 $S_{\ell^*} \setminus \{i^*\} \subseteq T_{\ell^*+1}$。由此我们可以给 $g(y')$ 取上界：

$$\begin{aligned} g(y') \le f(y') &= \min_{\ell\in\{0,\dots,k\}} m - |S_\ell| + \ell \\ &\le m - |T_{\ell^*+1}| + \ell^* + 1 \\ &\le m - (|S_{\ell^*}| - 1) + \ell^* + 1 \\ &= g(y) + 2. \end{aligned}$$

因此，我们完成了。 ■

4. 稳定均值估计

本节介绍用于非私有均值估计的算法 StableMean，并证明其稳定性保证。该算法基于一种“离群性”概念：内点与数据集中许多其他点的距离都足够近。更正式地说，可以将其视为一个图，其中 n 个顶点对应一个数据集 x；若顶点 i 和 j 之间的距离足够近，则它们相邻。按照这种解释，度低于某个阈值的点就是离群点。

ˆ StableMean 的输入包括数据集 x、用作预条件器的矩阵 Σ 、离群值阈值 λ_0、“离散化”参数 k以及一组“参考点” $R \subseteq [n]$。

我们分析该算法在同时改变单个数据点 ˆ 并轻微改变预条件器 Σ 时的稳定性。我们所说的“轻微改变”，具体是指在正定序意义下乘法意义上的接近，如引理 11 的推论所示。

参考点 $R \subseteq [n]$ 用于估计哪些点是离群点：我们不检查它们与数据集中其他每个 $n-1$ 点之间的距离，而是计算它们与少数随机选取的数据点之间的距离。可以将此过程理解为估计上述图中顶点的度：我们将论证，当 R 具有足够的代表性时，该算法是隐私的。标准的集中论证表明，随机选取的足够大样本 R 以至少 $1-\delta$ 的概率具有代表性。（Tsfadia 等人（2022）的工作中也采用过类似的论证，其应用与隐私估计类似。）

定义 32（度代表性）固定数据集 x、协方差 Σ、离群值阈值 λ_0和参考集 $R \subseteq [n]$。对于所有 i，令

$$N_i \stackrel{def}{=} \left\{ j \in [n] : \|x_i - x_j\|_\Sigma^2 \le e^2 \lambda_0 \right\}$$
$$\tilde{N}_i \stackrel{def}{=} \left\{ j \in R : \|x_i - x_j\|_\Sigma^2 \le e^2 \lambda_0 \right\}.$$

˜令 $z_i = |N_i|$ 和 $\tilde{z}_i = |N_i|$ 为集合的大小。我们称 R 具有相对于 x 的度代表性，如果对于每个索引 i，都有 $\left| \frac{1}{|R|}\tilde{z}_i - \frac{1}{n} z_i \right| \le \frac{1}{6}$。

本节的主要结果是关于相邻数据集上的非私有均值估计的以下稳定性保证。

引理 33（引理 13 重述）固定数据集大小 n、维度 d、离群值阈值 $\lambda_0 \ge 1$和离散化参数 $k \in \mathbb{N}$。使用满足 $|R| > 6k$的参考集 $R \subseteq [n]$ 。令 x 和 x'为大小为 n的相邻 d维数据集。令 Σ_1、$\Sigma_2 \in \mathbb{R}^{d \times d}$ 为满足$(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$ （对于 $\gamma = 16e^2 \frac{\lambda_0}{n}$）的正定矩阵。假设 $k \le \frac{1}{2\gamma}$。令

$$\hat{\mu}, \texttt{SCORE} \leftarrow \texttt{StableMean}(x, \Sigma_1, \lambda_0, k, R)$$
$$\hat{\mu}', \texttt{SCORE}' \leftarrow \texttt{StableMean}(x', \Sigma_2, \lambda_0, k, R).$$

如果 SCORE、SCORE′ $< k$ 和 R 对于 x 和 x' 都是度代表性的（见定义 *32*），那么$\|\hat{} \mu - \hat{} \mu'\|_{\Sigma_1}^2 \le (1+2\gamma)\frac{38e^2\lambda_0}{n^2}$。

如第 S 3 节中所述，在证明引理 33 之前，我们先给出另一个定义和几个引理。

在本节中，“异常值”是指度不足的点。下面我们将形式化这一概念（不明确引用图）。我们考虑加权子集，如第 3 节中所述。

算法 4： StableMean(x, $\hat{\Sigma}$, λ_0, k, R) 用于非私有均值估计

输入： 数据集 $x = (x_1, \ldots\ldots, x_n) \in \mathbb{R}^{n \times d}$，协方差 $\hat{\Sigma}$，离群值阈值 λ_0，离散化参数 $k \in \mathbb{N}$，参考集 $R \subseteq [n]$

对于 $\ell = 0, 1, \ldots, 2k$ **执行**

  $S_\ell \leftarrow \text{LargestCore}(x, \hat{\Sigma}, e^{\ell/k}\lambda_0, R)$;

end

SCORE $\leftarrow$ 最小值 $\{k$,最小值$_{0 \le \ell \le k} \{n - |S_\ell| + \ell\}\}$;

对于 $i = 1, \ldots, n$ **do**

  $c_i \leftarrow \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$;

end

$Z \leftarrow \sum_{i \in [n]} c_i$; /* 归一化常数 */

$\forall i \in [n], w_i \leftarrow \frac{c_i}{Z}$; /* 若 $Z = 0$，则设置 $w_i \leftarrow 0$ */

$\hat{\mu} \leftarrow \sum_{i \in [n]} w_i x_i$;

返回 $\hat{\mu}$，SCORE;

算法 5： LargestCore(x, $\hat{\Sigma}$, λ, τ, R)，均值估计子程序

输入： 数据集 $x = (x_1, \ldots, x_n) \in \mathbb{R}^{n \times d}$，协方差 $\hat{\Sigma}$，离群值阈值 λ，度阈值 τ，参考集 $R \subseteq [n]$

对于 $i \in [n]$ **执行**

  $N_i = \left\{ j \in R : \|x_i - x_j\|_\Sigma^2 \le \lambda \right\}$; /* 邻近点 in R */

end

return $\{i \in [n] : |N_i| \ge \tau\}$;

定义 34（加权核心）设 x 是大小为 n 的数据集。设 $\lambda > 0$、$\tau \in \mathbb{N}$ 和 $R \subseteq [n]$。若分布 w 定义在 $[n]$ 上，则当对于每个 $i \in \text{supp}(w)$，都存在一个大小至少为 τ 的集合 $N_i \subseteq R$，使得 $\forall j \in N_i$、$\|x_i - x_j\|_\Sigma^2 \leq \lambda$ 时，称其为关于 R 的 x 的-加权核心) $\Sigma\, \lambda\tau$(。若 w 在子集 S 上均匀分布，则称 S 为关于 R 的 x 的-核心) $\Sigma\, \lambda\tau$(。

备注35 对于给定的数据集 x、参考集 R 以及参数 τ、λ 和 Σ，可以直接找到 $S^* \subseteq [n]$，即关于 R 的 x $(\tau\, \lambda\, \Sigma)$-核心中的最大者（如果 y_i 在 R 中有 τ 个邻近点，则将 i 放入 S^*，否则将其省略）。与此相比，在第 *3* 节中，高效算法能够恢复最大的 λ-良好集合这一点并非显然。除了这种算法保证之外，引理 *26* 还证明了最大的 λ-良好集合是唯一的，并且包含所有其他 λ-良好集合。对于$(\tau\, \lambda\, \Sigma)$-核心，类似的观察显然成立。

证明引理 33 时，我们需要的第一个论断表明：如果两个加权核心（对应于相邻数据集）在 ℓ_1 距离下接近，那么它们的均值也接近。我们在附录B中给出推论 36（重述为推论 55）的证明；几乎相同的计算已出现在 BGSUZ 中。

推论 36 设 x 和 x' 是大小为 n 的相邻数据集。设 $\lambda > 0$、$\tau \in \mathbb{N}$ 和 $R \subseteq [n]$。设 Σ_1 和 Σ_2 是满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$（其中 $\gamma \leq \frac{1}{2}$）的正定矩阵。设 ~~w~~ 是关于 R、对应于 x 的 $(\tau, \lambda, \Sigma_1)$-加权核心，且 v 是关于 R、对应于 x' 的 $(\tau, \lambda, \Sigma_2)$-加权核心，其中 $\tau > \frac{n}{2} + 1$。若 $\|w - v\|_1 \leq \rho$，则

$$\|\mu_w - \mu_v\|_{\Sigma_1}^2 = \left\| \sum_{i \in [n]} w_i x_i - \sum_{i \in [n]} v_i x_i' \right\|_{\Sigma_1}^2 \leq (1 + 2\gamma)\lambda(\rho + \|w\|_\infty + \|v\|_\infty)^2.$$

为了应用推论 36，我们需要证明 StableMean 生成的权重实际上构成 x的加权核心。

引理 37 固定一组输入至 `StableMean`（算法 *4*）：数据集 x、协方差 Σ、离群值阈值 λ_0、离散化参数 $k \in \mathbb{N}$以及参考集 $R \subseteq [n]$。假设 $|R| > 6k$。如果`StableMean` 计算出 `SCORE` $< k$ 且 R 对于 x 是度代表性的，那么它生成的权重 $w \in [0,1]^n$ 是关于 $[n]$ 的 $\left(\frac{n}{2} + 1, e^2\lambda_0, \hat{\Sigma}\right)$-加权核心，其中 x。

此外，StableMean 在相邻数据集上生成的权重在 ℓ_1距离下必须彼此接近。

引理 38 固定离群值阈值 $\lambda_0 \geq 1$、离散化参数 $k \in \mathbb{N}$以及参考集 $R \subseteq [n]$。令 x 和 x' 为大小为 n 的相邻数据集。令 Σ_1 和 Σ_2 为满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$ 的正定矩阵，其中 $\gamma \leq \frac{1}{2}$。进一步假设 $k \leq \frac{1}{2\gamma}$。

令 w 为算法 *4* 在输入$(x, \Sigma_1, \lambda_0, k, R)$ 上计算得到的权重，而 w' 为其在输入$(x', \Sigma_2, \lambda_0, k, R)$上计算得到的权重。假设在这两种情况下，它都计算出`SCORE` $< k$。那么

$$\|w - w'\|_1 \leq \frac{4(1 + 2k/n)^2}{n}.$$

此外，$\|w\|_\infty$、$\|w'\|_\infty \leq \frac{1+2k/n}{n}$

现在我们可以证明非私有均值估计的稳定性保证。

证明 [引理 33] 的证明设 w 为 StableMean 在输入（x、Σ_1、λ_0、k、R）上计算得到的权重向量，w' 为在输入（x'、Σ_2、λ_0、k、R）上计算得到的权重。根据引理 37，以及 R对于 x具有度代表性的事实，w是 x的（$n/2+1$、$e^2\lambda_0$、Σ_1）加权核心。出于相同的原因，w'是 x'的（$n/2+1$、$e^2\lambda_0$、Σ_2）加权核心。根据引理 38，w和 w'满足 $\|w-w'\|_1 \le {}^{4}(1+2k/n)_{n}{}^{2}$以及 $\|w\|_\infty$、$\|w'\|_\infty \le (1+2k/n)_n$。因此，令 $\rho = {}^{4}(1+2k/n)_{n}{}^{2}$且令 $\lambda = e^2\lambda_0$，推论 36 告诉我们，这些均值必须彼此接近：

$$\begin{aligned}\left\|\hat{\mu}-\hat{\mu}'\right\|_{\Sigma_1}^2 &\le 2\lambda(\rho+\|w\|_\infty+\|v\|_\infty)^2 \\ &\le (1+2\gamma)e^2\lambda_0\cdot\left(\frac{6(1+2k/n)^2}{n}\right)^2.\end{aligned}$$

为整理该表达式，注意到 $\gamma = \frac{16e^2\lambda_0}{n}$和 $k \le \frac{1}{2\gamma}$ 可推出

$$\frac{2k}{n} \le \frac{1}{16e^2\lambda_0} \le \frac{1}{16e^2}.$$

因此 $(1+\frac{2k}{n})^4 \le 1.04$，从而得到 $\|\hat{\mu}-\hat{\mu}'\|_{\Sigma_1}^2 \le (1+2\gamma)\frac{38e^2\lambda_0}{n^2}$。 ■

4.1. 最大核心族

在给出引理 37 和 38 的证明之前，我们先建立一些关于 StableMean、算法 4 的操作的事实。

固定一个参考集 $R \subseteq [n]$和一个离散化参数 $k \in \mathbb{N}$。对于每个 $\ell = 0, 1, \ldots, 2k$，StableMean 调用 LargestCore（算法 5），以寻找相对于 R、针对 xˆ 的最大 $\left(|R|-\ell, e^{\ell/k}\lambda_0, \hat{\Sigma}\right)$-核心，其中 Σ 是由 StableCovariance 生成的非私有估计量。将 x 相对于 R的最大核心族记为$\{S_\ell\}_{\ell=0}^{2k}$。本节将讨论此类核心族的一些性质。与第 3 节和引理 27 中一样，我们将证明，相邻数据集上的核心族彼此紧密交织。

引理 39 固定 $R \subseteq [n]$、$\lambda_0 > 0$和 $k \in \mathbb{N}$。设 x 和 x' 是仅在索引 i^* 处不同的相邻数据集。设 Σ_1 和 Σ_2 是满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq {}_{1-\gamma}^{1}\Sigma_1$（对 $\gamma \le \frac{1}{2}$ 成立）的正定矩阵。进一步假设 $k \le \frac{1}{2\gamma}$。

设 $\{S_\ell\}_{\ell=0}^{2k}$ 为满足以下条件的集合族：对于 R，S_ℓ 是关于 x 的最大 $(|R|-\ell$、$e^{\ell/k}\lambda_0$、$\Sigma_1)$-核心。设 $\{T_\ell\}_{\ell=0}^{2k}$ 为关于 R 的 x' 对应的最大 $(|R|-\ell$、$e^{\ell/k}\lambda_0$、$\Sigma_2)$-核心集合族。以下性质成立：

(i) For any $0 \le \ell \le \ell' \le 2k$, we have $S_\ell \subseteq S_{\ell'}$.

(ii) For any $0 \le \ell < 2k$, we have $S_\ell \setminus \{i^*\} \subseteq T_{\ell+1}$.

证明命题(i)直接由我们对核心的定义得出：降低邻居阈值并提高离群值阈值会使条件限制更宽松，因此 S_ℓ是关于 R的、相对于 x 的$(|R|-\ell', e^{\ell'/k}\lambda_0, \Sigma_1)$-核心。由于 $S_{\ell'}$ 包含所有此类核心（回顾备注 35），$S_\ell \subseteq S_{\ell'}$。

当 $\Sigma_1 \succeq (1-\gamma)\Sigma_2$时，对于任意向量 v，我们有 $\|v\|_{\Sigma_1}^2 \leq \frac{1}{1-\gamma}\|v\|_{\Sigma_2}^2$。现在，$S_\ell \setminus \{i^*\}$ 是关于 R的、相对于 x 的$(|R|-\ell, e^{\ell/k}\lambda_0, \Sigma_1)$-核心：对于所有 $i \in S_\ell \setminus \{i^*\}$，都存在一个集合 $N_i \subseteq R$，使得

$$|N_i| \geq |R| - \ell \quad \text{and} \quad \forall j \in N_i,\ \|x_i - x_j\|_{\Sigma_1}^2 \leq e^{\ell/k}\lambda_0.$$

（该条件对所有 $i \in S_\ell$ 都成立，因此对所有 $i \in S_\ell \setminus \{i^*\}$ 也成立。）改变协方差后，我们看到，对于 x，相对于 R，$S_\ell \setminus \{i^*\}$ 是一个（$|R|-\ell$，λ'，Σ_2）-核心，其中 $\lambda' = \frac{1}{1-\gamma}e^{\ell/k}\lambda_0$。我们假设 $\gamma \leq \frac{1}{2}$ 和 $k \leq \frac{1}{2\gamma}$，因此 $\frac{1}{1-\gamma} \leq 1 + 2\gamma \leq e^{2\gamma} \leq e^{1/k}$，这意味着 $\lambda' \leq e^{(\ell+1)/k}\lambda_0$。

现在我们要论证 $S_\ell \setminus \{i^*\}$ 是 x' 的核心。这并非显而易见，因为 i^* 仍可能属于我们的参考集。我们必须将度阈值降低 1：对于所有 $i \in S_\ell \setminus \{i^*\}$，存在一个集合 $N_i' = N_i \setminus \{i^*\} \subseteq R$，使得

$$\left|N_i'\right| \geq |R| - (\ell+1) \quad \text{and} \quad \forall j \in N_i',\ \left\|x_i' - x_j'\right\|_{\Sigma_1}^2 \leq e^{(\ell+1)/k}\lambda_0,$$

其中我们用到了这样一个事实：对于所有 i 属于 $S_\ell \setminus \{i^*\}$ 或 $N_i' = N_i \setminus \{i^*\}$ 的情况，都有 $x_i = x_i'$。因此，$S_\ell \setminus \{i^*\}$是相对于 R而言、针对 x'的 $(|R|-(\ell+1)\ e^{(\ell+1)/k}\lambda_0, \Sigma_2)$-核心，这正如我们先前所论证的那样，意味着$S_\ell \setminus \{i^*\} \subseteq T_{\ell+1}$。 ■

4.2. 权重是良好的（引理 37 的证明）

引理 37 表明，当 StableMean 在输入 x 上计算 SCORE $< k$，且参考点集合具有度代表性时，所生成的权重是 x的一个核心。

证明 [引理 37] 的证明 对于 $\ell \in \{0, \ldots, 2k\}$，令 S_ℓ表示相对于 R而言、针对 x 的最大的 $(|R|-\ell, e^{\ell/k}\lambda_0, \hat{}\Sigma)$-核心。我们知道 S_{2k}包含所有此类核心（回顾备注35）。特别地，如果索引 i属于 w的支撑集，那么 $i \in S_{2k}$。这直接意味着 w是相对于 R而言、针对 x 的 $(|R|-2k, e^2\lambda_0, \hat{}\Sigma)$-加权核心。（作为一个技术细节，请注意，SCORE $< k$ 意味着存在一个 $\ell^* < k$ 满足 $S_{\ell^*} \neq \emptyset$，因此返回的权重不是零向量。）

现在反设权重 w 不是一个 $(n/2+1,\ e^2\lambda_0, \hat{}\Sigma)$-加权核心（针对 x，相对于 $[n]$）。那么存在一个索引 $i \in \text{supp}(w)$，使得 $z_i \leq n/2$（这里的 z_i 和 $\tilde{z}_i$ 与定义 32 中相同）。由于 R 是 x的度代表性，这意味着 $\frac{1}{|R|}\tilde{z}_i \leq \frac{2}{3}$。然而，由于 i 属于加权核心，它在 R中有许多邻居：$\tilde{z}_i$ 至少为 $|R|-2k > |R| - \frac{2|R|}{3} = \frac{2|R|}{3}$。因此我们得出了矛盾。 ■

4.3. 权重是稳定的（引理 38 的证明）

引理 38 表明，StableMean 在相邻输入上产生的权重向量几乎相同（假设分数不会太大）。

证明 [引理 38] 的证明对于 $\ell \in \{0, \ldots, 2k\}$，令 S_ℓ表示针对 $\ell \in \{0, \ldots, 2k\}$、相对于 R 的最大 $(|R|-\ell,\ e^{\ell/k}\lambda_0,\ \hat{}\Sigma)$-核心。类似地，令 T_ℓ表示针对x' 、相对于 R 的最大 $(|R|-\ell,\ e^{\ell/k}\lambda_0,\ \Sigma_2)$-核心。回忆算法 4 中的记号：对于每个 i，我们计算计数 $c_i = \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in S_\ell\}$、归一化常数 $Z = \sum_{i\in[n]} c_i$ 和权重 $w_i = \frac{c_i}{Z}$。令 c_i'，Z'和w_i' 表示在 x' 上计算得到的相同数值。与引理 25 的证明一样，令 $\Delta_i = c_i - c_i'$。

我们首先证明，在 x 和 x' 下计算得到的归一化常数相近。假设 x 和 x' 在索引 i^* 处不同。我们知道 $|\Delta_{i^*}| \leq k$，因为计数介于 0 和 k 之间。

引理 39 告诉我们，对于所有 $i \neq i^*$，$\Delta_i \in \{-1, 0, +1\}$。我们还对满足 $\Delta_i \neq 0$ 的索引 $i \neq i^*$ 的数量有一个上界。由于我们在两种情况下都计算了 SCORE $< k$，核心 S_k 的大小至少为 $n - k + 1$。此外，对于所有 $\ell > k$，我们知道 $S_k \setminus \{i^*\} \subseteq S_\ell$ 且 $S_k \setminus \{i^*\} \subseteq T_\ell$。因此，除至多 k 个索引 $i \neq i^*$（即不属于 $S_k \setminus \{i^*\}$ 的那些索引）之外，我们有 $\Delta_i = 0$。所以 $|Z - Z'| = |\sum_i c_i - c_i - \Delta_i| \leq 2k$，因为 i^* 项的绝对值可能为 k，且至多有 k 个索引的绝对值为 1。

同样的事实表明，Z' 很大：

$$\begin{aligned} Z' = \sum_{i=1}^{n} \sum_{\ell=k+1}^{2k} \mathbb{1}\{i \in T_\ell\} &= \sum_{\ell=k+1}^{2k} |T_\ell| \\ &\geq k \cdot |S_k \setminus \{i^*\}|. \end{aligned}$$

这对 Z 也成立，因此我们有 $Z,\ Z' \geq k(n-k)$。

现在，我们可以将权重的 ℓ_1 差异分为三种情况：i^*、那些 $i \neq i^*$ 满足 $\Delta_i = 0$ 的情况，以及那些 $i \neq i^*$ 满足 $\Delta_i \neq 0$ 的情况。

$$\begin{aligned} \left\|w - w'\right\|_1 &= \sum_{i \in [n]} \left| \frac{c_i}{Z} - \frac{c_i}{Z'} - \frac{\Delta_i}{Z'} \right| \\ &\leq \sum_{i \in [n]} c_i \left| \frac{1}{Z} - \frac{1}{Z'} \right| + \left| \frac{\Delta_i}{Z'} \right| \\ &\leq n \cdot \left| \frac{1}{Z} - \frac{1}{Z'} \right| \cdot \max_i \{c_i\} + \frac{2k}{Z'}, \end{aligned} \tag{7}$$

再次通过将满足 $|\Delta_i| = 1$ 的（至多为 k 个索引）与具有 $|\Delta_{i^*}| \leq k$ 的那个（即 i^*）索引相结合，得到 $2k$。由于 Z' 很大，第二项至多为 $\frac{2}{n-k} \leq \frac{2}{(1+2k/n)\, n}$。公式 ~~(7)~~ 中的~~第一项~~也很小：我们有

$$\begin{aligned} \left| \frac{1}{Z} - \frac{1}{Z'} \right| &= \left| \frac{Z'}{ZZ'} - \frac{Z}{ZZ'} \right| \\ &\leq \frac{2k}{k^2(n-k)^2}. \end{aligned}$$

由于 $c_i \leq k$ 我们可以将第一项界定为

$$n \cdot \left| \frac{1}{Z} - \frac{1}{Z'} \right| \cdot \max_i \{c_i\} \leq \frac{2nk^2}{k^2(n-k)^2} \leq \frac{2(1+2k/n)^2}{n}.$$

我们利用 $(1+2k/n)^2 \geq (1+2k/n)$ 将其与公式（7）中的第二项结合起来。最后，请注意所有权重至多为 $\frac{1}{n-k} \leq \frac{1+2k/n}{n}$。 ——— ■

4.4. SCORE 具有低敏感度

在本小节中，我们介绍函数 $g(\cdot)$，StableMean 通过该函数确定 SCORE 的值，并将其与均值估计一同返回。我们证明 SCORE 具有较低的敏感度，这使我们能够将其作为输入传递给 $\mathcal{M}^{\varepsilon,\ \delta}$ PTR，即我们的提出-测试-发布子程序。

定义40 固定 x、 Σ、 λ_0、 k和 $R \subseteq [n]$。对于 $\ell \in \{0,1,\ldots,k\}$，令 S_ℓ表示相对于 R的 x的最大（$|R|-\ell$、 $e^{\ell/k}\lambda_0$、 Σ）- 核心；如果不存在这样的核心，则令 $S_\ell = \emptyset$。定义

- $f(x,\Sigma) \stackrel{def}{=} \min_{\ell\in\{0,\ldots,k\}} (n - |S_\ell| + \ell)$ *and*
- $g(x,\Sigma) \stackrel{def}{=} \min\{f(x,\Sigma), k\}$.

引理41 固定 $\lambda_0 > 0$、 $k \in \mathbb{N}$和 $R \subseteq [n]$。设 x 和 x' 是相邻数据集。设 Σ_1 和 Σ_2 是满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$ 的正定矩阵，其中 $\gamma \leq \frac{1}{2}$。进一步假设 $k \leq \frac{1}{2\gamma}$。则 $|g(x, \Sigma_1) - g(x', \Sigma_2)| \leq 2$。

这个证明与引理41几乎完全相同，只是我们使用引理39而非引理27。为完整起见，附录B重述了该结果（作为断言56）并给出了证明。

5. 私有协方差估计与高斯分布的快速学习

如第1.1节所述，我们的隐私分析表明，对于相邻输入 x 和 x'，我们要么失败，要么得到协方差估计 Σ_1 和 Σ_2，使得 $\mathcal{N}(0,\Sigma_1) \approx_{(\varepsilon,\delta)} \mathcal{N}(0,\Sigma_2)$。正如 Alabi 等人（2022）所指出的，若有更强的稳定性保证，就可以为这些分布的多个独立抽样提供不可区分性保证。有了足够多这样的样本，就可以构造出准确的私有协方差矩阵估计。

下面我们将形式化这一联系。为简化表述，我们假设 $d = \Omega(\log n)$，并重点讨论高斯数据的保证，这些保证以较高的常数概率成立。我们给出谱范数方面的保证；利用不等式 $\|\cdot\|_F \leq \sqrt{d}\,\|\cdot\|_2$，可立即得到弗罗贝尼乌斯范数方面的保证。

定理 42 固定 $\varepsilon \in (0,1)$、 $\delta \in (0, \varepsilon/10)$以及 n、 $d \in \mathbb{N}$。假设 $d = \Omega(\log n)$。算法 *6* 的输入为一个包含 n 个点的数据集，每个点都属于 $\mathbb{R}^d$；此外还输入隐私参数 ε、 δ以及离群值阈值 λ_0。

- 对于任意 $\lambda_0 \geq 1$，算法 *6* 都满足$(\varepsilon,\ \delta)$-差分隐私。
- 假设 x 是从 $\mathcal{N}(0,\ \Sigma)$ 中独立同分布抽取的，其中 $\Sigma \in \mathbb{R}^{d\times d}$ 为正定矩阵。存在绝对常数 K_1、 K_2和 K_3，使得若 $\lambda_0 = K_1 d$ 且 $n \geq K_2 d\log(1/\delta)/\varepsilon$，则以较高的常数概率，算法 *6* 不会失败，而是返回一个半正定矩阵 $\tilde{\Sigma} \in \mathbb{R}^{d\times d}$，满足

$$\left\|\Sigma^{-1/2}\tilde{\Sigma}\Sigma^{-1/2} - \mathbb{I}\right\|_2 \leq K_3\left(\sqrt{\frac{d}{n}} + \frac{d^{3/2}\sqrt{\log 1/\delta}}{\varepsilon n}\right).$$

- 算法 *6* 可实现为需要：一个形如 A^TA 的乘积（其中 $A \in \mathbb{R}^{n\times d}$），一个形如 AB 的乘积（其中 $A \in \mathbb{R}^{n\times d}$ 和 $B \in \mathbb{R}^{d\times d}$），一个形如 A^TA 的乘积（其中$A \in \mathbb{R}^{N\times d}$），一个形如 AB 的乘积（其中 $A \in \mathbb{R}^{N\times d}$和 $B \in \mathbb{R}^{d\times d}$），对 $\mathbb{R}^{d\times d}$ 中的正定矩阵进行一次求逆和一次特征值分解，其比特复杂度为对数级，˜以及额外的计算开销 $O(nd/\varepsilon)$。其中， N 是合成样本的数量，在算法 *6* 中设定。

命题 43 将高斯参数估计（均值具有较小的马氏误差，协方差具有较小的弗罗贝尼乌斯误差）与以较小全变差距离学习分布本身联系起来。有了这一事实，算法 7（仅运行算法 1 和 6）的效用保证便可直接得到。隐私保证则遵循基本组合（事实 51）。

命题 43（Diakonikolas 等人（2016）） 存在一个常数 K，使得对于任意 $\alpha \le \frac{1}{2}$、向量μ_1、$\mu_2 \in \mathbb{R}^d$ 以及正定 Σ_1、$\Sigma_2 \in \mathbb{R}^{d\times d}$，如果 $\|\mu_1-\mu_2\|_{\Sigma_1} \le \alpha$ 且 $\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}-\mathbb{I}\right\|_F \le \alpha$，则 $\mathrm{TV}(\mathcal{N}(\mu_1、\ \Sigma_1)\ \mathcal{N}(\mu_2、\ \Sigma_2)) \le K\alpha$。

推论 44 固定 $\varepsilon \in (0,\ 1)$、$\delta \in (0,\ \varepsilon/10)$ 以及 n、$d \in \mathbb{N}$。假设 $d = \Omega(\log\ n)$。算法 7 接受一个包含 n 个点的数据集，其中每个点属于 $\mathbb{R}^d$，以及隐私参数 ε、δ 和离群值阈值 λ_0。

- 对于任意 $\lambda_0 \ge 1$，算法 7 都是 $(2\varepsilon, 2\delta)$-差分隐私的。
- 假设 $x \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma)$，其中 Σ 是正定的。存在绝对常数 K_1 和 K_2，使得如果 $\lambda_0 = K_1 d$ 且

$$n \ge K_2\left(\frac{d^2}{\alpha^2} + \frac{d^2\sqrt{\log 1/\delta}}{\alpha\varepsilon} + \frac{d\log 1/\delta}{\varepsilon}\right),$$

那么，以较高的常数概率，算法 7 不会失败，而是返回一对 $\tilde{}$ $\tilde{}$$(\tilde{\mu}, \Sigma)$，使得 $\mathrm{TV}(\mathcal{N}(\mu, \Sigma)\ \mathcal{N}(\tilde{\mu}, \Sigma)) \le \alpha$。

5.1. 私有协方差估计的讨论

近年来，差分隐私高斯协方差估计受到了广泛关注。在假设真实协方差满足 $\mathbb{I} \preceq \Sigma \preceq \kappa\mathbb{I}$（对于某个 $\kappa \ge 1$）的情况下，Kamath、Li、Singhal 和 Ullman（2019）提出了在谱范数和弗罗贝尼乌斯范数下进行学习的高效算法。这些算法几乎达到了最优，但需要以 κ 的形式提供先验信息。其误差仅以 κ 的多对数形式增长（与先前的方法形成对比，后者的误差为 Ω(poly(κ))，但这类先验知识通常并不可得）。在下文中，我们关注对不受限 协方差的学习，即不具备这类先验知识的情形。

使用 $n \gtrsim d^{3/2}$ 个样本，算法 6 返回的私有协方差估计在谱范数下是准确的。据我们所知，这是首个针对无约束高斯分布实现此类保证的差分隐私多项式时间算法，解决了 Alabi 等人（2022）提出的一个开放问题。在 $\alpha = O(1/\sqrt{d})$ 的范围内，这种对维度的依赖已知是最优的（Kamath 等人，2022），并且与在假设 $\mathbb{I} \preceq \Sigma \preceq O(1)\mathbb{I}$ 下应用的标准高斯机制相匹配。

在过去两年中，许多论文建立了隐私学习无约束高斯分布协方差的技术，使其在缩放后的弗罗贝尼乌斯范数（也称为矩阵的马氏范数）下达到 α 误差。这种估计量结合标准的均值估计程序后，可以将整个分布学习到 $O(\alpha)$ 的全变差距离（见命题 43）。这项任务的最优样本复杂度为

$$n \gtrsim \frac{d^2}{\alpha^2} + \frac{d^2}{\alpha\varepsilon} + \frac{\log 1/\delta}{\varepsilon}.$$

信息论上界由 Aden-Ali 等人（2021）给出；第一项是非私有估计所必需的，第三项即使对于已知方差且具有无约束均值的一维高斯分布的均值估计也是必需的（Karwa 和 Vadhan，2017），而第二项最近已被证明是紧的（Kamath 等人，2022）。Aden-Ali 等人（2021）的上界是非构造性的；随后 Liu 等人（2022）给出了一个具有几乎相同保证的指数时间算法。一组同期独立论文很快给出了多项式时间算法（Kamath 等人，2021；Ashtiani 和 Liaw，2021；Kothari 等人，2021），其中 Ashtiani 和 Liaw（2021）实现了最优的 d^2 维度依赖。（利用 Ashtiani 和 Liaw（2021）的工具，Tsfadia 等人（2022）的框架实现了类似的误差上界。）最近，Hopkins 等人（2022）和 Alabi 等人（2022）的同期工作给出了多项式时间且样本最优的估计量，同时还满足针对输入的对抗性扰动的鲁棒性。前者针对该任务的保证也是多项式时间内已知的最佳结果，与方程 5.1 相比，仅在 d, α, ε 和 $\log 1/\delta$ 中相差对数因子。

我们的工作（结合算法 1 和 6）在对数因子的范围内达到了最优样本复杂度，运行时间为 $\tilde{O}(nd^{\omega-1} + nd/\varepsilon)$。据我们所知，目前已知最快的算法运行时间为 $\tilde{O}(nd^{\omega-1} + d^{\omega+1}/\varepsilon)$，该算法结合了 Ashtiani 和 Liaw（2021）以及 Tsfadia 等人（2022）的思路。对于这项任务的典型参数设置 $n \gtrsim d^2/\varepsilon$，这两个表达式的渐近主导项都是第一项，即计算整个输入数据协方差所需的时间。

我们指出，特征值分解无法精确计算；相关的近似算法已得到充分研究（例如，参见 Trefethen 和 Bau，1997年）。

算法 6： 隐私协方差估计，$\mathcal{A}^{\varepsilon,\delta,\lambda_0}_{\text{协方差}}(x)$

输入： 数据集 $x \in \mathbb{R}^{n\times d}$，隐私参数 ε, δ，离群值阈值 λ_0

要求： $n \geq \frac{272e^2\lambda_0 \log 2/\delta}{\varepsilon} \left\lceil 4 \text{日志}_{\varepsilon}{}^{2/\delta} \right\rceil + 4$;

$k \leftarrow \left\lceil 4 \text{日志}_{\varepsilon}{}^{2/\delta} \right\rceil + 4$;

$N \leftarrow \left\lfloor \overline{10^{-6} \cdot \lambda_0^2}^{n^2\varepsilon^2} \text{日志 } 2/\delta \right\rfloor$;

$\hat{\Sigma}$,SCORE $\leftarrow$ StableCovariance(x, λ_0, k); /* $\hat{\Sigma} \in \mathbb{R}^{d\times d}$,SCORE $\in \mathbb{N}$ */

如果 $\mathcal{M}^{\varepsilon/2,\ \delta/2}_{\text{PTR}}$ (SCORE) $=$ 通过 **则**

  抽取 Z_1、.....、$Z_N \overset{iid}{\sim} \mathcal{N}(0,\ \hat{\Sigma})$;

  返回 $\frac{1}{N}\sum_{i=1}^{N} Z_i Z_i^T$;

else

  返回 失败

end

5.2. 定理 42 的证明

我们使用以下引理，可将其视为命题 12（在 BGSUZ 中使用）的类似结果；该命题使用高级组合，若要重复这一过程 N 次，大约需要多出 $\sqrt{N}$ 倍的样本。

断言 45（Alabi 等人（2022年），定理 5.1） 固定 $\varepsilon, \delta \in (0,1)$。设 Σ_1, Σ_2 为正定矩阵，满足

$$\left\|\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - \mathbb{I}\right\|_F, \left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}\right\|_F = \Delta$$

对于满足 $\Delta \leq \frac{\varepsilon}{8 \log(1/\delta)} \leq 1$ 的任意对象。对于任意 $N \leq \frac{\varepsilon^2}{8\Delta^2 \log(1/\delta)}$，我们有 $\mathcal{N}(0, \Sigma_1)^{\otimes N} \approx_{(\varepsilon,\delta)} \mathcal{N}(0, \Sigma_2)^{\otimes N}$，其中 $p^{\otimes N}$ 表示 N 重乘积分布，即 N 个 p 的独立副本。

证明 [定理 42] 的证明。令 $\varepsilon' = \varepsilon/2$ 和 $\delta' = \delta/2$。回顾引理 11，当分数低于 k 时，对于 $\gamma = 16e^2\lambda_0/n$，我们有 $\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \leq (1+2\gamma)\gamma$（交换 Σ_1 和 Σ_2 后也有同样的不等式）。因此，为应用断言 45，我们利用 $\|\cdot\|_F \leq \|\cdot\|_{\mathrm{tr}}$ 这一事实，并要求 $(1+2\gamma)\gamma \leq \frac{\varepsilon'}{8 \log(1/\delta)} = \frac{\varepsilon}{16 \log(2/\delta)}$。我们对 ε 和 δ 的假设意味着 $\gamma \leq 0.03$，因此 $(1+2\gamma) \leq 1.06$ 且 $(1+2\gamma)\gamma \leq 17e^2\lambda_0/n$。所以，要抽取单个样本，我们要求 $n \geq \frac{272e^2\lambda_0}{\log(2/\delta)\varepsilon}$；否则就会失败。这一假设强于 $k \leq \frac{n}{4e^2\lambda_0}$，而引理 11 也要求该条件。要抽取 N 个样本，我们要求 $\frac{n^2}{17^2e^4\lambda_0^2} \cdot \frac{32}{\varepsilon^2 \log(2/\delta)}$。（算法 6 中的常数为 10^6，大于 $17^2 \times e^4 \times 32$。）隐私论证的其余部分遵循提出-测试-发布的保证（引理 9 给出评分函数的敏感度，命题10 给出 $\mathcal{M}^{\varepsilon,\delta}_{\mathrm{PTR}}$ 的隐私保证）以及基本组合。

我们在第 2.3 节分析了 StableCovariance 的运行时间。为了生成私有$\hat{}$协方差估计，我们首先计算 Σ 的特征值分解，并由此生成$\hat{\Sigma}^{1/2}$。接着，我们从 $\mathcal{N}(0, \mathbb{I})$中抽取 N个样本（耗时 $\tilde{O}(Nd)$），用$\hat{\Sigma}^{1/2}$对其重新缩放，并计算缩放后各点的经验协方差。

关于准确性，首先注意到，我们假设 $n \gtrsim d\log(1/\delta)/\varepsilon$，因此满足对 n 的输入要求，不会立即失败。除此之外，准确性论证只需对 Claim 49 进行两次$\tilde{}$直接应用。设 y 的经验协方差为 Σ_y，我们发布的私有估计为 Σ。我们加上并减去 Σ_y，然后应用三角不等式：

$$\begin{aligned}
\left\|\Sigma^{-1/2}\tilde{\Sigma}\Sigma^{-1/2} - \mathbb{I}\right\|_2 &= \left\|\Sigma^{-1/2}\Big(\tilde{\Sigma} - \Sigma\Big)\Sigma^{-1/2}\right\|_2 \\
&= \left\|\Sigma^{-1/2}\Big(\tilde{\Sigma} - \Sigma_y + \Sigma_y - \Sigma\Big)\Sigma^{-1/2}\right\|_2 \\
&\leq \left\|\Sigma^{-1/2}\Big(\tilde{\Sigma} - \Sigma_y\Big)\Sigma^{-1/2}\right\|_2 + \left\|\Sigma^{-1/2}(\Sigma_y - \Sigma)\Sigma^{-1/2}\right\|_2 \\
&= \left\|\Sigma^{-1/2}\Big(\tilde{\Sigma} - \Sigma_y\Big)\Sigma^{-1/2}\right\|_2 + \left\|\Sigma^{-1/2}\Sigma_y\Sigma^{-1/2} - \mathbb{I}\right\|_2.
\end{aligned}$$

由于 $n \gtrsim d + \log 1/\beta$，以高概率 Σ_y 可逆且 $\left\|\Sigma^{1/2}\Sigma_y^{-1/2}\right\|_2 = 1 + o(1)$，因此应用柯西-施瓦茨不等式可得

$$\begin{aligned}
\left\|\Sigma^{-1/2}\Big(\tilde{\Sigma} - \Sigma_y\Big)\Sigma^{-1/2}\right\|_2 &= \left\|\Sigma^{-1/2}\Big(\Sigma_y^{1/2}\Sigma_y^{-1/2}\Big)\Big(\tilde{\Sigma} - \Sigma_y\Big)\Big(\Sigma_y^{-1/2}\Sigma_y^{1/2}\Big)\Sigma^{-1/2}\right\|_2 \\
&\lesssim \left\|\Sigma_y^{-1/2}\Big(\tilde{\Sigma} - \Sigma_y\Big)\Sigma_y^{-1/2}\right\|_2 \\
&= \left\|\Sigma_y^{-1/2}\tilde{\Sigma}\Sigma_y^{-1/2} - \mathbb{I}\right\|_2.
\end{aligned}$$

因此，我们需要控制两项：

$$\left\|\Sigma^{-1/2}\tilde{\Sigma}\Sigma^{-1/2}-\mathbb{I}\right\|_2 \lesssim \left\|\Sigma^{-1/2}\Sigma_y\Sigma^{-1/2}-\mathbb{I}\right\|_2+\left\|\Sigma_y^{-1/2}\tilde{\Sigma}\Sigma_y^{-1/2}-\mathbb{I}\right\|_2. \tag{8}$$

由于 Σ_y 是 Σ 的经验估计，且使用了 n 个样本，方程（8）中的第一项以高概率至多为 $\sqrt{d/n}$。

类似地，Σ 是 Σ_y 的经验估计，且使用了 N 个样本，因此方程 8 中的第二项以高概率至多为 $\sqrt{d/N}$。我们有 $N \approx d^2 n^2\varepsilon^2$ 对数 $1/\delta$，因此整理后可得上界 $\frac{d^{3/2}\sqrt{\log 1/\delta}}{\varepsilon n}$，如前所述。

■

算法 7： 无约束高斯分布的隐私学习器

输入： 数据集 $x \in \mathbb{R}^{n\times d}$，隐私参数 ε, δ，离群值阈值 λ_0

要求： $n \geq \frac{272e^2\lambda_0\log 2/\delta}{\varepsilon}$

$\tilde{\mu} \leftarrow \mathcal{A}^{\varepsilon,\delta,\lambda_0}$main (x); /* 算法 1 */

$\forall i \in [\lfloor n/2 \rfloor]$, $y_i \leftarrow \frac{1}{\sqrt{2}}(x_i - x_{i+\lfloor n/2 \rfloor})$; /* 配对并重新缩放 */

$\tilde{\Sigma} \leftarrow \mathcal{A}^{\varepsilon,\ \delta,\ \lambda_0}$协方差 (y); /* 算法 6*/

如果 $\tilde{\mu} =$ 失败或 $\tilde{\Sigma} =$ 失败**则**

  返回 失败;

else

  返回 $\tilde{\mu}$, $\tilde{\Sigma}$;

end

参考文献

Ishaq Aden-Ali、Hassan Ashtiani 和 Gautam Kamath。关于隐私学习无界高维高斯分布的样本复杂度。在《算法学习理论》中，第 185–216 页。PMLR，2021年。

Daniel Alabi、Pravesh K Kothari、Pranay Tankala、Prayaag Venkat 和 Fred Zhang。隐私估计高斯分布：高效、稳健且最优。*arXiv*预印本 *arXiv:2212.08018*，2022年。

Hassan Ashtiani 和 Christopher Liaw。用于学习高斯分布及其他分布的私有多项式时间算法。*arXiv* 预印本 *arXiv:2111.11320*，2021年。

Gavin Brown、Marco Gaboardi、Adam Smith、Jonathan Ullman 和 Lydia Zakynthinou。在不进行私有协方差估计的情况下进行协方差感知的私有均值估计。《神经信息处理系统进展》，34:7950–7964，2021年。

T Tony Cai、Yichen Wang 和 Linjun Zhang。《隐私成本：差分隐私下参数估计的最优收敛速率》。*arXiv* 预印本 *arXiv:1902.04495*，2019 年。

Clément L. Canonne、Gautam Kamath 和 Thomas Steinke。用于差分隐私的离散高斯分布。*CoRR*，abs/2004.00010，2020年。网址：https://arxiv.org/abs/2004.00010。

Ilias Diakonikolas、Gautam Kamath、Daniel M Kane、Jerry Li、Ankur Moitra 和 Alistair Stewart。在无需面对计算难处理性的高维稳健估计量。载于*2016 IEEE 第57*届计算机科学基础年会（*FOCS*），2016年。John Duchi、Saminul Haque 和 Rohith Kuditipudi。自适应隐私均值估计的快速算法，2023年。Cynthia Dwork 和 Jing Lei。差分隐私与稳健统计。载于第*41*届 *ACM* 计算理论研讨会论文集，STOC ’09，第371–380页。ACM，2009年。Samuel B Hopkins、Gautam Kamath、Mahbod Majid 和 Shyam Narayanan。鲁棒性蕴含统计估计中的隐私。*arXiv* 预印本 *arXiv:2212.05015*，2022年。Roger A Horn 和 Charles R Johnson。矩阵分析。剑桥大学出版社，2012年。Gautam Kamath 和 Jonathan Ullman。私有统计学入门。*arXiv* 预印本 *arXiv:2005.00010*，2020年。Gautam Kamath、Jerry Li、Vikrant Singhal 和 Jonathan Ullman。隐私学习高维分布。载于学习理论会议，第1853–1902页。PMLR，2019年。Gautam Kamath、Argyris Mouzakis、Vikrant Singhal、Thomas Steinke 和 Jonathan Ullman。用于无界高斯分布的私有且计算高效的估计量。*arXiv* 预印本 *arXiv:2111.04609*，2021年。Gautam Kamath、Argyris Mouzakis 和 Vikrant Singhal。隐私估计的新下界与推广的指纹引理。*arXiv* 预印本 *arXiv:2205.08532*，2022年。Vishesh Karwa 和 Salil Vadhan。有限样本差分隐私置信区间。*arXiv* 预印本 *arXiv:1711.03908*，2017年。Shiva Prasad Kasiviswanathan 和 Adam D. Smith。论差分隐私的“语义”：一种贝叶斯表述。《隐私与保密期刊》，2014年。Pravesh K Kothari、Pasin Manurangsi 和 Ameya Velingker。通过稳定化凸松弛实现隐私稳健估计。*arXiv* 预印本 *arXiv:2112.03548*，2021年。Xiyang Liu、Weihao Kong 和 Sewoong Oh。高维中的差分隐私与稳健统计。载于学习理论会议，第1167–1246页。PMLR，2022年。Vikrant Singhal 和 Thomas Steinke。隐私学习子空间。神经信息处理系统进展，34:1312–1324，2021年。Matthew Skala。超几何尾不等式：终结荒谬。*arXiv* 预印本 *arXiv:1311.5939*，2013年。Lloyd N Trefethen 和 David Bau。数值线性代数，第181卷。SIAM，1997年。Eliad Tsfadia、Edith Cohen、Haim Kaplan、Yishay Mansour 和 Uri Stemmer。Friendlycore：实用的差分隐私聚合。载于国际机器学习会议，第21828–21863页。PMLR，2022年。

Salil Vadhan。差分隐私的复杂度。载于密码学基础教程，第 347–450 页。Springer，2017。

Roman Vershynin。高维概率：数据科学中的应用导论，第 47 卷。剑桥大学出版社，2018。

附录 A. 预备知识

A.1. 记号与初等事实

除非另有说明，“$\gtrsim$”和“$\lesssim$”表示绝对常数。我们将点集 $x = (x_1, \ldots\ldots, x_n)$既视为有序的 n 元组，也视为矩阵 $x \in \mathbb{R}^{n\times d}$。向量均为列向量。对数以 e 为底。我们用 $[n]$ 表示集合 $\{1, \ldots, n\}$，用 $\mathbb{N} = \{1, 2, \ldots\}$ 表示自然数。对于向量 $v \in \mathbb{R}^d$，其支撑集为 $\operatorname{supp}(v) = \{i \in [d] \colon v_i \neq 0\}$。当谓词 P 为真时，指示函数 $\mathbb{1}\{P\} \in \{0, 1\}$ 取值为 1，否则取值为 0。对于零均值数据，我们常用“协方差”指二阶矩矩阵：$\Sigma_x = \frac{1}{n}\sum_{i=1}^n x_i x_i^T = \frac{1}{n} x^T x$。

在全文中，我们使用以下事实：$\frac{1}{1-x} \leq 1 + 2x$（对于 $x \in [0, 1/2]$）以及 $1 + x \leq e^x$（对于所有 x）。

如果一个对称矩阵 $A = \mathbb{R}^{d\times d}$ 满足：对于所有 $v \in \mathbb{R}^d$，$v^T A v \geq 0$，则称其为半正定矩阵。如果同一陈述中的不等式是严格不等式，则称 A 为正定矩阵。对于半正定矩阵 A 和 B，我们以标准方式表示半正定序（也称洛文纳序）：当且仅当 $A - B$ 为半正定矩阵时，$A \succeq B$。该记号可扩展到 $\preceq$、$\succ$和 $\prec$。矩阵求逆保持半正定序（Horn and Johnson（2012年），推论 7.7.4.a）：如果 $A,\ B \succ 0$ 则当且仅当 $A^{-1} \preceq B^{-1}$时，$A \succeq B$。

对于具有奇异值 $\sigma_1 \geq \cdots \geq \sigma_d$ 的矩阵 $A \in \mathbb{R}^{d\times d}$，我们使用谱范数 $\|A\|_2 = \sigma_1$、弗罗贝尼乌斯范数 $\|A\|_F = \sqrt{\sum_{i=1}^d \sigma_i^2}$和迹范数 $\|A\|_{\mathrm{tr}} = \sum_{i=1}^d \sigma_i$。我们有$\|A\|_2 \leq \|A\|_F \leq \|A\|_{\mathrm{tr}}$。

A.2. 次高斯随机变量与集中不等式

有关次高斯随机变量的进一步讨论，请参阅 Vershynin（2018）的教材。

定义 46（次高斯范数）设 $y \in \mathbb{R}$ 是一个随机变量。y 的次高斯范数记为 $\|y\|_{\psi_2}$，定义为 $\|y\|_{\psi_2} = $下确界 $\left\{t > 0 \colon \mathbb{E}\exp\left(y^2/t^2\right) \leq 2\right\}$。

定义 47（次高斯随机变量）设 $y \in \mathbb{R}^d$ 是一个均值为 μ、协方差为 Σ 的随机变量。若存在 $K \geq 1$，使得对于所有 $v \in \mathbb{R}^d$，都有，则称 y 具有参数 K 的次高斯性质。

$$\|\langle y - \mu, v\rangle\|_{\psi_2} \leq K\sqrt{v^T \Sigma v}.$$

高斯分布 $\mathcal{N}(\mu,\ \Sigma)$ 是参数为 $K = O(1)$ 的次高斯分布。

命题 48（范数集中）令 $y_1, \ldots, y_n$ 从一个参数为 K_y、均值为 μ 且协方差 Σ 为（满秩）的 d 维次高斯分布中独立同分布地抽取。存在一个常数 K_1，使得以至少 $1 - \beta$ 的概率，同时有

$$\|y_1 - \mu\|_\Sigma^2 \leq K_1 K_y^2 (d + \log 1/\beta) \quad \textit{and} \quad \left\|\frac{1}{n}\sum_{i=1}^n y_i - \mu\right\|_\Sigma^2 \leq K_1 K_y^2 \cdot \frac{d + \log 1/\beta}{n}.$$

Claim 49（协方差集中）令 $y_1, \ldots, y_n$ 从一个参数为 K_y、均值为 $\mu = 0$ 且协方差 Σ 为（满秩）的 d 维次高斯分布中独立同分布地抽取。令 $\hat{}\Sigma = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^T$ 为经验协方差。存在绝对常数 K_1 和 K_2，使得对于任意 $\beta \in$ （0，1），若 $n \geq K_2$（$d + \log 1/\beta$），则以至少 $1-\beta$ 的概率有

$$\left\| \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2} - \mathbb{I} \right\|_2 \leq K_1 K_y^2 \sqrt{\frac{d + \log 1/\beta}{n}}.$$

最后，我们使用以下关于超几何分布的尾界来论证：参考点的随机集合 R（在算法 1 中选取并提供给 StableMean）以高概率具有度代表性。

主张50（见 Skala (2013)）假设一个 *urn* 中有 N 个球，其中恰好有 k 个是黑球。令随机变量 y 表示从 *urn* 中均匀且不放回地抽取 n 个球时抽到的黑球数量。那么对于所有 $t \geq 0$，都有 概率 $[|y - kn/N| \geq tn] \leq 2e^{-2t^2 n}$。

A.3. 差分隐私

关于这些事实以及差分隐私的更多背景知识，请参阅 Vadhan (2017) 的专著。

事实 51（基本组合）对于所有 $1 \leq i \leq K$，假设机制 $\mathcal{M}_i$ 是（ε，δ）-差分隐私的。那么（$\mathcal{M}_1$，$\mathcal{M}_2$，…，$\mathcal{M}_K$）是（$K\varepsilon$，$K\delta$）-差分隐私的。此外，即使机制是自适应选择的，该结论仍然成立。

事实 52 假设对于某些 ε 和 δ，分布 p_1、p_2 和 p_3 满足 $p_1 \approx_{(\varepsilon,\delta)} p_2$ 和 $p_2 \approx_{(\varepsilon,\delta)} p_3$。那么 $p_1 \approx_{(2\varepsilon,(1+e^{\varepsilon})\delta)} p_3$。

附录B. 延后的证明：可识别性与敏感度

在本附录中，我们证明三个命题；这些命题的证明与我们在 BGSUZ 中给出的版本相比仅有细微改动。

我们还证明引理41（重述为引理 56）；其证明与引理 31 的证明几乎完全相同。我们也证明命题10（重述为命题57），即我们提出-测试-发布变体的保证。

引理 53（引理 23 的重述）设 $\lambda > 0$，并设 y 和 y' 是大小为 m 的相邻数据集。设 w 是 y 的 λ-良好加权，v 是 y' 的 λ-良好加权，且满足 $\|w - v\|_1 \leq \rho$ 和$\|v\|_\infty \|w\|_\infty \leq \eta$。设 Σ_w 和 Σ_v 为相应的加权协方差，如定义 *22* 所述。

那么，Σ_w 和 Σ_v 均为正定，并且对于 $\gamma = \lambda(\rho + 2\eta)$，满足

$$(1-\gamma)\Sigma_w \preceq \Sigma_v \preceq \frac{1}{1-\gamma}\Sigma_w. \tag{9}$$

此外，若 $\gamma \leq \frac{1}{2}$，则

$$\left\| \Sigma_v^{-1/2} \Sigma_w \Sigma_v^{-1/2} - \mathbb{I} \right\|_{\mathrm{tr}}, \left\| \Sigma_w^{-1/2} \Sigma_v \Sigma_w^{-1/2} - \mathbb{I} \right\|_{\mathrm{tr}} \leq (1+2\gamma)\gamma.$$

证明 不失一般性地假设 y 与 y' 仅在索引 1 处不同，并令 $z_1 \overset{\text{定义}}{=} y'_1$。那么我们有

$$\Sigma_v = \sum_{i\in[m]} v_i \cdot \left(y'_i\right)\left(y'_i\right)^T = \Sigma_w + v_1 \cdot z_1 z_1^T - w_1 \cdot y_1 y_1^T + \sum_{\substack{i\in[m]\\ i>1}} (v_i - w_i)\cdot y_i y_i^T \tag{10}$$
$$\succeq \Sigma_w + 0 - w_1 \cdot y_1 y_1^T + \sum_{\substack{i\in\text{supp}(w)\\ i>1}} (v_i - w_i)\cdot y_i y_i^T.$$

最后一行由以下事实得出：限制到 w 的支撑集后，我们只是从求和中去掉了半正定项（当 $w_i = 0$ 时，我们有 $v_i - w_i \geq 0$）。我们希望用 $(1-\gamma)\Sigma_w$ 给出它的下界，因此只需证明 $-w_1 \cdot y_1 y_1^T + \sum_{i\in\text{支撑集}(w)}(v_i - w_i)y_i y_i^T \succeq -\gamma\Sigma_w$，或者等价地证明 $w_1 \cdot y_1 y_1^T + \sum_{i\in\text{支撑集}(w)}(w_i - v_i)y_i y_i^T \preceq \gamma\Sigma_w$。这是算子范数的一个上界；为证明这一点，我们用 $\Sigma_w^{-1/2}$ 进行共轭，并应用三角不等式以及 w 是 λ-良好的假设。（另请回忆，对于任意向量 u，都有 $\left\|uu^T\right\|_2 = u^T u$。）

$$\begin{aligned}
&\left\|w_1 \cdot \Sigma_w^{-1/2} y_1 y_1^T \Sigma_w^{-1/2} + \sum_{i\in\text{supp}(w)} (w_i - v_i)\cdot \Sigma_w^{-1/2} y_i y_i^T \Sigma_w^{-1/2}\right\|_2 \\
&\qquad \leq |w_1| \cdot \left\|\Sigma_w^{-1/2} y_1 y_1^T \Sigma_w^{-1/2}\right\|_2 + \sum_{i\in\text{supp}(w)} |w_i - v_i| \cdot \left\|\Sigma_w^{-1/2} y_i y_i^T \Sigma_w^{-1/2}\right\|_2 \\
&\qquad = |w_1| \cdot \left\|\Sigma_w^{-1/2} y_1\right\|_2^2 + \sum_{i\in\text{supp}(w)} |w_i - v_i| \cdot \left\|\Sigma_w^{-1/2} y_i\right\|_2^2 \\
&\qquad \leq |w_1| \cdot \lambda + \sum_{i\in\text{supp}(w)} |w_i - v_i| \cdot \lambda.
\end{aligned}$$

这至多为 $\lambda(\|w\|_\infty + \|w - v\|_1)$，小于 $\lambda(2\eta + \rho)$。对称的论证表明 $(1-\gamma)\Sigma_v \preceq \Sigma_w$，由此得到 Σ_v 的上界。

通过直接论证即可得出迹范数不等式。由方程 (10) 出发，并以 $\Sigma_w^{-1/2}$进行共轭变换，我们有

$$\begin{aligned}
\Sigma_w^{-1/2}\Sigma_v\Sigma_w^{-1/2} - \mathbb{I} &= v_1 \cdot \Sigma_w^{-1/2} z_1 z_1^T \Sigma_w^{-1/2} - w_1 \cdot \Sigma_w^{-1/2} y_1 y_1^T \Sigma_w^{-1/2} \\
&\quad + \sum_{i=2}^m (v_i - w_i)\cdot \Sigma_w^{-1/2} y_i y_i^T \Sigma_w^{-1/2}.
\end{aligned}$$

与之前一样，我们应用三角不等式和 $\left\|uu^T\right\|_{\text{tr}} = \|u\|_2^2$：

$$\begin{aligned}
\left\|\Sigma_w^{-1/2}\Sigma_v\Sigma_w^{-1/2} - \mathbb{I}\right\|_{\text{tr}} &\leq |v_1| \cdot \left\|\Sigma_w^{-1/2} z_1\right\|_2^2 + |w_1| \cdot \left\|\Sigma_w^{-1/2} y_1\right\|_2^2 \\
&\quad + \sum_{i=2}^m |v_i - w_i| \cdot \left\|\Sigma_w^{-1/2} y_i\right\|_2^2.
\end{aligned}$$

我们无法以 λ 为上界统一界定这些平方范数，因为我们关于 w 良性的假设仅适用于 w 的支撑集中的点。然而，根据方程 (9)，对于支撑集中的所有 i 在 v，均有 $\left\|\Sigma_w^{-1/2} y'_i\right\|_2^2 \leq (1 + 2\gamma)\lambda$。

该论证结合 $\|w-v\|_1 \le \rho$，得到

$$\left\|\Sigma_w^{-1/2}\Sigma_v\Sigma_w^{-1/2} - \mathbb{I}\right\|_{\mathrm{tr}} \le (1+2\rho\lambda)\cdot(\|v\|_\infty + \|w\|_\infty + \rho)\cdot\lambda.$$

第二个迹范数不等式具有对称性。 ■

注意，当 $R=[n]$ 和度阈值 τ 至少为 $\frac{n}{2}+1$ 时，$\mathrm{supp}(w)$ 中的任意两个点必定有一个共同的“邻居”，因此它们彼此之间的距离不会太大。我们有如下断言：在 ℓ_1 范数（等价地，全变差距离）下彼此接近的加权核心具有接近的均值。此外，即使这些核心是在略有不同的协方差下计算得到的，这一结论仍然成立。

引理 54 令 $x = x_1, \dots, x_n$ 为一个数据集。令 $\lambda > 0, \tau \in \mathbb{N}$ 和 $R \subseteq [n]$。令 Σ_1 和 Σ_2 为满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$（对 $\gamma \le \frac{1}{2}$）的正定矩阵。令 w 为关于 R 的 x 的一个 (τ,λ,Σ_1)-加权核心，并令 v 为关于 R 的 x 的一个 (τ,λ,Σ_2)-加权核心。假设 $\tau > \frac{n}{2}$。如果 $\|w-v\|_1 \le \rho$，则

$$\|\mu_w - \mu_v\|_{\Sigma_1}^2 = \left\|\sum_{i\in[n]} w_i x - \sum_{i\in[n]} v_i x_i\right\|_{\Sigma_1}^2 \le (1+2\gamma)\rho^2\lambda.$$

证明 对于任意 $i,j \in \mathrm{supp}(w) \cup \mathrm{supp}(v)$，根据定义，存在大小至少为 τ 的“邻近”数据点集合$N_i, N_j \subseteq R$。由于 $\tau > \frac{n}{2}$，我们知道 N_i 和 N_j 有非空交集。因此，存在一点 $\iota \in R$，使得$\|x_i - x_\iota\|_{\Sigma_1}^2 \le \lambda$ 且$\|x_\iota - x_j\|_{\Sigma_2}^2 \le \lambda$。根据关于 Σ_1, Σ_2和 γ的假设，第二个不等式还推出$\|x_\iota - x_j\|_{\Sigma_1}^2 \le (1+2\gamma)\lambda$。现在，我们使用（未平方的）范数，先加上再减去 x_ι，并应用三角不等式。对于任意 $i,j \in \mathrm{supp}(w)\cup\mathrm{supp}(v)$,

$$\begin{aligned}\|x_i - x_j\|_{\Sigma_1} &= \|x_i - x_\iota + x_\iota - x_j\|_{\Sigma_1}\\ &\le \|x_i - x_\iota\|_{\Sigma_1} + \|x_\iota - x_j\|_{\Sigma_1}\\ &\le 2\sqrt{(1+2\gamma)\lambda}.\end{aligned}$$

设 C 是服从分布 w的多元随机变量，并设随机变量 D服从 v。因此，$\|\mu_w - \mu_v\|_{\Sigma_1}$ 就是 $\|\mathbb{E}[C-D]\|_{\Sigma_1}$。分布 w和 v之间的全变差距离至多为 $\rho/2$（因为全变差距离恰好是 ℓ_1 距离的一半），因此存在 w 和 v 的最大耦合，以及随机变量 C'、 D'，使得 $\mathbb{E}[C-D] = \mathbb{E}[C'-D']$ 且 $\Pr[C' \ne D'] \le \rho/2$。我们对范数应用 Jensen 不等式，并舍去期望中对应于 $C'=D'$的部分，因为该部分为零:

$$\begin{aligned}\|\mu_w - \mu_v\|_{\Sigma_1} &= \left\|\mathbb{E}[C'-D']\right\|_{\Sigma_1}\\ &\le \mathbb{E}\left[\left\|C'-D'\right\|_{\Sigma_1}\right]\\ &= \Pr[C'\ne D']\cdot\mathbb{E}\left[\left\|C'-D'\right\|_{\Sigma_1} \mid C'\ne D'\right].\end{aligned}$$

利用 $\mathbf{Pr}[C' \neq D]$ 的上界以及期望内部各取值的均匀分布上界，我们得到

$$\|\mu_w - \mu_v\|_{\Sigma_1} \leq \frac{\rho}{2} \cdot 2\sqrt{(1+2\gamma)\lambda} = \rho\sqrt{(1+2\gamma)\lambda}.$$

平方后得到上界 $(1+2\gamma)\rho^2\lambda$。 ■

推论 36 表明，稍作修改后，同样的论证也适用于在相邻数据集上计算核心的情形。请注意，我们采用了稍强的假设，要求 $\tau > \frac{n}{2} + 1$ 而不是 $\frac{n}{2}$。

推论 55（推论 36 的重述）设 x 和 x' 是大小为 n 的相邻数据集。设 $\lambda > 0$、$\tau \in \mathbb{N}$ 和 $R \subseteq [n]$。设 Σ_1 和 Σ_2 是满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$（对 $\gamma \leq \frac{1}{2}$）的正定矩阵。设 w 是相对于 R 的 x 的（τ、λ、Σ_1）- 加权核心，且 v 是相对于 R 的 x' 的（τ、λ、Σ_2）- 加权核心，其中 $\tau > \frac{n}{2} + 1$。若 $\|w - v\|_1 \leq \rho$，则

$$\|\mu_w - \mu_v\|_{\Sigma_1}^2 = \left\| \sum_{i\in[n]} w_i x_i - \sum_{i\in[n]} v_i x_i' \right\|_{\Sigma_1}^2 \leq (1+2\gamma)\lambda(\rho + \|w\|_\infty + \|v\|_\infty)^2.$$

证明 [推论 36] 的证明：由 w 和 v，我们构造出关于 $x \cup x'$ 的核心 w^+ 和 v^+，并应用引理 54。

令 $i^* \in [n]$ 为 x 与 x' 不同的索引。我们的数据集是有序的，并非严格意义上的集合，因此将 $x \cup x' = (x_1, x_2, \ldots\ldots, x_n, x'_{i^*})$ 视为大小为 $n+1$ 的数据集。令 $R^+ \leftarrow R \cup \{n+1\}$ 为相应扩展后的参考集。显然，$w^+ = (w_1, \ldots\ldots, w_n, 0)$ 是关于 R^+ 的 $x \cup x'$ 的 (m, λ, Σ_1)-核心。同理，

$$v^+ = (v_1, \ldots, v_{i-1}, 0, v_{i+1}, \ldots, v_n, v_i)$$

它是关于 R^+ 的 $x \cup x'$ 的 $(\tau, \lambda, \Sigma_2)$-核心。引理 54 要求 $m > \frac{n+1}{2}$，而这一点满足：我们已假设 $m > \frac{n}{2} + 1$。因此只需计算 $\|w^+ - v^+\|_1$，这很简单：

$$\begin{aligned}
\left\|w^+ - v^+\right\|_1 &= \sum_{j\in[n+1]} \left|w_j^+ - v_j^+\right| \\
&= \left|w_i^+ - v_i^+\right| + \left|w_{n+1}^+ - v_{n+1}^+\right| + \sum_{j\in[n+1]\setminus\{i,n+1\}} \left|w_j^+ - v_j^+\right| \\
&= \left|w_i^+ - 0\right| + \left|0 - v_{n+1}^+\right| + \sum_{j\in[n+1]\setminus\{i,n+1\}} \left|w_j^+ - v_j^+\right| \\
&\leq \|w\|_\infty + \|v\|_\infty + \|w - v\|_1 .
\end{aligned}$$

■

引理 56（引理41的重述）固定 $\lambda_0 > 0$、$k \in \mathbb{N}$ 和 $R \subseteq [n]$。设 x 和 x' 为相邻数据集。设 Σ_1 和 Σ_2 为满足 $(1-\gamma)\Sigma_1 \preceq \Sigma_2 \preceq \frac{1}{1-\gamma}\Sigma_1$（对 $\gamma \leq \frac{1}{2}$）的正定矩阵。进一步假设 $k \leq \frac{1}{2\gamma}$。则 $|g(x、\Sigma_1) - g(x'、\Sigma_2)| \leq 2$。

证明不失一般性，假设 $g(x, \Sigma_1) \leq g(x', \Sigma_2)$。我们将通过分析两种情况证明 $g(x', \Sigma_2) \leq g(x, \Sigma_1) + 2$。

情况 1： 假设 $g(x, \Sigma_1) = k$。那么，根据构造，$g(x', \Sigma_2) \leq g(x, \Sigma_1) \leq g(x, \Sigma_1) + 2$，因为 $g(\cdot)$ 的上限为 k。

情况 2： 假设 $g(x, \Sigma_1) < k$。这只有在 $g(x, \Sigma_1) < k$ 时才会发生，因此存在一个$\ell^* \in \{0, \ldots, k\}$和子集 S_{ℓ^*}，使得：(i) 相对于 R，S_{ℓ^*} 是 x 的$(|R| - \ell, e^{\ell/k}\lambda_0, \Sigma_1)$-核心；(ii) 满足 $n - |S_{\ell^*}| + \ell^* < k$。此外，我们知道 $\ell^* \neq k$，因为 $n - |S_{\ell^*}|$ 是非负的。

现在我们可以应用引理 39。对于 $\ell \in \{0, \ldots, k\}$，令 T_ℓ 为关于 R 的最大的$(|R| - \ell、e^{\ell/k}\lambda_0、\Sigma_2)$-核心。引理 39 表明，$S_{\ell^*} \setminus i^* \subseteq T_{\ell^*+1}$。这使我们能够对 $g(x', \Sigma_2)$ 给出上界：

$$\begin{aligned} g(x', \Sigma_2) \leq g(x', \Sigma_2) &= \min_{\ell \in \{0,\ldots,k\}} n - |T_\ell| + \ell \\ &\leq n - |T_{\ell^*+1}| + \ell^* + 1 \\ &\leq n - (|S_{\ell^*}| - 1) + \ell^* + 1 \\ &= g(x, \Sigma_1) + 2. \end{aligned}$$

因此证明完成。 ■

命题57（命题10的重述）固定 $0 < \varepsilon \leq 1$ 和 $0 < \delta \leq \frac{\varepsilon}{10}$。存在一个算法 $\mathcal{M}^{\varepsilon, \delta}{}_{\mathrm{PTR}} : \mathbb{R} \to \{\text{通过，失败}\}$，满足以下条件：

(1) 令 $\mathcal{U}$ 为一个集合，令 $g: \mathcal{U}^n \to \mathbb{R}_{\geq 0}$ 为一个函数。如果对于所有仅有一个条目不同的 x、$x' \in \mathcal{U}^n$，都有 $|g(x) - g(x')| \leq 2$，那么 $\mathcal{M}^{\varepsilon, \delta}{}_{\mathrm{PTR}}(g(\cdot))$ 是 (ε, δ)-差分隐私。

(2) $\mathcal{M}^{\varepsilon, \delta}{}_{\mathrm{PTR}}(0) =$通过。*(3)* 对所有 $z \geq 2^{\frac{1-\delta}{\log_\varepsilon 1/\delta} + 4}$，$\mathcal{M}^{\varepsilon, \delta}{}_{\mathrm{PTR}}(z) =$失败。

证明 令 $\tau = 2^{\frac{1-\delta}{\log_\varepsilon 1/\delta} + 4}$，并令 $p(z)$ 表示 $\mathcal{M}^{\varepsilon, \delta}{}_{\mathrm{PTR}}(z)$ 通过的概率，定义如下：

$$p(z) = \Pr\left[\mathcal{M}^{\varepsilon,\delta}_{\mathrm{PTR}}(z) = \texttt{PASS}\right] = \begin{cases} 1 & \text{if } z = 0 \\ 0 & \text{if } z \geq \tau \\ 1 - e^{\frac{\varepsilon}{2}(z-2)}\delta & \text{otherwise} \end{cases}.$$

$p(z)$ 必须具备哪些性质？根据构造，它满足第二和第三个条件，因此我们只需论证隐私性，并针对 $|z - z'| \leq 2$ 考虑 $p(z)$ 和 $p(z')$。不失一般性，假设 $z \leq z'$。由于 $p(\cdot)$ 单调递减（因此 $p(z) \geq p(z')$），我们必须证明

$$p(z) \leq e^{\varepsilon} p(z') + \delta.$$

由于 $z' \leq z + 2$，我们知道 $p(z') \geq p(z + 2)$，因此只需证明

$$p(z) \leq e^{\varepsilon} p(z + 2) + \delta.$$

我们将分情况证明这一点，首先处理边界（接近 0 和 τ 的部分）。（注意，为简便起见，引理陈述中的 τ 略大于 $2\log_{\varepsilon}\frac{1-\delta}{\delta}+4$；我们将证明更强的结论。）

情况 1： 假设 $z = 0$。则 $p(z) = 1$ 以及 $p(z+2) = 1 - e^{\frac{\varepsilon}{e}(2-2)\delta} = 1 - \delta$. 因此 $p(z) \le p(z+2) + \delta \le e^{\varepsilon}p(z+2) + \delta$。

情况 2： 假设 $z \ge \tau - 2$，因此 $p(z+2) = 0$。我们计算

$$\begin{aligned} p(z) \le p(\tau-2) &= 1 - e^{\frac{\varepsilon}{2}(\tau-2-2)\delta} \\ &= 1 - e^{\frac{\varepsilon}{2}\left(\frac{2\log\frac{1-\delta}{\delta}}{\varepsilon}\right)\delta} \\ &= 1 - \frac{1-\delta}{\delta}\cdot\delta = \delta. \end{aligned}$$

因此 $p(z) \le p(z+2) + \delta \le e^{\varepsilon}p(z+2) + \delta$。

情况 3： 现在假设 $0 < z < \tau - 2$ 并设 $\delta' \overset{\text{定义为}}{=} e^{\frac{\varepsilon}{2}(z-2)\delta}$，为简洁起见。由于 $0 < z$，我们有 $\delta' \ge e^{-\varepsilon}\delta$。我们计算

$$p(z) = 1 - \delta' \quad \text{and} \quad p(z+2) = 1 - e^{\frac{\varepsilon}{2}(z+2-2)}\delta = 1 - e^{\varepsilon}\delta'.$$

现在我们 $e^{\varepsilon}p(z+2) - p(z)$，并证明其为正：

$$\begin{aligned} e^{\varepsilon}p(z+2) - p(z) &= e^{\varepsilon}\left(1 - e^{\varepsilon}\delta'\right) - \left(1 - \delta'\right) \\ &= e^{\varepsilon} - e^{2\varepsilon}\delta' - 1 + \delta' \\ &= (e^{\varepsilon} - 1) - \delta'\left(e^{2\varepsilon} - 1\right). \end{aligned}$$

当 $\frac{e^{\varepsilon}-1}{e^{2\varepsilon}-1} \ge \delta' \ge e^{-\varepsilon}\delta$ 时，该值为正。由于 $\varepsilon \le 1$，该结论成立。注意：

$$\begin{aligned} \frac{e^{\varepsilon}-1}{e^{2\varepsilon}-1} &\ge \frac{\varepsilon}{e^{2\varepsilon}-1} \\ &\ge \frac{\varepsilon}{e^2-1}. \end{aligned}$$

当 $\delta \le \frac{e^{\varepsilon}}{e^2-1}\cdot\varepsilon$ 时，该值大于 δ'。利用 $\frac{e^{\varepsilon}}{e^2-1} \ge \frac{1}{10}$ 即可完成证明，因为我们假设了 $\delta \le \frac{\varepsilon}{10}$。 ■