Sparsity-Dimension Trade-Offs for Oblivious Subspace Embeddings

YI LI*, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
MINGMOU LIU*, School of Intelligent Software and Engineering, Nanjing University, China

An oblivious subspace embedding (OSE), characterized by parameters $m, n, d, \epsilon, \delta$, is a random matrix $\Pi \in \mathbb{R}^{m\times n}$ such that for any d-dimensional subspace $T \subseteq \mathbb{R}^n$, $\Pr_\Pi[\forall x \in T, (1-\epsilon)\|x\|_2 \leq \|\Pi x\|_2 \leq (1+\epsilon)\|x\|_2] \geq 1-\delta$. When an OSE has $s \leq 1/2.001\epsilon$ nonzero entries in each column, we show it must hold that $m = \Omega(d^2/(\epsilon^2 s^{1+O(\delta)}))$, which is the first lower bound with multiplicative factors of d^2 and $1/\epsilon$, improving on the previous $\Omega(d^2/s^{O(\delta)})$ lower bound due to Li and Liu (PODS 2022). When an OSE has $s = \Omega(\log(1/\epsilon)/\epsilon)$ nonzero entries in each column, we show it must hold that $m = \Omega((d/\epsilon)^{1+1/4.001\epsilon s}/s^{O(\delta)})$, which is the first lower bound with multiplicative factors of d and $1/\epsilon$, improving on the previous $\Omega(d^{1+1/(16\epsilon s+4)})$ lower bound due to Nelson and Nguyễn (ICALP 2014). This second result is a special case of a more general trade-off among d, ϵ, s, δ and m.

CCS Concepts: • **Theory of computation → Random projections and metric embeddings**.

Additional Key Words and Phrases: oblivious subspace embedding, sparse subspace embedding, lower bound

ACM Reference Format:
Yi Li and Mingmou Liu. 2026. Sparsity-Dimension Trade-Offs for Oblivious Subspace Embeddings. *Proc. ACM Manag. Data* 4, 2 (PODS), Article 116 (May 2026), 25 pages. https://doi.org/10.1145/3801912

1 Introduction

Subspace embedding is an essential dimensionality reduction tool for processing large datasets, such as clustering [2, 7], performing correlation analysis [1], and solving basic linear algebraic problems like regression and low-rank approximation [5]. Formally, given a subspace $T \subseteq \mathbb{R}^n$, a linear map $\Pi : \mathbb{R}^n \to \mathbb{R}^m$ (or, equivalently, a matrix $\Pi \in \mathbb{R}^{m\times n}$) is called an ϵ-subspace-embedding for T if $(1-\epsilon)\|x\|_2 \leq \|\Pi x\|_2 \leq (1+\epsilon)\|x\|_2$ for all $x \in T$ simultaneously. In this definition, Π may depend on the subspace T; however, in many applications, the data can be dynamically updated (and thus T changes), so it is preferable to have a Π that is independent of T, in which case Π is called an oblivious subspace embedding. It is clear that a fixed Π cannot be a subspace embedding for all subspaces T; a more reasonable definition thus includes randomness in Π. The formal definition is given below.

Definition 1. *An (n, d, ϵ, δ)-oblivious-subspace-embedding is a random matrix $\Pi \in \mathbb{R}^{m\times n}$ such that for any fixed d-dimensional subspace $T \subseteq \mathbb{R}^n$, it holds that*

$$\Pr\left[\forall x \in T, (1-\epsilon)\|x\|_2 \leq \|\Pi x\|_2 \leq (1+\epsilon)\|x\|_2\right] \geq 1-\delta. \tag{1}$$

*Both authors contributed equally to this research.

Authors' Contact Information: Yi Li, yili@ntu.edu.sg, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore; Mingmou Liu, lmm@nju.edu.cn, School of Intelligent Software and Engineering, Nanjing University, Suzhou, Jiangsu, China.

ACM 2836-6573/2026/5-ART116
https://doi.org/10.1145/3801912

In most applications, T is the column space of data matrix $A \in \mathbb{R}^{n\times d}$, so Eq. (1) can be written as

$$\Pr\left[\forall x \in \mathbb{R}^d, (1-\epsilon)\|Ax\|_2 \le \|\Pi Ax\|_2 \le (1+\epsilon)\|Ax\|_2\right] \ge 1-\delta. \tag{2}$$

When n and d are clear from the context, we may omit them in the parameters. We shall consider only oblivious subspace embeddings, so we will also drop the word "oblivious" and simply say "(ϵ,δ)-subspace-embedding".

A typical application of oblivious subspace embedding is linear regression, namely, solving $\min_{x\in\mathbb{R}^d} \|Ax-b\|_2$ for given $A \in \mathbb{R}^{n\times d}$ and $b \in \mathbb{R}^n$ with $n \gg d$. Suppose Π is an (ϵ,δ)-subspace-embedding for the column space of the concatenated matrix $(A\ b)$. This subspace has dimension at most $d+1$, so Π can have only $\mathrm{poly}(d/\epsilon) \ll n$ rows. The subspace embedding property implies that $\|\Pi Ax - \Pi b\|_2 = \|\Pi(Ax-b)\|_2 = (1\pm\epsilon)\,\|Ax-b\|_2$ for all x. Thus, the original large-scale problem $\|Ax-b\|_2$ is reduced to a much smaller one $\|\Pi Ax - \Pi b\|_2$, which can be solved much faster.

The design of Π aims to minimize the number m of rows while allowing ΠA to be computed efficiently. It has long been known that the optimal dimension is $m = \Theta((d+\log(1/\delta))/\epsilon^2)$ [11], but constructions achieving this bound are dense matrices and therefore slow to apply. Two main approaches have been developed to improve runtime: using structured transforms (e.g. FFT-based), or using sparse matrices with at most s nonzeros per column so that ΠA can be computed in $O(s\cdot \mathrm{nnz}(A))$ time. We focus on the latter approach, namely sparse oblivious subspace embeddings.

The sparsest case $s=1$ is now completely understood. The classical Count-Sketch construction places exactly one nonzero entry in each column at a uniformly random position, with its value given by a uniformly random sign in $\{-1,1\}$. It is known that $m = O(d^2/(\epsilon^2\delta))$ rows suffice for an (ϵ,δ)-subspace-embedding [9]. The quadratic dependence on d has long been known to be necessary [10], and recently Li and Liu established a fully tight bound of $m = \Omega(d^2/(\epsilon^2\delta))$ [8].

For larger s, the trade-off between sparsity and embedding dimension remains less well understood. The OSNAP construction [6, 9] shows (i) $m = O(d\log(d/\delta)/\epsilon^2)$ when $s = \Theta(\log(d/\delta)/\epsilon)$, or (ii) $m = O(d^{1+\gamma}\log(d/\delta)/\epsilon^2)$ when $s = \Theta(1/(\gamma\epsilon))$ for any constant $\gamma > 0$. Since the general lower bound is $m = \Omega((d+\log(1/\delta))/\epsilon^2)$ [11], case (i) is tight up to a logarithmic factor. Recent progress has largely closed this logarithmic gap. It was shown in [4] that $s = \Theta(\log^4(d/\delta)/\epsilon^6)$, with higher exponents of logarithmic factors and $1/\epsilon$, can achieve $m = O((d+\log(1/\delta))/\epsilon^2)$, which is optimal up to a constant factor. This was later improved in [3], for $\epsilon = e^{-\Omega(d)}$, to $s = O(\log^2(d/\delta)/\epsilon + \log^3(d/\delta))$ and optimal $m = O((d+\log(1/\delta))/\epsilon^2)$.

In case (ii), as s decreases by a logarithmic factor, m significantly increases while still maintaining a subquadratic dependence on d, namely, $d^{1+1/\Theta(\epsilon s)}$. This form matches the lower bound of $\Omega(d^{1+1/(16\epsilon s+4)})$ [11][1]. However, when s is a constant factor smaller, say, $s \le \alpha/\epsilon$ for some constant $\alpha > 0$, a quadratic dependence on d becomes necessary: it is known that $m = \Omega(\epsilon^2 d^2)$ for constant δ [11], improved to $m = \Omega(\epsilon^{O(\delta)} d^2)$ in [8]. Since Count-Sketch ($s=1$) already attains the tight bound $m = \Theta(d^2/\epsilon^2)$, no substantially better construction is known in the regime $s \le \alpha/\epsilon$, leaving a large gap between the upper and lower bounds.

We conjecture that a tight lower bound $m = \Omega(d^2/\epsilon^2)$ exists when $s \le \alpha/\epsilon$ and a lower bound $m = \Omega(d^{1+1/\Theta(\epsilon s)}/\epsilon^2)$ exists when $s \ge \alpha/\epsilon$, so there is a sharp transition for $s \in (\epsilon^{-1}, O(\epsilon^{-1}\log d))$. As a first step towards this goal, we ask the following question:

Is it possible to obtain a lower bound including both d^2 (resp. $d^{1+1/\Theta(\epsilon s)}$) and $1/\epsilon$ as multiplicative factors, such as $\Omega(d^2/\epsilon^{0.1})$ (resp. $\Omega(d^{1+1/\Theta(\epsilon s)}/\epsilon^{0.1})$), when $s \le \alpha/\epsilon$ (resp. $s \ge \alpha/\epsilon$)?

Our Results. In this paper, we answer the question affirmatively by providing two trade-offs between m and s. These are the first lower bounds with factors of both $d^{1+1/\Theta(\epsilon s)}$ and $1/\epsilon$.

[1]The constants such as 16 and 4 are not specified in the theorem statement in [11] but can be determined through a careful examination of the proof.

The first result states that any (ϵ, δ)-subspace-embedding with column sparsity at most $(1/2 - o(1))\epsilon$ must have $\Omega(d^2/(s^{1+O(\delta)}\epsilon^2 \operatorname{polylog} s))$ rows. The result is formally stated below.

THEOREM 2. *There exist absolute constants $\epsilon_0, \delta_0, c_0, K_0, K_1 > 0$ such that the following holds. For all $\epsilon \in (0, \epsilon_0)$, $\delta \in (0, \delta_0)$, $d \geq 1/\epsilon^2$ and $n \geq K_0 d^2/(\epsilon^2\delta)$, any (ϵ, δ)-subspace-embedding Π must have*

$$m \geq c_0 d^2/(\epsilon^2 s^{1+K_1\delta} \log^{13} s)$$

rows if the column sparsity s of Π satisfies $3 \leq s \leq \big((1-\epsilon)^2 + (2\epsilon - \epsilon^2)/\log s + 1/\log^2 s\big)/(2\epsilon)$.

Note that the upper bound for s is $(1/2 - o(1))\epsilon$ if $\epsilon = o(1)$ and $s = \omega(1)$.

Our second result is a general lower bound for sparse OSE: any OSE with column sparsity s must have $\Omega((d/s\epsilon)^{1+1/(\lfloor (4+o(1))\epsilon s\rfloor+1)} s^{1-O(\delta)}/\operatorname{polylog} s)$ rows.

THEOREM 3. *There exist absolute constants $\epsilon_0, \delta_0, c_0, K_0, K_1, K_2 > 0$ such that the following holds. For all $\epsilon \in (0, \epsilon_0)$, $\delta \in (0, \delta_0)$, $d \geq 1/\epsilon^7$, $n \geq K_0 d^2/(\epsilon^2\delta)$, any (ϵ, δ)-subspace-embedding Π must have*

$$m \geq \frac{c_0 \cdot \theta^{1/(\chi-1)}}{s^{K_1\delta\chi/(\chi-1)} \log^{6/(\chi-1)+2} s} \cdot \left(\frac{d}{\epsilon}\right)^{1+1/(\chi-1)},$$

rows, if the column sparsity s of Π satisfies $3 \leq s \leq K_2\epsilon^{-1}\log d$, where $\chi = \lfloor (4\epsilon+\theta)/\theta(1-1/\log s)\rfloor + 1$ and $\theta \geq (1-\epsilon)^2/s - 1/s\log s$.

Note that $4\epsilon/\theta = 4\epsilon s(1 + 2\epsilon/(1-\epsilon)^2 + O(1/\log s))$, and $\chi = \lfloor (4 + o(1))\epsilon s\rfloor + 2$ if $s = \omega(1)$ and $\epsilon = o(1)$. Thus for $\epsilon \leq 1/202$ and $s = \Omega(\log(1/\epsilon)/\epsilon) \cap O((\log d)/\epsilon)$, our lower bound is at least

$$m = \Omega\left(\left(\frac{d}{\epsilon}\right)^{1+\frac{1}{4(1+2\epsilon)\epsilon s}} \cdot s^{-O(\delta)}\right),$$

which is better than the previous $\Omega\left(d^{1+1/(16\epsilon s+4)}\right)$ lower bound of Nelson and Nguyễn [11] in the dependence of both d and ϵ.

We remark that the first lower bound is significantly stronger than the second when $1/4\epsilon < s < 1/2\epsilon$, as the two arguments rely on different contradiction principles. We further remark that the same lower bound can be obtained for $s = 1, 2$ (except that $\log s$ is replaced with some constant) with a minor modification to our proof.

2 Technique overview

We begin by applying Yao's minimax principle to construct a matrix distribution $\mathcal{D}$ (called the *hard instance*) and to show that any deterministic (ϵ, δ)-subspace-embedding Π for the column space of a random matrix $U \sim \mathcal{D}$, i.e. Eq. (2) holds for $U \sim \mathcal{D}$ and deterministic Π, must have many rows.

We adopt the hard instance $\widetilde{\mathcal{D}}$ from [8], formally defined in Definition 7. Informally, $U \sim \widetilde{\mathcal{D}}$ has a random parameter $\beta \in (0, 1]$ and takes the form $U = VW$. We denote $U \sim \mathcal{D}_\beta$ when β is fixed, and the overall construction intuitively forces Π to be an OSE for $U \sim \mathcal{D}_\beta$ for most values of β. The matrix V has d/β columns and each column $(\Pi V)_{\star,i}$ is a copy of a random column of Π.[2] The matrix W is a random matrix ensuring that any large inner product (in absolute value) between two columns of ΠV makes it unlikely that Π is a subspace embedding for $U = VW$ (see Lemma 8). Thus, as in [8, 11], the key task in the sparse case is to find a large inner product among the columns of ΠV.

Specifically, the proof for Theorem 2 identifies a pair of columns of ΠV with inner product exceeding 2ϵ. The random W randomizes $\|\Pi U v\|_2^2$ for some unit vector v, causing $\|\Pi U v\|_2^2$ to

[2]For a matrix A, we denote its i-th column vector by $A_{\star,i}$.

fluctuate over a range of length larger than 4ϵ. This contradicts the OSE guarantee that $\|\Pi U v\|_2^2 \in (1 \pm \epsilon)^2$, which is a range of length exactly 4ϵ.

For the dense case, the proof follows the some intuition but considers all pairs from $\chi \geq 2$ columns rather than just a single pair. We seek χ columns $X_1, \ldots, X_\chi$ of ΠV such that $\sum_{i \neq j} \langle X_i, X_j \rangle \geq 4\epsilon\chi$. Here, the random matrix W is no longer useful (the randomnization trick does not work for large χ). Instead, we show that $\sum_i \|X_i\|_2^2 \geq (1-\epsilon)^2\chi$ and then choose an appropriate unit vector v such that $\|\Pi U v\|_2^2 \cdot \chi = \|\sum_i X_i\|_2^2 = \sum_i \|X_i\|_2^2 + \sum_{i \neq j} \langle X_i, X_j \rangle > (1+\epsilon)^2\chi$, violating the OSE guarantee.

We remark that the strategy for Theorem 2 naturally leads to an upper bound of $1/(2\epsilon)$ for s, while the strategy for Theorem 3 yields a phase transition around $s = 1/(4\epsilon)$ from quadratic to subquadratic dependence on d.

In the following sections, we shall describe our approach for the sparse case in Sections 2.1 to 2.3, followed by the dense case in Section 2.4. Although the limitations in Section 2.1 and the intuition in Section 2.2 apply to both cases, for simplicity, they are described in a form tailored to sparse case.

2.1 Limitations of previous approaches

The previous proofs [8, 11] adopt the following strategy. For a row k of Π, let $S_k = \{j : \Pi_{k,j} \geq \sqrt{\kappa}\}$, where $\kappa \in (0,1)$. As noted earlier, we aim show that there exist two columns of ΠV with a large inner product. More precisely, the goal is to argue that $\langle (\Pi V)_{\star,i}, (\Pi V)_{\star,j} \rangle$ is large with a good small probability when $(\Pi V)_{\star,i}$ and $(\Pi V)_{\star,j}$ are drawn uniformly from the columns in S_k.

To see this, we split a vector into the k-th coordinate and the remaining coordinates, writing

$$\langle (\Pi V)_{\star,i}, (\Pi V)_{\star,j} \rangle = (\Pi V)_{k,i}(\Pi V)_{k,j} + \langle (\Pi'_k V)_{\star,i}, (\Pi'_k V)_{\star,j} \rangle \geq \kappa + \langle (\Pi'_k V)_{\star,i}, (\Pi'_k V)_{\star,j} \rangle,$$

where Π'_k is obtained from Π by zeroing out the k-th row. Thus, it suffices to show that the remainder $X = \langle (\Pi'_k V)_{\star,i}, (\Pi'_k V)_{\star,j} \rangle$ would not be too negative to cancel κ. For example, it can be shown that

$$\Pr[X \geq -\kappa/2] = \Omega(\kappa). \tag{3}$$

Taking $\kappa = \Theta(\epsilon)$ implies that two random columns of ΠV in S_k have inner product $\Omega(\epsilon)$ with probability $\Omega(\epsilon)$. It then requires finding $\Omega(1/\epsilon)$ such pairs of columns (which may come from S_k for different values of k) to obtain a large inner product with constant probability. A simple greedy strategy in [11] finds $\Omega(\sqrt{d^2/m})$ pairs, leading to a lower bound of $m = \Omega(\epsilon^2 d^2)$; a more sophisticated greedy strategy in [8] finds many more pairs and improves this to $\Omega(\epsilon^{\Theta(\delta)} d^2)$.

We remark that Eq. (3) cannot be improved for a general matrix Π. Furthermore, an example was given in [8, Remark 10], showing that this strategy cannot prove a lower bound beyond $\Omega(d^2)$. The example takes Π to be a horizontal concatenation of copies of $d^2 \times d^2$ block diagonal matrix, where each diagonal block is $\sqrt{8\epsilon}H$ with H being a Hadamard matrix of order $1/(8\epsilon)$. Hence, every nonzero entry of Π is $\pm\sqrt{8\epsilon}$. Assume that δ is a constant. The hard instance is $\mathcal{D}_1$ and the matrix V has thus d columns. The previous strategy applies Eq. (3) with $\kappa = 8\epsilon$. Now, two random columns of ΠV belong to the same S_k with probability $\Theta(1/(\epsilon d^2))$; when this occurs, they have a large inner product with probability $\Theta(\epsilon)$. Hence, two random columns of ΠV have an inner product $\Theta(\epsilon)$ with probability $\Theta(1/d^2)$. On the other hand, since ΠV has at most $O(d^2)$ column pairs, the previous strategy finds a large inner product with probability at most a small constant. By enlarging Π from d^2 to $\Theta(d^2)$ rows, this probability can be made smaller than δ, so no contradiction with the subspace embedding property arises. Hence this approach cannot yield lower bound beyond $\Omega(d^2)$.

In this example, however, two random columns of ΠV has a much larger inner product than $\Theta(\epsilon)$ whenever the inner product is nonzero; that is, X is likely to be much larger than κ. Specifically,

$$\Pr\{X \geq 1 - 8\epsilon\} = \Theta(\epsilon)$$

when $(\Pi'_k V)_{\star,i}$ and $(\Pi'_k V)_{\star,j}$ are drawn from the same S_k. This suggests that we could use a more challenging hard instance $\mathcal{D}_\beta$ with a smaller β, which would be conducive to a higher lower bound. For instance, when $\beta = \Theta(\epsilon)$, the inner product of two columns of ΠV will be scaled down by a factor of $\Theta(\epsilon)$, leading to an inequality comparable with (3). But there are now potentially many more column pairs than before, making a higher lower bound more likely.

2.2 The intuition

Recall that $X = \langle (\Pi'_k V)_{\star,i}, (\Pi'_k V)_{\star,j} \rangle$ for two random $i, j \in [n]$, where Π'_k is the matrix obtained from Π by zeroing out the k-th row. The central technical issue is how to exploit the fact that X can be much larger than κ. We characterize the phenomenon with Eq. (4) below. Note that $X \leq 1$ and $\mathbb{E}[X] \geq 0$. For all $\kappa \in (0,1)$, if $X < -3\kappa$ with probability at least $2/3$, then there is a nonnegative integer $\ell \leq \log(1/\kappa)$ such that

$$\Pr[X \approx 2^\ell \kappa] \approx 2^{-\ell}/\log(1/\kappa). \tag{4}$$

This quickly follows from applying the law of total expectation:

$$\begin{aligned}\kappa + \sum_{\ell=0}^{\log(1/\kappa)} \mathbb{E}\left[X \mid X \in (2^\ell\kappa, 2^{\ell+1}\kappa]\right] \cdot \Pr\left[X \in (2^\ell\kappa, 2^{\ell+1}\kappa]\right] &\geq \mathbb{E}[X \mid X \geq -3\kappa] \cdot \Pr[X \geq -3\kappa] \\ &= \mathbb{E}[X] - \mathbb{E}[X \mid X < -3\kappa] \cdot \Pr[X < -3\kappa] \geq 2\kappa.\end{aligned}$$

The formal statement and its proof can be found in Lemma 9. It is tempting to apply our new observation, Eq. (4), in the strategy used by [8, 11] as described in Section 2.1 to prove a higher lower bound. However, Eq. (4) is incompatible with the existing strategy. Instead, we devise a novel approach which is drastically different from the previous greedy strategies to lower bound the number of column pairs with large inner products.

2.3 Our approach for sparse case

To exploit Eq. (4), we adopt a straightforward strategy in a global manner. For description simplicity, we call a column pair *good* if the pair has a large inner product. We partition columns of ΠV into two disjoint sets of equal size $d/(2\beta)$ and show that

Claim 4 (Informal). *One of the following two cases must hold:*

(i) (good intra-partition pair) With $\Omega(1)$ probability, there exist $i, j \leq d/(2\beta)$ such that the inner product $\langle (\Pi V)_{\star,i}, (\Pi V)_{\star,j} \rangle$ is large;

(ii) (good inter-partition pair) With $\Omega(1)$ probability, there exist $i \leq d/(2\beta)$ and $j > d/(2\beta)$ such that the inner product $\langle (\Pi V)_{\star,i}, (\Pi V)_{\star,j} \rangle$ is large.

Claim 4 is proved by an incremental strategy. Define $S'_j \triangleq \{i \in [n] : \langle \Pi_{\star,i}, (\Pi V)_{\star,j} \rangle \text{ is large}\}$, the set of the columns of Π which has a large inner product with $(\Pi V)_{\star,j}$. We shall show

Claim 5 (Informal). *For any column set $S'' \subseteq [n]$ and a uniformly random column $J \in [n]$, with a good probability, either $J \in S''$, or $S'_J \setminus S''$ is not small.*

We first demonstrate how to prove Claim 4 using Claim 5. The first $d/(2\beta)$ columns of ΠV are examined sequentially. At step j ($j \leq d/(2\beta)$), let $S'' = S'_1 \cup \cdots \cup S'_{j-1}$ and let $J \in [n]$ be the index such that $\Pi_{\star,J} = (\Pi V)_{\star,j}$. If $J \in S''$, a good intra-partition pair is found; otherwise, the set S'' grows substantially. At the end of this procedure, if no good intra-partition pair is found, S'' must have grown at every step and $n' = |S'_1 \cup \cdots \cup S'_{d/(2\beta)}|$ must therefore be large. Note that n' counts the number of columns of Π which can form a good pair with the first half of ΠV. Recall that the second half of ΠV are $d/(2\beta)$ columns uniformly distributed over the n columns of Π; hence, a good inter-partition pair exists with probability at least $1-(1-n'/n)^{d/(2\beta)}$. Since n' is large, this probability is also large, and Claim 4 follows.

Now, we need to prove Claim 5 for an arbitrary Π. Directly analysing S'_J is hardly feasible, so instead we proceeds as follows. We choose a suitable value $\theta \in [\epsilon, 1]$ with an integer $\ell \in [0, \log(1/\theta)]$ and sample a random row k such that $\Pi_{k,J} \approx \sqrt{\theta}$. We then lower bound $|S'_J \setminus S''|$ by lower bounding the size of $\{i \in [n] \setminus S'' : \Pi_{k,i} \approx \sqrt{\theta}, \langle \Pi_{\star,J}, \Pi_{\star,i}\rangle \approx 2^\ell \theta\}$.

Great collision lemma. To this end, we introduce the *great collision lemma*, which is our core technical innovation for the sparse case.

Lemma 6 (Great collision lemma, informal version of Lemma 10). *Suppose that S is a finite multiset of vectors in the unit ℓ_2 ball and $\kappa \in (0, 1/2]$. There exists $\ell \le \lceil \log(1/\kappa)\rceil$ such that for every $S'' \subseteq S$ of size $O(|S|/\log(1/\kappa))$,*

$$\Pr_{u\sim \mathrm{Unif}(S)}\left[\left|\{v \in S \setminus S'' : \langle u, v\rangle \ge 2^\ell \kappa - 2\kappa\}\right| = \Omega\left(\frac{|S|}{2^\ell \log(1/\kappa)}\right)\right] = \Omega\left(\frac{1}{\log(1/\kappa)}\right).$$

Informally, the great collision lemma shows that, for any set S of unit vectors, for any $\kappa > 0$, there exists a scale $2^\ell\kappa$ such that $\approx |S|^2/(2^\ell \log(1/\kappa))$ pairs of vectors in S have inner product $\approx 2^\ell\kappa$.

The application of the great collision lemma is straightforward. By an averaging argument, we fix universal θ, ℓ such that (i) many entries of Π have value $\approx \sqrt{\theta}$, and (ii) for many rows $k \in [m]$, applying the great collision lemma with $\kappa = \theta/2$ to the vector set $\left\{(\Pi'_k)_{\star,j} : j \in [n], \Pi_{k,j} \approx \sqrt{\theta}\right\}$ returns the fixed ℓ.

We now provide more quantitative details of the argument which proves Claim 4 using Claim 5. Since a column of ΠV is uniformly distributed over all columns of Π, we write it as $\Pi_{\star,J}$, where J is uniform over $[n]$. We sample a row k such that $\Pi_{k,J} \approx \sqrt{\theta}$, then the joint distribution of (k, J) can be viewed as first sampling a row k (not necessarily uniformly) and then sampling a column J from the set $S_{k,\theta} \triangleq \{j \in [n] : \Pi_{k,j} \approx \sqrt{\theta}\}$. If the sampled row k does not satisfy condition (ii), we skip this column J, i.e. skipping a column of ΠV, so from now on we suppose k is such a row in (ii). If $|S'' \cap S_{k,\theta}| = \Omega(|S_{k,\theta}|/\log(1/\theta))$, then $J \in S''$ with probability $\Omega(1/\log(1/\theta))$ and a good intra-partition pair is found. Otherwise, since $\Pi_{k,j} \cdot \Pi_{k,j'} \approx \theta$ for all $j, j' \in S_{k,\theta}$, the (ii) and Eq. (4) imply that, with probability $\Omega(1/\log(1/\theta))$, many columns $j \in S_{k,\theta} \setminus S''$ satisfy $\langle \Pi_{\star,J}, \Pi_{\star,j}\rangle \approx 2^\ell\theta$ with J. Consequently, the size of S'' will increase by $\Omega(|S_{k,\theta}|/(2^\ell \log(1/\theta)))$. Thus, after examining the first $d/(2\beta)$ columns of ΠV, either a good intra-partition pair is found with $\Omega(1)$ probability, or $n' = |S''|$ is large enough to guarantee a good inter-partition pair with $\Omega(1)$ probability.

Proof sketch of Lemma 6. Let $\varphi(u, \ell, S'') \triangleq |\{v \in S \setminus S'' : \langle u, v\rangle \ge 2^\ell\kappa - 2\kappa\}|$. We shall prove that, for every not-too-large set $S'' \subseteq S$, there are many $u \in S$ with large $\varphi(u, \ell, S'')$. Due to the symmetry of the inner product, this intuitively claims that there are many $u \in S$ with large $\varphi(u, \ell, \emptyset)$. Our idea is to find $\log(1/\kappa)$ disjoint subsets $S_1, \ldots, S_{\log(1/\kappa)} \subset S$ of equal size, where each S_ℓ contains only the vectors $u \in S$ with large $\varphi(u, \ell, \emptyset)$. By the averaging principle, there is an $\ell \in [\log(1/\kappa)]$ such that the subset S_ℓ contains many u of large $\varphi(u, \ell, \emptyset)$. Informally, the procedure works as follows.

(1) let $S_0 \triangleq \emptyset$, enumerate $\ell \in \{0, \ldots, \log(1/\kappa)\}$ in ascending order.
(2) for each ℓ, let $S_\ell \triangleq \emptyset$. Repeat the following for $|S|/(6\log(1/\kappa))$ times: add to S_ℓ the $u \in S \setminus \cup_{i\le\ell} S_i$ with the largest $\varphi(u, \ell, \cup_{i\le\ell} S_i)$.
(3) we claim that there must exist an ℓ such that there are many $u \in S \setminus \cup_{i<\ell} S_i$ of large $\varphi(u, \ell, \cup_{i<\ell} S_i)$ (which is clearly larger than $\varphi(u, \ell, \emptyset)$).
(4) if the assertion does not hold for all ℓ, we end up with $\bar{S} \triangleq S \setminus \cup_{i\le\log(1/\kappa)} S_i$ of size $\frac{5}{6}|S|$, and for all ℓ, only a few pairs $u, v \in \bar{S}$ have inner product $\approx 2^\ell\kappa$. However, such an $\bar{S}$ cannot exist

due to Eq. (4): for any large vector set $\bar{S}$, by applying Eq. (4) with $X = \langle u, v\rangle$ for uniformly random $u, v \in \bar{S}$, there must be an ℓ such that many pairs $u, v \in \bar{S}$ are of inner product $\approx 2^\ell \kappa$.

However, this argument still does not work for an arbitrary S''. To see this, note that the threshold $\Theta(|S|/(2^\ell \log(1/\kappa)))$ for $\varphi(\cdot)$ could be much smaller than $|S''|$ and S'' may contain some vectors so that $\varphi(u, \ell, S'')$ is small for almost every u. To complete our proof, we slightly modify the procedure as follows: In Step (2), for each ℓ, we let S'_ℓ be the subset $S'' \subseteq S$ of size at most $|S|/(32\log(1/\kappa))$ which minimizes the number of $u \in S \setminus S''$ with large $\varphi(u, \ell, S'')$, then initialize $S_\ell \triangleq S'_\ell$ and add the u for $|S|/(6\log(1/\kappa))$ times. The formal proof can be found in Section 4.2.

2.4 Our approach for general case

When s is large, e.g. $s > 9/\epsilon$, an average nonzero entry becomes small in absolute value and the previous strategy of finding column pairs will not work. To see this, consider a random Π where each column contains s nonzero entries, placed uniformly at random, each equal to $1/\sqrt{s}$. If two columns of ΠV share a nonzero row, their inner product is $\approx 1/s < \epsilon/9$ only (in absolute value), which is too small to induce a contradiction by the argument from the previous section.

Instead, if we examine χ columns $X_1, \dots, X_\chi$ of ΠV simultaneously, where each distinct pair X_i, X_j share a nonzero row, we have

$$\left\|\sum_i X_i\right\|_2^2 = \sum_i \|X_i\|_2^2 + \sum_{i\neq j}\langle X_i, X_j\rangle \approx \chi + \frac{\chi(\chi-1)}{s}.$$

If there exists a unit vector v such that $\Pi U v = \sum_i X_i/\sqrt{\chi}$, then $\|\Pi U v\|_2^2 \approx 1 + (\chi-1)/s$. Thus, if we can find such a tuple $(X_1, \dots, X_\chi)$ with sufficiently large χ, we can hope to derive a contradiction.

A first observation is that the randomization trick (Lemma 8), which works in the sparse case, no longer works for large χ, such as $\chi = \omega(1)$. Indeed, if χ vectors are multiplied by independent random signs, the sum of the pairwise inner products may still be small with a high probability. For example, if all χ vectors are e_1, then the signed sum is $O(\sqrt{\chi \log \chi})$ in absolute value with probability $1 - O(1/\chi)$ by a Chernoff bound, but we need anticoncentration of magnitude $\Omega(\chi)$ to derive the contradiction for a good lower bound. This indicates that we should examine the full quantity $\|\Pi U v\|_2 = \|\sum_i X_i\|_2/\sqrt{\chi}$ instead of only a few pairwise inner products.

For simplicity, assume henceforth that every nonzero entry of Π has absolute value $1/\sqrt{s}$. We also remove the random signs from W so that every entry of U is nonnegative. It is easy to see that almost every column of Π has an ℓ_2-norm of $1 \pm \epsilon$; otherwise, Π would not be an OSE for $U \sim \mathcal{D}_1$. Therefore, $\sum_i \|X_i\|_2^2 = (1\pm\epsilon)^2\chi$, and the key is to show that $\sum_{i\neq j}\langle X_i, X_j\rangle \notin [-4\epsilon\chi, 4\epsilon\chi]$ (recall that $\|\Pi U v\|_2^2 \in [(1-\epsilon)^2, (1+\epsilon)^2]$, which is a range of length 4ϵ). A straightforward approach is to find a row r such that there are $\chi > 4\epsilon s + 1$ columns $X_1, \dots, X_\chi$ in ΠV which are nonzero in row r and have the same sign there. We can then decompose $\sum_{i\neq j}\langle X_i, X_j\rangle$ into the contribution from row r and the contribution from the remaining rows. The former part is exactly $\chi(\chi-1)/s > 4\epsilon\chi$, and the latter part can be bounded by applying our observation Eq. (4) as in the sparse case.

However, an immediate issue arises. To exploit the fact that the sum of the inner products is large with small probability in order to obtain a stronger lower bound, we should take $U \sim \mathcal{D}_\beta$ for some small β. This corresponds to splitting each coordinate of v so that ΠV has d/β columns and each coordinate of v corresponds to $1/\beta$ columns in ΠV when examining $\Pi U v$. For $\beta < 1$, we can find multiple tuples $(X_1, \dots, X_\chi)$ such that the event in Eq. (4) happens for one of these χ-tuples with good probability, but we cannot find a unit vector v such that $\Pi U v$ covers exactly one of desired χ-tuples. Nevertheless, we can choose v to be a normalized sum of χ standard basis vectors such that $\Pi U v$ covers $X_1, \dots, X_\chi$ and another $\chi/\beta - \chi$ columns. Consider $\|\Pi U v\|_2^2$ via the expansion $(\chi/\beta)\,\|\Pi U v\|_2^2 = \|X\|_2^2 + \|Y\|_2^2 + 2\langle X, Y\rangle$, where $X \triangleq \sum_i X_i$ and Y is the sum of the remaining $\chi/\beta - \chi$ columns. We wish to show that the latter two terms do not cancel the first term. The middle term is easy to control as Y is independent of X but the last term is problematic as $\langle X, Y\rangle$ depends on X. A straightforward idea is to apply Eq. (4) by including $\|Y\|_2^2 + 2\langle X, Y\rangle$ together with the coordinates of X_i outside row r. However, it does not work owing to a similar issue encountered when combining Eq. (4) with previous approaches in the sparse case.

Since Y is the sum of $\chi/\beta - \chi$ uniformly random columns of Π, it is easy to obtain that $\|Y\|_2 = (1 \pm \epsilon)(\chi/\beta - \chi)$ by requiring Π to be an OSE for $U \sim \mathcal{D}_{\beta'}$, where $1/\beta' = 1/\beta - 1$. Next, we focus on the inner product $\langle X, Y\rangle$, which can be written as the sum of inner products of two columns which are drawn uniformly from all columns of Π. By the linearity of expectation, we can see that $\mathbb{E}\langle X, Y\rangle \geq 0$. Thus it is tempting to use Eq. (4) again to lower bound the inner product. However, a chicken-and-egg issue arises: given a distribution of $\langle X, Y\rangle$, applying Eq. (4) yields an ℓ, which in turn determines our choice of β. As a result, the distribution of $\langle X, Y\rangle$ changes according to the new value of β, since Y is sum of $\chi/\beta - \chi$ random columns of Π. To see how the distribution changes, write $\langle X, Y\rangle = \sum_i \langle X, Y_i\rangle$, where each Y_i is an independent random column of Π. Thus, certain values of $\langle X, Y_i\rangle$ would be barely present in the sum when $1/\beta$ is small while they are more likely to appear when $1/\beta$ is large.

We resolve this issue by arguing in the reverse direction. We show that, by including a constant factor more columns in Y, $\langle X, Y\rangle$ could be smaller than $-8\epsilon\chi$ and in this case would not be cancelled by $\sum_{i\neq j}\langle X_i, X_j\rangle$. Specifically, if $\Pr[\langle X, Y\rangle < -\epsilon\chi] > 1/2$, which means that $\langle X, Y\rangle$ cancels $\sum_{i\neq j}\langle X_i, X_j\rangle$, then $\Pr[\langle X, Y''\rangle < -8\epsilon\chi] > 1/256$, where Y'' is the sum of 8 independent copies of Y. Therefore, $\sum_{i\neq j}\langle X_i, X_j\rangle$ will not cancel $\langle X, Y\rangle$ if $U \sim \mathcal{D}_{\beta/8}$.

To summarize, we prove the lower bound with the following ingredients:

(1) set $\chi \approx 4\epsilon s$,
(2) apply Eq. (4) to a random χ-tuple of columns of Π that share same nonzero row, obtaining ℓ,
(3) set $\beta \approx 2^{-\ell}$ so that with $\Omega(1)$ probability, at least one χ-tuple satisfies $\sum_{i\neq j}\langle X_i, X_j\rangle \approx 4\epsilon\chi/\beta$,
(4) force Π to be an OSE for $U \sim \mathcal{D}_{\beta'}$ for four different values of β' simultaneously: (a) $\beta' = \beta$, (b) $1/\beta' = 1/\beta - 1$, (c) $\beta' = \beta/8$, (d) $1/\beta' = 8/\beta - 1$, so that $\sum_{i\neq j}\langle X_i, X_j\rangle + 2\langle X, Y\rangle$ is either $> 4\epsilon\chi/\beta$ when $U \sim \mathcal{D}_\beta$, or $< -4\epsilon\chi/(\beta/8)$ when $U \sim \mathcal{D}_{\beta/8}$, thereby yielding a contradiction.

The full proof can be found in Section 5.

3 Notation

For $x, y \in \mathbb{R}$ and $\theta > 0$, we write $x = y \pm \theta$ if $x \in [y - \theta, y + \theta]$. For a matrix A, we denote its i-th column vector by $A_{\star,i}$, and denote by A'_k the matrix obtained by zeroing out the k-th row. For a finite multiset S, we denote by $\mathrm{Unif}(S)$ the uniform distribution on S. For a random variable X and a probability distribution $\mathcal{D}$, we write $X \sim \mathcal{D}$ to denote that X follows $\mathcal{D}$.

When d and ϵ are clear from the context, we abbreviate the event in the probability of Eq. (2) to "Π is a subspace embedding for A".

4 Lower Bound for Sparse Case

4.1 Preliminaries

Our hard instance $\widetilde{\mathcal{D}}$ is based on the same construction as in [8]. It is a mixture distribution of $\mathcal{D}_\beta$, parameterized by β, on $n \times d$ matrices. With probability $1/2$, $\widetilde{\mathcal{D}} = \mathcal{D}_1$; with probability $1/2$, $\widetilde{\mathcal{D}}$ is $\mathcal{D}_\beta$, where $1/\beta = 2^\ell$ and $\ell \sim \mathrm{Unif}(\{0, \ldots, \lceil \log s \rceil\})$. Recall the definition of $\mathcal{D}_\beta$ below.

Definition 7 (Distribution $\mathcal{D}_\beta$ [8]). *The distribution $\mathcal{D}_\beta$ ($0 < \beta \leq 1$) is defined on matrices $U \in \mathbb{R}^{n\times d}$ as follows. The matrix U is decomposed as $U = VW$, where $V \in \mathbb{R}^{n\times d/\beta}$ and $W \in \mathbb{R}^{d/\beta\times d}$. The matrix V has i.i.d. columns, each $V_{\star,i}$ ($i = 1, \ldots, d/\beta$) is chosen uniformly from the n standard* *basis vectors in* $\mathbb{R}^n$. *The matrix* W *is distributed as follows: For each* $i = 1, \ldots, d$, *set* $W_{j,i} \triangleq \sigma_j \sqrt{\beta}$ *for* $j = (i-1)/\beta + 1, \ldots, i/\beta$, *where* $\sigma_j \in \{-1, 1\}$ *are independent Rademacher variables; set the remaining entries of* W *to zero.*

It follows from the same argument in [8] that $U = VW$ is an isometry with probability at least $1 - \delta/(2K)$ when $n \geq Kd^2/(\beta^2\delta)$. Hence we assume that K is large enough and restrict our attention to the case U is an isometry (i.e., V has independent columns and full column rank.)

Our starting point is the same as [8]. Observe that if two columns of ΠV have a large inner product (in absolute value) then Π cannot be a subspace embedding for U. This is formalized in the following lemma.

Lemma 8 ([8, Lemma 4]). *Suppose that* $|\langle A_{\star,p}, A_{\star,q}\rangle| \geq \lambda\epsilon/\beta$ *for some distinct columns* p, q *of a matrix* $A \in \mathbb{R}^{m \times d/\beta}$, *where* $\lambda > 2$, *and* W *is as in Definition 7. Then there exists a unit vector* $u \in \mathbb{R}^d$ *such that* $\|AWu\|_2^2 \notin [(1-\epsilon)^2, (1+\epsilon)^2]$ *with probability at least* $1/4$ *(over* W*).*

It remains to find a pair of columns with a large inner product. A key tool for this task was given in [8, Lemma 3], which strengthens the earlier results [11, Lemmata 8 and 9]. It states that given a finite collection of vectors of length at most $1 + \epsilon$, there always exists a small fraction of vector pairs whose inner product is not too small ("small" here means being negative and far from 0). The following lemma, which refines [8, Lemma 3], further extends the inner products to large positive numbers.

Lemma 9 (Refined from [8, Lemma 3]). *Suppose that* S *is a finite multiset of vectors in the* ℓ_2*-ball of radius* $1 + \epsilon$ *and* u, v *are independently sampled from* $\mathrm{Unif}(S)$. *Further suppose that* $\kappa, \Delta > 0$ *satisfy* $\kappa \leq \Delta \leq 1$, *and* $L = \lceil \log((1+\epsilon)^2/\Delta) \rceil$. *Then one of the following must hold: (i)* $\Pr[-\kappa \leq \langle u, v\rangle \leq \Delta] \geq \kappa/(2\Delta(L+1))$; *(ii) there exists* $i \in \{0, \ldots, L-1\}$ *such that* $\Pr[2^i\Delta < \langle u, v\rangle \leq 2 \cdot 2^i\Delta] \geq \kappa/(4 \cdot 2^i\Delta(L+1))$.

Proof. Let $X \triangleq \langle u, v\rangle$. Then $\mathbb{E}[X] = \mathbb{E}\langle u, v\rangle = \langle \mathbb{E}\, u, \mathbb{E}\, v\rangle = \langle \mathbb{E}\, u, \mathbb{E}\, u\rangle \geq 0$. If $\Pr[X \geq -\kappa] \geq 1/2$ there is nothing to prove, we henceforth assume that $\Pr[X < -\kappa] \geq 1/2$.

As $\mathbb{E}[X] \geq 0$, by the law of total expectation

$$\mathbb{E}[X|X \geq 0] \cdot \Pr[X \geq 0] \geq -\mathbb{E}[X|X < 0] \cdot \Pr[X < 0] \geq \frac{\kappa}{2}.$$

Since $X \leq (1 + \epsilon)^2$, we have that

$$\mathbb{E}[X|X \in [0, \Delta]] \cdot \Pr[X \in [0, \Delta]] + \sum_{i=0}^{L-1} \mathbb{E}[X|X \in (2^i\Delta, 2^{i+1}\Delta]] \cdot \Pr[X \in (2^i\Delta, 2^{i+1}\Delta]] \geq \frac{\kappa}{2}.$$

By the pigeonhole principle, at least one of the $L + 1$ terms on the LHS is at least $\kappa/(2(L + 1))$. If it is the first term, we conclude with (i); otherwise, we conclude with (ii). □

When applying Lemma 9, we always assume that it returns an $\ell \in \{-1, \ldots, \lceil \log((1 + \epsilon)^2/\Delta) \rceil\}$, i.e. the case (i) is referred to as $\ell = -1$.

4.2 Great collision lemma

Lemma 9 shows that among n vectors sampled from a set S, there exists an ℓ such that the expected number of pairs with inner product $\approx 2^\ell\kappa$ is at least $\approx n^2/2^\ell$. In this section, we prove something stronger: the number of such pairs is at least $\approx n^2/2^\ell$ with high probability (over the choice of the n vectors). This result, which we call the *great collision lemma*, is our major technical innovation and is formally stated in Lemma 10 below.

To find vector pairs with large inner products, we use Algorithm 1, a greedy procedure. For each ℓ, it collects vectors that have a large inner product with many other vectors. If there are many

Algorithm 1: Finding a great collision

1 $j \leftarrow 1, S_1 \leftarrow S$;
2 $L \leftarrow \lceil \log((1+\epsilon)^2/\Delta) \rceil + 1$;
3 $K \leftarrow L\Delta/\kappa$;
4 **foreach** $\ell \in \{-1, \dots, L-1\}$ **do**
5 $\quad$ $p \leftarrow -\kappa$ if $\ell = -1$ and $p \leftarrow 2^\ell \Delta$ otherwise;
6 $\quad$ $S'_\ell \leftarrow S_j$;
7 $\quad$ break ties arbitrarily: $S''_\ell \leftarrow \underset{S' \subseteq S_j : |S'| \le \frac{|S|}{32K}}{\arg\min} \sum_{c \in S_j \setminus S'} \mathbf{1}_{\left\{ |\{c' \in S_j \setminus S' : \langle c, c' \rangle \ge p\}| \ge \frac{|S|}{2^{\ell+5}K} \right\}}$;
8 $\quad$ $S_{j+1} \leftarrow S'_\ell \setminus S''_\ell$;
9 $\quad$ $j \leftarrow j + 1$;
10 $\quad$ $i \leftarrow 1$;
11 $\quad$ **while** $i \le |S|/(6K)$ *and* $\left|\left\{(c, c') \in S_j \times S_j : \langle c, c' \rangle \ge p\right\}\right| \ge |S|^2/(2^{\ell+3}K)$ **do**
12 $\quad\quad$ $c \leftarrow \underset{c' \in S_j}{\arg\max} \left|\left\{c'' \in S_j : \langle c', c'' \rangle \ge p\right\}\right|$;
13 $\quad\quad$ **if** $\left|\left\{c' \in S_j : \langle c, c' \rangle \ge p\right\}\right| < \kappa |S|/(2^\ell \Delta)$ **then**
14 $\quad\quad\quad$ **return** (ℓ, S'_ℓ, S_j);
15 $\quad\quad$ $S_{j+1} \leftarrow S_j \setminus \{c\}$;
16 $\quad\quad$ $j \leftarrow j + 1$;
17 $\quad\quad$ $i \leftarrow i + 1$;
18 $\quad$ **if** $i > |S|/(6K)$ **then**
19 $\quad\quad$ **return** (ℓ, S'_ℓ);

such vectors, the task is done (the case in Line 19). Otherwise, each remaining vector has a large inner product with not-too-many other vectors, and by Lemma 9 and the averaging principle, there will still be sufficiently many vector pairs with a large inner product (the case in Line 14).

Lemma 10 (Great collision lemma). *Suppose that S is a finite multiset of vectors in the ℓ_2-ball of radius $1+\epsilon$, and that $0 < \kappa \le \Delta \le 1$, $L = \lceil \log((1+\epsilon)^2/\Delta) \rceil + 1$ and $K = L\Delta/\kappa$. Then, there exists an integer $\ell \in \{-1, \dots, L-1\}$ such that for every $S' \subseteq S$ with $|S'| \le |S|/(32K)$ it holds that*

$$\Pr_{c \sim \mathrm{Unif}(S)} \left[\Pr_{c' \sim \mathrm{Unif}(S \setminus S')} [\langle c, c' \rangle \ge p] \ge \frac{1}{2^{\ell+5}K} \right] \ge \frac{3}{31K},$$

where $p \triangleq -\kappa$ if $\ell = -1$, and $p \triangleq 2^\ell \Delta$ if $\ell \ge 0$.

Proof. For notational convenience, let $\lambda \triangleq 3/31$. Recall that $p \triangleq -\kappa$ if $\ell = -1$ and $p \triangleq 2^\ell \Delta$ if $\ell \ge 0$. Consider Algorithm 1. We first claim that it must return an ℓ. Line 8 removes in total at most $|S|/(32K)$ columns from S_j and Line 15 at most $|S|/(6K)$ columns. Hence, if the algorithm does not return an ℓ, we would end up with an S_j such that (i) $|S_j| \ge \frac{77}{96}|S|$, and (ii) $\left|\left\{(c, c') \in S_j \times S_j : \langle c, c' \rangle \ge p\right\}\right| < |S|^2/(2^{\ell+3}K) < |S_j|^2/(2^{\ell+2}K)$. But, by Lemma 9, (i) and (ii) cannot hold simultaneously.

We now know that the algorithm will return an ℓ. There are two cases.

Case 1: ℓ returned in Line 14. The algorithm returns S_j, S'_ℓ and ℓ, which means that the "ground" set is S'_ℓ and we can find "good" vectors in S_j. The claimed result can be rewritten as

$$\sum_{c \in S} \mathbf{1}_{\left\{ |\{c' \in S \setminus S' : \langle c, c' \rangle \ge p\}| \ge \frac{|S|}{2^{\ell+5}K} \right\}} \ge \frac{\lambda |S|}{K}$$

for all $S' \subseteq S$ with $|S'| \le |S|/(32K)$. Recall that $S'_\ell \subseteq S$ in Line 6. To prove the claimed result, it suffices to show that for all $S' \subseteq S$ with $|S'| \le |S|/(32K)$,

$$P \triangleq \sum_{c \in S'_\ell \setminus S'} \mathbf{1}_{\left\{\left|\{c' \in S'_\ell \setminus S' : \langle c, c' \rangle \ge p\}\right| \ge \frac{|S|}{2^{\ell+5}K}\right\}} \ge \frac{\lambda |S|}{K}, \tag{5}$$

and we can assume without loss of generality that $S' \subseteq S'_\ell$. The definition of S''_ℓ guarantees that S''_ℓ is the S' which minimizes

$$\left|\left\{c \in S'_\ell \setminus S' : \left|\left\{c' \in S'_\ell \setminus S' : \langle c, c' \rangle \ge p\right\}\right| \ge \frac{|S|}{2^{\ell+5}K}\right\}\right|,$$

regardless of how we break the tie in Line 7. Recall that $S_j \subseteq S'_\ell \setminus S''_\ell$. Algorithm 1 guarantees in Line 13 that

$$\left|\{c' \in S_j : \langle c, c' \rangle \ge p\}\right| < \frac{\kappa |S|}{2^\ell \Delta}, \quad \forall c \in S_j.$$

On the other hand, combining the above equation with Eq. (5),

$$\left|\{(c, c') \in S_j \times S_j : \langle c, c' \rangle \ge p\}\right| \le \frac{\kappa |S|}{2^\ell \Delta} \cdot P + \frac{|S|}{2^{\ell+5}K} \cdot (|S_j| - P). \tag{6}$$

Algorithm 1 also guarantees in Line 11 that

$$\left|\left\{(c, c') \in S_j \times S_j : \langle c, c' \rangle \ge p\right\}\right| \ge \frac{|S|^2}{2^{\ell+3}K}. \tag{7}$$

Combining Eq. (6) with Eq. (7) yields the desired result Eq. (5).

Case 2: ℓ returned in Line 19. The algorithm returns S'_ℓ and ℓ, which means that the "ground" set is S'_ℓ and the "good" vectors are the elements c found in Line 12.

Recall that $S'_\ell \subseteq S$ in Line 6. As in the previous case Eq. (5), it suffices to show that

$$\sum_{c \in S'_\ell \setminus S'} \mathbf{1}_{\left\{\left|\{c' \in S'_\ell \setminus S' : \langle c, c' \rangle \ge p\}\right| \ge \frac{|S|}{2^{\ell+5}K}\right\}} \ge \frac{\lambda |S|}{K} \tag{8}$$

for all $S' \subseteq S$ with $|S'| \le |S|/(32K)$, and we can assume without loss of generality that $S' \subseteq S'_\ell$. The definition of S''_ℓ guarantees that S''_ℓ is the S' which minimizes

$$\left|\left\{c \in S'_\ell \setminus S' : \left|\left\{c' \in S'_\ell \setminus S' : \langle c, c' \rangle \ge p\right\}\right| \ge \frac{|S|}{2^{\ell+5}K}\right\}\right|,$$

regardless of how we break the tie in Line 7. After removing S''_ℓ from S'_ℓ, the while-loop is executed for at least $|S|/(6K)$ times, so there are at least $|S|/(6K)$ elements $c \in S'_\ell \setminus S''_\ell$ satisfying that $\left|\{c' \in S'_\ell \setminus S''_\ell : \langle c, c' \rangle \ge p\}\right| \ge \kappa |S|/(2^\ell \Delta)$. This implies that

$$\sum_{c \in S'_\ell \setminus S''_\ell} \mathbf{1}_{\left\{\left|\{c' \in S'_\ell \setminus S''_\ell : \langle c, c' \rangle \ge p\}\right| \ge \frac{\kappa |S|}{2^\ell \Delta}\right\}} \ge \frac{|S|}{6K}.$$

It follows that for a general S' (letting $N(c) = \left|\{c' \in S'_\ell \setminus S' : \langle c, c' \rangle \ge p\}\right|$ for brevity)

$$\sum_{c \in S'_\ell \setminus S'} \mathbf{1}_{\left\{N(c) \ge \frac{|S|}{2^{\ell+5}K}\right\}} \ge \sum_{c \in S'_\ell \setminus S''_\ell} \mathbf{1}_{\left\{N(c) \ge \frac{|S|}{2^{\ell+5}K}\right\}} \ge \sum_{c \in S'_\ell \setminus S''_\ell} \mathbf{1}_{\left\{N(c) \ge \frac{\kappa |S|}{2^\ell \Delta}\right\}} \ge \frac{|S|}{6K},$$

whence we see that Eq. (8) holds automatically. □

4.3 Fixing the parameters

We shall prove a lower bound on the number of rows for an arbitrary unknown matrix Π which is an (ϵ, δ)-subspace-embedding for $U = VW \sim \widetilde{\mathcal{D}}$. To analyse ΠV, we shall fix some parameters of Π and U so that the great collision lemma (Lemma 10) can be applied with suitable parameters.

First, we set $\kappa = \theta/\log s$ and $\Delta = \theta$ in Lemma 10, where $\sqrt{\theta}$ is the most common magnitude among the nonzero entries of Π. Then, for each row k, we find a set $S^+_{k,\theta}$ of columns which will work perfectly with Lemma 10, and apply the lemma to obtain a value $\ell = \ell(k)$. We then find the most frequent value ℓ_θ among the $\ell(k)$, and focus on the corresponding rows in Π, i.e., the rows k with $\ell(k) = \ell_\theta$. Since our overall hard distribution is a mixture of hard instances corresponding to different values of β, we also fix the hard instance so that the pair of β and ℓ_θ induces a high lower bound. Specifically, we define a parameter ℓ'_θ based on ℓ_θ and assert that Π must be a subspace embedding for $U \sim \mathcal{D}_\beta$ with $\beta = 2^{-\ell'_\theta}$.

Let $L \triangleq \lceil \log s \rceil$. Recall that our hard distribution $\widetilde{\mathcal{D}}$ is $\mathcal{D}_1$ with probability $1/2$ and is $\mathcal{D}_{2^{-\ell}}$ with probability $1/2$, where $\ell \sim \mathrm{Unif}(\{0, \dots, L\})$. We assume that Π is an (ϵ, δ)-subspace-embedding for $U \sim \widetilde{\mathcal{D}}$. By Lemma 11, it suffices to consider only the columns of Π whose ℓ_2-norms lie in $1 \pm \epsilon$.

Lemma 11 (Refined from [8, Lemma 6]). *If $\Pi \in \mathbb{R}^{m \times n}$ is an (ϵ, δ)-subspace-embedding for $U \sim \widetilde{\mathcal{D}}$, then a $(1 - 2\delta/d)$-fraction of the columns of Π have ℓ_2-norm in $1 \pm \epsilon$.*

The next lemma finds a large fraction of entries with the same order of magnitude. The proof is an easy counting argument.

Lemma 12. *Suppose that Π is a matrix with column sparsity at most $s \geq 3$. Then there exists a $\theta \geq (1-\epsilon)^2/s - 1/s\log s$ such that*

$$\mathop{\mathbb{E}}_{j \sim \mathrm{Unif}([n])} \left[\sum_{k : |\Pi_{k,j}|^2 \in [\theta, 2\theta)} \left|\Pi_{k,j}\right|^2 \,\middle|\, \left\|\Pi_{\star,j}\right\|_2 = 1 \pm \epsilon \right] = \Omega\left(\frac{1}{\log^2 s}\right). \tag{9}$$

Proof. Let $\theta \triangleq (1-\epsilon)^2/s - 1/s\log s$ and $I \triangleq \lceil \log_2((1+\epsilon)^2/\theta) \rceil$. Since the column sparsity is s,

$$\sum_{k : |\Pi_{k,j}|^2 \geq \theta} |\Pi_{k,j}|^2 > (1-\epsilon)^2 - s\theta \geq \frac{1}{\log s}.$$

By the linearity of expectation, this implies that

$$\sum_{i=0}^{I} \mathop{\mathbb{E}}_{j \sim \mathrm{Unif}([n])} \left[\sum_{k : |\Pi_{k,j}|^2 \in [2^i\theta, 2^{i+1}\theta)} \left|\Pi_{k,j}\right|^2 \,\middle|\, \left\|\Pi_{\star,j}\right\|_2 = 1 \pm \epsilon \right] \geq \frac{1}{\log s}.$$

Therefore there exists i such that

$$\mathop{\mathbb{E}}_{j \sim \mathrm{Unif}([n])} \left[\sum_{k : |\Pi_{k,j}|^2 \in [2^i\theta, 2^{i+1}\theta)} \left|\Pi_{k,j}\right|^2 \,\middle|\, \left\|\Pi_{\star,j}\right\|_2 = 1 \pm \epsilon \right] = \Omega\left(\frac{1}{\log^2 s}\right).$$

Let θ in the lemma statement be $2^i\theta$. The proof is complete. □

4.3.1 Applying the great collision lemma. Let θ be as guaranteed by Lemma 12. For each row k, let $S_{k,\theta} \triangleq \{j \in [n] : |\Pi_{k,j}|^2 \geq \theta \text{ and } \left\|\Pi_{\star,j}\right\|_2 = 1 \pm \epsilon\}$ denote the set of column vectors $\Pi_{\star,j}$ such that $|\Pi_{k,j}|^2 \geq \theta$. For each row k, either $|\{j \in S_{k,\theta} : \Pi_{k,j} \geq \sqrt{\theta}\}| \geq |S_{k,\theta}|/2$ or $|\{j \in S_{k,\theta} : \Pi_{k,j} \leq -\sqrt{\theta}\}| \geq |S_{k,\theta}|/2$. In the former case, we let $S^+_{k,\theta} \triangleq \{j \in S_{k,\theta} : \Pi_{k,j} \geq \sqrt{\theta}\}$; in the latter case, we let $S^+_{k,\theta} \triangleq \{j \in S_{k,\theta} : \Pi_{k,j} \leq -\sqrt{\theta}\}$.

Lemma 13. *Suppose that* Π *is a matrix with column sparsity at most* s ($s \geq 3$)*. There is an* $\ell_\theta \in [-1, \lceil \log((1+\epsilon)^2/\theta) \rceil + 1]$ *and a set* S_{ℓ_θ} *of rows such that the following holds.*

(i) $\mathbb{E}_{j \sim \mathrm{Unif}([n])} \left[\left| \left\{ k \in S_{\ell_\theta} : j \in S^+_{k,\theta} \right\} \right| \,\middle|\, \left\| \Pi_{\star,j} \right\|_2 = 1 \pm \epsilon \right] = \Omega\left(\frac{1}{\theta \log^3 s} \right)$;

(ii) For each row $k \in S_{\ell_\theta}$*, let* $S'_k \triangleq \{\Pi_{\star,j}\}_{j \in S^+_{k,\theta}}$*, then for every* $S' \subseteq S'_k$ *satisfying that* $|S'| \leq |S'_k|/(32\lceil \log(1/\theta) \rceil \log s)$,

$$\Pr_{c \sim \mathrm{Unif}(S'_k)} \left[\Pr_{c' \sim \mathrm{Unif}(S'_k \setminus S')} [\langle c, c' \rangle \geq p + \theta] = \Omega\left(\frac{1}{2^{\ell_\theta} \log^2 s} \right) \right] = \Omega\left(\frac{1}{\log^2 s} \right),$$

where $p \triangleq -\theta/\log s$ *if* $\ell_\theta = -1$*, and* $p \triangleq 2^{\ell_\theta}\theta$ *if* $\ell_\theta \geq 0$*.*

Proof. We first show guarantee (i). Intuitively, we apply the great collision lemma (Lemma 10) to each column, then take the majority of the returned ℓ's.

For each $k \in [m]$, recall that matrix Π'_k is obtained from Π by zeroing out the k-th row of Π. For each nonempty $S^+_{k,\theta}$, note that $\|(\Pi'_k)_{\star,j}\|_2 \leq 1 + \epsilon$ for each $j \in S^+_{k,\theta}$. Thus, for each row $k \in [m]$, we can apply Lemma 10 to $\{(\Pi'_k)_{\star,j}\}_{j \in S^+_{k,\theta}}$ with $\kappa = \theta/\log s, \Delta = \theta$ and obtain an $\ell_k \leq \lceil \log((1+\epsilon)^2/\theta) \rceil + 1$. Applying the averaging principle to Eq. (9), there is an ℓ_θ such that

$$\mathop{\mathbb{E}}_{j \sim \mathrm{Unif}([n])} \left[\left| \left\{ k \in S_{\ell_\theta} : j \in S^+_{k,\theta} \right\} \right| \,\middle|\, \left\| \Pi_{\star,j} \right\|_2 = 1 \pm \epsilon \right] = \Omega\left(\frac{1}{\theta \log^3 s} \right),$$

where $S_{\ell_\theta} \triangleq \{k \in [m] : \ell_k = \ell_\theta\}$. This completes the proof of item (i).

Next we show guarantee (ii). Since ℓ_θ is obtained by applying Lemma 10 to $\{(\Pi'_k)_{\star,j}\}_{j \in S^+_{k,\theta}}$ with $\kappa = \theta/\log s$ and $\Delta = \theta$ for each $k \in S_{\ell_\theta}$, it holds that for every $S' \subseteq S'_k$ with $|S'| \leq |S'_k|/(32\lceil \log((1+\epsilon)^2/\theta) \rceil \log s)$ that

$$\Pr_{c \sim \mathrm{Unif}(S'_k)} \left[\Pr_{c' \sim \mathrm{Unif}(S'_k \setminus S')} [\langle c, c' \rangle \geq p] = \Omega\left(\frac{1}{2^{\ell_\theta} \log^2 s} \right) \right] = \Omega\left(\frac{1}{\log^2 s} \right),$$

where $p \triangleq -\theta/\log s$ if $\ell_\theta = -1$ and $p \triangleq 2^{\ell_\theta}\theta$ if $\ell_\theta \geq 0$. Guarantee (ii) follows immediately by noticing that $\Pi_{j,k}\Pi_{j',k} \geq \theta$ for all $j, j' \in S^+_{k,\theta}$, for every row $k \in S_{\ell_\theta}$. □

Let $S_\theta \triangleq \{(k, j) \in [m] \times [n] : j \in S^+_{k,\theta}$ and $k \in S_{\ell_\theta}\}$. We say an entry (i, j) of Π is *good* if $(i, j) \in S_\theta$. Lemma 13(i) can be rephrased as

$$\mathop{\mathbb{E}}_{j \sim \mathrm{Unif}([n])} \left[|\{k \in [m] : (k, j) \in S_\theta\}| \,\middle|\, \left\| \Pi_{\star,j} \right\|_2 = 1 \pm \epsilon \right] = \Omega\left(\frac{1}{\theta \log^3 s} \right). \tag{10}$$

The following corollary follows easily from combining Eq. (10) and Lemma 11.

Corollary 14. *If* $\Pi \in \mathbb{R}^{m \times n}$ *is an OSE for* $U \sim \widetilde{\mathcal{D}}$*, then* Π *has at least* $\Omega(n/(\theta \log^3 s))$ *good entries.*

4.3.2 Fixing the hard instance. Recall that $L = \lceil \log s \rceil$. Also recall that our hard instance $\widetilde{\mathcal{D}}$ is a mixture of $\mathcal{D}_\beta$ with different values of β and Π is an (ϵ, δ)-subspace-embedding for $U \sim \widetilde{\mathcal{D}}$. We have assumed $n \geq Kd^2/(\epsilon^2\delta)$ for large enough constant K so that $U \sim \widetilde{\mathcal{D}}$ is an isometry with probability $1 - \delta/(2K)$. By Markov's inequality, for $(1-\gamma)$-fraction of $\ell \in \{0, \ldots, L\}$, Π is an $(\epsilon, 2\delta/\gamma)$-subspace-embedding for $U \sim \mathcal{D}_{2^{-\ell}}$, where $\gamma \triangleq K\delta < 1/2$, so that $2\delta/\gamma = 2/K$ is a small constant. Hence, for each $\ell \in \{0, \ldots, L\}$, there exists an $\ell' \in [\max\{0, (1-\gamma)L - \ell\}, L - \ell]$ such that

$$\begin{cases} 2^{-\ell-\ell'} \in [2^{-L}, (2^{-L})^{1-\gamma}], \text{ and} \\ \Pi \text{ is an } (\epsilon, 2\delta/\gamma)\text{-subspace-embedding for } U \sim \mathcal{D}_{2^{-\ell'}}. \end{cases} \tag{11}$$

Now, if $\ell_\theta < 0$, we let $\ell'_\theta = 0$; otherwise let ℓ'_θ be the ℓ' as guaranteed by Eq. (11) with $\ell = L - (\ell_\theta + \log(\theta/\epsilon))$. We shall apply Π to $U \sim \mathcal{D}_{2^{-\ell'_\theta}}$ and show that there is a unit vector v such that $\|\Pi U v\|_2 \neq 1 \pm \epsilon$ with probability larger than $2\delta/\gamma$. Note that

$$2^{\ell'_\theta} = \Omega(\theta 2^{\ell_\theta}/\epsilon s^\gamma). \tag{12}$$

4.4 The lower bound

We prove the lower bound in this section. Recall that we assume

$$3 \leq s \leq \left((1-\epsilon)^2 + (2\epsilon - \epsilon^2)/\log s + 1/\log^2 s\right)/(2\epsilon). \tag{13}$$

Let $(C_1, C_2, \ldots, C_{d'}) \in [n]^{d'}$ be the columns chosen by V, where $d' \triangleq 2^{\ell'_\theta} d$, listed in the order in which they are sampled. We partition $\{C_i\}_{i\in[d']}$ into the first half $\{C_i\}_{i\leq d'/2}$ and the second half $\{C_i\}_{i>d'/2}$. Our goal is to show that the set

$$\left\{c \in [n] : \exists i \leq d'/2, \langle c, C_i\rangle \geq 2^{\ell_\theta}\theta\right\} \tag{14}$$

is large with high probability. To this end, we examine the columns $\Pi_{\star,C_i}$ sequentially and maintain a growing set $S' \subseteq [n]$ consisting of the columns which have a large inner product with one of the examined columns $\Pi_{\star,C_1}, \ldots, \Pi_{\star,C_i}$.

We wish to apply Lemma 10 to lower bound the increment of S' after examining a new column $\Pi_{\star,C_i}$, but Lemma 10 only works in the following scenario (as the inner product at least $-\kappa$ when $\ell = -1$): There is a row r and a set of columns S_r such that (i) $\Pi_{r,i} \cdot \Pi_{r,j} \geq 2\epsilon + \kappa$ for any columns $i, j \in S_r$, and (ii) the column c in Lemma 10 is sampled uniformly from S_r. Then we apply Lemma 10 to the set of column vectors S_r with row r removed, to find $S'_r \subseteq S_r$ such that $\langle c', c\rangle \geq p$ for any $c' \in S'_r$ after removing row r. Consequently, $\langle c', c\rangle \geq p + 2\epsilon + \kappa \geq 2\epsilon$ by condition (i).

To apply Lemma 10 (recall the aforesaid (i) and (ii)), we employ rejection sampling and double counting in Section 4.4.1, which allows for viewing a random column C_i as being sampled from the aforesaid S_r with a sampled row r. Then, in Section 4.4.2, we combine everything and follow the proof idea in Section 2.3 to derive the desired lower bound.

4.4.1 Sampling entries. Let $s'_{\max} \triangleq \max_{j\in[n]} |\{k \in [m] : (k,j) \in S_\theta\}|$. To lower bound the size of the set in Eq. (14), every time we examine a column $C_i \sim \mathrm{Unif}([n])$, we choose a row k at random by the following procedure: (1) let s' be the columns sparsity of C_i; (2) sample a random number $\alpha \in [0,1]$; (3) if $\alpha \geq s'/s'_{\max}$, discard C_i; otherwise, choose a row k among the set $\{k' \in [m] : (k', C_i) \in S_\theta\}$ uniformly at random. If we do not discard C_i, we include the columns j with properties $(k,j) \in S_\theta$ and $\langle \Pi_{\star,j}, \Pi_{\star,C_i}\rangle \geq 2^{\ell_\theta}\theta$ to the set in Eq. (14).

By a double counting argument, it is easy to see that, if we do not discard C_i, then (k, C_i) is uniformly distributed at random over all the $(k,j) \in S_\theta$.

We call a column *good* if the column is sampled by V and is not discarded by the procedure above. Let g be the number of good columns. By combining Corollary 14 with the fact $s'_{\max} \leq 2/\theta$, any single column $(\Pi V)_i$ is discarded with probability at most $O(1/\log^3 s)$. By a Chernoff bound, $g = \Omega(d'/\log^3 s)$ with high probability. Consequently we assume $g \geq d'/\log^3 s$ in the remainder of this paper. Let $(C'_1, C'_2, \ldots, C'_g) \in [n]^g$ be the remaining columns, and $(R'_1, R'_2, \ldots, R'_g) \in [m]^g$ the rows corresponding to the remaining columns. Note that for any $i \in [g]$, (R'_i, C'_i) is uniformly at random over all the $(k,j) \in S_\theta$. The uniform distribution can be rephrased as following:

(1) Let $R'_i = r$ with probability proportional to $|\{j \in [n] : (r,j) \in S_\theta\}|$;
(2) Let C'_i be the column of a nonzero entry in $\{j \in [n] : (r,j) \in S_\theta\}$ uniformly at random.

Algorithm 2: Collecting columns

Let g be the number of good columns which are sampled by V and are not discarded by us;
Let $(C'_1, C'_2, \ldots, C'_g) \in [n]^g$ be the good columns, and $(R'_1, R'_2, \ldots, R'_g) \in [m]^g$ the rows correspond to the good columns chosen by us;
$S' \leftarrow \emptyset$, $i \leftarrow 1$;
$p \leftarrow -\theta/\log s$ if $\ell_\theta = -1$ and $p \leftarrow 2^{\ell_\theta}\theta$ if $\ell_\theta \geq 0$;
while $i \leq g/2$ **do**
  if $C'_i \in S'$ **then**
    return *Collision*;
  $S' \leftarrow S' \cup \left\{ c \in [n] : (R'_i, c) \in S_\theta \text{ and } \langle \Pi_{\star, C'_i}, \Pi_{\star, c} \rangle \geq p + \theta \right\}$;

4.4.2 Finding a large inner product. We use Algorithm 2 to find a pair of columns which has a large inner product. The algorithm examines the columns from first half of the good columns one by one, and checks if one of them can be used to form a large inner product with one of the examined column. If this fails, we shall show that one of the remaining half good columns can be used to form a large inner product with one of the examined columns with at least a (small) constant probability. The detailed analysis is given below.

Let $S_{R'_i} \triangleq \{c \in [n] : (R'_i, c) \in S_\theta\}$ be the set column indices corresponding to entries of S_θ in row R'_i. If $|S' \cap S_{R'_i}| \geq |S_{R'_i}|/(32\log(1/\theta)\log s)$, then there is a $j < i$ such that $\langle \Pi_{\star, C'_j}, \Pi_{\star, C'_i} \rangle$ is large with probability at least $1/(32\log(1/\theta)\log s)$ (recall that C'_i is uniformly random over $S_{R'_i}$). Otherwise, $|S' \cap S_{R'_i}| \leq |S_{R'_i}|/(32\log(1/\theta)\log s)$. Recall the definition of a good entry and the fact that $\ell_{R'_i}$ is obtained by applying Lemma 10 to the column set $\{(\Pi'_{R'_i})_{\star, c}\}_{c \in S_{R'_i}}$. It follows that $|S'|$ will increase by $\Omega(|S_{R'_i}|/(2^{\ell_\theta}\log^2 s))$ with probability at least $\Omega(1/\log^2 s)$.

Recall that S_θ has at least $\Omega(n/(\theta \log^3 s))$ elements by Corollary 14. Therefore, with probability at least $1/2$, R'_i is a row which has at least $\Omega(n/(m\theta\log^3 s))$ entries in S_θ. Hence, with high probability, in the while-loop, at least $g/4$ times we have a row R'_i which has at least $\Omega(n/(m\theta\log^3 s))$ entries in S_θ. Now we consider only such i.

If $|S' \cap S_{R'_i}| \geq |S_{R'_i}|/(32\log(1/\theta)\log s)$ happens at least $g/4$ times, we find a pair of columns with a large inner product with probability at least

$$
\begin{aligned}
1 - \left(1 - \frac{1}{32\log(1/\theta)\log s}\right)^{g/4} &\geq 1 - \exp\left(-\frac{g}{128\log(1/\theta)\log s}\right) \\
&\overset{(a)}{\geq} 1 - \exp\left(-\Omega\left(\frac{d'}{\log^4 s}\right)\right) \geq 1 - \exp\left(-\Omega\left(\frac{2^{\ell'_\theta} d}{\log^4 s}\right)\right) \overset{(b)}{\geq} \frac{2}{3},
\end{aligned}
$$

where step (a) follows from $g \geq d'/(18\log^2 s)$ and $1/\theta \leq 2s$, and step (b) follows from Eq. (12), $d \geq 1/\epsilon^2$ and $s \leq 1/2\epsilon$.

Otherwise, $|S' \cap S_{R'_i}| \leq |S_{R'_i}|/(32\log(1/\theta)\log s)$ happens at least $g/4$ times. Combining Lemma 13(ii) and the fact that $|S_{R'_i}| = \Omega(n/(m\theta\log^3 s))$, we have that $|S'|$ increases by at least $\Omega(n/(2^{\ell_\theta} m\theta \log^5 s))$ with probability $\Omega(1/\log^2 s)$ each time. Hence, by a Chernoff bound, with probability at least $1 - \exp(-\Omega(\frac{g}{\log^2 s})) = 1 - o(1)$, $|S'|$ increases at least $\Omega(g/\log^2 s)$ times and thus

$$
|S'| = \Omega\left(\frac{gn}{2^{\ell_\theta} m\theta \log^7 s}\right). \tag{15}
$$

To summarize, when Algorithm 2 terminates, either (i) we have found a pair of columns with a large inner product with probability at least $1/2$, or (ii) we have with probability at least $1/2$ a large S' satisfying Eq. (15).

In case (ii), there remain at least $g/2$ columns which have not been examined. By the construction of S', each unexamined column c has a large inner product with some examined column C'_i ($i \le g/2$) with probability at least $\Pr_{c\sim \mathrm{Unif}(G)}[c \in S'] = |S'|/n = \Omega(g/(2^{\ell_\theta} m\theta \log^7 s))$. Hence, there exists a pair of columns with a large inner product with probability at least

$$1 - \left(1 - \Omega\left(\frac{g}{2^{\ell_\theta} m\theta \log^7 s}\right)\right)^{g/2} \ge 1 - \exp\left(-\Omega\left(\frac{g^2}{2^{\ell_\theta} m\theta \log^7 s}\right)\right). \tag{16}$$

Recall that $g = \Omega(d'/\log^3 s)$, $d' = 2^{\ell'_\theta} d$ and Eq. (12). The $\Omega(\cdot)$ term on the right-hand side of Eq. (16) is thus

$$\Omega\left(\frac{d'^2}{2^{\ell_\theta} m\theta \log^{13} s}\right) = \Omega\left(\frac{2^{2\ell_\theta}\theta^2 d^2}{2^{\ell_\theta} m\theta \log^{13} s \cdot \epsilon^2 s^{2\gamma}}\right) = \Omega\left(\frac{2^{\ell_\theta}\theta d^2}{m s^{2\gamma}\epsilon^2 \log^{13} s}\right) = \Omega\left(\frac{d^2}{m s^{1+2\gamma}\epsilon^2 \log^{13} s}\right).$$

If $m \le c_0 d^2/(s^{1+2\gamma}\epsilon^2 \cdot \log^{13} s)$ for some constant c_0, then the probability in Eq. (16) is at least $2/3$. Combining all the "with high probability" events and both cases (i) and (ii), we conclude that with probability at least $1/2$, there exists a pair of columns of ΠV whose inner product is at least $p+\theta$ in absolute value.

If $\ell_\theta = -1$, we have $\ell'_\theta = 0$. In this case, we have found a pair of columns whose inner product is at least $p+\theta$ in absolute value. Recall that $p = -\theta/\log s$ and $\theta \ge ((1-\epsilon)^2 - 1/\log s)/s$ by Lemma 12, then $p + \theta > 2\epsilon$ by Eq. (13), i.e. our assumption on s. By applying Lemma 8 with $1/\beta = 1$, Π is a subspace embedding for $U = VW \sim \mathcal{D}_{2^{-\ell'_\theta}}$ with probability at most $1/8$, this contradicts Eq. (11) since $1 - 2\delta/\gamma \ge 1 - 2/K$.

If $\ell_\theta \ge 0$, we have again found a pair of columns whose inner product is at least $p+\theta$ in absolute value. Recall that $p = 2^{\ell_\theta}\theta$ and $\theta \ge ((1-\epsilon)^2 - 1/\log s)/s$ by Lemma 12, then $p + \theta > 2\epsilon 2^{\ell_\theta}$ by Eq. (13), i.e. our assumption on s. Applying Lemma 8 with $1/\beta = 2^{\ell'_\theta}$ shows that Π is a subspace embedding for $U = VW \sim \mathcal{D}_{2^{-\ell'_\theta}}$ with probability at most $1/8$, again contradicting Eq. (11).

We can therefore conclude that $m = \Omega(d^2/(s^{1+2\gamma}\epsilon^2 \cdot \log^{13} s))$. Applying Yao's minimax principle then completes the proof of Theorem 2.

5 Lower Bound for General Case

Our goal in this section is to prove Theorem 3, for OSEs with general column sparsity s. Recall that ϵ is assumed to be at most a sufficiently small constant.

We modify the definition of the hard instance $\mathcal{D}_\beta$ by replacing Rademacher variables to +1s in $U \sim \mathcal{D}_\beta$. We also define an additional hard instance $\mathcal{D}'_\beta = \mathcal{D}_{\beta/(\beta+1)}$ (under this modified definition of $\mathcal{D}_\beta$) so that $U \sim \mathcal{D}'_\beta$ has exactly $(1 + 1/\beta)$ nonzero entries per column. For notation convenience, when we say $1/\beta = 0$, it is understood that $\mathcal{D}'_\beta$ is $\mathcal{D}_1$.

Definition 15 (Modified definition of $\mathcal{D}_\beta$). *The distribution $\mathcal{D}_\beta$ ($0 < \beta \le 1$) is defined on matrices $U \in \mathbb{R}^{n\times d}$ as follows. The matrix U is decomposed as $U = VW$, where $V \in \mathbb{R}^{n\times d/\beta}$ and $W \in \mathbb{R}^{d/\beta\times d}$. The matrix V has i.i.d. columns, each $V_{\star,i}$ ($i = 1, \ldots, d/\beta$) is uniformly distributed among the n canonical basis vectors in $\mathbb{R}^n$. The matrix W is distributed as follows: For each $i = 1, \ldots, d$ and $j = (i-1)/\beta + 1, \ldots, i/\beta$, set $W_{j,i} \triangleq \sqrt{\beta}$; set the remaining entries of W to zero.*

5.1 Preliminaries

We first state a simple proposition which will be needed later. It can be proven easily by Chebyshev's inequality.

Claim 16. *Let $(X_1, \dots, X_n) \in \{0, a_1\} \times \cdots \times \{0, a_n\}$ be a tuple of bivalued random variables. If X_i's are negatively correlated and $a_i \le 1$ for all i, then $\Pr[\sum_i X_i \in [\mu/2, 3\mu/2]] \ge 1 - 4/\mu$, where $\mu \triangleq \mathbb{E}[\sum_i X_i]$.*

Proof. Let $p_i \triangleq \Pr[X_i = a_i]$. Since X_i's are negatively correlated, and $a_i \le 1$ for all i,

$$\mathrm{Var}\left[\sum_i X_i\right] \le \sum_i \mathrm{Var}[X_i] = \sum_i (a_i^2 p_i - a_i^2 p_i^2) \le \sum_i a_i p_i = \mathbb{E}\left[\sum_i X_i\right].$$

By Chebyshev's inequality,

$$\Pr\left[\left|\sum_i X_i - \mu\right| \ge \mu/2\right] \le 4 \cdot \mathrm{Var}\left[\sum_i X_i\right] / \mu^2 \le 4/\mu. \qquad \square$$

We shall apply the following proposition to show that $\|X\|_2^2$ may be large. The corollary is a trivial extension of Lemma 9.

Corollary 17. *Suppose that S is a finite multiset of vectors in the ℓ_2-ball of radius $1 + \epsilon$ and $v_1, v_2, \dots, v_k$ are k vectors independent sampled from $\mathrm{Unif}(S)$. Further suppose that $\kappa, \Delta > 0$ satisfy that $\kappa \le \Delta \le 1$, and $L = \lceil \log((1+\epsilon)^2/\Delta) \rceil$. Then one of the following must hold:*

(i)

$$\Pr\left[-k^2\kappa \le \sum_{i,j \in [k],\ i \ne j} \langle v_i, v_j \rangle \le k^2 \Delta\right] \ge \frac{\kappa}{2\Delta(L+1)};$$

(ii) there exists $i \in \{0, \dots, L-1\}$ such that

$$\Pr\left[2^i k^2 \Delta < \sum_{i,j \in [k],\ i \ne j} \langle v_i, v_j \rangle \le 2 \cdot 2^i k^2 \Delta\right] \ge \frac{\kappa}{4 \cdot 2^i \Delta(L+1)}.$$

Proof. Let $X \triangleq \sum_{i \ne j \in [k]} \langle v_i, v_j \rangle$ for simplicity. Note that if $\Pr[X \ge -k^2\kappa] \ge 1/2$ there is nothing to prove, we henceforth assume that $\Pr[X < -k^2\kappa] \ge 1/2$.

As $\mathbb{E}[X] \ge 0$, by the law of total expectation

$$\mathbb{E}[X | X \ge 0] \cdot \Pr[X \ge 0] \ge -\mathbb{E}[X | X < 0] \cdot \Pr[X < 0] \ge k^2\kappa/2.$$

Since $X \le (1+\epsilon)^2 k^2$, we have that

$$\mathbb{E}[X | X \in [0, k^2\Delta]] \cdot \Pr[X \in [0, k^2\Delta]] + \sum_{i=0}^{\lceil \log((1+\epsilon)/\Delta) \rceil - 1} \mathbb{E}[X | X \in (2^i k^2 \Delta, 2^{i+1} k^2 \Delta]] \cdot \Pr[X \in (2^i k^2 \Delta, 2^{i+1} k^2 \Delta]] \ge k^2\kappa/2.$$

By the pigeonhole principle, at least one of the $\lceil \log((1+\epsilon)/\Delta) \rceil + 1$ terms on the LHS is at least $k^2\kappa/2(\lceil \log((1+\epsilon)/\Delta) \rceil + 1)$. If it is the first term in the preceding inequality, we conclude with (i); otherwise, we conclude with (ii). $\square$

Similarly to before, when applying this corollary, we always assume that it returns an $\ell \in \{-1, \dots, L\}$, i.e. the case (i) is referred to as $\ell = -1$.

5.2 A lower bound under assumption

In this subsection we assume that every nonzero entry of Π is $\pm 1/\sqrt{s}$, each column contains exactly s nonzeroes, and Π is an OSE for every $\mathcal{D}'_\beta$ with $1/\beta \in \{0, \ldots, 1/\epsilon\}$. We shall prove the following lower bound in this subsection with these assumptions on Π and remove the assumptions in Section 5.3.

Theorem 18. *Suppose that (i) every nonzero entry of Π is $\pm 1/\sqrt{s}$, (ii) each column contains exactly $s \in (2, O(\epsilon^{-1} \log d))$ nonzeroes, and (iii) Π is an OSE for U of distribution $\mathcal{D}'_\beta$ for every $1/\beta \in \{0, \ldots, 1/\epsilon\}$. Then Π must have at least*

$$m = \Omega\left(\frac{s^{-1/(\chi-1)}}{\log^{2/(\chi-1)} s} \cdot \left(\frac{d}{\epsilon}\right)^{1+1/(\chi-1)}\right)$$

rows, where $\chi = \lfloor 4\epsilon s \rfloor + 4$.

Note that for $s = \Omega(\log(1/\epsilon)/\epsilon) \cap O((\log d)/\epsilon)$, the lower bound becomes

$$m = \Omega\left(\left(\frac{d}{\epsilon}\right)^{1+\frac{1}{\lfloor 4\epsilon s\rfloor+4}} \cdot \left(\frac{\log^3(1/\epsilon)}{\epsilon}\right)^{\Theta\left(\frac{1}{\log(1/\epsilon)}\right)}\right) = \Omega\left(\left(\frac{d}{\epsilon}\right)^{1+\frac{1}{\lfloor 4\epsilon s\rfloor+4}}\right).$$

From now on, we let $\chi \triangleq \lfloor 4\epsilon s \rfloor + 4$.

5.2.1 Partitioning OSE matrix. We partition submatrices according to the following claim.

Claim 19. *The matrix $\Pi \in \mathbb{R}^{m\times n}$ can be partitioned into $4m/s$ (combinatorial) submatrices $\Pi_1, \ldots, \Pi_{4m/s} \in \mathbb{R}^{m\times n/(4m/s)}$ such that for each Π_i with $i \leq 2m/s$ there is a row r such that every entry in row r is nonzero and has the same sign.*

The proof is easy. Whenever a matrix has at least $n/2$ columns, by the averaging principle, there must be a row which contains at least $ns/4m$ entries of $1/\sqrt{s}$ or at least $ns/4m$ entries of $-1/\sqrt{s}$. We can just partition Π in a greedy manner.

If we choose d/β columns of Π uniformly at random, then there will be $\approx (m/s)\binom{d/\beta}{\chi}(s/m)^\chi$ submatrices which receive at least χ of the d/β random columns. We are going to examine each of such submatrices from which at least χ columns are chosen. For each of such submatrices Π_i, we can choose χ canonical basis vectors $e'_1, \ldots, e'_\chi$ so that $U(e'_1 + \cdots + e'_\chi)$ covers at least χ columns in submatrix Π_i. Let X denote the sum of the χ columns in Π_i and Y the sum of the remaining columns outside of Π_i. Then

$$\|\Pi U(e_1 + \cdots + e_\chi)/\sqrt{\chi}\|_2^2 = \|X\|_2^2/(\chi/\beta + \chi) + 2\langle X, Y\rangle/(\chi/\beta + \chi) + \|Y\|_2^2/(\chi/\beta + \chi).$$

Recall that Y is simply the sum of $1/\beta$ columns of Π, it is not difficult to show that $\|Y\|_2 = (\chi/\beta)(1 \pm \epsilon)$ with probability $1 - \delta$. Thus, it suffices to focus on the first two terms.

For each $i \leq 2m/s$, let Π'_i denote the set of columns obtained from the columns in Π_i by removing row r of Π_i. We apply Corollary 17 to each of Π'_i for $i \leq 2m/s$ with $k = \chi$, $\kappa = 1/(s\log s)$, $\Delta = 1/s$ and obtain an ℓ for each Π'_i. Let ℓ_1 be the majority of these $2m/s$ values of ℓ and $\mathcal{B} \subseteq [2m/s]$ the index set of i's such that Corollary 17 returns ℓ_1 when applied to Π'_i. Note that $|\mathcal{B}| = \Omega(m/(s\cdot\log^2 s))$ with constant probability. Let X be a random variable which is obtained by sampling $i \in \mathcal{B}$ uniformly at random then summing over χ uniformly random columns in Π_i. Let $\mathcal{A}$ be the event that

$$\chi^2(1/s - 1/s\log s) \leq \|X\|_2^2 - (1 - 1/s)\chi \leq 2\chi^2/s$$

if $\ell_1 = -1$, or

$$(2^{\ell_1} + 1)\chi^2/s < \|X\|_2^2 - (1 - 1/s)\chi \leq (2^{\ell_1+1} + 1)\chi^2/s$$

if $\ell_1 \geq 0$. Note that $\Pr[\mathcal{A}] = \Omega(1/(2^{\ell_1}\log^2 s))$ if $i \in \mathcal{B}$ by Corollary 17 and Claim 19.

5.2.2 Obtaining lower bound via anticoncentration. Let $\mathcal{A}_i$ be the event that U chooses at least χ columns in Π_i, Y' a uniformly random column over Π and $\bar{Y}^\ell$ the sum of $\chi 2^\ell$ independent copies of Y'. We are going to invoke the following lemma to complete the proof.

Lemma 20. *Assume that $s = O(\log d/\epsilon)$ and $s \geq 3, d \geq 1/\epsilon^9$. If there exists $\beta = 2^{-\ell}$ such that Π is an OSE for $\mathcal{D}'_\beta$ and*

$$\mathop{\mathbb{E}}_{i\sim \mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X + \bar{Y}^\ell\right\|_2^2 \notin (1\pm\epsilon)^2(\chi/\beta+\chi)\middle|\mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^\ell\log^2 s}\right), \tag{17}$$

then it must hold that

$$m = \Omega\left(\frac{2^\ell \cdot s^{-1/(\chi-1)}}{\log^{2/(\chi-1)} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right).$$

Proof. Let $\mathcal{B}' \subseteq \mathcal{B}$ denote the set of submatrices which receive at least χ columns from U. Let X_i be indicator random variable for $i \in \mathcal{B}$ having at least χ columns from U. Note that X_i's are negatively correlated, by Claim 16, with constant probability

$$|\mathcal{B}'| = \Theta\left(|\mathcal{B}|\cdot\binom{d/\beta+d}{\chi}\left(\frac{s}{m}\right)^\chi\right). \tag{18}$$

If the RHS is $O(2^\ell\log^2 s)$, we are done. In the remainder of the proof, we thus assume that $|\mathcal{B}'| = \Omega(2^\ell\log^2 s)$ with constant probability. Choose a subset $\mathcal{B}'' \subseteq \mathcal{B}'$ of size $\Theta(2^\ell\log^2 s)$ uniformly at random over $\mathcal{B}'$. Conditioned on $|\mathcal{B}''| = k$, it is easy to see that $\mathcal{B}''$ is uniformly distributed over all k-tuples of $[2m/s]$ without replacement by the symmetry of distribution $\widetilde{\mathcal{D}}$.

Each submatrix Π_i with $i \in \mathcal{B}''$ may receive more than χ columns from U. Thus, we choose χ columns $\{C_{i,j}\}_{j\in[\chi]}$ among these columns uniformly at random. Note that $\{C_{i,j}\}_{j\in[\chi]}$ is distributed uniformly at random over $(C_i)^\chi$, where C_i is the set of all columns of submatrix Π_i. For each $\{C_{i,j}\}_{j\in[\chi]}$ with $i \in \mathcal{B}''$, let $\{e_{i,j}\}_{j\in[\chi]}$ be the set of canonical basis vectors so that $\{Ue_{i,j}\}_{j\in[\chi]}$ covers all the $\{C_{i,j}\}_{j\in[\chi]}$. Note that we may be able to cover them with less than χ canonical basis vectors, in that case we choose another canonical basis vector uniformly at random. We shall show that the set $\cup_{i\in\mathcal{B}''}\{e_{i,j}\}_{j\in[\chi]}$ is of size $|\mathcal{B}''|\chi$ with probability $1-o(1)$, i.e. $\{e_{i,j}\}_{j\in[\chi]}$'s are mutually disjoint.

Fix any selection of U over all columns of Π, which is a multiset of $d/\beta + d$ columns of Π. Note that this fixes all the $\{C_{i,j}\}_{i\in\mathcal{B}'',j\in[\chi]}$ as well. Then the relationship between e_j's and the aforesaid multiset behaves in the following manner: each e_j chooses $1/\beta + 1$ columns from the multiset uniformly at random, and any two distinct e_j, e_k never choose the same column from the multiset. Notice that (i) $|\{C_{i,j}\}_{i\in\mathcal{B}'',j\in[\chi]}| = O(\chi 2^\ell\log^2 s) = O(s^2\epsilon\log^2 s) = O(\log^3 d/\epsilon^2)$, (ii) for each e_j, Ue_j covers precisely $1/\beta + 1 = O(\log d/\epsilon)$ columns, (iii) U chooses $d/\beta + d = \omega(\log^8 d/\epsilon^6)$ columns in total, any e_j is associated with at most one column in $\{C_{i,j}\}_{i\in\mathcal{B}'',j\in[\chi]}$ with probability $1-o(1)$ by a simple calculation.

Therefore, we conclude that with constant probability, there is a set $\mathcal{B}'' \subseteq [2m/s]$ of indices of submatrices such that

(1) $|\mathcal{B}''| = \Theta(2^\ell\log^2 s)$;
(2) $\mathcal{B}''$ distributed over $[2m/s]$ uniformly at random without replacement;
(3) for each $i \in \mathcal{B}''$, there is a set $\{e_{i,j}\}_{j\in[\chi]}$ of χ canonical basis vectors such that $\{Ue_{i,j}\}_{j\in[\chi]}$ covers χ vectors in Π_i, and the χ columns are distributed uniformly at random over all the columns of Π_i with replacement, the remaining χ/β columns are distributed uniformly at random over all the columns of Π with replacement;

(4) $\cup_{i\in\mathcal{B}''}\{e_{i,j}\}_{j\in[\chi]}$ is of size $|\mathcal{B}''|\cdot\chi$, i.e. $\{e_{i,j}\}_{j\in[\chi]}$'s are mutually disjoint;
(5) $\{Ue_{i,j}\}_{j\in[\chi]}$'s are mutually independent for different $i\in\mathcal{B}''$.

For simplicity, let $p_i \triangleq \Pr[\|X+\bar{Y}^\ell\|_2 \notin (1\pm\epsilon)(\chi/\beta+\chi)|\mathcal{A}_i]$. Now we examine each of $\Pi U(e_{i,1}+\cdots+e_{i,\chi})/\sqrt{\chi}$ for $i\in\mathcal{B}''$. Recall that none of the canonical basis vectors is examined more than once, therefore, X and Y from different $i\in\mathcal{B}''$ are mutually independent by properties (4) and (5). By properties (2) and (3), there exists $i\in\mathcal{B}''$ such that $\|X+\bar{Y}^\ell\|_2\notin(1\pm\epsilon)(\chi/\beta+\chi)$ with probability

$$1-\prod_{i\in\mathcal{B}''}(1-p_i)\ge 1-\prod_{i\in\mathcal{B}''}e^{-p_i}=1-\exp\left(\sum_{i\in\mathcal{B}''}p_i\right)=\Omega\left(\min\left\{1,\sum_{i\in\mathcal{B}''}p_i\right\}\right).$$

Define for each i a random variable X_i such that $X_i=p_i$ for $i\in\mathcal{B}''$ and $X_i=0$ otherwise, then $\sum_i X_i=\sum_{i\in\mathcal{B}''}p_i$. The assumption (17) states that $\mathbb{E}_{i\sim\mathrm{Unif}(\mathcal{B})}\,p_i=\Omega(1/(2^\ell\log^2 s))$. Note that X_i's are negatively correlated, and $\mathbb{E}[X_i]=\Omega(|\mathcal{B}''|p_i/|\mathcal{B}|)$, so $\mathbb{E}[\sum_{i\in\mathcal{B}''}p_i]=\Omega(|\mathcal{B}''|/2^\ell\log^2 s)$. As $|\mathcal{B}''|=\Theta(2^\ell\log^2 s)$, by property (2) and Claim 16, $\sum_{i\in\mathcal{B}''}p_i=\Omega(|\mathcal{B}'|/(2^\ell\log^2 s))$ with probability $1-o(1)$. Therefore, as long as there exists such a $\mathcal{B}''$ of size $\Theta(2^\ell\log^2 s)$, there is an $i\in\mathcal{B}''$ such that $\|X+\bar{Y}^\ell\|_2\notin(1\pm\epsilon)(\chi/\beta+\chi)$ with constant probability. Hence it must hold that

$$|\mathcal{B}'|=O(2^\ell\log^2 s).$$

Plugging in (18) and $|\mathcal{B}|=\Omega(m/(s\cdot\log^2 s))$, we obtain that

$$m=\Omega\left(\frac{2^\ell}{(s\log^2 s)^{1/(\chi-1)}}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right).\qquad\square$$

5.2.3 *The anticoncentration.* Conditioned on event $\mathcal{A}$, we have $\|X\|_2^2\approx\chi+2^{\ell_1}\chi^2/s$. We now examine $\langle X,\bar{Y}^\ell\rangle$ to find a good parameter β to meet the conditions of Lemma 20. Recall that $\chi=\lfloor 4\epsilon s\rfloor+4$, so $(1-1/\log s)\chi/s>2\epsilon+\epsilon^2+1/s$.

Case 1: $\ell_1\ge 3$ and $\Pr[\langle X,\bar{Y}^{\ell_1-4}\rangle\ge-(1/4)\cdot 2^{\ell_1}\chi^2/s|\mathcal{A}]\ge 1/2$. By a union bound, (i) $\langle X,\bar{Y}^{\ell_1-3}\rangle$ does not cancel $\|X\|_2^2-(1+\epsilon)^2\chi$, and (ii) $\|Y\|_2\ge(1-\epsilon)^2(\chi/\beta)$ happen simultaneously with probability $1/2-\delta$. To summarize, if $U\sim\mathcal{D}_\beta$ with $1/\beta=2^{\ell_1-3}$, conditioned on $\mathcal{A}$,

$$\begin{aligned}\|X\|_2^2+2\langle X,Y\rangle+\|Y\|_2^2&\ge(1-1/s)\chi+(2^{\ell_1}+1)\chi^2/s-2\cdot(1/4)\cdot 2^{\ell_1}\chi^2/s+(\chi/\beta)(1-\epsilon)^2\\&>(1-\epsilon)^2(\chi/\beta+\chi)+(1/2)\cdot 2^{\ell_1}\chi^2/s\\&\ge(1-\epsilon)^2(\chi/\beta+\chi)+2(\chi/s)(\chi/\beta+\chi)\\&>(1+\epsilon)^2(\chi/\beta+\chi)\qquad\text{since }\chi/s>4\epsilon\end{aligned}$$

with probability at least $1/2-\delta$. It can be rephrased as

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X+\bar{Y}^{\ell_1-3}\right\|_2^2>(1+\epsilon)^2(\chi/\beta+\chi)\middle|\mathcal{A}_i,\mathcal{A}\right]\right]\ge\frac{1}{2}-\delta,$$

which implies that

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X+\bar{Y}^{\ell_1-3}\right\|_2^2>(1+\epsilon)^2(\chi/\beta+\chi)\middle|\mathcal{A}_i\right]\right]=\Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

Case 2: $\ell_1<3$. We follow the argument in Case 1 but apply $U\sim\mathcal{D}_\beta$ with $1/\beta=0$ to Π, then obtain that conditioned on $\mathcal{A}$,

$$\|X\|_2^2\ge(1-1/s)\chi+\chi^2(1/s-1/s\log s)>(1+\epsilon)^2\chi,$$

where the second inequality is because $(1 - 1/\log s)\chi/s > 2\epsilon + \epsilon^2 + 1/s$. Then we conclude

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\|X\|_2^2 > (1+\epsilon)^2\chi \middle| \mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

Case 3: $\ell_1 \geq 3$ *and* $\Pr[\langle X, \bar{Y}^{\ell_1-3}\rangle < -(1/4)\cdot 2^{\ell_1}\chi^2/s|\mathcal{A}] \geq 1/2$. It then follows that

$$\begin{aligned}\Pr[\langle X, \bar{Y}^{\ell_1}\rangle < -2\cdot 2^{\ell_1}\chi^2/s] &= \mathbb{E}[\Pr[\langle X, \bar{Y}^{\ell_1}\rangle < -2\cdot 2^{\ell_1}\chi^2/s]|X]\\ &\geq \mathbb{E}[\Pr[\langle X, \bar{Y}^{\ell_1-3}\rangle < -(1/4)\cdot 2^{\ell_1}\chi^2/s]^8|X]\\ &\geq \mathbb{E}[\Pr[\langle X, \bar{Y}^{\ell_1-3}\rangle < -(1/4)\cdot 2^{\ell_1}\chi^2/s]|X]^8\\ &= \Pr[\langle X, \bar{Y}^{\ell_1-3}\rangle < -(1/4)\cdot 2^{\ell_1}\chi^2/s]^8\\ &\geq 1/256,\end{aligned}$$

where the second inequality follows from Jensen's inequality and the convexity of $x \mapsto x^8$. By a union bound, with probability at least $1/256 - \delta$, (i) $\langle X, \bar{Y}^{\ell_1}\rangle$ dominates $\|X\|_2 - (1-\epsilon)\chi$, and (ii) $\|Y\|_2 \leq (1+\epsilon)(\chi/\beta - \chi)$ happen simultaneously. To summarize, if $U \sim \mathcal{D}_\beta$ with $1/\beta = 2^{\ell_1}$, conditioned on $\mathcal{A}$,

$$\begin{aligned}\|X\|_2^2 + 2\langle X, Y\rangle + \|Y\|_2^2 &\leq (1-1/s)\chi + (2^{\ell_1+1}+1)\chi^2/s - 2\cdot 2\cdot 2^{\ell_1}\chi^2/s + (1+\epsilon)^2(\chi/\beta)\\ &< (1+\epsilon)^2(\chi/\beta + \chi) - (2\cdot 2^{\ell_1} - 1)\chi^2/s\\ &\leq (1 + 2\epsilon + \epsilon^2 - (5/3)\chi/s)(\chi/\beta + \chi)\\ &< (1 - 3\epsilon + \epsilon^2)(\chi/\beta + \chi)\end{aligned}$$

with probability at least $1/256 - \delta$, where the last inequality is because $\chi/s > 4\epsilon$. It can be rephrased as

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X + \bar{Y}^{\ell_1}\right\|_2^2 < (1-\epsilon)^2(\chi/\beta)\middle|\mathcal{A}_i, \mathcal{A}\right]\right] \geq \frac{1}{256} - \delta,$$

which implies that

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X + \bar{Y}^{\ell_1}\right\|_2^2 < (1-\epsilon)^2(\chi/\beta)\middle|\mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

5.3 Removing the assumption

We find the most "popular" value of the entries of Π with the following lemma.

Lemma 21 (Refined from Lemma 12). *Suppose that* Π *is a matrix with column sparsity at most* $s \geq 3$. *Then there exists a* $\theta \geq (1-\epsilon)^2/s - 1/s\log s$ *such that*

$$\mathop{\mathbb{E}}_{j}\left[\sum_{k:|\Pi_{k,j}|^2\in[\theta,2\theta)}\left|\Pi_{k,j}\right|^2 \middle| \left\|\Pi_{\star,j}\right\|_2 = 1\pm\epsilon\right] = \Omega(1/\log^2 s). \tag{19}$$

We apply the preceding lemma to Π, then let $\chi \triangleq \lfloor(4\epsilon+\theta)/\theta(1-1/\log s)\rfloor + 1 \geq 2$.

Let distribution $\widetilde{\mathcal{D}}$ be a mixture of $\mathcal{D}_\beta$ and $\mathcal{D}'_\beta$, which are parameterized by β, on $n\times d$ matrices. With probability $1/3$, $\widetilde{\mathcal{D}} = \mathcal{D}_1$; with probability $1/3$, $\widetilde{\mathcal{D}}$ is a $\mathcal{D}'_\beta$ for $\beta \sim \mathrm{Unif}([\lceil\log s\rceil])$; and with probability $1/3$, $\widetilde{\mathcal{D}}$ is a $\mathcal{D}_\beta$ for $\beta \sim \mathrm{Unif}([\lceil\log s\rceil])$.

We shall prove the following theorem in this subsection.

Theorem 22. *Suppose that (i) $s = O(\epsilon^{-1}\log d)$ and $s \geq 3$, and (ii) Π is an OSE for $U \sim \widetilde{\mathcal{D}}$. Then Π must have at least*

$$m = \Omega\left(\frac{\theta^{1/(\chi-1)}}{s^{8\gamma/(\chi-1)}\log^{6/(\chi-1)+2} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right)$$

rows, where $\chi = \lfloor(4\epsilon+\theta)/(\theta(1-1/\log s))\rfloor + 1 \geq 2$ and $\theta \geq (1-\epsilon)^2/s - 1/s\log s$.

Note that $4\epsilon/\theta = 4\epsilon s(1+2\epsilon+O(\epsilon^2)+O(1/\log s))$. Thus for $\epsilon \leq 1/202$ and $s = \Omega(\log(1/\epsilon)/\epsilon) \cap O((\log d)/\epsilon)$, the lower bound is at least

$$m = \Omega\left(\left(\frac{d}{\epsilon}\right)^{1+\frac{1}{4\epsilon(1+\epsilon)s}}\cdot\frac{1}{s^{O(\delta)}}\right).$$

5.3.1 Partitioning OSE matrix. Since Π is an OSE for $U \sim \widetilde{\mathcal{D}}$, Π is also an OSE for $U \sim \mathcal{D}_1$. Thus $\|\Pi_{\star,j}\|_2 = 1 \pm \epsilon$ for at least $n(1-3\delta/d)$ columns j of Π by Lemma 23.

Lemma 23 (Refined from [8, Lemma 6]). *If $\Pi \in \mathbb{R}^{m\times n}$ is an (ϵ,δ)-subspace-embedding for $U \sim \widetilde{\mathcal{D}}$, then $(1-3\delta/d)$-fraction of the columns of Π has ℓ_2-norm of $1 \pm \epsilon$.*

Now we partition Π with new parameters.

Claim 24 (Refined from Claim 19). *Assume $\Pi \in \mathbb{R}^{m\times n}$ is of column sparsity $s \geq 3$. Π can be partitioned into $0.05m\theta\log^2 s$ submatrices $\Pi_1,\ldots,\Pi_{0.05m\theta\log^2 s} \in \mathbb{R}^{m\times 20n/(m\theta\log^2 s)}$ such that for each Π_i with $i \leq 0.024m\theta$ (1) every column has ℓ_2-norm $1 \pm \epsilon$, and (2) there is a row r such that every entry in row r (i) is in $[\sqrt{\theta},\sqrt{2\theta}]$ and (ii) has the same sign.*

For each $i \leq 0.024m\theta$, let Π'_i denote the set of columns obtained from the set of columns in Π_i after removing row r. We apply Corollary 17 to each Π'_i for $i \leq 0.024m\theta$ with $k = \chi$, $\kappa = \theta/\log s$, $\Delta = \theta$, and obtain an ℓ for each Π'_i. Let ℓ_1 be the majority of these $0.024m\theta$ values of ℓ and $\mathcal{B} \subseteq [0.024m\theta]$ the index set of i's such that Corollary 17 returns ℓ_1 when applied to Π'_i with $i \leq 0.024m\theta$. Note that $|\mathcal{B}| = \Omega(m\theta/\log^2 s)$. Let X be a random variable which is obtained by sampling $i \in \mathcal{B}$ uniformly at random then summing over χ uniformly random column vectors in Π_i, $X'_1,\ldots,X'_\chi$ the χ random column vectors with zeroing out r-th row, $x_1,\ldots,x_\chi$ the values of the r-th row in Π_i. Then

$$\|X\|_2^2 = \sum_{i,j\in[\chi]}(\langle X'_i,X'_j\rangle + x_ix_j) = \sum_{i,j\in[\chi]:i\neq j}(\langle X'_i,X'_j\rangle + x_ix_j) + \sum_i(\|X'_i\|_2^2 + x_i^2).$$

Recall that each column in Π_i is of ℓ_2-norm $1\pm\epsilon$, so $\sum_i(\|X'_i\|_2^2+x_i^2) = (1\pm\epsilon)^2\chi$. Recall that each entry in row r and in Π_i is in $[\sqrt{\theta},\sqrt{2\theta}]$ and has the same sign, so $\sum_{i\neq j}x_ix_j \in [\chi(\chi-1)\theta, 2\chi(\chi-1)\theta]$. Finally, $\sum_{i,j\in[\chi]:i\neq j}(\langle X'_i,X'_j\rangle$ is characterized by Corollary 17. Let $\mathcal{A}$ be the event that

$$(1-\epsilon)^2\chi - \theta\chi + \chi^2(\theta-\theta/\log s) \leq \|X\|_2^2 \leq (1+\epsilon)^2\chi - \theta\chi + 3\chi^2\theta$$

if $\ell_1 = -1$, or

$$(1-\epsilon)^2\chi - \theta\chi + (2^{\ell_1}+1)\chi^2\theta < \|X\|_2^2 \leq (1+\epsilon)^2\chi - \theta\chi + (2^{\ell_1+1}+2)\chi^2\theta$$

if $\ell_1 \geq 0$. Note that $\Pr[\mathcal{A}] = \Omega(1/(2^{\ell_1}\log^2 s))$ if $i \in \mathcal{B}$ by Corollary 17 and Claim 24.

5.3.2 Obtaining lower bound via anticoncentration. Let $\mathcal{A}_i$ be the event that U chooses at least χ columns in Π_i, Y' a uniformly random column over Π, $\bar{Y}^\ell$ the sum of $\chi 2^\ell$ independent copies of Y'. We are going to invoke the following lemma to complete the proof.

Lemma 25. *Assume that $s = O(\epsilon^{-1}\log d)$, $s \geq 3$, and $d \geq 1/\epsilon^7$. If there exists $\ell, 1/\beta \geq 0$ such that Π is an OSE for $\mathcal{D}'_\beta$ and*

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X+\bar{Y}^{\ell}\right\|_2^2 \notin (1\pm\epsilon)^2(\chi/\beta+\chi)\middle|\mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^\ell\log^2 s}\right), \tag{20}$$

then it must hold that

$$m = \Omega\left(\frac{2^{-\ell/(\chi-1)}(1+1/\beta)^{\chi/(\chi-1)}}{\theta^{-1/(\chi-1)}\log^{6/(\chi-1)+2} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right).$$

Proof. The proof is almost identical with the proof of Lemma 20, except the calculations. In this case, we have

$$|\mathcal{B}'| = \Theta\left(|\mathcal{B}|\cdot\binom{d(1+1/\beta)}{\chi}\left(\frac{1}{m\theta\log^2 s}\right)^\chi\right).$$

Also, β is not fixed to 2^{ℓ_1}, so

$$|\mathcal{B}|\cdot\binom{d(1+1/\beta)}{\chi}\left(\frac{1}{m\theta\log^2 s}\right)^\chi = O(2^\ell\log^2 s).$$

Algebraic manipulation then leads to the claimed lower bound of m. □

5.3.3 The anticoncentration. Recall that Π is an (ϵ,δ)-subspace-embedding for $U\sim\widetilde{\mathcal{D}}$. Recall that we assume $n \geq K(sd)^2/\delta$ for large enough constant K so that $U\sim\widetilde{\mathcal{D}}$ is an isometry with probability $1-\delta/(2K)$. By Markov's inequality, for $(1-\gamma)$-fraction of $\ell\in[L]$, Π is an $(\epsilon, 2\delta/\gamma)$-subspace-embedding for $U\sim\mathcal{D}_{2^{-\ell}}$, where $\gamma\triangleq K\delta = O(\delta)$, so that $2\delta/\gamma = 2/K$ is a small constant. Similarly, for $(1-\gamma)$-fraction of $\ell\in[L]$, Π is an $(\epsilon,2\delta/\gamma)$-subspace-embedding for $U\sim\mathcal{D}'_{2^{-\ell}}$, where $\gamma\triangleq K\delta = O(\delta)$, so that $2\delta/\gamma = 2/K$ is a small constant.

Conditioned on $\mathcal{A}$, we have $\|X\|_2 \approx \chi + 2^{\ell_1}\chi^2\theta$. We now examine $\langle X, \bar{Y}^\ell\rangle$ to find a good parameter β to meet the conditions of Lemma 25. Recall that $\chi = \lfloor(4\epsilon+\theta)/(\theta(1-1/\log s))\rfloor + 1 > 4\epsilon/\theta$, so

$$-\theta + \chi\theta - \chi\theta/\log s > 4\epsilon. \tag{21}$$

We follow the same argument as in the previous subsection. To this end, we shall find an integer $\ell_2 \leq \ell_1$ such that Π is an $(\epsilon, 2/K)$-subspace-embedding for $U\sim\mathcal{D}_{2^{\ell_2}}$, $U\sim\mathcal{D}'_{2^{\ell_2}}$, $U\sim\mathcal{D}_{2^{\ell_2-3}}$ and $U\sim\mathcal{D}'_{2^{\ell_2-3}}$ simultaneously. By the pigeonhole principle, for $\ell_1 \geq 8\gamma\log s + 3$, there is an integer $\ell_2\in[\ell_1 - 8\gamma\log s, \ell_1]$ which meets the preceding requirement. Note that

$$2^{\ell_2} = \Omega(2^{\ell_1}/s^{8\gamma}). \tag{22}$$

Case 1: $\ell_1 > 8\gamma\log s + 3$ and $\Pr[\langle X, \bar{Y}^{\ell_2-3}\rangle \geq -(1/4)\cdot 2^{\ell_1}\chi^2\theta] \geq 1/2$. By our assumption, if we examine ΠU with $1/\beta = 2^{\ell_2-3} \leq 2^{\ell_1}/8$

$$\begin{aligned}\|X\|_2^2 + \|Y\|_2^2 + 2\langle X, Y\rangle \geq &(1-\epsilon)^2(\chi/\beta+\chi) - \theta\chi + (2^{\ell_1-1}+1)\chi^2\theta\\ \geq &(1-2\epsilon+\epsilon^2+2(\chi\theta))(\chi/\beta+\chi) > (1+\epsilon)^2(\chi/\beta+\chi)\end{aligned}$$

with probability at least $1/2 - 2/K$, where the last inequality is because $\chi\theta > 4\epsilon$. So we have for $1/\beta = 2^{\ell_2-3} \geq 2^{\ell_1}/8s^{8\gamma}$ that

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X+\bar{Y}^{\ell_1-3}\right\|_2^2 > (1+\epsilon)^2(\chi/\beta+\chi)\middle|\mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

Case 2: $\ell_1 \leq 8\gamma \log s + 3$. We follow the argument in Case 1 but apply $U \sim \mathcal{D}'_\beta$ with $1/\beta = 0$ to Π, then obtain that conditioned on $\mathcal{A}$,

$$\|X\|_2^2 \geq (1-\epsilon)^2\chi - \theta\chi + \chi^2(\theta - \theta/\log s) > (1+\epsilon)^2\chi,$$

where the second inequality is due to Eq. (21). Then we conclude for $1/\beta = 0 \geq 2^{\ell_1}/8s^{8\gamma} - 1$

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\|X\|_2^2 > (1+\epsilon)^2\chi \middle| \mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

Case 3: $\ell_1 > 8\gamma \log s + 3$ *and* $\Pr[\langle X, \bar{Y}^{\ell_2-3}\rangle < -(1/4)\cdot 2^{\ell_1}\chi^2\theta] \geq 1/2$. By our assumption, if we examine ΠU with $1/\beta = 2^{\ell_2}$

$$\begin{aligned}\|X\|_2^2 + \|Y\|_2^2 + 2\langle X, Y\rangle \leq& (1+\epsilon)^2(\chi/\beta + \chi) - \theta\chi - (2\cdot 2^{\ell_1} - 2)\chi^2\theta \\ \leq& (1 + 2\epsilon + \epsilon^2 - (5/3)(\chi\theta))(\chi/\beta + \chi) < (1-\epsilon)^2(\chi/\beta + \chi)\end{aligned}$$

with probability at least $1/256 - 2/K$, where the last inequality is because $\chi\theta > 4\epsilon$. So we have for $1/\beta = 2^{\ell_2} \geq 2^{\ell_1}/s^{8\gamma}$ that

$$\mathop{\mathbb{E}}_{i\sim\mathrm{Unif}(\mathcal{B})}\left[\Pr\left[\left\|X + \bar{Y}^{\ell_1}\right\|_2^2 < (1-\epsilon)^2(\chi/\beta + \chi)\middle| \mathcal{A}_i\right]\right] = \Omega\left(\frac{1}{2^{\ell_1}\log^2 s}\right).$$

Applying Lemma 25 with the values of β in Cases 1 and 3, we obtain

$$m = \Omega\left(\frac{2^{-\ell_1/(\chi-1)}(1+2^{\ell_2})^{\chi/(\chi-1)}}{\theta^{-1/(\chi-1)}\log^{6/(\chi-1)+2} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right)$$

which by Eq. (22) is at least

$$m = \Omega\left(\frac{2^{\ell_1}\cdot\theta^{1/(\chi-1)}}{s^{8\gamma\chi/(\chi-1)}\log^{6/(\chi-1)+2} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right).$$

In Case 2, we have

$$m = \Omega\left(\frac{\theta^{1/(\chi-1)}}{(8s^{8\gamma})^{1/(\chi-1)}\log^{6/(\chi-1)+2} s}\cdot\left(\frac{d}{\epsilon}\right)^{\chi/(\chi-1)}\right).$$

Acknowledgments

Yi Li was supported in part by Singapore Ministry of Education AcRF Tier 2 grant MOE-T2EP20122-0001 and Tier 1 grant RG21/25. Part of this work was carried out while Mingmou Liu was a postdoctoral researcher at the University of Copenhagen, supported by VILLUM Foundation grant 16582. Mingmou Liu was also supported in part by startup funding from Nanjing University, the Nanjing University "Miaopu Program" Leading Talent Grant, and the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM118).

References

[1] Haim Avron, Christos Boutsidis, Sivan Toledo, and Anastasios Zouzias. 2014. Efficient Dimensionality Reduction for Canonical Correlation Analysis. *SIAM Journal on Scientific Computing* 36, 5 (2014), S111–S131. doi:10.1137/130919222

[2] Christos Boutsidis, Anastasios Zouzias, Michael W. Mahoney, and Petros Drineas. 2015. Randomized Dimensionality Reduction for k-Means Clustering. *IEEE Trans. Inf. Theory* 61, 2 (2015), 1045–1062. doi:10.1109/TIT.2014.2375327

[3] Shabarish Chenakkod, Michał Dereziński, and Xiaoyu Dong. 2025. Optimal Oblivious Subspace Embeddings with Near-Optimal Sparsity. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 334)*, Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 55:1–55:20. doi:10.4230/LIPIcs.ICALP.2025.55

[4] Shabarish Chenakkod, Michał Dereziński, Xiaoyu Dong, and Mark Rudelson. 2024. Optimal Embedding Dimension for Sparse Subspace Embeddings. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (Vancouver, BC, Canada) *(STOC 2024)*. Association for Computing Machinery, New York, NY, USA, 1106–1117. doi:10.1145/3618260.3649762

[5] Kenneth L. Clarkson and David P. Woodruff. 2017. Low-Rank Approximation and Regression in Input Sparsity Time. *Journal of the ACM* 63, 6 (feb 2017), 1–45. doi:10.1145/3019134

[6] Michael B. Cohen. 2016. Nearly Tight Oblivious Subspace Embeddings by Trace Inequalities. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (Arlington, Virginia) *(SODA '16)*. Society for Industrial and Applied Mathematics, USA, 278–287.

[7] Michael B. Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. 2015. Dimensionality Reduction for k-Means Clustering and Low Rank Approximation. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* (Portland, Oregon, USA) *(STOC '15)*. Association for Computing Machinery, New York, NY, USA, 163–172. doi:10.1145/2746539.2746569

[8] Yi Li and Mingmou Liu. 2022. Lower Bounds for Sparse Oblivious Subspace Embeddings. In *Proceedings of the 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (Philadelphia, PA, USA) *(PODS '22)*. Association for Computing Machinery, New York, NY, USA, 251–260. doi:10.1145/3517804.3526224

[9] Jelani Nelson and Huy L. Nguyen. 2013. OSNAP: Faster Numerical Linear Algebra Algorithms via Sparser Subspace Embeddings. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (Berkeley, CA, USA) *(FOCS'13)*. IEEE Computer Society, Washington, DC, USA, 117–126. doi:10.1109/FOCS.2013.21

[10] Jelani Nelson and Huy L. Nguyen. 2013. Sparsity lower bounds for dimensionality reducing maps. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (Palo Alto, California, USA) *(STOC '13)*. Association for Computing Machinery, New York, NY, USA, 101–110. doi:10.1145/2488608.2488622

[11] Jelani Nelson and Huy L. Nguyen. 2014. Lower Bounds for Oblivious Subspace Embeddings. In *Automata, Languages, and Programming*, Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 883–894.

Received December 2025; accepted February 2025