arXiv:2601.02257v1 [cs.CR] 5 Jan 2026

通过矩阵分解提升全动态流中的隐私持续基数估计精度

Joel Daniel Andersson* Palak Jain† Satchit Sivakumar†

2026年1月6日

摘要

我们研究全动态持续观测模型中的差分隐私统计，其中每个时间步都可能到达多个更新，并且对数据流的更新既可以涉及插入，也可以涉及删除某个项目。早期工作（例如，Jain 等人发表于 NeurIPS 2023 的不同元素计数研究；Raskhodnikova与Steiner发表于 PODS 2025 的带边更新的三角形计数研究）将相应的基数估计问题归约为对与输入流真实函数值相关的<style id='3'>差分流进行持续计数。在此类归约中，原始流的一次变化可能导致差分流发生多次变化，这给应用私有持续计数算法以获得最优误差界带来了挑战。我们通过研究所得差分流所对应的 ℓ_p敏感度向量并提取其性质，提升了若干此类归约的精度。

我们证明，我们的框架在不同元素计数、度直方图估计以及三角形计数（采用略微宽松的隐私模型）方面给出了更优的界，从而为流式环境中的隐私持续基数估计提供了一种通用方法。我们的精度提升源于对该场景下计数矩阵已知分解机制的紧致分析；关键技术挑战在于论证，我们可以将最先进的分解用于具有我们所归纳出的性质的敏感度向量集。特别是，我们证明，对于近似差分隐私，著名的计数矩阵平方根分解（Henzinger 等人，SODA 2023；Fichtenberger等人，ICML 2023）可用于实现相较于以往基于二叉树机制的工作在精度上的具体提升。我们还对这些敏感度模式下带减法的 b 叉树机制进行了紧致分析，其中包括一种用于计算 ℓ_p 敏感度的多项式时间例程，从而为纯差分隐私和近似差分隐私提供时间与空间高效的算法。通过实证和分析，我们证明，在广泛的参数范围内，我们改进后的误差界显著提升了基数估计问题的精度。

*奥地利科学技术研究所。joel.andersson@ist.ac.at。
†波士顿大学计算机科学系。{palakj,satchit}@bu.edu。

目录

1 引言 **3**

2 我们的结果 42.1 低空间误差界.6

3 技术概述 **7**
3.1 框架概述.7 3.2 证明中使用的技术.10 3.3 分析分解机制的挑战.11

4 相关工作 **12**

5 预备知识 14
5.1 因子分解机制.14 5.2 感兴趣的问题.16

6 全动态流的敏感度向量集 17
6.1 基数问题中敏感度向量的性质.18 6.2 有界贡献下的敏感度向量集.20 6.3 根据持续计数的误差界推导基数估计的误差界 23

7 从$\mathcal{S}_{D,k}$ 到 $\mathcal{S}_{1,k}$的归约 **25**

8 托普利兹分解 **28**
8.1 作为二次型的敏感度与矩阵性质.29 8.2 敏感度上界.30 8.3 平方根敏感度的下界.34

9 基于树的分解 **39**
9.1 b元树机制简介.39 9.2 通过奇偶性计数研究敏感度.42 9.3 误差界.47

10A 朴素分解 49

11 界的解析比较 **50**

12 结论 **51**

1 引言

隐私是处理敏感数据的系统面临的一项核心挑战，尤其是在数据持续变化、必须随时间进行监测或使用的场景中。差分隐私的持续观测模型 [13, 23] 正是在这类场景下为个体隐私提供了严格保证。在本研究中，我们重新探讨了持续观测条件下的差分隐私基数估计问题，研究的场景颇具挑战性：完全动态 场景，其中更新可能涉及插入或删除。

我们的主要贡献是证明，对分解机制进行新的分析，可以在差分隐私的持续观测模型中，为若干问题带来切实的准确性提升。为此，我们：（1）定义并分析一类新的结构化敏感度向量，这类向量出现在将基数估计问题归约为持续计数的过程中，例如不同元素计数（CountDistinct）、发布度直方图（DegreeCount）和发布三角形计数（TriangleCount）；（2）针对这类新的敏感度向量，证明基于托普利兹分解的持续计数分解机制的误差界；（3）表明，即使是基于树聚合的分解机制，在适配我们所识别的结构化敏感度向量后，也能获得更好的误差保证。

持续计数的改进机制 比特求和，也称为“持续计数”，是各种数据分析任务中的基本构件。针对这一问题的二叉树机制 [23, 13] 具有渐近误差，其与最优误差相差不超过对数因子。然而，尽管二叉树机制被广泛使用，其误差保证中的常数并非最优 [30, 40], ，这在实际应用中可能产生显著影响 [46, 20]。

因此，长期以来，许多研究致力于提高持续计数中差分隐私算法的实际精度。无论是否以这种形式明确呈现，所有这些方法都可以理解为分解机制 [50]。[1] 在这一框架中，持续计数问题的目标工作负载（时间范围 T内的前缀和查询）由下三角全1矩阵 $A \in \mathbb{R}^{T\times T}$表示。然后，对于选定的分解 $LR = A$，机制通过高斯机制估计中间查询 R，再使用线性变换 L 对其进行后处理，从而得到原始查询的答案。

这类研究有两种方向：第一种研究与 A的稀疏分解相对应的层次化方法。也就是说，这类方法类似于二叉树机制，通常结合更巧妙的聚合函数，并对树结构 [33, 54, 41, 2, 4]加以改进。第二类研究则探讨稠密分解 $A = LR$ [20, 30, 40, 21, 3, 38, 35]，其中 R 缺乏简单的组合结构，而是具有其他形式的结构（例如，R 可能具有托普利兹结构）。这些方法能够在持续计数中实现近乎最优的实际精度。然而，正如下文所述，我们无法将这些改进直接应用于基数估计。

无法直接使用改进机制对于某些应用（例如差分隐私机器学习），持续计数算法运行在这样的数据流上：对于相邻输入流，输入仅在恰好一个（或少数几个）时间步上存在有界差异。在这种情况下，可以直接利用持续计数的改进机制，以获得更好的效用。

在不同元素计数或发布图的度数列表等基数估计问题中，归约到持续计数的过程更加复杂。这类问题常用的一种策略是构造差分流$d_f[t] = f(\vec{\mathbf{x}}[0\colon t]) - f(\vec{\mathbf{x}}[0\colon t-1])$，并在该流上运行连续计数算法 [56, 29, 43, 55]。

其中一项工作是 Jain 等人 [43]，将这一策略用于不同元素计数。不幸的是，对于这一任务，在一个时间步对输入流的更改 $\vec{\mathbf{x}}[t]$ 可能会导致差分流中多达 T处更改（在元素既可以插入也可以删除的全动态流中）。因此，一个朴素的 ε-差分隐私机制，即使用 ε/T-差分隐私机制进行持续计数，以发布该差分流的前缀和（通过群体隐私提供 ε-差分隐私保证），将导致误差相较于持续计数增加一个倍数

[1] 有时也称为矩阵机制。

T。其他基数估计问题也会出现类似现象，包括图流中的三角形计数和度列表估计。[2]

为了解决这一问题，Jain 等人 [43] 首次针对不同元素计数提出了参数化的准确度保证。他们定义了数据流的最大翻转度概念，并证明最大翻转度有界于 k 意味着，在某个时间步对输入流进行一次更改 $\vec{\mathbf{x}}[t]$，最多可能导致差分流发生 k 次更改。随后，类似的群体隐私论证将使最大翻转度具有保证上界 k 的数据流产生 k 倍的误差放大。[3] Jain 等人通过以非黑盒方式使用二叉树机制，将最大翻转度的影响降低至 $\sqrt{k}$ 倍。

遗憾的是，他们的分析专门针对二叉树机制（该机制的误差界常数并不理想）；正如他们在论文中指出的那样，目前尚不清楚是否可以用噪声添加更少的机制（例如复杂的托普利兹分解机制）替代该机制。他们的分析也针对不同元素计数进行了定制，未能为其他基数估计问题（如度数计数）获得类似的准确度保证提供思路。这促成了本文的主要问题。

我们能否通过用其他持续计数机制替代二叉树机制，提高持续观测下差分隐私基数估计的准确性？

我们对此给出了肯定的回答，并显著提高了若干基数估计问题的准确性。

基数估计问题的机制是更复杂任务的基础构件，因此，为这些问题实现更优的误差界，对于推动差分隐私的实际应用至关重要。例如，不同元素计数用于检测网络中的异常 [1],、计算物种间的遗传差异 [6] 等；三角形计数用于检测和监测社交网络中社群的凝聚性 [25], 、检测网页图中是否存在垃圾信息活动 [7] 等；度数计数用于追踪性传播疾病的传播信息 [14], 、计算在线社交网络的参数[19] 等。事实上，对于标准持续计数，通过新颖的分解机制 [46, 20, 30, 40, 21, 52] 提升准确性的常数因子改进，已推动其在隐私机器学习系统中的应用 [58, 32]。

2 我们的结果

我们的结果着重于持续发布机制 $\mathcal{M}$在每个时间步 $t \in [T]$ 处理流并生成答案 a_t时，针对函数[4] f 的两种误差概念：平方误差期望均值的平方根 $\mathrm{MeanSE}(\mathcal{M}, T) = \sqrt{\mathbb{E}\left[\frac{1}{T}\sum_{t\in[T]}(f(\vec{\mathbf{x}}[0\colon t]) - a_t)^2\right]}$，以及平方误差期望最大值的平方根 $\mathrm{MaxSE}(\mathcal{M}, T) = \sqrt{\max_{t\in[T]}\mathbb{E}\left[(f(\vec{\mathbf{x}}[0\colon t]) - a_t)^2\right]}$。在本结果概述中，我们着重讨论 MaxSE；相关结果汇总于表1。均方根误差的结果类似，将在本文后文讨论。我们的隐私保证以零集中差分隐私（zCDP）表示，形式化定义见第 5 节。

不同元素计数。针对项目级隐私下的不同元素计数（相邻流在单个元素的出现次数上有所不同）， [43] 的先前工作采用二叉树机制，得到了一种 $1/2$-零集中差分隐私算法，其平方误差的均方根最大值约为 $\sqrt{k}\log T$，其中 k表示流的最大翻转度（任一元素在出现与

[2]注意，这种算法的误差会大于简单基线：在每个时间步向函数值添加高斯噪声，并为 T 个时间步的组合付出代价。[3]注意，这样便能为最大翻转度不超过 k 的数据流提供隐私和准确性；通过预处理步骤，可以将隐私保证扩展到所有数据流。[4]此处的定义针对一维函数 f。如果 f 是向量值的（例如，度数计数），那么均方根误差和均方根最大误差定义为：将机制输出限制在每个坐标上，分别计算均方根误差或均方根最大误差，再取这些值中的最大值。

表1：我们的方法（MaxSE）与先前工作在长度为 T 的数据流上针对 $1/2$-零集中差分隐私机制所得的准确度上界比较。这些结果通过结合定理 8.1、定理 6.14、定理 6.15 和定理 6.16 得出。

问题	论文	准确度所用参数 度	准确度 上界	隐私 for all 数据流?
不同元素计数	[43]	最大翻转度 k_{cd}	$\sqrt{k_{\mathrm{cd}}}$对数 $T, \sqrt{T}$	Yes
	我们的工作	最大翻转度 k_{cd}	$\frac{\pi^{\frac{1}{\text{对数}_2 e}\sqrt{k_{\mathrm{cd}}}\text{对数}}}{T}$	Yes
度数计数	[55]	流长度 T	$\sqrt{T}$	Yes
	我们的工作	度数贡献 k_{deg} (定理 6.9)	$\frac{\sqrt{2}}{\pi \log_2 e}\sqrt{k_{\mathrm{deg}}}\log T$	Yes
三角形计数	[55]	流长度 T	$T^{3/2}$	Yes
	我们的工作	三角形贡献 k_{tri} (定理 6.11), 度 D	$\frac{\pi^{\frac{1}{\text{对数}_2 e}\sqrt{k_{\mathrm{tri}}D}\text{对数}}}{T}$	No

未出现之间切换的最大次数）。对于现实流，我们预计 $k \ll T$——例如，流媒体服务上的用户账户在“已登录”和“已注销”之间切换的次数，远少于状态更新的总次数。

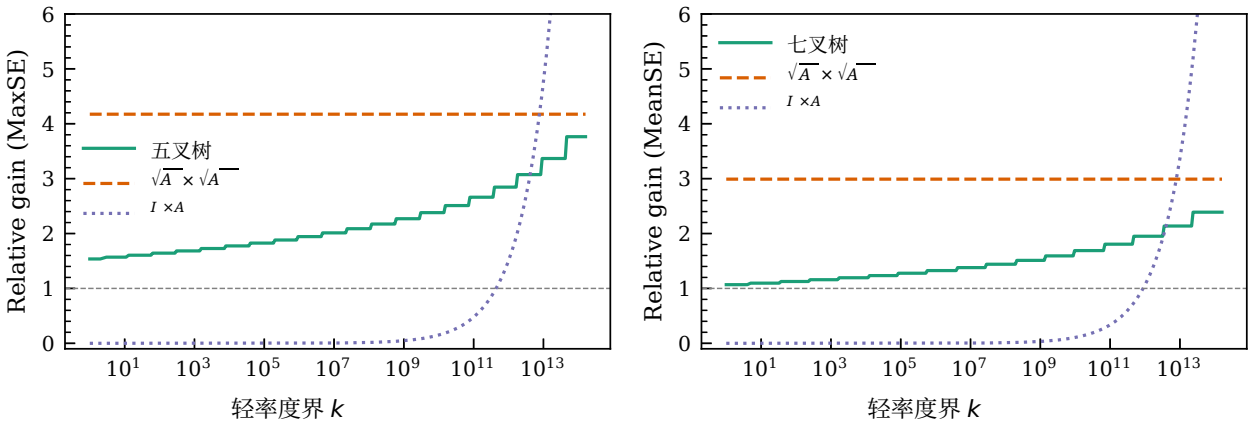
我们的工作以相同参数给出了上界，但将常数改进了约 4 倍——确切的相对比较见图1。

度直方图。 在项目级隐私下发布度直方图时（相邻图流在单条边的所有出现位置上均不同），此前最优的算法依赖朴素组合（在每个时间步重新计算直方图），其误差随 T 呈多项式增长。此外，对于与 MaxSE [5] 密切相关的误差概念， [55] 的先前工作以时间范围 T 为参数，确立了严格的多项式精度下界。因此，我们的工作转而基于一个新参数实现精度：我们引入的<我方最大度贡献>，即一条边在出现与缺失之间切换的最大次数。在许多自然图流中，这一数量远小于时间步数 T。例如，在随时间记录用户购物车重叠情况、以此捕捉用户相似度的数据流中，任意两名用户在共享与不共享商品之间切换的时间步数，通常远少于时间步总数。

三角形计数。 在项目级隐私下发布三角形计数（相邻图数据流在单条边的所有出现位置上均不同），在我们的工作之前，已知的最佳算法依赖于组合（误差为 T 的多项式级别）；而 [55] 的先前工作则针对与 MaxSE 密切相关的误差概念，证明了关于时间范围 T 的强精度下界。我们的工作则基于一个新定义的参数来获得精度：最大三角形贡献，即在整个数据流中，包含某条特定边的三角形数量的总变化最大值。对于度为 D 、经历 T 次更新的图，我们预计该值远小于 DT。

表1所反映的一项局限是，我们的保证并不能为所有 图提供隐私保护：它仅适用于度受公开参数 D 限制、且三角形贡献受公开参数 k_{tri}限制的图。相比之下，采用朴素组合会导致精度显著降低，但能为所有数据流提供隐私保护。值得注意的是，这一结果与一系列关于仅插入数据流图隐私的研究相呼应 [29],：这些研究仅为有界度数的图提供隐私保护。

[5]他们采用了误差的 ℓ_∞ 概念—— $\mathbb{E}[\max_{t\in[T]} \|f(\vec{\mathbf{x}}[0:t]) - a_t\|_\infty]$。我们的上界在这一误差概念下也能带来更小的误差；我们将在第 4 节进一步讨论。



(a) MaxSE 的相对改进。

(b) MeanSE 的相对改进。

图1：在固定轻率度界 k 的情况下，不同机制对 CountDistinct 进行 $T = 2^{50}$ 个时间步处理时，根节点最大预期平方误差和根节点平均预期平方误差的比较。误差以 [43] 达到的精确误差界为基准绘制。对于带减法的 b 叉树，使用定理 9.15 绘制其上界（包括低阶项），其中每种树均取渐近最优的 b。平方根分解的上界使用定理 8.1（包括低阶项）；朴素分解的误差为精确值（见定理 10.2）。

我们通过统一的分析得到上述结果。该分析提炼出基数估计问题中敏感度向量的一组性质，并通过一种新颖的分析说明，如何将最先进的托普利兹分解用于具有这类敏感度向量的问题。具体而言，我们预计，我们的结果将有助于提升上述问题以外的基数估计问题的精度上界（例如，统计满足特定性质的输入元素数量（比如，获得超过 1000 名用户五星评价的电影数量），或统计某个图中某种子图的出现次数等）。

2.1 低空间误差界

我们注意到，上述使用基于托普利兹分解的机制得到的结果有一个缺点：实现它需要 $\Omega(T)$ 空间，而基于树的方法只需 $O(\log T)$ 空间即可实现。[6] 因此，我们给出了利用改进的基于树的方法和新颖分析技术的结果，在将空间维持在 $O(\log T)$ 的同时提升精度。结果汇总于表2。

需要注意的一个重要问题是：要在 CountDistinct 和 DegreeCount 中为所有数据流提供隐私保护，同时在现实流上保持较高精度，需要通过预处理步骤跟踪数据流中感兴趣的参数，并相应地修改数据流，使其满足该参数的上界。目前尚不清楚如何在使用低于时间范围 T 的多项式阶空间的同时跟踪这些感兴趣的参数。因此，仅使用基于树的方法，无法在这些问题中同时实现对所有数据流的隐私保护和对数空间复杂度。

尽管如此，这些结果仍意义重大，因为它们将问题简化为在低空间内跟踪参数。值得注意的是，对于不同元素计数，近期研究 [18] 提出了一种跟踪最大翻转度变体的方法[7]，所需空间为 $\tilde{O}(T^{1/3})$；而采用本节介绍的基于树的方法，则可以提高精度

[6]有实证研究 [52] 表明，对于 T 的广泛取值范围，具备我们所需性质的托普利兹分解可以在多对数空间内实现。在某些情形下也已知相应的理论结果 [21], ，但这些研究中的分解并不满足我们所需的性质。[7]Cummings 等人 [18] 将这个量称为“出现次数”，它大致对应于单个元素的最大更新次数。

表2：使用 $1/2$-零集中差分隐私树聚合机制，对长度为 T 的数据流进行估计时，我们的（MaxSE）准确性上界比较。这些结果结合定理 9.15、定理 6.14、定理 6.15 和定理 6.16 得出。

问题	论文	准确性中使用的参数	准确性上界
不同元素计数	[43]	最大翻转度 k_{cd}	$\sqrt{k_{\mathrm{cd}}}$对数 $T, \sqrt{T}$
	我们的工作	最大翻转度 k_{cd}	$0.609\sqrt{k_{\mathrm{cd}}\text{ 对数 }T\text{ 对数}(T/k_{\mathrm{cd}})}$
度数计数	[55]	流长度 T	$\sqrt{T}$
	我们的工作	度数贡献 k_{deg}（定理 6.9）	$0.609\sqrt{2k_{\mathrm{deg}}\text{ 对数 }T\text{ 对数}(T/k_{\mathrm{deg}})}$
三角形计数	[55]	流长度 T，度 D	$T^{3/2}$
	我们的工作	三角形贡献 k_{tri}（定理 6.11），度 D	$0.609\sqrt{k_{\mathrm{tri}}D\text{ 对数 }T\text{ 对数}(DT/k_{\mathrm{tri}})}$

，同时保持亚线性空间；相比之下，像上一节那样采用托普利兹分解，则无法实现亚线性空间。

我们对表2中的结果作如下说明，该表以与表1相同的参数给出了界。首先，主导常数约为 0.6，优于通过二叉树机制得到的结果（见图1）。相比之下，该常数比通过托普利兹分解得到的结果大约差 2 倍。不过，我们对基于树的算法的分析相较于以往对基于树的算法的分析（以及我们通过托普利兹分解得到的上界）有渐近改进：对 T 的依赖为 $\sqrt{\log T \log T/k}$，而不是 $\log T$。对于较小的 k，这种差异并不显著；但对于较大的 k（相对于 T 为多项式级），这意味着基于树的界在渐近意义上优于我们对托普利兹分解得到的上界。我们将在第 11 节中更详细地比较这些界。

我们还证明，基于树聚合的分解机制针对 $(\varepsilon,0)$-差分隐私给出了当前最优界（其界渐近优于通过群体隐私所能得到的界）。有关这些界的详细说明，请参阅第 9.3 节。

下一节将讨论我们用于获得上述所有结果的技术。

3 技术概述

3.1 框架概述

为了证明上一节的结果，我们首先刻画与我们关注的问题相关的敏感度向量的性质。随后，我们将证明适用于敏感度向量满足这些性质的问题的一般结果，由此构建一个统一框架，并据此获得所有基数估计问题的结果。

对于函数 f，设 $\vec{\mathbf{x}}$和 $\vec{\mathbf{y}}$是两个相邻流，并设 df是一个函数，它接收流$\vec{\mathbf{x}}$并生成相应的差分流 $df(\vec{\mathbf{x}})[t] = f(\vec{\mathbf{x}}[0\colon t]) - f(\vec{\mathbf{x}}[0\colon t-1])$。由于我们希望将持续计数机制应用于差分流，因此研究中的一个关键对象是敏感度向量 $\vec{\Delta}_f = df(\vec{\mathbf{x}}) - df(\vec{\mathbf{y}})$。请注意，若函数 f是向量值函数，那么$df[t]$和 $\vec{\Delta}_f[t]$也都是向量值函数。

基数估计问题的敏感度向量集（第 6 节）。我们的首要贡献是刻画敏感度向量的性质，并利用这些性质分析我们关注的持续计数机制。

• 不同元素计数、度数计数和三角形计数的敏感度向量具有有界的区间和。对于不同元素计数和度数计数，我们发现了敏感度向量的一个关键性质，并在误差界的证明中加以利用——它们具有**交替的**性质。也就是说，非零元素交替出现；如果 $\vec{\Delta}_{\text{不同元素计数}}$ 的一个非零元素为 1，那么下一个非零元素必为 -1 （反之亦然）。类似地，对于度数计数，我们发现存在两个索引 a 和 b，使得敏感度向量 $\vec{\Delta}_{\text{度数计数}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 限制在这两个索引中的任意一个上时，也具有相同的交替性质。对于其他所有索引 c、$\vec{\Delta}_{\text{度数计数}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})[c] = 0^T$。我们注意到，“交替”性质也等价于敏感度向量的所有区间和都介于 -1 和 1 之间（定理 6.1）。对于三角形计数函数，我们证明，对于度为 D 的图序列，尽管该函数的差分流的敏感度向量不具有“交替”性质，但它们满足一个更一般的性质：所有区间和都介于 $-D$ 和 D 之间。

•现实不同元素计数、度数计数和三角形计数数据流的敏感度向量具有有界的ℓ_1范数。我们利用这一点为现实数据流获得更好上界的关键在于：单个元素的更新集合对输出统计量具有有界贡献。我们证明，这种有界贡献由敏感度向量的 ℓ_1 范数刻画。

对于全动态流模型中的不同元素计数，先前的工作 [43] 引入了流的最大翻转度$\text{Flip}(\vec{\mathbf{x}})$，其定义为某个元素在流中从出现变为缺失（或反之）所经历的最大时间步数。对于最大翻转度至多为 k 的数据流，我们观察到，相应的敏感度向量至多具有 k 个非零元素（k-稀疏）。对于度数计数，我们定义了一个新参数，将其称为边的“最大度贡献”，并证明，若最大度贡献有界于 k，则对于所有节点 c，$\vec{\Delta}_{\text{DegreeCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})[c]$ 都是 k-稀疏的。注意，对于这些具有 $\vec{\Delta} \in \{-1, 0, 1\}^T$、$k$-稀疏性的向量，其 ℓ_1 范数有界于 k。对于三角形计数，我们还定义了一个新参数，即数据流的“最大三角形贡献”，并证明该参数恰好等于某个敏感度向量的最大 ℓ_1 范数。

受我们所揭示性质的启发，我们将区间和介于 $-D$ 和 D之间、ℓ_1 范数不超过 k 的整数向量集合定义为敏感度向量集 $\mathcal{S}_{D,\ k}$。我们进一步将两个整数值求和流定义为：若它们的差属于$\mathcal{S}_{D,\ k}$，则它们是$\mathcal{S}_{D,\ k}$-邻接的。本文的技术论证将聚焦于$\mathcal{S}_{D,\ k}$-流上的持续计数；我们将其定义为相对于$\mathcal{S}_{D,\ k}$-邻接关系定义差分隐私的持续计数。

利用托普利兹分解机制（第8节）。现在，我们通过分解机制研究$\mathcal{S}_{D,\ k}$-流上的持续计数，从而为我们关注的基数估计问题给出上界。在这类机制中，时间范围 T 内的前缀和由下三角$T \times T$全1矩阵 A表示，因此，流 df的前缀和对应于矩阵向量乘积 $A \cdot df$。分解机制考虑分解 $A = LR$，并发布 $L(R \cdot d_f + \vec{\mathbf{z}})$，其中 $\vec{\mathbf{z}}$是独立同分布的高斯噪声向量，其方差按敏感度 $\text{sens}_2(R,\ \mathcal{S}_{D,\ k}) \equiv \max_{\vec{\Delta} \in \mathcal{S}_{D,\ k}} \|R\vec{\Delta}\|_2$成比例缩放。对此类机制的大量研究（见第 5 节）表明，采用适当的参数时，上述方法满足差分隐私，且可以在线方式实现。此外，此类机制的误差完全由敏感度 $\text{sens}_2(R,\ \mathcal{S}_{D,\ k})$和 L上的矩阵范数决定。我们的技术论证将分析前者，而后者则依赖已有研究成果。

我们证明以下主定理，该定理在对右分解矩阵 R 的某些限制条件下给出了 $\text{sens}_2(R,\ \mathcal{S}_{D,\ k})$ 的上界。

定理 3.1。设 T、k、$D \in \mathbb{N}$。设 R为一个下三角托普利茨矩阵，其下对角线元素非递增且非负。设 $\vec{\Delta} \in \mathcal{S}_{D,\ k}$。则，[8]

$$\|R\vec{\Delta}\|_2 \leq \sqrt{kD} \cdot \|R\|_{1\to 2}\,.$$

[8]注意，若 $k \leq D$，则 $\vec{\Delta}$ 上的区间和界是空泛的。在这种情况下，对于任意矩阵 R，$\text{sens}_2(R,\ \mathcal{S}_{D,\ k}) = k\|R\|_{1\to 2} \leq \sqrt{kD}\|R\|_{1\to 2}$ （例如，第一步可参见 [40, Fact 2.1] 的证明）。因此，该定理的有意义范围是 $k > D$。

其中， $\|R\|_{1\to 2}$ 是矩阵 R中任意一列的最大范数。

请注意， $k = D = 1$ 时的情形已在先前的工作中得到广泛研究 [30, 40, 39, 21, 35]，且该工作已证明，平方根分解满足 $L = R = \sqrt{A}$ ，具备应用定理 3.1 所需的性质，从而得到下述结果中的上界。

定理 3.2 （关于在 $\mathcal{S}_{D,\ k}$-流上持续计数的非正式误差界）。 固定 T、 k、 $D \in \mathbb{N}$。令 A 为下三角全 *1* 矩阵（大小为 $T \times T$）。那么，具有 $L = R = \sqrt{A}$ 的因子分解机制，在长度为 T 的 $\mathcal{S}_{D,\ k}$-流上实现 $1/2$-零集中差分隐私持续计数时，具有以下误差界：

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \le \left(\frac{1}{\pi \log(e)} + o(1)\right) \sqrt{kD} \log T\,,$$
$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \le \left(\frac{1}{\pi \log(e)} + o(1)\right) \sqrt{kD} \log T\,.$$

此外，如果 $k = O\left(T^{1/3}\right)$ 且 $k \ge D$，则有

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \ge \left(\frac{1}{\pi \log(e)} - o(1)\right) \sqrt{kD \log(DT/k) \log T}\,,$$
$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \ge \left(\frac{1}{\pi \log(e)} - o(1)\right) \sqrt{kD \log(DT/k) \log T}\,.$$

请注意，下界表明，对于 k 和 D 的广泛取值范围，我们对平方根分解的分析除低阶项外是紧的。我们猜测，下界对 T 的依赖（具体而言，$\sqrt{\log(DT/k)\log T}$ 与 $\log T$）是紧的；直接分析平方根矩阵（而不是通过我们针对一大类托普利兹矩阵的一般结果进行分解）或许能得到这一改进的上界。我们还注意到，当 $k = D = 1$ 时，我们的结果恢复了标准持续计数 [35] 中关于平方根分解的最紧上界。

将上述一般性结果谨慎应用于我们关注的问题，可得表1中的结果。

利用基于树的分解机制（第 9 节）。如第 2.1 节所述，为了在保持较低空间开销（关于时间范围呈对数级）的同时改善误差界，我们除了研究稠密托普利兹分解外，还研究树聚合分解机制。基于 [2,3]关于改进树聚合机制的研究，并利用我们^ ^ ^设定中树的额外冗余，我们引入一种分解 $A = L_b R_b$，并研究量 $\mathsf{sens}_2(R_b,\ \mathcal{S}_{D,\ k})$，从而刻画我们关注的误差概念。

定理 3.3 (基于树的机制的非正式误差界)。 设 $k > D, T, b \in \mathbb{N}$。对于长度为 T 的 $\mathcal{S}_{D,\ k}$-数据流上的 $1/2$-*zCDP* 持续计数，采用减法的 b 叉树机制（带有相关因子 ^ L_b、^ R_b）可达到以下误差界：

$$\mathsf{MaxSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) \le \left(\frac{\sqrt{b-1}}{\sqrt{2}\log b} + o(1)\right) \sqrt{Dk \log(DT/k) \log(T)}\,,$$
$$\mathsf{MeanSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) \le \left(\frac{\sqrt{b(1-1/b^2)}}{2\log b} + o(1)\right) \sqrt{Dk \log(DT/k) \log(T)}\,.$$

其中，括号中的常数分别在 $b = 5$ 和 $b = 7$ 处取得最小值，对应的值为0.609 和0.466。

我们给出了 $\mathsf{sens}_2(\hat{}R_b, \mathcal{S}_{D,k})$ 的闭式下界，与我们的上界仅相差低阶项。我们还给出了通过动态规划精确计算 $\mathsf{sens}_2(\hat{}R_b, \mathcal{S}_{D,k})$ 的多项式时间方法，在实践中可利用该方法获得精确上界。完整细节见第 9 节

3.2 证明中使用的技术

从 $S_{D,k}$ 数据流归约到 $S_{1,k}$ 数据流。 </style}贯穿我们证明过程的一项关键要素，是将对 ,$_k$ 中的数据流进行推理，归约为对来自$S_{1,k}$的数据流进行推理。我们在第 7 节证明如下结论：

定理 3.4. 设 $1 \le k \le D \le T$、 $1 \le m \le T \in \mathbb{N}$。设 R 为任意 $m \times T$ 矩阵。那么，

$$D \cdot \mathsf{sens}_2(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) \le \mathsf{sens}_2(R, \mathcal{S}_{D,k}) \le \max_{k_1,\ldots,k_D : \sum_{i=1}^{D} k_i = k} \sum_{i=1}^{D} \mathsf{sens}_2(R, \mathcal{S}_{1,k_i})$$

该结果的下界很直接，可由范数缩放的简单事实推出。 ⃗上界则更为复杂，需要将敏感度流 Δ 从 $S_{D,k}$ 拆分为至多 D个不同的敏感度流 $\vec{\Delta}^{(i)}$ 从 S_{1,k_i}，其中 $\sum_{i=1}^{D} k_i \le k$。每个流对应于 $\vec{\Delta}$ 的前缀和可能取到的一个不同值，并且这些流经过精心构造，使得 $\sum_{i=1}^{D} \vec{\Delta}^{(i)} = \vec{\Delta}$。结合这一分解的结构性质与基本范数不等式，即可得到上界。详情见第 7 节。

这一归约在我们的分析中至关重要，因为 $S_{1,k}$ 中的流比 $S_{D,k}$中的流具有严格得多的结构：它们是交替的，也就是说，它们的非零元素的大小为 1，且符号交替。

托普利兹分解机制的技术。 我们在此概述用于托普利兹分解的技术（完整细节请参见第8节）。

上界概述。 我们以不同风格给出定理 3.1 的两种证明。一种证明是组合证明，另一种则更偏代数。利用敏感度 $\mathsf{sens}_2(R,\mathcal{S}_{D,k})$ 与分解误差之间的已知联系，将其应用于平方根分解即可得到定理 3.2 的误差上界。

令 $c(i,j) = (R^\top R)[i,j]$。两种证明的出发点都是分析二次型 $\|R\vec{\Delta}\|_2^2$。作为前置内容，我们证明， R 的下对角线元素的单调性会诱导出$c(\cdot,\cdot)$的有用单调性，这些单调性将在两种证明中使用。

组合证明首先考虑交替的情形 $S_{1,k}$。我们将二次型分解为： （i）形式为 $c(i,i)\vec{\Delta}[i]^2$ 的“平方项”；以及（ii）涉及 $c(i,j)\vec{\Delta}[i]\vec{\Delta}[j]$（其中 $i \ne j$）的“交叉项”。关键步骤是证明交叉项非正。利用交替结构，当 $\vec{\Delta}[i]\vec{\Delta}[j] = 1$（其中 $i < j$）时，我们可以将这一对与另一个 $i < j'$ 配对，其中 $\vec{\Delta}[i]\vec{\Delta}[j'] = -1$ ，且j' 更接近 i。随后， $c(\cdot,\cdot)$ 的单调性使我们能够证明每个配对对交叉项的贡献均为非正。这有助于我们证明对于 $D = 1$ 成立的定理 3.1；随后，我们利用定理 3.4 中从 $S_{D,k}$ 数据流到 $S_{1,k}$ 数据流的归约，将其扩展到所有 D 。详细内容见第 8.2 节。

第二个证明采用线性代数的方式，直接对 $S_{D,k}$ 进行推理；我们认为这种方法更加优雅，但直观性要差得多。我们处理的一个关键量是 $\sum_{j=0}^{T-1} c(i,j)\vec{\Delta}[j]$。记作 $SUM_j = \sum_{i=0}^{j} \vec{\Delta}[i]$后，我们将 $\vec{\Delta}[j]$ 写成相邻前缀和 $SUM_{j+1} - SUM_j$之差，然后改变求和式-$\sum_{j=0}^{T-1} c(i,j)(SUM_{j+1} - SUM_j) = \sum_{j=0}^{T-1}[c(i,j) - c(i,j+1)] \cdot SUM_j$中的指标，这一变换通常称为阿贝尔求和。事实证明，利用 $c(\cdot,\cdot)$的单调性，可以更容易地对后一个量进行推理；我们谨慎地运用这一点来证明定理。详细内容请参见第 8.2 节。

下界概述。 对于平方根因子分解的敏感度 $\mathsf{sens}_2(R, \mathcal{S}_{D,k})$ 的下界，我们采用概率方法。首先考虑 $S_{1,k}$ 流：从 $S_{1,k}$ 中均匀随机抽取敏感度向量 Δ （以 $\|\vec{\Delta}\|_1 = k$ 为条件），并分析 $\mathbb{E}\left[\|\sqrt{A} \cdot \vec{\Delta}\|_2^2\right]$。我们将期望拆分为一个涉及 $\mathbb{E}[\vec{\Delta}[i]^2]$ 的平方项，以及一个涉及 $\mathbb{E}[\vec{\Delta}[i]\vec{\Delta}[j]]$ 的交叉项，其中 $i \ne j$。前者较易分析，因为 Δ 的每个坐标都以相同的概率为非零。证明的关键挑战在于界定交叉项，这归结为分析在条件为它们均非零时， $\vec{\Delta}[i] \ne \vec{\Delta}[j]$ 的概率。这对应于落在 i 和 j 之间的非零值数量为偶数的概率，也就是超几何随机变量的奇偶性。因此，我们证明了一个刻画该概率的结果（该结果或许本身也有独立价值），并用它来处理交叉项。控制交叉项的绝对值最终需要复杂的解析推导；详情请参阅读者

第 8.3 节。由此得到 $S_{1,\ k}$ 流的结果；推广到 $S_{D,\ k}$ 流同样需要用到定理 3.4。

基于树聚合的分解机制技术。 对于树聚合，我们直接利用树的组合结构，并再次从 $S_{1,\ k}$ 数据流入手，利用其交替性质。我们将敏感度分析（针对 $\mathsf{sens}_2(R,\ \mathcal{S}_{D,\ k})$）归约为一个新的组合问题，称为树上的奇偶计数：给定一棵具有 b 个分支的完全树，其包含 T 个叶节点，并将 k 个球放置在叶节点上，有多少个内部节点的子树中包含奇数个球？由于$S_{1,\ k}$数据流的交替结构，最大化该数量即可对树聚合矩阵的敏感度 $\mathsf{sens}_2(R,\ \mathcal{S}_{D,\ k})$ 给出精确刻画。

我们利用这一刻画推导出在低阶项意义下紧的上下界，并设计一种能够在多项式时间内精确计算 $\mathsf{sens}_2(R,\ \mathcal{S}_{D,\ k})$ 的动态规划算法。我们的组合视角还启发了一种针对当前场景设计的稀疏化树因式分解：由于标准树因式分解仅使用节点的 $(1-1/b)$ 比例，因此在计算敏感度时，可以忽略子树中包含奇数个球的节点中相同比例的节点。将这些洞见与针对 $S_{1,\ k}$ 数据流的归约相结合，即可得到定理 3.3 及相关结果；详见第 9 节。

3.3 分析因子分解机制的挑战

对于 $S_{D,k}$-数据流上的持续计数，$k = D = 1$ 的情形已在以往工作中得到广泛研究。在这种情况下，平方根分解可实现近似最优界 [30, 40]。因此，人们自然会思考，这些技术是否也能自然地推广到 $k > 1$（即使 $D = 1$）。本节将说明为什么有必要采用不同的分析方法。

任意因子分解机制的误差取决于 $\max_{\vec{\Delta}} \|R\vec{\Delta}\|_2$，从而形成一个自然的优化问题。当 $k = 1$时，该优化问题有一个简单的解；$\max_{\vec{\Delta}} \|R\vec{\Delta}\|_2$ 恰好等于 $\|R\|_{1\to 2}$，即 R各列上的最大 ℓ_2 范数，这一数值可以通过遍历各列高效计算。此外，对于平方根机制等结构化因式分解，R 是一个对角线单调递减的下三角托普利茨矩阵。这意味着最坏情况敏感度向量始终为 $[1, 0, \ldots, 0]$，因而可以直接计算敏感度 $\mathsf{sens}_2(R,\mathcal{S}_{1,1})$。处理更大的 k 时，一个自然的初步尝试可能是借助 R的算子范数，将问题归约为 $k = 1$。然而，这会导致对 k 的次优线性依赖（而我们实现的是平方根依赖）。

更高 k的计算挑战。为了在更大的 k 上取得进展，我们首先将注意力限定在交替敏感度向量这一类上——这些向量的元素均为整数，绝对值至多为 1，且非零值的符号交替变化；这一设定源于我们的基数估计问题所呈现的结构。然而，即使在这一简化设定下，直接计算 $\max_{\vec{\Delta}} \|R\vec{\Delta}\|_2$ 在计算上仍然难以承受：朴素的暴力枚举必须遍历 $\Theta\big(T^{\min(k,\,n-k)}\big)$ 个候选敏感度向量，而随着 k 增大，这显然不可行。一般而言，在离散集合上求解凸二次最大化问题可能是 NP 困难的。因此，与 $k = 1$ 设定不同，我们无法指望通过枚举来求解该优化问题。

通过矩阵结构识别最难的敏感度向量。 另一种思路是直接识别“最难”的敏感度向量，这在平方根分解中是可行的，条件是 $k = 1$。事实上，对不同 k 值下的平方根矩阵所做的实验呈现出一个令人鼓舞的启发式规律：最优交替 k-稀疏向量的非零元素往往分布得相当均匀。（关于 $T = 30$ 和 $k \in [1, 10]$，见图 2。）

遗憾的是，这些向量似乎没有简洁的闭式描述。交替约束会引入微妙的相互作用：请注意，为了优化 $\|R\vec{\Delta}\|_2$，我们关注的是 k 列的线性组合的范数（其中一些列相加，另一些列相减）。在 $\vec{\Delta}$ 的特定位置插入一个 -1，可能会增大 $R\vec{\Delta}$ 某些坐标的绝对值（例如，使原本较小的负值变得更负），同时又会减小其他坐标的绝对值（例如，通过抵消正向贡献）。这些影响如何权衡，取决于矩阵 R中的数值，且难以明晰。由于缺乏可用的闭式刻画，我们最终放弃了先前工作中的直接优化方法，

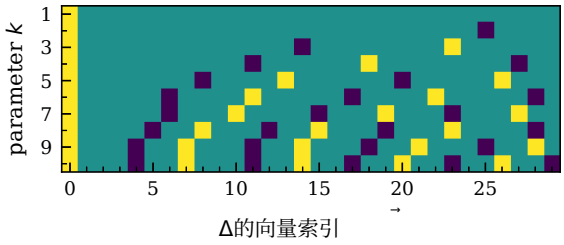


图 2：热力图展示了交替的 k-稀疏敏感度向量 $\vec{\Delta}$，它们对每个$k \in [1, 10]$ 和 $n = 30$均使 $\|\sqrt{A}\vec{\Delta}\|_2$ 达到最大值。黄色方格表示 1，紫色方格表示 -1，青绿色方格表示 0。

转而开发了前文所述的组合与代数上界技术，以及概率下界技术。

4 相关工作

持续观测下的差分隐私。 如引言所述，Dwork等人 [23] 和 Chan等人 [13]的开创性工作定义了差分隐私的持续观测模型：输入以流的形式接收，目标是在保护隐私的同时，持续跟踪某项统计量。他们研究的模型版本设定每个时间步到达一个元素，并在每个时间步发布一个输出。我们考虑一种更一般的模型：每个时间步可以到达多个元素／发生多次更新，但仍在每个时间步发布一个输出。这种更一般的模型已在研究图问题隐私保护持续观测的论文中得到探讨 [56, 29, 45]。

持续计数与二叉树机制。 持续计数（或比特求和）是许多数据分析任务中使用的基本原语。定义持续观测模型的论文（[13] 和 [23]）还提出了二叉树机制，这是一种用于持续计数的机制，其误差随时间范围 T呈多对数级缩放。

持续计数机制是持续观测领域中为许多任务构建机制的主要基础模块，例如，直方图的持续观测 [11, 44, 36], 差分隐私机器学习的最先进机制 [46, 20, 16, 15, 58, 53], 数据流上的各种频率统计量 [9, 31, 43, 27, 18], 各种图统计量的持续观测 [56, 29, 45, 55,26], 以及聚类问题的持续观测 [28, 57]。因此，接下来将讨论的近期研究试图提出能够改进二叉树机制常数的持续计数机制（尽管它们实现了相同的渐近误差）。其中大多数机制都是分解机制，通过利用巧妙的分解来为私有持续计数实现更优的误差界。需要注意的是，所有这些工作研究的邻接概念都是：相邻数据流在某一个时间步仅有一个比特不同。

持续计数的因子分解机制。受基数估计应用的启发，我们确定了新的邻接关系，并在这些关系下研究私有持续计数。我们的工作实质上是研究在这些新的邻接概念下，将因子分解机制 [50] 应用于全 1 下三角矩阵 $A \in \mathbb{R}^{T\times T}$。关于在常规邻接概念 [20, 2, 30, 40, 21, 3, 38, 34, 5, 42]下用于私有持续计数的最优矩阵分解，已有大量文献进行研究。下面我们将更详细地介绍这些研究。

基于树的分解 [23, 13, 2, 4] 是首类实现 $O(\log\ T)$ 误差的分解方法，并且能够高效地实现这一点，所需空间为 $O(\log\ T)$，时间为 $O(T)$。其关键结构是稀疏性，这不仅使计算更高效，还能降低 ℓ_2 和 ℓ_1 敏感度，使其在.....方面也达到最先进水平

针对 ε-差分隐私持续计数这一独立问题。人们也提出了对基于树的机制进行“去噪”的技术：通过使左因子更加稠密，以牺牲效率 [41, 33, 46]为代价，换取常数倍的改进。

随后有人提出了更一般的稠密分解 [20, 30, 40, 38, 35], ，在牺牲一定效率的代价下，降低了误差的首要常数。我们尤其强调平方根分解 $L = R = \sqrt{A}$，其中 $\sqrt{A}$ 是托普利兹下三角矩阵，其元素 $\sqrt{A}[i,j] = \binom{2(i-j)}{i-j}/4^{i-j}$ 满足 $i \geq j$。该方法达到$\ln(T)\ _{\pi\sqrt{2\rho}} \pm O(1)$的误差；对于我们的两种误差概念而言，除加性常数外，这是最优的 [30, 40, 51]。后续工作展示了如何利用有理函数逼近 [21]，通过 $O(\log^2 T)$空间和 $O(T\log^2 T)$时间，借助下三角托普利兹分解将该误差逼近至加性常数范围内。

“非标准”邻接关系下的分解机制。最后需要指出的是，我们并非首个研究在持续观测下，为发布前缀而针对“非标准”邻接关系设计分解机制的工作。为了允许用户在机器学习流水线中贡献多个数据点 [46], ，研究者引入了 b-最小间隔 k-重复参与 的概念 [16, 15]。在这种关系下，任何训练样本最多可以出现 k 次，且每个样本之间至少间隔 b 个时间步。对于分解机制设计问题，这对应于一个敏感度向量集 $\mathcal{S}_{(k,b)} \subseteq [-1,1]^T$，其中向量是 k-稀疏的，且非零元素之间的间距至少为 b。由于我们的 k-稀疏交替敏感度向量集（$S_{1,k}$）是 $\mathcal{S}_{(k,1)}$的真子集，因此，在 $\mathcal{S}_{(k,1)}$-参与关系下，分解机制的任何误差上界都可以推广到我们的设置中。关键在于， $\mathcal{S}_{(k,1)}$ 上不存在类似的结构来确保有界前缀和，因此这些上界在渐近意义上劣于我们给出的结果。具体而言，基于带状（逆）平方根分解 [47, 49] 的分解，对于该问题而言在渐近意义下是最优的，其上界为 $O(\sqrt{kT})$ [49, 定理 2]。相比之下，在我们的设置中，误差仅对 T 呈对数依赖。

与直接相关工作的比较。最直接相关的工作是 [43]的研究，他们考察了全动态流模型中的不同元素计数，并引入了最大翻转度参数。他们的设定与我们的略有不同（按照我们的定义，他们所说的相邻流之间的距离至多为 2），但他们的算法可以扩展到我们的模型中，我们也将由此得到的上界与我们的结果进行了比较。如图1所示，在最大翻转度参数的整个取值范围内，我们的误差都优于他们的结果。Henzinger 等人 [37]提出了基于总翻转度的另一种参数化方式；虽然我们关注的是最大翻转度——在现实流中，它通常较小——但我们的改进同样适用于这种设定。

对于三角形计数和度数计数，Raskhodnikova 和 Steiner [55] 在项目级差分隐私下，为期望 ℓ_∞ 误差给出了关于 T的强多项式下界。他们的模型假设每个时间步只有一次更新，因此，虽然他们的下界仍适用于我们的设定，但基于周期性重新计算的上界并不适用。据我们所知，这些问题此前最好的上界来自于在每个时间步朴素地运行一次批处理差分隐私算法，并应用组合定理。我们引入了在现实流中显著小于时间范围的新参数，并据此获得了大幅改进的误差保证。尽管我们的准确度保证以 MaxSE 和 MeanSE 表示，但通过标准的高斯尾界（例如参见 [21]第 2.4 节），它们也能改进 ℓ_∞ 误差，从而为现实流规避 [55] 中的下界提供了一种方法。

我们还注意到，另有一些关于发布三角形计数和度数计数的不可比工作 [56, 29, 45]）。这些工作聚焦于仅插入设置，其要求低于我们研究的完全动态情形；不过，其中一些工作（例如 [45]）探讨了更具挑战性的节点隐私模型，而我们关注的是边隐私。此外， [45] 给出了一种适用于所有仅插入图序列的三角形计数机制，而我们针对全动态流中三角形计数的机制仅对一类受限的图序列满足隐私要求。

最后，我们强调，我们的关注点与上述所有工作都不同：我们的目标是获得强渐近保证并优化常数，而以往的工作只关注渐近行为。

5 预备知识

在全文中，我们令 $[i,j]=\{i,i+1,\ldots,j-1,j\}$ 并使用简写 $[n]=[0,n-1]$。我们对向量采用从零开始的索引，也就是说，对于 $\vec{\mathbf{x}}\in\mathbb{R}^n$, $\vec{\mathbf{x}}[i]$ 是 $i\in[n]$对应的第 i^{th} 个元素。我们还会方便地使用简写 $\vec{\mathbf{x}}[i:j]$，其中 $0\le i\le j\le n-1$ 表示 $\mathbb{R}^{i-j+1}$ 中的向量，其元素为 $(\vec{\mathbf{x}}[i:j])[k]=\vec{\mathbf{x}}[i+k]$。此外，我们令 $\vec{\mathbf{x}}[:k]=\vec{\mathbf{x}}[0:k]$ 以及 $\vec{\mathbf{x}}[k:]=\vec{\mathbf{x}}[k:n-1]$。类似地，对于矩阵 $B\in\mathbb{R}^{m\times n}$的索引，我们用 $B[i,j]$ 表示 $(i,j)\in[m]\times[n]$。我们用 $\log_b(x)$ 表示以 b为底的 x 的对数，用 $\ln(x)=\log_e(x)$ 表示自然对数，并用 $\log(x)=\log_2(x)$。

我们的主要结果针对零集中差分隐私，但也讨论纯差分隐私。我们指出，所有差分隐私概念都量化了随机化算法 $\mathcal{A}$ 在敏感相邻输入上运行时，其输出的不可区分性： $\mathcal{X}^n\to\mathcal{Y}$。本文中，我们用记号 $x\sim x'$ 表示这两个输入，并默认根据上下文可知它们属于 $\mathcal{X}^n$。

下面给出所有相关的隐私定义。

定义 5.1 ((ε,δ)-差分隐私 [22, 24])。 如果对于所有 $S\subseteq\mathrm{Range}(\mathcal{M})$ 以及所有相邻输入 $x,x'\in\mathcal{X}^n$，随机化算法 $\mathcal{M}:\mathcal{X}^n\to\mathcal{Y}$均满足以下条件，则称其为$(\varepsilon,\delta)$-差分隐私：

$$\Pr[\mathcal{M}(x)\in S]\le\exp(\varepsilon)\cdot\Pr[\mathcal{M}(x')\in S]+\delta\,.$$

We 将$(\varepsilon,0)$-差分隐私称为 ε-差分隐私，或纯 差分隐私 .

定义 5.2 (ρ-零集中差分隐私 [10])。 如果对于所有相邻输入 $x,x'\in\mathcal{X}^n$，随机化算法 $\mathcal{M}:\mathcal{X}^n\to\mathcal{Y}$均满足以下条件，则称其满足 ρ-零集中差分隐私：

$$\forall\alpha>1:D_\alpha(\mathcal{M}(x)\|\mathcal{M}(x'))\le\rho\alpha\,,$$

其中， $D_\alpha(\mathcal{M}(x)\|\mathcal{M}(x'))$ 是 α-*Rényi*散度，衡量 $\mathcal{M}(x)$ 与 $\mathcal{M}(x')$之间的差异。

我们注意到，对于任意 $\delta>0$ [10]， ρ-零集中差分隐私还意味着$(\rho+2\sqrt{\rho\ln(1/\delta)},\delta)$-差分隐私。

定义 5.3 (ℓ_p 敏感度)。 设 $f:\ \mathcal{X}^n\to\mathbb{R}^n$ 是定义在敏感输入上的函数，其邻接关系为“$\sim$”。则我们将（全局） ℓ_p 敏感度定义为 f

$$\mathsf{sens}_p(f)=\max_{x\sim x'}\|f(x)-f(x')\|_p\,,$$

有时我们使用$\mathsf{sens}(f)=\mathsf{sens}_2(f)$作为简写。

定义 5.4 (高斯机制 [10])。 设 $f:\ \mathcal{X}^n\to\mathbb{R}^n$ 是定义在敏感输入上的函数，并定义（随机化）算法 $\mathcal{M}(x)=f(x)+\vec{\mathbf{z}}$，其中 $\vec{\mathbf{z}}\sim\mathcal{N}(0,0.5\cdot\mathsf{sens}_2(f)^2/\rho)^n$ 是零均值、 n维的独立同分布高斯噪声。则 $\mathcal{M}$满足 ρ-零集中差分隐私。

定义 5.5 (拉普拉斯机制 [22])。 设 $f:\mathcal{X}^n\to\mathbb{R}^n$ 是定义在敏感输入上的函数，并定义随机化算法 $\mathcal{M}(x)=f(x)+\vec{\mathbf{z}}$，其中 $\vec{\mathbf{z}}\sim\mathrm{Lap}(0,\mathsf{sens}_1(f)/\varepsilon)^n$ 是零均值、 n维的独立同分布拉普拉斯噪声。则 $\mathcal{M}$ 满足 ε-差分隐私。

5.1 因子分解机制

因子分解机制（有时也称为矩阵机制 [50]）是我们贡献的核心。给定查询矩阵 $A\in\mathbb{R}^{T\times T}$ 和一组满足邻接关系“$\sim$”的敏感输入 $\mathcal{D}\subseteq\mathbb{R}^T$，它提供了一种方法，可对任意 $\vec{\mathbf{x}}\in\mathcal{D}$私下发布 $A\vec{\mathbf{x}}$。对于将 A 分解为矩阵 $L\in\mathbb{R}^{T\times m},R\in\mathbb{R}^{m\times T}$的任意一种分解，使得 $A=LR$，它会导出一个机制 $\mathcal{M}_{LR}(\vec{\mathbf{x}})$。给定 R的敏感度 ℓ_2

$$\mathsf{sens}_2(R)=\max_{\substack{(\vec{\mathbf{x}},\vec{\mathbf{y}})\in\mathcal{D}^2\\ \vec{\mathbf{x}}\sim\vec{\mathbf{y}}}}\|R\vec{\mathbf{x}}-R\vec{\mathbf{y}}\|_2\,,\ \max_{\substack{(\vec{\mathbf{x}},\vec{\mathbf{y}})\in\mathcal{D}^2\\ \vec{\mathbf{x}}\sim\vec{\mathbf{y}}}}\|R(\vec{\mathbf{x}}-\vec{\mathbf{y}})\|_2\,,$$

该机制的输出

$$\mathcal{M}_{LR}(\vec{\mathbf{x}}) = A\vec{\mathbf{x}} + L\vec{\mathbf{z}} \qquad \text{where} \qquad \vec{\mathbf{z}} \sim \mathcal{N}(0, 0.5 \cdot \mathsf{sens}_2(R)^2/\rho)^m$$

以提供 ρ-零集中差分隐私保证。该机制背后的直觉是，使用高斯机制将 $R\vec{\mathbf{x}}$ 发布为 $R\vec{\mathbf{x}} + \vec{\mathbf{z}}$，然后从左侧通过 L 对其进行后处理，得到 $\mathcal{M}_{LR}(\vec{\mathbf{x}}) = L(R\vec{\mathbf{x}} + \vec{\mathbf{z}})$。

由于我们的主要贡献是研究非标准邻接关系下的分解机制，因此我们在符号表示中更明确地体现这种依赖关系。我们定义敏感度向量集 $\mathcal{S}$，并将相关的 ℓ_p R 的敏感度表示为 $\mathsf{sens}_p(R,\ \mathcal{S})$:

$$\mathcal{S} = \{\vec{\mathbf{x}} - \vec{\mathbf{y}} : \vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathcal{D} \text{ and } \vec{\mathbf{x}} \sim \vec{\mathbf{y}}\}, \quad \mathsf{sens}_p(R, \mathcal{S}) = \max_{\vec{\Delta} \in \mathcal{S}} \|R\vec{\Delta}\|_p = \mathsf{sens}_p(R).$$

在本文 中，我们将采用两种常用于因子分解机制的误差度量:

$$\mathsf{MaxSE}(L, R, \mathcal{S}) = \sqrt{\max_{t \in [T]} \mathbb{E}\left[(A\vec{\mathbf{x}} - \mathcal{M}_{LR}(\vec{\mathbf{x}}))_t^2\right]} = \frac{\|L\|_{2\to\infty} \cdot \mathsf{sens}_2(R, \mathcal{S})}{\sqrt{2\rho}}$$

$$\mathsf{MeanSE}(L, R, \mathcal{S}) = \sqrt{\mathbb{E}\left[\frac{1}{T}\sum_{t \in [T]} (A\vec{\mathbf{x}} - \mathcal{M}_{LR}(\vec{\mathbf{x}}))_t^2\right]} = \frac{\|L\|_F \cdot \mathsf{sens}_2(R, \mathcal{S})}{\sqrt{2T\rho}}.$$

其中，$\|\cdot\|_F$ 是*Frobenius* 范数，$\|\cdot\|_{2\to\infty}$ 是矩阵所有行中最大的 ℓ_2 范数。稍后我们还将使用 $\|\cdot\|_{1\to 2}$ 范数，它等于矩阵所有列中最大的 ℓ_2 范数。这两种误差度量分别对应于所有输出的平方误差期望最大值的平方根 和平方误差期望均值的平方根 。由于因子分解机制在期望意义下是精确的，这些误差与输入无关，可以视为所有输出的最大方差和平均方差的平方根。

我们还要指出，同一框架也可用于推导满足 ε-差分隐私的机制。唯一的变化是，现在使用与 ℓ_1 敏感度成比例的拉普拉斯机制来发布 $R\vec{\mathbf{x}}$ 。相应的误差度量可通过将 $\mathsf{sens}_2(R, \mathcal{S})/\sqrt{2}\rho$ 替换为 $\sqrt{2}\mathsf{sens}_1(R, \mathcal{S})/\varepsilon$来推导。除非另有明确说明，本文中的所有结果和论述均针对 ρ-零集中差分隐私。

持续观测 当查询矩阵 A 为下三角矩阵时，可以研究在持续观测下私密发布 $A\vec{\mathbf{x}}$ 的问题，这也称为（加权）持续计数 [23, 13]。可以采用之前的同一邻接关系，但此时 $\vec{\mathbf{x}}$ 以每次一个条目的方式流式输入，并使用更新后的目标函数:

$$\text{On receiving } \vec{\mathbf{x}}[t] \text{ at time } t, \text{ immediately output } (A\vec{\mathbf{x}})[t].$$

该问题对应的因子分解机制可以自然地扩展:

$$\text{On receiving } \vec{\mathbf{x}}[t], \text{ immediately output } \mathcal{M}_{LR}(\vec{\mathbf{x}})[t] = (A\vec{\mathbf{x}})[t] + (L\vec{\mathbf{z}})[t].$$

在持续观测下，即使对于更强的自适应 差分隐私 [20]，它仍满足 ρ-零集中差分隐私。在这种情况下，输入$\vec{\mathbf{x}}[t]$ 可以依赖于过去的私密输出 $\mathcal{M}(\vec{\mathbf{x}})[t']$，其中 $t' < t$。特别地，对于这一更强的定义， L 和 R 均不受结构约束。最终，我们的目标是将基数估计问题归约到这一设定。

计算模型 因子分解机制的时间和空间开销也值得关注。由于$A\vec{\mathbf{x}}$ 易于计算，资源占用取决于输出 $L\vec{\mathbf{z}}$的难度。与以往关于持续计数的研究 [2, 21, 3], 一致，我们将注意力限制在输出 $L\vec{\mathbf{z}}$的线性 算法上。空间用需要存储的实数值数量来衡量，时间则用加法和乘法的次数来衡量。

5.2 感兴趣的问题

在本节中，我们将介绍本文研究的基数估计问题，以及相关的邻接关系。我们指出，对于函数 f，我们考虑的两种误差度量是平方误差期望最大值的平方根

$$\mathsf{MaxSE}(\mathcal{M}, T) = \sqrt{\max_{t\in[T]} \mathbb{E}\left[(f(\vec{\mathbf{x}}[0:t]) - \mathcal{M}(\vec{\mathbf{x}}[0:t]))^2\right]},$$

以及平方误差期望均值的平方根

$$\mathsf{MeanSE}(\mathcal{M}, T) = \sqrt{\mathbb{E}\left[\frac{1}{T}\sum_{t\in[T]} (f(\vec{\mathbf{x}}[0:t]) - \mathcal{M}(\vec{\mathbf{x}}[0:t]))^2\right]}.$$

对于具有向量输出的 DegreeCount，我们将考虑这些误差指标的一种自然变体，详见下文关于图函数的章节。

持续观测下的不同元素计数：我们研究转入转出（完全动态）数据流中的不同元素问题。在这个问题中，元素来自全集 $\mathcal{U}$。该问题的数据流 $\vec{\mathbf{x}}$ 对应一系列更新（每次更新是一组元素的插入或删除操作，也可以是不执行任何操作），也就是说，对于所有 $t \in [T]$, $\vec{\mathbf{x}}[t] \in \{\{\{+,-\} \times \mathcal{U}\} \cup \perp\}^*$。我们将用 $\vec{\mathbf{x}}[t][i]$ 表示时间步 t的 i^{th} 更新。通俗地说，我们用 $+u$ 表示插入一个元素 u，用 $-u$表示删除一个元素 u。现在，我们可以正式定义不同元素计数问题了。不同元素计数是一个函数，它以数据流 $\vec{\mathbf{x}}$ （一系列 T 更新）作为输入，并输出一个由 T 非负整数组成的数据流。对于 $t \in [T]$,

$$\mathsf{CountDistinct}(\vec{\mathbf{x}})[t] = \sum_{u\in\mathcal{U}} \mathbb{1}(u \text{ has been inserted more often than deleted in } \vec{\mathbf{x}}[0:t])\,.$$

请注意，在持续观测下研究这一问题的先前工作 [43, 18] 考虑的是每个时间步只有一次更新的情况。我们考虑更一般的模型，即每个时间步可以有任意多次更新，因为这与实际系统更加直接地对应。

在我们的工作中，我们将考虑以下邻接关系。隐私是相对于这一邻接关系定义的。

定义 5.6（不同元素计数的元素邻接关系）。对于不同元素计数，如果至多存在一个元素 $u \in \mathcal{U}$，使得将其中一个数据流中与 u （即 $-u$ 或 $+u$）对应的所有更新替换为 $\perp$ 后得到另一个数据流，则称数据流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$ 是邻接的。

这一定义（在我们看来）比先前工作 [43]中使用的定义更简单。在那个定义中，邻接关系对应于将与 u 对应的更新子集 替换为 $\perp$。从语义上讲，我们认为，我们的邻接概念与先前工作中研究的邻接关系一样，能够很好地捕捉泄露有关元素 u 的信息所带来的风险。此外，我们注意到，在他们的邻接概念下相邻的数据流，在我们的邻接概念下对应于 2-邻接，因此，我们的结果（增加这一额外的 2 倍因子后）同样适用于他们的邻接关系。

持续观测下发布图函数：我们研究边差分隐私下全动态图流中的度直方图问题和三角形计数问题。

对于这个问题，流 $\vec{\mathbf{x}}$ 可以看作由一系列图 $G_1, \ldots, G_T$组成，它们都具有相同的 n节点。流中的每个元素都对应一组边更新（即在图中两个节点 a 和 b 之间插入或删除一条边 $e = (a, b)$，或者表示空操作的 $\perp$ 操作），这些更新描述了如何由时间步 $t-1$的图生成时间步 t 的图。我们将用 $\vec{\mathbf{x}}[t][i]$

表示时间步 t的 i^{th} 更新。我们假设 G_{-1} 为空图。非正式地，我们用 $+(a,b)$ 表示插入一条边 (a,b)，用 $-(a,b)$ 表示删除一条边 (a,b)。为方便起见，我们还会滥用记号，根据需要将流交替称为边更新序列或图序列。

现在，我们可以正式定义度估计问题了。度数计数是一个函数，它以数据流 $\vec{\mathbf{x}}$（由 T图 $G_0,\ldots,G_{T-1}$组成的序列）为输入，并输出一个由 T非负整数向量组成的数据流，每个向量的大小为 n。对于 $t\in[T]$,

$$\mathsf{DegreeCount}(\vec{\mathbf{x}})[t][i]=\sum_{a\in[n]}\mathbb{1}[\text{edge }(i,a)\in G_t]\,.$$

接下来，我们将定义三角形计数问题。在图中，三角形对应于三条边 $\{(a,b),(b,c),(a,c)\}$ 的存在，这三条边连接图中的三个节点 a,b,c。三角形计数是一个函数，它以数据流 $\vec{\mathbf{x}}$（由 T图 $G_0,\ldots,G_{T-1}$组成的序列）为输入，并输出一个由 T非负整数组成的数据流。对于 $t\in[T]$,

$$\mathsf{TriangleCount}(\vec{\mathbf{x}})[t]=\sum_{a,b,c\in[n]}\mathbb{1}[a,b,c\text{ form a triangle in }G_t]\,.$$

接下来，我们将定义所关注的邻接关系，隐私的定义将以此为依据。

定义 5.7 (图问题的边邻接)。 我们称针对三角形计数和度数计数 的两个数据流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$ 为边邻接，当且仅当存在至多一条边(a,b)，使得将其中一个数据流中所有对应于(a,b)（即 $-(a,b)$ 或 $+(a,b)$）的更新替换为 $\perp$ 后，得到另一个数据流。

请注意，先前研究持续观测下这一问题的工作 [55] 考虑的是每个时间步仅更新一条边的情况。我们考虑每个时间步可更新任意数量边的情况。

请注意，我们关注的是数据流中所有图的某个参数（例如度）均受值 k 限制的情况。我们可以将 k-受限邻居定义为满足上述边邻接要求的图序列，并额外要求两个图序列中的所有图的最大参数均受 k 限制。基于这样的邻接关系，可以定义 k-受限边隐私。

最后请注意，由于度数计数是向量值的，因此我们的误差概念对其定义不明确。我们将对误差度量作如下扩展：将该问题的均方根误差定义为各坐标上输出的均方根误差的最大值，均方根最大误差也作类似定义。

6 完全动态数据流的敏感度向量集

本节将正式介绍我们关注的敏感度向量集，并说明它们与基数估计问题之间的关系。具体而言，我们将解释，针对所定义的敏感度向量集进行持续计数时的误差界，如何对应于不同元素计数、度数计数和三角形计数的误差界。

我们考虑的问题有一个共同点：其算法都涉及在差分流上运行持续计数机制。形式化地，对于计数函数 PropCount，定义差分流$d\mathsf{PropCount}(\vec{\mathbf{x}})[t]=\mathsf{PropCount}(\vec{\mathbf{x}})[t]-\mathsf{PropCount}(\vec{\mathbf{x}})[t-1]$。

首先，在第 6.1 节中，我们将刻画不同元素计数、度数计数和三角形计数所对应的差分流的敏感度向量性质。随后，在第 6.2 节中，我们将探讨如何通过限制这些数据流中元素的“贡献”，为敏感度向量集增加额外结构，以供我们利用。

6.1 基数问题中敏感度向量的性质 s

考虑定义在全集 $\mathcal{U}$ 上的计数问题 PropCount 所对应的相邻流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$。与这些相邻流相关联的敏感度向量，是相应差分流的减法：

$$\vec{\Delta}_{\mathsf{PropCount}}(\vec{\mathbf{x}},\vec{\mathbf{y}}) = d_{\mathsf{PropCount}}(\vec{\mathbf{x}}) - d_{\mathsf{PropCount}}(\vec{\mathbf{y}})\,.$$

敏感度向量 集是相邻流对上所有此类向量的集合 rs:

$$\vec{\Delta}_{\mathsf{PropCount}} = \{\vec{\Delta}_{\mathsf{PropCount}}(\vec{\mathbf{x}},\vec{\mathbf{y}}) \mid \vec{\mathbf{x}},\vec{\mathbf{y}} \in \mathcal{U}^T \text{ s.t. } \vec{\mathbf{x}} \;\&\; \vec{\mathbf{y}} \text{ are neighbours}\}\,.$$

我们首先关注取值于 $\{-1,0,1\}^T$ 的敏感度向量，并将非零元素符号交替的向量定义为交替的向量。例如，向量 (1，0，0，-1，0，1) 是交替向量。在定理 6.2 和 6.3 中，我们证明了不同元素计数和度数计数的敏感度向量集是交替的。

我们通过定义集合 $\mathcal{S}_D$ 对这一概念进行推广，该集合包含所有具有 D 有界区间和的向量，

$$\mathcal{S}_D = \left\{ \vec{s} \in \mathbb{Z}^T \;\middle|\; \max_{i \le j \in [T]} \left| \sum_{t=i}^{j} \vec{s}[t] \right| \le D \right\}.$$

对于本文所考虑的所有基数估计问题（不同元素计数、度数计数和三角形计数），我们证明其敏感度向量集包含于 $\mathcal{S}_D$，其中 D 的取值适当。事实上，当 $D=1$ 时，集合 $\mathcal{S}_1$ 恰好是交替向量的集合，正如下面的定理 6.1 所证明的那样。

命题 6.1. $\vec{\Delta} \in \{-1,0,1\}^T$ 是交替的 $\iff$ $\vec{\Delta} \in \mathcal{S}_1$

证明。注意，如果敏感度向量是交替的，那么任意区间和至多为 1，因为区间内相邻的 1 和 -1 会相互抵消，最多只留下一个未抵消的多余值。另一方面，如果敏感度向量不是交替的，那么就存在两个符号相同的非零值，它们之间有零值；对应于这两个非零值索引的区间和的绝对值将大于 1。因此，该敏感度向量不属于 $\mathcal{S}_1$。 □

敏感度向量 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in \mathcal{S}_1$ 适用于所有相邻计数流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$。 考虑 CountDistinct 的相邻流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$。我们将针对项 u 的受限流 $\vec{\mathbf{x}}|_u$ 定义如下。

$$\vec{\mathbf{x}}|_u[t][i] = \begin{cases} +u & \text{if } \vec{\mathbf{x}}[t][i] = +u\,, \\ -u & \text{if } \vec{\mathbf{x}}[t][i] = -u\,, \\ \perp & \text{o/w}\,. \end{cases} \tag{1}$$

定理 6.2. 对于所有相邻流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}} \in (\mathcal{U} \cup \{\perp\})^T$，敏感度流 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 都是交替的。（即 $\vec{\Delta}_{\mathsf{CountDistinct}} \in \mathcal{S}_1$。）

证明。考虑相邻流 $\vec{\mathbf{x}}$、$\vec{\mathbf{y}}$，它们在某个元素 u 的数据上有所不同。不失一般性，我们可以假设 u 出现在 $\vec{\mathbf{x}}$ 中，而不出现在 $\vec{\mathbf{y}}$ 中。首先注意，$\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 仅取决于 $-u,u$ 的出现次数，因为这些是 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$ 唯一不同的输入。也就是说，

$$\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}},\vec{\mathbf{y}}) = d_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u) - d_{\mathsf{CountDistinct}}(\vec{\mathbf{y}}|_u)\,.$$

Mo此外，$\vec{\mathbf{y}}|_u = \{\perp\}^T$ 这意味着 $d_{\mathsf{CountDistinct}}(\vec{\mathbf{y}}|_u) = \vec{0}$。由此得到以下 g:

$$\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}},\vec{\mathbf{y}}) = d_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)\,.$$

现在考虑 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})[t] = d_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t]$，并证明它是交替的。为此，我们构造表 3，其中展示了如何根据 $\mathsf{CountDistinct}(\vec{\mathbf{x}}|_u)[t]$ 和 $\mathsf{CountDistinct}(\vec{\mathbf{x}}|_u)[t-1]$ 计算 $d_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t]$。

不同元素计数($\vec{\mathbf{x}}|_u$)[$t-1$] \ 不同元素计数($\vec{\mathbf{x}}|_u$)[t]	**0**	**1**
0	0	1
1	−1	0

表 3：基于不同元素计数($\vec{\mathbf{x}}|_u$)计算$d_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t]$ 。

我们可以从表 3 看出，只有当 CountDistinct($\vec{\mathbf{x}}|_u$)$[t]=1$ 时，$\mathrm{d}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t]=1$ 才会如此。在后续时间步中，这个计数要么保持不变，要么减少到 0。当计数保持不变时，$\mathrm{d}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)$ 将为 0；当计数在时间 t^* 减少到 0 时，d CountDistinct($\vec{\mathbf{x}}|_u$)$[t^*]$ 将为 -1。类似地，只有当 CountDistinct($\vec{\mathbf{x}}|_u$)$[t]=0$ 时，$\mathrm{d}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t]=-1$ 才会如此。在后续时间步中，这个计数要么保持不变，要么增加到 1。当计数保持不变时，$\mathrm{d}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)$ 将为 0；当计数在时间 t^* 增加到 1 时，$\mathrm{d}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}|_u)[t^*]$ 将为 1。 □

刻画敏感度向量度数计数。考虑边邻接流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$ 对应度数计数。我们将边 (u,v)上的受限流 $\vec{\mathbf{x}}|_{(u,v)}$ 定义如下。

$$\vec{\mathbf{x}}_{(u,v)}[t][i]=\begin{cases}\vec{\mathbf{x}}[t][i] & \text{if } \vec{\mathbf{x}}[t][i]=\pm(u,v)\,,\\ \perp & \text{o/w}\,.\end{cases}$$

定理　第 6.3 节。对于所有边邻接图数据流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，存在两个节点 u,v 使得

1. $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][u]\in\mathcal{S}_1$, $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][v]\in\mathcal{S}_1$. *2.* 对所有节点 $c\neq u,v$，有 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][c]=0^T$。

证明。考虑边邻接流 $\vec{\mathbf{x}}$, $\vec{\mathbf{y}}$ ，它们在某条边 (u,v) 的数据上有所不同。不失一般性，我们可以假设 (u,v) 出现在 $\vec{\mathbf{x}}$ 中，而不出现在 $\vec{\mathbf{y}}$中。

首先，注意 $\vec{\Delta}_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 的非零值仅取决于 $-(u,v)$ $+(u,v)$ 的出现情况。这是因为边 u,v 只影响节点 u 和 v 的度，并且它对这些节点度的影响独立于图其余部分的状态。因此

$$\vec{\Delta}_{\mathsf{DegreeCount}}(\vec{\mathbf{x}},\vec{\mathbf{y}})=d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})-d_{\mathsf{DegreeCount}}(\vec{\mathbf{y}}|_{(u,v)})\,.$$

此外 ， $\vec{\mathbf{y}}|_{(u,v)}=\left\{\vec{\perp}\right\}^T$这意味着 $d_{\mathsf{DegreeCount}}(\vec{\mathbf{y}}|_{(u,v)})=\vec{0}$。由此可得如下 g:

$$\vec{\Delta}_{\mathsf{DegreeCount}}(\vec{\mathbf{x}},\vec{\mathbf{y}})=d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})\,. \tag{2}$$

由于边 (u,v) 只会影响与其相接的顶点 u 和 v 的度，因此对于所有 $w\neq u,v$，都有 $d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t][w]=0$。这证明了定理的第二项。此外，由于对边 (u,v) 的更新会以相同的方式影响 u 和 v 的度，因此我们有

$$d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t](u)=d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t](v)\,.$$

我们现在关注更新流 $d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t][u]$ ，并证明它必定是交替的。首先，若 $d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t][u]=1$，我们证明该更新流中下一个非零项必定为 -1。随后，我们可以用类似的论证证明其逆命题。

请注意，只有当 DegreeCount($\vec{\mathbf{x}}|_{(u,v)}$)$[t][u]=1$ 时，$d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t][u]=1$ （见表4）。在未来的时间步中，这个计数只能保持不变或降为 0。当 DegreeCount 保持不变时，$d_{\mathsf{DegreeCount}}$ 将为 0；而当 DegreeCount 降为 0 时，$d_{\mathsf{DegreeCount}}$ 将为 -1。

同理，$d_{\mathsf{DegreeCount}}(\vec{\mathbf{x}}|_{(u,v)})[t][u]=-1$ 仅当 DegreeCount($\vec{\mathbf{x}}|_{(u,v)}$)$[t][u]=0$时成立。在后续时间步中，该计数可以保持不变，也可以增加到 1。当 DegreeCount 保持不变时，$d_{\mathsf{DegreeCount}}$ 将为 0；当计数增加到 1 时，$d_{\mathsf{DegreeCount}}$ 将为 1。这就完成了定理第一部分的证明。 □

度数计数($\vec{\mathbf{x}}|_{(u,v)}$)$[t-1][u]$ \ 度数计数($\vec{\mathbf{x}}|_{(u,v)}$)$[t][u]$	**0**	**1**
0	0	1
1	−1	0

表4：基于度数计数($\vec{\mathbf{x}}|_{(u,v)}$)计算$d_{\mathsf{DegreeCount}(\vec{\mathbf{x}}|_{(u,v)})[t][u]}$。

适用于所有相邻图流的敏感度向量 $\vec{\Delta}_{\mathsf{TriangleCount}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})} \in \mathcal{S}_D$ **$\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$度数上界为** D。考虑用于三角形计数的相邻流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$，其度数上界为 D。我们将边 (a,b) 对应的受限流 $\vec{\mathbf{x}}|_{(a,b)}$ 定义如下。

$$\vec{\mathbf{x}}_{(a,b)}[t][i] = \begin{cases} \{\vec{\mathbf{x}}[t][i] & \text{if } \vec{\mathbf{x}}[t][i] \in \big\{(a,\cdot),(b,\cdot),(\cdot,a),(\cdot,b)\big\}, \\ \perp & \text{o/w}. \end{cases}$$

定理 6.4. 固定相邻图流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，其度数上界为 D。敏感度流为 $\vec{\Delta}_{\mathsf{TriangleCount}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})} \in \mathcal{S}_D$。

证明。考虑相邻图 $\vec{\mathbf{x}}$和 $\vec{\mathbf{y}}$，它们仅在边 (a,b) 上不同。不失一般性，我们可以假设边 (a,b) 出现在 $\vec{\mathbf{x}}$ 中，而不出现在 $\vec{\mathbf{y}}$中。我们使用以下三个命题证明定理 6.4。

命题6.5. $\vec{\Delta}_{\mathsf{TriangleCount}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})} = d_{\text{三角形计数}(\vec{\mathbf{x}})} - d_{\text{三角形计数}(\vec{\mathbf{y}})} = d_{\mathsf{TriangleCount}\left(\vec{\mathbf{x}}|_{(a,b)}\right)} - d_{\text{三角形计数}\left(\vec{\mathbf{y}}|_{(a,b)}\right)} = d_{\text{三角形计数}\left(\vec{\mathbf{x}}|_{(a,b)}\right)}$。

这是由以下事实得出的：$d_{\text{三角形计数}(\vec{\mathbf{x}})} - d_{\text{三角形计数}(\vec{\mathbf{y}})}$ 仅取决于包含边 (a,b) 的三角形，并且这类三角形只存在于 $\vec{\mathbf{x}}$ 中（即，$d_{\mathsf{TriangleCount}(\vec{\mathbf{y}}|_{(a,b)})} = \vec{0}$）。

命题6.6。那么，对所有 $i < j \in [T]$:

$$\sum_{t=i}^{j} d_{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})}[t] = \mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[j] - \mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[i-1]\,.$$

这是由以下事实得出的：对所有 $i < j \in [T]$:

$$\begin{aligned} \sum_{t=i}^{j} d_{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})}[i] &= \sum_{t=i}^{j} -\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[t-1] + \mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[t] \\ &= -\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[i-1] + \cancel{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[i]} - \cancel{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[i]} \ldots \\ &\quad + \cancel{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[j-1]} - \cancel{\mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[j-1]} + \mathsf{TriangleCount}(\vec{\mathbf{x}}|_{(a,b)})[j]\,. \end{aligned}$$

以下命题将完成此证明。

命题6.7。那么，对所有 $i < j \in [T]$:

$$|\mathsf{TriangleCount}[j] - \mathsf{TriangleCount}[i-1]| \le D\,.$$

Th这可直接由以下事实得出：三角形计数$[j]$，三角形计数$[i-1] \in \{0,\ldots,D\}$. □

6.2 有界贡献下的敏感度向量集

本节将阐述限制特定元素或项对敏感度向量集的贡献变化所产生的影响。

Let $\mathcal{S}_{D,\ k} \subseteq \mathcal{S}_D$ 是具有额外保证的向量集：其范数受 k 限制 s:

$$\mathcal{S}_{D,k} = \left\{\vec{s} \in \mathbb{Z}^T \;\middle|\; \vec{s} \in S_D,\ \|\vec{s}\|_1 \le k\right\}.$$

我们将证明，[43] 中提出的用于不同元素计数的有界最大翻转度概念，对应于将敏感度向量集限制为 $\mathcal{S}_{1,k}$ 的一个子集，其中 k 是最大翻转度的上界。对于度直方图和三角形计数，我们将提出能够体现有界贡献的新概念，分别称为最大度贡献和最大三角形贡献，并论证它们具有类似效果：将敏感度向量集限制为$\mathcal{S}_{D,k}$ 的一个子集，其中 k 是相应参数的上界。

不同元素计数的有界贡献变化通过最大翻转度：注意，对于不同元素计数的任意相邻流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，敏感度向量的 ℓ_1 范数可能大到 $T(\vec{\mathbf{x}} = \{+u, -u, +u, -u, \ldots, +u, -u\}, \vec{\mathbf{y}} = \{\perp, \perp, \ldots, \perp\})$。然而，这样的流 $\vec{\mathbf{x}}$ 似乎并不现实。为此，先前的研究 [43] 提出了最大翻转度这一概念，用来刻画实际数据流的特征，并证明对于这类数据流，可以实现更好的误差界。

我们先回顾 [43] 中最大翻转度的定义。对于不同元素计数，流 $\vec{\mathbf{x}}$ 中某个元素$u \in \mathcal{U}$ 的翻转度 $\mathrm{flip}(u)$ 定义为该元素 u 在流 $\vec{\mathbf{x}}$中从出现变为不出现，或从不出现变为出现的时间步数。这对应于该元素对 $\vec{\mathbf{x}}$ 的不同元素计数值的贡献发生变化的时间步数。最大翻转度定义为 $\max_{u\in\mathcal{U}} \mathrm{flip}(u)$。

定理 6.8。 设 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$ 是用于 不同元素计数 的相邻流，且最大翻转度至多为 k。则，$\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in \mathcal{S}_{1,k}$。

证明。 考虑一个流 $\vec{\mathbf{x}}$，其最大翻转度由 k所界定；再考虑一个相邻流 $\vec{\mathbf{y}}$，其最大翻转度同样由 k 所界定。不失一般性，假设 $\vec{\mathbf{x}}$ 具有与项目 $u \in \mathcal{U}$ 对应的更新，而 $\vec{\mathbf{y}}$ 没有。

考虑将 $\vec{\mathbf{x}}|_u$ 限制于项目 u 的受限流（定义提醒见公式（1））。注意，$d_{\text{不同元素计数}}(\vec{\mathbf{x}}|_u)$至多有 k 个非零元素。另一方面，$d_{\mathsf{CountDistinct}}(\vec{\mathbf{y}}|_u)$为 0^T。因此，敏感度向量 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$是 k-稀疏的。

由于 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 中的每个值都属于 $\{1, 0, -1\}$，这进一步意味着 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 的 ℓ_1 范数有上界 k。结合定理 6.2（$\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in \mathcal{S}_1$），可得 $\vec{\Delta}_{\mathsf{CountDistinct}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in \mathcal{S}_{1,k}$ 如所需。 □

度数计数的有界贡献变化：**通过最大度贡献：**本文为图序列中一条边（a，b）的度数贡献给出了一个新定义，该图序列包含 n 个节点。请注意，在没有进一步假设的情况下，限制在对应于节点a 或节点 b 的索引上的敏感度向量，其范数可能高达 T。为说明这一点，考虑由对边（a，b）进行交替插入和删除组成的 $\vec{\mathbf{x}}$，以及一个相邻流 $\vec{\mathbf{y}}$，其中 $\perp^T$。我们定义流的最大度贡献，以避免此类病态情况。

定义 6.9 (最大度贡献**)**。 流 $\vec{\mathbf{x}}$ 中一条边 (a, b) 的度数贡献定义为该边在整个数据流中被更新的次数 $\vec{\mathbf{x}}$。流的最大度贡献是其所有边的度数贡献中的最大值。

接下来我们说明，对流的最大度贡献设置上界，能够进一步刻画敏感度向量的结构。

定理 6.10。 设 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$ 是关于 度数计数 的相邻流，且最大度贡献有上界 k。那么，存在两个节点 a, b，使得：

1. $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][a] \in \mathcal{S}_{1,k}$，$\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][b] \in \mathcal{S}_{1,k}$。
2. 对所有节点 $c \neq a, b$，都有 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][c] = 0^T$。

证明。 不失一般性，设 $\vec{\mathbf{x}}$ 包含与边 (a, b) 对应的更新，而 $\vec{\mathbf{y}}$ 不包含与边 (a, b) 对应的更新。第 2 项直接由定理 6.3 得出。对于第 1 项，对于节点 a 和 b，根据定理 6.3 的第 1 项，我们有 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][a] \in \mathcal{S}_1$ 和 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][b] \in \mathcal{S}_1$。

最后，注意到若 $\vec{\Delta}_{\mathsf{DegreeCount}}[t][a] = d^{\mathsf{DegreeCount}}(\vec{\mathbf{x}})[t][a] - d^{\text{度数计数}}(\vec{\mathbf{y}})[t][a] \neq 0$，则存在一个与节点 a 在 $\vec{\mathbf{x}}$ 中的更新对应，而在 $\vec{\mathbf{y}}$中不存在对应的更新。这意味着存在一个与边 (a, b) 对应的更新。因此，由于最大度贡献的上界为 k，$\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][a]$ 中的非零元素数量的上界为 k。对于节点 b，也可作类似论证。这意味着 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][a] \in \mathcal{S}_{1,k}$ 和 $\vec{\Delta}_{\mathsf{DegreeCount}}[\cdot][b] \in \mathcal{S}_{1,k}$，如所需。

□

有界贡献变化：三角形计数 **通过最大三角形贡献：** 本文针对图序列中一条边（a，b）的三角形贡献给出了一个新定义，该图序列包含 n个节点且最大度数为 D。我们首先说明该定义的动机。首先注意，在没有更多假设的情况下，存在相邻图序列 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，使得 $\|\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})\|_1 = DT$。 [9]

然而，我们预计，对于现实流 $\vec{\mathbf{x}}$，不会观察到如此极端的情况。我们定义最大三角形贡献来刻画这一点。

定义的首次尝试： 类似于翻转度，我们可以尝试将一条边的三角形贡献定义为包含该边的三角形数量发生变化的时间步数（a，b）。这种定义的一个细微之处在于，即使更新不涉及该边 （a，b），三角形数量也可能发生变化。一旦定义了三角形贡献，就可以通过对边取最大值，以类似于最大翻转度的方式定义最大三角形贡献。

考虑用于三角形计数的两个相邻流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$，其中 $\vec{\mathbf{x}}$ 包含涉及边 (a, b) 的更新，而 $\vec{\mathbf{y}}$ 不包含此类更新。注意，根据定理 6.5，我们有 $\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) = d^{\mathsf{TriangleCount}}(\vec{\mathbf{x}}|_{(a,b)})$。现在，由于图的最大度数为 D，注意在包含 (a, b) 的三角形数量发生变化的时间步，最多有 D 个涉及边 (a, b) 的三角形可能受到影响。此外，根据定义，注意 $\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})$ 是 k-稀疏的。因此，如果最大三角形贡献有上界 k，则有 $\|\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}})\|_1 \leq kD$。

我们的定义：在上述论证中，当我们对 ℓ_1 范数取上界时，每当包含边 (a, b) 的三角形数量发生变化时，都必须考虑最坏情况，即有 D 个三角形受到影响。为了避免考虑这种最坏情况，我们提出了一个不同的定义，直接对 ℓ_1 范数取上界。

定义 6.11（最大三角形贡献）。 在任意时间步 t，考虑包含一条边 （a, b）的三角形数量变化，即 $d^{\mathsf{TriangleCount}}(\vec{\mathbf{x}}|_{(a,b)})[t]$。边 （$a, b$）的三角形贡献定义为所有时间步中这些变化的绝对值之和，即 $\sum_{t\in[T]} |d^{\text{三角形计数}}(\vec{\mathbf{x}}|_{(a,b)})[t]|$。最大三角形贡献定义为所有边的三角形贡献中的最大值。

现在，我们可以直接给出敏感度向量的 ℓ_1 范数上界。

定理 6.12。 考虑相邻图流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，其度数上界为 D，且最大三角形贡献的上界为 k。敏感度流为 $\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in \mathcal{S}_{D,\ k}$。

证明。 考虑用于三角形计数的两个相邻流 $\vec{\mathbf{x}}$ 和 $\vec{\mathbf{y}}$，其中 $\vec{\mathbf{x}}$ 包含涉及边 $(a,\ b)$ 的更新，而 $\vec{\mathbf{y}}$ 不包含。根据定理 6.4，我们有 $\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) \in S_D$。

此外，请注意，根据定理 6.5，我们有 $\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}\ \vec{\mathbf{y}}) = d^{\mathsf{TriangleCount}}(\vec{\mathbf{x}}|_{(a,b)})$。这意味着

$$\|\vec{\Delta}_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}, \vec{\mathbf{y}})\|_1 = \|d_{\mathsf{TriangleCount}}(\vec{\mathbf{x}}|_{(a,b)})\|_1.$$

根据最大三角形贡献的定理 6.11，我们有 $\|d^{\mathsf{TriangleCount}}(\vec{\mathbf{x}}|_{(a,b)})\|_1 \leq k$。证明至此完成。

□

[9]例如，考虑 $\vec{\mathbf{x}}$ 表示在第一个时间步添加一条边 (a, b)，随后在偶数时间步添加所有其他边 $(a, \cdot)$,$(b, \cdot)$，并在非第一个时间步的奇数时间步删除所有形式为 $(a, \cdot), (b, \cdot)$ 的边（边 (a, b) 除外）。 $\vec{\mathbf{y}}$ 与 $\vec{\mathbf{x}}$ 相同，唯一不同的是，它在第一个时间步的更新为 $\perp$。

6.3 基数估计的误差界（摘自《持续计数的误差界》）

考虑具有整数值的数据流 $\vec{s}, \vec{t}$。如果这两个数据流满足 $\vec{s} - \vec{t} \in \mathcal{S}$，我们就称它们是 $\mathcal{S}$-邻接的。如果机制 M 满足关于 $\mathcal{S}$-邻接关系的 ρ-零集中差分隐私，我们就称它是在 $\mathcal{S}$-数据流上的 ρ-零集中差分隐私持续计数机制。

我们用于所有关注问题（不同元素计数、度数计数、三角形计数）的主要算法，采用相同思路将问题归约为持续计数。对于原始数据流 $\vec{\mathbf{x}}$，我们将为基数估计问题 PropCount 构造差分流 $d^{\mathsf{PropCount}}(\vec{\mathbf{x}})$，然后在 $\mathcal{S}$-流上使用适当的差分隐私持续计数机制，其中 $\mathcal{S}$ 是与问题 PropCount 相关联的敏感度向量集合。形式化描述见算法 1。

Al算法 1 算法 $\mathcal{A}$：将 PropCount 归约为持续计数机制 M

输入： 基数估计问题 PropCount、PropCount 的输入流 $\vec{\mathbf{x}}$ 、时间范围 T、持续计数机制 M。**输出：** 输出流 $o_1, \ldots, o_T \in \mathbb{R}^T$1: 令 $\vec{\mathbf{x}}[-1] = \perp$。2: **对于** $t \in [T]$： **执行**3: 令 $d^{\mathsf{PropCount}}(\vec{\mathbf{x}})[t] = \mathsf{PropCount}(\vec{\mathbf{x}}[-1 : t]) - \mathsf{PropCount}(\vec{\mathbf{x}}[-1 : t-1])$。4: 将 $d^{\mathsf{PropCount}}(\vec{\mathbf{x}})[t]$ 发送给机制 M，并输出结果 o_t。

我们证明以下定理，该定理将持续计数机制M 的隐私性和准确性与 PropCount 对应算法的隐私性和准确性联系起来。我们关注问题 PropCount 中机制 $\mathcal{M}$ 的平方误差期望均值的平方根和最大平方误差期望均值的平方根，并分别记为 MaxSE($\mathcal{M}$, T) 和 MeanSE($\mathcal{M}$, T)。对于度数计数，每个时间步都会有向量值输出（函数本身也是向量值的），因此我们关注最大平方误差期望均值的平方根和最大平方误差期望最大值的平方根（这两种情况下的最大值均取自向量的各个索引）。在适用时，我们也将 MaxSE($\mathcal{M}$, T) 和 MeanSE($\mathcal{M}$, T) 的记号扩展为表示这一含义。

我们通常会关注某一特定输入集合上的这一误差，并会在陈述中明确说明。

定理 6.13。 设 $\mathcal{S}$ 为与基数估计问题PropCount相关联的敏感度向量集合。令 $\mathcal{A}$ 表示算法 *1*。

1. 对于任意 $\rho > 0$，如果 M 关于 $\mathcal{S}$ 邻接关系满足 ρ-零集中差分隐私，则 $\mathcal{A}$ 满足 ρ-零集中差分隐私 .

2. 设 M 是一种用于持续计数的因子分解机制，使用因子分解矩阵 L、 R ，对应敏感度向量集 $\mathcal{S}$。那么，

$$\mathsf{MaxSE}(\mathcal{A}, T) = \mathsf{MaxSE}(L, R, \mathcal{S}),$$

$$\mathsf{MeanSE}(\mathcal{A}, T) = \mathsf{MeanSE}(L, R, \mathcal{S}).$$

证明。对于定理的第 1 项，考虑问题 PropCount 中相邻的数据流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$ 。那么，根据敏感度向量集的定义，我们有 $d^{\mathsf{PropCount}}(\vec{\mathbf{x}}) - d^{\mathsf{PropCount}}(\vec{\mathbf{y}}) \in \mathcal{S}$。注意，对数据流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$运行算法 1 时，持续计数机制 M 将分别作用于差分流 $d^{\mathsf{PropCount}}(\vec{\mathbf{x}})$ 和 $d^{\mathsf{PropCount}}(\vec{\mathbf{y}})$。因此，由于 M 关于 $\mathcal{S}$ 邻接关系满足 ρ- 零集中差分隐私，我们有算法 1 满足 ρ-零集中差分隐私。

现在我们证明定理中的第 2 项（对于均方根最大误差，同样的论证也适用于均方根误差）。注意，对于所有 $t \in [T]$，都有 $\sum_{i=0}^{t} d^{\mathsf{PropCount}}(\vec{\mathbf{x}})[t] = \mathsf{PropCount}(\vec{\mathbf{x}})[t]$。因此，由于 $\mathcal{A}$ 和持续计数机制 M 的输出相同（在算法 1 中）， $\mathcal{A}$ 的均方根最大误差恰好等于机制 M 的均方根最大误差，按定义即为 MaxSE(L, R, S)。 □

现在，我们将前面几个小节对敏感度向量集的刻画与定理 6.13 相结合，说明这些结果对我们关注的问题有何影响。

误差界：不同元素计数**：**我们注意到，直接应用定理 6.13，并结合定理 6.8（刻画不同元素计数的敏感度向量集），可以为两个翻转度均不超过 k的相邻数据流提供隐私保护，但不一定能为所有数据流提供隐私保护。

不过，我们注意到，借助 [43]的一项技术，可以将这一方法扩展到为所有数据流提供隐私保护。下面我们回顾这项技术。我们对数据流进行预处理，以确保处理后的数据流的翻转度界至多为 k。第一步是维护一个计数器，用于跟踪数据流中所有元素的翻转度。对于 $\vec{\mathbf{x}}$ 导致某个元素的翻转度越过阈值 k的任何更新，我们只需将其替换为 $\perp$。将所得数据流记为 trunc($\vec{\mathbf{x}}$)。注意，trunc($\vec{\mathbf{x}}$)的最大翻转度至多为 k。此外，如果$\vec{\mathbf{x}}$ 本身的最大翻转度至多为 k，则trunc($\vec{\mathbf{x}}$) $= \vec{\mathbf{x}}$。现在，我们可以使用这一预处理步骤，在为所有数据流提供隐私保护的同时，仍然确保翻转度至多为 k的数据流具有相同的准确性。

推论 6.14. 固定 $k > 0$。令 M 为一种用于持续计数的因子分解机制，使用因子分解矩阵 L, R 来处理敏感度向量集 $\mathcal{S}_{1,\ k}$。令 $\mathcal{A}$ 表示算法 *1*。我们有：

- 对于任意 $\rho > 0$, $\mathcal{A}(\mathsf{trunc}(\cdot))$，若机制 M 相对于 $\mathcal{S}_{1,\ k}$-邻接关系满足 ρ-零集中差分隐私，则其满足 ρ-零集中差分隐私。

- 对于最大翻转度至多为 k 的不同元素计数数据流，我们有：

$$\mathsf{MaxSE}(\mathcal{A}(\mathsf{trunc}(\cdot)), T) = \mathsf{MaxSE}(L, R, \mathcal{S}_{1,k}),$$

$$\mathsf{MeanSE}(\mathcal{A}(\mathsf{trunc}(\cdot)), T) = \mathsf{MeanSE}(L, R, \mathcal{S}_{1,k}).$$

我们还要指出，与 [43],中的情况完全相同，一旦我们有了一个对给定翻转度阈值而言准确、但对所有数据流均满足差分隐私的机制，就可以通过谨慎使用稀疏向量技术，将其扩展到事先不知道翻转度界的情形。这会得到能够适应数据流最大翻转度的准确性上界。详情请参阅 [43]的第 3.2 节。

误差界：度数计数**：** 要对度数计数运行算法 1，我们可以运行 n 个并行副本，每个副本对应一个节点的度数。注意，若直接运行，只有在最大度贡献有界的数据流上才能保证隐私。为了将其扩展到所有数据流上的隐私，我们采用类似于不同元素计数的预处理步骤。具体来说，我们跟踪所有边的度数贡献；如果某条边的度数贡献超过上界 k，就将此边后续的所有更新替换为 $\perp$。将得到的数据流称为 trunc($\vec{\mathbf{x}}$)。注意，trunc($\vec{\mathbf{x}}$) 的最大度贡献至多为 k。此外，如果 $\vec{\mathbf{x}}$ 本身的最大度贡献至多为 k，那么 trunc($\vec{\mathbf{x}}$) $= \vec{\mathbf{x}}$。

将定理 6.13 与定理 6.10 结合（并依据上述讨论），即可得到以下结果（其中我们利用了这样一个事实：一条边只能影响其端点的度数；此外，还利用组合来处理两个节点的敏感度向量属于 $S_{1,k}$，而其他节点的敏感度向量均为0^T的情况）。

推论 6.15。 固定 $k > 0$。令 M 为一个使用因子分解矩阵 L, R 针对敏感度向量集 $\mathcal{S}_{1,\ k}$进行持续计数的因子分解机制。令 $\vec{\mathbf{x}}$ 为对应于 度数计数的流，并令 $\mathcal{A}$表示并行运行算法 *1* 共 n 次，每次用于估计图中一个节点的度。我们有：

- 对于任意 $\rho > 0$, $\mathcal{A}(\mathsf{trunc}(\cdot))$，如果机制 M 相对于$\mathcal{S}_{1,\ k}$-邻接关系满足 ρ-零集中差分隐私，则其满足 2ρ-零集中差分隐私。

- 对于最大度贡献至多为 k的度数计数数据流，我们有

$$\mathsf{MaxSE}(\mathcal{A}(\mathsf{trunc}(\cdot)), T) = \mathsf{MaxSE}(L, R, \mathcal{S}_{1,k}),$$

$$\mathsf{MeanSE}(\mathcal{A}(\mathsf{trunc}(\cdot)), T) = \mathsf{MeanSE}(L, R, \mathcal{S}_{1,k}).$$

三角形计数的误差界：直接应用定理 6.13，并结合定理 6.12（刻画三角形计数的敏感度向量集），可得以下结果。

推论 6.16。 固定 $k, D > 0$。设 M 是一个使用因子分解矩阵 L, R 对敏感度向量集 $\mathcal{S}_{D,k}$进行持续计数的因子分解机制。令 $\mathcal{A}$ 表示算法 *1*。若对于三角形计数 的数据流 $\vec{\mathbf{x}}$ $\vec{\mathbf{y}}$，其中两者的每个图的最大三角形贡献均不超过 k，且度数均不超过 D，则称它们是(D, k)-相邻的。我们有：

- 对于任意 $\rho > 0$，如果机制 M 关于$\mathcal{S}_{D,k}$-邻接关系满足 ρ-零集中差分隐私，则 $\mathcal{A}$ 关于(D, k)-邻接关系满足 ρ- 零集中差分隐私。
- 对于最大三角形贡献至多为 k且最大度数至多为 D的三角形计数数据流，我们有：

$$\mathsf{MaxSE}(\mathcal{A}, T) = \mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) ,$$
$$\mathsf{MeanSE}(\mathcal{A}, T) = \mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) .$$

上述推论的一个缺点是，隐私性仅针对 (D, k)-相邻图流成立。与前述情形不同，目前尚不清楚如何扩展该算法，使其对所有图都具有隐私性。关键难点在于：当某条边的三角形贡献达到其上界后，若忽略对该边的更新，可能会对其他边产生连锁影响，因为这也会影响这些边的三角形贡献。这可能导致相邻图在截断操作后变得极为不同。 [45] 的这项工作研究的正是这一关键问题，他们为此设计了稳定投影。然而，他们的技术仅适用于仅插入数据流，无法轻易扩展到全动态设置。我们将把 TriangleCount 算法扩展为对所有图都具有隐私性，留作一个非常值得探究的问题。

我们再次强调，在我们的设定下，现有的 TriangleCount 上界随T呈多项式增长。我们的参数化使我们能够使用分解机制，在差分流上执行持续计数，并实现形式为 $\sqrt{kD}\log\ T$ 的显著更高精度。因此，我们认为，针对这一问题所取得的准确度保证提升（即使是在受限的隐私概念下）是一项有价值的贡献。

7 从$\mathcal{S}_{D、\ k}$ 到 $\mathcal{S}_{1、\ k}$的归约

回顾一下，令 $\mathcal{S}_D$ 为向量集合，其部分和是 D-有界的，

$$\mathcal{S}_D = \left\{ \vec{s} \in \mathbb{Z}^T \;\middle|\; \max_{i \le j \in [T]} \left| \sum_{t=i}^{j} \vec{s}[t] \right| \le D \right\},$$

并令 $\mathcal{S}_{D、\ k} \subseteq \mathcal{S}_D$ 为具有额外保证的向量集合，即其 k-范数是 ℓ_1-有界的，

$$\mathcal{S}_{D,k} = \left\{ \vec{s} \in \mathbb{Z}^T \;\middle|\; \vec{s} \in \mathcal{S}_D,\ \|\vec{s}\|_1 \le k \right\}.$$

我们已经展示了如何将基数估计问题的机制设计（以 D 和 k为参数）归约为持续计数机制的设计，但适用对象是敏感度向量集为 $\mathcal{S}_{D、\ k}$ 的数据流。具体而言，我们关注这一空间中的分解机制 ，其敏感度分析并非易事。直截了当地说，对于合理选择的矩阵 $R \in \mathbb{R}^{m \times T}$:

$$\textit{how do we compute } \mathsf{sens}_2(R, \mathcal{S}_{D,k}) = \max_{\vec{\Delta} \in \mathcal{S}_{D,k}} \|R\vec{\Delta}\|_2 \textit{ ?}$$

我们注意到，对于比特的标准持续计数，如果两个数据流仅在一个位置上不同，则它们互为邻居，相应的敏感度向量集为 $\mathcal{S}_{1,\ 1}$。这里的答案很简单：

$$\mathsf{sens}_2(R, \mathcal{S}_{1,1}) = \|R\|_{1 \to 2} ,$$

算法 2 将 $\vec{\Delta}$ 分解为 $\vec{\Delta}^{(1)}, \ldots, \vec{\Delta}^{(D)}$ 的过程

输入：向量 $\vec{\Delta} \in \mathcal{S}_{D,k} \subseteq \mathbb{R}^T$。输出：向量 $\vec{\Delta}^{(1)}, \ldots, \vec{\Delta}^{(D)} \in \mathbb{R}^T$。1：初始化 $\forall d \in [-D, D]$: $\vec{\Delta}^{(d)} \leftarrow \vec{0}$和 $p \leftarrow 0$。2：对于 $t \leftarrow 0$ 到 $T-1$，执行 3：对于 $i \leftarrow 1$ 到 $|\vec{\Delta}[t]|$，执行 4：如果 $\mathrm{sign}(\vec{\Delta}[t]) > 0$，则 5：设置 $p \leftarrow p+1$6：设置 $\vec{\Delta}^{(p)}[t] = 1$7：否则 8：设置 $\vec{\Delta}^{(p)}[t] = -1$9：设置 $p \leftarrow p-1$10：重新标记集合 $\{\vec{\Delta}^{(d)} : \ d \in [-D, D]\}$，使得 $\vec{\Delta}^{(1)}, \ldots, \vec{\Delta}^{(D)}$ 包含所有非零向量。11：返回 $\vec{\Delta}^{(1)}, \ldots, \vec{\Delta}^{(D)}$。

也就是说， ℓ_2 敏感度等于 R 中任意一列的最大 ℓ_2 范数。

为了回答关于 k 和 D 的其他取值的问题，我们将给出一个归约。在本节中，我们将说明，研究 $\mathcal{S}_{1,\ k}$ 数据流通常就足够了，也就是说，研究来自 $\{0, \pm 1\}^T$、具有交替非零项的 k-稀疏数据流即可。通过研究更简单、结构更规整的 $\mathcal{S}_{1,\ k}$ 数据流，我们可以紧确估计 R 对于 $\mathcal{S}_{D,\ k}$ 数据流的敏感度。我们首先给出这一方向的一个下界。

引理 7.1。设 $1 \le k \le T$ 和 D， $p \ge 1$，其中 $\mathcal{S}_{D,\ k} \subset \mathbb{R}^T$，且 $R \in \mathbb{R}^{m \times T}$ 对于 $m \ge 1$ 成立。那么

$$\mathsf{sens}_p(R, \mathcal{S}_{D,k}) \ge D \cdot \mathsf{sens}_p(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) .$$

证明。 考虑达到最大范数的 $\Delta^* \in \mathcal{S}_{1,\ \lfloor k/D \rfloor}$，也就是说，

$$\mathsf{sens}_p(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) = \max_{\vec{\Delta} \in \mathcal{S}_{1,\lfloor k/D \rfloor}} \|R\vec{\Delta}\|_p = \|R\vec{\Delta}^*\|_p .$$

注意到 $\vec{\Delta}' = D\vec{\Delta}^* \in \mathcal{S}_{D,\ D\lfloor k/D \rfloor} \subseteq \mathcal{S}_{D,\ k}$，我们可以推出

$$\mathsf{sens}_p(R, \mathcal{S}_{D,k}) \ge \mathsf{sens}_p(R, \mathcal{S}_{D,D\lfloor k/D \rfloor}) \ge \|R\vec{\Delta}'\|_p = D\|R\vec{\Delta}^*\|_p = D \cdot \mathsf{sens}_p(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) ,$$

完成证明。 □

接下来我们将展示 另一个方向上也存在类似的上界。下面的引理至关重要 l.

引理 7.2. 设 $T \ge k \ge 1$ 和 $D \ge 1$，其中 $\mathcal{S}_{D,\ k} \subset \mathbb{R}^T$。那么，对于任意 $\vec{\Delta} \in \mathcal{S}_{D,\ k}$,

$$\exists k_1, \ldots, k_d \ge 0 : \quad \sum_{d=1}^{D} k_d = k \qquad such\ that \qquad \vec{\Delta} = \sum_{d=1}^{D} \vec{\Delta}^{(d)} \qquad where \qquad \vec{\Delta}^{(d)} \in \mathcal{S}_{1,k_d} .$$

证明。 可以通过多种方式识别该分解，例如，使用贪心的“栈着色”论证。对于固定的 $\Delta \in \mathcal{S}_{D,\ k}$，我们将通过算法 2 给出的过程进行论证。根据 $\mathcal{S}_{D,\ k}$ 的定义， $\vec{\Delta}$ 上的任意区间和都在范围 $[-D, D]$ 且 前缀至多取 $D+1$ 个不同的值。为说明第二个结论，注意到如果前缀取到的不同值更多，那么必然存在一个区间和，其取值落在 $[-D, D]$之外。双重 for 循环执行完毕后、Line 10 执行前，我们对这些变量作几点说明:

1. $\{\vec{\Delta}^{(d)}$ 中至多有 D 个不同的向量： $d \in [-D, D]\}$，它们均非零。

2. 所有 $\vec{\Delta}^{(d)} \in \{0, \pm 1\}^T$ 均为交替向量。
3. $\sum_{d\in[-D,D]} \vec{\Delta}^{(d)} = \vec{\Delta}$。

第一条说明显然成立，因为不同前缀值最多有 $D+1$ 个，而且第 4–5 行与第 8–9 行中 p 和 $\vec{\Delta}^{(p)}$ 的更新顺序不同。例如，如果前缀的取值范围是 $[a,b]$，那么只有 $\{\vec{\Delta}^{(d)}:\ d \in [a, b+1]\}$。第二条说明也显然成立：如果在时间 t 向 $\vec{\Delta}^{(d)}$ 中插入了一个 $+1$，那么在时间 $t' > t$ 插入的下一个元素必定是一个 -1，因为前缀的值必须减小才能再次达到 d。第三条说明同样显然成立：读取 $\vec{\Delta}[t]$ 时，同一个（带符号的）单位权重会平均分配给 $|\vec{\Delta}[t]|$ 个向量。

在第 10 行重新标记之后，可知算法 2 将返回 D 交替向量 $\vec{\Delta}^{(1)}$, ..., $\vec{\Delta}^{(D)}$，它们来自 $\mathcal{S}_{1,\ k}$。我们可以将陈述细化为：每个 $\vec{\Delta}^{(d)} \in \mathcal{S}_{1,\ k_d}$，其中 $k_d = \|\vec{\Delta}^{(d)}\|_1$。由此我们必然可以断言 $\sum_{d\in[D]} k_d = k$，从而完成证明。 □

引理 7.3。设 $1 \le k \le T$ 和 D、$p \ge 1$，其中 $\mathcal{S}_{D,\ k} \subset \mathbb{R}^T$，且 $R \in \mathbb{R}^{m\times T}$ 对任意 $m \ge 1$ 均成立。那么

$$\mathsf{sens}_p(R, \mathcal{S}_{D,k}) \le \max_{k_1,\dots,k_D \ge 0:\sum_{d=1}^D k_i = k} \sum_{d=1}^{D} \mathsf{sens}_p(R, \mathcal{S}_{1,k_d}) \le D \cdot U(\lceil k/D \rceil)\ .$$

其中，最后一个上界对任意凹且非递减的函数 U：$[0,T] \to \mathbb{R}$ 都成立，只要 敏感度$_p(R, \mathcal{S}_{1,\ k}) \le U(k)$。

证明。我们从界定目标敏感度开始。

$$\mathsf{sens}_p(R, \mathcal{S}_{D,k}) = \max_{\vec{\Delta}\in\mathcal{S}_{D,k}} \|R\vec{\Delta}\|_p = \|R\vec{\Delta}^*\|_p \le \sum_{d=1}^{D} \|R\vec{\Delta}^{(d)}\|_p \le \sum_{d=1}^{D} \mathsf{sens}_p(R, \mathcal{S}_{1,k_d})\,.$$

第一个不等式使用定理 7.2，将 $\vec{\Delta}^*$ 分解为对 $\{\vec{\Delta}^{(d)} \in \mathcal{S}_{1,k_d}$ 的求和：$d \in [D]\}$，然后使用三角不等式。第二个不等式将每一项都以上界敏感度。为了得到引理中的第一个上界，我们对所有满足 $\sum_{d\in[D]} k_d = k$ 的整数 k_1，$\dots k_d \ge 0$ 取上界中的最大值。为完成证明，我们使用 U：

$$\mathsf{sens}_p(R, \mathcal{S}_{D,k}) \le \max_{k_1,\dots,k_D \ge 0:\sum_{d=1}^D k_d = k} \sum_{d=1}^{D} \mathsf{sens}_p(R, \mathcal{S}_{1,k_d}) \le \max_{k_1,\dots,k_D \ge 0:\sum_{d=1}^D k_d = k} \sum_{i=1}^{D} U(k_d)$$

由于 U 具有凹性且非递减，我们得到，近似均分的划分使该表达式最大，即 ，

$$\sum_{d=1}^{D} U(k_d) \le (D-r)U(q) + rU(q+1)$$

其中 $q = \lfloor k/D \rfloor$ 且 $r = k - Dq$。之所以如此，直观上是因为，将较大的参数减一并将较小的参数加一，永远不会使总和的值减小。形式化地，对于整数 $0 \le a < b$

$$U(a+1) + U(b-1) - (U(a) + U(b)) = \underbrace{U(a+1) - U(a)}_{\Delta U(a)} - \underbrace{U(b) - U(b-1)}_{\Delta U(b-1)} \ge 0$$

其中，不等式成立是因为前向差分 $\Delta U(\cdot)$ 递减不增（ U的凹性）。为完成证明，我们有

$$(D-r)U(q) + rU(q+1) \le D \cdot U(q+1)\,.$$

由于 U 是非递减的 。此外， $q+1 = \lceil k/D \rceil$，除非 $r = 0$，但在这种情况下 $q = \lfloor k/D \rfloor = \lceil k/D \rceil$ 且

$$(D-r)U(q) + rU(q+1) = DU(q) = DU(\lceil k/D \rceil)\,.$$

因此，

$$\mathsf{sens}_p(R, \mathcal{S}_{k,D}) \le D \cdot U(\lceil k/D \rceil)\,.$$ □

自然会想到，$\mathsf{sens}_p(R,\ \mathcal{S}_{1,\ k}) = f(k)$ 本身是否是关于 k 的凹函数，以及引理是否因此可以简化。遗憾的是，存在一些矩阵 R 使得情况并非如此：例如，

$$R = \begin{bmatrix}1 & 0 & 1\end{bmatrix} \in \mathbb{R}^{1\times 3}, \qquad \text{where} \qquad f(1)=1\,, \quad f(2)=1\,, \quad f(3)=2^{1/p}\,,$$

这意味着 $\frac{f(1)+f(3)}{2} = \frac{1+2^{1/p}}{2} \geq 1 = f(2)$，从而否定了 f 的凹性。然而，对于我们所考虑的所有分解和右因子 R，都存在这样的凹且紧的上界。

我们通过在一个定理中总结定理 7.1 和定理 7.3 来结束本节。

定理 7.4。设 $1 \leq k \leq T$ 和 $D, p \geq 1$，其中 $\mathcal{S}_{D,k} \subset \mathbb{R}^T$ 且 $R \in \mathbb{R}^{m\times T}$ 对于 $m \geq 1$。则

$$D \cdot \mathsf{sens}_p(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) \leq \mathsf{sens}_p(R, \mathcal{S}_{D,k}) \leq \max_{k_1,\ldots,k_D \geq 0 : \sum_{d=1}^D k_i = k} \sum_{d=1}^{D} \mathsf{sens}_p(R, \mathcal{S}_{1,k_d}) \leq D \cdot U(\lceil k/D \rceil)\ .$$

其中，最后一个上界对任意满足以下条件的凹非递减函数 U：$[0,T] \to \mathbb{R}$ 均成立：$\mathsf{sens}_p(R, \mathcal{S}_{1,k}) \leq U(k)$。

8 托普利兹分解

本节将证明 $\mathcal{S}_{D,k}$ 数据流上持续计数的误差界，其中使用的分解机制满足某些特殊性质。与之前一样，令 A 为 $T \times T$ 全 1 下三角矩阵（表示持续计数的查询矩阵）。我们使用平方根分解 [8, 30, 40, 21, 35]展示持续计数的新误差界（在新的邻接关系下），其中 A 满足 $A = LR$，且 $L = R = \sqrt{A}$（形式化定义见定理 8.5）。我们关注的敏感度向量集（对应于这些邻接关系）正是第 6 节所描述的集合，并且自然地来源于全动态流模型中在持续观测下发布函数（例如不同元素计数和三角形计数）。

我们证明的主要结果是以下误差界，它们是我们所建立的关于 R 的敏感度界（即关于 $\mathsf{sens}_2(R,\ \mathcal{S}_{D,k})$ 的界）的推论。误差度量的定义见第 5 节。

推论 8.1。设 $A \in \mathbb{R}^{T\times T}$ 为任意查询矩阵，并考虑分解 $A = LR$，其中 R 是对角线元素非递增的下三角托普利兹矩阵。固定 D、$k \geq 1$。则

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \leq \sqrt{Dk} \cdot \mathsf{MaxSE}(L, R, \mathcal{S}_{1,1})\,,$$
$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \leq \sqrt{Dk} \cdot \mathsf{MeanSE}(L, R, \mathcal{S}_{1,1}),\,.$$

特别地，若 A 为全 1 下三角矩阵且 $L = R = \sqrt{A}$，则

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \leq \left(\frac{\ln(T)}{\pi} + 1.067\right)\sqrt{\frac{Dk}{2\rho}}\,,$$
$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \leq \left(\frac{\ln(T)}{\pi} + 0.908 + o(1)\right)\sqrt{\frac{Dk}{2\rho}}\,.$$

本质上，我们的敏感度分析使我们能够直接套用现有的、用于标准持续计数的最优主常数因子分解，适用于 $\mathcal{S}_{1,1}$-流，代价是乘上因子 $\sqrt{Dk}$。我们还证明，特别是在平方根分解的情形下，对于较大范围的 D and k，我们的分析实际上是紧的。

推论 8.2. 设 $A \in \mathbb{R}^{T\times T}$ 为全 *1* 下三角矩阵，并考虑分解$L = R = \sqrt{A}$。取定 $D, k \geq 1$，其中 $k/D \leq O(T^{1/3})$ 且 T充分大。则

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \geq \left(\frac{1}{\pi} - o(1)\right)\sqrt{\frac{D^2\lfloor k/D\rfloor \ln(T)\ln\left(\frac{T}{\lfloor k/D\rfloor}\right)}{2\rho}},$$

$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \geq \left(\frac{1}{\pi} - o(1)\right)\sqrt{\frac{D^2\lfloor k/D\rfloor \ln(T)\ln\left(\frac{T}{\lfloor k/D\rfloor}\right)}{2\rho}}.$$

本节其余部分的组织如下。首先，在第 8.1 节中，我们证明矩阵R^TR 的性质，这些性质对我们的分析至关重要。接下来，在第 8.2 节中，我们将证明定理 8.1，并特别论证平方根分解具有所需的性质。对于 $\mathsf{sens}_2(R, \mathcal{S}_{D,k})$ 的关键上界，我们给出两种相互独立的证明：一种是组合证明（基于匹配论证），另一种基于对二次型的直接运算。最后，在第 8.3 节中，我们证明定理 8.2，说明对于 D 和 k的较大取值范围，我们关于敏感度 $\mathsf{sens}_2(R, \mathcal{S}_{D,k})$的上界是紧的。我们通过概率方法完成这一证明：对于从$\mathcal{S}_{1,k}$的某个子集均匀采样的 $\vec{\Delta}$，我们给出 $\mathbb{E}[\|\sqrt{A}\vec{\Delta}\|_2^2]$的下界。

8.1 敏感度作为二次型与矩阵性质

设 $R \in \mathbb{R}^{T\times T}$ 为一个下三角托普利兹矩阵，其对角线为 r_0、……、r_{T-1}，且单调不增并且非负。注意，在这样的矩阵中，对于 $i \geq j$，有 $R[i,j] = r_{i-j}$。接下来我们将分析 $\|R\vec{\Delta}\|_2^2$，因为这对应于一个在代数上更易于推理的二次型。根据定义，我们有

$$\|R\vec{\Delta}\|_2^2 = \vec{\Delta}^T R^T R\vec{\Delta}.$$

我们首先分析矩阵 R^TR 的性质，然后对上述二次型进行推理，从而推导出敏感度 $\mathsf{sens}_2(R, \mathcal{S}_{D,k})$ 的上界。令 $c(i, j) \equiv R^TR[i,j]$。我们将建立 $c(i, j)$ 的一些关键单调性性质，这些性质将有助于后续分析。

命题 8.3。令 $c(i,j)$ 如上文所定义。那么，

1. 对称性。**对于所有** $0 \leq i, j \leq T-1$，**有** $c(i,j) = c(j,i)$。**2. 沿行单调性（超出对角元素）**。**令** $0 \leq i \leq j \leq T-1$。**对于** $0 \leq m \leq T-j-1$，**有** $c(i,j) \geq c(i, j+m)$。**3. 对角元素单调性**。**令** $0 \leq m \leq j \leq T - max(i,j) - 1$。**有** $c(i,j) \geq c(i+m, j+m)$。

证明。首先，我们写出 $c(i, j)$ 的闭式表达式：

$$\begin{aligned} c(i,j) = (R^TR)[i,j] &= \sum_{k=0}^{T-1} R^T[i,k]R[k,j] \\ &= \sum_{k=\max(i,j)}^{T-1} R[k,i]R[k,j] = \sum_{k=\max(i,j)}^{n-1} r_{k-i}r_{k-j} \\ &= \sum_{\ell=0}^{n-\max(i,j)-1} r_\ell r_{\ell+|i-j|}, \end{aligned} \tag{3}$$

其中，第二个等式由矩阵乘法的定义得出；第三个等式是因为 R 为下三角矩阵，并依据转置的定义得出；第四个等式是因为 R 为

托普利兹矩阵，因此 $R[i,j] = r_{i-j}$，最后一个等式则是代入 $\ell = k - \max(i,\ j)$ 并用 ℓ 重写求和得出。

现在我们证明该断言的每一项。

1. **对称性**。 $c(i,\ j) = c(j,\ i)$：根据转置运算的性质，这是因为 $(R^T R)^T = R^T R$。

2. **沿行单调性（超出对角线元素）**。令 $0 \le i \le j \le T-1$。对于 $0 \le m \le T-j-1$，考虑 $c(i,\ j) - c(i,\ j+m)$。根据公式（3），我们有

$$\begin{aligned} c(i,j) - c(i,j+m) &= \sum_{\ell=0}^{T-j-1} r_\ell r_{\ell+j-i} - \sum_{\ell=0}^{T-j-m-1} r_\ell r_{\ell+j+m-i} \\ &= \sum_{\ell=0}^{T-j-m-1} r_\ell [r_{\ell+j-i} - r_{\ell+j+m-i}] + \sum_{\ell=T-j-m}^{T-j-1} r_\ell r_{\ell+j-i} \\ &\ge 0\,, \end{aligned}$$

   其中，最后一个不等式源于以下事实：对于我们所关注的矩阵，$r_s \ge r_t$ for $s \le t$，并且所有 r 值均为非负数。

3. **对角线元素单调性**。令 $0 \le i \le j \le T-1$。对于 $0 \le m \le T - \max(i,\ j) - 1$，考虑$c(i,\ j) - c(i+m,\ j+m)$。根据公式（3），我们有

$$\begin{aligned} c(i,j) - c(i+m,j+m) &= \sum_{\ell=0}^{T-j-1} r_\ell r_{\ell+j-i} - \sum_{\ell=0}^{T-j-m-1} r_\ell r_{\ell+j+m-i-m} \\ &= \sum_{\ell=T-j-m}^{T-j-1} r_\ell r_{\ell+j-i} \ge 0\,, \end{aligned}$$

   其中最后一个不等式成立，是因为每个 r 值都是非负的。

□

现在我们可以分析所关注的二次型了。

$$\|R\vec{\Delta}\|_2^2 \vec{\Delta}^T R^T R \vec{\Delta} = \sum_{i,j\in[T]} (R^T R)[i,j]\vec{\Delta}[i]\vec{\Delta}[j] = \sum_{i,j\in[T]} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j]\,. \tag{4}$$

现在，我们针对所关注的敏感度向量集（即第 6 节所述的敏感度向量集）分析这一量。

8.2 敏感度上界

下面的定理是本节的关键结果，给出了 $\mathcal{S}_{D,k}$-流的 ℓ_2 敏感度上界。

定理 8.4。设 $R \in \mathbb{R}^{T\times T}$ 为一个对角线元素非递增且非负的下三角托普利茨矩阵。固定 $D, k \ge 1$。则，

$$\mathsf{sens}_2(R, \mathcal{S}_{D,k}) \le \sqrt{Dk} \cdot \|R\|_{1\to 2}\,.$$

我们将为该定理给出两种不同的证明。第一种证明是组合性的，依赖于一个匹配论证。可以说，相较于第二种更简短、更直接的证明，第一种证明更直观。

通过匹配论证证明定理 *8.4*。 首先，考虑较简单的情形 $D = 1$。考虑任意 $\vec{\Delta} \in \mathcal{S}_{1,\ k}$。根据公式 (4)，

$$\|R\vec{\Delta}\|_2^2 = \sum_{i,j\in[T]} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j] = \sum_{i=0}^{T-1} c(i,i)\vec{\Delta}[i]^2 + 2\sum_{0\le i<j\le T-1} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j]$$
$$\le \sum_{i=0}^{k-1} c(i,i) + 2\sum_{0\le i<j\le T-1} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j]\,. \tag{5}$$

其中，第二个等式由定理 8.3 的第 1 项得出（$c(i,j) = c(j,i)$）；第二个不等式则由 $\vec{\Delta} \in \{-1,0,1\}^T$、$\vec{\Delta}$ 是 k-稀疏的，以及定理 8.3 的第 3 项得出（对于 $0 \le m \le T - \max(i,j) - 1$，$c(i,j) \ge c(i+m, j+m)$）。

我们将通过一个组合论证说明上述求和中的第二项为负，从而将结果上界为第一项。

考虑以下匹配有序对的匹配操作 M：对于每个满足 $\vec{\Delta}[i] \ne \vec{\Delta}[j]$ 且 $\vec{\Delta}i, \vec{\Delta}[j] \ne 0$ 的 $i < j$，将 (i,j) 与 (i,j') 匹配，其中 j' 是大于 j 且满足 $\vec{\Delta}[i] = \vec{\Delta}[j']$ 的最小值。如果这样的 j' 不存在，则将 (i,j) 留作未匹配。

令 包含匹配项（$i,.$）的集合为 M_i。令未匹配有序对（$i,.$）的集合为 N $_i$.
我们观察到该匹配 M 的两个性质：

- **所有有序对** $(i,j) \in N_i$ **都满足** $x_i \ne x_j$。我们首先论证，满足 $\vec{\Delta}[i], \vec{\Delta}[j] \ne 0$ 且 $\vec{\Delta}[i] = \vec{\Delta}[j]$ 的每个 (i,j) 都已匹配。为此，考虑这样的 i 和 j。由于 Δ 是交替的，存在一个 $i < j_1 < j$ 使得 $\vec{\Delta}[j_1] \ne 0$、$\vec{\Delta}[i] \ne \vec{\Delta}[j_1]$，并且在 $\vec{\Delta}[j_1]$ 与 $\vec{\Delta}[j]$ 之间不存在非零项。注意，这意味着 (i,j) 和 (i,j_1) 彼此匹配。这意味着所有满足 $x_i \ne x_j$ 的有序对 $(i,j) \in N_i$。
- **匹配对的贡献非正**。我们还要说明，在每个匹配 $\{(i,\ j_1),\ (i,\ j_2)\} \in M_i$ (其中 $\vec{\Delta}[j_1] \ne \vec{\Delta}[i]$ 和 $\vec{\Delta}[j] = \vec{\Delta}[i]$)中，该匹配内有序对对所关注总和的总贡献非正，即 $c(i,\ j_1)\vec{\Delta}[i]\vec{\Delta}[j_1] + c(i,\ j_2)\vec{\Delta}[i]\vec{\Delta}[j_2] \le 0$。这是因为 $i < j_1 < j_2$，根据匹配的构造方式可知，由定理 8.3 的第 2 项可得$c(i,\ j_2) \le c(i,\ j_1)$.

接着，分析公式（5）右侧的第二项，并将其化归为分别针对每个指标 i进行推理，我们得到

$$\begin{aligned}
&2\sum_{0\le i<j\le T-1} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j] \\
&= 2\sum_{0\le i\le T-1}\left(\sum_{(i,a)\in N_i} c(i,a)\vec{\Delta}[i]\vec{\Delta}[a] + \sum_{\{(i,s),(i,t)\}\in M_i} c(i,s)\vec{\Delta}[i]\vec{\Delta}[s] + c(i,t)\vec{\Delta}[i]\vec{\Delta}[t]\right) \\
&\le 0
\end{aligned}$$

其中，最后一个不等式由关于匹配的两个性质以及$c(.,.)$始终非负这一事实得出。因此，总体而言，由公式（5）可得

$$\|R\vec{\Delta}\|_2 \le \sqrt{\sum_{i=0}^{k-1} c(i,i)} \le \sqrt{kc(0,0)} = \sqrt{k}\cdot \mathsf{sens}_2(R, \mathcal{S}_{1,1}),$$

其中，最后一个不等式由对角元素的单调性（定理 8.3 的第 3 项）得出。特别地，我们已经证明

$$\mathsf{sens}_2(R, \mathcal{S}_{1,k}) \le \sqrt{k}\cdot \mathsf{sens}_2(R, \mathcal{S}_{1,1})\,.$$

为了完成　　证明中，我们现在考虑 $D > 1$。在第一步应用定理 7.3，可得

$$\begin{aligned}\mathsf{sens}_2(R, \mathcal{S}_{D,k})^2 &\leq \left(\max_{k_1,\dots,k_D : \sum_{i=1}^D k_i = k} \sum_{i=1}^D \mathsf{sens}_2(R, \mathcal{S}_{1,k_i}) \right)^2 \\ &\leq \max_{k_1,\dots,k_D : \sum_{i=1}^D k_i = k} \left(\sum_{i=1}^D \sqrt{k_i} \|R\|_{1\to 2} \right)^2 \leq Dk \cdot \|R\|_{1\to 2}^2\end{aligned}$$

其中，第二步使用了我们关于 $D = 1$ 的结果，最后一步使用了柯西-施瓦茨不等式。取平方根即可完成证明。□

除了这个直观的组合证明之外，我们还给出了一种更具解析性的证明。虽然本文没有利用这一点，但该证明还有一个额外优点：即使将 $\mathcal{S}_{D,\,k}$ 定义在 $\mathbb{R}^T$ 上而非 $\mathbb{Z}^T$ 上，它仍然成立。

通过阿贝尔求和证明定理 *8.4*。任取 $\vec{\Delta} \in \mathcal{S}_{D,\,k} \subseteq \mathbb{R}^T$。我们可以写成

$$\|R\vec{\Delta}\|_2^2 = \sum_{i,j\in[T]} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j] = \sum_{i=0}^{T-1} \vec{\Delta}[i] \underbrace{\sum_{j=0}^{T-1} c(i,j)\vec{\Delta}[j]}_{Q_i},$$

其中我们继续处理内层求和。定义具有有界支撑的无限序列 $(f_t)_{t\in\mathbb{Z}}$，$(g_t)_{k\in\mathbb{Z}}$：

$$f_t = \begin{cases} c(i,t) & \text{for } 0 \leq t \leq T-1\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad g_t = \begin{cases} \vec{\Delta}[t] & \text{for } 0 \leq t \leq T-1\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

定义 $G_m = \sum_{t=-\infty}^{m} g_t$，我们可以写成

$$Q_i = \sum_{j=0}^{T-1} c(i,j)\vec{\Delta}[j] = \sum_{t\in\mathbb{Z}} f_t(G_t - G_{t-1}) = \sum_{t\in\mathbb{Z}} f_t G_t - \sum_{t\in\mathbb{Z}} f_t G_{t-1} = \sum_{t\in\mathbb{Z}} (f_t - f_{t+1}) G_t\,,$$

其中，最后一步改变了第二个求和中的索引，随后将两个求和合并。这种变换有时称为阿贝尔求和。根据定理 8.3 中 $c(i,\ j)$ 的单调性、非负性和对称性，我们有 $c(i,\ j) - c(i,\ j+1) \geq 0$ 当 $j \geq i$ 时，以及 $c(i,\ j) - c(i,\ j+1) \leq 0$ 当 $j < i$ 时。因此可以写成

$$\begin{aligned} Q_i &= \sum_{j\geq i} [c(i,j) - c(i,j+1)]\, G_j - \sum_{j<i} [c(i,j+1) - c(i,j)]\, G_j \\ &\leq \sup_{j\in\mathbb{Z}}(G_j) \sum_{j\geq i} [c(i,j) - c(i,j+1)] - \inf_{j\in\mathbb{Z}}(G_j) \sum_{j<i} [c(i,j+1) - c(i,j)] \\ &= \left(\sup_{j\in\mathbb{Z}}(G_j) - \inf_{j\in\mathbb{Z}}(G_j) \right) c(i,i)\,, \end{aligned}$$

其中，不等式利用了对于非负序列 $(a_k)_{\ k}$，$\inf_k(b_k) \sum_k a_k \leq \sum_k a_k b_k \leq \sup_k(b_k) \sum_k a_k$；最后一个等式则利用了求和项会逐项抵消这一点。通过类似的论证（交换求和项的作用），我们还可以推导出一个下界：

$$Q_i \geq \left(\inf_{j\in\mathbb{Z}}(G_j) - \sup_{j\in\mathbb{Z}}(G_j) \right) c(i,i)\,.$$

结合这两个界，我们得到

$$|Q_i| \le \left(\sup_{j\in\mathbb{Z}}(G_j) - \inf_{j\in\mathbb{Z}}(G_j)\right) c(i,i) \le D\cdot c(i,i)\,,$$

其中，不等式利用了 G_j 是 $\vec{\Delta}$ 上的前缀，因此差值 $\sup_{j\in\mathbb{Z}}(G_j) - \inf_{j\in\mathbb{Z}}(G_j)$ 是 Δ 上的（带符号）区间和，故根据 $\mathcal{S}_{D,k}$ 的定义，它有 D 上界。

因此，我们得到（对被加数取绝对值）

$$\|R\vec{\Delta}\|_2^2 = \sum_{i=0}^{T-1}\vec{\Delta}[i]Q_i \le D\sum_{i=0}^{T-1}|\vec{\Delta}[i]|c(i,i)\,.$$

利用 $c(i,i) \le c(0,0) = \|R\|_{1\to 2}^2$，以及根据定义 $\|\vec{\Delta}\|_1 \le k$：

$$\|R\vec{\Delta}\|_2^2 \le D\cdot\|R\|_{1\to 2}^2\sum_{i=0}^{T-1}|\vec{\Delta}[i]| \le Dk\cdot\|R\|_{1\to 2}^2\,.$$

代入敏感度的最终上界

$$\mathsf{sens}_2(R,\mathcal{S}_{D,k}) = \max_{\vec{\Delta}\in\mathcal{S}_{D,k}}\|R\vec{\Delta}\|_2 \le \sqrt{Dk}\cdot\|R\|_{1\to 2}$$

即可完成证明。 □

我们注意到，先前工作中的平方根分解 [8, 30, 40, 21, 35]（其中 $L = R = \sqrt{A}$，且 A 是下三角全一矩阵）满足上述 R 的性质。下面将对此作正式陈述，并给出该分解的一些有用性质。

定理 8.5 [35] $A \in \mathbb{R}^{T\times T}$ $\sqrt{A}\in$ $\mathbb{R}^{T\times T}$ ()。设为下三角全一矩阵。将其平方根定义为一个下三角托普利兹矩阵，其对角线元素为 t^{th} $r_t = \binom{2t}{t}/4^t$。也就是说，$\sqrt{A}[i,j] = r_{i-j}$ 对于 $i \ge j$。令 $c(i,j) = \left(\sqrt{A}^T\sqrt{A}\right)[i,j]$。那么，以下陈述成立：

1. $(r_t)_{t\in\mathbb{N}}$ 是一个递减序列，且满足 $r_0 = 1$，对于整数 $t \ge 1$，

$$\frac{1}{\sqrt{\pi(t+1)}} \le r_t \le \frac{1}{\sqrt{\pi t}} \qquad \text{[35, Lemma 4]}$$

2. 对于 $i \in [T]$，

$$1 + \frac{\ln(T-i)}{\pi} \le c(i,i) \le 1.067 + \frac{\ln(T-i)}{\pi} \qquad \text{[35, Theorem 3]}$$

3. 对于 $1/2$-零集中差分隐私，误差满足以下结果，

$$\mathsf{MaxSE}(\sqrt{A},\sqrt{A},\mathcal{S}_{1,1}) = \|\sqrt{A}\|_{1\to 2}^2 = \|\sqrt{A}\|_{2\to\infty}^2 = c(0,0)\,,$$
$$\mathsf{MeanSE}(\sqrt{A},\sqrt{A},\mathcal{S}_{1,1}) = 0.908 + \frac{\ln(T)}{\pi} + o(1) \le c(0,0) \qquad \text{[35, Theorem 4]}$$

4. 此外，

$$\frac{\|\sqrt{A}\|_F}{\sqrt{T}} = \sqrt{\frac{\ln(T)}{\pi}} + O(1)\,. \qquad \text{[35, Proof of Theorem 4]}$$

根据定理 8.5，我们现在可以证明定理 8.1。

定理 *8.1* 的证明。在 $\mathsf{MaxSE}(L,\ R,\ \mathcal{S}_{D,\ k})$ 和 $\mathsf{MeanSE}(L,\ R,\ \mathcal{S}_{D,\ k})$ 的定义中，利用定理 8.4 给出的 $\mathsf{sens}_2(R,\ \mathcal{S}_{D,\ k})$ 上界，即可得到推论陈述的第一部分。根据定理 8.5 的 (1)，我们还可以得出，$R = \sqrt{A}$ 满足定理 8.4 的单调性条件。将定理 8.5 的 (3) 中的误差界代入定理 8.1 的第一条陈述，即可证明最后一条陈述。 □

8.3 平方根敏感度的一个下界

对于平方根分解的情形，定理 8.4 允许我们给出一个上界 of

$$\mathsf{sens}_2(\sqrt{A}, \mathcal{S}_{D,k}) \le \sqrt{Dk}\|\sqrt{A}\|_{1\to 2} = \left(\frac{1}{\sqrt{\pi}} + o(1)\right)\sqrt{Dk\ln(T)}$$

对于 ℓ_2 敏感度。自然会问，对于 D 和 k的一般取值，我们的上界是否是紧的。我们证明，至少对于足够小的 k/D，该上界是紧的。

定理 8.6。 对于 $1 \le D \le k$、 T 足够大且 $k/D = O(T^{1/3})$，有

$$\mathsf{sens}_2(\sqrt{A}, \mathcal{S}_{D,k}) \ge \left(\frac{1}{\sqrt{\pi}} - o(1)\right)\sqrt{D^2\lfloor k/D\rfloor \ln\left(\frac{T}{\lfloor k/D\rfloor}\right)},$$

其中 $o(1)$趋于 零，当 $T \to \infty$时。如果进一步有 $k/D = \omega(1)$，则有以下更简单的上界

$$\mathsf{sens}_2(\sqrt{A}, \mathcal{S}_{D,k}) \ge \left(\frac{1}{\sqrt{\pi}} - o(1)\right)\sqrt{Dk\ln(DT/k)}\,.$$

本质上，对于 $k/D = O(T^{1/3})$，我们的敏感度上界具有正确的渐近行为。此外，如果 k/D 是次多项式的，并且满足以下任一条件：(1) $D = 1$，或 (2) k/D 是超常数，则上界中的主导常数也是紧的。

该定理通过概率方法证明：对于 $\|\vec{\Delta}\|_1 = k$为条件），我们证明 $\vec{\mathbb{E}}[\|\sqrt{A}\vec{\Delta}\|_2^2] \le \mathsf{sens}_2(\sqrt{A}, \mathcal{S}_{1,k})^2$ 必须很大。利用定理 7.1，我们进而能够将这一结果推广到$\mathcal{S}_{D,k}$。

我们需要以下两个引理来证明定理 8.6。第一个引理给出了平方根在 $i \ne j$时 $c(i, j)$的上界。证明该引理所用的技术是标准的，过去研究平方根分解的工作中也出现过类似的结论，参见例如 [30, 40, 47, 21, 48]。第二个引理给出了超几何随机变量奇偶性的期望上界。此类结果很可能是众所周知的，但我们未能找到相关参考文献。

引理 8.7。 令 $R = \sqrt{A} \in \mathbb{R}^{T\times T}$为平方根矩阵，并定义 $c(i,j) \equiv (R^TR)[i,j]$，其中 $i,j \in [T]$。对于 $i \ne j$，以下上界成立：

$$c(i,j) \le \frac{1}{\pi}\ln\left(1 + \frac{T}{|i-j|}\right) + \frac{3}{2}\,.$$

引理 8.8。 令 $N, K, n \ge 1$和 $X \sim$超几何分布(N, K, n)。也就是说， X表示从大小为 N、其中有 K 个成功元素的总体中不放回抽样，抽取 n次时成功抽取的次数。于是

$$|\mathbb{E}\left[(-1)^X\right]| \le \exp(-2nK/N) + n(n-1)/N\,.$$

定理 *8.7* 的证明。 回顾一下，平方根矩阵的元素为 $\sqrt{A}[i,j] = r_{i-j}$，其中 $r_t = \binom{2t}{t}/4^t$，并且特别满足以下上界

$$\frac{1}{\sqrt{\pi(t+1)}} \le r_t \le \frac{1}{\sqrt{\pi t}}$$

对于所有 $t \ge 1$ （定理 8.5）。

不失一般性，令 $j < i$。定义 $d = i - j > 0$，并令 $N = T - 1 - \max(i,j) = T - 1 - i$

$$c(i,j) = \sum_{t=0}^{N} r_t r_{t+d} \le \frac{1}{\sqrt{\pi d}} + \frac{1}{\pi}\sum_{t=1}^{N}\frac{1}{\sqrt{t(t+d)}} \le \frac{1}{\sqrt{\pi d}} + \frac{1}{\pi\sqrt{d+1}} + \frac{1}{\pi}\int_1^N \frac{\mathrm{dz}}{\sqrt{z(z+d)}}$$

我们继续计算积分：

$$\int_1^N \frac{\mathrm{dz}}{\sqrt{z(z+d)}} = \left[2 \cdot \operatorname{asinh}\left(\sqrt{\frac{z}{d}}\right)\right]_{z=1}^{z=N} \leq 2 \cdot \operatorname{asinh}\left(\sqrt{N/d}\right) ,$$

其中 $\operatorname{asinh}(x) \equiv \ln(x + \sqrt{1+x^2})$。给定 $x \geq 0$，我们有 $\sqrt{1+x^2} \leq 1+x$，因此 $\operatorname{asinh}(x) \leq \ln(1+2x)$。此外，注意到 $1 + 2\sqrt{x} \leq 3\sqrt{x}$ if $x > 0$，因此 $\ln(1+2\sqrt{x}) \leq \ln(3) + \frac{1}{2}\ln(x) \leq \ln(3) + \frac{1}{2}\ln(1+x)$ 也成立。同样地，$1 + 2\sqrt{x} \leq 3$ for $0 < x < 1$，因此 $\ln(1+2\sqrt{x}) \leq \ln(3) \leq \ln(3) + \frac{1}{2}\ln(1+x)$。因此我们可以写成

$$\operatorname{asinh}\left(\sqrt{N/d}\right) \leq \ln(1+2\sqrt{N/d}) \leq \ln(3) + \frac{1}{2}\ln(1+N/d) .$$

综合以上各项

$$\begin{aligned} c(i,j) &\leq \frac{1}{\sqrt{\pi d}} + \frac{1}{\pi\sqrt{d+1}} + \frac{2}{\pi}\left(\ln(3) + \frac{1}{2}\ln(1+N/d)\right) \\ &= \frac{1}{\sqrt{\pi d}} + \frac{1}{\pi\sqrt{d+1}} + \frac{2\ln(3)}{\pi} + \frac{1}{\pi}\ln(1+N/d) \\ &\leq \underbrace{\frac{1}{\sqrt{\pi}} + \frac{1}{\pi\sqrt{2}} + \frac{2\ln(3)}{\pi}}_{\approx 1.489} + \frac{1}{\pi}\ln(1+N/d) \\ &\leq \frac{3}{2} + \frac{\ln\left(1+\frac{N}{d}\right)}{\pi} . \end{aligned}$$

由于 $c(i,j) = c(j,i)$，最终的引理表述可通过令 $d = |i-j|$ 并将 $N = T - \max(i,j) - 1 \leq T$ 取上界而得。□

定理 8.8 的证明。 我们将通过用二项分布近似该表达式来给出其上界。定义 $X \sim \mathrm{Bin}(n, p)$，其中 $p = K/N$。我们有

$$\mathbb{E}\left[(-1)^{\hat{X}}\right] = \sum_{\ell=0}^{n} (-1)^\ell \binom{n}{\ell} p^\ell (1-p)^{n-\ell} = (1-2p)^n \leq \exp(-2pn)$$

其中第二个等式由 $(1 - p + (-p))^n$ 的二项式展开得到。注意，X 描述的是ˆ无放回抽样，而 X 描述的是有放回 抽样。考虑改为从 N 中有放回地抽取 n 个样本的过程。避免碰撞（多次抽到同一元素）的概率为：

$$\Pr[\#\text{collisions} = 0] = \frac{N}{N} \cdot \frac{N-1}{N} \cdot \ldots \cdot \frac{N-n+1}{N} = \prod_{\ell=0}^{n-1}(1-\ell/N)$$

因此

$$\Pr[\#\text{collisions} \geq 1] = 1 - \prod_{\ell=0}^{n-1}(1-\ell/N) \leq \sum_{\ell=0}^{n-1}\frac{\ell}{N} = \frac{n(n-1)}{2N} ,$$

ˆ其中不等式是由并集界得到的。现在我们使用耦合技巧。令 X 表示从 N 中有放回地抽取这ˆn 个样本的结果。如果这 n 个样本之间没有碰撞，则定义 $X = X$。否则，如果发生碰撞，则根据无ˆ放回抽取 n 个新元素的结果设定 X。显然，X 和 X 分别服从正确的超几何分布和二项分布，但此时 $\Pr[X \neq \hat{X}] \leq \Pr[\#\text{碰撞} \geq 1] \leq {}^{n}_{(n-1)\ 2N}$。

为完成证明，我们证明这两个期望值很接近：

$$\begin{aligned} \left|\mathbb{E}\left[(-1)^X\right] - \mathbb{E}\left[(-1)^{\hat{X}}\right]\right| &= \left|\mathbb{E}\left[(-1)^X - (-1)^{\hat{X}}\right]\right| \leq \mathbb{E}\left[\left|(-1)^X - (-1)^{\hat{X}}\right|\right] \\ &= 2 \cdot \Pr[X + \hat{X} \text{ is odd}] \leq 2 \cdot \Pr[X \neq \hat{X}] \\ &\leq n(n-1)/N . \end{aligned}$$

最后的结论由应用三角不等式得出：

$$\left|\mathbb{E}\left[(-1)^X\right]\right| \le \left|\mathbb{E}\left[(-1)^{\hat{X}}\right]\right| + n(n-1)/N \le \exp(-2nK/N) + n(n-1)/N\,. \quad \square$$

我们接下来证明定理 8.6。

定理 8.6 的证明。 令 $R = \sqrt{A} \in \mathbb{R}^{T\times T}$。考虑 $D = 1$ 和 $k \ge 3$的情形；证明末尾我们会再讨论其他情形。我们将采用概率方法进行论证。考虑从 $\mathcal{S}_{1,k}$中均匀随机抽取的 Δ，并以 $\|\vec{\Delta}\|_1 = k$为条件。将此分布记为 $\mathcal{U}$。如果我们能证明 $\|R\vec{\Delta}\|_2^2$的期望有一个下界，那么这就意味着（平方）ℓ_2 敏感度也有一个下界，即

$$\mathsf{sens}_2(R, \mathcal{S}_{1,k}) \ge \sqrt{\mathbb{E}_{\vec{\Delta}\sim\mathcal{U}}\left[\|R\vec{\Delta}\|_2^2\right]}\,.$$

从现在起，我们省略期望的下标；所有期望均针对 $\mathcal{U}$取值。我们有

$$\mathbb{E}\left[\|R\vec{\Delta}\|_2^2\right] = \mathbb{E}\left[\sum_{(i,j)\in[T]^2} c(i,j)\vec{\Delta}[i]\vec{\Delta}[j]\right] = \sum_{i=0}^{T-1} c(i,i)\mathbb{E}\left[\vec{\Delta}[i]^2\right] + 2\sum_{j<i} c(i,j)\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]\,,$$

根据期望的线性性。我们先来看第一项。注意到 $\mathbb{E}[\vec{\Delta}[i]^2] = k/T$，我们可以直接利用定理 8.5 中关于 $c(i,i)$的上界来对其进行界定：

$$\sum_{i=0}^{T-1} c(i,i)\mathbb{E}\left[\vec{\Delta}[i]^2\right] \ge \frac{k}{T}\left(T + \sum_{i=1}^{T-1}\frac{\ln(T-i)}{\pi}\right) \ge \frac{k}{T}\left(\frac{(\pi-1)T}{\pi} + \frac{T\ln(T)}{\pi}\right) = \frac{k(\ln(T)+\pi-1)}{\pi}\,,$$

其中，最后一个不等式是将斯特林近似应用于 $\ln(T!) \ge T\ln(T) - T$ 得出的。

对于第二项，我们先讨论 $\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]$。令 $X \sim \mathrm{Hypergeom}(N, K, n)$ 表示一个服从超几何分布的随机变量，其中我们从总体中抽取 $n = k-2$ 个样本，总体大小为 $N = T-2$，其中 $K = i-j-1$ 个被标记为成功。我们断言

$$\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right] = \mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\middle|\vec{\Delta}[i]\vec{\Delta}[j]\ne 0\right]\Pr\left[\vec{\Delta}[i]\vec{\Delta}[j]\ne 0\right] = -\mathbb{E}\left[(-1)^X\right]\cdot\frac{k(k-1)}{T(T-1)}\,.$$

要理解最后一步，请注意，$\vec{\Delta}[i]$ 和 $\vec{\Delta}[j]$ 均非零的概率等于第二个因子；在以该事件为条件时，它们乘积的符号由 Δ 中介于 i 和 j 之间的个非零元素的数量的奇偶性决定，也就是说，由 X 的奇偶性决定。

根据定理 8.8，我们得到以下上界。

$$\begin{aligned}\left|\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]\right| &\le \frac{k(k-1)}{T(T-1)}\left(\exp\left(-\frac{2(k-2)(i-j-1)}{T-2}\right) + \frac{(k-2)(k-3)}{T-2}\right)\\ &\le \frac{k^2}{T^2}\exp\left(-\frac{2(k-2)(i-j-1)}{T-2}\right) + \frac{k^4}{T^3}\end{aligned}$$

现在我们可以开始处理非对角项了。

$$\begin{aligned}\left|2\sum_{j<i} c(i,j)\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]\right| &\le 2\sum_{j<i} c(i,j)\left|\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]\right|\\ &\le \underbrace{\frac{2k^2}{T^2}\sum_{j<i}\exp\left(-\frac{2(k-2)(i-j-1)}{T-2}\right)c(i,j)}_{S_{exp}} + \underbrace{\frac{2k^4}{T^3}\sum_{j<i}c(i,j)}_{S_{poly}}\end{aligned}$$

我们首先将论证 $S_{poly} = o(k \log k)$。注意，

$$\sum_{j<i} c(i,j) \le \sum_{i,j\in[T]} c(i,j) = \|\sqrt{A}\mathbf{1}\|_2^2 = O(T^2)\,.$$

最后一步需要进行一些计算，但可由 ($O = O\ k\quad k = O\ T^{1/3}$ — $\sqrt{A}\mathbf{1}\ [\ell] = \sum_{t=0}^{\ell} r_t = O\ \sqrt{\ell}$ $S_{poly} =$) ( ) 得出。因此，(k^4/T) ( ) 可由 ( ) 得出。

我们继续处理 S_{exp}。我们将使用 $c(i,\ j) \le \frac{3}{2} + \frac{1}{\pi}\ln(1 + \frac{T}{|i-j|})$，这来自定理 8.7。

$$\begin{aligned}
S_{exp} &= \frac{2k^2}{\pi T^2}\sum_{j<i}\exp\left(-\frac{2(k-2)(i-j-1)}{T-2}\right)\left(\frac{3\pi}{2} + \ln\left(1+\frac{T}{i-j}\right)\right)\\
&= \frac{2k^2}{\pi T^2}\sum_{j=0}^{T-2}\sum_{d=1}^{T-1-j}\exp\left(-\frac{2(k-2)(d-1)}{T-2}\right)\left(\frac{3\pi}{2} + \ln\left(1+\frac{T}{d}\right)\right)\\
&= \frac{2k^2}{\pi T^2}\sum_{d=1}^{T-1}(T-d)\exp\left(-\frac{2(k-2)(d-1)}{T-2}\right)\left(\frac{3\pi}{2} + \ln\left(1+\frac{T}{d}\right)\right)\\
&\le \frac{2k^2}{\pi T}\sum_{d=1}^{T-1}\exp\left(-\frac{2(k-2)(d-1)}{T-2}\right)\left(\frac{3\pi}{2} + \ln\left(1+\frac{T}{d}\right)\right)\\
&= \frac{3k^2}{T}\underbrace{\sum_{d=1}^{T-1} e^{-\alpha(d-1)}}_{Q_1} + \frac{2k^2}{\pi T}\underbrace{\sum_{d=1}^{T-1} e^{-\alpha(d-1)}\ln\left(1+\frac{T}{d}\right)}_{Q_2},
\end{aligned}$$

其中 $\alpha = {}^{2}{}_{(k-2)\ T-2}$。我们首先给 Q_1 求上界：

$$Q_1 = \sum_{d=1}^{T-1} e^{-\alpha(d-1)} \le \sum_{d=0}^{\infty} e^{-\alpha d} = \frac{1}{1-e^{-\alpha}} \le 1 + \frac{1}{\alpha} = \Theta(T/k)\,.$$

其中第二个不等式利用了对于任意 $x \ge 0,\ \ 1 - e^{-x} \ge x/(1+x)$ 这一事实。对于 Q_2，我们需要进行更多工作。令 $f(x) = e^{-\alpha x}\ln\left(1 + \frac{T}{x+1}\right)$，当 $x \ge 0$ 时该式为正。计算其导数可得，

$$f'(x) = -e^{-\alpha x}\left[\alpha\ln\left(1+\frac{T}{x+1}\right) + \frac{T}{(x+1)^2 + T(x+1)}\right],$$

我们可以验证它在 $[0,\infty)$ 上为负，因此 f 在此范围内严格递减。由此，我们可以对 Q_2 求上界：

$$Q_2 = \sum_{d=0}^{T-2} f(d) \le \sum_{d=0}^{\infty} f(d) \le f(0) + \int_0^{\infty} f(x)\,\mathrm{dx} = \ln(1+T) + \int_0^{\infty} e^{-\alpha x}\ln\left(1+\frac{T}{x+1}\right)\mathrm{dx}\,.$$

我们继续处理该积分：

$$\int_0^{\infty} e^{-\alpha x}\ln\left(1+\frac{T}{x+1}\right)\mathrm{dx} \le \int_0^{\infty} e^{-\alpha x}\ln(1+T/x)\,\mathrm{dx} \le \frac{1}{\alpha}\int_0^{\infty} e^{-y}\ln(1+K/y)\,\mathrm{dy}\,,$$

其中，最后一步进行了代换 $y = \alpha x$，并令 $K = \alpha T = \Theta(k)$。重新改写被积函数，可得

$$\ln(1+K/y) = \ln K - \ln y + \ln(1+y/K)\,.$$

因此，

$$\int_0^\infty e^{-y}\ln(1+K/y)\,\mathrm{d}y = \ln K \underbrace{\int_0^\infty e^{-y}\,\mathrm{d}y}_{=1} - \underbrace{\int_0^\infty e^{-y}\ln y\,\mathrm{d}y}_{=-\gamma} + \int_0^\infty e^{-y}\ln(1+y/K)\,\mathrm{d}y\,,$$

其中 $\gamma \approx 0.5772$ 是欧拉–马歇罗尼常数。对于最后一个积分，我们对 $u \geq 0$ 使用上界 $\ln(1+u) \leq u$：

$$\int_0^\infty e^{-y}\ln(1+y/K)\,\mathrm{d}y \leq \frac{1}{K}\int_0^\infty ye^{-y}\,\mathrm{d}y = \frac{1}{K}\,.$$

现在我们可以对 Q_2 取上界：

$$Q_2 \leq \ln(1+T) + \frac{1}{\alpha}\int_0^\infty e^{-y}\ln(1+K/y)\,\mathrm{d}y \leq \ln(1+T) + \frac{1}{\alpha}\left(\ln K + \gamma + \frac{1}{K}\right).$$

As $\frac{2k}{T-2} \leq \alpha = \frac{2\,(k-2}{T-2} \leq \frac{2k}{T}$)，因此 $2(k-2) \leq K = \alpha T \leq 2k$，我们进一步简化这个上界：

$$Q_2 \leq \frac{T-2}{2k}\Bigg(\ln(2k) + \underbrace{\frac{2k\ln T}{T-2} + \gamma + \frac{1}{2(k-2)}}_{O(1)}\Bigg) = \frac{T\ln k}{2k} + O(T/k)$$

其中 $k\ln(T)/T = O(1)$，当 $k = o(T)$ 时。接下来我们对 S_{exp} 取上界：

$$S_{exp} \leq \frac{k^2}{\pi T}(3\pi Q_1 + 2Q_2) \leq \frac{k^2}{\pi T}\left(\frac{T\ln k}{k} + O(T/k)\right) = \frac{k\ln k}{\pi} + O(k)\,.$$

现在我们可以把所有部分整合起来了。

$$\begin{aligned}
\mathsf{sens}_2(R,\mathcal{S}_{1,k})^2 &\geq \mathbb{E}\left[\|R\vec{\Delta}\|_2^2\right] = \sum_{i=0}^{T-1} c(i,i)\mathbb{E}\left[\vec{\Delta}[i]^2\right] + 2\sum_{j<i} c(i,j)\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right] \\
&\geq \frac{k\ln(T)}{\pi} + O(k) - \left|2\sum_{j<i} c(i,j)\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right]\right| \\
&\geq \frac{k\ln(T)}{\pi} + O(k) - S_{exp} - S_{poly} \\
&\geq \frac{k\ln(T)}{\pi} - \frac{k\ln(k)}{\pi} - O(k) \\
&= (1-o(1))\cdot\frac{k\ln(T/k)}{\pi}\,.
\end{aligned}$$

要得到最终结论，取平方根。

$$\mathsf{sens}_2(R,\mathcal{S}_{1,k}) \geq \sqrt{(1-o(1))\cdot\frac{k\ln(T/k)}{\pi}} = (1-o(1))\cdot\sqrt{\frac{k\ln(T/k)}{\pi}}\,.$$

最后，我们讨论 k 和 D 更一般的情形。注意，对于 $D=1$，$k=1$，问题是平凡的，因为 $\mathsf{sens}_2(R, S_{1,\,1}) = \|R\|_{1\to 2}$。对于 $D=1$，$k=2$，则需要多做一些工作。现在可以直接计算非对角项

$$2\sum_{j<i} c(i,j)\mathbb{E}\left[\vec{\Delta}[i]\vec{\Delta}[j]\right] = -\frac{2\cdot 2\cdot 1}{T(T-1)}\sum_{j<i} c(i,j) = O(1)$$

根据与之前相同的论证，最后的求和为 $O(T^2)$，而与对角贡献 $k\ln(T)/\pi$ 相比，这一项的阶更低。对于 $1 < D \le k$，我们使用定理 7.1：

$$\mathsf{sens}_2(R, \mathcal{S}_{D,k}) \ge D \cdot \mathsf{sens}_2(R, \mathcal{S}_{1,\lfloor k/D \rfloor}) \ge \left(\frac{1}{\sqrt{\pi}} - o(1)\right) \cdot \sqrt{D^2 \lfloor k/D \rfloor \ln\left(T/\lfloor k/D \rfloor\right)}$$

这使我们能够化归到先前的情形，其中 $D = 1$。此外，若 $k/D = \omega(1)$，则 $\lfloor k/D \rfloor = (1 - o(1))k/D$，该表达式还可以进一步简化一步。

$$\mathsf{sens}_2(R, \mathcal{S}_{D,k}) \ge \left(\frac{1}{\sqrt{\pi}} - o(1)\right) \cdot \sqrt{Dk \ln(DT/k)}\,. \qquad \square$$

定理 8.2 的证明。将敏感度 $\mathsf{sens}_2(R, \mathcal{S}_{D,k})$ 乘以 $\|\sqrt{A}\|_{2\to\infty}$ 或 $\|\sqrt{A}\|_F/\sqrt{T}$（分别对应 MaxSE 和 MeanSE），再除以 $\sqrt{2\rho}$，即可得到该推论。由于 $\|\sqrt{A}\|_{2\to\infty}$ 和 $\|\sqrt{A}\|_F/\sqrt{T}$ 均为 $\left(\frac{1}{\sqrt{\pi}} + o(1)\right)\sqrt{\ln(T)}$（参见定理 8.5 中的 (2)、(3) 和 (4)），该推论结论显然成立。

$\square$

9 基于树的分解

首个能够为私有持续计数实现非平凡误差界的机制是二叉树机制 [23, 13]。本节将研究其自然推广，即 b 叉树机制；先前的工作 [54, 17, 12, 4] 已对其进行研究。尽管我们关注的是 ρ-零集中差分隐私，与在 $\mathcal{S}_{1,1}$-数据流上进行标准持续计数的情形一样，基于树的分解对于纯差分隐私也能给出当前最优的误差界。本节末尾将给出这两种隐私变体的误差界。

在第 9.1 节中，我们介绍在常规$\mathcal{S}_{1,1}$计数背景下研究的基于树的分解。这包括陈述关于左因子的必要（已知）上界，我们也为完整起见给出证明。接下来，在第 9.2 节中，我们研究$\mathcal{S}_{D,k}$数据流的敏感度计算。我们展示了在树的背景下，相对于$\mathcal{S}_{1,k}$ 的右因子进行的敏感度计算可归结为一个奇偶计数问题，随后可利用定理 7.4 将其扩展到$\mathcal{S}_{D,k}$ 。最后，第 9.3 节给出了针对一般$\mathcal{S}_{D,k}$数据流的基于树的分解的紧确误差界。

9.1 b元树机制引言

我们先简要总结步数为 $T = b^h$ （其中某个正整数 h）时的情形。全文将使用以下记号

$$\mathcal{T}_{b,h} = \bigcup_{\ell=0}^{h} \mathcal{T}_{b,h}^{\ell} \quad \text{where} \quad \mathcal{T}_{b,h}^{\ell} = \{[j \cdot b^{\ell}, (j+1) \cdot b^{\ell}] : \ 0 \le j \le b^{h-\ell} - 1\}\,,$$

表示高度为 h、具有 $m = (b^{h+1} - 1)/(b - 1)$个节点的完全 b元树。 b元树机制可以表示为矩阵 $L_b \in \{0,1\}^{T \times m}$, $R_b \in \{0,1\}^{m \times T}$的一种因子分解机制。定义节点到索引的双射 $i : \mathcal{T}_{b,h} \to [m]$后，分解 $L_b R_b = A$ 满足

1. For each node $v \in \mathcal{T}_{b,h}$: $(R\vec{\mathbf{x}})_{i(v)} = \sum_{j \in v} \vec{\mathbf{x}}[j]$.
2. 对于每个时间步 ，将 $\mathcal{Q}_b(t)$ 定义为满足条件的节点 $\mathcal{Q}_b(t) \subseteq \mathcal{T}_{b,h}$ 的（唯一）极小集合 g

$$\bigcup_{v \in \mathcal{Q}_b(t)} v = [0, t] \qquad \text{and} \qquad \forall v, v' \in \mathcal{Q}_b(t) : v \cap v' = \emptyset \quad \text{or} \quad v = v'\,.$$

然后对于每个 $v \in \mathcal{Q}_b(t)$: $L_b[t, i(v)] = 1$。

直观地说， b 叉树机制会发布节点 $R_b \vec{\mathbf{x}}$ 的加噪版本，并通过在每一步将最少数量的节点相加，将这棵树解码为一个前缀。由于 $|\mathcal{Q}_b(t)|$ 描述了在时间 t 相加的节点数量，因此可以分别根据 $|\mathcal{Q}_b(t)|$ 的最大值和平均值计算出 $\|L_b\|_{2\to\infty}$ 和 $\|L_b\|_F$。为完整起见，我们概述以下已知引理的证明（例如参见 [13, 4]）。

Le引理 9.1。 对于 $T = b^h$、 $\|L_b\|_{2\to\infty} = \sqrt{(b-1)h}$ 以及 $\|L_b\|_F/\sqrt{T} = \sqrt{(b-1)h/2 + b^{-h}}$。

我们将使用关于 $\mathcal{Q}_b(t)$ 的以下观察，例如参见 [13, 2]。

断言 9.2。 对于 $b \geq 2$，令 $\vec{\mathbf{w}}(i) \in [b]^h$ 表示整数 $i \leq b^h - 1$ 在基数 b下编码为一个 h 位数。则 $|\mathcal{Q}_b(t)| = \|\vec{\mathbf{w}}(t+1)\|_1$。

定理 *9.1* 的证明概要。 注意， $\|L_b\|_{2\to\infty}^2 =$ 最大值 $_{t\in[T]}|\mathcal{Q}_b(t)|$ 以及 $\|L_b\|_F^2 = \sum_{t\in[n]}|\mathcal{Q}_b(t)|$，因此计算归结为讨论 $\mathcal{Q}_b(t)$ 的最大规模和平均规模。这两点都将通过定理 9.2 进行论证。关于 $\|L_b\|_{2\to\infty}$ 的结果可由以下事实得到：编码 $\vec{\mathbf{w}}(b^h - 1)$ 具有 h 个数位，其数值为 $b-1$。对于 $\|L_b\|_F$，注意 $\vec{\mathbf{w}}(t+1)$ 在 T 步内遍历所有 h 位数，但不包括 $\vec{\mathbf{w}}(0)$；此外，还有一个 $(h+1)$ 位数 $\vec{\mathbf{w}}(T)$，除了最高位为 1 外其余均为 0。由于平均的 h 位数具有 ℓ_1 的权重 $h(b-1)/2$，因此得到 $\sum_{t\in[T]}|\mathcal{Q}_b(t)| = hT(b-1)/2 + 1$。除以 T 并取根即可完成证明。

□

结合这一结果以及 $\mathrm{sens}_2(R_b, \mathcal{S}_{1,\,1}) = \sqrt{h+1}$ 允许恢复常规持续计数中 MeanSE(L_b, R_b) 和 MaxSE(L_b, R_b) 的标准 $O(\log T)$ 界。通过对 $\mathrm{sens}_2(R_b, \mathcal{S}_{D,\,k})$ 进行紧确分析，我们的目标是将这些误差界扩展到 $\mathcal{S}_{D,\,k}$-流。

9.1.1 通过减法实现常数因子改进。

不过，在此之前我们注意到，固定右因子 R_b后，左因子 L_b 的选择并不唯一。近期研究 [4] 提出了另一种左因子选择方案，可在不牺牲时间或内存效率的情况下改善主导常数——对于常数 b，它们仍分别为 $O(T)$ 和 $O(\log T)$。尽管他们研究的是纯差分隐私，同一思路也可推广到 ρ-零集中差分隐私。接下来我们概述他们的构造。

设分支因子 b 为奇数。在每一层，他们的构造要么添加 $\leq (b-1)/2$ 个最左侧子节点，要么减去 $\leq (b-1)/2$ 个最右侧子节点，以形成一个前缀。树的根节点只能被添加，而每个父节点的中间子节点都不会被使用。这唯一确定了每一步所用的节点。将时间 t 前缀所用的（唯一）节点集合记为添加节点与减去节点的并集，即 $\hat{\mathcal{Q}}_b(t) = \hat{\mathcal{Q}}_b^+(t) \cup \hat{\mathcal{Q}}_b^-(t)$，于是有

$$(A\vec{\mathbf{x}})[t] = \sum_{i=0}^{t} \vec{\mathbf{x}}[i] = \sum_{v\in\hat{\mathcal{Q}}_b^+(t)} \sum_{i\in v} \vec{\mathbf{x}}[i] - \sum_{v\in\hat{\mathcal{Q}}_b^-(t)} \sum_{i\in v} \vec{\mathbf{x}}[i]\,.$$

因此，$\hat{L}_b$ 的每一行由 $\hat{\mathcal{Q}}_b(t)$ 定义，其中 $\hat{L}_b[t, i(v)] = 1$ 当 $v \in \hat{\mathcal{Q}}_b^+(t)$ 时等于 -1，当 $v \in \hat{\mathcal{Q}}_b^-(t)$ 时等于 -1，其他情况下为零。以下引理改编自 [4]。

引理 9.3。 设 $b \geq 3$ 为奇数，且 $T = b^h$ 对 $h \geq 1$成立。那么

$$\|\hat{L}_b\|_{2\to\infty} = \sqrt{\frac{(b-1)h+2}{2}}\,, \qquad \frac{\|\hat{L}_b\|_F}{\sqrt{T}} = \frac{\sqrt{b(1-1/b^2)h + 2(1+b^{-h})}}{2}\,.$$

为证明该结论，我们将使用来自 [4], 的以下论断，它将每步使用的节点数与 b进制编码中偏移量 的权重联系起来。

命题 9.4。 对于奇数 $b \geq 3$，令 $\tilde{\mathbf{w}}(i) \in [-(b-1)/2, (b-1)/2]^h$ 表示整数 $i \leq (b^h-1)/2$ 在以 b 为基数且带偏移数字的 h 位数中的编码，即$\hat{\quad}$ $\hat{\quad}$。则 $|\mathcal{Q}_b(t)| = \|\vec{\mathbf{w}}(t+1)\|_1$。

定理 *9.3* 的证明。 和之前一样，论证归结为分析

$$\max_{t\in[T]}|\hat{\mathcal{Q}}_b(t)| \quad \text{and} \quad \frac{1}{T}\sum_{t\in[T]}|\hat{\mathcal{Q}}_b(t)|\,.$$

最大节点数等于 $1+h(b-1)/2$，并且在包含根节点，且每一层都在 $t=T-2$ 处减去 $(b-1)/2$ 个节点时达到。平均节点数的推导需要更复杂的论证，但几乎可以直接从 [4] 中引理 14 的证明得出。根据定理 9.4，节点数再次与 1、.....、T 的编码的 ℓ_1 权重相关，但现在它们被编码为$h+1$位数，数位取自 $[-(b-1)/2, (b-1)/2]$。在所有 T 步中，最低有效的 h 个数位将遍历所有可用 b 位数表示的数，且恰好 遍历一次。单个数位的平均 ℓ_1 权重为$b(1-1/b^2)\ 4$，因此这些低位数位在 T 步内平均贡献 $^{b}{}_{(1-1/b^2)_4}{}^{h}$。考虑到树的根节点仅在最后 $(b^h+1)/2$ 步中使用，我们得到最终的平均节点数

$$\frac{b(1-1/b^2)h}{4}+\frac{(b^h+1)/2}{b^h}=\frac{b(1-1/b^2)h+2(1+b^{-h})}{4}\,,$$

完成证明。 □

对于最终的误差界，我们还需要考虑步数 T 不是 b 的整数次幂的情况。这里我们考虑 $h=\lceil\log_b T\rceil$，并将 L_b 定义为对应于 $T'=b^h$ 步的矩阵中前 T 行。根据 [4], ，这种情况下存在上界是隐含成立的，但该工作并未正式证明这一点。我们接下来给出证明。

推论 9.5. 设 $b\geq 3$ 为奇数， $T\geq b$，且 $h=\lceil\log_b(T)\rceil$ 满足 $b^{h-1}<T\leq b^h$。则

$$\|\hat{L}_b\|_{2\to\infty}\leq\sqrt{\frac{(b-1)h+2}{2}}\,,\qquad \frac{\|\hat{L}_b\|_F}{\sqrt{T}}\leq\frac{\sqrt{b(1-1/b^2)h+2(1+b^2+b^{-h})}}{2}\,.$$

此外，如果 $T\geq b^{2b}$，我们还有下界

$$\|\hat{L}_b\|_{2\to\infty}\geq\sqrt{\frac{(b-1)(h-1)+2}{2}}\,,\qquad \frac{\|\hat{L}_b\|_F}{\sqrt{T}}\geq\frac{\sqrt{b(1-1/b^2)(h-1)+2(1-b^2+b^{-h})}}{2}\,.$$

证明。 我们先考虑最大节点数。显然，该数可以分别以下界和上界取为时间 $t'=b^{h-1}-2<T-1$ 和 $t=b^h-2$ 使用的节点数。每一层都对应于添加根节点并减去 $(b-1)/2$ 个节点，树的高度为 $h-1$ 和 h。因此

$$1+\frac{(h-1)(b-1)}{2}\leq\max_{t\in[T]}|\hat{Q}_b(t)|\leq 1+\frac{h(b-1)}{2}\,,$$

最终的上界通过取平方根得到。

对于平均误差，我们将分析每一层的贡献偏离其平均值的程度。如前所述，最低的 h 数位在平均每 b^h 步中，每步贡献 $b(1-1/b^2)/4$ 个节点。注意，这些 h 个数位会循环出现，其中最低有效位的周期为 $b^{\ell+1}$。通过对 T 进行对抗性选择，我们可能会遇到一个不完整周期，其贡献高于或低于平均值。因此，这个数位贡献的节点平均数量最多会使平均值偏移 $0.5(b-1)b^{\ell+1}/T<0.5(b-1)b^{\ell-h+2}$。因此，最低 h 个数位的平均 ℓ_1 权重总偏移为 $<\sum_{\ell=0}^{h-1}0.5(b-1)b^{\ell-h+2}=b^2/2$。

对于上界，注意到在完整树中，根节点仅用于最后 $(b^h+1)/2$ 步，因此其在整棵树中的平均贡献本身就是一个上界。因此:

$$\frac{1}{T}\sum_{t\in[T]}\hat{Q}_b(t)\leq\frac{b^2}{2}+\frac{1}{b^h}\sum_{t\in[b^h]}\hat{Q}_b(t)=\frac{b(1-1/b^2)h+2(1+b^2+b^{-h})}{4}\,.$$

对于下界，我们取时间 $t = b^{h-1}$ 的精确平均值，并再次利用时间 T 时平均ℓ_1 权重的总偏移至多为 $b^2/2$，即：

$$\frac{1}{T}\sum_{t\in[T]}\hat{Q}_b(t) \geq -\frac{b^2}{2} + \frac{1}{b^{h-1}}\sum_{t\in[b^h]}\hat{Q}_b(t) = \frac{b(1-1/b^2)(h-1)+2(1-b^2+b^{-(h-1)})}{4}\,.$$

可以验证，当 $h \geq 1+2b$ 时，右侧为正。当 $T > b^{h-1} \geq b^{2b}$ 时，这一条件成立。对各个上界分别开平方，即可得到该结论。 □

定理 9.5 的核心意思是，当放宽到任意 T 时，定理 9.3 中的缩放关系仍然成立。对于足够大的 T 和常数 b，两个表达式均为 $\Theta(\sqrt{\log(T)})$，且具有显式的紧致首要常数。

9.2 通过奇偶性计数求敏感度

在给出左因子矩阵的上界后，我们将注意力转向本研究设定中特有的挑战：计算敏感度 $\mathsf{sens}_2(R_b,\ \mathcal{S}_{D,\ k})$。鉴于树的组合结构，研究 $\mathcal{S}_{1,\ k}$ 的问题将证明是可行的，其中敏感度向量具有 $\{0,\pm1\}^T$，交替且 k-稀疏的性质。在给出此情形下的紧界后，我们通过定理 7.4 将其推广至 $\mathcal{S}_{D,\ k}$。先前的工作 [43, 引理 B.1] 给出了一个上界

$$\mathsf{sens}_2(R_2, \mathcal{S}_{1,k}) \leq \sqrt{k(1+\log T)}$$

对于二叉树及其 ℓ_2 敏感度（将其邻接关系转换为我们的设定之后）。我们将把这个上界渐近地收紧为 $\Theta(\sqrt{k\log_b(T/k)})$，适用于 b叉树。

我们从组合角度着手，将问题化简为树上的奇偶性计数。暂且假设$T = b^h$ 为正整数 h，因此我们处理的是一棵完全树$\mathcal{T}_{b,\ h}$。称子集 $\mathcal{B} \subseteq [T]$为球的一种配置。对于任意节点$v \in \mathcal{T}_{b,\ h}$，我们将其子树视为一个箱，并定义 $o_v(\mathcal{B}) = |\mathcal{B}\cap v|$(模 2)为指示变量，用来表示节点 v 的箱中是否有奇数个球。定义

$$F_b(h,k) = \max_{\mathcal{B}\subseteq\{0,1\}^T:\,|\mathcal{B}|=k}\ \sum_{v\in\mathcal{T}_{b,h}} o_v(\mathcal{B})\,,$$

也就是说，$F_b(h,\ k)$是对$\mathcal{T}_{b,\ h}$ 中所有 k 个球的放置方式取最大值后，奇节点的数量。以下引理是我们分析的核心。

引理 9.6。对于 $p \geq 1$，敏感度$p(R_b,\ \mathcal{S}_{1,\ k}) = \sqrt[p]{\text{最大值}_{k'\leq k}\, F_b(h,\ k')}$。

证明。回想一下，$\mathsf{sens}^p(R_b,\ \mathcal{S}_{1,\ k}) = \max_{\vec{\Delta}\in\mathcal{S}_{1,\ k}} \|R_b\vec{\Delta}\|_p$。对于 $\vec{\Delta}\in\mathcal{S}_{1,\ k}$，注意到 $R_b\vec{\Delta}\in\{0,\pm1\}^m$，这是因为树中的每个节点都是 Δ 上的区间和，而且 Δ 的非零元素符号交替。令 $\mathcal{B}$ 等于 Δ 的非零索引，因此我们有

$$\|R_b\vec{\Delta}\|_p^p = \sum_{v\in\mathcal{T}_{b,h}} |(R_b\vec{\Delta})[i(v)]|^p = \sum_{v\in\mathcal{T}_{b,h}} o_v(\mathcal{B})\,,$$

因此，$\mathsf{sens}^p(R_b,\ \mathcal{S}_{1,\ k})^p = \max_{\mathcal{B}\in\{0,1\}^T:\ |\mathcal{B}|\leq k} \sum_{v\in\mathcal{T}_{b,\ h}} o_v(\mathcal{B}) = \max_{k'\leq k} F_b(h,\ k')$。 □

本质上，奇偶性计数精确描述了树的敏感度。接下来我们将展示如何使用动态规划计算 $F(h,\ k)$。

引理 9.7（动态规划）。令 $h \geq 1$ 和 $k \leq b^h$。那么

$$F_b(h,k) = \mathbb{1}_{[k\,\text{odd}]} + \max_{\substack{k_1+\cdots+k_b=k\\ \forall i:0\leq k_i\leq b^{h-1}}} \left\{\sum_{i=1}^{b} F_b(h-1,k_i)\right\}, \quad \text{where} \quad F_b(0,k) = k\,.$$

此外，$F_b(h,k)$ 可以通过动态规划计算，表的大小为 $O(k)$，时间复杂度为 $O(hk^{b-1})$。

证明概要。 根据问题描述，递推式的正确性显而易见。为了最大化奇节点的数量，我们可以枚举球分配给根节点的子节点的所有可能方式并递归处理，而每个叶节点对最终计数的贡献至多为 1。该递推式可以通过“自底向上”的动态规划求解，每层需要 $O(k)$ 的内存，并且可以在不同层之间重复使用。对于时间复杂度，注意到最大化过程枚举了将 k划分为 b 个（非负）整数部分的所有整数划分。此类整数划分的数量可以用将 k 划分为 b 个非负部分的弱组合数量作为上界： $\binom{k+b-1}{b-1} = O(k^{b-1})$（对于常数 b）。这种计数至多将整数划分的数量多算了一个常数因子 $b!$。在所有层上执行这一最大化过程，即得到所述的 $O(hk^{b-1})$ 时间复杂度。

□

我们指出，通过动态规划估计树的敏感度是一个自然的想法，类似的想法也已在相关但不同的多重参与背景下得到探索 [46]。当 b为常数时，时间复杂度关于 k 是多项式的，但一旦考虑更大的 b，就会很快变得不可行。受此启发，接下来我们将紧确地给出 $F_b(h,\ k)$ 的上界。

引理 9.8。 设 $b \geq 2$、$h \geq 1$、$1 \leq k \leq b^h$ 和 $\alpha_k = \lceil \log_b k \rceil$。则

$$k(h - \alpha_k + 1) \leq F_b(h,k) \leq k(h - \alpha_k + 1) + \frac{b^{\alpha_k} - 1}{b-1}\,,$$

其中，上界随 k 单调递增。

证明。 根据定义，层 $\ell^* = h - \alpha_k$ 是树中具有 k 个或更多节点的最高层。这由 $b^{h-(\ell^*-1)}$ $= b^{\alpha_k - 1} < k \leq b^{\alpha_k} = b^{h-\ell^*}$ 得出。对于上界，注意树中的层 ℓ 至多有 $\min(k,\ b^{h-\ell})$ 个奇节点，因此

$$F_b(h,k) \leq \sum_{\ell=0}^{h} \min(k, b^{h-\ell}) = \sum_{\ell=0}^{\ell^*} k + \sum_{\ell=\ell^*+1}^{h} b^{h-\ell} = k(h - \alpha_k + 1) + \frac{b^{\alpha_k} - 1}{b-1}\,,$$

其中，k 中的单调性可直接由第一个不等式之后的表达式得出。对于下界，我们通过向根位于第 ℓ^* 层的每棵子树至多添加一个球，将球分配给叶节点，从而在直至第 ℓ^* 层的每一层都获得 k 的贡献：

$$F_b(h,k) \geq k(\ell^* + 1) = k(h - \alpha_k + 1)\,,$$

于是结论得证。 □

推论 9.9. 令 $b \geq 2,\ h \geq 1,\ 1 \leq k \leq b^h$ 且 $\alpha_k = \lceil \text{对数}_b\, k \rceil$。则

$$\max\left(k(h - \alpha_k + 1), b^{\alpha_k - 1}(h - \alpha_k + 2)\right) \leq \mathsf{sens}_p(R_b, \mathcal{S}_{1,k})^p \leq k(h - \alpha_k + 1) + \frac{b^{\alpha_k} - 1}{b-1}\,.$$

证明。 由定理 9.6 可知，$\mathsf{sens}_p(R_b,\ \mathcal{S}_{1,\ k})^p = \max_{k' \leq k} F_b(h,\ k')$。利用定理 9.8 中的上界，可得

$$\mathsf{sens}_p(R_b, \mathcal{S}_{1,k})^p \leq \max_{k' \leq k} \left\{ k'(h - \alpha_{k'} + 1) + \frac{b^{\alpha_{k'}} - 1}{b-1} \right\} = k(h - \alpha_k + 1) + \frac{b^{\alpha_k} - 1}{b-1}\,,$$

其中 th 最后一步由 k' 中的单调性得出。下界的证明类似，我们得到 et

$$\mathsf{sens}_p(R_b, \mathcal{S}_{1,k})^p \geq \max_{k' \leq k} k'(h - \alpha_{k'} + 1) = \max_{k' \leq k} L(k')$$

但在这种情况下，$L(k')$ 并非关于 k' 单调递增。不过，它在每个区间（$b^{j-1},\ b^j$]（其中 $j \geq 1$ 为整数）上单调递增，并在此处达到最大值 $M_j = b^j(h - j + 1)$。容易验证，M_j 关于 j 非递减，例如，

$$\frac{M_{j+1}}{M_j} = b \cdot \frac{h-j}{h-j+1} \geq 2 \cdot \frac{1}{2} \geq 1\,.$$

由此可知，$L(k')$ 的最大值出现在 $k' = k$ 或 $k' = b^{\alpha_k - 1}$ 处，从而证明了该命题。 □

当 $k = o(T)$ t 上界中的第二项是低阶项，此时我们已经证明了 t

$$\mathsf{sens}_p(R_b\mathcal{S}_{1,k})^p = (1+o(1))\cdot k(h - \lceil \log_b k \rceil + 1)\,,$$

其渐近表现为 $\Theta(k\log(T/k))$.

9.2.1 一种敏感度降低技巧。

我们注意到，无需额外成本，就有可能进一步改善低阶项。关键观察是，大多数基于树的分解实际上只使用树中大约 $1-1/b$ 的节点，因此，在 $F_b(h,\ k)$ 中，同样比例的奇数节点计数在敏感度计算中可以忽略。例如，标准的 b 叉树机制从不使用任何父节点最右侧的子节点；减法变体从不使用中间子节点。换句话说，$1-1/b$ 的列在 L_b（以及 L_b）中恒为零，因此我们可以将 R_b 中对应的行置零，得到 R_b，从而获得一个敏感度更低的新分解 $L_bR_b = A$。这种优化是我们这一设定所独有的，对标准的 $\mathcal{S}_{1,\ 1}$ 持续计数并无改进。

ˆ 接下来，我们将推导 R_b 的紧界。其分析与前面完整树的情况几乎完全相同，其中定理 9.12 对应于定理 9.9。令 $F(h,k)$ 表示相应的奇数节点数，其中只有 $b-1$ 个子节点（共 b 个）能够作出贡献。为完整起见，我们给出一个用于 $F(h,k)$ 的匹配动态ˆ规划。

引理 9.10 （第二个动态规划）。令 $h \ge 1,\ k \le b^h$ 和 $w \in \{0,1\}$。定义 $F_b(h,k,w)$ vi a

$$F_b(h,k,w) = w\cdot \mathbb{1}_{[k\,\mathrm{odd}]} + \max_{\substack{k_1+\cdots+k_b=k\\ \forall i:0\le k_i\le b^{h-1}}} \left\{ F_b(k_1,h-1,0) + \sum_{i=2}^{b} F_b(k_i,h-1,1) \right\},$$

$$F_b(k,0,w) = w\cdot k\,.$$

那么对于实数 $p \ge 1$，$\mathsf{sens}_p(\hat{}R_b, \mathcal{S}_{1,k}) = \sqrt[p]{\max_{k'\le k} \hat{}F_b(h,k')} = \sqrt[p]{\max_{k'\le k} F_b(k,h,1)}$。此外，$F_b(k,h,1)$ 可以通过动态规划计算，其表大小为 $O(k)$，所需时间为 $O(hk^{b-1})$。

证明概要。注意，递推式与定理 9.7 中的递推式类似，只是现在多了一个变量 w。当对应节点对总奇数节点数有贡献时，该变量设为 1，否则设为 0。任意节点的 b 个子节点中，恰有 $b-1$ 个被赋值为 $w=1$；而只有节点满足 $w=1$ 时，$w\cdot\mathbb{1}_{[k\text{ 奇数}]}$ 才会产生贡献。该敏$_p$($\hat{}R_b$, $\mathcal{S}_{1,\ k}$)^p =max$_{k'\le k}$ˆF_b(h, k') 可由与定理 9.6 类似的论证得出：ˆ$R_b\vec{\Delta}$ 仍然只能取 $\{0,\pm1\}$ 中的值，因此，最大化缩减树的奇数节点数就意味着最大化范数。

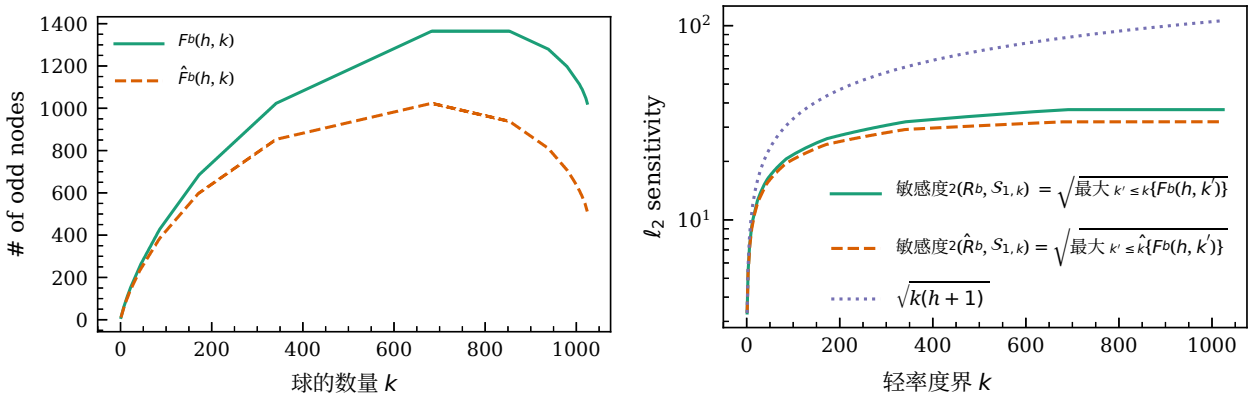
对于动态规划的实现，请注意，将 k 拆分为 b 个部分的弱（有序）加法组合数，仍是固定 k, b 时需要优化的配置数的上界。此外，引入变量 w 至多会使表的大小增加一倍。因此，其空间和时间开销在渐近意义上与定理 9.7 的相同。 □

在正式给出 ˆ$F(h,k)$ 的上界之前，我们先通过图 3 展示敏感度的改进效果。我们的方法相较于 [43], 给出的上界有显著提升，但使用“缩减”分解带来的收益并不明显。此外，b 的值越小，预期归约幅度越大，因为树中可从 R_b （大约 $1/b$）中丢弃的节点比例更高。结果表明，即使对于二叉树，除非 $k = \Omega(T)$，这种改进的阶也较低。下面将对此进行证明。

引理 9.11. 设 $h \ge 1$、$b \ge 2$、$1 \le k \le b^h$ 和 $\alpha_k = \lceil$对数$_b\, k\rceil$，则

$$k(h-\alpha_k+1-1/b) \le \hat{F}_b(h,k) \le k(h-\alpha_k+1) + b^{\alpha_k-1}\,,$$

且上界随 k 单调递增。



(a) 奇数节点数的比较。

(b) ℓ_2 敏感度的比较。

图 3: $b = 2$ 和 $h = 10$的奇数节点数及相应 ℓ_2 敏感度的比较。$\hat{}F_b(h,k)$ 和 $F_b(h,k)$ 根据定理 9.7 和 9.10 中的表述，使用动态规划计算得到。图 3a 直接绘制了 $F_b(h,k)$ 和 $\hat{}F_b(h,k)$ 的值，图 3b 则将它们分别与 R_b 和 $\hat{}R_b$ 的精确敏感度联系起来。图 3b 中顶部的绿色线使用的是 [43,引理 B.1]，在我们设定下转换得到的上界。

证明。 对于 $k = 1$，注意到 $\hat{}F_b(h, 1) = h + 1$ 同时满足两个上界。证明的其余部分聚焦于 $k \ge 2$。按照我们的 α_k 选择，与定理 9.8 中一样，我们再次得到，层 $\ell^* = h - \alpha_k$ 是树中含有 k 个或更多节点的最高层。对 k 的限制保证了 $\ell^* < h$。对于上界，树中的层 $\ell < h$ 至多有 $\min(k, (b-1)b^{h-\ell-1})$ 个奇节点，因此

$$\hat{F}_b(k,h) \le 1 + \sum_{\ell=0}^{h-1} \min(k, (b-1)b^{h-\ell-1}) = 1 + \sum_{\ell=0}^{\ell^*} k + (b-1) \sum_{\ell=\ell^*+1}^{h-1} b^{h-\ell-1} = k(h - \alpha_k + 1) + b^{\alpha_k - 1}.$$

上界的单调性由第一次不等式后的形式直接得到。对于下界，类似于定理 9.8，我们将 k 个球分配给每个根位于第 ℓ^*层的子树，并选择每个叶节点，使得我们从每一层都获得 k 的贡献。然而，在第 ℓ^*层，由于可能少于 k 个节点产生贡献，我们得到的贡献为 $\min(k,(b-1)b^{\alpha_k - 1})$。因此，

$$\hat{F}_b(h,k) \ge k(h - \alpha_k) + \min(k, (b-1)b^{\alpha_k - 1}) \ge k(h - \alpha_k + 1 - 1/b),$$

其中最后一步用到了 $(b-1)b^{\alpha_k - 1} \ge (1 - 1/b)k$。 □

推论 9.12. 令 $b \ge 3, h \ge 1, 1 \le k \le b^h$ 和 $\alpha_k = \lceil \log_b k \rceil$。则

$$\max\left(k(h - \alpha_k + 1 - 1/b), b^{\alpha_k - 1}(h - \alpha_k + 2 - 1/b)\right) \le \text{sens}_p(\hat{R}_b, \mathcal{S}_{1,k})^p \le k(h - \alpha_k + 1) + b^{\alpha_k - 1}.$$

证明。 根据定理 9.10，我们有 $\text{sens}_p(\hat{}R_b, \mathcal{S}_{1,k})^p = \max_{k' \le k} \hat{}F_b(h,k')$。应用定理 9.11 中的上界，我们有

$$\text{sens}_p(\hat{R}_b, \mathcal{S}_{1,k})^p \le \max_{k' \le k} \left\{k'(h - \alpha_{k'} + 1) + b^{\alpha_{k'} - 1}\right\} = k(h - \alpha_k + 1) + \frac{b^{\alpha_k} - 1}{b - 1},$$

这源于最大化目标关于 k' 是单调的。对于下界，同理可得

$$\text{sens}_p(R_b, \mathcal{S}_{1,k})^p \ge \max_{k' \le k} k'(h - \alpha_{k'} + 1 - 1/b) = \max_{k' \le k} L(k')$$

但在这种情况下， $L(k')$ 关于 k' 并非单调递增。然而，它在每个区间 $(b^{j-1}, b^j]$ （其中 $1 \le j \le h$ 为整数）上都是单调递增的，并在此处取得最大值 $M_j = b^j(h-j+1)$。我们直接验证 M_j 关于 j 是非递减的，

$$\frac{M_{j+1}}{M_j} = b \cdot \frac{h-j-1/b}{h-j+1-1/b} \ge b \cdot \frac{1-1/b}{2-1/b} \ge 3 \cdot \frac{1-1/3}{2-1/3} = 6/5\,,$$

这里我们用到了 $b \ge 3$。因此， $L(k')$ 的最大值要么在 $k' = k$ 处取得，要么在 $k' = b^{\alpha_k - 1}$ 处取得，从而证明了该结论。 □

请注意，该推论不支持 $b = 2$ （不过，上界支持）。

9.2.2 一般 T 的上界。

继续之前，我们还给出 T 不是 b 的整数次幂时敏感度的上界。令 $h = \lceil \log_b T \rceil$，并从 $\hat{}\, L_b$ 中删去最后 $b^h - T$ 行，使得 $\hat{}\, L_b \hat{}\, R_b = A = \mathbb{R}^{T \times T}$。请注意，这可能会 $\hat{}\,\hat{}$ 导致额外的零列，在 L 中，这使得可以将 R_b 中更多元素置零——我们在此不会利用这一点，但请注意，对于某些 k 和 T 的取值，这可以带来额外的（低阶）归约。

推论 9.13。令 $3 \le b \le T$， $1 \le k \le T$， $h = \lceil \log_b T \rceil$ 且 $\tilde{k} = \min(k,\ b^{h-1})$。则

$$\left[\tilde{k}(\lceil \log_b T \rceil - \lceil \log_b \tilde{k} \rceil - 1/b)\right]^{1/p} \le \mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{1,k}) \le \left[k(\lceil \log_b (T/k) \rceil + 2)\right]^{1/p}.$$

此外，若 $k \le T/b$，则可得到以下简化的下界

$$\mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{1,k}) \ge \left[k(\lfloor \log_b (T/k) \rfloor - 1/b)\right]^{1/p}\,.$$

证明。 对于上界，注意到考虑 $b^h \ge T$ 个叶节点上的完整树时，敏感度不会降低，因此可以直接应用定理 9.5 中的上界。于是

$$\mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{1,k})^p \le k(\lceil \log_b T \rceil - \lceil \log_b k \rceil + 1) + b^{\lceil \log_b k \rceil - 1} \le k(\lceil \log_b (T/k) \rceil + 2)$$

其中，最后一个不等式利用了对于任意 $a, b \ge 0$， $\lceil a \rceil - \lceil b \rceil \le \lceil a - b \rceil$ 和 $\lceil a \rceil \le a + 1$。对于下界，我们考虑 $b^{h-1} < T$ 个叶节点上的完整树，并令其对应的（约化）矩阵为 $\hat{}\, R'_b \in \mathbb{R}^{m \times b^{h-1}}$，其中某个 $m \ge 1$。定义 $\tilde{k} = \min(k, b^{h-1})$。在较小的树上，并在受限的 k 范围内进行范数最大化，必然会得到更小的范数，因此

$$\mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{1,k})^p \ge \mathsf{sens}_p(\hat{R}'_b, \mathcal{S}_{1,\tilde{k}}) \ge \tilde{k}(\lceil \log_b T \rceil - \lceil \log_b \tilde{k} \rceil - 1/b)$$

这里我们应用定理 9.5 中的下界，并选取最大化中的第一项。这个下界也严格为正。对于更受限的情形，即 $k \le T/b$，可得 $k = k$，并且下界还可以进一步简化为

$$\mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{1,k})^p \ge k(\lceil \log_b T \rceil - \lceil \log_b k \rceil - 1/b) \ge k(\lfloor \log_b (T/k) \rfloor - 1/b)$$

其中我们使用了 $\lceil a \rceil - \lceil b \rceil \ge \lfloor a - b \rfloor$。对于这种情况，我们有 $\log_b(T/k) \ge 1$，因此可以保证下界为正数。对所有（正的）上界取 p 次根，即可得到最终结论。 □

只要 $k \le T/b$（实用范围），我们就已证明 $\mathsf{sens}_p(\hat{}\, R_b,\ \mathcal{S}_{1,\ k})^p = \Theta(k \log(T/k))$。我们稍后会看到，对于 $k = \Omega(T)$，因式分解有更好的选择。

9.2.3 扩展至$\mathcal{S}_{D,\,k}$。

现在我们可以给出基于树的因式分解在 $\mathcal{S}_{D,\,k}$ 上敏感度的最终结果。为方便起见，我们只陈述 $k \le T/b$ 的情形，此时下界的形式（相对）简洁。我们还假设 $D \le k$，否则 $\mathcal{S}_{D,\,k}$ 会退化为仅包含范数有界的向量的集合。

定理 9.14。设 $3 \le b \le T$ 且 $1 \le D \le k \le T/b$ 那么

$$D\left\lceil\left\lfloor\frac{k}{D}\right\rfloor\left(\left\lfloor\log_b\left(\frac{DT}{k}\right)\right\rfloor-\frac{1}{b}\right)\right\rceil^{1/p} \le \mathsf{sens}_p(\hat{R}_b,\mathcal{S}_{D,k}) \le D\left\lceil\left\lceil\frac{k}{D}\right\rceil\left(\left\lceil\log_b\left(\frac{DT}{k}\right)\right\rceil+2\right)\right\rceil^{1/p}.$$

证明。 对于上界，我们首先说明上界

$$\mathsf{sens}_p(\hat{R}_b,\mathcal{S}_{1,k}) \le U(k) = \lceil k(\lceil\log_b(T/k)+2)\rceil^{1/p}$$

定理 9.13 中的上界是 k 在 $[1,T]$上的凹、非递减函数。由于 $g(x) = x^{1/p}$ 对于 $p \ge 1$ 是凹且非递减的，只需说明 $f(x) = x(\lceil\log_b(x/k)\rceil+2)$是凹函数，即可证明$U = g\circ f$的凹性。我们通过非递增的前向差分 $\Delta f(x) = f(x+1) - f(x)$来证明:

$$\begin{aligned}
\Delta f(x) &\le (x+1)\left(\lceil\log_b(T/x)\rceil+2\right) - x\left(\lceil\log_b(T/x)\rceil+2\right) \le \lceil\log_b T/x\rceil + 2\,,\\
\Delta f(x) &\ge (x+1)\left(\left\lceil\log_b\left(\frac{T}{x+1}\right)\right\rceil+2\right) - x\left(\left\lceil\log_b\left(\frac{T}{x+1}\right)\right\rceil+2\right) = \left\lceil\log_b\left(\frac{T}{x+1}\right)\right\rceil+2\,,\\
\Delta^2 f(x) &= \Delta f(x+1) - \Delta f(x) \le \left\lceil\log_b\left(\frac{T}{x+1}\right)\right\rceil+2-\left(\left\lceil\log_b\left(\frac{T}{x+1}\right)\right\rceil+2\right) = 0\,.
\end{aligned}$$

因此，我们可以应用定理 7.3，得到

$$\mathsf{sens}_p(\hat{R}_b,\mathcal{S}_{D,k}) \le D\cdot U(\lceil k/D\rceil) \le D\left\lceil\left\lceil\frac{k}{D}\right\rceil\left(\left\lceil\log_b\left(\frac{T}{\lceil k/D\rceil}\right)\right\rceil+2\right)\right\rceil^{1/p},$$

其中，定理陈述是应用 $T/\lceil k/D\rceil \le DT/k$后的结果。对于下界，我们直接应用定理 7.1，得到

$$\mathsf{sens}_p(\hat{R}_b,\mathcal{S}_{D,k}) \ge D\cdot\mathsf{sens}_p(\hat{R}_b,\mathcal{S}_{1,\lfloor k/D\rfloor}) \ge D\left\lceil\left\lfloor\frac{k}{D}\right\rfloor\left(\left\lfloor\log_b\left(\frac{T}{\lfloor k/D\rfloor}\right)\right\rfloor-1/b\right)\right\rceil^{1/p},$$

类似地，定理结论由 $T/\lfloor k/D\rfloor \ge DT/k$ 得出。 □

关于定理 9.14，可以作几点说明。首先，注意到，对于 ℓ_1 和 ℓ_2 的敏感度，上界是紧的：

$$\begin{aligned}
\mathsf{sens}_1(\hat{R}_b,\mathcal{S}_{D,k}) &= \Theta\left(k\log\left(DT/k\right)\right),\\
\mathsf{sens}_2(\hat{R}_b,\mathcal{S}_{D,k}) &= \Theta\left(\sqrt{Dk\log\left(DT/k\right)}\right).
\end{aligned}$$

它们还 严格推广了定理 9.13，也就是说，它们恢复了 $D = 1$ 的上界 .

9.3 误差界

完成敏感度的讨论后，我们终于可以推导误差保证了。我们分别给出 ρ-零集中差分隐私和 ε-差分隐私下的上界。我们仅陈述采用减法的 b 叉树情形，因为与仅使用加法相比，它能降低主导常数。请注意，删除对 k/D 的条件也可以，但陈述会更加繁琐。

定理 9.15 (ρ-零集中差分隐私树)。 设 $b \geq 3$ 为奇数， $T \geq b^{2b}$、 $1 \leq D \leq k \leq T/b$ 和 $\rho > 0$。另假设 $k/D = o(T)$ 以及 $k/D = \omega(1)$。那么，用于对 $\mathcal{S}_D$ 上的 k-数据流进行计数、并产生 T 个输出的 ρ-零集中差分隐私 b 叉树减法机制，具有以下渐近误差界:

$$\mathsf{MaxSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) = \Theta\left(\sqrt{Dk\log(DT/k)\log(T)/\rho}\right),$$
$$\mathsf{MeanSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) = \Theta\left(\sqrt{Dk\log(DT/k)\log(T)/\rho}\right).$$

在较弱的限制条件 $T \geq b$ 和 $k \leq T$下，以下上界同样成立 :

$$\mathsf{MaxSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) \leq \left(\frac{\sqrt{b-1}}{\sqrt{2}\log(b)} + o(1)\right) \cdot \sqrt{\frac{Dk\log(DT/k)\log(T)}{2\rho}},$$
$$\mathsf{MeanSE}(\hat{L}_b, \hat{R}_b, \mathcal{S}_{D,k}) \leq \left(\frac{\sqrt{b(1-1/b^2)}}{2\log(b)} + o(1)\right) \cdot \sqrt{\frac{Dk\log(DT/k)\log(T)}{2\rho}},$$

其中， $b = 5$ 和 $b = 7$ 的首要常数分别取最小值，最小值分别为 0.609 和0.466 。对于常数 b， $o(1)$ 项在 $T \to \infty$ 时趋于零。对于常数 b，其空间复杂度为 $O(\log T)$ ，时间复杂度为 $O(T)$ 。

证明。 渐近误差界可直接由定理 9.5 和定理 9.14 得出。对于更精细的上界，我们首先对定理 9.14 中的敏感度给出上界，但首先注意到，这里不施加 $k \leq T/b$ ，因为该条件仅用于该定理的下界。我们有

$$\mathsf{sens}_p(\hat{R}_b, \mathcal{S}_{D,k}) \leq D \cdot \sqrt{\left\lceil \frac{k}{D} \right\rceil \left(\left\lceil \log_b\left(\frac{DT}{k}\right)\right\rceil + 2\right)} \leq (1+o(1)) \cdot \sqrt{Dk\log_b\left(\frac{DT}{k}\right)}$$

其中，第二个不等式对常数 $a \geq 0$和 $x = \omega(1)$ 两次使用了 $\sqrt{\lceil x\rceil + a} \leq \sqrt{x+a} + 1 \leq \sqrt{x} + \sqrt{a} \mp 1 = (1+o(1))\sqrt{x}$ 。一次用于 $x = \log_b(DT/k)$ 和 $a = 2$，另一次用于 $x = k/D$、 $a = 0$，并结合 $(1+o(1))^2 = 1 + o(1)$。对于左因子的范数（定理 9.5），通过类似的推导可得

$$\|\hat{L}_b\|_{2\to\infty} \leq \sqrt{\frac{(b-1)\lceil \log_b(T)\rceil + 2}{2}} \leq (1+o(1)) \cdot \sqrt{\frac{(b-1)\log_b(T)}{2}},$$
$$\frac{\|\hat{L}_b\|_F}{\sqrt{T}} \leq \frac{\sqrt{b(1-1/b^2)\lceil \log_b(T)\rceil + 2(1+b^2+b^{-\lceil \log_b T\rceil})}}{2} \leq (1+o(1)) \cdot \frac{\sqrt{b(1-1/b^2)\log_b(T)}}{2}.$$

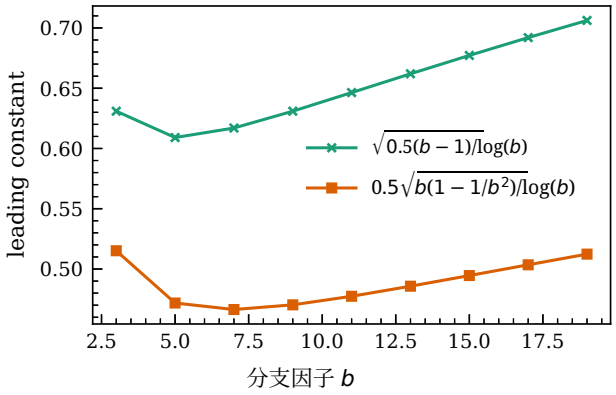
要得到定理陈述中更精细的上界，将相应的范数上界相乘，除以 $\sqrt{2}\rho$ ，并提取主导常数。主导常数相对于 b 的变化绘于图 4a 中。关于时间和空间用量的陈述已在 [4] 中证明。

□

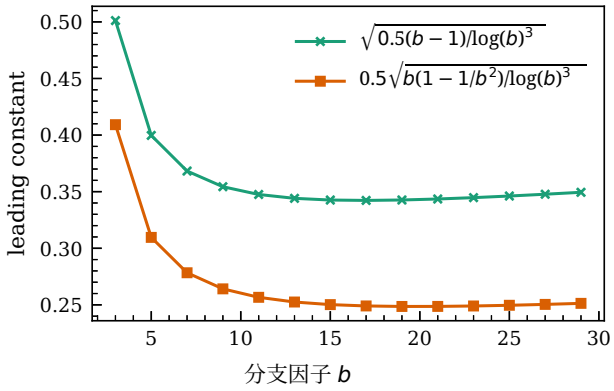
ˆ 要得到纯差分隐私的结果，我们只需改为发布 $R_b\vec{\mathbf{x}}$ ，并使用拉普拉斯机制，这样添加的噪声尺度为 $\sqrt{2}\cdot\mathsf{sens}_1(\hat{R}_b, \mathcal{S}_{D,k})/\varepsilon$，而非 $\mathsf{sens}_2(\hat{R}_b, \mathcal{S}_{D,k})/\sqrt{2\rho}$。在这种情况下，我们仅给出更精细的上界，但也可以为此情形推导出类似的紧渐近界。

定理 9.16 （ ε-差分隐私的树）。 设 $b \geq 3$ 为奇数， $T \geq b$、 $1 \leq D \leq k \leq T$和 $\varepsilon > 0$。另外假设 $k/D = o(T)$ 以及 $k/D = \omega(1)$ 。然后，使用减法进行计数、在 $\mathcal{S}_{D,\ k}$-数据流上生成 T 个输出的 ε-差分隐私 b元树机制可达到以下误差界:

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) \leq \left(\sqrt{\frac{b-1}{2\log(b)^3}} + o(1)\right) \cdot \frac{\sqrt{2\log(T)}k\log(DT/k)}{\varepsilon},$$
$$\mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) \leq \left(\sqrt{\frac{b(1-1/b^2)}{4\log(b)^3}} + o(1)\right) \cdot \frac{\sqrt{2\log(T)}k\log(DT/k)}{\varepsilon}.$$



(a) ρ-零集中差分隐私相对于 b 的主导常数。



(b) ε-差分隐私相对于 b 的主导常数。

图 4： ρ-零集中差分隐私（定理 9.15）和 ε-差分隐私（定理 9.16）关于 MaxSE($\hat{L}_b$, $\hat{R}_b$, $\mathcal{S}_{1,\ k}$) 和 MeanSE($\hat{L}$, $\hat{R}_b$, $\mathcal{S}_{1,\ k}$) 的上界的主导常数，按奇数 b 绘制。在图 4a 中，它们分别在 $b = 5$ 和 $b = 7$ 处达到最小值，最小值分别为 0.609 和 0.466。在图 4b 中，它们分别在 $b = 17$ 和 $b = 19$ 处达到最小值，最小值分别为 0.342 和 0.249。

其中， $b = 17$ 和 $b = 19$ 对应的主导常数分别达到最小值，相应的主导常数为 0.342 和 0.249。当 b为常数时， o（1）项随 $T \to \infty$ 趋于无穷而趋于零。对于常数 b，其空间复杂度为 O（log T），时间复杂度为 O（T）。

证明。 证明过程与定理 9.15 的证明完全相同，不同之处在于依赖于 ℓ_1 敏感度。我们调用定理 9.14（同样，不强制要求 $k \le T/b$，因为它只影响该定理的下界）来给敏感度取上界：

$$\mathsf{sens}_1(\hat{R}_b, \mathcal{S}_{D,k}) \le D \cdot \left\lceil \frac{k}{D} \right\rceil \left(\left\lceil \log_b \left(\frac{DT}{k} \right) \right\rceil + 2 \right) \le (1 + o(1)) \cdot k \log_b \left(\frac{DT}{k} \right)$$

其中，第二个不等式对常数 $a \ge 0$ 和 $x = \omega(1)$ 两次使用了 $\lceil x \rceil + a \le x + a + 1 \le (1 + o(1))x$。一次用于 $x = \log_b(DT/k)$ 和 $a = 2$，另一次用于 $x = k/D$、 $a = 0$，并结合 $(1 + o(1))^2 = 1 + o(1)$。与 ρ-zCDP 情形相同，我们根据定理 9.5 对左因子的范数进行上界估计，得到

$$\|\hat{L}_b\|_{2\to\infty} \le (1 + o(1)) \cdot \sqrt{\frac{(b-1)\log_b(T)}{2}}, \qquad \frac{\|\hat{L}_b\|_F}{\sqrt{T}} \le (1 + o(1)) \cdot \frac{\sqrt{b(1-1/b^2)\log_b(T)}}{2}.$$

对于最终的定理陈述，将相应的上界相乘，再乘以 $\sqrt{2/\varepsilon}$，并提取主导常数，即可得到最终表达式。关于时间和空间用量的陈述同样在 [4] 中得到证明。主导常数同样是 b 的凸函数，并绘于图 4b 中。 □

在这两种情形下，主导常数的优化均独立于 k 和 D。对于 ρ-zCDP，基于树的分解可实现误差缩放 $O\left(\sqrt{Dk}\log(DT/k)\log(T)\right)$。我们将其与第 11 节中所考虑的其他分解的渐近行为进行比较。

10 朴素分解

鉴于基于树的因式分解在 k 较大时能够得到更优的上界，一个自然的问题是：在这种情况下，基于树的因式分解是否是正确的构造。我们证明，答案并不简单：当 k 足够大时，朴素分解 $A = I \times A$ 的表现优于它。

引理 10.1。设1 $\leq D \leq k \leq T$、 $p \geq 1$和 $A \in \mathbb{R}^{T\times T}$为全 *1* 下三角矩阵。 n

$$\mathsf{sens}_p(A, \mathcal{S}_{D,k}) = D \cdot T^{1/p}\,.$$

证明。注意，对于任意 $\vec{\Delta} \in \mathcal{S}_{D,\ k}$，根据定义有 $A\vec{\Delta} \in [-D, D]^T$。由于在 , 中始终存在向量 $\vec{\Delta}'$，其 $\vec{\Delta}'[0] = D$ $\mathcal{S}_{D\ k}$ $_p$ A $\mathcal{S}_{D\ k}$ $= \|A\vec{\Delta}'\|_p =$ 且其他每个条目均为零，因此我们 $DT^{1/p}$ 立即得到 sens（，，）。 □

定理 10.2。 对于 $1 \leq D \leq k \leq T$ 和 $\rho > 0$，考虑在 $\mathcal{S}_{D,\ k}$-数据流上进行计数的 ρ-零集中差分隐私分解机制，其中 $L = I$ 且 $R = A$。该机制达到的误差为

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{D,k}) = \mathsf{MeanSE}(L, R, \mathcal{S}_{D,k}) = \frac{D\sqrt{T}}{\sqrt{2\rho}}\,.$$

类似地，给定 $\varepsilon > 0$， ε-差分隐私分解机制（使用拉普拉斯机制发布 $R\vec{\ }\mathbf{x}$ ）可达到以下误差：

$$\mathsf{MaxSE}(L, R, \mathcal{S}_{k,D}) = \mathsf{MeanSE}(L, R, \mathcal{S}_{k,D}) = \frac{\sqrt{2}DT}{\varepsilon}\,,$$

两个版本的运行时间均为 $O(T)$ ，空间用量为 $O(1)$ 。

证明。 显然有 $\|I\|_{2\to\infty} = \|I\|_F/\sqrt{T} = 1$，因此关于误差的论述只需对 ℓ_2 和 ℓ_1 敏感度应用定理 10.1。关于时间和空间用量的论述：该机制在每个时间步都会加入新的独立噪声，每步只需常数时间即可完成，且无需跨时间步存储噪声。 □

聚焦于 ρ-零集中差分隐私设置，结论显而易见。通过使用高斯机制单独发布每个前缀，并为组合付出代价，始终可以达到 $\Theta(\min(D, k)\sqrt{T})$ 的误差。我们将在第 11 节中将其与其他分解的误差界渐近行为进行比较。

11 上界的解析比较

在此，我们对论文中的上界进行解析比较，并讨论相关结果。我们重点关注敏感度向量集 $\mathcal{S}_{1,k}$ 对应的结果（即 DegreeCount 和不同元素计数的结果）。

在这一分析性比较中，对于本文所述的平方根分解机制和基于树的机制，我们将采用各自上界中的主导项；对于每个时间步添加高斯噪声并承担组合代价的朴素机制，则采用其精确上界。与引言中一样，我们将考虑 $\rho = 1/2$ （由于所有上界对 ρ 的依赖都相同，因此不影响比较）。

相应的上界为： $c_1\sqrt{k}\log T$（对于平方根分解机制），其中 c_1 是可由定理 8.1 中的上界推导出的常数； $c_2\sqrt{k}\log(T/k)\log\ T$（对于基于树的分解机制），其中在 b 取最优值时， c_2 可由定理 9.15 中的上界推导得出；以及 $\sqrt{T}$，这是朴素机制的精确上界（见定理 10.2）。注意， $c_2 > c_1$。

比较这些上界后，我们可以得出以下结论。注意，当 $k \leq {}_{c_1^2}{}^{T}{}_{\log^2 T}$ 时，平方根机制优于朴素界。类似地，当 $k \leq T^{1-(c_1/c_2)^2}$ 时，平方根分解的上界优于基于树的机制的上界；否则，基于树的机制的上界更优。最后请注意，只要 $k\log(T/k) \leq {}_{c_2^2}{}^{T}{}_{\log T}$，基于树的机制的上界就优于朴素机制的上界。令等式成立时 k 的取值为 k^*。

注意到，这些上界意味着，对于 $k \in [1, T^{1-(c_1/c_2)^2}]$,，基于平方根分解的上界占优；对于 $k \in [T^{1-(c_1/c_2)^2},\ k^*]$（其中 $k^* >$ ${}^{T}{}_{\log^2 T}$），基于树的机制的上界占优；

而对于 $k \in [k^*, T]$，基于朴素组合机制的上界占优。我们预计，在大多数实际应用中，k 会远小于 T，因此，对于大多数具有实际意义的 k值，我们的上界都能带来改进。不同 k 对应的最优界的精确刻画留待未来工作。

12 结论

我们的工作为一系列基数估计问题给出了更优的误差界。为此，我们将全动态流上的私有基数估计重新表述为针对其差分流的私有线性查询，并详细研究了相关的敏感度向量集。未来还有许多值得探索的方向。

适用于所有图的隐私三角形计数。 如前所述，将三角形计数简化为在差分流上进行持续计数所提供的隐私保障，以图数据流的度数上界为 D 且最大三角形贡献上界为 k_{tri} 为前提。如何设计一种能够保障所有图的隐私，同时对满足参数约束的数据流仍能达到与我们的方法相当的准确度的算法，是一个自然的开放问题。

其他线性查询。 虽然本文仅研究了应用于差分流的持续计数，但探索这些数据流上的其他线性查询或许也很有意义。例如，我们可能不想精确计算不同元素的数量，而只想知道不同元素的数量最近是在增加还是减少。这种统计量可以通过对运行计数进行后处理自然地得出，但针对目标统计量直接调整查询矩阵或许会更高效。通过我们的框架研究此类查询矩阵，是未来值得探索的一个自然方向。

空间效率。 有证据表明，我们采用的平方根分解不允许任何次线性空间算法 [5]。近期研究 [21] 提出了一种托普利兹分解，可在空间 $O(\log^2 T)$ 内将其误差近似到一个加性常数范围内，但其构造并未对右因子的对角线施加单调性条件。对于标准持续计数，设计一种满足此单调性约束且误差近乎最优的空间高效托普利兹分解仍是一个未解问题；若能解决这一问题，结合第8节的结果，便可得到 $\mathcal{S}_{D,k}$-流上持续计数的误差界。

下界。 最后，我们提出一个问题：如何为 $\mathcal{S}_{D,k}$-流上的持续计数，尤其是分解机制，建立一般下界。我们已经知道，对于 $\mathcal{S}_{1,1}$，平方根分解在我们的误差度量 [51, 30, 40], 下具有最优的主导常数，但显然它对于 $\mathcal{S}_{1,T}$ 并非最优（见定理 10.2 ）。因此，不同的分解预计会在不同的参数范围内达到最优。

致谢

我们感谢 Nikita Kalinin 和 Adam Smith 对本文早期草稿提出的宝贵意见，也感谢 Sofya Raskhodnikova 就 [55] 的结果进行的宝贵讨论。

Joel Daniel Andersson 的研究经费由欧盟提供。具体而言，他获得了欧洲研究委员会 (ERC) 通过欧盟地平线2020研究与创新计划提供的资助（MoDynStruct，编号 101019564），此外还获得了 Providentia 的资助，以及诺和诺德基金会提供的数据科学杰出研究者资助。部分工作是在访问波士顿大学期间完成的。

所表达的观点和意见仅代表作者本人，并不一定反映任何赞助方的立场。这包括欧盟和欧洲研究理事会执行机构。

参考文献

[1] Aditya Akella、Ashwin R. Bharambe、Michael K. Reiter 和 Srinivasan Seshan。《检测 ISP 网络中的 DDoS 攻击》。载于：数据流管理与处理研讨会。2003年。[2] Joel Daniel Andersson 和 Rasmus Pagh。《用于高效私有持续观测的平滑二元机制》。载于：神经信息处理系统进展第*36*卷：神经信息处理系统年度会议 *2023*（*NeurIPS 2023*），美国路易斯安那州新奥尔良，*2023*年*12*月*10*日至*16*日。2023年。网址：http : / / papers . nips . cc / paper % 5C _ files / paper / 2023 / hash / 99c41fb9fd53abfdd4a0259560ef1c9d-Abstract-Conference.html。[3] Joel Daniel Andersson 和 Rasmus Pagh。《通过分箱实现流式隐私持续计数》。载于：*IEEE*安全可信机器学习会议，*SaTML 2025*，丹麦哥本哈根，*2025*年*4*月*9*日至*11*日。IEEE，2025年，第575–589页。doi：10.1109/SATML64287.2025.00038。网址：https: //doi.org/10.1109/SaTML64287.2025.00038。[4] Joel Daniel Andersson、Rasmus Pagh、Teresa Anna Steiner 和 Sahel Torkamani。《相信年长者：拉普拉斯噪声与高斯噪声》。载于：第六届负责任计算基础研讨会（*FORC 2025*），*2025*年*6*月*4*日至*6*日，美国加利福尼亚州斯坦福大学。第329卷。LIPIcs。Schloss Dagstuhl - Leibniz 信息中心，2025年，10:1–10:24。doi：10 . 4230 / LIPICS . FORC . 2025 . 10。网址：https://doi.org/10.4230/LIPIcs.FORC.2025.10。[5] Joel Daniel Andersson 和 Amir Yehudayoff。《在线卷积的空间复杂度》。载于：*CoRR*abs/2505.00181（2025年）。doi：10 . 48550 /ARXIV . 2505 . 00181。arXiv：2505 .00181。网址：https://doi.org/10.48550/arXiv.2505.00181。[6] Daniel N. Baker 和 Ben Langmead。《Dashing：使用 HyperLogLog 快速、准确地计算基因组距离》。载于：*Genome Biology* 20（2018年）。[7] Luca Becchetti、Paolo Boldi、Carlos Castillo 和 Aristides Gionis。《用于大规模图局部三角形计数的高效半流式算法》。载于：第*14*届 *ACM SIGKDD*知识发现与数据挖掘国际会议论文集，美国内华达州拉斯维加斯，*2008*年*8*月*24*日至*27*日。Ying Li、Bing Liu 和 Sunita Sarawagi 编。ACM，2008年，第16–24页。doi：10.1145/1401890.1401898。网址：https://doi.org/10.1145/1401890.1401898。[8] G. Bennett。《舒尔乘子》。载于：*Duke Mathematical Journal* 44.3（1977年），第603–639页。doi：10. 1215/S0012-7094-77-04426-X。网址：https://doi.org/10.1215/S0012-7094-77-04426-X。[9] Jean Bolot、Nadia Fawaz、S. Muthukrishnan、Aleksandar Nikolov 和 Nina Taft。《流上的隐私衰减谓词和》。载于：*2013*年*EDBT/ICDT*联合会议，*ICDT '13* 论文集，意大利热那亚，*2013*年*3*月*18*日至*22*日。Wang-Chiew Tan、Giovanna Guerrini、Barbara Catania 和 Anastasios Gounaris 编。ACM，2013年，第284–295页。doi：10.1145/2448496.2448530。网址：https://doi.org/10. 1145/2448496.2448530。[10] Mark Bun 和 Thomas Steinke。《集中差分隐私：简化、扩展与下界》。载于：密码学理论。Martin Hirt 和 Adam Smith 编。柏林、海德堡：Springer Berlin Heidelberg，2016年，第635–658页。ISBN：978-3-662-53641-4。[11] Adrian Rivera Cardoso 和 Ryan Rogers。《持续观测下的差分隐私直方图：面向未知的流式选择》。载于：人工智能与统计国际会议，*AISTATS 2022*，*2022*年*3*月*28*日至*30*日，线上活动。Gustau Camps-Valls、Francisco J. R. Ruiz 和 Isabel Valera 编。第151卷。机器学习研究论文集。PMLR，2022年，第2397–2419页。网址：https://proceedings.mlr.press/v151/rivera-cardoso22a.html。[12] Adrian Rivera Cardoso 和 Ryan Rogers。《持续观测下的差分隐私直方图：面向未知的流式选择》。载于：人工智能与统计国际会议，*AISTATS 2022*，*2022*年*3*月*28*日至*30*日，线上活动。Gustau Camps-Valls、Francisco J. R. Ruiz 和 Isabel Valera 编。第151卷。机器学习研究论文集。PMLR，2022年，第2397–2419页。网址：https://proceedings.mlr.press/v151/rivera-cardoso22a.html。

[13] T.-H. Hubert Chan、Elaine Shi 和 Dawn Song。“统计数据的私密持续发布”。发表于：*ACM* 信息与系统安全汇刊 14.3（2011年11月），26:1–26:24。ISSN：1094-9224。DOI：10.1145/2043621.2043626。网址：https://doi.org/10.1145/2043621.2043626（访问于2023年1月26日）。[14] Christina Chandra、Martina Morris、Connor Van Meter、Steven M. Goodreau、Travis Sanchez、Patrick Janulis、Michelle Birkett 和 Samuel M. Jenness。“男男性行为者性网络平均活跃度测量指标的比较”。发表于：$medRxiv$（2022年）。DOI：10.1101/2022. 02.11.22270855。预印本：https://www.medrxiv.org/content/early/2022/05/16/2022.02.11.22270855.full.pdf。网址：https://www.medrxiv.org/content/early/2022/05/16/2022.02.11. 22270855。[15] Christopher A. Choquette-Choo、Arun Ganesh、Ryan McKenna、H. Brendan McMahan、John Rush、Abhradeep Guha Thakurta 和 Zheng Xu。“（放大）带状矩阵分解：隐私训练的统一方法”。发表于：神经信息处理系统进展 *36*：神经信息处理系统年会 *2023*，*NeurIPS 2023*，美国路易斯安那州新奥尔良，*2023*年*12*月*10*日至*16*日。Alice Oh、Tristan Naumann、Amir Globerson、Kate Saenko、Moritz Hardt 和 Sergey Levine 编。2023年。网址：http://papers.nips.cc/paper%5C_files/paper/2023/hash/ecc28b4ce9b39f5f23c3efb03e25b7bf-Abstract-Conference.html。[16] Christopher A. Choquette-Choo、Hugh Brendan McMahan、J. Keith Rush 和 Abhradeep Guha Thakurta。“用于隐私保护机器学习的多轮矩阵分解机制”。发表于：国际机器学习大会，*ICML 2023*，美国夏威夷州檀香山，*2023*年*7*月*23*日至*29*日。Andreas Krause、Emma Brunskill、Kyunghyun Cho、Barbara Engelhardt、Sivan Sabato 和 Jonathan Scarlett 编。第202卷。《机器学习研究论文集》。PMLR，2023年，第5924–5963页。网址：https://proceedings.mlr.press/v202/choquette-choo23a.html。[17] Graham Cormode、Tejas Kulkarni 和 Divesh Srivastava。“在局部差分隐私下回答范围查询”。发表于：*VLDB*基金会会刊 12.10（2019年6月），第1126–1138页。ISSN：2150-8097。DOI：10.14778/3339490.3339496。网址：https://doi.org/10.14778/3339490.3339496。[18] Rachel Cummings、Alessandro Epasto、Jieming Mao、Tamalika Mukherjee、Tingting Ou 和 Peilin Zhong。“旋转门模型中不同元素计数的差分隐私空间高效算法”。发表于：$CoRR$ abs/2505.23682（2025年）。DOI：10.48550/ARXIV.2505.23682。arXiv：2505.23682。网址：https://doi.org/10.48550/arXiv.2505.23682。[19] Anirban Dasgupta、Ravi Kumar 和 Tamás Sarlós。“关于平均度的估计”。发表于：第*23*届国际万维网大会，*WWW '14*，韩国首尔，*2014*年*4*月*7*日至*11*日。Chin-Wan Chung、Andrei Z. Broder、Kyuseok Shim 和 Torsten Suel 编。ACM，2014年，第795–806页。DOI：10.1145/2566486.2568019。网址：https://doi.org/10.1145/2566486.2568019。[20] Sergey Denisov、H. Brendan McMahan、John Rush、Adam D. Smith 和 Abhradeep Guha Thakurta。“通过自适应数据流上的最优隐私线性算子改进随机梯度下降的差分隐私”。发表于：$NeurIPS$。2022年。网址：http://papers.nips.cc/paper%5C_files/paper/2022/hash/271ec4d1a9ff5e6b81a6e21d38b1ba

Abstract-Conference.html.[21] Krishnamurthy Dvijotham、H. Brendan McMahan、Krishna Pillutla、Thomas Steinke 和 Abhradeep Thakurta。“流式差分隐私的高效近最优噪声生成”。载于：$CoRR$ abs/2404.16706（2024）。doi: 10.48550/ARXIV.2404.16706。arXiv: 2404.16706。url: https://doi.org/10.48550/arXiv.2404.16706。[22] Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith。“私有数据分析中的噪声敏感度校准”。载于：第三届密码学理论会议论文集。TCC'06。纽约州纽约市：Springer-Verlag，2006年，第265–284页。isbn: 3540327312。doi: 10.1007/11681878_14。url: https://doi.org/10.1007/11681878_14。

[23] Cynthia Dwork、Moni Naor、Toniann Pitassi 和 Guy N. Rothblum。“持续观测下的差分隐私”。载于：第四十二届 *ACM* 计算理论研讨会论文集。STOC ’10。美国纽约州纽约市：计算机协会，2010年6月，第715–724页。isbn: 978-1-4503- 0050-6。doi: 10.1145/1806689.1806787。url: https://doi.org/10.1145/1806689.1806787（访问日期：2023年1月13日）。[24] Cynthia Dwork 和 Aaron Roth。“差分隐私的算法基础”。en。载于：基础与趋势® 理论计算机科学领域9.3-4（2013），第211–407页。issn: 1551-305X, 1551-3068。doi: 10.1561/0400000042。[25] Jean-Pierre Eckmann 和 Elisha Moses。“共链的曲率揭示了万维网上隐藏的主题层”。载于：美国国家科学院院刊99.9（2002），第5825–5829页。doi: 10.1073/pnas.032093399。eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.032093399。url: https://www.pnas.org/doi/abs/10.1073/pnas.032093399。[26] Alessandro Epasto、Quanquan C. Liu、Tamalika Mukherjee 和 Felix Zhou。“持续发布模型中图稀疏化的力量”。载于：$CoRR$ abs/2407.17619（2024）。doi: 10.48550/ ARXIV.2407.17619。arXiv: 2407.17619。url: https://doi.org/10.48550/arXiv.2407.17619。[27] Alessandro Epasto、Jieming Mao、Andres Muñoz Medina、Vahab Mirrokni、Sergei Vassilvitskii 和 Peilin Zhong。“流式频率矩估计的差分隐私持续发布”。载于：第*14*届理论计算机科学创新会议（*ITCS 2023*），*2023*年*1*月*10*日至*13*日，美国马萨诸塞州剑桥市麻省理工学院。Yael Tauman Kalai 编。第251卷。LIPIcs。Schloss Dagstuhl - Leibniz 信息中心，2023年，48:1–48:24。doi: 10 . 4230 / LIPICS . ITCS . 2023 . 48。url: https://doi.org/10.4230/LIPIcs.ITCS.2023.48。[28] Alessandro Epasto、Tamalika Mukherjee 和 Peilin Zhong。“数据流中的差分隐私聚类”。载于：$CoRR$ abs/2307.07449（2023）。doi: 10.48550/ARXIV.2307.07449。arXiv: 2307.07449。url: https://doi.org/10.48550/arXiv.2307.07449。[29] Hendrik Fichtenberger、Monika Henzinger 和 Lara Ost。“持续观测下图的差分隐私算法”。载于：第*29*届欧洲算法研讨会（*ESA 2021*），*2021*年*9*月*6*日至*8*日，葡萄牙里斯本（线上会议）。Petra Mutzel、Rasmus Pagh 和 Grzegorz Herman 编。第204卷。LIPIcs。Schloss Dagstuhl - Leibniz 信息中心，2021年，42:1–42:16。doi: 10.4230/LIPICS.ESA.2021.42。url: https://doi.org/10.4230/LIPIcs.ESA.2021.42。[30] Hendrik Fichtenberger、Monika Henzinger 和 Jalaj Upadhyay。“常数很重要：差分隐私持续观测的精细误差界”。载于：第*40*届国际机器学习会议论文集。Andreas Krause、Emma Brunskill、Kyunghyun Cho、Barbara Engel- hardt、Sivan Sabato 和 Jonathan Scarlett 编。第202卷。机器学习研究论文集。PMLR，2023年7月，第10072–10092页。url: https://proceedings.mlr.press/v202/fichtenberger23a.html。[31] Badih Ghazi、Ravi Kumar、Jelani Nelson 和 Pasin Manurangsi。“时间窗口内不同项目及出现k次项目的私有计数”。载于：第*14*届理论计算机科学创新会议（*ITCS 2023*），*2023*年*1*月*10*日至*13*日，美国马萨诸塞州剑桥市麻省理工学院。Yael Tauman Kalai 编。第251卷。LIPIcs。Schloss Dagstuhl - Leibniz 信息中心，2023年，55:1–55:24。doi: 10.4230/LIPICS.ITCS.2023.55。url: https://doi.org/10.4230/LIPICS.ITCS.2023.55。[32] Google Research。借助 *JAX Privacy* 大规模实现差分隐私机器学习。https :/ / research . google / blog / differentially - private - machine - learning - at - scale - with - jax - privacy/。Google Research 博客。访问日期：2025-12-05。2025年。[33] Michael Hay、Vibhor Rastogi、Gerome Miklau 和 Dan Suciu。“通过一致性提升差分隐私直方图的准确性”。载于：*VLDB*基金会会刊 3.1–2（2010年9月），第1021–1032页。issn: 2150-8097。doi: 10.14778/1920841.1920970。url: https://doi.org/10.14778/1920841.1920970。[34] Monika Henzinger、Nikita P. Kalinin 和 Jalaj Upadhyay。“用于差分隐私持续计数的分箱群代数分解”。载于：$CoRR$ abs/2504.04398（2025年）。doi: 10.48550/ARXIV. 2504.04398。arXiv: 2504.04398。url: https://doi.org/10.48550/arXiv.2504.04398。

[35] Monika Henzinger、Nikita P. Kalinin 和 Jalaj Upadhyay。《归一化平方根：差分隐私持续计数的更精确矩阵分解界</style》。*2025*年。*arXiv*：*2509.cs.DS*。网址：*https://arxiv.org/abs/2509.14334*。*Monika Henzinger*、*A. R. Sricharan* 和 *Teresa Anna Steiner*。“直方图及相关查询的差分隐私持续发布”。载于：人工智能与统计国际会议，*AISTATS 2025*，泰国迈考，*2025*年*5*月*3*日至*5*日。Yingzhen Li、Stephan Mandt、Shipra Agrawal 和 Mohammad Emtiyaz Khan 编。第258卷。机器学习研究论文集。PMLR，2025年，第1990–1998页。网址：https://proceedings.mlr.press/v258/henzinger25a.html。[37] Monika Henzinger、A. R. Sricharan 和 Teresa Anna Steiner。“旋转门模型中不同元素的隐私计数及其扩展”。载于：近似、随机化与组合优化：算法与技术，*APPROX/RANDOM 2024*，伦敦政治经济学院，英国伦敦，*2024*年*8*月*28*日至*30*日。Amit Kumar 和 Noga Ron-Zewi 编。第317卷。LIPIcs。Schloss Dagstuhl - Leibniz 信息中心，2024年，40:1–40:21。doi：10.4230/LIPICS.APPROX/RANDOM.2024.40。网址：https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2024.40。[38] Monika Henzinger 和 Jalaj Upadhyay。“利用群代数改进差分隐私持续观测”。载于：*2025*年*ACM-SIAM*离散算法研讨会论文集，*SODA 2025*，美国路易斯安那州新奥尔良，*2025*年*1*月*12*日至*15*日。Yossi Azar 和 Debmalya Panigrahi 编。SIAM，2025年，第2951–2970页。doi：10.1137/1.9781611978322.95。网址：https://doi.org/10.1137/1.9781611978322.95。[39] Monika Henzinger、Jalaj Upadhyay 和 Sarvagya Upadhyay。“持续观测下差分隐私求和的统一框架”。载于：离散算法研讨会（*SODA*）论文集。SIAM。2024年，第995–1018页。[40] Monika Henzinger、Jalaj Upadhyay 和 Sarvagya Upadhyay。“差分隐私持续计数的近乎紧确误差界”。载于：*2023*年*ACM-SIAM*离散算法研讨会论文集，*SODA 2023*，意大利佛罗伦萨，*2023*年*1*月*22*日至*25*日。Nikhil Bansal 和 Viswanath Nagarajan 编。SIAM，2023年，第5003–5039页。doi：10.1137/1.9781611977554.CH183。网址：https://doi.org/10.1137/1.9781611977554.ch183。[41] James Honaker。“差分隐私二叉树的高效使用”。载于：差分隐私理论与实践（*TPDP 2015*），英国伦敦2（2015年），第26–27页。[42] Ben Jacobsen 和 Kassem Fawaz。“无界数据流的隐私持续计数”。载于：$CoRR$abs/2506.15018（2025年）。doi：10.48550/ARXIV.2506.15018。arXiv：2506.15018。网址：https://doi.org/10.48550/arXiv.2506.15018。[43] Palak Jain、Iden Kalemaj、Sofya Raskhodnikova、Satchit Sivakumar 和 Adam D. Smith。“持续观测下差分隐私保护的旋转门模型不同元素计数”。载于：神经信息处理系统进展第*36*卷：*2023*年神经信息处理系统年会，*NeurIPS 2023*，美国路易斯安那州新奥尔良，*2023*年*12*月*10*日至*16*日。Alice Oh、Tristan Naumann、Amir Globerson、Kate Saenko、Moritz Hardt 和 Sergey Levine 编。2023年。网址：http: / / papers .nips . cc / paper %5C _ files / paper / 2023 / hash / 0ef1afa0daa888d695dcd5e9513bafa3 - Abstract-Conference.html。[44] Palak Jain、Sofya Raskhodnikova、Satchit Sivakumar 和 Adam D. Smith。“持续观测下差分隐私的代价”。载于：国际机器学习会议，*ICML 2023*，美国夏威夷州檀香山，*2023*年*7*月*23*日至*29*日。Andreas Krause、Emma Brunskill、Kyunghyun Cho、Barbara Engelhardt、Sivan Sabato 和 Jonathan Scarlett 编。第202卷。机器学习研究论文集。PMLR，2023年，第14654–14678页。网址：https://proceedings.mlr.press/v202/jain23b.html。[45] Palak Jain、Adam D. Smith 和 Connor Wagaman。“时间感知投影：持续观测下真正的节点隐私图统计量”。载于：*IEEE*安全与隐私研讨会，*SP 2024*，美国加利福尼亚州旧金山，*2024*年*5*月*19*日至*23*日。IEEE，2024年，第127–145页。doi：10.1109/SP54263.2024.00196。网址：https://doi.org/10.1109/SP54263.2024.00196。

[46] Peter Kairouz、Brendan Mcmahan、Shuang Song、Om Thakkar、Abhradeep Thakurta 和 Zheng Xu。“无需采样或打乱的实用隐私（深度）学习”。英文。载于：第*38*届国际机器学习会议论文集。ISSN：2640-3498。PMLR，2021年7月，第5213–5225页。网址：https://proceedings.mlr.press/v139/kairouz21b.html（访问于2023年1月24日）。[47] Nikita Kalinin 和 Christoph H. Lampert。“用于差分隐私模型训练的带状平方根矩阵分解”。载于：神经信息处理系统进展第*38*卷：*2024*年神经信息处理系统年会，*NeurIPS 2024*，加拿大不列颠哥伦比亚省温哥华，*2024*年*12*月*10*日至*15*日。Amir Globersons、Lester Mackey、Danielle Belgrave、Angela Fan、Ulrich Paquet、Jakub M. Tomczak 和 Cheng Zhang 编。2024年。网址：http://papers.nips.cc/paper%5C_files/paper/ 2024/hash/1f7b3b0dd7710af02aac0db5be4cfc8d-Abstract-Conference.html。[48] Nikita P. Kalinin 和 Joel Daniel Andersson。《利用矩阵分解为隐私训练调度学习率》。2025年。arXiv：2511.17994 [cs.LG]。网址：https://arxiv.org/abs/2511.17994。[49] Nikita P. Kalinin、Ryan McKenna、Jalaj Upadhyay 和 Christoph H. Lampert。“回到平方根：多轮差分隐私随机梯度下降的矩阵分解误差最优界”。载于：*CoRR* abs/2505.12128（2025年）。doi：10.48550/ARXIV.2505.12128。arXiv：2505.12128。网址：https://doi.org/10.48550/arXiv.2505.12128。[50] Chao Li、GeromeMiklau、MichaelHay、Andrew McGregor 和 VibhorRastogi。“矩阵机制：差分隐私下优化线性计数查询”。英文。载于：*VLDB*期刊24.6（2015年12月），第757–781页。ISSN：0949-877X。doi：10.1007/s00778-015-0398-x。网址：https://doi.org/10.1007/s00778-015-0398-x（访问于2023年1月11日）。[51] Jiří Matoušek、Aleksandar Nikolov 和 KunalTalwar。“分解范数与遗传差异度”。载于：国际数学研究通讯2020.3（2020年），第751–780页。[52] H. Brendan McMahan、Zheng Xu 和 Yanxiang Zhang。“实用隐私学习的简便算法：别用树聚合，改用BLT”。载于：*CoRR* abs/2408.08868（2024年）。doi：10.48550/ARXIV.2408.08868。arXiv：2408.08868。网址：https://doi.org/10.48550/arXiv.2408.08868。[53] Krishna Pillutla、Jalaj Upadhyay、Christopher A. Choquette-Choo、Krishnamurthy Dvijotham、Arun Ganesh、Monika Henzinger、Jonathan Katz、Ryan McKenna、H. Brendan McMahan、Keith Rush、Thomas Steinke 和 Abhradeep Thakurta。“用于差分隐私学习的相关噪声机制”。载于：*CoRR* abs/2506.08201（2025年）。doi：10.48550/ARXIV.2506.08201。arXiv：2506.08201。网址：https://doi.org/10.48550/arXiv.2506.08201。[54] Wahbeh Qardaji、Weining Yang 和 Ninghui Li。“理解差分隐私直方图的层次化方法”。载于：*VLDB*基金会会刊6.14（2013年9月），第1954–1965页。ISSN：2150-8097。doi：10.14778/2556549.2556576。网址：https://doi.org/10.14778/2556549.2556576。[55] Sofya Raskhodnikova 和 Teresa Anna Steiner。“具有边差分隐私的图数据库全动态算法”。载于：*ACM*数据管理会议录 3.2（2025年），99:1–99:28。doi：10.1145/3725236。网址：https://doi.org/10.1145/3725236。[56] Shuang Song、Susan Little、Sanjay Mehta、Staal Amund Vinterbo 和 Kamalika Chaudhuri。“图统计量的差分隐私持续发布”。载于：*CoRR* abs/1809.02575（2018年）。arXiv：1809.02575。网址：http://arxiv.org/abs/1809.02575。[57] Max Dupré la Tour、Monika Henzinger 和 David Saulpic。“旧物新用：差分隐私聚类的统一算法”。载于：第四十一届国际机器学习会议，*ICML 2024*，奥地利维也纳，*2024*年*7*月*21*日至*27*日。OpenReview.net，2024年。网址：https://openreview.net/forum?id=3ajK5xplDL。

[58] Zheng Xu、Yanxiang Zhang、Galen Andrew、Christopher A. Choquette-Choo、PeterKairouz、H. Brendan McMahan、Jesse Rosenstock 和 Yuanbo Zhang。《采用差分隐私的 Gboard语言模型联邦学习》。载于：《计算语言学协会第 *61* 届年会论文集：工业赛道，*ACL 2023*》，*2023* 年 *7* 月 *9—14* 日，加拿大多伦多。Sunayana Sitaram、Beata Beigman Klebanov 和 Jason D. Williams 编。计算语言学协会，2023 年，第 629–639 页。doi：10.18653/V1/2023.ACL-INDUSTRY.60。url：https://doi.org/10.18653/v1/2023.acl-industry.60。