arXiv:2302.06165v1 [cs.DS] 13 Feb 2023

重新审视稀疏降维

Mikael Møller H øgsgaard * Lior Kamma † Kasper Green Larsen ‡ Jelani Nelson §

Chris Schwiegelshohn ¶

摘要

稀疏 Johnson–Lindenstrauss 变换是降维中的核心技术之一。它支持将集合中的 n 个点从 $\mathbb{R}^d$ 嵌入到 $m = O(\varepsilon^{-2} \lg n)$ 维度，同时将所有成对距离保持在 $1 \pm \varepsilon$ 以内。每个输入点 x 都被嵌入为 Ax，其中 A 是一个每列含有 s 个非零元素的 $m \times d$ 矩阵，因此嵌入时间为 $O(s\|x\|_0)$。

由于 A 的稀疏度决定了嵌入所需的时间，因此许多研究致力于降低稀疏度s。Kane 和 Nelson (JACM'14) 给出的当前最佳结果表明，$s = O(\varepsilon^{-1} \lg n)$ 已经足够。Nelson 和 Nguyen (STOC'13) 给出的下界为 $s = \Omega(\varepsilon^{-1} \lg n/\lg(1/\varepsilon))$，与之几乎匹配。因此，先前的研究表明，我们已经得到了近乎最优的嵌入。

在这项工作中，我们重新审视了稀疏嵌入，并发现了下界中的一个漏洞。具体而言，该下界要求 $d \geq n$，这在许多应用中并不现实。我们利用这一漏洞，在 $d = o(n)$ 时给出一种更稀疏的嵌入，其稀疏度达到 $s = O(\varepsilon^{-1}(\lg n/\lg(1/\varepsilon) + \lg^{2/3} n \lg^{1/3} d))$。我们还进一步加强了对 Nelson 和 Nguyen 下界的分析，使其在 $d \ll n$ 时同样成立，从而与我们新的稀疏度上界中的首项相匹配。最后，对于达到最优嵌入维度的情形，我们也降低了最佳无关子空间嵌入的稀疏度。

*计算机科学系。奥胡斯大学。 hogsgaard@cs.au.dk。

†计算机科学学院。特拉维夫雅法学术学院。 liorkm@mta.ac.il。

‡计算机科学系。奥胡斯大学。 larsen@cs.au.dk。§电气工程与计算机科学系。加州大学伯克利分校。 minilek@berkeley.edu。本研究由美国国家科学基金会资助（CCF-1951384），并获美国海军研究办公室资助（N00014-18-1-2562 和 DORECG 奖项 N00014-17-1-2127）。

¶计算机科学系。奥胡斯大学。 schwiegelshohn@cs.au.dk。

1 引言

降维是加速大规模数据分析算法、降低存储内存消耗的一项核心技术。通俗地说，欧几里得距离保持降维是指将高维欧几里得空间嵌入低维空间，同时近似保持任意两点之间的欧几里得距离。一个里程碑式的成果是 Johnson-Lindenstrauss 变换 [JL84], ，它表明，任意一组包含 n 个点的数据，都可以嵌入到仅有 $m = O(\varepsilon^{-2} \lg n)$ 维的 d 维空间中，同时使点之间所有成对的欧几里得距离保持在因子 $(1 \pm \varepsilon)$ 以内。构造这类降维映射最简单的方法是随机构造，称为分布式约翰逊-林登施特劳斯引理：对随机 $m \times d$ 矩阵 A 进行采样，其条目要么服从独立同分布的 $\mathcal{N}(0, 1)$，要么为均匀分布的拉德马赫变量（以 $1/2$ 的概率取 -1 或 1）。对于一个包含 n 个点的集合$X \subset \mathbb{R}^d$，以至少 $1 - 1/n$ 的概率，有 $L = A/\sqrt{m}$ 满足

$$\forall x, y \in X : \|Lx - Ly\|_2^2 \in (1 \pm \varepsilon)\|x - y\|_2^2 . \tag{1}$$

我们称满足 (1) 的矩阵 L 是针对 X 的 ε-JL 矩阵。值得注意的是，有些研究要求 ε-JL 矩阵满足不带欧几里得范数平方的条件 (1)。这两种定义只相差 ε 中的一个常数缩放因子；我们采用前一种定义，因为它能简化计算。

虽然已知 $m = O(\varepsilon^{-2} \lg n)$ 的目标维数是最优的 [JW13, LN17], ，即使在 $d = O(m)$ 的情况下也是如此，但使用上述构造计算点 x 的嵌入 Lx 需要 $\Omega(md) = \Omega(\varepsilon^{-2} d \lg n)$ 次运算。在某些应用中，这可能成为计算瓶颈，因此人们投入了大量工作来设计更快的嵌入算法。这些研究大致可分为两种方法：（1）使用具有快速矩阵-向量乘法算法的结构化嵌入矩阵的构造；（2）使用稀疏嵌入矩阵的构造。

前一种方法的经典例子是 Ailon 和 Chazelle 提出的 FastJL 变换 [AC09]。他们的构造通过计算乘积 $PHDx$ 将点 x 嵌入，其中 D 是一个对角线上带有随机符号的对角矩阵， H 是一个 $d \times d$ 哈达玛矩阵，而 P 是一个随机稀疏矩阵，其中每个元素仅以较小的概率非零。该构造的核心思想是， HD 将向量 x 的质量均匀地“铺开”到各个坐标上，从而可以使用非常稀疏的 $m \times d$ 嵌入矩阵 P。此外，哈达玛矩阵具有一个 $O(d \lg d)$ 矩阵-向量乘法算法。FastJL 变换的分析，尤其是目标维数与嵌入时间之间的正确权衡，已得到广泛研究（例如参见 [DGC+09、KW11、FL20、JPS20]）。Fandina、Høgsgaard 和 Larsen [FHL22] 给出的最新紧确分析表明，嵌入时间可以界定为$O(d \lg d + \min\{\varepsilon^{-1} d \lg n, m \lg n \cdot \max\{1, \varepsilon \lg n / \lg(1/\varepsilon)\}\})$。

后一种方法则设计每列仅含 $s \ll m$ 个非零元素的嵌入矩阵。给定这样的稀疏嵌入矩阵，可以在 $O(ds)$ 时间内嵌入一个点，而不是 $O(dm)$，因此，最小化 s 一直是大量研究工作的重点。当前最稀疏的嵌入构造归功于 Kane 和 Nelson [KN14], ，其稀疏度上界达到 $s = O(\varepsilon^{-1} \lg n)$。Nelson 和 Nguyen [NN13a]为任意稀疏 ε-JL 矩阵给出了 $s = \Omega(\varepsilon^{-1} \lg n / \lg(m / \lg n))$ 的下界，几乎确定了 Kane 和 Nelson 构造的最优性。对于最优目标维数 $m = \Theta(\varepsilon^{-2} \lg n)$，这可简化为 $s = \Omega(\varepsilon^{-1} \lg n / \lg(1/\varepsilon))$。尽管 $O(d\varepsilon^{-1} \lg n / \lg(1/\varepsilon))$ 通常大于 FastJL 所达到的近似 $O(d \lg d)$ 嵌入时间，但稀疏嵌入有一项显著优势，即还可以利用输入点的稀疏性。具体而言，容易看出，点 x 的嵌入时间为$O(s\|x\|_0)$，其中 $\|x\|_0$ 是 x的非零条目数。在许多应用中，例如嵌入文档的词袋和词频-逆文档频率表示时，相比定义域大小 d，输入点确实非常稀疏（文档中每个词在 x 中对应一个非零条目，其中 d 是词典中不同词语的数量）。

少维度的大集合。乍看之下， ε-JL 矩阵稀疏度的已知上界与下界之间相差 $\lg(1/\varepsilon)$，似乎已没有多少改进空间，但我们发现 Nelson 和 Nguyen [NN13a] 的下界存在一个缺陷。具体而言，他们证明中的困难实例是 $\{e_1, \ldots, e_n\}$ 标准单位向量的集合。然而，在许多理论应用中，原始

维度 d 远小于向量集合的大小 n。在这些情形下，这个困难实例并不存在，因此下界会退化为 $s = \Omega(\varepsilon^{-1} \lg d / \lg(m/\lg d))$。然而，Kane 和 Nelson 的上界分析无法利用这一点 $d \ll n$，其上界仍为 $O(\varepsilon^{-1} \lg n)$。

除了拓展我们对稀疏降维的理论认识外，我们还发现，$d \ll n$ 也是一种自然的实际应用场景，即使与稀疏输入点结合时也是如此。以 Sentiment140 数据集为例，它包含 160 万条推文，[GBH09]。采用词袋或词频-逆文档频率表示这些推文，并剔除在 $1.6M$ 推文中出现次数少于 10 次的所有词语后，得到的数据集具有 $n = 1.6 \cdot 10^6$、$d = 37$、129 的特征，每条推文平均包含 12 个词语/非零元素。因此，这些向量极其稀疏，其 d 比 n小 43 倍。类似地，Kosarak 数据集 [BKT18] 包含来自匈牙利在线新闻门户的匿名点击流，其中有$n = 900$、002 条交易记录，每条交易记录由若干项目组成。该数据集共有 $d = 41$、270 个不同项目，每条交易记录平均包含 8.1 个项目。在这里，我们同样有一个 n，其规模是 d的 22 倍，而且输入点非常稀疏。一般而言，在考虑词袋和词频-逆文档频率时，人们会假设词典大小 d固定，而数据点数量 n 可以任意大，这进一步说明了在稀疏度界中区分 n 和 d 的必要性。

因此，人们或许可以期望给出依赖于 $\lg d$ 而非 $\lg n$的上界。这正是我们这项工作的核心观点。

1.1 主要结果

我们的第一个主要结果是改进 Kane 和 Nelson [KN14]随机稀疏嵌入的分析，将在 $d \ll n$情况下稀疏度的 $O(\varepsilon^{-1} \lg n)$ 上界降低。具体而言，我们证明如下结论。

定理 1. 设 $0 < \varepsilon < \varepsilon_0$，其中 ε_0为常数。存在一个定义在 $\mathbb{R}^{m \times d}$ 中的 s-稀疏矩阵分布，满足$m = O(\varepsilon^{-2} \lg n)$，并且

$$s = O\left(\frac{1}{\varepsilon} \cdot \left(\frac{\lg n}{\lg(1/\varepsilon)} + \lg^{2/3} n \lg^{1/3} d\right)\right),$$

使得对于任意一组 n 个向量 $X \subset \mathbb{R}^d$，以至少 $1 - O(1/d)$ 的概率，采样得到的矩阵都是针对 X的一个 ε-Johnson–Lindenstrauss 矩阵。

虽然第一项可能与 Nelsonand Nguyen [NN13a], 给出的下界相似，但他们的下界不适用于集合 X 的大小显著大于维度 d的情形，因此该集合不能仅由 $\mathbb{R}^d$的标准基组成。

我们的第二个结果为定理 1 中的上界提供了补充，给出了 ε-JL 矩阵稀疏度的紧下界。我们证明，若 m 相较于 d 足够小，那么每个将 d 维向量嵌入到 $\mathbb{R}^m$ 中的 ε-JL 矩阵都必须具有相对稠密的列。形式化地说，我们证明如下结论。

定理 2. 设 $0 < \varepsilon < 1/4$，并令 m 满足 $m = \Omega(\varepsilon^{-2} \lg n)$ 且 $m \le (\varepsilon d/\lg n)^{1-o(1)}$。则存在一组 n 个向量 $X \subset \mathbb{R}^d$，使得任何将 X 嵌入到 $\mathbb{R}^m$ 中的 ε-JL 矩阵 A，都必须有一列的稀疏度为 s，且满足

$$s = \Omega\left(\frac{\lg n}{\varepsilon \lg(m/\lg n)}\right).$$

对于最优的 $m = \Theta(\varepsilon^{-2} \lg n)$，这可简化为 $s = \Omega(\varepsilon^{-1} \lg n/\lg(1/\varepsilon))$。

请回想 [NN13a] 中相应的下界，它专门针对 $n = d$ 的情形。结合改进后的上界，当 $\lg n \ge \lg d \cdot \lg^3(1/\varepsilon)$ 时，我们现在对稀疏降维有了完全紧确的认识。尽管这些改进在渐近意义上或许不大，但它们是十年来首次取得的改进，也表明可以利用输入数据的维度来加速嵌入。

子空间嵌入。给定一个 k 维子空间 $V \subset \mathbb{R}^d$，ε-子空间嵌入 [Sar06] 是满足以下条件的矩阵 $A \in \mathbb{R}^{m \times d}$：对所有 $x \in V$，$\|Ax\|_2^2 \in (1 \pm \varepsilon)\|x\|_2^2$。已知存在一个大小为 $O(1)^k$ 的子集$V' \subset V$，使得如果 A 将 V' 中每个向量的 ℓ_2 范数保持在 $(1 + \varepsilon/2)$ 范围内，那么 A

则它是一个 ε-子空间嵌入 [AHK06]。因此，JL 引理表明可以取 $m = O(k/\varepsilon^2)$，事实上，当 A 取自 $\mathbb{R}^{m\times d}$ 上一个与 V [NN14] 独立的固定分布时，这是最优的（这类嵌入称为无关子空间嵌入（OSE））。OSE 可用于加速近似回归、低秩近似以及数值线性代数中的大量其他问题；参见 Woodruff 的专著 [Woo14]。

作为一个简单的例子，考虑近似线性回归问题：我们希望找到一个 $\tilde{\beta}$ 使 $\|X\beta - y\|_2^2$ 近似最小化，其中 $X \in \mathbb{R}^{n\times d}$ 是给定的。通过写出奇异值分解 $X = U\Sigma V^\top$，再令 $\beta_{\mathrm{LS}} := V\Sigma^{-1}U^\top y$，可以在 $O(nd^2)$ 时间内精确求解该问题。于是，$X\beta_{\mathrm{LS}} = UU^\top y$ 是 y 在 X 的列空间上的投影，它使误差最小。[Sar06], 在一项分析中提出的草图与求解范式建议，将 A 设为 $span\{y,\mathrm{cols}(X)\}$ 的子空间嵌入（其维度至多为 $d+1$），然后将 $\tilde{\beta}$ 设为 $\|AX\beta - Ay\|_2^2$ 的最小化解。注意，AX 现在是一个小得多的矩阵，因此可以更快地计算 β。不过，我们还需要 A 要么是稀疏的，要么具有特定结构，这样才能快速计算 AX。否则，如果 A 是任意的非结构化矩阵，计算 AX 所需的时间将超过精确计算 β_{LS}！

注意，如果 A 的每一列都有 s 个非零条目，那么 AX 可以在 $O(s\|X\|_0)$ 时间内计算，其中 $\|X\|_0$ 是 X 中非零条目的数量。直接使用 SparseJL 变换 [KN14] 会得到 $m = O(k/\varepsilon^2)$，$s = O(k/\varepsilon)$。Clarkson 和 Woodruff [CW13] 证明了 $m = O(k^2/\varepsilon^2)$、$s = O(1)$ 是可以实现的；对于无关子空间嵌入（OSE）而言，这是最优的 [NN14, LL22]。然而，如果我们不希望将 m 增加到超过最优上界 $O(k/\varepsilon^2)$，该怎么办？在不牺牲降维的渐近质量的情况下，能够实现的最佳稀疏度 s 是多少？Nelson 和 Nguyen 证明了 $m = O((k/\varepsilon^2)\cdot poly(\varepsilon^{-1}\lg k))$ 可以通过 $s = poly(\lg(k/\varepsilon))/\varepsilon$ [NN13b], 实现，并猜想只需 $s = O((\lg k)/\varepsilon)$，且 $m = O(k/\varepsilon^2)$。Cohen 给出了改进的界，证明 $m = O((k\lg k)/\varepsilon^2)$、$s = O((\lg k)/\varepsilon)$ 即已足够 [Coh16],，这至今仍是已知的最佳界。特别地，对于 $m = O(k/\varepsilon^2)$，尽管有 [NN13b], 的猜想，目前尚不知道比 $s = O(k/\varepsilon)$ 更好的稀疏度界；该界可由直接将 SparseJL 作为黑盒应用得到。在本文中，我们首次证明，在保持 $m = O(k/\varepsilon^2)$ 的同时，可以得到一个为 $o(k/\varepsilon)$ 的稀疏度界。具体而言，我们实现了 $s = O(k/(\varepsilon\lg(1/\varepsilon)) + \sqrt[3]{k^2}\lg k/\varepsilon)$。形式化地说，我们证明如下结论。

定理 3。令 0 $< \varepsilon < 1$。在 $\mathbb{R}^{m\times d}$ 中存在关于 s-稀疏矩阵的分布，其 $m = O(\varepsilon^{-2}k)$ 为一个 d

$$s = O\left(\frac{1}{\varepsilon}\cdot\left(\frac{k}{\lg(1/\varepsilon)} + k^{2/3}\lg^{1/3} k\right)\right),$$

对于任意 k 维子空间 $V \subseteq \mathbb{R}^d$，以至少 $1 - 2^{-k^{2/3}}$ 的概率，采样得到的矩阵是一个针对 V 的 ε-JL 矩阵。

尽管这与猜想的最优上界 $O((\lg k)/\varepsilon)$ 相去甚远，但它首次给出了这样一种分析：在保持最优 m 的同时，其稀疏度 s 严格优于将 SparseJL 作为黑盒应用的结果。

2 技术概述

本节将介绍我们新成果中采用的核心思想。我们首先概述改进后的上界分析，然后介绍下界的主要思想，最后介绍新的子空间嵌入结果。为简化记号，以下我们用 $\|x\|$ 表示 $\|x\|_2$。

更稀疏的降维。Kane 和 Nelson [KN14] 提出的一种构造稀疏 JL 矩阵的方法基于 CountSketch 算法 [CCF04]。通过将 m 行划分为 s 组、每组包含 m/s 个条目，来采样嵌入矩阵 A。在 A 的每一列中，从每组均匀随机地采样一个条目，并将其均匀设置为 $1/\sqrt{s}$ 或 $-1/\sqrt{s}$。所有其他条目均设为 0。随后，Kane 和 Nelson 证明，如果 $s = \Omega(\varepsilon^{-1}\lg(1/\delta))$，那么对于每个单位向量 x，以至少 $1-\delta$ 的概率有 $\|Ax\|^2 \in 1\pm\varepsilon$。令 $\delta = n^{-3}$，利用 A 的线性性，并对输入点/向量集合 X 中的 $z = (y-x)/\|y-x\|$ 应用并集界，即可完成他们的证明。下文将重点说明，以良好的概率，A 能保持集合 X 中 n^2 个单位向量的范数。Kane 和 Nelson

还给出了一个简短的论证，说明他们的分析对于标准单位向量 e_1、……、 e_d之间的距离是紧的。

然而，我们的关键观察是，如果 $d \ll n$，那么对 X 中所有 n^2 对向量直接使用并集界可能会过于宽松。具体来说，这种最坏情况形式的向量远少于 n^2 个。尤其是，当 $d \ll n$ 时，基数为 n 的集合 X 中的大多数向量都必须有许多幅值较小的条目。关于特征哈希 [WDL+09、DKT17、FKL18、Jag19] 以及 FastJL 变换 [AC09、FHL22] 的研究早已表明，与最坏情况向量相比，x 中 $\|x\|_\infty$ 与 $\|x\|$ 的比值较小的向量更容易嵌入。例如，对于最优的 $m = \Theta(\varepsilon^{-2} \lg n)$，Jagadeesan [Jag19] 证明，只要 $s = \Omega(\varepsilon^{-1} \lg n / \lg(1/\varepsilon))$，且比值 $\nu = \|x\|_\infty / \|x\|$ 满足 $\nu \le \sqrt{\varepsilon s / \lg n}$，那么 SparseJL 以至少 $1 - 1/n^3$ 的概率将 x 的范数保持在 $1 \pm \varepsilon$ 以内。

为了利用较小的维度 d，我们将每个向量 $x \in X$ 拆分为两个支撑集不相交的向量，分别称为头部和尾部，其中头部包含 x 中排名前 ℓ 的条目，尾部则包含其余条目。也就是说，我们写作 $x = x_{head} + x_{tail}$。然后

$$\|Ax\|^2 = \|Ax_{head}\|^2 + \|Ax_{tail}\|^2 + 2\langle Ax_{head}, Ax_{tail}\rangle \,.$$

现在我们分别处理这三项。证明以高概率有 $\|Ax_{head}\|^2 \in (1 \pm \varepsilon)\|x_{head}\|^2$，$\|Ax_{tail}\|^2 \in (1 \pm \varepsilon)\|x_{tail}\|^2$，以及 $|\langle Ax_{head},\ Ax_{tail}\rangle| \le \varepsilon$ （因为 $\langle x_{head},\ x_{tail}\rangle = 0$）。技术上的关键在于界定交叉项。

为了界定头部，主要观察是头部位置的选择大约有 $\binom{d}{\ell} \le d^\ell$ 种。选定位置后，我们进一步用一个基数为 $(1/\varepsilon)^{O(\ell)}$ 的 ε-网来近似头部。由于 $d \ge m = \Omega(\varepsilon^{-2} \lg n)$，我们需要考虑的头部总数为$d^\ell(1/\varepsilon)^{O(\ell)} = d^{O(\ell)}$。利用 Kane 和 Nelson 在 $\delta = d^{-O(\ell)}$ 情况下的分析可知，只需满足$s = \Omega(\varepsilon^{-1}\ell \lg d)$，就能以高概率得到所需的界。

至于尾部，至多有 n^2 个不同的尾部，且它们具有$\|x_{tail}\|_\infty \le 1/\sqrt{\ell} \le \|x_{tail}\|_2/\sqrt{\ell}$。因此，我们可以利用 Jagadeesan 的结果证明，只要 s 同时满足 $s = \Omega(\varepsilon^{-1} \lg n / \lg(1/\varepsilon))$ 和 $(1/\sqrt{\ell}) \le \sqrt{\varepsilon s / \lg n}$，就有 $\|Ax_{tail}\|^2 \in (1 \pm \varepsilon)\|x_{tail}\|^2$；而这可由 $s = \Omega(\varepsilon^{-1} \lg(n)/\ell)$ 推出。

主要挑战在于界定交叉项，证明 $|\langle Ax_{head}, Ax_{tail}\rangle| \le \varepsilon$。先前的结果，尤其是 Kane 和 Nelson [KN14] 以及 Jagadeesan [Jag19] 的上述结果，都无法直接应用：一方面，配对数量非常庞大，具体而言，它关于 n 呈多项式级增长；另一方面，由于头部包含大元素，相应向量的 ℓ_∞/ℓ_2 比值无法得到上界。为界定交叉项，我们针对 Kane 和 Nelson 基于 CountSketch 的构造，给出了新的集中界。我们首先证明，对于最优维数 $m = O(\varepsilon^{-2} \lg n)$、稀疏度 $s \le \varepsilon m$ 和 $\ell \le \varepsilon^{-1/2}$，以高概率，对于每个 $x \in X$，在 A 中，只有少数几行与 x_{head} 的支撑集重合的非零条目超过 5 个。反过来说，这意味着 Ax_{head} 的大多数条目都不会太大，具体而言，不会超过 $\sqrt{5/s}$。接下来，我们分析存在某个 $x \in X$ 时的概率，

$$\langle Ax_{head}, Ax_{tail}\rangle = \sum_{i\in[m]} (Ax_{head})_i (Ax_{tail})_i = \sum_{i\in[m]} \left((Ax_{head})_i \sum_{j \in \mathrm{supp}(x_{tail})} a_{ij} x_j \right)$$

至多为 ε。为此，我们将求和拆分为两部分：第一部分处理满足 $(Ax_{head})_i$ 和 x_j 较大的项（i, j），第二部分处理其余项。正如刚才所论证的，我们在此利用了对于大多数 i，$(Ax_{head})_i$都很小这一点。此外，由于 x 是单位长度，因此 j 中使x_j 较大的取值也很少。因此，第一部分可以利用求和中的项数较少这一点来处理，而第二部分由于各项都较小，具有很强的集中性。

更强的下界。为了改进 Nelson 和 Nguyen [NN13a], 给出的下界，我们首先需要定义一个更难的输入实例。具体而言，他们使用了标准单位向量 $e_1, \ldots, e_n$；正如前文所述，这只有在 $d \ge n$时才是有效的输入。

我们的困难实例 X 则由所有形如 $v_S = \sum_{i\in S} e_i/\sqrt{|S|}$ 的向量组成，其中子集 $S \subseteq [d]$ 的基数为 $\lg n / \lg d$；还包括所有标准单位向量 $e_1, \ldots, e_d$，以及原点 0。

现在考虑一个 $m \times d$ 嵌入矩阵 A，使得 A 的每一列至多有 s 个非零元素，并且 A 是适用于 X 的 ε-Johnson–Lindenstrauss 矩阵。由于 $e_1, \ldots, e_d$ 和 0 都在输入中，因此 A 的每一列 a_j 的范数都必须属于 $1 \pm \varepsilon \leq 2$。现在为简便起见，假设 A 的所有列都恰好有 s 个非零元素，且这些元素的取值为 $\{-1/\sqrt{s}, 1/\sqrt{s}\}$。对于由 t 个条目构成的子集 $T \subseteq [m]$，以及由 t 个符号组成的列表 $\sigma = (\sigma_1, \ldots, \sigma_t)$，如果 a_j 在 T 对应的每个坐标上均非零，且其在 T 内的坐标符号为 σ，我们就称它具有特征标记 (T, σ)。那么，任意一列都会有 $\binom{s}{t}$ 种不同的特征标记。由于共有 $\binom{m}{t}2^t$ 种特征标记和 d 列，根据平均法，必然存在一种特征标记，至少由 $d\binom{s}{t}/\binom{m}{t}2^t \approx d(s/m)^t$ 列共享。我们将 t 大致设为 $c\lg d/\lg(m/s)$，其中 $c > 0$ 为一个较小的常数，由此至少有 $\mathrm{poly}(d)$ 列共享相同的特征标记。

现在固定这样一个特征标记，并令 S 为 A 中具有该特征标记的列子集。如果 $|S| = \mathrm{poly}(d) \geq \varepsilon^{-1}\lg n/\lg d$，那么我们可以从 S 中选出 ε^{-1} 个互不相交的子集 $S_1, \ldots, S_{\varepsilon^{-1}}$，每个子集的基数均为 $\lg n/\lg d$。对于每个这样的子集 S_i，我们知道向量 v_{S_i} 属于 X。现在，在 T 中的坐标上，S_i 中的所有列均非零且符号相同。因此，由于这些列相加，Av_{S_i} 在 T 中的条目在绝对值上为 $\sqrt{\lg n/(s\lg d)}$。此外，对于不同的 Av_{S_i} 和 Av_{S_j}，T 中的条目也具有相同的符号。

现在，如果我们从所有这样的 Av_{P_i} 中删去 T 中的条目，剩下的向量范数至多为 $1 + \varepsilon < 2$。此外，由于 v_{S_i} 和 v_{S_j} 的支撑集互不相交，它们在嵌入前彼此正交，因此，为了保持它们之间的距离，Av_{S_i} 和 Av_{S_j} 的内积必须为 $O(\varepsilon)$。删去 T 中的条目会使这些内积减少 $|T|\lg n/(s\lg d) = t\lg n/(s\lg d) \approx \lg n/(s\lg(m/s))$。若将所得向量记为 $\tilde{A}v_{S_i}$，则必须有 $0 \leq \|\sum_{i=1}^{\varepsilon^{-1}} \tilde{A}v_{S_i}\|^2 = \sum_{i=1}^{\varepsilon^{-1}} \|\tilde{A}v_{S_i}\|^2 + \sum_i \sum_{j \neq i} \langle \tilde{A}v_{S_i}, \tilde{A}v_{S_j} \rangle \leq 2\varepsilon^{-1} + \varepsilon^{-1}(\varepsilon^{-1} - 1)(O(\varepsilon) - \lg n/(s\lg(m/s)))$。两边乘以 ε 并对 s 求解，得到 $s = \Omega(\varepsilon^{-1}\lg n/\lg(m/s))$。由于 $m = \Omega(\varepsilon^{-2}\lg n)$，这等价于 $s = \Omega(\varepsilon^{-1}\lg n/\lg(m/\lg n))$。

为了处理 A 中不属于 $\{-1/\sqrt{s}, 0, 1/\sqrt{s}\}$ 形式的列，我们将特征重新定义为坐标的子集，其中 a_j 限制在这些坐标上时具有较大的范数。此外，我们不再使用符号 σ，而是在 T 坐标上构造一个 $1/4$-网，并令最近的网点替代这些符号。

将我们的方法与 Nelson 和 Nguyen [NN13a], 的方法进行比较，关键区别在于：我们对 A 中具有相同特征的多个列求和。为确保这些列的和对应于 X 中的一个向量，我们将每个由 $\lg n/\lg d$ 列 v_S 组成的和都加入输入中。

子空间嵌入。对于子空间嵌入，我们注意到，要证明稀疏度为 $s = O(\varepsilon^{-1}k)$，经典方法是构造一个 $1/2$-网 $\mathcal{N}_{\frac{1}{2}} \subset V$，覆盖 k 维子空间 V。随后可以（粗略地）证明，如果线性嵌入矩阵 A 保持所有网点的范数，那么它就会保持 V 中所有点之间的两两距离。由于这样的网的基数为 $2^{O(k)}$，根据 Kane 和 Nelson 的 $s = O(\varepsilon^{-1}\lg(1/\delta))$（其中 $\delta = 2^{-O(k)}$），即可得到所声称的稀疏度。

改进这一结果的一个初步尝试，是将 $n = 2^{O(k)}$ 直接代入上文改进后的稀疏嵌入中。这样得到的稀疏度为 $s = O(\varepsilon^{-1}(k/\lg(1/\varepsilon) + k^{2/3}\lg^{1/3} d))$。第一项没有问题，但后一项取决于 d，这会使该界无法与仅取决于 k 和 ε 的先前结果相比较。因此，我们进一步考察对 d 的依赖是如何产生的。

回顾一下，对于一个包含 n 个向量的集合（例如网 $\mathcal{N}_{\frac{1}{2}}$），我们将向量 $w \in \mathcal{N}_{\frac{1}{2}}$ 划分为头部和尾部：$w = w_{head} + w_{tail}$ 其中，w_{head} 包含 w 中最大的 ℓ 个条目。随后我们观察到，对于嵌入矩阵 A，有

$$\|Aw\|^2 = \|Aw_{head}\|^2 + \|Aw_{tail}\|^2 + 2\langle Aw_{head}, Aw_{tail} \rangle.$$

随后我们证明了 $\|Aw_{head}\|^2 \in (1 \pm O(\varepsilon))\|w_{head}\|^2$，$\|Aw_{tail}\|^2 \in (1 \pm O(\varepsilon))\|w_{tail}\|^2$ 以及 $|\langle Aw_{head}, Aw_{tail} \rangle| = O(\varepsilon)$。对于第二项，我们利用了 $\|w_{tail}\|_\infty \leq 1/\sqrt{\ell}$，并将其与 Jagadeesan 关于具有较小 $\|\cdot\|_\infty$ 的嵌入向量的结果结合起来。由这一项得出的对 s 的要求是 $s = \Omega(\varepsilon^{-1}\lg n/\lg(1/\varepsilon)) = \Omega(\varepsilon^{-1}k/\lg(1/\varepsilon))$，以及 $s = \Omega(\varepsilon^{-1}\lg(n)/\ell) = \Omega(\varepsilon^{-1}k/\ell)$。因此，此处不依赖于 d。类似地，对于交叉项，我们得到的要求是 $s = \Omega(\varepsilon^{-1}\lg(n)/\sqrt{\ell}) = \Omega(\varepsilon^{-1}k/\sqrt{\ell})$。因此，对 d 的依赖仅来自对头部范数的保持。

对于头部，我们论证了头部位置有 $\binom{d}{\ell}$ 种选择，随后需要在选定的 ℓ 个位置上构造一个 ε-网。这使得网中有 $d^{O(\ell)}$ 个头部，接着我们使用 Kane

和 Nelson 的分析，得到 $s = O(\varepsilon^{-1}\lg(1/\delta))$，其中 $\delta = d^{-O(\ell)}$。因此，我们需要对头部数量给出更紧的上界，以避免依赖于 d。

第一个想法是将头部 w_{head} 的定义改为 w 中满足$|w_i| \geq 1/\sqrt{\ell}$的所有条目 w_i。与之前将头部定义为前 ℓ 个条目的方式相比，这是一个虽小却至关重要的改变。为区分二者，我们改用 w_{heavy} 表示大元素，用 $w_{light} = w - w_{heavy}$表示其余条目。

接下来，我们论证 w_{heavy}中至多 ℓ 个条目的位置必须属于一个较小的坐标集合：

引理 4。设 V 是 $\mathbb{R}^d$的一个 k维子空间。对于每个 $\ell \geq 1$，存在一个由 $S \subseteq [d]$ 组成的坐标集合，其满足 $|S| \leq k\ell$；对于每个单位向量 $v \in V$，所有满足 $|v_i| < 1/\sqrt{\ell}$的坐标 $i \in [d] \setminus S$ 均如此。

引理 4 表明，所有 w_{heavy} 中非零元素的位置都必须包含在一个较小的集合 S 中，而该集合的基数仅为 $|S| = k\ell$。因此， w_{heavy}中非零元素的位置只有 $\binom{|S|}{\ell} = 2^{O(\ell \lg k)}$种可能。接下来，我们还将论证，一旦确定了大元素的位置，只需在所选位置上使用 $1/2$-网即可。因此，我们将 w_{heavy}的数量减少到仅 $2^{O(\ell \lg k)}$，并消除了对 d 的依赖。现在应用 Kane 和 Nelson 可知，我们需要 $s = \Omega(\varepsilon^{-1}\ell \lg k)$。将其与 $s = \Omega(\varepsilon^{-1}k/\sqrt{\ell})$平衡后，得到 $\ell = (k/\lg k)^{2/3}$。因此，最终界为 $s = O(\varepsilon^{-1}(k/\lg(1/\varepsilon) + k^{2/3}\lg^{1/3} k))$，如所述。

3 稀疏度上界

本节证明定理 1。设 d 为整数， $\varepsilon \in \in (0, 1)$，并设 $X \subseteq \mathbb{R}^d$ 是某个有限的 n向量集。令 $m = O(\varepsilon^{-2}\lg n)$，并令 $s = O(\varepsilon^{-1}\lg n/\lg(1/\varepsilon) + \varepsilon^{-1}\lg^{2/3} n \lg^{1/3} d$。我们将证明，若按 Kane 和 Nelson [KN14], 中的方式对 A 进行采样，则 A 是一个 ε-JL 矩阵，对于 X，其概率至少为 $1 - O(1/d)$。

为简单起见，我们实际上只证明它是一个 $O(\varepsilon)$-JL 矩阵。将 ε 乘以一个常数因子进行简单缩放，即可得到所需结果。

由于 A 是线性变换，且 n 出现在对数内的所有项中，因此只需证明以下命题（将 X 替换为包含 $x_{i,j} = (x_i - x_j)/\|x_i - x_j\|$ 的 X'，对所有 x_i、$x_j \in X$ 均成立）。

命题 5. 假设 A 按照 Kane 和 Nelson *[KN14]*的方式采样，其中 $m = O(\varepsilon^{-2}\lg n)$ 且 $s = O(\varepsilon^{-1}\lg n/\lg(1/\varepsilon)+\varepsilon^{-1}\lg^{2/3} n\lg^{1/3} d)$，那么对于每个由 n 个单位向量组成的集合 $X \subseteq \mathbb{R}^d$，以至少 $1-O(1/d)$ 的概率，对所有满足 $\|Ax\|^2 \in (1 \pm O(\varepsilon))$ 的 $x \in X$，结论均成立。

因此，在本节余下部分，我们将证明命题 5，并首先引入以下记号。

记号 1. 设 $x \in \mathbb{R}^d$，并设 $\ell \in [d]$。用 $x_{head(\ell)}$ 表示从 x 得到的向量，其中除前 ℓ 个条目外，其余条目均置为零。记 $x_{tail(\ell)} = x - x_{head(\ell)}$。

令 $\ell = \left\lceil \min\left\{\varepsilon^{-1/2}, \left(\frac{\lg n}{\lg d}\right)^{2/3}\right\}\right\rceil$ 为一个整数。对于每个 $T \in \binom{[d]}{\ell}$，令 $\mathcal{Y}_T$为所有满足 $\|y\| \leq 1$ 且 $\mathrm{supp}(y) \subseteq T$ 的向量$y \in \mathbb{R}^d$组成的集合。令 $\mathcal{Y} = \bigcup_{T \in \binom{[d]}{\ell}} \mathcal{Y}_T$。注意，对于每个 $i \in [d]$， $e_i \in \mathcal{Y}$。

固定一些 $X \subseteq \mathbb{R}^d$ n 单位向量。定义 $\mathcal{E}_1$ 为所有满足以下条件的矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in \mathcal{Y}$，$\|Ax\|^2 \in (1 \pm \varepsilon)\|x\|^2$。定义 $\mathcal{E}_2$ 为所有满足以下条件的矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in X$，$\|Ax_{tail(\ell)}\|^2 \in \|x_{tail(\ell)}\|^2 \pm \varepsilon$。定义 $\mathcal{E}_3$ 为所有满足以下条件的矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in X$，要么 $\|Ax_{head(\ell)}\|^2 > 2$，要么$\left|\langle Ax_{head(\ell)},\ Ax_{tail(\ell)}\rangle\right| < \varepsilon$。

命题 6. 假设 $A \in \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$。那么对于每个 $x \in X$， $\|Ax\|^2 \in (1 \pm O(\varepsilon))$。

证明。令 $x \in X$，则 $\|Ax\|^2 = \|Ax_{head(\ell)}\|^2 + \|Ax_{tail(\ell)}\|^2 + 2\langle Ax_{head(\ell)}, Ax_{tail(\ell)}\rangle$。若 $A \in \mathcal{E}_1$，则 $\|Ax_{head(\ell)}\|^2 \in (1 \pm \varepsilon)\|x_{head(\ell)}\|^2$。具体而言， $\|Ax_{head(\ell)}\|^2 < 2$，因此，由于我们还有 $A \in \mathcal{E}_3$，必定有 $\left|\langle Ax_{head(\ell)}, Ax_{tail(\ell)}\rangle\right| < \varepsilon$。因此

$$\|Ax\|^2 \leq (1 + \varepsilon)\|x_{head(\ell)}\|^2 + \|x_{tail(\ell)}\|^2 + \varepsilon + 2\varepsilon \leq (1 + O(\varepsilon))$$

类似地

$$\|Ax\|^2 \geq (1-\varepsilon)\|x_{head(\ell)}\|^2 + \|x_{tail(\ell)}\|^2 - \varepsilon - 2\varepsilon \geq (1 - O(\varepsilon))$$

□

还需证明 $\Pr[A \in \mathcal{E}_1 \cap \mathcal{E} \cap \mathcal{E}_3] \geq 1 - O(1/d)$。这由接下来的三个断言推出，它们分别界定了各个事件的概率。

断言 7。$\Pr[A \in \mathcal{E}_1] \geq 1 - d^{-1}$。

证明。记 $\delta = 2^{-\Omega(\sqrt[3]{\lg^2 n \lg d})}$。由于 $n \geq d$，我们得到 $m \geq \Omega(\varepsilon^{-2}\lg(1/\delta))$ 和 $s \geq \Omega(\varepsilon^{-1}\lg(1/\delta))$。根据 Kane 和 Nelson [KN14,]，对于每个单位向量 $x \in \mathbb{R}^d$，我们有 $\Pr\left[\|Ax\|^2 \in (1 \pm \varepsilon)\right] \geq 1 - \delta$。记 $\hat{\mathcal{Y}}$ 为满足以下条件的所有向量 $y \in \mathcal{Y}$ 的集合：对于每个 $i \in T$，$d^3 y_i$ 都是整数。于是，对于每个 $y \in \hat{\mathcal{Y}}$，以及每个 $i \in \mathrm{supp}(y)$，$d^3 y_i \in \{-d^3, -d^3+1, \ldots, d^3\}$，因此

$$|\hat{\mathcal{Y}}| \leq \binom{d}{\ell}(2d^3)^\ell \leq (2d^4)^\ell \leq 2^{O(\ell \lg d)} = 2^{O(\sqrt[3]{\lg^2 n \lg d})} \leq \frac{1}{\sqrt{\delta}} .$$

因此，以至少 $1 - \sqrt{\delta} \geq 1 - d^{-1}$ 的概率，我们得到对所有 $y \in \hat{\mathcal{Y}}$，$\|Ay\|^2 \in (1 \pm \varepsilon)$。

因此，假设对所有 $y \in \hat{\mathcal{Y}}$，$\|Ay\|^2 \in (1 \pm \varepsilon)$。令 $x \in \mathcal{Y}$，并令 $y \in \hat{\mathcal{Y}}$ 是 $\hat{\mathcal{Y}}$ 中距离 x 最近的向量。于是 $\|x - y\|^2 \leq \ell/d^3$。由于 $\|A\|_F^2 = d$，我们得到

$$\|A(x-y)\| \leq \|A\|_F \|x-y\| \leq \sqrt{\frac{sd\ell}{d^3}} \leq O\left(\sqrt{\frac{(\lg n)^{5/3}}{\varepsilon d^2}}\right) = O(\varepsilon) ,$$

其中最后一个不等式是因为 $d \geq \lg_{\varepsilon^2}{}^n$。因此

$$\|Ax\| \leq \|Ay\| + \|A(x-y)\| \leq 1 + \varepsilon + O(\varepsilon) \leq 1 + O(\varepsilon),$$

类似地，$\|Ax\| \geq 1 - O(\varepsilon)$。 □

对于第二个断言，我们利用 Jagadeesan [Jag19] 的以下结果。

定理 8（Jagadeesan [Jag19]）。对于任意 $0 < \delta < 1$ 和 $0 < \varepsilon < \varepsilon_0$，若某个常数为 ε_0，假设 A 按照 Kane 和 Nelson *[KN14]* 的方式进行采样，且满足 $m \geq \Theta(\varepsilon^{-2}\lg(1/\delta))$ 和 $m \geq s\exp(\max\{1,\ln(1/\delta)/(\varepsilon s)\})$，则对于任意向量 v，有

$$\frac{\|v\|_\infty}{\|v\|} = O\left(\sqrt{\frac{\varepsilon s \ln(m\varepsilon^2/\ln(1/\delta))}{\ln(1/\delta)}}\right)$$

我们有 $\Pr[\|Av\|^2 \in (1 \pm \varepsilon)\|v\|^2] \geq 1 - \delta$。

利用定理 8，我们现在转而界定 $\mathcal{E}_2$ 的概率。

断言 9。$\Pr[A \in \mathcal{E}_2] \geq 1 - n^{-1}$。

证明。固定 $x \in X$。记 $v = x_{tail(\ell)}$，并令 $\hat{\varepsilon} = \max\{\varepsilon, \lg n \left(\frac{\|v\|_\infty}{\|v\|}\right)^2\}$。我们希望应用定理 8，因此首先验证所选参数满足该定理中的约束。代入适当的常数可得 $m \geq \Theta(\varepsilon^{-2}\lg n) \geq \Theta(\hat{\varepsilon}^{-2}\lg n)$。此外

$$\begin{aligned} se^{\Theta(\max\{1,(\hat{\varepsilon}s)^{-1}\lg n\})} &\leq se^{\Theta(\max\{1,(\varepsilon s)^{-1}\lg n\})} \leq se^{\max\{\Theta(1),-\lg\varepsilon\}} = s \cdot \max\{e^{\Theta(1)}, \frac{1}{\varepsilon}\} \\ &= O\left(\frac{1}{\varepsilon}\left(\sqrt[3]{\lg^2 n \lg d} + \lg n/\lg(1/\varepsilon)\right)\right)\max\{e^{\Theta(1)}, \frac{1}{\varepsilon}\} \leq m. \end{aligned}$$

最后请注意

$$\sqrt{\hat{\varepsilon} s \frac{\lg \frac{m\hat{\varepsilon}^2}{\lg n}}{\lg n}} \geq \sqrt{\frac{\lg n}{s}\left(\frac{\|v\|_\infty}{\|v\|}\right)^2 \cdot s\frac{1}{\lg n}} \geq \frac{\|v\|_\infty}{\|v\|}$$

因此，定理 8 告诉我们，以概率 $\geq 1-\frac{1}{n^2}$，有 $\|Av\|^2 \in (1\pm\hat{}\varepsilon)\|v\|^2$。也就是说，$\|Av\|^2 \in |v\|^2 \pm \hat{}\varepsilon\|v\|^2 = \|v\|^2 \pm \max\{\varepsilon\|v\|^2, \lg_s^n \|v\|_\infty^2\}$。首先注意到 $\varepsilon\|v\|^2 < \varepsilon$。接下来，由于 $v = x_{tail(\ell)}$，且 $\|x\| = 1$，我们有 $\|v\|_\infty \leq \frac{1}{\sqrt{\ell}}$，因此 $\lg_s^n \|v\|_\infty^2 \leq \frac{\lg n}{s\ell} = O\left(\frac{1}{s}\cdot\max\{\lg^{1/3} n \lg^{2/3} d,\ \varepsilon^{1/2}\lg n\}\right) = O(\varepsilon)$。我们得出结论，以概率 $\geq 1-\frac{1}{n^2}$，有 $\|Ax_{tail(\ell)}\|^2 \in \|x_{tail(\ell)}\|^2 \pm O(\varepsilon)$。应用并集界，我们得到 $\Pr[A \in \mathcal{E}_2] \geq 1-\frac{1}{n}$。□

下面的断言基本表明，对于 A 的选择，以高概率而言，交叉项很小；它构成了上界结果的技术核心，而其证明需要对 Kane 和 Nelson [KN14] 的构造进行更加仔细的考察。

断言 10。$\Pr[A \in \mathcal{E}_3] \geq 1 - n^{-1}$。

回顾一下，对于 m 和 s 的一种选择，该构造通过将 A 的行分成 s 个块，每个块包含 m/s 个连续行，即 $[1, m/s]$、$[m/s+1, 2m/s]$ 等来实现。对于每一列，在每个块中选择一个均匀随机的条目，并同时选择一个独立的均匀随机符号 σ。随后将该条目设为 $\sigma/\sqrt{s}$。因此，每一列在每个行块中恰好有一个非零元素。

本节的其余部分将用于证明断言 10。我们首先证明，$x_{head(\ell)}$ 经常具有某种理想性质。具体来说，我们定义如下

定义 1. 如果集合 $J \in \binom{[d]}{\ell}$ of ℓ 列 A 是良态的，则不存在多于 $6\lg n/\lg(1/\varepsilon)$ 行 $i \in [m]$ 使得 $|\{j \in J : a_{ij} \neq 0\}| \geq 6$。

命题11. 设 A 按照 Kane 和 Nelson *[KN14]* 进行采样，其中 $m = O(\varepsilon^{-2}\lg n)$ 且 $s \leq \varepsilon m$。那么对于 A 的每个 $J \in \binom{[d]}{\ell}$ 列集合 ℓ，以至少 $1 - n^{-3}$ 的概率有 J 是良态的。

证明。设 $J \in \binom{[d]}{\ell}$，并记 $\beta = 6\lg n/\lg(1/\varepsilon)$。对于 β 的每个 $I = \{i_1, \ldots, i_\beta\} \in \binom{[m]}{\beta}$ 行子集，以及 J 的大小均为 6 的子集所组成的每个序列 V_1、……、$V_\beta \in \binom{J}{6}$，定义事件 $\mathcal{E}_{I,\ V_1,\ldots,\ V_\beta}$ 为满足以下条件的所有矩阵 A 的集合：对于每个 $i \in I$，对于每个 $j \in V_i$，$a_{i,\ j} \neq 0$。

首先注意，如果条目 $\{a_{i,j}\}_{i\in I,\ j\in V_i}$ 不独立，那么必定有两个这样的条目位于同一列和同一行子集中的位置。在这种情况下，$\Pr[\mathcal{E}_{I,V_1,\ldots,V_\beta}] = 0$ 至多有一个非零。否则，A 在 $\mathcal{E}_{I,\ V_1,\ \ldots,\ V_\beta}$ 中考虑的所有 6β 条目彼此独立，因此 $\Pr[\mathcal{E}_{I,V_1,\ldots,V_\beta}] \leq \left(\frac{s}{m}\right)^{6\beta}$。由于 $s \leq \varepsilon m$ 和 $\ell \leq \varepsilon^{-1/2}$，并且应用并集界，我们得到

$$\Pr[J \text{ is not well-behaved}] \leq \sum_{I\in\binom{[m]}{\beta}} \sum_{V_1,\ldots,V_\beta\in\binom{J}{6}} \Pr[\mathcal{E}_{I,V_1,\ldots,V_\beta}] \leq \left(\frac{me}{\beta}\right)^\beta \cdot \ell^{6\beta}\varepsilon^{6\beta} \leq (O(1)\varepsilon^{-2}\lg(1/\varepsilon))^\beta \cdot \varepsilon^{3\beta}.$$

当 ε 小于某个常数时，这至多为 $\varepsilon^{\beta/2} \leq n^{-3}$。□

接下来我们证明，如果 $x_{head(\ell)}$ 的支撑集是列的一个良态子集，那么 $Ax_{head(\ell)}$ 只有少量“较大”条目（注意 $|\text{supp}(x_{head(\ell)})| \leq \ell$）。

命题 12. 设 $x \in X$，并假设 $x_{head(\ell)}$ 的支撑集是良态的。那么，$Ax_{head(\ell)}$ 中超过 $\sqrt{5/s}$ 的条目至多有 $6\lg n/\lg(1/\varepsilon)$ 个。此外，我们有 $\|Ax_{head(\ell)}\|_\infty \leq \sqrt{\ell/s}$。

证明。令 $I \subseteq [m]$ 为满足 $|\{j \in \text{supp}(x_{head(\ell)}) : a_{ij} \neq 0\}| \geq 6$ 的行集合 $i \in [m]$。考虑一个 $i \in [m] \setminus I$。满足 $a_{ij} \neq 0$ 的列 $j \in [d]$ 数量至多为 5，且对其中每一列，我们都有 $|a_{ij}| \leq 1/\sqrt{s}$。因此，由于 $\|x_{head(\ell)}\| \leq 1$，可得 $|(Ax_{head(\ell)})_i| \leq \sqrt{5/s}$。由于 $\text{supp}(x_{head(\ell)})$ 是良态的，因此 $|I| \leq 6\lg n/\lg(1/\varepsilon)$，结论得证。界 $\|Ax_{head(\ell)}\|_\infty \leq \sqrt{\ell/s}$ 直接由 $x_{head(\ell)}$ 的支撑集仅有基数 ℓ 以及 $\|x_{head(\ell)}\| \leq 1$ 得出。□

对于每个 $x \in X$，令 $\mathcal{E}_{3,x}$ 为满足 $\|Ax_{head(\ell)}\|^2 \geq 2$ 或 $|\langle Ax_{head(\ell)}, Ax_{tail(\ell)}\rangle| < \varepsilon$ 的矩阵 A 的集合。则 $\mathcal{E}_3 = \cap_{x\in X}\mathcal{E}_{3,x}$。我们的目标是证明 $\Pr[\mathcal{E}_{3,x}] \geq 1 - O(1/n^2)$；对所有$x \in X$ 应用并集界即可完成断言 10 的证明。为此，定义 $\mathcal{W}_x$ 为所有满足 $x_{head(\ell)}$ 的支撑集是良态列集的矩阵 A 的集合。命题11意味着 $\Pr[\mathcal{W}_x] \geq 1 - n^{-3}$。因此，只需证明 $\Pr[\mathcal{E}_{3,x} \mid \mathcal{W}_x] \geq 1 - O(n^{-2})$，因为 $\Pr[\mathcal{E}_{3,x}] \geq \Pr[\mathcal{E}_{3,x} \mid \mathcal{W}_x]\Pr[\mathcal{W}_x]$。由此，下面的引理即完成断言 10 的证明；本节余下部分将用于证明该引理。

引理13. $\Pr[\mathcal{E}_{3,x} \mid \mathcal{W}_x] \geq 1 - O(n^{-2})$。

由于 $\Pr[\mathcal{E}_{3,x} \mid \mathcal{W}_x \wedge \|Ax_{head(\ell)}\|^2 \geq 2] = 1$，只需给出 $\Pr[\mathcal{E}_{3,x} \mid \mathcal{W}_x \wedge \|Ax_{head(\ell)}\|^2 < 2]$ 的上界。注意，由于 $x_{head(\ell)}$ 和 $x_{tail(\ell)}$ 的支撑集不相交，向量 $Ax_{head(\ell)}$ 和 $Ax_{tail(\ell)}$ 相互独立。事实上，$Ax_{tail(\ell)}$ 与 A 中支撑集为 $x_{head(\ell)}$ 的所有列完全独立。因此，我们将证明，在给定 $\mathcal{W}_x \wedge \|Ax_{head(\ell)}\|^2 < 2$ 的条件下，$|\langle Ax_{head(\ell)}, Ax_{tail(\ell)}\rangle| = O(\varepsilon)$ 在对支撑集为 $x_{tail(\ell)}$ 的随机列进行选择时，以至少 $1 - O(1/n^2)$ 的概率成立。因此，我们可以固定 $u = Ax_{head(\ell)}$ 的某个取值，使得 $\mathrm{supp}(x_{head(\ell)})$ 也是良态的。

对于每个 $i \in [m]$ 和 $j \in \mathrm{supp}(x_{tail(\ell)})$，定义 b_{ij} 为伯努利随机变量：当 A 的条目 (i,j) 非零时取值为 1，否则取值为 0。此外，令 σ_{ij} 表示均匀随机且相互独立的符号。于是 $\langle u, Ax_{tail(\ell)}\rangle = \sum_{i=1}^{m} u_i \sum_{j\in\mathrm{supp}(x_{tail(\ell)})} b_{ij}\sigma_{ij}x_j/\sqrt{s}$。为界定该和，我们将其拆分为两个和，并分别界定每一部分不超过 $O(\varepsilon)$ 的概率。记

$$R = \{(i,j) \in [m] \times \mathrm{supp}(x_{tail(\ell)}) : |u_i| > \sqrt{5/s} \text{ and } |x|_j > 1/(\sqrt{\ell}\lg^2(1/\varepsilon))\}\text{, and}$$
$$S = ([m] \times \mathrm{supp}(x_{tail(\ell)})) \setminus R$$

命题14。$\Pr\left[\left|\sum_{i\,j\,(,)\in R} u_i \cdot b_{ij}\sigma_{ij}x_j/\sqrt{s}\right| \leq O(\varepsilon)\right] \geq 1 - n^{-2}$。

证明。回顾 $\|u\|_\infty \leq \sqrt{\ell/s}$ 和 $\|x_{tail(\ell)}\|_\infty \leq 1/\sqrt{\ell}$。因此

$$\left|\sum_{(i,j)\in R} u_i \cdot b_{ij}\sigma_{ij}x_j/\sqrt{s}\right| \leq \frac{1}{\sqrt{s}}\sum_{(i,j)\in R}|u_i| \cdot b_{ij}|\sigma_{ij}x_j| \leq \frac{1}{s}\sum_{(i,j)\in R} b_{ij}\,.$$

为完成证明，我们将证明以至少 $1-n^{-2}$ 的概率，有 $\sum_{(i,j)\in R} b_{ij} \leq O(s\varepsilon) \leq c\lg n/\lg(1/\varepsilon)$。由于 $\mathrm{supp}(x_{head(\ell)})$ 是良态的，因此至多有 $6\lg n/\lg(1/\varepsilon)$ 行 $i \in [m]$ 满足 $|u_i| > \sqrt{5/s}$；并且由于 $\|x_{tail(\ell)}\| = 1$，至多有 $\ell\lg^4(1/\varepsilon)$ 列 $j \in \mathrm{supp}(x_{tail(\ell)})$ 满足 $|x_j| \geq 1/(\sqrt{\ell}\lg^2(1/\varepsilon))$。因此 $|R| \leq 6\ell\lg n\lg^3(1/\varepsilon)$，从而 $\mu := \mathbb{E}\left[\sum_{(i,j)\in R} b_{ij}\right] \leq (s/m) \cdot 6\ell\lg n\lg^3(1/\varepsilon) \leq \varepsilon^{1/2}\lg n\lg^3(1/\varepsilon)$，其中最后一个不等式是由 $s \leq \varepsilon m$ 和 $\ell \leq \varepsilon^{-1/2}$ 得出的。当 ε 小于某个常数时，我们得到期望至多为 $\mu \leq \varepsilon^{1/4}\lg n/\lg(1/\varepsilon)$。通过直接计算可得以下观察结果，其证明推迟至附录。

观察 15。对于每个 $t > 0$, $\mathbb{E}\left[\exp\left(t\sum_{(i,j)\in R} b_{ij}\right)\right] \leq \prod_{(i,j)\in R}\mathbb{E}\left[\exp\left(tb_{ij}\right)\right]$。

利用观察 15，我们可以对 $\sum_{(i,j)\in R} b_{ij}$ 较大的概率应用类 Hoeffding 不等式。具体而言，对于足够大的常数 c，令 $\delta = c\varepsilon^{-1/4} - 1$ 和 $t = \ln(1+\delta)$，由马尔可夫不等式可得

$$\Pr\left[\sum_{(i,j)\in R} b_{ij} > \frac{c\lg n}{\lg(1/\varepsilon)}\right] = \Pr\left[\exp\left(t\sum_{(i,j)\in R} b_{ij}\right) > \exp\left(\frac{tc\lg n}{\lg(1/\varepsilon)}\right)\right] \leq \frac{e^{\delta\mu}}{(1+\delta)^{c\lg n/\lg(1/\varepsilon)}}\,.$$

由于 $(1+\delta) = c\varepsilon^{-1/4}$ 和 $\mu \leq \varepsilon^{1/4}\lg n/\lg(1/\varepsilon)$，我们得到：如果 c 足够大

$$\Pr\left[\sum_{(i,j)\in R} b_{ij} > \frac{c\lg n}{\lg(1/\varepsilon)}\right] \leq \left(e\varepsilon^{1/4}\right)^{c\lg n/\lg(1/\varepsilon)} \leq n^{-2}\,.$$

命题 16. $\Pr\left[\left|\sum_{i\,j}\sum_{(,)\in S} u_i \cdot b_{ij}\sigma_{ij}x_j/\sqrt{s}\right| \le O(\varepsilon)\right] \ge 1 - O(n^{-2})$。 □

证明。首先注意到，这个和可以看作由 $(i,j) \in S$ 索引的两个向量之间的内积。具体而言，令 $\sigma, w \in \mathbb{R}^S$ 如下定义。对于每个 $(i,j) \in S$, $\sigma_{(i,j)} = \sigma_{ij}$ 和 $w_{(i,j)} = c_{(i,j)}b_{ij}$，其中 $c_{(i,j)} = u_i x_j/\sqrt{s}$。由于 σ 和 w 是独立的，根据霍夫丁不等式，对于每个 $c > 0$

$$\Pr\left[|\langle w, \sigma\rangle| > c\varepsilon \mid \|w\|\right] \le 2\exp\left(-\frac{(c\varepsilon)^2}{2\|w\|^2}\right).$$

因此，只需证明以至少 $1 - O(n^{-2})$ 的概率有 $\|w\|^2 = O(\varepsilon^2/\lg\ n)$ 成立。首先注意到

$$\mathbb{E}\left[\|w\|^2\right] = \mathbb{E}\left[\sum_{(i,j)\in S} c_{(i,j)}^2 b_{ij}^2\right] = \frac{1}{s}\sum_{(i,j)\in S} u_i^2 x_j^2 \mathbb{E}[b_{ij}] = \frac{1}{m}\sum_{(i,j)\in S} u_i^2 x_j^2 = \frac{1}{m}\|x_{tail(\ell)}\|^2\|u\|^2.$$

由于我们是在给定 $\|u\|^2 < 2$ 的条件下进行分析，并且 $\|x_{tail(\ell)}\|^2 \le 1$，因此有 $\mathbb{E}\left[\|w\|^2\right] \le 2/m = O(\varepsilon^2/\lg\ n)$。因此，我们的目标是对某个常数 $\delta > 0$，给出 $Pr[\|w\|_2 > (1+\delta)\mathbb{E}[\|w\|^2]]$ 的上界。与前一个证明类似，我们使用以下观察，其证明推迟到附录中。

观察 17. 对于每个 $t > 0$, $\mathbb{E}\left[\exp\left(t\sum_{(i,j)\in S} c_{(i,j)}^2 b_{ij}\right)\right] \le \prod_{(i,j)\in S}\mathbb{E}\left[\exp\left(tc_{(i,j)}^2 b_{ij}\right)\right]$。

我们首先对系数 $c_{(i,j)}$进行界定。回顾一下，$\|u\|_\infty \le \sqrt{\ell/s}$ 和 $\|x_{tail(\ell)}\|_\infty \le 1/\sqrt{\ell}$，并令 $(i,j) \in S$。那么，要么 $|u_i| \le \sqrt{5/s}$，要么 $|x_j| \le 1/(\sqrt{\ell}\lg^2(1/\varepsilon))$。在前一种情况下，$|u_i x_j/\sqrt{s}| \le \sqrt{5}/(s\sqrt{\ell})$，根据 s和 ℓ的选取，我们得到 $|u_i x_j/\sqrt{s}| \le O(\varepsilon/\lg\ n)$。在后一种情况下，$|u_i x_j/\sqrt{s}| \le 1/s\lg(1/\varepsilon) = O(\varepsilon/\lg\ n)$。因此，对于所有 $(i,j) \in S$，都有 $|c_{(i,j)}| = |u_i x_j/\sqrt{s}| \le O(\varepsilon/\lg\ n)$。令 $\mu = \mathbb{E}[\|w\|^2]$,
$\alpha = O$ n $t =$ $1+\delta$ δ $((\varepsilon/\lg\ )^2)$，并令 $\ln(\ )/\alpha$，其中 为某个足够大的常数，则由马尔可夫不等式可得

$$\Pr\left[\|w\|^2 > (1+\delta)/m\right] \le \frac{\mathbb{E}\left[\exp\left(t\sum_{(i,j)\in S} c_{(i,j)}^2 b_{ij}\right)\right]}{\exp(t(1+\delta)/m)} \le \frac{\prod_{(i,j)\in S}\mathbb{E}\left[\exp\left(tc_{(i,j)}^2 b_{ij}\right)\right]}{(1+\delta)^{(1+\delta)/(\alpha m)}} \tag{2}$$

现在注意，对于每个 (i,j)，都有 $\in S$

$$\mathbb{E}\left[\exp\left(tc_{(i,j)}^2 b_{ij}\right)\right] = \frac{s}{m}e^{tc_{(i,j)}^2} + \left(1 - \frac{s}{m}\right) = 1 + \frac{s}{m}\left(e^{tc_{(i,j)}^2} - 1\right) = 1 + \frac{s}{m}\left((1+\delta)^{c_{(i,j)}^2/\alpha} - 1\right)$$

由于 $c_{(i,j)}^2 \le \alpha$，我们得到 $(1+\delta)^{c^2\ (i,\ j)\ /\alpha} \le 1 + \delta c_{(i,j)}^2/\alpha$。因此

$$\mathbb{E}\left[\exp\left(tc_{(i,j)}^2 b_{ij}\right)\right] \le 1 + \frac{sc_{(i,j)}^2\delta}{\alpha m} \le \exp\left(\frac{\delta}{\alpha}\cdot\frac{sc_{(i,j)}^2}{m}\right).$$

代入（2）式可得

$$\Pr\left[\|w\|^2 > (1+\delta)\mu\right] \le \frac{\prod_{(i,j)\in S}\exp\left(\frac{\delta}{\alpha}\cdot\frac{sc_{(i,j)}^2}{m}\right)}{(1+\delta)^{(1+\delta)/(\alpha m)}} = \frac{e^{\delta\mu/\alpha}}{(1+\delta)^{(1+\delta)/(\alpha m)}} \le \left(\frac{e^{2\delta}}{(1+\delta)^{1+\delta}}\right)^{1/(\alpha m)},$$

其中，最后一个不等式是由于 $\mu \le 2/m$。由于 $2/(\alpha m) = \Omega(\lg n)$，因此当常数 δ 足够大时，概率至多为 n^{-2} □

4 稀疏度下界

在本节中，我们证明下界结果——定理 2。令 $0 < \varepsilon < 1/4$。我们首先在 $\mathbb{R}^d$ 中定义一组困难输入向量。令 $\ell = \lg n/\lg(ed/\ell)$。对于每个大小为 ℓ 的坐标子集 $S \subseteq [d]$，构造向量 $x_S = \sum_{i\in S} e_i/\sqrt{\ell}$。这些向量连同零向量以及 e_1、……、e_d，构成我们的困难输入实例 X，其基数为 $|X| \le \binom{d}{\ell} + 1 + d \le (ed/\ell)^\ell + n \le 2n$。

假设 A 是一个 $m \times d$ 矩阵，其中每列至多有 s 个非零元素，并且 A 满足 $\|Au - Av\|^2 \in (1\pm\varepsilon)\|u - v\|^2$，对所有 u、$v \in X$ 均成立。我们还假设，关于 m，已有 $\Omega(\varepsilon^{-2}\lg n)$ 这样的下界。我们根据这些假设证明 s 的一个下界。在整个证明过程中，我们假设 $s \le m/2$，否则问题已经解决。

令 a_j 表示 A 的第 j 列。我们首先注意到，对所有 j，都有 $\|a_j\|^2 \in (1\pm\varepsilon)$，因为 $\|a_j\|^2 = \|Ae_j\|^2 = \|Ae_j - A0\|^2 \in (1\pm\varepsilon)\|e_j - 0\|^2 = (1\pm\varepsilon)$。

下一步是找出一个子集 $T \subseteq [m]$，使得 A 的许多列在 T 中都有较大的条目。为此，我们证明以下引理:

引理 18。设 $v \in \mathbb{R}^m$ 是一个至多包含 $s \le m/2$ 个非零元素的向量。对于任意 $t \le s/8$，至少存在 $\min\{\binom{m-1}{t-1}, (s/(8t))^t\}$ 个不同的、基数为 $|T| = t$ 的 $T \subseteq [m]$ 子集，使得 $\sum_{i\in T} v_i^2 \ge t\|v\|^2/(2s)$。

我们将 证明推迟到本节末尾，先进行下界论证 .

设 t 为待定参数。A 中有 d 列，根据引理 18，并在 $[m]$ 的所有 t 元子集中取平均可知，存在一个 T，满足 $|T| = t$，使得至少有 $d \min\{\binom{m-1}{t-1}, (s/(8t))^t\}/\binom{m}{t} \ge d \quad \{t/m,\ s/\ 8t\ ^t/\ em/t\ ^t\} = d \quad \{t/m,\ s/\ 8em\ ^t\}$ a_j A $\sum_{i\in T} a_{i\,j}^2 \ge 1-\varepsilon\ t/\ 2s\ \ge \min($ $t/\ 4s$ T A_T T $)) ( ) \min (( )) $ 列满足 $, ( ) ( ) ( )$。固定这样一个，并令 为满足关于该 的上述条件的列子集。

令 $\mathcal{N}_{\frac{1}{4}}$ 为 $\mathbb{R}^t$ 中单位向量集合的一个 $1/4$-网，即对于任意满足 $\|x\| = 1$ 的 $x \in \mathbb{R}^t$，都存在一个满足 $\|x - x'\| \le 1/4$ 和 $\|x'\| = 1$ 的 $x' \in \mathcal{N}_{\frac{1}{4}}$。标准结果表明，存在这样一个基数为 $2^{O(t)}$ 的 $\mathcal{N}_{\frac{1}{4}}$。对于每个 $a_j \in A_T$，令 a_j^T 表示将 a_j 限制为 T 中的 t 个条目，并令 $w(a_j)$ 表示 $\mathcal{N}_{\frac{1}{4}}$ 中距离 $a_j^T/\|a_j^T\|$ 最近的向量。通过平均法，存在一个向量 $w \in \mathcal{N}_{\frac{1}{4}}$，使得至少有 $d\min\{t/m, (s/m)^t\}2^{-O(t)}$ 个向量 $a_j \in A_T$ 以 w 作为距离 $a_j^T/\|a_j^T\|$ 最近的向量。固定这样一个 w，并令 $A_{T,\ w}$ 为 A 中满足这些条件的列子集。

现在固定 $t = (1 - o(1))\lg(\varepsilon d/\ell)/\lg(m/s)$。首先假设对于这一选择，我们有 $(s/m)^t \le t/m$。那么，由于 $s = o(m)$（否则下界证明即可完成），我们有

$$|A_{T,w}| \ge d(s/m)^t 2^{-O(t)} = d(s/m)^{(1+o(1))t} \ge \ell/\varepsilon.$$

由 $A_{T,\ w}$，构造 ε^{-1} 个不相交集合，每个集合包含 ℓ 个向量。对于每个这样的集合 S，向量 $\sum_{a_j\in S} e_j/\sqrt{\ell}$ 属于 X。令 v_1、……、$v_{\varepsilon^{-1}}$ 表示这些向量。由于它们具有不相交的支撑集，我们有 $\langle v_i, v_j\rangle = 0$（对于 $i \ne j$），因此 $\|v_i - v_j\|^2 = 2$。这也意味着 $\|Av_i - Av_j\|^2 \in 2 \pm 2\varepsilon$。由于 $\|Av_i - Av_j\|^2 = \|Av_i\|^2 + \|Av_j\|^2 - 2\langle Av_i, Av_j\rangle$ 且 $\|Av_i\|^2$、$\|Av_j\|^2 \in 1 \pm \varepsilon$，必然有 $\langle Av_i, Av_j\rangle \in \pm 2\varepsilon$。

另一方面，我们有 $\langle Av_i\ Av_j\rangle = \sum_{a_h\in S_i}\sum_{a_k\in S_j}\langle a_h\ a_k\rangle/\ell \qquad \sum_{a_h\in S_i}\sum_{a_k\in S_j}\langle a_h\ a_k\rangle/\ell \le$ ， 2ε $i \in T$ 0 A 。因此， 。现在，将 的所有列中的所有条目都设为 。将所得的列称为 $\hat{a}_j$，将所得的矩阵称为 $\hat{}A$。那么，对于两列 a_h、a_k，我们有 $\langle\hat{a}_h, \hat{a}_k\rangle = \langle a_h, a_k\rangle - \langle a_h^T, a_k^T\rangle$。对于任意两个 a_h、$a_k \in A_{T,\ w}$，我们有 $\langle a_h^T, a_k^T\rangle = \|a_h^T\|\|a_k^T\|\langle w + (a_h^T/\|a_h^T\| - w)\ w + ($ $a_k^T/\|a_k^T\| - w)\rangle$。我们有

$$\begin{aligned}
\langle w + (a_h^T/\|a_h^T\| - w), w + (a_k^T/\|a_k^T\| - w)\rangle &= \\
\|w\|^2 + \langle w, (a_h^T/\|a_h^T\| - w)\rangle + \langle w, (a_k^T/\|a_k^T\| - w)\rangle + \langle (a_h^T/\|a_h^T\| - w), (a_k^T/\|a_k^T\| - w)\rangle &\ge \\
\|w\|^2 - \|w\|\|a_h^T/\|a_h^T\| - w\| - \|w\|\|a_k^T/\|a_k^T\| - w\| - \|a_h^T/\|a_h^T\| - w\|\|a_k^T/\|a_k^T\| - w\| &\ge \\
1 - 1/4 - 1/4 - 1/16 &\ge \\
1/4.
\end{aligned}$$

因此，对于任意两个 $a_h, a_k \in A_{T,\ w}$，都有 $\langle\hat{a}_h,\ \hat{a}_k\rangle \le \langle a_h, a_k\rangle - \|a_h^T\|\|a_k^T\|/4 \le \langle a_h, a_k\rangle - t(1-\varepsilon)/(8s) \le\hat{}\ \hat{}$ $\langle a_h, a_k\rangle - t/(16s)$。由此我们得出结论：$\langle Av_i, Av_j\rangle \le \sum_{a_h\in S_i}\sum_{a_k\in S_j}(\langle a_h, a_k\rangle - t/(16s))/\ell \le 2\varepsilon - \ell t/(16s)$。

最后，考虑向量 $z = \sum_{i=1}^{\varepsilon^{-1}} \hat{} Av_i$。我们有 $\|z\|^2 = \sum_{i=1}^{\varepsilon^{-1}} \|\hat{} Av_i\|^2 + \sum_{i \neq j} \langle \hat{} Av_i, \hat{} Av_j \rangle \leq \varepsilon^{-1}(1+\varepsilon) + \varepsilon^{-1}(\varepsilon^{-1}-1)(2\varepsilon - \ell t/(16s))$。由于 $\|z\|^2 \geq 0$，因此必有 $(\varepsilon^{-1}-1)\ell t/(16s) \leq 1+\varepsilon+(\varepsilon^{-1}-1)2\varepsilon$。由于 $\varepsilon^{-1} - 1 \geq \varepsilon^{-1}/2$ and $1 + \varepsilon + (\varepsilon^{-1} - 1)2\varepsilon \leq 4$，我们得出 $s \geq \varepsilon^{-1}\ell t/128$。由于 $d \geq m = \Omega(\varepsilon^{-2} \lg n)$，对于某个常数 $c' > 0$，有 $\lg(ed/\ell) \leq c' \lg(\varepsilon d/\ell)$。因此

$$s = \Omega(\varepsilon^{-1} \lg n / \lg(m/s)) = \Omega(\varepsilon^{-1} \lg n / \lg(m/\lg n)).$$

这仅在假设 $(s/m)^t \leq t/m$ for $t = (1 - o(1)) \lg(\varepsilon d/\ell) / \lg(m/s)$ 的情况下成立。这由 $(\ell/(\varepsilon d))^{1-o(1)} \leq (1 - o(1)) \lg(\varepsilon d/\ell) / (m \lg(m/s))$ 推出。特别地，它由 $m \leq (\varepsilon d/\ell)^{1-o(1)}$ 推出。因此，施加约束 $m \leq (\varepsilon d/\lg n)^{1-o(1)}$ 即完成了证明。

引理 18 的证明。首先考虑 v 至少有一个坐标 j 满足 $v_j^2 \geq t\|v\|^2/(2s)$ 的情形。在这种情况下，T 至少有 $\binom{m-1}{t-1}$ 种有效选择。

如果所有坐标 j 都满足 $v_j^2 < t\|v\|^2/(2s)$，我们就根据 v 的坐标大小将其划分到不同的桶中。具体而言，对于每个 $i = 0, \ldots, \lg t - 1$，令 V_i 表示满足 $v_j^2 \in [\|v\|^2 2^{i-1}/s, \|v\|^2 2^i/s)$ 的坐标 j 子集。注意，j 中满足 $v_j^2 < \|v\|^2/(2s)$ 的所有坐标对 $\|v\|^2$ 的贡献至多为$s\|v\|^2/(2s) = \|v\|^2/2$ 。此外，对于满足 $j \in V_i$ 的坐标 j，若某个 V_i 满足$|V_i| \leq s/(4t)$，则这些坐标的贡献不超过 $\sum_{i=0}^{\lg t-1} (s/(4t))\|v\|^2 2^i/s \leq \|v\|^2/4$。因此 $\sum_{i:\ |V_i| > s/(4t)} \sum_{j \in V_i} v_j^2 \geq \|v\|^2/4$。这意味着我们也有 $\sum_{i:\ |V_i| > s/(4t)} |V_i| \cdot \|v\|^2 2^i/s \geq \|v\|^2/4$。

对于每个满足 $|V_i| > s/(4t)$ 的 i，令 $t_i := \lceil 4t|V_i|/s \rceil$。则 $t_i \leq 4t|V_i|/s + 1 \leq 4t|V_i|/s + 4t|V_i|/s \leq 8t|V_i|/s$。考虑所有满足以下条件的集合 T：对于所有满足 $|V_i| > s/(4t)$ 的 i，均有 $|T \cap V_i| = t_i$。任何这样的 T 都满足$\sum_{j \in T} v_j^2 \geq \sum_{i:|V_i| > s/(4t)} t_i \|v\|^2 2^{i-1}/s \geq (2t/s) \sum_{i:|V_i| > s/(4t)} |V_i| \cdot \|v\|^2 2^i/s \geq (t/(2s))\|v\|^2$。此类 T 的数量至少为 $\binom{m/2}{t - \sum_i t_i} \prod_{i:|V_i| > s/(4t)} \binom{|V_i|}{t_i}$。对于 $|V_i| > s/(4t)$，我们有 $\binom{|V_i|}{t_i} \geq (|V_i|/t_i)^{t_i} \geq (|V_i|/(8t|V_i|/s))^{t_i} = (s/(8t))^{t_i}$。因此，有效 T 的数量至少为 $\binom{m-s}{t - \sum_i t_i} \prod_{i:|V_i| > s/(4t)} (s/(8t))^{t_i} \geq (m/(2t))^{t - \sum_i t_i} (s/(8t))^{\sum_i t_i} \geq (s/(8t))^t$。

□

5 子空间嵌入

在本节中，我们将证明，对于任意 k 维子空间 $V \subset \mathbb{R}^d$，按照 Kane 和 Nelson [KN14], 中的方法采样、具有稀疏度 $s = \Theta(\varepsilon^{-1}(k/\lg(1/\varepsilon) + k^{2/3} \lg^{1/3} k))$ 的嵌入矩阵 A，以高概率将 V 中每个向量的范数保持在 $1 \pm \varepsilon$ 以内，从而证明定理 3。

为简化证明，我们将再次论证范数保持在 $1 \pm O(\varepsilon)$ 以内。与第 3 节一样，只需将 ε 乘以一个常数因子进行简单缩放，即可得到该结果。

我们首先证明，只要 A 能近似保持由子空间上的 $1/2$-网定义的有限集合的范数，就足够了。以下引理是已知结果，曾出现在以往的研究中。为完整起见，我们给出证明；证明推迟至附录第 B 节。

引理 19. 设 A 是一个矩阵，V 是 $\mathbb{R}^d$ 的一个子空间。设 $\mathcal{N}_{\frac{1}{2}}$ 是 V 的一个 $1/2$-网，且 $\mathcal{N}_{\frac{1}{2}}^+ = \{x + y : x, y \in \mathcal{N}_{\frac{1}{2}} \cup \{0\}\}$。假设对于所有 $v \in \mathcal{N}_{\frac{1}{2}}^+$，$\|Av\|^2 \in (1 \pm O(\varepsilon))\|v\|^2$，则对于所有单位向量 $x \in V$，$\|Ax\|^2 \in (1 \pm O(\varepsilon))\|x\|^2$。

如技术概述中所述，我们还使用引理 4。该引理给出了 $\mathbb{R}^d$ 的子空间的一种组合性质。

引理 4. 设 V 是 $\mathbb{R}^d$ 的一个 k-维子空间。对于每个 $\ell \geq 1$，都存在一个坐标集合 $S \subseteq [d]$，满足 $|S| \leq k\ell$，使得对于每个单位向量 $v \in V$，所有坐标 $i \in [d] \setminus S$ 都满足$|v_i| < 1/\sqrt{\ell}$。

证明。设 $v^1, \ldots, v^k$ 是 V 的一组标准正交基。取任意单位向量 $u \in V$，并将其写为 $u = \sum_j \alpha_j v^j$，其中 $\sum_j \alpha_j^2 = 1$。于是 $u_i = \sum_j \alpha_j v_i^j$。根据柯西-施瓦茨不等式，有 $|u_i| \leq \sqrt{\sum_j \alpha_j^2} \cdot \sqrt{\sum_j (v_i^j)^2} = \sqrt{\sum_j (v_i^j)^2}$。现在令 $S \subseteq [d]$ 表示所有满足存在单位向量 $u \in V$ 且 $|u_i| \geq 1/\sqrt{\ell}$ 的坐标。那么对于所有 $i \in S$，必有 $\sum_j (v_i^j)^2 \geq 1/\ell$。但是 $\sum_j \sum_i (v_i^j)^2 = k$，因此 $|S| \leq k\ell$。

□

借助上述两个引理，我们现在可以证明关于子空间嵌入的主要结果，该结果由定理 3 给出。与定理 1 的证明类似，我们定义如下记号。

记号 2。令 $x \in \mathbb{R}^d$。对于每个 $\ell \in [d]$，将 $x_{heavy(\ell)}$ 记为由 x 得到的向量，其中所有幅值不严格大于 $1/\sqrt{\ell}$ 的条目都置为零。记 $x_{light(\ell)} = x - x_{heavy(\ell)}$。

令 $\mathcal{N}_{\frac{1}{2}}$为 V中单位球上的一个 $1/2$-网，并定义 $\mathcal{N}_{\frac{1}{2}}^{+} = \mathcal{N}_{\frac{1}{2}} \cup \{x+y:\ x, y \in \mathcal{N}_{\frac{1}{2}} \cup \{0\}\}$。可以构造一个 $1/2$-网，使得 $|\mathcal{N}_{\frac{1}{2}}| \leq 4^k$。令 $n = |\mathcal{N}_{\frac{1}{2}}^{+}| \leq 8^k$。令 $\ell = \left\lceil \min \left\{ \varepsilon^{-1/2},\ \left(\frac{\lg n}{\lg k}\right)^3 \right\} \right\rceil$ 为一个整数，并令 S 按引理 4 中的定义给出。我们将 $\mathcal{Y}$ 定义为满足 $\text{supp}(y) \subseteq S$，$|\text{supp}(y)| \leq \ell$ 且 $\|y\| \leq 1$的所有向量 $y \in \mathbb{R}^d$ 构成的集合。

定义 $\mathcal{E}_1$ 为满足以下条件的所有矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in \mathcal{Y}$, $\|Ax\|^2 \in (1 \pm \varepsilon)\|x\|^2$。定义 $\mathcal{E}_2$为满足以下条件的所有矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in \mathcal{N}_{\frac{1}{2}}^{+}$, $\|Ax_{light(\ell)}\|^2 \in \|x_{light(\ell)}\|^2 \pm \varepsilon$。定义 $\mathcal{E}_3$为满足以下条件的所有矩阵 $A \in \mathbb{R}^{m \times d}$ 的集合：对于所有 $x \in \mathcal{N}_{\frac{1}{2}}^{+}$, $\left|\left\langle Ax_{heavy(\ell)}, Ax_{light(\ell)} \right\rangle\right| < \varepsilon$。

断言 20. 假设 $A \in \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$。那么对于每个单位向量 $x \in V$, $\|Ax\|^2 \in (1 \pm O(\varepsilon))$。

证明。根据引理 19，并利用 A的线性性，只需对 $\mathcal{N}_{\frac{1}{2}}^{+}$ 中的所有向量 z 证明该断言。因此，令 x 为任意一个这样的单位向量。那么 $\|Ax\|^2 = \|Ax_{heavy(\ell)}\|^2 + \|Ax_{light(\ell)}\|^2 + 2\left\langle Ax_{heavy(\ell)}, Ax_{light(\ell)} \right\rangle$。由于 $\|x\| = 1$，且 $x_{heavy(\ell)}$ 的每个元素的大小至少为 $1/\sqrt{\ell}$，根据 S 的定义，我们有 $\text{supp}(x_{heavy(\ell)}) \subseteq S$ 且 $|\text{supp}(x_{heavy(\ell)})| \leq \ell$。因此 $x_{heavy(\ell)} \in \mathcal{Y}$，于是

$$\|Ax\|^2 \leq (1+\varepsilon)\|x_{heavy(\ell)}\|^2 + (1+\varepsilon)\|x_{light(\ell)}\|^2 + \varepsilon + 2\varepsilon \leq (\|x\|^2 + O(\varepsilon))$$

同理

$$\|Ax\|^2 \geq (1-\varepsilon)\|x_{heavy(\ell)}\|^2 + (1-\varepsilon)\|x_{light(\ell)}\|^2 - \varepsilon - 2\varepsilon \geq (\|x\|^2 - O(\varepsilon))$$

□

与定理 1 的证明一样，接下来只需给出 $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ 的概率下界。再次分别对每个事件的概率进行界定。

断言 21. $\Pr[A \in \mathcal{E}_1] \geq 1 - 2^{-k^{2/3}}$。

证明。记 $\delta = 2^{-\Omega(\sqrt[3]{\lg^2 n \cdot \lg k})}$。我们得到 $m \geq \Omega(\varepsilon^{-2}\lg(1/\delta))$ 且 $s \geq \Omega(\varepsilon^{-1}\lg(1/\delta))$。根据 Kane 和 Nelson 的结果 [KN14],，对每个单位向量 $x \in \mathbb{R}^d$，有 $\Pr\left[\|Ax\|^2 \in (1 \pm O(\varepsilon))\right] \geq 1-\delta$。

接下来，对每个满足 $|T| = \ell$ 的 $T \subseteq S$，令 $\mathcal{Y}_T = \{y \in \mathbb{R}^d:\ \|y\| \leq 1 \ \ and \text{supp}(y) \subseteq T\}$，则 $\mathcal{Y}_T$ 是 $\mathbb{R}^d$ 中一个 ℓ 维子空间的单位球，因此存在一个针对 $\mathcal{Y}_T$ 的 $1/2$-网 $\hat{} \mathcal{Y}_T$，使得 $|\hat{}\mathcal{Y}_T| \leq 4^\ell$。注意，在这些记号中，$\mathcal{Y} = \bigcup_{T \in \binom{S}{\ell}} \mathcal{Y}_T$，并另外记 $\hat{}\mathcal{Y} = \bigcup_{T \in \binom{S}{\ell}} \hat{} \mathcal{Y}_T$。那么

$$|\hat{\mathcal{Y}}| \leq \binom{|S|}{\ell} 4^\ell \leq (4ek)^\ell = 2^{\Omega(\ell \lg(4ek))}.$$

对于 $k > 1$，我们有 $|\hat{}\mathcal{Y}| \leq 2^{\Omega(\ell \text{对数 } k)} = 2^{\Omega(\sqrt[3]{\text{对数}^2 n \text{对数 } k)} = \delta^{-1/2}}$。因此，以至少 $1 - \sqrt{\delta} \geq 1 - 2^{-k^{2/3}}$ 的概率，我们得到对所有 $y \in \hat{}\mathcal{Y}$, $\|Ay\|^2 \in (1 \pm O(\varepsilon))$ 均成立。

因此，假设对所有 $y \in$ 令 $x \in \mathcal{Y}$，则存在 $T \subseteq S$ 使得 $\hat{}|T| = \ell$ 且 $\text{supp}(x) \subseteq T$，因此 $x \in \mathcal{Y}_T$。由于 $\mathcal{Y}_T$ 是 $\mathcal{Y}_T$ 的一个 $1/2$-网，且 $\mathcal{Y}_T$ 是 $\mathbb{R}^d$ 中一个 ℓ 维子空间的单位球，引理 19 表明 $\|Ax\|^2 \in (1 \pm O(\varepsilon))\|x\|^2$。因此 $\Pr[A \in \mathcal{E}_1] \geq 1 - 2^{-k^{2/3}}$。 □

下面的论断完成了定理 3 的证明。证明 $\mathcal{E}_2$ 和 $\mathcal{E}_3$的概率界分别类似于论断 9 和 10 的证明，因此省略。

断言 22。$\Pr[A \in \mathcal{E}_2] \geq 1 - \frac{1}{n}$ and $\Pr[A \in \mathcal{E}_3] \geq 1 - \frac{1}{n}$。

参考文献

[AC09] N. Ailon 和 B. Chazelle。《快速 Johnson-Lindenstrauss 变换与近似最近邻》。SIAM 计算杂志，39(1):302–322，2009。[AHK06] S. Arora、E. Hazan 和 S. Kale。《用于矩阵稀疏化的快速随机采样算法》。载于第 10 届随机化与计算国际研讨会（RANDOM）论文集，第 272–279 页，2006。[BKT18] A. R. Benson、R. Kumar 和 A. Tomkins。《用于子集选择的离散选择模型》。载于第十一届 ACM 网络搜索与数据挖掘国际会议（WSDM）论文集。ACM，2018。[CCF04] M. Charikar、K. C. Chen 和 M. Farach-Colton。《在数据流中查找频繁项》。理论计算机科学，312(1):3–15，2004。[Coh16] M. B. Cohen。《通过迹不等式实现近乎紧致的无关子空间嵌入》。载于第二十七届 ACM-SIAM 离散算法研讨会（SODA）论文集，第 278–287 页，2016。[CW13] K. L. Clarkson 和 D. P. Woodruff。《输入稀疏时间内的低秩近似与回归》。载于第 45 届 ACM 计算理论研讨会（STOC）论文集，第 81–90 页，2013。[DGC+09] T. T. Do、L. Gan、Y. Chen、N. Nguyen 和 T. D. Tran。《使用结构随机矩阵实现快速高效的降维》。载于 2009 年 IEEE 声学、语音与信号处理国际会议论文集，第 1821–1824 页，2009。doi:10.1109/ICASSP.2009.4959960。[DKT17] S. Dahlgaard、M. Knudsen 和 M. Thorup。《用于相似度估计和降维的实用哈希函数》。载于《神经信息处理系统进展》第 30 卷，第 6615–6625 页。Curran Associates, Inc.，2017。[FHL22] O. N. Fandina、M. M. Høgsgaard 和 K. G. Larsen。《快速 Johnson-Lindenstrauss 变换还能更快》，2022。可从以下网址获取：https://arxiv.org/abs/2204.01800。[FKL18] C. B. Freksen、L. Kamma 和 K. G. Larsen。《全面理解哈希技巧》。载于《神经信息处理系统进展》第 31 卷：神经信息处理系统 2018 年度会议，NeurIPS 2018，2018年12月3日至8日，蒙特利尔，加拿大，第 5394–5404 页，2018。[FL20] C. B. Freksen 和 K. G. Larsen。《关于使用 Toeplitz 矩阵和循环矩阵进行 Johnson-Lindenstrauss 变换》。Algorithmica，82(2):338–354，2020。[GBH09] A. Go、R. Bhayani 和 L. Huang。《利用远程监督进行 Twitter 情感分类》。Processing，150，2009 年 1 月。[Hoe63] W. Hoeffding。《有界随机变量之和的概率不等式》。美国统计学会会刊，58:13–30，1963。[Jag19] M. Jagadeesan。《理解用于特征哈希的稀疏 JL》。载于《神经信息处理系统进展》第 32 卷：神经信息处理系统 2019 年度会议，NeurIPS 2019，2019年12月8日至14日，温哥华，不列颠哥伦比亚省，加拿大，第 15177–15187 页，2019。[JL84] W. Johnson 和 J. Lindenstrauss。《将 Lipschitz 映射扩展到希尔伯特空间》。载于《现代分析与概率会议》（康涅狄格州纽黑文，1982），《当代数学》第 26 卷，第 189–206 页。美国数学会，1984。

[JPS20] V. Jain、N. S. Pillai 和 A. Smith。《Kac meets johnson and lindenstrauss：一种内存最优的快速 Johnson-Lindenstrauss 变换》。CoRR，abs/2003.10069，2020 年。[JW13] T. S. Jayram 和 D. P. Woodruff。《Johnson-Lindenstrauss 变换以及具有次常数误差的流式问题的最优界》。ACM 算法汇刊，9(3):26:1–26:17，2013 年。[KN14] D. M. Kane 和 J. Nelson。《更稀疏的 Johnson-Lindenstrauss 变换》。J. ACM，61(1):4:1–4:23，2014 年。[KW11] F. Krahmer 和 R. Ward。《通过受限等距性质实现的新型及改进的 Johnson–Lindenstrauss 嵌入》。SIAM 数学分析杂志，43(3):1269–1281，2011 年。[LL22] Y. Li 和 M. Liu。《稀疏无关子空间嵌入的下界》。载于第 41 届年度数据管理国际会议（PODS）论文集，第 251–260 页，2022 年。[LN17] K. G. Larsen 和 J. Nelson。《 Johnson-Lindenstrauss 引理的最优性》。载于第 58 届 IEEE 年度计算机科学基础研讨会（FOCS 2017），美国加利福尼亚州伯克利，2017 年 10 月 15–17 日，第 633–638 页，2017 年。[MS90] S. Montgomery-Smith。《Rademacher 和的分布》。美国数学学会会刊，109:517–522，1990 年。[NN13a] J. Nelson 和 H. L. Nguyen。《降维映射的稀疏度下界》。载于第 45 届年度 ACM 计算理论研讨会（STOC）论文集，第 101–110 页，2013 年。[NN13b] J. Nelson 和 Nguyˆen。《OSNAP：通过更稀疏的子空间嵌入实现更快速的数值线性代数算法》。载于第 54 届年度 IEEE 计算机科学基础研讨会（FOCS）论文集，第 117–126 页，2013 年。[NN14] J. Nelson 和 H. L. Nguyˆen。《无关子空间嵌入的下界》。载于第 41 届自动机、语言与编程国际研讨会（ICALP）论文集，第 883–894 页，2014 年。[Sar06] T. Sarl´os。《通过随机投影改进大型矩阵的近似算法》。载于第 47 届年度 IEEE 计算机科学基础研讨会（FOCS）论文集，第 143–152 页，2006 年。[WDL+09] K. Weinberger、A. Dasgupta、J. Langford、A. Smola 和 J. Attenberg。《用于大规模多任务学习的特征哈希》。载于第 26 届年度国际机器学习会议（ICML '09）论文集，第 1113–1120 页，2009 年。[Woo14] D. P. Woodruff。《草图化：数值线性代数的工具》。理论计算机科学基础与趋势，10(1-2):1–157，2014 年。

观察15和17的证明

为求完整起见，我们证明以下引理，该引理可推出观察15和17。

引理23. 设 $I \subseteq [m] \times [d]$，并令 $\{c_{(i,j)}\}_{(i,j)\in I}$ 为一组非负常数。对于每个 $(i,j) \in I$，定义 b_{ij} 为伯努利随机变量，当且仅当 $a_{ij} \neq 0$ 时取值为 1，则

$$\mathbb{E}\left[\exp\left(\sum_{(i,j)\in I} c_{(i,j)} b_{ij}\right)\right] \leq \prod_{(i,j)\in I} \mathbb{E}[\exp(c_{(i,j)} b_{ij})]$$

证明. 回想一下， A 的行被划分为 s 个块 I_1、.....、I_s，每块包含 m/s 个连续行。也就是说，对于每个 $p \in [s]$， $I_p = [(p-1)(m/s)+1, pm/s]$。采用这些记号，

$$\sum_{(i,j)\in I} c_{(i,j)} b_{ij} = \sum_{j\in[d]} \sum_{p\in S} \sum_{i\in I_p:(i,j)\in I} c_{(i,j)} b_{ij} .$$

由于列 A的列以及每列中的不同块彼此独立，因此我们得到

$$\mathbb{E}\left[\exp\left(\sum_{(i,j)\in I} c_{(i,j)} b_{ij}\right)\right] \le \prod_{j\in[d]} \prod_{p\in[s]} \mathbb{E}\left[\prod_{i\in I_p:(i,j)\in I} \exp(c_{(i,j)} b_{ij})\right]$$

固定 $j \in [d]$ 和 $p \in [s]$，并记 $C = \{i \in I_p : (i,j) \in I\}$。对于每个 $i \in C$, $c_{(i,j)} \ge 0$，因此 $e^{c_{i,j}} \ge 0$。所以

$$\begin{aligned}\mathbb{E}\left[\prod_{i\in C} \exp(c_{(i,j)} b_{ij})\right] &= \sum_{i\in C} \frac{e^{c_{(i,j)}}}{(m/s)} + (1 - \frac{|C|}{(m/s)}) = 1 + \sum_{i\in C} \frac{e^{c_{(i,j)}} - 1}{(m/s)} \\ &\le \prod_{i\in C} \left(1 + \frac{e^{c_{(i,j)}} - 1}{(m/s)}\right) = \prod_{i\in C} \mathbb{E}(\exp(c_{(i,j)} b_{ij}))\end{aligned}$$

□

B A 1/2-网就足够了

这里给出引理 19 的延后证明

引理 19 的证明。设 $x \in V$为一个单位向量。我们通过归纳构造一个由向量 $\{x_i\}_{i=0}^{\infty}$ 组成的序列，它们属于$\mathcal{N}_{\frac{1}{2}}$；并构造一个由非负实数 $\{\alpha_i\}_{i=0}^{\infty}$ 组成的序列，使得 $x = \sum_{i=0}^{\infty} \alpha_i x_i$，并且对所有 $i \ge 0$，还有 $\alpha_i \le 2^{-i}$。设 x_0 为 x在 $\mathcal{N}_{\frac{1}{2}}$中的最近向量，并令 $\alpha_0 = 1$。那么 $x = \alpha_0 x_0 + (x - \alpha_0 x_0)$。显然，若 $x - \alpha_0 x_0 = 0$，则结论成立，因为我们可以对所有 $i \ge 1$定义 $\alpha_i = 0$。否则，记 $\alpha_1 = \|x - \alpha_0 x_0\|$ 和$v_1 = \alpha_1^{-1}(x - \alpha_0 x_0)$，则 $\alpha_1 \le 1/2$, v_1 是一个单位向量，并且 $x = \alpha_0 x_0 + \alpha_1 v_1$。接下来进行归纳，令$p \in \mathbb{N}$，并假设存在向量 $x_0, \ldots, x_p \in \mathcal{N}_{\frac{1}{2}}$、数 $\alpha_0, \ldots, \alpha_{p+1}$ 以及一个单位向量 v_{p+1}，使得 $x = \sum_{i=0}^{p} \alpha_i x_i + \alpha_{p+1} v_{p+1}$，且对所有 $i \le p+1$ 都有 $\alpha_i \le 2^{-i}$。设 x_{p+1} 为 v_{p+1}在 $\mathcal{N}_{\frac{1}{2}}$中的最近向量。则 $v_{p+1} = x_{p+1} + (v - x_{p+1})$。若 $v - x_{p+1} = 0$，则结论成立，因为我们可以对所有 $i \ge p+2$定义 $\alpha_i = 0$。否则，记 $\beta = \|v - x_{p+1}\|$, $\alpha_{p+2} = \alpha_{p+1}\beta$ 和 $v_{p+2} = \beta^{-1}(v - x_{p+1})$，则 $\alpha_{p+2} \le 2^{-p+1}$, v_{p+2} 是一个单位向量，并且

$$x = \sum_{i=0}^{p} \alpha_i x_i + \alpha_{p+1} v_{p+1} = \sum_{i=0}^{p} \alpha_i x_i + \alpha_{p+1}(x_{p+1} + (v - x_{p+1})) = \sum_{i=0}^{p+1} \alpha_i x_i + \alpha_{p+2} v_{p+2} .$$

至此，序列的构造完成。接下来注意到

$$\|x\|^2 = \left\|\sum_{i=0}^{\infty} \alpha_i x_i\right\|^2 = \sum_{i=0}^{\infty} \alpha_i \|x_i\|^2 + \sum_{i<j} 2\alpha_i \alpha_j x_i^t x_j .$$

同理，我们得到

$$\|Ax\|^2 = \left\|\sum_{i=0}^{\infty} \alpha_i A x_i\right\|^2 = \sum_{i=0}^{\infty} \alpha_i \|Ax_i\|^2 + \sum_{i<j} 2\alpha_i \alpha_j x_i^t A^t A x_j .$$

由于对所有 i 都有 $x_i \in \mathcal{N}_{\frac{1}{2}} \subseteq \mathcal{N}^+_{\frac{1}{2}}$，因此我们有 $\|Ax_i\|^2 \in 1 \pm O(\varepsilon)$。此外，对所有 $i < j$，$2x_i^t x_j = \|x_i + x_j\|^2 - \|x_i\|^2 - \|x_j\|^2$。由于 x_i、x_j、$x_i + x_j \in \mathcal{N}^+_{\frac{1}{2}}$，我们有

$$2x_i^t A^t A x_j = \|Ax_i + Ax_j\|^2 - \|Ax_i\|^2 - \|Ax_j\|^2 = \|A(x_i + x_j)\|^2 - \|Ax_i\|^2 - \|Ax_j\|^2 \in 2x_i^t x_j \pm O(\varepsilon)$$

因此

$$\begin{aligned}
\|Ax\|^2 &= \sum_{i=0}^{\infty} \alpha_i \|Ax_i\|^2 + \sum_{i<j} 2\alpha_i \alpha_j x_i^t A^t A x_j \\
&\in \sum_{i=0}^{\infty} \alpha_i (\|x_i\|^2 \pm O(\varepsilon)) + \sum_{i<j} 2\alpha_i \alpha_j (x_i^t x_j \pm O(\varepsilon)) \\
&\subseteq \sum_{i=0}^{\infty} \alpha_i \|x_i\|^2 + \sum_{i<j} 2\alpha_i \alpha_j x_i^t x_j + O(\varepsilon) \left(\sum_{i=0}^{\infty} \alpha_i + \sum_{i<j} 2\alpha_i \alpha_j \right) \subseteq 1 \pm O(\varepsilon)
\end{aligned}$$

□