

Well-Conditioned Oblivious Perturbations in Linear Space*

Shabarish Chenakkod† Michał Dereziński† Xiaoyu Dong‡ Mark Rudelson†

Abstract. Perturbing a deterministic n -dimensional matrix with small Gaussian noise is a cornerstone of smoothed analysis of algorithms [Spielman and Teng, JACM 2004], as it reduces the condition number of the input to $O(n)$, and with it the complexity of many matrix algorithms. However, when deployed algorithmically, these perturbations are expensive due to the cost of generating and storing n^2 Gaussian random variables. We propose a perturbation that requires generating and storing $O(n)$ random numbers in $O(\log n)$ bits of precision, and reduces the condition number of any deterministic matrix to $O(n)$, matching Gaussian perturbations. Our result takes a step towards derandomizing matrix perturbations in numerical linear algebra, and in particular, implies that we can solve an $n \times n$ linear system to within ε backward error in $O(n^2 \log(n)/\varepsilon)$ time using only $O(n \log n)$ random bits.

In our construction, we introduce the concept of a pattern matrix, which is a dense deterministic matrix that maps all sparse vectors into dense vectors in a way that goes beyond the classical uncertainty principle of the Fourier transform. We combine this with a sparse perturbation whose entries are dependent and located in a non-uniform fashion. In order to analyze this construction, we develop new techniques for lower bounding the smallest singular value of a random matrix with dependent entries.

1 Introduction Going beyond the worst-case guarantees is a central challenge in closing the theory-practice gap for modern algorithm analysis. One of the most effective tools towards achieving this goal has been *smoothed analysis*. Proposed by Spielman and Teng [33], this framework considers the performance of an algorithm on an arbitrary input perturbed randomly by small Gaussian noise. Such small perturbations are known to escape the hard corner cases that arise in the analysis of numerous algorithms, which has led to new guarantees for classical methods such as the simplex algorithm in linear programming [33] or Gaussian elimination for solving systems of linear equations [30], as well as a new understanding of other popular algorithms, including in machine learning [18, 24], game theory [11], and computational linear algebra [7, 9].

Most techniques in smoothed analysis require ensuring that a random perturbation of the input escapes certain degenerate corner cases, which are typically instances where an appropriate matrix is nearly singular. This is attained by leveraging tools from random matrix theory to show that a perturbed matrix is sufficiently well-conditioned to ensure good performance of the algorithm. The simplest and most classical formulation of this perturbation property, which directly arises as a non-degeneracy condition for many algorithms in computational linear algebra, is the following.

DEFINITION 1.1. *An $n \times n$ random matrix R is a $(\kappa, \varepsilon, \delta)$ -(well-conditioned oblivious) perturbation if for any deterministic matrix $A \in \mathbb{R}^{n \times n}$ with unit spectral norm (denoted $\|A\| = 1$), we have:*

$$\mathbb{P}\left(\kappa(A + \varepsilon R) \leq \kappa \quad \text{and} \quad \|R\| \leq 1\right) \geq 1 - \delta, \quad \text{where} \quad \kappa(M) = \|M\| \|M^{-1}\|.$$

In other words, a $(\kappa, \varepsilon, \delta)$ -perturbation is one that, for any matrix, with probability $1 - \delta$ introduces at most ε relative error and reduces its condition number to at most κ . We refer to this as an *oblivious* perturbation because the same distribution over R has to apply to any input matrix A . Sankar, Spielman, and Teng [30] showed that an appropriately scaled Gaussian matrix is a $(\kappa, \varepsilon, \delta)$ -perturbation with condition number that scales linearly with n , namely $\kappa = O(n/\varepsilon\delta)$, and they used this to provide smoothed analysis of Gaussian elimination in finite-precision arithmetic. The condition number κ also controls the complexity of many iterative numerical algorithms such

*The full version of the paper can be accessed at <https://arxiv.org/abs/0000.00000>

†University of Michigan, Ann Arbor, MI, USA (shabari@umich.edu, derezin@umich.edu, rudelson@umich.edu).

‡National University of Singapore, Singapore (xdong@nus.edu.sg).

as Krylov subspace methods [15] and interior-point methods [25], and is a necessary component in the stability analysis of others, such as algorithms for matrix diagonalization [3].

Smoothed analysis can be viewed from a purely algorithmic perspective, where one actually performs the perturbation before solving the problem instance. This viewpoint has been particularly prevalent in numerical linear algebra [30, 12, 21, 23], where a predominant way to evaluate an algorithm is through *backward error*, i.e., the distance to the nearest problem instance solved by its output [16, 13]. In particular, an algorithm that applies a $(\kappa, \varepsilon, \delta)$ -perturbation and then solves the resulting task incurs at most ε backward error. The notion of backward error dates back to von Neumann and Goldstine [38] in the early days of computing, and it was developed as a rigorous framework by Wilkinson in the late 1950s [39, 40] in order to analyze the effect of perturbations occurring as a result of rounding errors. The success of smoothed analysis has shown that injecting deliberate perturbations into the data that go beyond the effect of rounding errors can in fact be beneficial in attaining strong backward error performance in modern algorithms.

While Gaussian perturbations are often the most convenient for the analysis, they are typically not the most practical (or sometimes not even tractable) due to their high cost in both randomness and space requirements. For instance, an $n \times n$ Gaussian matrix requires generating n^2 Gaussian random variables, each taking $O(\log n)$ random bits when properly discretized, which is expensive to generate, and it may also exceed the memory requirements of the input data (e.g., when it is sparse or otherwise structured). This motivates the broader goal of derandomizing matrix perturbations, which was raised recently by Bürgisser and Srivastava [1] during a workshop at the Simons Institute in Berkeley. As a step towards that goal, we ask:

How many random bits do we need to generate an $n \times n$ well-conditioned oblivious perturbation?

A natural strategy for reducing the random bit complexity of a random matrix is to sparsify it by replacing most of its entries with zeros. This places our question in the context of a line of works in random matrix theory studying the spectral properties of sparse matrices. In this context, Tao and Vu [35] showed that a $\text{poly}(n)$ -conditioned perturbation can be obtained from a matrix with $n^{1+\theta}$ randomly placed non-zero entries for any fixed $\theta > 0$. Later, Shah, Srivastava, and Zeng [32] optimized their analysis, reducing the sparsity to $O(n \log^2 n)$ non-zero entries, i.e., $O(n \log^3 n)$ random bits since $\log n$ bits are used for each entry, at the cost of a super-polynomial dependence in the condition number, $\kappa = n^{O(\frac{\log n}{\log \log n})}$. They also observed that $n \log n$ non-zeros and $n \log^2 n$ random bits is the natural limit for this approach since below that limit the resulting perturbation becomes singular due to coupon collector problem, and thus cannot be well-conditioned.

1.1 Main result In this work, we improve upon the prior results in two ways. First, we construct a well-conditioned oblivious perturbation using only $O(n \log n)$ random bits, overcoming the coupon collector barrier from previous approaches. In particular, our perturbation can be stored in linear space, i.e., using $O(n)$ numbers of $\text{poly}(n)$ size. Second, we show that the condition number of this perturbation is $O(n)$, i.e., it scales linearly with the dimension, matching dense Gaussian matrices.

THEOREM 1.2. *Given $n \geq 1$ and $\varepsilon, \delta \in (0, 1)$, using $n \log(n) \cdot \text{poly}(1/\delta)$ random bits we can generate an $n \times n$ $(\kappa, \varepsilon, \delta)$ -well-conditioned oblivious perturbation with condition number $\kappa = n \cdot \text{poly}(1/\varepsilon\delta)$.*

In order to overcome the coupon collector barrier for uniformly sparse perturbations, we develop a new construction, which adds together two crucial components:

1. A dense deterministic matrix whose rows and columns are multiplied by $O(n)$ random signs;
2. A sparse random matrix with $O(n)$ random sign entries placed in a non-uniform fashion.

Our analysis of this construction is centered on lower bounding the smallest singular value of the perturbed matrix, $s_n(M)$ for $M = A + \varepsilon R$, which is the infimum of $\|Mx\|_2$ over all x in the unit sphere. Our starting point is a standard approach of decomposing the sphere into compressible vectors (those that are sufficiently close to a sparse vector) and incompressible vectors (all others).

In order to handle the compressible vectors, we introduce the concept of a *pattern matrix*, which is a dense deterministic matrix that maps all sparse vectors into dense vectors in a certain strong sense. At a high level, this idea is widely known as the *uncertainty principle* [14], which is attained for instance by the Discrete Fourier Transform and the Walsh-Hadamard matrix. However, our concept of a pattern matrix defines a strictly stronger property which is required for our guarantees and is not attained by these classical transforms. While the

existence of a pattern matrix is relatively straightforward (e.g., a dense random sign matrix), obtaining one with high probability using $O(n \log n)$ random bits requires a more elaborate construction. Equipped with such a matrix, we then randomize it by multiplying its rows and columns by independent random signs, and show that this perturbation is sufficient to handle all compressible vectors (see Section 2.1 for more details).

In order to handle the incompressible vectors, we return to the sparse perturbations, and show that $O(n)$ non-zero entries are sufficient in this case. However, a naïve implementation that places the entries uniformly at random hits an obstacle: it leads to a few *heavy rows*, i.e., those with disproportionately more non-zeros than the rest, which blows up the norm of the perturbation. To address this, we can simply zero out those heavy rows, but this introduces a non-trivial dependency between the entries, which is not handled by standard techniques. Here, our key contribution is developing a novel argument for analyzing sparse matrices with such dependencies (see Section 2.2 for more details).

1.2 Application: Solving linear systems in backward error As a motivating application for our main result, we consider solving a large system of linear equations, $Ax = b$, for an $n \times n$ matrix A and an n -dimensional vector b . Controlling the condition number κ of A is crucial both for the stability analysis of direct solvers such as Gaussian elimination [30], as well as for the convergence analysis of iterative solvers such as Conjugate Gradient [29]. By adding a perturbation to A , we can gain strong control over the condition number while introducing only a small backward error into the solution.

To illustrate this, we show how Theorem 1.2 can be used to reduce the random bit complexity of solving a linear system in backward error. We build on recent work by Dereziński, Nakatsukasa, and Rebrova [13], who show that when seeking a backward error approximation to a linear system, it is possible to overcome traditional computational barriers, such as the current cost of matrix multiplication, through the use of randomization. The following result combines our perturbation from Theorem 1.2 with their deterministic $O(\log(\kappa)/\varepsilon)$ -degree polynomial approximation to a linear system in backward error [13, Theorem 5.1]. The proof follows the same steps as their Theorem 5.3.

THEOREM 1.3. *Given $\varepsilon \in (0, 1)$, $A \in \mathbb{R}^{n \times n}$, and $b \in \mathbb{R}^n$, using $O(n^2 \log(n)/\varepsilon)$ time and $O(n \log n)$ random bits we can compute x such that, with probability 0.9,*

$$\tilde{A}x = b \quad \text{for some } \tilde{A} \text{ such that } \|\tilde{A} - A\| \leq \varepsilon \|A\|.$$

Remark 1.4. The guarantee is meaningful when $\varepsilon \gg 1/n^{\omega-2}$, where $\omega \approx 2.371$ is the current matrix multiplication exponent. Otherwise, an $O(n^\omega)$ time deterministic solver becomes more efficient.

Remark 1.5. While Theorem 1.3 is stated in exact arithmetic (Real RAM), the stability analysis of adding our structured perturbation to a matrix or applying it to a vector is straightforward. We leave the stability analysis of the polynomial approximation algorithm from [13] to future work.

This result matches the best known time complexity for solving an $n \times n$ linear system in backward error [13, Theorem 5.3], while reducing the amount of randomness required from the previous $O(n^{1+\theta})$ random bits (for any $\theta > 0$), which is attained by the sparse perturbation result of Tao and Vu [35]. Alternatively, one could use a discretized version of the sparse perturbation of Shah, Srivastava and Zeng [32], which requires $O(n \log^3 n)$ random bits but attains a sub-optimal $O(n^2 \log^2(n)/\varepsilon)$ time complexity. Our perturbation construction has the added advantage that it takes linear space, i.e., only as much as it takes to store the output vector x , which means that it does not introduce any additional memory overhead into the algorithm. We remark that it is crucial for the above application that the linear system may be non-symmetric (for symmetric systems a simple diagonal shift is sufficient), since backward error is not preserved under standard symmetrizations such as Hermitian dilation or the normal equations.

1.3 Further discussion of related work One of the settings where the analysis of the smallest singular value of zero mean sparse random matrices with i.i.d. entries perturbed by deterministic shifts has been previously considered is in the context of proving the circular law for random matrices [35, 27, 5]. Formally, these look at the smallest singular value of a (complex valued) matrix $A + R$, for a deterministic matrix A and a random matrix R with independent entries distributed as $x \cdot \delta$ where x is some complex valued random variable and $\delta \sim \text{Bernoulli}(\rho)$. To prove the circular law, it suffices to consider a diagonal shift $A = zI$ for $z \in \mathbb{C}$. However, as discussed earlier, Tao and Vu [35] prove a more general result with arbitrary shift which applies to our setting and requires $n^{1+\theta}$

random bits. Notably, [27, 5] attain better sparsity, but their approaches only work for a diagonal shift, and thus do not apply to our setting.

Shah, Srivastava and Zeng [32] study sparse perturbations in the context of fast matrix diagonalization algorithms. In this setting, they require bounding the so-called *eigenvector condition number* and the *minimum eigenvalue gap* of a perturbed matrix, in order to attain the so-called pseudospectral shattering [3]. A smallest singular value bound for a perturbed matrix, such as the one in our main result, appears as a central step in their argument. The problem of derandomizing pseudospectral shattering perturbations was recently posed by Bürgisser and Srivastava [1]. Our results can be viewed as a step in that direction.

Another approach towards using less randomness and less space in smoothed analysis is taken by Shah and Silwal [31], who consider *low-rank perturbations*. However, since a low-rank perturbation is inherently singular, their results are not directly comparable with our setting.

Randomized perturbations arise as part of several algorithmic approaches in computational linear algebra. For solving sparse linear systems, Peng and Vempala [23] use a random perturbation to improve the stability of a Krylov method. A similar strategy is used by Chen, Epperly, Meyer, Musco, Musco, and Rao [21, 8] for Krylov-based low-rank approximation. These works rely on a diagonal perturbation of a symmetric matrix to ensure a sufficient minimum eigenvalue gap via the result of Minami [22]. Finally, as discussed in Section 1.2, Dereziński, Nakatsukasa, and Rebrova [13] use random perturbations to avoid condition number dependence when solving dense non-symmetric linear systems in backward error.

2 Overview of the Techniques In this section, we provide a high-level overview of the construction and analysis for the well-conditioned oblivious perturbation that gives rise to Theorem 1.2.

Recall that our goal is to control the condition number of the perturbed matrix $M = A + \varepsilon R$ for a random perturbation that we will define gradually throughout the section. By assumption, the starting matrix A has $\|A\| \leq 1$ and we will ensure by construction that $\|R\| \leq 1$ with high probability, which means that $\|M\| \leq 2$. Thus to control the condition number, it suffices to look at the smallest singular value. The following is our main technical result.

THEOREM 2.1. *Let $0 < \varepsilon < 1$ and $0 < \delta < 1$. Then there exist $n \times n$ random matrices R_n and constants $c_{2.1.1}, c_{2.1.2}, c_{2.1.3}$ such that $\|R_n\| \leq 1$ with probability $1 - \delta$ and for any $n \times n$ deterministic matrix A with $\|A\| = 1$ and $n \geq c_{2.1.1}\delta^{-5}$, we have*

$$\mathbb{P}(s_n(A + \varepsilon R_n) \geq c_{2.1.2} \varepsilon^3 \delta^2 n^{-1}) \geq 1 - \delta$$

and R_n can be generated from $c_{2.1.3}\delta^{-2}n \log n$ random bits.

The standard approach [19, 28] to lower bounding $s_n(M) = \inf_{x \in \mathbb{S}^{n-1}} \|Mx\|_2$ involves separately considering the infimum over compressible vectors and incompressible vectors for some α and ν ,

$$\inf_{x \in \text{Comp}(\alpha, \nu)} \|Mx\|_2 \quad \text{and} \quad \inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2,$$

where

$$\text{Comp}(\alpha, \nu) = \{x \in \mathbb{S}^{n-1} \mid \|x - v\|_2 \leq \nu \text{ for some } v \in \mathbb{S}^{n-1} \text{ with } \text{supp}(v) \leq \alpha n\},$$

$$\text{Incomp}(\alpha, \nu) = \mathbb{S}^{n-1} \setminus \text{Comp}(\alpha, \nu).$$

In each case, we use a different property of R to obtain a high probability lower bound for the infimum, and to get these properties, we define $R := R_1 + R_2$ as a sum of two independent matrices, a dense matrix R_1 and a sparse matrix R_2 . For the infimum over compressible vectors x , we condition on the matrix R_2 and treat R_1 as the perturbation, obtaining a high probability lower bound for $\|(A_1 + \varepsilon R_1)x\|_2$ where $A_1 = A + \varepsilon R_2$. For incompressible vectors, we condition on R_1 and treat R_2 as the perturbation.

2.1 Construction of R_1 and invertibility control for compressible vectors The standard analysis of $\inf_{x \in \text{Comp}(\alpha, \nu)} \|Mx\|_2$ involves obtaining a lower bound on $\|Mx\|_2$ for a single appropriately sparse vector and then using a net argument to extend the lower bound to the whole set $\text{Comp}(\alpha, \nu)$. When the coordinates of the perturbation are independent, Mx has independent coordinates and tensorization of a small ball probability estimate for each coordinate provides the lower bound. However, using this approach for a *sparse* perturbation may not give us a sufficiently strong tail probability estimate for $\|Mx\|_2$ to be able to take the union bound. Indeed, this is what prevents us from relying only on the sparse component R_2 in our construction.

2.1.1 Dense perturbation via pattern matrices To address this challenge, we introduce a novel construction for R_1 . We define

$$R_1 = \frac{1}{\rho\sqrt{n}} \cdot D_1 V D_2$$

for $D_1 = \text{diag}(\eta_{1,1}, \dots, \eta_{1,n})$ and $D_2 = \text{diag}(\eta_{2,1}, \dots, \eta_{2,n})$ where $\eta_{1,1}, \dots, \eta_{1,n}$ and $\eta_{2,1}, \dots, \eta_{2,n}$ are i.i.d. Rademacher random variables, and V is any deterministic *pattern* matrix. Here, a pattern matrix is any $n \times n$ matrix V with ± 1 entries which has $\|V\| = O(\sqrt{n})$ and, for any sparse unit vector x within a specified sparsity threshold, both Vx and $V^T x$ have $\Omega(n)$ many coordinates bigger than some absolute constant (See Definition 4.1 for the full details). Thus, the pattern matrix transforms a sparse vector to one with many large coordinates.

After conditioning on R_2 and D_2 , if $i \in [n]$ is such that $(VD_2x)_i > \beta$ for some absolute constant $\beta > 0$, then

$$(Mx)_i = ((A + \varepsilon R_2)x)_i + (\varepsilon/\rho\sqrt{n})\eta_{1,i}(VD_2x)_i$$

is greater than $(\varepsilon\beta/\rho\sqrt{n})$ with probability more than half. Tensorizing this lower bound (Lemma 6.2) across all such i , of which there are $\Omega(n)$ many by the definition of a pattern matrix, we get a high probability lower bound for $\|Mx\|_2$ with the tail probability decaying exponentially in n . This is sufficient to take the union bound over an appropriate net of sparse vectors and complete the proof of the lower bound over all compressible vectors (D_2 does not play a role in this part of the argument, but rather, for the lower bound over incompressible vectors).

2.1.2 Constructing pattern matrices There exist well known linear transforms that map sparse vectors to ones with many nonzero coordinates, such as the Discrete Fourier Transform or the Walsh-Hadamard matrix. They satisfy the so-called *uncertainty principle* according to which the product of the support of a vector and the support of its transform is at least n [14]. Unfortunately, this uncertainty principle is tight, in the sense that one can find a $\sqrt{n}$ -sparse vector whose transform is also $\sqrt{n}$ -sparse (for completeness, we prove this in Proposition 5.1). This is an issue since we require $\Omega(n)$ large (and not just nonzero) coordinates for all cn -sparse vectors, for some small c . By considering randomized Hadamard matrices, one can get stronger uncertainty principle for a fixed sparse vector. To extend the norm lower bound from a fixed sparse vector x to all sparse vectors via a union bound, we need the bound to hold with probability $1 - \exp(-cn)$ whereas the stronger uncertainty principle (e.g., Lemma 3.3, Tropp, AADA 2011) only holds with probability $1 - n^{-O(1)}$. In fact, observe that if z is a bad vector for H as in Proposition 5.1, then Dz is also a bad vector for the randomized-Hadamard HD , so applying randomization does not exclude any of the bad vectors identified in Proposition 5.1.

A stronger additive version of the uncertainty principle, which has the sum (instead of product) of supports equal to n would work for our needs. This has been shown for the DFT over finite fields by Tao [34]. However, the proof of this result relies heavily on finite field arithmetic and does not readily extend to real numbers.

Therefore, we consider a randomized construction for V . While one can show that a matrix with independent Rademacher entries is a pattern matrix with high probability, we introduce a construction which requires much fewer random bits, and therefore can be stored efficiently.

This construction is inspired by one due to Lovett and Sodin [20], who in a different context construct a randomized embedding from ℓ_2 into ℓ_1 using the idea of considering an entry-wise (Hadamard) product of two random matrix models with ± 1 entries, with each model having its own desirable properties, to get a random matrix model which enjoys both sets of properties. We consider a model where $V = V_1 \odot V_2 \odot V_3^T$ for the Hadamard product $\odot$ of three random matrices V_1, V_2 and V_3 with Rademacher entries. Here, V_1 has $2\log(n)$ -wise independent entries, V_2 has independent rows with 4-wise independent coordinates and V_3 is an i.i.d. copy of V_2 , i.e., V_3^T has independent columns with 4-wise independent coordinates (See Figure 2.1). In Proposition 5.2, we prove that a random matrix generated with this model is a pattern matrix with high probability. We use each of the three components in the definition of V to prove the three properties we want of the pattern matrix by conditioning on the other two components:

1. After conditioning on V_2 and V_3 , we use the moment method to get a $O(\sqrt{n})$ bound for $\|V_1\|$, and consequently, for $\|V\|$. Without V_1 , using just the independence of the rows or columns, we would yield a poorer norm bound that we would have to scale R_1 by, which would result in a suboptimal smallest singular value estimate.
2. After conditioning on V_1 and V_3 , for any fixed unit vector x , Vx has independent coordinates. Each coordinate $(Vx)_i$ is a linear combination of the coordinates of x with 4-wise independent Rademacher

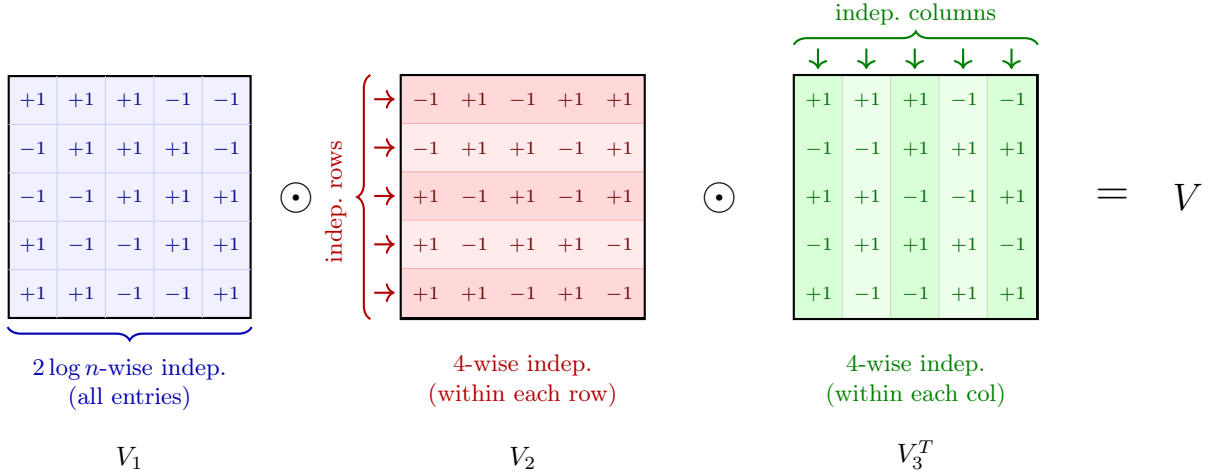


Figure 2.1: Illustration of the pattern matrix generation model for $n = 5$.

coefficients. This is enough to show that $|(Vx)_i|$ is greater than some constant with constant probability. The independence of rows allows us to use concentration arguments to claim that $|(Vx)_i|$ is greater than a constant for sufficiently many i , with a probability that decays exponentially in n , which is crucial to be able to take a union bound across all sparse vectors.

3. We use V_3 similarly as V_2 , in order to show that $V^T x$ has $\Omega(n)$ large coordinates.

Necessity of including a sparse random perturbation The perturbation R_1 is a random ± 1 matrix which is easy to generate. Unfortunately, this matrix alone cannot provide any meaningful smallest singular value guarantee. In fact, in Lemma 4.2, we show that for any $n \times n$ deterministic matrix $\hat{V}$ with $\|\hat{V}\| = 1$, one can exhibit an A and a sufficiently small ε for which $A + \varepsilon D_1 \hat{V} D_2$ is singular with high probability. To address this problem we introduce another random perturbation.

2.2 Construction of R_2 and invertibility control for incompressible vectors To obtain a lower bound on $\|Mx\|_2$ over incompressible vectors, we construct the sparse random matrix R_2 . Briefly speaking, in R_2 , we choose nonzero positions (where we place i.i.d. random signs) uniformly in each column, and then trim the rows. We start with a simple construction where the non-zeros are placed uniformly at random, and then we show how to modify this construction via row trimming to control the norm of the perturbation.

2.2.1 Simple construction with i.i.d. sparse entries To minimize the number of random bits needed, the most natural choice is to consider a random matrix with sparse Rademacher entries. We first construct a hashing matrix $\bar{H}$ where the columns of $\bar{H}$ are independent $\{0, 1\}$ random vectors containing K ones to fix the positions of our random signs, where K will be a constant that only depends on ε and δ , and does not depend on n (for simplicity, we ignore the dependence on ε and δ in this overview). More formally, for $i \in [n]$, let J_i be independent subsets of $[n] = \{1, \dots, n\}$ of size K , uniformly distributed among such subsets. Then, for $i, j \in [n]$, we set $\bar{h}_{j,i} = \mathbb{1}_{J_i}(j)$.

Next, we just put random signs on the nonzero positions determined by $\bar{H}$ with appropriate scaling. More formally, we define a sequence $\{X_{l,i} : i \in [n], l \in [K]\}$ of Kn i.i.d. Rademacher random variables, and enumerate the elements of each random set $J_i, i \in [n]$:

$$J_i = \{j(i, 1), \dots, j(i, K)\} \text{ where } j(i, 1) < j(i, 2) < \dots < j(i, K)$$

Then we can define the matrix $\bar{R}_2$ by

$$[\bar{R}_2]_{j(i,l),i} = \begin{cases} \frac{1}{L} \cdot X_{l,i}, & \text{for } l \in [K] \\ 0 & \text{otherwise} \end{cases}$$

where $L > K$ is a parameter to be chosen later.

The random matrix $\bar{R}_2$ indeed gives us desired control of the smallest singular value (nevertheless, $\bar{R}_2$ is not completely sufficient for our purpose, as explained later).

First of all, we observe that by conditioning on R_1 , the problem reduces to bounding $\|(A_1 + \varepsilon R_2)x\|_2$ over incompressible vectors where $A_1 = A + \varepsilon R_1$ can be considered deterministic because of conditioning. Next, we can use the standard invertibility by distance method as in [28, 4], which gives optimal control of $\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2$ (here optimal means as good as for a dense gaussian perturbation). This involves reducing the problem of finding the infimum over incompressible vectors to finding a lower bound for the distance between a column of M and the subspace spanned by the remaining columns of M . For concreteness, since the columns are identically distributed, we consider the distance between the last column and the subspace spanned by the first $n - 1$ columns. The techniques for analyzing this distance use the independence between columns to condition on the first $n - 1$ columns, which fixes the normal (i.e., orthogonal) vector to the subspace spanned by them, and then use the randomness of the last column to control the inner product between the normal vector and the last column, i.e., $\mathbb{P}(|\langle Y_n, Z \rangle| < t)$, where Y_n is the last column of $A_1 + \varepsilon \bar{R}_2$ and Z is a random normal vector of the subspace spanned by the first $n - 1$ columns. Although [28, 4] only considered the case without the shift A_1 , adding this shift will not change the argument because the small ball probability bound we use to control $\mathbb{P}(|\langle Y_n, Z \rangle| < t)$ is not affected by deterministic shifts. Note that Coupon Collector argument shows that if $\bar{R}_2$ were the whole perturbation, then with only $K = O(1)$ nonzeros per column it misses about $e^{-K}n$ rows, so many coordinates receive no randomness at all, and this is why a purely sparse $O(n)$ -nonzero perturbation cannot reliably fix every bad direction. However, using $R_1 + \bar{R}_2$ avoids this obstruction because R_1 already controls the sparse directions, and R_2 is used only when the relevant vector is spread over αn coordinates, where hitting a few useful coordinates is easy even though covering all n rows is impossible.

2.2.2 Controlling the norm and row trimming The natural construction $\bar{R}_2$ with i.i.d. sparse entries gives optimal control on the smallest singular value with $K = O(1)$ nonzero entries per column. However, in this case, the norm of R_2 grows with n since some rows may have more than $O(1)$ many entries with absolute value ≥ 1 (in other words, some of the rows are heavy). Since we want our perturbations to have $O(1)$ norm, we can scale by the norm of $\bar{R}_2$, but this makes the resulting smallest singular value bound suboptimal by a $\log n$ factor, and cannot give us the $O(n)$ bound on the condition number to match Gaussian perturbations.

We resolve this issue by an intuitive modification, which is to truncate the matrix $\bar{R}_2$ by simply removing the problematic heavy rows. Formally, we choose an $L > K$ such that $L = O(1)$ and consider rows with more than L nonzero entries, calling them *heavy* rows. We show that the number of heavy rows is not too large, and zero them out. After this step, all columns have at most K nonzero entries and all rows have at most L nonzero entries, all with absolute value 1. We can now scale the resulting matrix by $1/L$ to ensure a norm bound of 1 without degrading the smallest singular value estimate.

Formally, we set $\text{Row}_j(H) = \text{Row}_j(\bar{H})$ if the number of non-zero entries in $\text{Row}_j(\bar{H})$ does not exceed L and $\text{Row}_j(H) = 0$ otherwise. Note that the number of non-zero entries in each row and column of the resulting matrix H is at most L . Finally, we set $R_2 = H \odot \bar{R}_2$. In summary, the nonzero positions (where we place i.i.d. random signs) are uniformly chosen in each column in R_2 , and then it is the rows that are being trimmed (see Figure 2.2).

2.2.3 Modifying the invertibility by distance argument for row-trimmed matrix R_2 Unfortunately, these modifications break the previous conditioning argument since the entries of the first $n - 1$ columns depend on the last column. The main novelty of our argument for controlling

$$\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2$$

is a modification of the invertibility by distance argument to make it work here for our matrix R_2 which has dependent columns. In this direction, the first attempt is to condition on the positions of the nonzero entries in the last column (and not the random signs) in addition to the first $n - 1$ columns. This way, we separate out the random signs $X_{l,n}$ in the last column which are independent from the part that we condition on. More precisely, recalling that Y_n denotes the last column of M and Z is a unit vector orthogonal to the subspace spanned by the first $n - 1$ columns, we can write

$$\langle Y_n, Z \rangle = \langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \text{ will not be removed}\}}$$

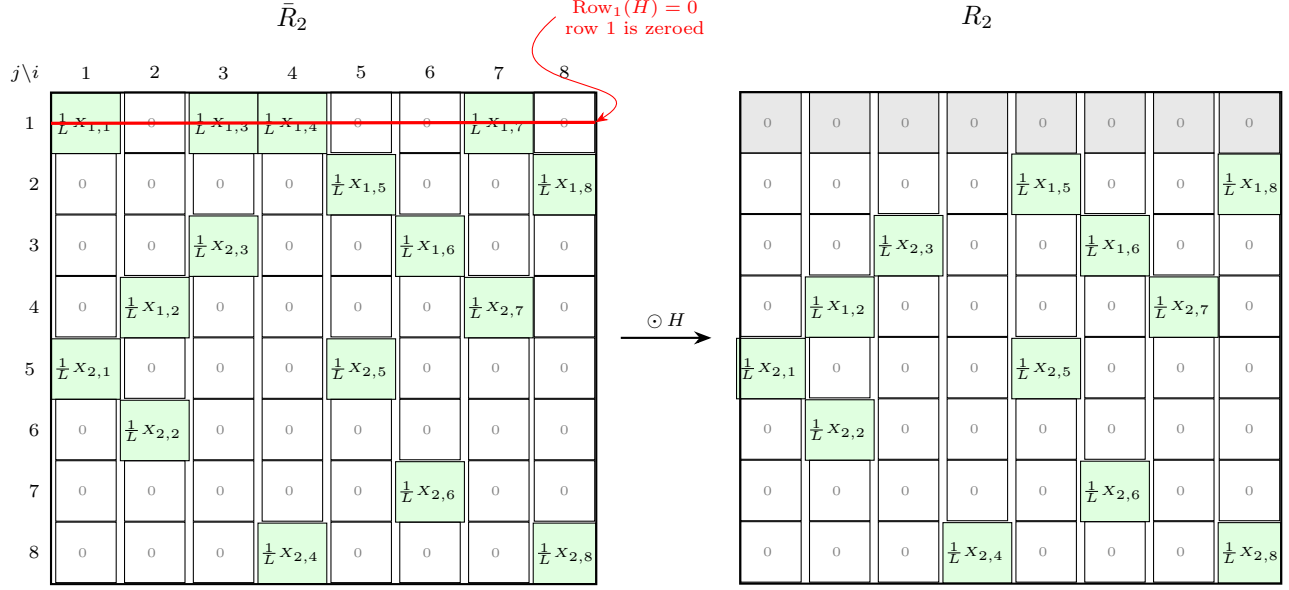


Figure 2.2: Illustration of $R_2 = H \odot \bar{R}_2$ for one realization. Since $\text{Row}_1(H) = 0$, the first row of $\bar{R}_2$ is zeroed in R_2 .

which, after conditioning, becomes a linear combination of independent random signs $X_{l,n}$ with coefficients $W_l = Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \text{ will not be removed}\}}$ (with a deterministic shift $\langle \text{Col}_n(A + \varepsilon R_1), Z \rangle$ which does not affect Levy concentration).

Nevertheless, separating out the random signs $X_{l,n}$ does not make the argument work directly, because to control the small ball probability, we need to have sufficiently many good coordinates (here good means that the corresponding coefficient is nonzero and with magnitude of order $1/\sqrt{n}$) in the coefficients $(W_1, \dots, W_K)$, which requires additional effort to show.

First, we show that the random vector Z is incompressible with high probability (this is where the other random diagonal matrix D_2 from R_1 is needed). This means Z has many coordinates of order $1/\sqrt{n}$, called “spread” coordinates (Lemma 7.3), denoted by Θ_0 . Then, the set of the good coordinates is the intersection of Θ_0 , the nonzero positions selected by J_n , and the row indices that will not be removed. The difficulty is that these three sets are dependent. To unravel this dependence, we define the index set of heavy rows of $\bar{H}$ up to column $n - 1$ and the index set of the light rows up to column $n - 1$ as

$$I_{\text{heavy},n} = \left\{ i \in [n] : \sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L \right\} \quad \text{and} \quad I_{\text{light},n} = \left\{ i \in [n] : \sum_{j=1}^{n-1} \bar{h}_{i,j} < L \right\},$$

and observe that a nonzero position selected by J_n will eventually reset the whole row to zero if and only if it is in a row that is already heavy enough in the $1, \dots, n - 1$ columns. Therefore, the set of good coordinates must contain $\Theta_0 \cap J_n \cap I_{\text{light},n}$. We observe that Z , and hence Θ_0 (the spread coordinates of Z), depends only on the first $n - 1$ columns of R_2 , which, in turn only depend on the first $n - 1$ columns of $\bar{R}_2$ and $J_n \cap I_{\text{heavy},n}$, because the nonzero entries in $J_n \cap I_{\text{light},n}$ in the last column (i.e., what J_n selects in $I_{\text{light},n}$) will not cause a row deletion in any case. This allows us to improve the naive conditioning argument described above. We condition on the first $n - 1$ columns of $\bar{R}_2$ and only on the set $J_n \cap I_{\text{heavy},n}$ instead of the whole set J_n . This conditioning fixes the set Θ_0 . Moreover, by choosing an appropriately large L , we can ensure that the set $\Theta_0 \cap I_{\text{light},n}$ is large with high probability. After that, we can take advantage of the extra randomness in the choice of the remaining part of J_n to guarantee that, with high probability, sufficiently many of elements of the set $J_n \cap I_{\text{light},n}$ fall in $\Theta_0 \cap I_{\text{light},n}$.

More formally, we can show that, conditioning on the first $n - 1$ columns of $\bar{R}_2$ and $J_n \cap I_{\text{heavy},n}$, the set $J_n \cap I_{\text{light},n}$ is a random uniform subset of $I_{\text{light},n}$ with cardinality $K - |J_n \cap I_{\text{heavy},n}|$ (see Lemma 7.5 and Corollary 7.6 for details). And therefore, we can show that, with high probability, we have sufficiently many good

coordinates in $(W_1, \dots, W_K)$ to control the small ball probability of

$$\langle Y_n, Z \rangle = \langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} W_l$$

and therefore prove invertibility for incompressible vectors (see Theorem 7.8 for details).

2.3 Outline of the paper Section 3 explains the notation used in the paper and provides details of some basic mathematical estimates we use in the rest of the paper. Section 4 provides the full details of our perturbation model. Section 5 establishes that our model for pattern matrices achieves the required properties with high probability. Section 6 and Section 7 prove the singular value lower bound over compressible and incompressible vectors, respectively. Section 8 combines these lower bounds to prove Theorem 2.1.

3 Preliminaries

3.1 Notation. The following notation and terminology will be used in the paper. The notation $[n]$ is used for the set $\{1, 2, \dots, n\}$. Also, for two integers a and b with $a \leq b$, we use the notation $[a : b]$ for the set $\{k \in \mathbb{Z} : a \leq k \leq b\}$. For $x \in \mathbb{R}$, we use the notation $\lfloor x \rfloor$ to denote the greatest integer less than or equal to x and $\lceil x \rceil$ to denote the least integer greater than or equal to x . In $\mathbb{R}^n$ (or $\mathbb{R}^m$ or $\mathbb{R}^d$), the l th coordinate vector is denoted by e_l .

We use the notation $\mathbb{P}$ for the standard probability measure, and the notation $\mathbb{E}$ for the expectation with respect to this standard probability measure. The covariance of two random variables X and Y is denoted by $\text{Cov}(X, Y)$. The standard L_q norm of a random variable ξ is denoted by $\|\xi\|_{L_q(\mathbb{P})}$, for $1 \leq q \leq \infty$. When there is no possibility of confusion, we will use the short notation $\|\xi\|_q$ for the L_q norm of ξ . For convenience of iterated integration, we introduce the following notations. Let X, Y be two random elements taking values in measurable spaces $\mathcal{S}_1$ and $\mathcal{S}_2$. Let $\phi : \mathcal{S}_1 \times \mathcal{S}_2 \rightarrow \mathbb{R}$ be a measurable function. Assume that $\phi(X, Y)$ is integrable. Define the function $\psi_2 : \mathcal{S}_2 \rightarrow \mathbb{R}$ such that $\psi_2(y) = \mathbb{E} \phi(X, y)$. Then the notation $\mathbb{E}_X(\phi(X, Y))$ stands for $\psi_2(Y)$. Similarly, we define the function $\psi_1 : \mathcal{S}_1 \rightarrow \mathbb{R}$ such that $\psi_1(x) = \mathbb{E} \phi(x, Y)$ and the notation $\mathbb{E}_Y(\phi(X, Y))$ stands for the random variable $\psi_1(X)$.

All matrices considered in this paper are real valued and the space of $m \times n$ matrices with real valued entries is denoted by $M_{m \times n}(\mathbb{R})$. Also, for a matrix $X \in M_{d \times d}(\mathbb{R})$, the notation $\text{Tr}(X)$ denotes the trace of the matrix X , and $\text{tr}(X) = \frac{1}{d} \text{Tr}(X)$ denotes the normalized trace. The $d \times d$ identity matrix is denoted by I_d . We write the operator norm of a matrix X as $\|X\|$, and it is also denoted by $\|X\|_{op}$ in some places where other norms appear for clarity. The spectrum of a matrix X is denoted by $\text{spec}(X)$. For matrices A and B with same dimensions, $A \odot B$ stands for the entry-wise (Hadamard) product of A and B , i.e., $(A \odot B)_{ij} = A_{ij} \cdot B_{ij}$.

Throughout the paper, the symbols $c_1, c_2, \dots$, and $Const, Const', \dots$ denote absolute constants.

3.2 Decomposition of the sphere Our method to bound the smallest singular value of the perturbed matrix uses the standard approach that separately control the invertibility for compressible vectors and for incompressible vectors. So we introduce these formal definition.

DEFINITION 3.1. Let $\alpha, \nu \in (0, 1)$. A vector $x \in \mathbb{R}^n$ is called sparse if $|\text{supp}(x)| \leq \alpha n$, and the set of sparse vectors is denoted by $\text{Sparse}(\alpha)$. A vector $x \in \mathbb{S}^{n-1}$ is called compressible if x is within Euclidean distance ν from the set of all sparse vectors. A vector $x \in \mathbb{S}^{n-1}$ is called incompressible if it is not compressible. The sets compressible and incompressible vectors will be denoted by

$$\begin{aligned} \text{Comp}(\alpha, \nu) &= \{x \in \mathbb{S}^{n-1} \mid \|x - v\|_2 \leq \nu \text{ for some } v \in \text{Sparse}(\alpha)\} \\ \text{Incomp}(\alpha, \nu) &= \mathbb{S}^{n-1} \setminus \text{Comp}(\alpha, \nu) \end{aligned}$$

COROLLARY 3.2. Let $0 < \alpha \leq 1$ and $\nu > 0$. If $x \in \text{Comp}(\alpha, \nu)$, then there exists $S \subset [n]$ such that

$$|S| \leq \alpha n \quad \text{and} \quad \|x_{[n] \setminus S}\|_2 \leq \nu.$$

Proof. Since $x \in \text{Comp}(\alpha, \nu)$, by definition, there exists $v \in \text{Sparse}(\alpha)$ such that

$$\|x - v\|_2 \leq \nu.$$

Set $S = \text{supp}(v)$. Then $|S| \leq \alpha n$. Since $v_{[n] \setminus S} = 0$, we have

$$x_{[n] \setminus S} = (x - v)_{[n] \setminus S}.$$

Therefore, we conclude

$$\|x_{[n] \setminus S}\|_2 = \|(x - v)_{[n] \setminus S}\|_2 \leq \|x - v\|_2 \leq \nu.$$

□

LEMMA 3.3 (Counting sparse supports). *For every $n \geq 1$ and every $0 < \alpha \leq 1/2$,*

$$|\{S \subset [n] : |S| \leq \alpha n\}| \leq \exp\left(\alpha n \log \frac{e}{\alpha}\right).$$

Proof. The results follow from direct counting. □

3.3 Covering numbers We will need the following facts for covering numbers. For a set $K \subset \mathbb{R}^n$ and $\varepsilon > 0$, $\mathcal{N}(K, \varepsilon)$ denotes the smallest cardinality of an ε -net of K in the Euclidean metric, and $\mathcal{P}(K, \varepsilon)$ denotes the largest cardinality of an ε -separated subset of K in the Euclidean metric. The notation $|\cdot|$ denotes Lebesgue volume.

PROPOSITION 3.4 (Proposition 4.2.12 in [37]). *Let K be a subset of $\mathbb{R}^n$ and $\varepsilon > 0$. Then*

$$\frac{|K|}{|\varepsilon B_2^n|} \leq \mathcal{N}(K, \varepsilon) \leq \mathcal{P}(K, \varepsilon) \leq \frac{|K + (\varepsilon/2)B_2^n|}{|(\varepsilon/2)B_2^n|}.$$

Here $|\cdot|$ denotes the volume in $\mathbb{R}^n$, B_2^n denotes the unit Euclidean ball in $\mathbb{R}^n$, so εB_2^n is a Euclidean ball with radius ε .

3.4 Lévy concentration As is explained in the outline, the control of invertibility for incompressible vectors needs to use the control for Lévy concentration. So we introduce the formal definition and standard results below.

DEFINITION 3.5 (Lévy concentration function). *For a real-valued random variable Z and $r \geq 0$, its concentration function is*

$$(3.1) \quad \mathcal{L}(Z, r) := \sup_{u \in \mathbb{R}} \mathbb{P}\{|Z - u| \leq r\}.$$

LEMMA 3.6 (Scaling of Lévy concentration). *Let Z be real-valued and let $a \neq 0$. Then for all $r \geq 0$,*

$$(3.2) \quad \mathcal{L}(aZ, r) = \mathcal{L}\left(Z, \frac{r}{|a|}\right).$$

In particular, if $a > 0$ then $\mathcal{L}(Z/a, r) = \mathcal{L}(Z, ar)$.

Proof. The result follows from straightforward calculation.

LEMMA 3.7 (Rogozin's theorem [26, Theorem 1]). *Let $\xi_1, \dots, \xi_n$ be independent real-valued random variables, let $r > 0$, and let $r_1, \dots, r_n \in (0, r]$. Assume that*

$$\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i)) r_i^2 > 0.$$

Then

$$(3.3) \quad \mathcal{L}\left(\sum_{i=1}^n \xi_i, r\right) \leq \frac{c_{3.7} r}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i)) r_i^2}},$$

where $c_{3.7} > 0$ is a universal constant.

3.5 Measure concentration We will need the following standard bounded difference inequality from [6].

LEMMA 3.8 (Bounded differences inequality). *Let $X_1, \dots, X_m$ be independent random variables taking values in a product space $\mathcal{X}_1 \times \dots \times \mathcal{X}_m$. Let $f : \mathcal{X}_1 \times \dots \times \mathcal{X}_m \rightarrow \mathbb{R}$ and assume that there exist constants $b_1, \dots, b_m \geq 0$ such that for every $k \in [m]$ and every two points x, x' that differ only in the k th coordinate,*

$$|f(x) - f(x')| \leq b_k.$$

Define

$$v := \frac{1}{4} \sum_{k=1}^m b_k^2.$$

Then for every $t > 0$, if $Z := f(X_1, \dots, X_m)$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} \leq \exp\left(-\frac{t^2}{2v}\right), \quad \mathbb{P}\{Z - \mathbb{E}Z < -t\} \leq \exp\left(-\frac{t^2}{2v}\right).$$

3.6 Random sampling without replacement Random sampling without replacement plays an important role in our construction of R_2 and in our proof for the invertibility for incompressible vectors. So we state the formal definitions (see e.g., [2]) and standard results below.

DEFINITION 3.9 (Ordered random sampling without replacement). *Let $k \leq n$ be integers. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\mathcal{Z}$ be the set of injections from $[k]$ to $[n]$. Let $T : \Omega \rightarrow \mathcal{Z}$ be a random variable such that for any $z \in \mathcal{Z}$, we have*

$$\mathbb{P}(T = z) = \frac{1}{n \cdot (n-1) \cdots (n-k+1)}$$

Then we say T is a random sampling of an ordered k -element sequence uniformly from $[n]$ without replacement. Also, if $\mathcal{S} = (s_i)_{i \in [n]} = (s_1, \dots, s_n)$ is a finite ordered sequence, and $S(\omega) = s \circ T(\omega)$, i.e., $S(l)(\omega) = s_{T(l)(\omega)}$ for $l = 1, 2, \dots, k$. Then $S = (S(1), \dots, S(k))$ is called a random sampling of an ordered k -element sequence uniformly from population $\mathcal{S}$ without replacement.

We will use the following results from [17] to get concentration results for sum of random sampling without replacement.

LEMMA 3.10 (Theorem 4 in [17]). *Let $\mathcal{S} = (s_1, \dots, s_n)$ be a finite population of n real points. Let $(X(1), \dots, X(k))$ denote a random sampling of an ordered k -element sequence uniformly from population $\mathcal{S}$ without replacement, and let $(Y(1), \dots, Y(k))$ denote a random sample with replacement from $\mathcal{S}$ (i.i.d. random samples from $\mathcal{S}$). If $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous and convex, then*

$$\mathbb{E} f\left(\sum_{i=1}^k X(i)\right) \leq \mathbb{E} f\left(\sum_{i=1}^k Y(i)\right).$$

Combining Lemma 3.10 with standard Hoeffding inequality for i.i.d. sampling (i.e., with replacement), we obtain the following Hoeffding's inequality for random sampling without replacement.

PROPOSITION 3.11 (Hoeffding's inequality for random sampling without replacement). *Let $\mathcal{S} = (s_1, \dots, s_n)$ be a finite population of n real points and let $(S(1), \dots, S(k))$ be a random sampling of an ordered k -element sequence uniformly from population $\mathcal{S}$ without replacement. Let*

$$a = \min_{1 \leq i \leq n} s_i, \quad b = \max_{1 \leq i \leq n} s_i, \quad \mu = \frac{1}{n} \sum_{i=1}^n s_i.$$

Then for all $t > 0$, we have

$$(3.4) \quad \mathbb{P}\left(\frac{1}{k} \sum_{i=1}^k S(i) - \mu \geq t\right) \leq \exp\left(-\frac{2kt^2}{(b-a)^2}\right).$$

Remark 3.12 (Lower tail). Applying Proposition 3.11 to the population $(-s_1, \dots, -s_n)$ we can also obtain the lower tail bound

$$(3.5) \quad \mathbb{P}\left(\frac{1}{k} \sum_{i=1}^k S(i) - \mu \leq -t\right) \leq \exp\left(-\frac{2kt^2}{(b-a)^2}\right), \quad t > 0.$$

DEFINITION 3.13 (Unordered random sampling without replacement). *Let $k \leq n$ be integers. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\mathcal{Z}$ be the set of injections from $[k]$ to $[n]$. Let $\mathcal{S} = (s_i)_{i \in [n]} = (s_1, \dots, s_n)$ is a finite ordered sequence. And assume that $s_1, \dots, s_n$ are pairwise distinct. Let $\mathcal{J} = \text{Range}(\mathcal{S}) = \{s_1, \dots, s_n\}$. Let $T : \Omega \rightarrow \mathcal{Z}$ be a random sampling of an ordered k -element sequence uniformly from $[n]$ without replacement. Define*

$$J(\omega) = \{s_{T(\omega)(1)}, \dots, s_{T(\omega)(k)}\} = \text{Range}(s \circ T(\omega)).$$

Then J is called a random sampling of an unordered k -element subset uniformly from $\mathcal{J}$ without replacement.

Remark 3.14. Let $\mathcal{S} = (s_i)_{i \in [n]} = (s_1, \dots, s_n)$ is a finite ordered sequence. And assume that $s_1, \dots, s_n$ are pairwise distinct. Let $\mathcal{J} = \text{Range}(\mathcal{S}) = \{s_1, \dots, s_n\}$. Then the followings hold:

(i) If J is a random sampling of an unordered k -element subset uniformly from $\mathcal{J}$ without replacement, then for every $B \subset \mathcal{J}$ with $\text{card}(B) = k$, we have

$$\mathbb{P}(J = B) = \frac{1}{\binom{n}{k}},$$

(ii) If J is a random set such that for every $B \subset \mathcal{J}$ with $\text{card}(B) = k$, we have

$$\mathbb{P}(J = B) = \frac{1}{\binom{n}{k}},$$

then (in a expanded probability space if necessary) J is a random sampling of an unordered k -element subset uniformly from $\mathcal{J}$ without replacement.

LEMMA 3.15 (Conditional law of Γ_2 given Γ_1). *Fix $n \in \mathbb{N}$ and $K \in \{0, 1, \dots, n\}$. Let J be a random sampling of an unordered K -element subset uniformly from $[n]$ without replacement. and let $I \subset [n]$ be deterministic. Define*

$$\Gamma_1 := J \cap I, \quad \Gamma_2 := J \cap ([n] \setminus I).$$

Let $\mu(\theta_1, B)$ be a regular conditional probability of Γ_2 given Γ_1 , i.e., μ is a probability kernel such that for any measurable set B (in the space where Γ_2 takes values), we have $\mu(\Gamma_1, B) = \mathbb{P}(\Gamma_2 \in B | \Gamma_1)$ almost surely. Then for every fixed $\theta_1 \subset I$ with $|\theta_1| \leq K$:

1. *If $K - |\theta_1| > |[n] \setminus I|$, then $\mathbb{P}(\Gamma_1 = \theta_1) = 0$.*
2. *If $K - |\theta_1| \leq |[n] \setminus I|$, then on the event $\{\Gamma_1 = \theta_1\}$, for every $\theta_2 \subset [n] \setminus I$,*

$$\mu(\theta_1, \{\theta_2\}) = \begin{cases} \left(\frac{|[n] \setminus I|}{K - |\theta_1|}\right)^{-1}, & \text{if } |\theta_2| = K - |\theta_1|, \\ 0, & \text{otherwise.} \end{cases}$$

In other words, conditional on Γ_1 , the random set Γ_2 is a random sampling of an unordered $(K - |\Gamma_1|)$ -element subset uniformly from $[n] \setminus I$ without replacement, or more precisely, any random element with law of distribution $\mu(\theta_1, \cdot)$ is a random sampling of an unordered $(K - |\theta_1|)$ -element subset uniformly from $[n] \setminus I$ without replacement (in an expanded probability space if necessary).

Proof. Fix $\theta_1 \subset I$ with $|\theta_1| \leq K$ and put $s := K - |\theta_1|$. If $s > |[n] \setminus I|$ then there is no K -subset J with $J \cap I = \theta_1$, hence $\mathbb{P}(\Gamma_1 = \theta_1) = 0$.

Assume henceforth that $s \leq |[n] \setminus I|$. Let $\theta_2 \subset [n] \setminus I$. If $|\theta_2| \neq s$, then the event $\{\Gamma_1 = \theta_1, \Gamma_2 = \theta_2\}$ is empty, so $\mathbb{P}(\Gamma_2 = \theta_2, \Gamma_1 = \theta_1) = 0$ and therefore $\mu(\theta_1, \{\theta_2\}) = 0$ on $\{\Gamma_1 = \theta_1\}$.

If $|\theta_2| = s$, then $\Gamma_1 = \theta_1$ and $\Gamma_2 = \theta_2$ is equivalent to $J = \theta_1 \cup \theta_2$ (disjoint union). Since J is uniform on $\binom{[n]}{K}$, we have

$$\mathbb{P}(\Gamma_2 = \theta_2, \Gamma_1 = \theta_1) = \mathbb{P}(J = \theta_1 \cup \theta_2) = \binom{n}{K}^{-1}.$$

In addition, we have

$$\mathbb{P}(\Gamma_1 = \theta_1) = \sum_{\substack{\theta'_2 \subseteq [n] \setminus I \\ |\theta'_2| = s}} \mathbb{P}(J = \theta_1 \cup \theta'_2) = \binom{|[n] \setminus I|}{s} \binom{n}{K}^{-1}.$$

Therefore,

$$\mathbb{P}(\Gamma_2 = \theta_2 \mid \Gamma_1 = \theta_1) = \frac{\mathbb{P}(\Gamma_2 = \theta_2, \Gamma_1 = \theta_1)}{\mathbb{P}(\Gamma_1 = \theta_1)} = \binom{|[n] \setminus I|}{s}^{-1}.$$

Since Γ_1 takes values in a finite set, a version of μ can be chosen so that $\mu(\theta_1, \{\theta_2\}) = \mathbb{P}(\Gamma_2 = \theta_2 \mid \Gamma_1 = \theta_1)$ whenever $\mathbb{P}(\Gamma_1 = \theta_1) > 0$. $\square$

4 Oblivious Perturbation Model In this section, we provide the full details of our oblivious perturbation model. Recall that the perturbation R is defined as a sum of two independent random matrices a dense matrix R_1 and a sparse matrix R_2 .

4.1 The dense perturbation R_1 As mentioned in Section 2.1,

$$(4.1) \quad R_1 = \frac{1}{\rho\sqrt{n}} \cdot D_1 V D_2$$

for $D_1 = \text{diag}(\eta_{1,1}, \dots, \eta_{1,n})$ and $D_2 = \text{diag}(\eta_{2,1}, \dots, \eta_{2,n})$ where $\eta_{1,1}, \dots, \eta_{1,n}$ and $\eta_{2,1}, \dots, \eta_{2,n}$ are i.i.d. Rademacher random variables, and V is any deterministic *pattern* matrix with parameters α, β, γ , and ρ as defined in the following Definition 4.1,

DEFINITION 4.1. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho > 0$. An $n \times n$ matrix V with ± 1 entries will be called an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$ if all of the following hold.

1. $\|V\| \leq \rho\sqrt{n}$,

2. For any $\eta_1, \dots, \eta_n \in \{-1, 1\}$, and for any αn -sparse vector $x \in \mathbb{S}^{n-1}$, we have

$$|\{j \in [n] : |e_j^T V \cdot \text{diag}(\eta_1, \dots, \eta_n) x| \geq \beta\}| \geq \gamma n$$

and

$$|\{j \in [n] : |e_j^T V^T \cdot \text{diag}(\eta_1, \dots, \eta_n) x| \geq \beta\}| \geq \gamma n$$

For discussion about the existence of pattern matrices, see Section 5.

4.2 Necessity of a second perturbation Before describing the construction of the second matrix R_2 , we note via the following lemma that setting $R = R_1$ alone does not suffice to obtain a condition number bound.

LEMMA 4.2. Let $\hat{V}$ be an $n \times n$ deterministic matrix with $\|\hat{V}\| = 1$. Let D_1 and D_2 be as in (4.1). Then, for $\varepsilon = 1/\sqrt{n}$, there exists a deterministic $n \times n$ matrix A with $\|A\| \leq 1$ such that

$$\mathbb{P}\left(s_n(A + \varepsilon D_1 \hat{V} D_2) = 0\right) \geq \frac{1}{2}$$

Proof. Write

$$D_1 = \text{diag}(\eta_{1,1}, \dots, \eta_{1,n}) \quad \text{and} \quad D_2 = \text{diag}(\eta_{2,1}, \dots, \eta_{2,n}),$$

where the two families of signs are independent.

Choose $u \in \mathbb{S}^{n-1}$ orthogonal to $\text{Col}_2(\hat{V}), \dots, \text{Col}_n(\hat{V})$, and define the random vector

$$v_{\eta_1} = D_1 u.$$

We observe that $v_{\eta_1} \in \mathbb{S}^{n-1}$, and, for every $k \geq 2$, we have

$$\begin{aligned} \langle \text{Col}_k(D_1 \hat{V} D_2), v_{\eta_1} \rangle &= \eta_{2,k} \langle D_1 \text{Col}_k(\hat{V}), D_1 u \rangle \\ &= \eta_{2,k} \langle \text{Col}_k(\hat{V}), u \rangle = 0. \end{aligned}$$

For the first column, we similarly have

$$\langle \text{Col}_1(D_1 \hat{V} D_2), v_{\eta_1} \rangle = \eta_{2,1} \langle \text{Col}_1(\hat{V}), u \rangle.$$

Therefore, we obtain

$$v_{\eta_1}^T D_1 \hat{V} D_2 = (\eta_{2,1} \langle \text{Col}_1(\hat{V}), u \rangle, 0, \dots, 0).$$

Choose $j \in [n]$ such that $|u_j| \geq 1/\sqrt{n}$, and define

$$A = \frac{\langle \text{Col}_1(\hat{V}), u \rangle}{\sqrt{n} u_j} e_j e_1^T.$$

Then A is deterministic and

$$\|A\| = \frac{|\langle \text{Col}_1(\hat{V}), u \rangle|}{\sqrt{n} |u_j|} \leq \|\text{Col}_1(\hat{V})\| \leq \|\hat{V}\| = 1.$$

Furthermore, we observe

$$v_{\eta_1}^T A = \left(\frac{\eta_{1,j}}{\sqrt{n}} \langle \text{Col}_1(\hat{V}), u \rangle, 0, \dots, 0 \right).$$

Therefore, for $\varepsilon = 1/\sqrt{n}$, we have

$$v_{\eta_1}^T (A + \varepsilon D_1 \hat{V} D_2) = \frac{\langle \text{Col}_1(\hat{V}), u \rangle}{\sqrt{n}} (\eta_{1,j} + \eta_{2,1}, 0, \dots, 0).$$

On the event

$$\eta_{1,j} + \eta_{2,1} = 0,$$

the nonzero vector v_{η_1} is a left null vector of $A + \varepsilon D_1 \hat{V} D_2$, so this matrix is singular. Since $\eta_{1,j}$ and $\eta_{2,1}$ are independent Rademacher random variables, we conclude that

$$\begin{aligned} \mathbb{P}(s_n(A + \varepsilon D_1 \hat{V} D_2) = 0) &\geq \mathbb{P}(\eta_{1,j} + \eta_{2,1} = 0) \\ &= \frac{1}{2}. \end{aligned} \quad \square$$

4.3 The sparse perturbation R_2 . We now explain the construction of R_2 once again with additional figures and conclude with a full definition of our perturbation model R (Definition 4.3) for completeness. R_2 is a sparse random matrix defined with parameters K and L with $K < L$ by first specifying the positions of the nonzero entries by a hashing matrix H and populating these positions by independent scaled Rademacher random variables. Before constructing H , we first construct an auxillary hashing matrix $\bar{H}$. The columns of the matrix $\bar{H}$ are independent $\{0, 1\}$ random vectors containing K ones. More formally, for $i \in [n]$, let J_i be independent subsets of $[n]$ of size K uniformly distributed among such subsets. Then, for $i, j \in [n]$, we set $\bar{h}_{j,i} = \mathbb{1}_{J_i}(j)$ (the matrix $\bar{H}$ is illustrated in Figure 4.1).

Next, define a sequence $\{X_{l,i} : i \in [n], l \in [K]\}$ of Kn i.i.d. Rademacher random variables, and enumerate the elements of each random set $J_i, i \in [n]$:

$$J_i = \{j(i, 1), \dots, j(i, K)\} \text{ where } j(i, 1) < j(i, 2) < \dots < j(i, K)$$

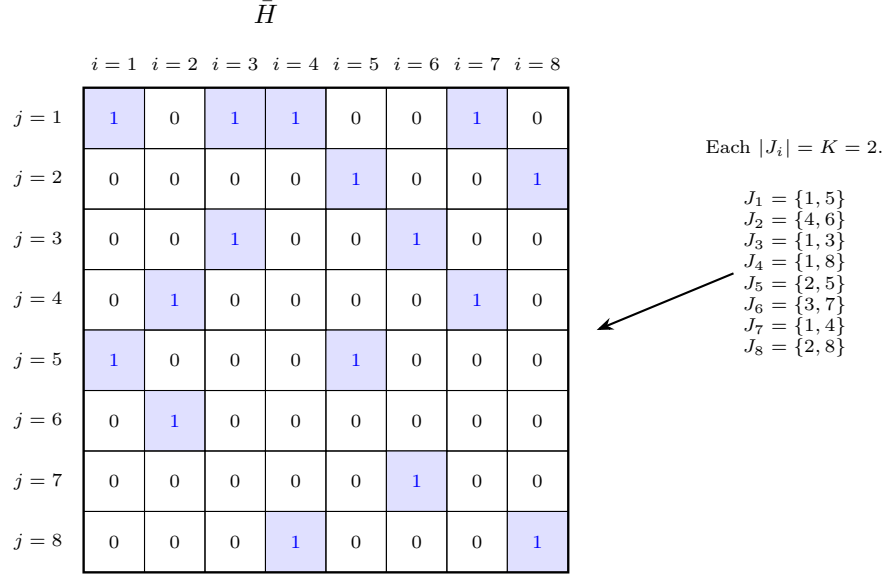


Figure 4.1: Illustration of $\bar{H}$ with $n = 8$, $K = 2$, $L = 3$.
Each column i selects an independent uniform subset $J_i \subset [n]$ of size K and sets $\bar{h}_{j,i} = \mathbf{1}_{J_i}(j)$.

Define the matrix $\bar{R}_2$ by

$$[\bar{R}_2]_{j(i,\ell),i} = \begin{cases} \frac{1}{L} \cdot X_{\ell,i}, & \text{for } \ell \in [K] \\ 0 & \text{otherwise} \end{cases}$$

In other words, we use $\bar{H}$ as a hashing matrix to create $\bar{R}_2$, which is illustrated in Figure 4.2. As mentioned

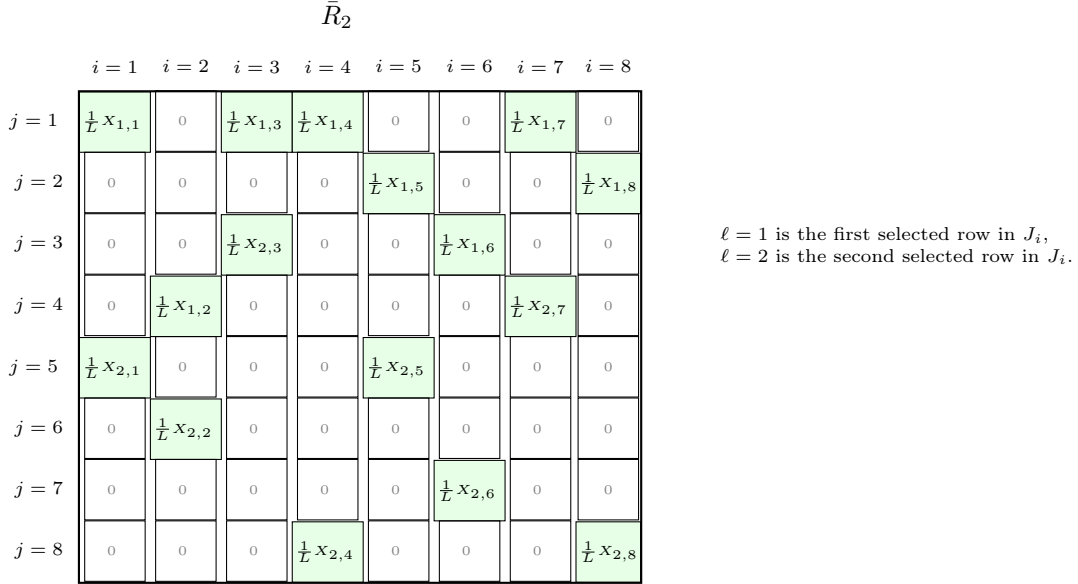


Figure 4.2: Illustration of $\bar{R}_2$ with $n = 8$, $K = 2$, $L = 3$.
Here, we enumerate $J_i = \{j(i, 1) < \dots < j(i, K)\}$ and set $\bar{r}_{j(i,\ell),i} = \frac{1}{L} X_{\ell,i}$ for $\ell \in [K]$, else 0.

before in Section 2, this construction $\bar{R}_2$ is not sufficient for our purpose because $\|\bar{R}_2\|$ may be large.

To reduce the norm of the random perturbation, we remove the heavy rows of $\bar{H}$ and $\bar{R}_2$ to construct the matrices H and R_2 . Formally, we set $\text{Row}_j(H) = \text{Row}_j(\bar{H})$ if the number of non-zero entries in $\text{Row}_j(\bar{H})$ does not exceed L and $\text{Row}_j(H) = 0$ otherwise. Note that the number of non-zero entries in each row and column of the matrix H is at most L . The process of transforming $\bar{H}$ to H is illustrated in Figure 4.3.

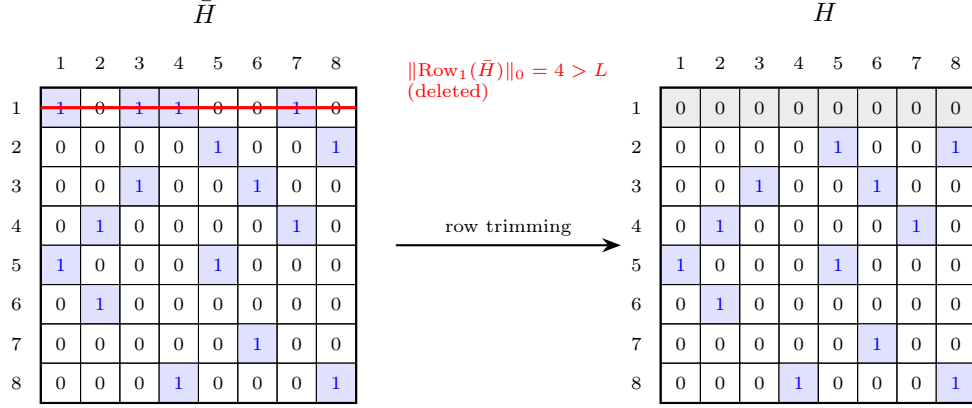


Figure 4.3: Illustration of constructing H from $\bar{H}$ with $n = 8$, $K = 2$, $L = 3$. Here, we keep row j if $\|\text{Row}_j(\bar{H})\|_0 \leq L$, else set it to 0.

Finally, set $R_2 = H \odot \bar{R}_2$ where $\odot$ stands for the entry-wise (Hadamard) product, which was illustrated earlier in Figure 2.2 in Section 2. Formally, we have the following definition.

DEFINITION 4.3. Let V be an n -dimensional deterministic pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let $\eta_{1,1}, \dots, \eta_{1,n}$ and $\eta_{2,1}, \dots, \eta_{2,n}$ be i.i.d. Rademacher random variables. Let

$$R_1 = \frac{1}{\rho\sqrt{n}} \cdot D_1 V D_2$$

where $D_1 = \text{diag}(\eta_{1,1}, \dots, \eta_{1,n})$ and $D_2 = \text{diag}(\eta_{2,1}, \dots, \eta_{2,n})$. Let K and L be positive integers with $1 \leq K \leq n$ and $L > K$. For $i \in [n]$, let J_i be independent subsets of $[n]$ of size K uniformly distributed among such subsets. Let $\bar{H}$ be the random matrix with entries $\bar{h}_{j,i}$ such that, for $i, j \in [n]$, we have

$$\bar{h}_{j,i} = \mathbb{1}_{J_i}(j)$$

We define the matrix H by setting $\text{Row}_j(H) = \text{Row}_j(\bar{H})$ if the number of non-zero entries in $\text{Row}_j(\bar{H})$ does not exceed L and setting $\text{Row}_j(H) = 0$ otherwise. Let $\{X_{l,i} : i \in [n], l \in [K]\}$ be i.i.d. Rademacher random variables, and enumerate the elements of each random set $J_i, i \in [n]$:

$$J_i = \{j(i, 1), \dots, j(i, K)\} \text{ where } j(i, 1) < j(i, 2) < \dots < j(i, K)$$

The signs $\eta_{1,j}$, $\eta_{2,j}$, $X_{l,i}$, and the subsets J_i are mutually independent. Let $\bar{R}_2$ be the random matrix with entries

$$[\bar{R}_2]_{j(i,l),i} = \begin{cases} \frac{1}{L} \cdot X_{l,i}, & \text{for } l \in [K] \\ 0 & \text{otherwise} \end{cases}$$

Let $R_2 = H \odot \bar{R}_2$. Then $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ is called a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ , K , and L .

The random perturbation constructed in the above manner has an $O(1)$ norm bound.

Remark 4.4 (Norm of the perturbation). Let $0 < \alpha, \beta, \gamma < 1$ and $\rho > 0$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ , K , and L . Then, $\|R_1 + R_2\| \leq 2$.

Proof. Note that the property $\|V\| \leq \rho\sqrt{n}$ leads to $\|R_1\| \leq 1$, and by the Schur test,

$$\|R_2\| \leq \max_{j \in n} \{\|\text{Row}_j(R_2)\|_1, \|\text{Col}_j(R_2)\|_1\} \leq 1$$

Thus, we have $\|R_1 + R_2\| \leq \|R_1\| + \|R_2\| \leq 2$. $\square$

5 Pattern Matrices In this section, we first give an example of a $\sqrt{n}$ -sparse vector whose Walsh-Hadamard transform is also $\sqrt{n}$ -sparse and then provide details of the randomized construction of pattern matrices mentioned in Section 2.1.

PROPOSITION 5.1. *Let $n = 2^{2k}$ for $k > 0$, and let $H = H_n$ be the n -dimensional Walsh Hadamard matrix. Then, there exists $x \in \mathbb{R}^n$ with $\text{supp}(x) = 2^k = \sqrt{n}$ such that $\text{supp}(Hx) = 2^k = \sqrt{n}$.*

Proof. Let $n = 2^{2k}$ for $k > 0$. Consider an $n \times n$ Walsh Hadamard matrix H and index its rows and columns with tuples from $\mathbb{F}_2^{2k} \cong \mathbb{F}_2^k \times \mathbb{F}_2^k := \{(a_1, \dots, a_k, b_1, \dots, b_k) | a_i, b_i \in \{0, 1\}\}$. That is to say, the $(i, j)^{\text{th}}$ entry of H is referenced as the $((i_a, i_b), (j_a, j_b))^{\text{th}}$ entry of H , where

$$\begin{aligned} (i_a, i_b) &= (i_{a_1}, \dots, i_{a_k}, i_{b_1}, \dots, i_{b_k}) \in \{0, 1\}^{2k} \\ (j_a, j_b) &= (j_{a_1}, \dots, j_{a_k}, j_{b_1}, \dots, j_{b_k}) \in \{0, 1\}^{2k} \end{aligned}$$

are the binary representations of i and j in $2k$ bits. One can then check that $H_{(i,j)} = H_{((i_a, i_b), (j_a, j_b))} = (-1)^{i_a \cdot j_a + i_b \cdot j_b}$ where $\cdot$ represents the usual component wise inner product of tuples.

Now consider the vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ given by

$$x_j = x_{(j_a, j_b)} = \begin{cases} 1 & \text{if } j_a = (0, \dots, 0) \\ 0 & \text{otherwise} \end{cases}$$

x is only supported on 2^k coordinates, so is $\sqrt{n}$ -sparse, and,

$$\begin{aligned} (Hx)_i &= (Hx)_{(i_a, i_b)} \\ &= \sum_{j_a, j_b} (-1)^{i_a \cdot j_a + i_b \cdot j_b} \mathbb{1}_{j_a = (0, \dots, 0)} \\ &= \sum_{j_b} (-1)^{i_b \cdot j_b} \end{aligned}$$

Observe that the sum is non-zero if and only if $i_b = (0, \dots, 0)$. Thus, Hx is supported on 2^k coordinates as well. $\square$

The following proposition proves that the model for pattern matrices described in Section 2.1 results in matrices that satisfy the relevant properties (as in Definition 4.1) with high probability.

PROPOSITION 5.2. *Let V_1 be a matrix with Rademacher entries that are $2q$ -wise independent where $q = \lceil 5 \log n \rceil$. Let W be a vector with 4 -wise independent Rademacher coordinates. Let V_2 and V_3 be a random matrices whose rows are independent copies of W . Let*

$$V = V_1 \odot V_2 \odot V_3^T$$

where we recall that $\odot$ stands for the entry-wise (Hadamard) product and V_1, V_2, V_3 are independent of each other. There exists constants $c_{5.2.1}, c_{5.2.2}, c_{5.2.3}, c_{5.2.4}, c_{5.2.5} > 0$ with $c_{5.2.3}, c_{5.2.5} < 1$, such that for any $n \geq c_{5.2.1}$, with probability at least $1 - 1/n^2$, V satisfies the following properties:

1. $\|V\| \leq c_{5.2.2}\sqrt{n}$,
2. For any $\eta_1, \dots, \eta_n \in \{-1, 1\}$, and for any αn -sparse vector $x \in \mathbb{S}^{n-1}$ with $\alpha \leq c_{5.2.3}$, we have

$$|\{j \in [n] : |e_j^T V \text{diag}(\eta_1, \dots, \eta_n)x| \geq c_{5.2.4}\}| \geq c_{5.2.5}n$$

and

$$|\{j \in [n] : |e_j^T V^T \text{diag}(\eta_1, \dots, \eta_n)x| \geq c_{5.2.4}\}| \geq c_{5.2.5}n$$

Proof. We first look at the norm. Conditioned on V_2 and V_3 , V is a random matrix with Rademacher entries that are $2q$ -wise independent. Since $q = \lceil 5 \log n \rceil$, we can apply in [20, Lemma 2] and obtain

$$\begin{aligned} \mathbb{P}(\|V\| \geq 3\sqrt{n}|V_2, V_3) &\leq 2n \left(\frac{3}{2}\right)^{-2q} \\ &\leq 2n^{1-10 \log(3/2)} \\ &\leq \frac{2}{n^3}. \end{aligned}$$

Taking expectation over V_2 and V_3 , we obtain

$$\mathbb{P}(\|V\| \geq 3\sqrt{n}) \leq \frac{2}{n^3}.$$

Next, we consider the number of "good" rows. Fix $x \in \mathbb{S}^{n-1}$, $\eta_1, \dots, \eta_n \in \{-1, 1\}$ and $j \in [n]$, and let $X_j = X_j(\eta_1, \dots, \eta_n, \beta, x)$ be the indicator random variable for the event that $|e_j^T V D_\eta x| \leq \beta$, for some $\beta > 0$ where $D_\eta = \text{diag}(\eta_1, \dots, \eta_n)$. Conditioned on V_1 and V_3 , the rows of V are independent and each row has 4-wise independent Rademacher coordinates. Let $Z_j = (e_j^T V D_\eta x)^2$. Then, $\mathbb{E} Z_j = 1$. By Paley-Zygmund inequality, we have

$$\mathbb{P}(Z_j > \theta) \geq \frac{(1 - \theta)^2}{\mathbb{E} Z_j^2}$$

Since the expression for Z_j^2 only has products of order 4 in the entries of V , and the coordinates in each row of V are 4-wise independent, $\mathbb{E} Z_j^2$ remains unchanged if we assume that the rows of V have independent ± 1 entries. In this case, by using Hoeffding's inequality and sub-gaussianity, we have that $1 \leq \mathbb{E} Z_j^2 \leq c_3$. Therefore, there exist c_4 and c_5 such that $\mathbb{P}(|e_j^T V D_\eta x| \leq c_4) = \mathbb{E} X_j(\eta_1, \dots, \eta_n, c_4, x) \leq c_5 < 1$. By Hoeffding's inequality, we can find $c_6 < 1$ and $c_7 > 0$ so that

$$\mathbb{P}\left(\sum_{j=1}^n X_j(\eta_1, \dots, \eta_n, c_4, x) \geq c_6 n\right) \leq \exp(-c_7 n)$$

which implies

$$\mathbb{P}(|\{j \in [n] : |e_j^T V D_\eta x| \leq c_4\}| \geq c_6 n) \leq \exp(-c_7 n)$$

If we further assume that x is αn -sparse, then $e_j^T V D_\eta x$ depends on only those η_i for which i is in the support of x . Thus, using the union bound, we obtain

$$\begin{aligned} \mathbb{P}(\exists \eta \in \{-1, 1\}^n : |\{j \in [n] : |e_j^T V D_\eta x| \leq c_4\}| \geq c_6 n) &\leq 2^{\alpha n} \exp(-c_7 n) \\ &\leq \exp(-c_7 n/2) \end{aligned}$$

when $\alpha < c_7/2 \log 2$.

Let $\mathcal{N}$ be a ν -net of αn -sparse unit vectors for the set of all αn -sparse unit vectors. This means that for every αn -sparse $y \in \mathbb{S}^{n-1}$, there exists an αn -sparse unit $x \in \mathcal{N}$ such that $\|x - y\|_2 \leq \nu$.

When $\log |\mathcal{N}| \leq c_7 n/4$ (we shall later show that this is the case when α is small enough), by a union bound over $\mathcal{N}$, we get

$$\mathbb{P}(\exists x \in \mathcal{N}, \eta \in \{-1, 1\}^n : |\{j \in [n] : |e_j^T V D_\eta x| \leq c_4\}| \geq c_6 n) \leq \exp(-c_7 n/4)$$

Define the event $\mathcal{E} := \{|\{j \in [n] : |e_j^T V D_\eta x| \geq c_4\}| \geq (1 - c_6)n \quad \forall x \in \mathcal{N}, \eta \in \{-1, 1\}^n\}$. We have shown that $\mathbb{P}(\mathcal{E}) \geq 1 - \exp(-c_7 n/4)$.

Now let y be an arbitrary αn -sparse vector. Then there is $x \in \mathcal{N}$ such that $\|x - y\| \leq \nu$. Conditioning on the event that $\|V\| \leq c_2 \sqrt{n}$, we have $\|V D_\eta(x - y)\| \leq c_2 \nu \sqrt{n}$ for all η . Then,

$$|\{j \in [n] : |e_j^T V D_\eta(x - y)| \geq \lambda c_2 \nu\}| \leq \frac{n}{\lambda^2}$$

Let λ be such that $1/\lambda^2 = (1 - c_6)/2$. Once λ has been chosen, let ν be chosen such that $\lambda c_2 \nu = c_4/2$. Then, conditioned on $\mathcal{E}$ and on the event that $\|V\| \leq c_2 \sqrt{n}$, we have, for all $\eta \in \{-1, 1\}^n$,

$$\begin{aligned} |\{j \in [n] : |e_j^T V D_\eta x| \geq c_4 \text{ and } |e_j^T V D_\eta (x - y)| \leq c_4/2\}| &\geq (1 - c_6)n/2 \\ \implies |\{j \in [n] : |e_j^T V D_\eta y| \geq c_4/2\}| &\geq (1 - c_6)n/2 \end{aligned}$$

We now show that $\log|\mathcal{N}| \leq (c_7/4)n$ when α is small enough. From [37, Corollary 4.2.11] we have,

$$\begin{aligned} |\mathcal{N}| &\leq \binom{n}{\lfloor \alpha n \rfloor} \left(\frac{3}{\nu}\right)^{\lfloor \alpha n \rfloor} \\ &\leq \left(\frac{3e}{\nu \alpha}\right)^{\alpha n} \end{aligned}$$

When $\alpha < \min\{\nu, 1/3e\}$,

$$\begin{aligned} \log|\mathcal{N}| &\leq \alpha n \log\left(\frac{3e}{\nu \alpha}\right) \\ &\leq \alpha n \log(3e) + \alpha n \log\left(\frac{1}{\nu}\right) + \alpha n \log\left(\frac{1}{\alpha}\right) \\ &\leq 3\alpha n \log\left(\frac{1}{\alpha}\right) \end{aligned}$$

We further let α be small enough that $3\alpha \log(\frac{1}{\alpha}) \leq c_7/4$.

We can repeat the same arguments by looking at V^T with V_3 in place of V_2 , to get the third property. Finally, by union bound, the probability that any of the required properties fails is at most

$$\frac{2}{n^3} + 2 \exp(-c_7 n/4) \leq \frac{1}{n^2}$$

□

for all sufficiently large n .

COROLLARY 5.3. *For any $n \geq c_{5.2.1}$, $\rho \geq c_{5.2.2}$, $0 < \alpha \leq c_{5.2.3}$, $0 < \beta \leq c_{5.2.4}$, and $0 < \gamma \leq c_{5.2.5}$, there exists an n -dimensional pattern matrix V with parameters $\alpha, \beta, \gamma, \rho$. In addition, there exists an algorithm using $O(n \log n)$ random bits that can generate such a pattern matrix with success probability at least $1 - 1/n^2$.*

Proof. We only need to justify the claim about random bit complexity. Considering the construction of V , we need to generate n^2 many $2q$ -wise independent signs, and also $2n$ independent families where each family contains n many 4-wise independent signs.

By [36, Corollary 3.34], we can choose the entries of V_1 and the distribution of W so that the number of random bits needed is at most

$$2q \lceil \log_2(n^2) \rceil + 8n \lceil \log_2 n \rceil = O(n \log n).$$

We use independent seeds for V_1 and for all the rows of V_2 and V_3 . The probability estimate follows from Proposition 5.2. □

6 Invertibility for compressible vectors This section is devoted to proving the following proposition, which establishes a lower bound for the infimum of $\|Mx\|_2$ over compressible vectors.

THEOREM 6.1. *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \overline{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \overline{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Then exist constants $c_{6.1.1} > 0$ and $0 < c_{6.1.2}, c_{6.1.3}, c_{6.1.4} < 1$ such that, for $\alpha < \frac{\gamma}{c_{6.1.1} \log\left(\frac{c_{6.1.1} \rho}{\beta \gamma^{3/2}}\right)}$ and $\nu = c_{6.1.2} \varepsilon \frac{\beta \sqrt{\gamma}}{\rho}$, we have*

$$\mathbb{P}\left(\inf_{v \in \text{Comp}(\alpha, \nu)} \|Mv\|_2 \geq c_{6.1.3} \varepsilon \frac{\beta \sqrt{\gamma}}{\rho}\right) \geq 1 - \exp(-c_{6.1.4} \gamma n)$$

Proof. Fix $x \in \mathbb{S}^{n-1}$ with $|\text{supp}(x)| \leq \alpha n$. Condition on D_2 and R_2 and let $A_1 = A + \varepsilon R_2$. Then, $Mx = A_1x + \frac{\varepsilon}{\rho\sqrt{n}} \cdot D_1 V D_2 x$. Let $J(x) = \{j \in [n] : |\text{Row}_j(V) D_2 x| \geq \beta\}$. By the definition of a pattern matrix, we have $|J(x)| \geq \gamma n$. Then, for $j \in J(x)$, we have

$$\mathbb{P}\left(|(Mx)_j| \geq \frac{\varepsilon\beta}{\rho\sqrt{n}} \middle| R_2, D_2\right) = \mathbb{P}\left(\left|(A_1x)_j + \frac{\varepsilon}{\rho\sqrt{n}} \cdot \eta_{1,j} \text{Row}_j(V) D_2 x\right| \geq \frac{\varepsilon\beta}{\rho\sqrt{n}} \middle| R_2, D_2\right) \geq 1/2$$

To see why observe that, since $j \in J(x)$, we have

$$\left|\frac{\varepsilon}{\rho\sqrt{n}} \text{Row}_j(V) D_2 x\right| \geq \frac{\varepsilon\beta}{\rho\sqrt{n}},$$

and therefore the two numbers

$$(A_1x)_j + \frac{\varepsilon}{\rho\sqrt{n}} \text{Row}_j(V) D_2 x, \quad (A_1x)_j - \frac{\varepsilon}{\rho\sqrt{n}} \text{Row}_j(V) D_2 x$$

cannot both have absolute value strictly smaller than $\varepsilon\beta/(\rho\sqrt{n})$.

Since the coordinates of Mx are independent after conditioning on D_2 and R_2 , we can use the following tensorization lemma,

LEMMA 6.2 ([28], Lemma 2.2 (2)). *Let $X_1, \dots, X_N$ be independent random variables satisfying*

$$\mathbb{P}(|X_i| \geq 1) \geq c_{6.2.1}$$

for some $c_{6.2.1} \in (0, 1)$. Then, there exists constants $c_{6.2.2}, c_{6.2.3} > 0$, depending on $c_{6.2.1}$, so that for the random vector $X := (X_1, \dots, X_N)$, we have,

$$\mathbb{P}\left(\|X\|_2 \leq c_{6.2.2} \sqrt{N}\right) \leq e^{-c_{6.2.3} N}$$

Applying Lemma 6.2 with $X_j = \frac{\rho\sqrt{n}}{\varepsilon\beta} (Mx)_j$ for $j \in J(x)$, and observing that

$$\|Mx\| \geq \sqrt{\sum_{j \in J(x)} (Mx)_j^2},$$

we get

$$\mathbb{P}_{D_1}\left(\|Mx\|_2 \leq c_{6.2.2} \frac{\beta\varepsilon}{\rho\sqrt{n}} \sqrt{|J(x)|}\right) \leq e^{-c_{6.2.3} |J(x)|}$$

or equivalently we can write

$$\mathbb{P}_{D_1}\left(\|Mx\|_2 \leq c_1 \varepsilon \frac{\beta\sqrt{\gamma}}{\rho}\right) \leq e^{-c_2 \gamma n}$$

for absolute constants c_1 and c_2 , using the fact that $|J(x)| \geq \gamma n$.

Define

$$(6.1) \quad G_{A_1} = \{v \in \mathbb{R}^n : \|v\|_2 \leq 1 \text{ and } \|A_1 v\|_2 \leq 2\varepsilon\}.$$

For each $I \subset [n]$, the set $G_{A_1} \cap E_I$ is convex and about the origin in the coordinate subspace E_I .

Choose

$$(6.2) \quad \eta = c_1 \frac{\beta\sqrt{\gamma}}{96\rho}.$$

Since $0 < c_1, \beta, \gamma < 1$ and $\rho > 1$, we have $0 < \eta < 1$.

Let $G_I = G_{A_1} \cap E_I$. By the standard volumetric estimate as in Proposition 3.4, applied in the subspace E_I , the set G_I can be covered by at most

$$\left(1 + \frac{2}{\eta}\right)^{|I|}$$

translates of ηG_I . To see why, if $\{z_k\}_{k=1}^N \subset G_I$ is a maximal set such that the translates $z_k + \frac{\eta}{2}G_I$ are pairwise disjoint, then maximality implies that the translates $z_k + \eta G_I$ cover G_I . Also we have

$$\bigcup_{k=1}^N \left(z_k + \frac{\eta}{2}G_I\right) \subset G_I + \frac{\eta}{2}G_I = \left(1 + \frac{\eta}{2}\right)G_I.$$

Since we can identify E_I with $\mathbb{R}^{|I|}$ by an isometry, we can define $|I|$ dimensional volume vol_I , i.e., Lebesgue measure, on E_I , such that $\text{vol}_I(\frac{\eta}{2}G_I) = \left(\frac{\eta}{2}\right)^{|I|} \text{vol}_I(G_I)$. Therefore we have

$$N \left(\frac{\eta}{2}\right)^{|I|} \text{vol}_I(G_I) \leq \left(1 + \frac{\eta}{2}\right)^{|I|} \text{vol}_I(G_I)$$

which implies

$$N \leq \left(1 + \frac{2}{\eta}\right)^{|I|}.$$

Now we select one point from each translate which intersects

$$(S^{n-1} \cap E_I) \cap G_{A_1},$$

to form a net $\mathcal{N}_I$. Then every point of $\mathcal{N}_I$ is a unit vector supported in I , and

$$(6.3) \quad |\mathcal{N}_I| \leq \left(1 + \frac{2}{\eta}\right)^{|I|} \leq \left(\frac{3}{\eta}\right)^{|I|} = \left(\frac{288\rho}{c_1\beta\sqrt{\gamma}}\right)^{|I|}.$$

We claim that, for every $x \in (S^{n-1} \cap E_I) \cap G_{A_1}$ there exists $y \in \mathcal{N}_I$ such that

$$(6.4) \quad x - y \in 2\eta(G_{A_1} \cap E_I).$$

To explain why, if x lies in a translate $z + \eta(G_{A_1} \cap E_I)$, and $y \in \mathcal{N}_I$ is the chosen representative from the same translate, then $x = z + \eta u$ and $y = z + \eta w$ for some $u, w \in G_{A_1} \cap E_I$. Since $G_{A_1} \cap E_I$ is convex and centrally symmetric, $(u - w)/2 \in G_{A_1} \cap E_I$, and hence $x - y = 2\eta((u - w)/2) \in 2\eta(G_{A_1} \cap E_I)$.

Fix $I \subset [n]$ with $|I| \leq \alpha n$. Suppose that there exists $x \in S^{n-1} \cap E_I$ such that

$$(6.5) \quad \|Mx\|_2 \leq \frac{1}{16}c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon.$$

We first show that $x \in G_{A_1} \cap E_I$. Since $M = A_1 + \varepsilon R_1$ and $\|R_1\| \leq 1$,

$$\|A_1 x\|_2 \leq \|Mx\|_2 + \varepsilon \|R_1 x\|_2 \leq \frac{1}{16}c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon + \varepsilon.$$

Since $0 < c_1, \beta, \gamma < 1$ and $\rho > 1$, we have $\beta\sqrt{\gamma}/\rho \leq 1$. Therefore, we have

$$\|A_1 x\|_2 \leq \frac{17}{16}\varepsilon \leq 2\varepsilon.$$

So $x \in G_{A_1} \cap E_I$.

Choose $y \in \mathcal{N}_I$ satisfying (6.4). From $x - y \in 2\eta(G_{A_1} \cap E_I)$ and the definition of G_{A_1} , we obtain

$$\|x - y\|_2 \leq 2\eta, \quad \|A_1(x - y)\|_2 \leq 4\eta\varepsilon.$$

Using $M = A_1 + \varepsilon R_1$ and $\|R_1\| \leq 1$, we get

$$\|M(x - y)\|_2 \leq \|A_1(x - y)\|_2 + \varepsilon \|R_1(x - y)\|_2 \leq 4\eta\varepsilon + 2\eta\varepsilon = 6\eta\varepsilon.$$

By definition (6.2), we have

$$6\eta\varepsilon = \frac{1}{16}c_1 \frac{\beta\sqrt{\gamma}}{\rho}\varepsilon.$$

Combining this with (6.5), we have

$$(6.6) \quad \|My\|_2 \leq \|Mx\|_2 + \|M(x - y)\|_2 \leq \frac{1}{8}c_1 \frac{\beta\sqrt{\gamma}}{\rho}\varepsilon.$$

In conclusion, for every support I , the existence of a bad vector $x \in S^{n-1} \cap E_I$ implies the existence of a unit net point $y \in \mathcal{N}_I$ satisfying (6.6).

Therefore, by union bound, for fixed I , we have

$$\begin{aligned} & \mathbb{P}_{D_1} \left\{ \exists x \in S^{n-1} \cap E_I : \|Mx\|_2 \leq \frac{1}{16}c_1 \frac{\beta\sqrt{\gamma}}{\rho}\varepsilon \right\} \\ & \leq \sum_{y \in \mathcal{N}_I} \mathbb{P}_{D_1} \left\{ \|My\|_2 \leq \frac{1}{8}c_1 \frac{\beta\sqrt{\gamma}}{\rho}\varepsilon \right\} \leq \left(\frac{288\rho}{c_1\beta\sqrt{\gamma}} \right)^{|I|} \exp(-c_2\gamma n). \end{aligned}$$

Choose

$$B = \frac{288e\rho}{c_1\beta\gamma^{3/2}}.$$

Since $0 < c_1, \beta, \gamma < 1$ and $\rho > 1$, we have $B \geq 288e$ and therefore $\log B > 5$.

Choose the constant $c_{6.1.1} > 0$ sufficiently large so that

$$c_{6.1.1} \geq \frac{288e}{c_1}$$

and, for all $t \geq e^5$, we have

$$(6.7) \quad \frac{1}{c_{6.1.1}} \frac{\log(c_{6.1.1}t \log t)}{\log t} \leq \frac{c_2}{2}.$$

This is possible because

$$\frac{\log(c_{6.1.1}t \log t)}{c_{6.1.1} \log t} \leq \frac{\log c_{6.1.1}}{5c_{6.1.1}} + \frac{2}{c_{6.1.1}}$$

for all $t \geq e^5$, and the right-hand side tends to 0 as $c_{6.1.1} \rightarrow \infty$.

By the assumption on α and by the choice $c_{6.1.1} \geq 288e/c_1$, we have

$$\alpha < \frac{\gamma}{c_{6.1.1} \log \left(\frac{c_{6.1.1}\rho}{\beta\gamma^{3/2}} \right)} \leq \frac{\gamma}{c_{6.1.1} \log B} < \frac{1}{2}.$$

Using Lemma 3.3 and taking the union bound over all $I \subset [n]$ with $|I| \leq \alpha n$, we have

$$\begin{aligned} & \mathbb{P}_{D_1} \left\{ \exists x \in S^{n-1} : |\text{supp}(x)| \leq \alpha n, \|Mx\|_2 \leq \frac{1}{16}c_1 \frac{\beta\sqrt{\gamma}}{\rho}\varepsilon \right\} \\ & \leq \exp \left(\alpha n \log \frac{e}{\alpha} \right) \left(\frac{288\rho}{c_1\beta\sqrt{\gamma}} \right)^{\alpha n} \exp(-c_2\gamma n) \\ (6.8) \quad & = \exp \left(\alpha n \log \left(\frac{288e\rho}{c_1\alpha\beta\sqrt{\gamma}} \right) - c_2\gamma n \right). \end{aligned}$$

We now verify that $\alpha n \log \left(\frac{288e\rho}{c_1\alpha\beta\sqrt{\gamma}} \right)$ can be absorbed by $c_2\gamma n$. The function

$$f(s) = s \log \left(\frac{288e\rho}{sc_1\beta\sqrt{\gamma}} \right), \quad 0 < s \leq 1,$$

is increasing on $(0, 1]$, because

$$f'(s) = \log \left(\frac{288e\rho}{sc_1\beta\sqrt{\gamma}} \right) - 1 \geq \log(288e) - 1 > 0$$

where we use

$$\frac{288e\rho}{c_1\beta\sqrt{\gamma}} \geq 288e.$$

Therefore, we have

$$(6.9) \quad \alpha \log \left(\frac{288e\rho}{c_1\alpha\beta\sqrt{\gamma}} \right) \leq \frac{\gamma}{c_{6.1.1} \log B} \log \left(\frac{288e\rho}{c_1\beta\sqrt{\gamma}} \cdot \frac{c_{6.1.1} \log B}{\gamma} \right) = \frac{\gamma}{c_{6.1.1} \log B} \log (c_{6.1.1} B \log B) \leq \frac{c_2\gamma}{2}$$

where the last inequality follows from (6.7) and $\log B > 5$.

Combining (6.8) and (6.9) we have

$$(6.10) \quad \mathbb{P}_{D_1} \left\{ \inf_{x \in S^{n-1}: |\text{supp}(x)| \leq \alpha n} \|Mx\|_2 \leq \frac{1}{16} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon \right\} \leq \exp \left(-\frac{c_2\gamma n}{2} \right).$$

Finally, we extend the estimates from the sparse vectors to all compressible vectors. Let

$$\nu = \frac{1}{192} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon.$$

On the complement of the event in (6.10), every αn -sparse unit vector x satisfies

$$\|Mx\|_2 > \frac{1}{16} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon.$$

Let $v \in \text{Comp}(\alpha, \nu)$. Then there exists an αn -sparse vector x with $\|v - x\|_2 \leq \nu$. Since $\|x\|_2 \geq 1 - \nu > 0$, the vector $x/\|x\|_2$ is a sparse unit vector. We also have

$$\begin{aligned} \left\| v - \frac{x}{\|x\|_2} \right\|_2 &\leq \|v - x\|_2 + \left\| x - \frac{x}{\|x\|_2} \right\|_2 \\ &= \|v - x\|_2 + |\|x\|_2 - 1| \\ &\leq 2\|v - x\|_2 \leq 2\nu. \end{aligned}$$

By Remark 4.4, we have $\|M\| \leq 3$. Therefore, we have

$$\begin{aligned} \|Mv\|_2 &\geq \left\| M \frac{x}{\|x\|_2} \right\|_2 - \|M\| \left\| v - \frac{x}{\|x\|_2} \right\|_2 \\ &> \frac{1}{16} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon - 3 \cdot 2 \cdot \frac{1}{192} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon \\ &\geq \frac{1}{16} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon - \frac{1}{32} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon \\ &= \frac{1}{32} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon. \end{aligned}$$

Therefore, by iterated expectation, we have

$$\begin{aligned} &\mathbb{P} \left\{ \inf_{v \in \text{Comp}(\alpha, \nu)} \|Mv\|_2 < \frac{1}{32} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon \right\} \\ &= \mathbb{E}_{D_2, R_2} \mathbb{P}_{D_1} \left\{ \inf_{v \in \text{Comp}(\alpha, \nu)} \|Mv\|_2 < \frac{1}{32} c_1 \frac{\beta\sqrt{\gamma}}{\rho} \varepsilon \right\} \\ &\leq \exp \left(-\frac{c_2\gamma n}{2} \right). \end{aligned}$$

□

7 Invertibility for Incompressible Vectors To bound $\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2$, we use the following invertibility by distance result from [28].

LEMMA 7.1 (Lemma 3.5 in [28]). *Let M be any random matrix. Let $Y_1, \dots, Y_n$ denote the column vectors of M , and let H_k denote the span of all column vectors except the k th. Then for every $\alpha, \nu \in (0, 1)$ and every $t > 0$, one has*

$$(7.1) \quad \mathbb{P} \left(\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2 < t \nu n^{-1/2} \right) \leq \frac{1}{\alpha n} \sum_{k=1}^n \mathbb{P}(\text{dist}(Y_k, H_k) < t).$$

We will apply Lemma 7.1 to our case $M = A + \varepsilon R$. Let $Y_1, \dots, Y_n$ denote the column vectors of M , and let H_k denote the span of all column vectors except the k th. Without loss of generality, we analyze

$$\mathbb{P}(\text{dist}(Y_n, H_n) < t)$$

and all the other terms in the sum

$$\sum_{k=1}^n \mathbb{P}(\text{dist}(Y_k, H_k) < t)$$

can be bounded by exactly the same way.

By constructing a measurable random normal vector Z of the subspace H_n , we have

$$\text{dist}(Y_n, H_n) \geq |\langle Y_n, Z \rangle|$$

and therefore it suffices to bound

$$\mathbb{P}(|\langle Y_n, Z \rangle| < t)$$

To this end, we first prove that the random normal vector Z is incompressible with high probability.

LEMMA 7.2 (Random normal is incompressible). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho > 0$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Let $Y_1, \dots, Y_n$ be the columns of M . Let ζ be a measurable function such that for every $(y_1, \dots, y_{n-1}) \in (\mathbb{R}^n)^{n-1}$,*

$$\zeta(y_1, \dots, y_{n-1}) \in S^{n-1} \cap \text{span}(y_1, \dots, y_{n-1})^\perp.$$

Let $Z = \zeta(Y_1, \dots, Y_{n-1})$. Then there exists an absolute constant $c_{7.2.1} > 0$ such that, if

$$n \geq \frac{2}{\gamma}, \quad \alpha < \frac{\gamma/2}{c_{6.1.1} \log\left(\frac{c_{6.1.1}\rho}{\beta(\gamma/2)^{3/2}}\right)}, \quad \nu = c_{6.1.2} \varepsilon \frac{\beta\sqrt{\gamma/2}}{\rho},$$

then

$$\mathbb{P}(Z \in \text{Comp}(\alpha, \nu)) \leq \exp(-c_{7.2.1}\gamma n).$$

Proof. Let M' be the $(n-1) \times n$ matrix such that M' is M^T with last row erased. We have

$$\{Z \in \text{Comp}(\alpha, \nu)\} \subset \{\exists x(x \in \text{Comp}(\alpha, \nu), M'x = 0)\}$$

Condition on D_1 and R_2 , and put $A_1 = A + \varepsilon R_2$. Since

$$M^T x = A_1^T x + \frac{\varepsilon}{\rho\sqrt{n}} D_2 V^T D_1 x,$$

the j th row of M' satisfies, for $j \in [n-1]$,

$$(M'x)_j = (A_1^T x)_j + \frac{\varepsilon}{\rho\sqrt{n}} \eta_{2,j} \text{Row}_j(V^T) D_1 x.$$

For a fixed αn -sparse $x \in S^{n-1}$, define

$$J(x) = \{j \in [n-1] : |\text{Row}_j(V^T)D_1x| \geq \beta\}.$$

By definition of pattern matrix, before erasing the last row there are at least γn such coordinates. Therefore

$$|J(x)| \geq \gamma n - 1 \geq \frac{\gamma n}{2},$$

where the last inequality uses $n \geq 2/\gamma$. For each $j \in J(x)$,

$$\begin{aligned} & \mathbb{P} \left(|(M'x)_j| \geq \frac{\varepsilon\beta}{\rho\sqrt{n}} \mid R_2, D_1 \right) \\ &= \mathbb{P} \left(\left| (A_1^T x)_j + \frac{\varepsilon}{\rho\sqrt{n}} \eta_{2,j} \text{Row}_j(V^T)D_1x \right| \geq \frac{\varepsilon\beta}{\rho\sqrt{n}} \mid R_2, D_1 \right) \geq \frac{1}{2}. \end{aligned}$$

The coordinates indexed by $J(x)$ are conditionally independent because the signs $\eta_{2,j}$ are independent after conditioning on (R_2, D_1) . Tensorization therefore gives absolute constants $c_1, c_2 > 0$ such that

$$\mathbb{P} \left(\|M'x\|_2 \leq c_1 \varepsilon \frac{\beta\sqrt{\gamma/2}}{\rho} \mid R_2, D_1 \right) \leq \exp(-c_2 \gamma n). \quad \square$$

The results then follows from the same ellipsoid covering argument used in Theorem 6.1.

Now we have proved that the random normal vector Z is an incompressible vector with high probability. This is useful because incompressible vectors are spread according to the following lemma in [28].

LEMMA 7.3 (Incompressible vectors are spread). *Let $0 < \alpha, \nu < 1$, and let $x \in \text{Incomp}(\alpha, \nu)$. Then the set*

$$\Theta = \left\{ i \in [n] : \frac{\nu}{\sqrt{2n}} < |x_i| \leq \sqrt{\frac{2}{\alpha n}} \right\}$$

satisfies

$$|\Theta| > \frac{\alpha n}{2}.$$

Proof. Define

$$A = \left\{ i \in [n] : |x_i| \leq \frac{\nu}{\sqrt{2n}} \right\}.$$

Let $P_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denote the orthogonal projection onto the coordinate subspace supported on A , i.e.,

$$(P_A x)_i = \begin{cases} x_i, & i \in A, \\ 0, & i \notin A. \end{cases}$$

Then

$$\|P_A x\|_2^2 = \sum_{i \in A} |x_i|^2 \leq |A| \cdot \frac{\nu^2}{2n} \leq n \cdot \frac{\nu^2}{2n} = \frac{\nu^2}{2}.$$

Hence

$$\|P_A x\|_2 \leq \frac{\nu}{\sqrt{2}} < \nu.$$

We claim that

$$|A^c| > \alpha n.$$

Indeed, suppose for contradiction that $|A^c| \leq \alpha n$. Put

$$y = P_{A^c} x.$$

Then $y \neq 0$, because otherwise $x = P_A x$ and therefore

$$1 = \|x\|_2 = \|P_A x\|_2 \leq \frac{\nu}{\sqrt{2}} < 1,$$

a contradiction. Now define

$$z = \frac{y}{\|y\|_2}.$$

Then $z \in S^{n-1}$ and

$$|\text{supp}(z)| \leq |A^c| \leq \alpha n.$$

Set

$$\delta = \|x - y\|_2 = \|P_A x\|_2.$$

Since $x = y + P_A x$ and the two summands have disjoint supports, they are orthogonal. Because $\|x\|_2 = 1$, we get

$$\|y\|_2^2 = 1 - \delta^2.$$

Therefore

$$\langle x, z \rangle = \left\langle y + P_A x, \frac{y}{\|y\|_2} \right\rangle = \|y\|_2,$$

and so

$$\|x - z\|_2^2 = 2 - 2\langle x, z \rangle = 2 - 2\|y\|_2 = 2(1 - \sqrt{1 - \delta^2}).$$

Since $0 \leq \delta \leq 1$,

$$1 - \sqrt{1 - \delta^2} = \frac{\delta^2}{1 + \sqrt{1 - \delta^2}} \leq \delta^2.$$

Hence

$$\|x - z\|_2^2 \leq 2\delta^2 = 2\|P_A x\|_2^2 \leq \nu^2.$$

Thus $\|x - z\|_2 \leq \nu$, with $z \in S^{n-1}$ and $|\text{supp}(z)| \leq \alpha n$. This shows that $x \in \text{Comp}(\alpha, \nu)$, contradicting $x \in \text{Incomp}(\alpha, \nu)$. The claim follows.

Now define

$$B = \left\{ i \in [n] : |x_i| > \sqrt{\frac{2}{\alpha n}} \right\}.$$

If $|B| \geq \alpha n/2$, then $B \neq \emptyset$ and therefore

$$1 = \|x\|_2^2 \geq \sum_{i \in B} |x_i|^2 > |B| \cdot \frac{2}{\alpha n} \geq \frac{\alpha n}{2} \cdot \frac{2}{\alpha n} = 1,$$

a contradiction. Hence

$$|B| < \frac{\alpha n}{2}.$$

Finally, let

$$\Theta = A^c \setminus B.$$

Then

$$|\Theta| \geq |A^c| - |B| > \alpha n - \frac{\alpha n}{2} = \frac{\alpha n}{2}.$$

Moreover, if $i \in \Theta$, then $i \notin A$ and $i \notin B$, so

$$\frac{\nu}{\sqrt{2n}} < |x_i| \leq \sqrt{\frac{2}{\alpha n}}.$$

This proves the claim. □

The standard argument (e.g., [28]) to bound the probability

$$\mathbb{P}(\text{dist}(Y_n, H_n) < t)$$

uses the fact

$$\text{dist}(Y_n, H_n) \geq |\langle Y_n, Z \rangle|$$

and then conditions on Z .

In our case, we cannot use this approach directly because $\sigma(Y_n)$ and $\sigma(Y_1, \dots, Y_{n-1})$ are not mutually independent. Nevertheless, the pure random signs $X_{i,n}$ will be independent with $\sigma(Y_1, \dots, Y_{n-1})$. To make the argument work, we need to show that there are sufficiently many random signs in the last column which will not be removed eventually ($h_{ij} \neq 0$). To understand when a random sign in the last column will be removed eventually, one key observation is that, a nonzero position in the n th column of $\bar{R}_2$ will be eventually be reset to zero if and only if it is in a row that is already heavy enough in the $1, \dots, n-1$ columns.

More formally, we define index set of the heavy rows of $\bar{H}$ up to column $n-1$ as

$$(7.2) \quad I_{\text{heavy},n} = \left\{ i \in [n] : \sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L \right\}$$

and the index set of the light rows up to column $n-1$

$$(7.3) \quad I_{\text{light},n} = \left\{ i \in [n] : \sum_{j=1}^{n-1} \bar{h}_{i,j} < L \right\}.$$

So the heavy rows may be removed if there is an extra nonzero entry in the column n , but the light rows will never be removed no matter what happens in the column n . With the above observation, we can write

$$\langle Y_n, Z \rangle = \langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}$$

We will write $\langle Y_n, Z \rangle$ as a function of two independent random elements. To this end, we first define the random elements

$$\Xi_1 = (R_1, (X_{l,j})_{l \in [K], j \in [n-1]}, J_1, \dots, J_{n-1}), \quad \Xi_2 = J_n. \quad \Xi_3 = \{X_{l,n} : l \in [K]\}$$

In other words, (Ξ_1, Ξ_2) contains the information of R_1 , the position of all nonzero entries of R_2 , and the value of random signs for the entries of R_2 only in columns $1, \dots, n-1$. The random element Ξ_3 contains the signs in column n . By construction, Ξ_1, Ξ_2, Ξ_3 are mutually independent, so it suffices to write $\langle Y_n, Z \rangle$ and a function of Ξ_1, Ξ_2, Ξ_3 so that we can use conditioning (iterated expectation). In particular, $I_{\text{heavy},n}$ and $I_{\text{light},n}$ are completely determined by Ξ_1 and the random normal vector Z is completely determined by (Ξ_1, Ξ_2) (or more precisely, completely determined by Ξ_1 and $J_n \cap I_{\text{heavy},n}$). Therefore, in the expression

$$\langle Y_n, Z \rangle = \langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}.$$

the parts, $\langle \text{Col}_n(A + \varepsilon R_1), Z \rangle$ and $Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}$ will be fixed after conditioning on (Ξ_1, Ξ_2) , and then bounding $\langle Y_n, Z \rangle$ becomes the Littlewood-Offord problem. To get a good bound, we need sufficiently many good coordinates to show up in the coefficient vector $\{Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}\}_{l \in [K]}$. To this end, the first step is to show there are enough light rows, or in other words, there cannot be too many heavy rows.

LEMMA 7.4 (Heavy rows are not too many). *Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters K and L . Let*

$$I_{\text{heavy},n} = \left\{ i \in [n] : \sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L \right\}.$$

be the index set of heavy rows up to column $n - 1$. Then, for $L > K$, we have

$$(7.4) \quad \mathbb{E}|I_{\text{heavy},n}| \leq ne^{-K} \left(\frac{eK}{L} \right)^L.$$

and for any $\lambda \in (0, 1]$, we have

$$(7.5) \quad \mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{\lambda}{2}n\right\} \leq \exp\left(-\frac{(\max\{(\frac{\lambda}{2}n - \mathbb{E}|I_{\text{heavy},n}|), 0\})^2}{2nK^2}\right)$$

$$(7.6) \quad \leq \exp\left(-\frac{(\max\{\frac{\lambda}{2}n - ne^{-K}(eK/L)^L, 0\})^2}{2nK^2}\right).$$

Proof. The idea of the proof is to first upper bound $\mathbb{E}|I_{\text{heavy},n}|$ and then obtain concentration inequalities for $|I_{\text{heavy},n}|$.

Fix $i \in [n]$ and $j \in [n - 1]$. Since J_j is uniform among all K -subsets of $[n]$, we have

$$(7.7) \quad \mathbb{P}(\bar{h}_{i,j} = 1) = \mathbb{P}(i \in J_j) = \frac{\binom{n-1}{K-1}}{\binom{n}{K}} = \frac{K}{n}.$$

Because $J_1, \dots, J_{n-1}$ are independent, the random variables $\bar{h}_{i,1} = \mathbf{1}_{\{i \in J_1\}}, \dots, \bar{h}_{i,n-1} = \mathbf{1}_{\{i \in J_{n-1}\}}$ are independent Bernoulli random variables with success probability K/n .

Therefore, for each row i , we have $\sum_{j=1}^{n-1} \bar{h}_{i,j}$ follows Binomial distribution with parameters $n - 1$ and K/n . In particular, the distribution does not depend on i . Let S be a Binomial random variable with parameters $n - 1$ and K/n . Observe that $\mathbb{E}[S] = K(1 - 1/n)$. Assume $L > K$. By Chernoff bound, we have

$$\mathbb{P}\{S \geq L\} \leq e^{-K} \left(\frac{eK}{L} \right)^L,$$

Since

$$|I_{\text{heavy},n}| = \sum_{i=1}^n \mathbf{1}_{\{\sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L\}},$$

by linearity of expectation, we have

$$(7.8) \quad \mathbb{E}|I_{\text{heavy},n}| = \sum_{i=1}^n \mathbb{P}\left\{\sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L\right\} = n \mathbb{P}\{S \geq L\} \leq ne^{-K} \left(\frac{eK}{L} \right)^L$$

To obtain concentration inequalities for $|I_{\text{heavy},n}|$, we can write

$$T = |I_{\text{heavy},n}| = \sum_{i=1}^n \mathbf{1}_{\{\sum_{j=1}^{n-1} \bar{h}_{i,j} \geq L\}} = f(J_1, \dots, J_{n-1}).$$

Let $k \in [n - 1]$ be fixed. Consider two sequences of column-sets

$$(J_1, \dots, J_{n-1}) \quad \text{and} \quad (J_1, \dots, J_{k-1}, J'_k, J_{k+1}, \dots, J_{n-1})$$

that differ only in the k th coordinate. For any fixed row index $i \in [n]$, replacing J_k by J'_k changes the row sum $\sum_{j=1}^n \bar{h}_{i,j}$ by at most 1, and it changes it only if $i \in J_k \Delta J'_k$ (symmetric difference). Therefore, the indicator $\mathbf{1}_{\{\sum_{j=1}^n \bar{h}_{i,j} \geq L\}}$ can change only for $i \in J_k \Delta J'_k$. So the total count T can change by at most

$$(7.9) \quad |J_k \Delta J'_k| \leq |J_k| + |J'_k| = 2K,$$

because both J_k and J'_k have exactly K elements. Therefore, T satisfies the bounded differences condition of Lemma 3.8 with $m = n - 1$ and

$$(7.10) \quad b_1 = \dots = b_{n-1} = 2K.$$

By Lemma 3.8 and (7.10), we have

$$v = \frac{1}{4} \sum_{k=1}^{n-1} b_k^2 = \frac{1}{4} (n-1) (2K)^2 \leq nK^2.$$

Therefore, for every $t > 0$, we have

$$(7.11) \quad \mathbb{P}\{T - \mathbb{E}T > t\} \leq \exp\left(-\frac{t^2}{2nK^2}\right).$$

To bound $\mathbb{P}\{T > \frac{\lambda}{2}n\}$, we choose

$$t = \frac{\lambda}{2}n - \mathbb{E}T$$

so that

$$\{T > \frac{\lambda}{2}n\} = \{T - \mathbb{E}T > t\}$$

If $t \leq 0$, then $\frac{\lambda}{2}n \leq \mathbb{E}T$ and the inequality (7.5) holds trivially because its right-hand side equals 1. So it suffices to consider $t > 0$. so (7.11) implies

$$\mathbb{P}\left\{T > \frac{\lambda}{2}n\right\} \leq \exp\left(-\frac{t^2}{2nK^2}\right) = \exp\left(-\frac{(\frac{\lambda}{2}n - \mathbb{E}T)^2}{2nK^2}\right).$$

□

Combining the cases $t \leq 0$ and $t > 0$ proves (7.5).

With the above preparation, we can now show that sufficiently many good coordinates show up in the coefficient vector $(Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}})_{l \in [K]}$.

LEMMA 7.5 (Good coordinates). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon < 1$. Let $0 < \alpha, \beta, \gamma, \nu < 1$, and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Let $Y_1, \dots, Y_n$ be the columns of M . Let ζ be a measurable function such that for every $(y_1, \dots, y_{n-1}) \in (\mathbb{R}^n)^{n-1}$, we have*

$$\zeta(y_1, \dots, y_{n-1}) \in S^{n-1} \cap \text{span}(y_1, \dots, y_{n-1})^\perp.$$

Let $Z = \zeta(Y_1, \dots, Y_{n-1})$. Let

$$\Theta_0 = \{l \in [n] : \frac{\nu}{\sqrt{2n}} \leq |Z_l| \leq \sqrt{\frac{2}{\alpha n}}\}$$

Let $0 < \lambda < 1$. Then for any $t \in (0, \lambda/8)$, we have

$$\begin{aligned} \mathbb{P}\left\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\right\} &\leq \mathbb{P}\{|\Theta_0| < \lambda n\} + \mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{\lambda}{2}n\right\} \\ &\quad + \exp\left(-2K\left(\frac{3}{4} - \frac{\lambda}{2}\right)^2\right) + \exp\left(-\frac{K}{2}\left(\frac{\lambda}{2} - 4t\right)^2\right). \end{aligned}$$

Proof. The first step is to consider the two good events

$$\mathcal{E}_1 = \{|\Theta_0| \geq \lambda n\}, \quad \mathcal{E}_2 = \left\{|I_{\text{heavy},n}| \leq \frac{\lambda}{2}n\right\}.$$

and then we can bound $|J_n \cap \Theta_0 \cap I_{\text{light},n}|$ under the good event $\mathcal{E}_1 \cap \mathcal{E}_2$. Formally, by union bound, we obtain

$$(7.12) \quad \mathbb{P}\left(\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\}\right) \leq \mathbb{P}(\mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_2^c) + \mathbb{P}\left(\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\} \cap \mathcal{E}_1 \cap \mathcal{E}_2\right).$$

It remains to bound the term $\mathbb{P}\left(\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\} \cap \mathcal{E}_1 \cap \mathcal{E}_2\right)$.

As is explained in the overview, the key idea to show that $\mathbb{P}\left(\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\} \cap \mathcal{E}_1 \cap \mathcal{E}_2\right)$ is small is to condition on R_1 , the first $n-1$ columns of $\bar{R}_2$, and the set $J_n \cap I_{\text{heavy},n}$, which in combination completely determines Θ_0 . To do this formally, let

$$\Xi_1 = (R_1, (X_{l,j})_{l \in [K], j \in [n-1]}, J_1, \dots, J_{n-1}),$$

which determines R_1 and the first $n-1$ columns of $\bar{R}_2$, and let

$$\Xi_2 = J_n.$$

By construction, Ξ_1 and Ξ_2 are independent. Moreover, $I_{\text{light},n}$ and $I_{\text{heavy},n}$ are functions of only Ξ_1 . We observe that the first $n-1$ columns of R_2 only depends on Ξ_1 and $J_n \cap I_{\text{heavy},n}$ because J_n can affect the first $n-1$ columns of R_2 only by causing deletions and what happens in $I_{\text{light},n}$ will not cause deletion in any case. More formally, for $i \in [n]$ and $j \in [n-1]$, we can explicitly write

$$(r_2)_{ij} = (\bar{r}_2)_{ij} \mathbf{1}_{([n] \setminus (J_n \cap I_{\text{heavy},n})) \cap \{s: \sum_{q=1}^{n-1} \bar{h}_{s,q} \leq L\}}(i).$$

Since Θ_0 depends only on the first $n-1$ columns of R_2 and therefore and only depends on $(\Xi_1, J_n \cap I_{\text{heavy},n})$, there exists a measurable map ψ such that

$$(7.13) \quad \Theta_0 = \psi(\Xi_1, J_n \cap I_{\text{heavy},n}).$$

Conditioning on R_1 and the first $n-1$ columns of $\bar{R}_2$ is relatively easy because of independence between Ξ_1 and Ξ_2 . To separate the randomness from Ξ_1 and from Ξ_2 and do conditioning, we write

$$\mathbb{1}_{\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK, |\Theta_0| \geq \lambda n, |I_{\text{heavy},n}| \leq \frac{\lambda}{2}n\}} = \phi(\Xi_1, \Xi_2)$$

for some measurable ϕ . Let $\phi_1(\xi_1) = \mathbb{E}_{\Xi_2} \phi(\xi_1, \Xi_2)$. By independence,

$$(7.14) \quad \mathbb{P}\left(|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK, |\Theta_0| \geq \lambda n, |I_{\text{heavy},n}| \leq \frac{\lambda}{2}n\right) = \mathbb{E} \phi_1(\Xi_1).$$

Now we fix (condition on) the randomness from Ξ_1 . Formally, we fix ξ_1 with $|I_{\text{heavy},n}(\xi_1)| \leq \frac{\lambda}{2}n$, because otherwise we have $|I_{\text{heavy},n}(\xi_1)| > \lambda n/2$, and then $\phi_1(\xi_1) = 0$, which reduces to a trivial case. Then $I_{\text{light},n}(\xi_1)$ and $I_{\text{heavy},n}(\xi_1)$ are deterministic. To also condition on $J_n \cap I_{\text{heavy},n}$, we define

$$\Gamma_1 = J_n \cap I_{\text{heavy},n}(\xi_1), \quad \Gamma_2 = J_n \cap I_{\text{light},n}(\xi_1).$$

By equation (7.13) (with $\Xi_1 = \xi_1$), we can write $\Theta_0 = \psi(\xi_1, \Gamma_1)$.

To ensure that there is still enough randomness after we conditioning on Ξ_1 and Γ_1 , we need to show $|\Gamma_1|$ is not too large. To separate the bad event that $|\Gamma_1|$ is large (which will happen with low probability), we write (for fixed ξ_1 with $|I_{\text{heavy},n}(\xi_1)| \leq \frac{\lambda}{2}n$)

$$\begin{aligned} \phi(\xi_1, \Xi_2) &= \mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \Gamma_2| \leq tK, |\psi(\xi_1, \Gamma_1)| \geq \lambda n\}} \\ &\leq \mathbb{1}_{\{|\Gamma_1| > \frac{3}{4}K\}} + \mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \Gamma_2| \leq tK, |\psi(\xi_1, \Gamma_1)| \geq \lambda n, |\Gamma_1| \leq \frac{3}{4}K\}}. \end{aligned}$$

Taking expectations over Ξ_2 gives

$$(7.15) \quad \phi_1(\xi_1) \leq \mathbb{P}\left(|\Gamma_1| > \frac{3}{4}K\right) + \mathbb{E}\left[\mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \Gamma_2| \leq tK, |\psi(\xi_1, \Gamma_1)| \geq \lambda n, |\Gamma_1| \leq \frac{3}{4}K\}}\right].$$

Now we will show that with high probability $|\Gamma_1|$ is not too large. And the idea is to use Hoeffding's inequality for random sampling without replacement (Proposition 3.11). We continue working with fixed ξ_1 with $|I_{\text{heavy},n}(\xi_1)| \leq \frac{\lambda}{2}n$. Since J_n is an unordered random sampling without replacement, by definition 3.13, there exists a random sequence T_1 which is a random sampling of an ordered K -element sequence uniformly from $[n]$ without replacement, such that

$$J_n(\omega) := \{T_1(\omega)(1), \dots, T_1(\omega)(k)\} = \text{Range}(T_1(\omega)).$$

Then the sequence $S_1(1), \dots, S_1(K)$ with $S_1(j) = \mathbb{1}_{I_{\text{heavy},n}(\xi_1)}(T_1(\omega)(j))$ is a random sampling of an ordered K -element sequence uniformly from population $\mathcal{S}_1$ without replacement, where

$$\mathcal{S}_1 = (s_{1,1}, \dots, s_{1,n})$$

with

$$s_{1,i} = \mathbb{1}_{I_{\text{heavy},n}(\xi_1)}(i) \in \{0, 1\}.$$

Then $\sum_{j=1}^K S_1(j)$ is the number of sampled indices in J_n that fall in $I_{\text{heavy},n}(\xi_1)$, i.e.

$$\sum_{j=1}^K S_1(j) = |J_n \cap I_{\text{heavy},n}(\xi_1)| = |\Gamma_1|.$$

Here $a = 0$, $b = 1$ in the sense of Proposition 3.11, and

$$\frac{1}{n} \sum_{i=1}^n s_{1,i} = \frac{|I_{\text{heavy},n}(\xi_1)|}{n} \leq \frac{\lambda}{2}$$

Therefore, applying Proposition 3.11 with $t = \frac{3}{4} - \frac{1}{n} \sum_{i=1}^n s_{1,i}$, we obtain

$$\begin{aligned} \mathbb{P}(|\Gamma_1| > \tfrac{3}{4}K) &= \mathbb{P}\left(\frac{1}{K} \sum_{j=1}^K S_1(j) - \frac{1}{n} \sum_{i=1}^n s_{1,i} \geq \tfrac{3}{4} - \frac{1}{n} \sum_{i=1}^n s_{1,i}\right) \\ (7.16) \quad &\leq \exp\left(-2K\left(\tfrac{3}{4} - \frac{1}{n} \sum_{i=1}^n s_{1,i}\right)^2\right) \leq \exp\left(-2K\left(\tfrac{3}{4} - \tfrac{\lambda}{2}\right)^2\right). \end{aligned}$$

Now, (after conditioning on R_1 , the first $n-1$ columns of $\bar{R}_2$, and the set $J_n \cap I_{\text{heavy},n}$) we can take advantage of the extra randomness in the choice of Γ_2 to guarantee that, with high probability, sufficiently many of elements of the set $J_n \cap I_{\text{light},n}$ fall in $\Theta_0 \cap I_{\text{light},n}$. More formally, we will show that, conditioning on the first $n-1$ columns of $\bar{R}_2$ and $J_n \cap I_{\text{heavy},n}$, the set $J_n \cap I_{\text{light},n}$ is a random uniform subset of $I_{\text{light},n}$ with cardinality $K - |J_n \cap I_{\text{heavy},n}|$. To this end, let $\mu(\theta_1, B)$ be a regular conditional probability of Γ_2 given Γ_1 (so intuitively, $\mu(\theta_1, B) = \mathbb{P}(B|\Gamma_1 = \theta_1)$), and then we have

$$\begin{aligned} (7.17) \quad &\mathbb{E}\left[\mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \Gamma_2| \leq tK, |\psi(\xi_1, \Gamma_1)| \geq \lambda n, |\Gamma_1| \leq \tfrac{3}{4}K\}} \mid \Gamma_1\right] \\ &= \mathbb{1}_{\{|\psi(\xi_1, \Gamma_1)| \geq \lambda n, |\Gamma_1| \leq \tfrac{3}{4}K\}} \int \mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \theta_2| \leq tK\}} \mu(\Gamma_1, d\theta_2). \end{aligned}$$

Fix an index set $\theta_1 \subset I_{\text{heavy},n}(\xi_1)$ with $|\theta_1| \leq \frac{3}{4}K$. By Lemma 3.15 applied with $J = J_n$ and $I = I_{\text{heavy},n}(\xi_1)$, the conditional law $\mu(\theta_1, \cdot)$ is the probability distribution of a random sampling of an unordered $(K - |\theta_1|)$ -element subset uniformly from $I_{\text{light},n}(\xi_1)$ without replacement. In other words, we have

$$\int \mathbb{1}_{\{|\psi(\xi_1, \theta_1) \cap \theta_2| \leq tK\}} \mu(\theta_1, d\theta_2) = \mathbb{P}_{\Upsilon}(\{|\psi(\xi_1, \theta_1) \cap \Upsilon| \leq tK\})$$

where Υ is the probability distribution of a random sampling of an unordered $(K - |\theta_1|)$ -element subset uniformly from $I_{\text{light},n}(\xi_1)$ without replacement, in some appropriate auxiliary probability space. Enumerate the elements in $I_{\text{light},n}(\xi_1)$ so that $I_{\text{light},n}(\xi_1) = \{s_{2,1}, \dots, s_{2,|I_{\text{light},n}(\xi_1)|}\}$.

By definition 3.13, there exists a random sequence T_2 , which is a random sampling of an ordered $(K - |\theta_1|)$ -element sequence uniformly from $|I_{\text{light},n}(\xi_1)|$ without replacement (also in the auxiliary probability space), such that

$$\Upsilon(\omega) = \{s_{2,T_2(\omega)(1)}, \dots, s_{2,T_2(\omega)(K-|\theta_1|)}\} = \text{Range}(s \circ T_2(\omega)).$$

Set

$$A = \psi(\xi_1, \theta_1) \cap I_{\text{light},n}(\xi_1).$$

For fixed ξ_1 and θ_1 with $|\psi(\xi_1, \theta_1)| \geq \lambda n$ and $|I_{\text{heavy},n}(\xi_1)| \leq \frac{\lambda}{2}n$, we have

$$(7.18) \quad |A| \geq |\psi(\xi_1, \theta_1)| - |I_{\text{heavy},n}(\xi_1)| \geq \frac{\lambda}{2}n.$$

Then the sequence $S_3(1), \dots, S_3(K - |\theta_1|)$ with $S_3(j) = \mathbb{1}_A(s_{2,T_2(\omega)(j)})$ is a random sampling of an ordered $(K - |\theta_1|)$ -element sequence uniformly from population $\mathcal{S}_3$ without replacement, where

$$\mathcal{S}_3 = (s_{3,1}, \dots, s_{3,|I_{\text{light},n}(\xi_1)|})$$

with

$$s_{3,i} = \mathbb{1}_A(s_{2,i}) \in \{0, 1\}.$$

Also, we have

$$\frac{1}{K - |\theta_1|} \sum_{j=1}^{K-|\theta_1|} S_3(j) = \frac{|A \cap \Upsilon|}{K - |\theta_1|} = \frac{|\psi(\xi_1, \theta_1) \cap \Upsilon|}{K - |\theta_1|}, \quad \mu = \frac{|A|}{|I_{\text{light},n}(\xi_1)|} \geq \frac{|A|}{n} \geq \frac{\lambda}{2},$$

where we used (7.18) and $|I_{\text{light},n}(\xi_1)| \leq n$. Moreover, $|\theta_1| \leq 3K/4$ implies $K - |\theta_1| \geq K/4$, hence $tK/(K - |\theta_1|) \leq 4t$. Thus the event $\{|\psi(\xi_1, \theta_1) \cap \Upsilon| \leq tK\}$ implies

$$\frac{1}{K - |\theta_1|} \sum_{j=1}^{K-|\theta_1|} S_3(j) \leq 4t, \quad \text{i.e.} \quad \frac{1}{K - |\theta_1|} \sum_{j=1}^{K-|\theta_1|} S_3(j) - \mu \leq -(\mu - 4t).$$

Since $t < \lambda/8$, we have $\mu - 4t \geq \lambda/2 - 4t > 0$. Applying (3.5), we obtain

$$(7.19) \quad \begin{aligned} \int \mathbb{1}_{\{|\psi(\xi_1, \theta_1) \cap \Upsilon| \leq tK\}} \mu(\theta_1, d\theta_2) &= \mathbb{P}_{\Upsilon}(\{|\psi(\xi_1, \theta_1) \cap \Upsilon| \leq tK\}) \\ &\leq \exp(-2(K - |\theta_1|)(\mu - 4t)^2) \\ &\leq \exp\left(-2 \cdot \frac{K}{4} \left(\frac{\lambda}{2} - 4t\right)^2\right) \\ &= \exp\left(-\frac{K}{2} \left(\frac{\lambda}{2} - 4t\right)^2\right). \end{aligned}$$

Inserting (7.19) into (7.17) and taking expectations gives

$$(7.20) \quad \mathbb{E} \left[\mathbb{1}_{\{|\psi(\xi_1, \Gamma_1) \cap \Gamma_2| \leq tK, |\psi(\xi_1, \Gamma_1)| \geq \lambda n, |\Gamma_1| \leq \frac{3}{4}K\}} \right] \leq \exp\left(-\frac{K}{2} \left(\frac{\lambda}{2} - 4t\right)^2\right).$$

Combining (7.15), (7.16), and (7.20), we conclude

$$\phi_1(\xi_1) \leq \exp\left(-2K\left(\frac{3}{4} - \frac{\lambda}{2}\right)^2\right) + \exp\left(-\frac{K}{2} \left(\frac{\lambda}{2} - 4t\right)^2\right).$$

This bound is uniform in ξ_1 with $|I_{\text{heavy},n}(\xi_1)| \leq \frac{\lambda}{2}n$ (and trivial otherwise since then $\phi_1(\xi_1) = 0$). Therefore, taking expectations in (7.14) gives

$$\mathbb{P}\left(|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK, |\Theta_0| \geq \lambda n, |I_{\text{heavy},n}| \leq \frac{\lambda}{2}n\right) \leq \exp\left(-2K\left(\frac{3}{4} - \frac{\lambda}{2}\right)^2\right) + \exp\left(-\frac{K}{2} \left(\frac{\lambda}{2} - 4t\right)^2\right).$$

Combining this with (7.12), we obtain the desired result. $\square$

COROLLARY 7.6 (Sufficiently many good coordinates). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon, \delta < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \overline{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \overline{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Let $Y_1, \dots, Y_n$ be the columns of M . Let ζ be a measurable function such that for every $(y_1, \dots, y_{n-1}) \in (\mathbb{R}^n)^{n-1}$, we have*

$$\zeta(y_1, \dots, y_{n-1}) \in S^{n-1} \cap \text{span}(y_1, \dots, y_{n-1})^\perp.$$

Let $Z = \zeta(Y_1, \dots, Y_{n-1})$. Let $W \in \mathbb{R}^K$ be such that $W_l = Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}$. Assume

$$n \geq \frac{2}{\gamma}, \quad \alpha < \frac{\gamma/2}{c_{6.1.1} \log\left(\frac{c_{6.1.1}\rho}{\beta(\gamma/2)^{3/2}}\right)}, \quad \nu = c_{6.1.2} \varepsilon \frac{\beta\sqrt{\gamma/2}}{\rho}.$$

If

$$L > eK, \quad K \geq \log\left(\frac{8}{\alpha}\right) + 8 \log\left(\frac{16}{\alpha\delta}\right) + \frac{128}{\alpha^2} \log\left(\frac{16}{\alpha\delta}\right),$$

$$n \geq \frac{1}{c_{7.2.1}\gamma} \log\left(\frac{16}{\alpha\delta}\right), \quad \frac{n}{K^2} \geq \frac{128}{\alpha^2} \log\left(\frac{16}{\alpha\delta}\right),$$

then

$$\mathbb{P}\left(\text{card}\left\{l \in [K] : \frac{\nu}{\sqrt{2n}} \leq |W_l| \leq \sqrt{\frac{2}{\alpha n}}\right\} \geq \frac{\alpha K}{32}\right) \geq 1 - \frac{1}{4}\alpha\delta.$$

Proof. Let

$$\Theta_0 = \{l \in [n] : \frac{\nu}{\sqrt{2n}} \leq |Z_l| \leq \sqrt{\frac{2}{\alpha n}}\}$$

Combining Lemma 7.3 and Lemma 7.2, we have

$$\mathbb{P}(|\Theta_0| < \frac{1}{2}\alpha n) \leq \exp(-c_{7.2.1}\gamma n)$$

By Lemma 7.4, we have

$$\mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{\alpha}{4}n\right\} \leq \exp\left(-\frac{\left(\frac{\alpha}{4}n - ne^{-K}(eK/L)^L\right)_+^2}{2nK^2}\right)$$

We observe that $(eK/L)^L < 1$ if we require $L > eK$. In this case, if we also require $K > -\log(\frac{\alpha}{8})$, then we have

$$\mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{\alpha}{4}n\right\} \leq \exp\left(-\frac{\left(\frac{\alpha}{8}n\right)^2}{2nK^2}\right) = \exp\left(-\frac{\alpha^2}{128K^2}n\right)$$

By Lemma 7.5, for any $t \in (0, \lambda/8)$, we have

$$(7.21) \quad \mathbb{P}\left\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq tK\right\} \leq \mathbb{P}\{|\Theta_0| < \lambda n\} + \mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{\lambda}{2}n\right\}$$

$$(7.22) \quad + \exp\left(-2K\left(\frac{3}{4} - \frac{\lambda}{2}\right)^2\right) + \exp\left(-\frac{K}{2}\left(\frac{\lambda}{2} - 4t\right)^2\right).$$

Choosing $\lambda = \frac{1}{2}\alpha < 1$ and $t = \frac{\lambda}{16} = \frac{1}{32}\alpha$, we have

$$(7.23) \quad \mathbb{P}\left\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq \frac{1}{32}\alpha K\right\} \leq \mathbb{P}\{|\Theta_0| < \frac{1}{2}\alpha n\} + \mathbb{P}\left\{|I_{\text{heavy},n}| > \frac{1}{4}\alpha n\right\}$$

$$(7.24) \quad + \exp\left(-\frac{1}{8}K\right) + \exp\left(-\frac{K}{2}\left(\frac{1}{8}\alpha\right)^2\right)$$

$$(7.25) \quad \leq \exp(-c_{7.2}n/2) + \exp\left(-\frac{\alpha^2}{128K^2}n\right)$$

$$(7.26) \quad + \exp\left(-\frac{1}{8}K\right) + \exp\left(-\frac{K}{2}\left(\frac{1}{8}\alpha\right)^2\right)$$

Therefore, if we require $n \geq \frac{1}{c_{7.2.1}\gamma} \log(16/(\alpha\delta))$, $\frac{n}{K^2} \geq \frac{128}{\alpha^2} \log(16/(\alpha\delta))$, $K \geq \log(\frac{8}{\alpha}) + 8 \log(16/(\alpha\delta)) + \frac{128}{\alpha^2} \log(16/(\alpha\delta))$, $L > eK$, then each of the four terms above is at most $\alpha\delta/16$. So we conclude

$$\mathbb{P}\left\{|J_n \cap \Theta_0 \cap I_{\text{light},n}| \leq \frac{1}{32}\alpha K\right\} \leq \frac{1}{4}\alpha\delta.$$

Since

$$\text{card}\left\{l \in [K]: \frac{\nu}{\sqrt{2n}} \leq |W_l| \leq \sqrt{\frac{2}{\alpha n}}\right\} = |J_n \cap \Theta_0 \cap I_{\text{light},n}|,$$

we obtain

$$\mathbb{P}\left(\text{card}\left\{l \in [K]: \frac{\nu}{\sqrt{2n}} \leq |W_l| \leq \sqrt{\frac{2}{\alpha n}}\right\} \geq \frac{\alpha K}{32}\right) \geq 1 - \frac{1}{4}\alpha\delta. \quad \square$$

Next, we prove a small ball probability result that will be used after we condition on (Ξ_1, Ξ_2) .

LEMMA 7.7 (Small ball probability). *Fix integers $n, K \geq 1$. Let $w = (w_1, \dots, w_K) \in \mathbb{R}^K$ be deterministic. Fix parameters $\alpha > 0$ and $\nu > 0$ and define the index set*

$$(7.27) \quad \Theta = \left\{l \in [K]: \frac{\nu}{\sqrt{2n}} \leq |w_l| \leq \sqrt{\frac{2}{\alpha n}}\right\}.$$

Assume that $\Theta \neq \emptyset$. Let $(X_{l,n})_{l \in \Theta}$ be independent random signs (Rademacher variables), i.e.

$$(7.28) \quad \mathbb{P}\{X_{l,n} = 1\} = \mathbb{P}\{X_{l,n} = -1\} = \frac{1}{2} \quad (l \in \Theta).$$

Fix deterministic constants $\varepsilon > 0$ and $L > 0$. Then there exists a constant $c_{7.7.1} > 0$ such that for every $t \geq \varepsilon \nu / L$,

$$\mathcal{L}\left(\frac{\varepsilon}{L} \sum_{l \in \Theta} w_l X_{l,n}, \frac{t}{\sqrt{n}}\right) \leq \frac{c_{7.7.1} L t}{\varepsilon \nu \sqrt{|\Theta|}}.$$

Proof. Let

$$T = \sum_{l \in \Theta} w_l X_{l,n}, \quad r = \frac{L t}{\varepsilon \sqrt{n}}.$$

By Lemma 3.6 we have

$$(7.29) \quad \mathcal{L}\left(\frac{\varepsilon}{L} \sum_{l \in \Theta} w_l X_{l,n}, \frac{t}{\sqrt{n}}\right) = \mathcal{L}\left(\frac{\varepsilon T}{L}, \frac{t}{\sqrt{n}}\right) = \mathcal{L}(T, r).$$

Let $m = \text{card}(\Theta)$ and enumerate $\Theta = \{l_1, \dots, l_m\}$. Also, let

$$\zeta_i = w_{l_i} X_{l_i,n}, \quad i = 1, \dots, m.$$

Then $T = \sum_{i=1}^m \zeta_i$. Since the $X_{l,n}$ are independent, the ζ_i are independent as well.

$$s := \min\left\{r, \frac{\nu}{2\sqrt{2n}}\right\}.$$

For every $i \in [m]$, the definition of Θ gives

$$|w_{l_i}| \geq \frac{\nu}{\sqrt{2n}},$$

so $s \leq |w_{l_i}|/2$. Since ζ_i takes the two values $\pm w_{l_i}$ with probability 1/2, any interval of radius s can contain at most one atom of ζ_i . Therefore

$$\mathcal{L}(\zeta_i, s) = \frac{1}{2}, \quad i \in [m].$$

Applying Lemma 3.7 with $r_i = s$ for all i , there is an absolute constant $c_1 > 0$ such that

$$\begin{aligned}\mathcal{L}(T, r) &\leq \frac{c_1 r}{\sqrt{\sum_{i=1}^m (1 - \mathcal{L}(\zeta_i, s)) s^2}} \\ &= \frac{c_1 r}{\sqrt{m s^2/2}} = \frac{\sqrt{2} c_1}{\sqrt{m}} \frac{r}{s}.\end{aligned}$$

By the definition of s , we have

$$\frac{r}{s} = \max \left\{ 1, \frac{2\sqrt{2n}}{\nu} r \right\} = \max \left\{ 1, \frac{2\sqrt{2} Lt}{\varepsilon \nu} \right\}.$$

Since $t \geq \varepsilon \nu / L$, we have

$$\max \left\{ 1, \frac{2\sqrt{2} Lt}{\varepsilon \nu} \right\} \leq \frac{(1 + 2\sqrt{2}) Lt}{\varepsilon \nu}.$$

Combining this with (7.29), we conclude that

$$\mathcal{L} \left(\frac{\varepsilon}{L} \sum_{l \in \Theta} w_l X_{l,n}, \frac{t}{\sqrt{n}} \right) \leq \frac{\sqrt{2}(1 + 2\sqrt{2}) c_1 Lt}{\varepsilon \nu \sqrt{m}}.$$

□

With all the above preparation, we can now implement the conditioning argument to bound $\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2$.

THEOREM 7.8 (Invertibility for incompressible vectors). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon < 1$ and $0 < \delta < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \overline{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \overline{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Then there exist constants $c_{7.8.1}, c_{7.8.2}, c_{7.8.3}$ such that for any*

$$\begin{aligned}\alpha &< \frac{\gamma/2}{c_{6.1.1} \log \left(\frac{c_{6.1.1} \rho}{\beta(\gamma/2)^{3/2}} \right)}, \quad \nu = c_{6.1.2} \varepsilon \frac{\beta \sqrt{\gamma/2}}{\rho}, \\ K &= \left\lceil \frac{c_{7.8.2}}{\delta^2 \alpha^3} \right\rceil, \quad L = \lceil 2eK \rceil,\end{aligned}$$

and

$$n \geq \max \left\{ \frac{c_{7.8.1}}{\delta^5 \alpha^9}, \frac{1}{c_{7.2.1} \gamma} \log \left(\frac{16}{\alpha \delta} \right), \frac{2}{\gamma} \right\},$$

we have

$$\mathbb{P} \left(\inf_{v \in \text{Incomp}(\alpha, \nu)} \|Mv\|_2 \geq c_{7.8.3} \frac{\beta^2 \gamma}{\rho^2} \varepsilon^3 \delta^2 \alpha^3 n^{-1} \right) \geq 1 - \delta/2.$$

Proof. As in the statement of Corollary 7.6, let $Y_1, \dots, Y_n$ be the columns of M . Let ζ be a measurable function such that for every $(y_1, \dots, y_{n-1}) \in (\mathbb{R}^n)^{n-1}$, we have

$$\zeta(y_1, \dots, y_{n-1}) \in S^{n-1} \cap \text{span}(y_1, \dots, y_{n-1})^\perp.$$

Let $Z = \zeta(Y_1, \dots, Y_{n-1})$. By assumption we have $0 < \nu < 1$. Let $Z = \zeta(Y_1, \dots, Y_{n-1})$. Let $W \in \mathbb{R}^K$ be such that $W_l = Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}$. Consider the event

$$\text{card}(\{l \in [K] : \frac{\nu}{\sqrt{2n}} \leq |W_l| \leq \sqrt{\frac{2}{\alpha n}}\}) \geq \frac{\alpha}{32} K$$

Since this event is completely determined by (Ξ_1, Ξ_2) , we can represent this event as

$$\{\omega : (\Xi_1(\omega), \Xi_2(\omega)) \in \mathcal{E}\}$$

for some measurable set $\mathcal{E}$ in the corresponding measurable space where $(\Xi_1(\omega), \Xi_2(\omega))$ takes values.

By Corollary 7.6, if $n \geq \frac{1}{c_{7.2.1}\gamma} \log(16/(\alpha\delta))$, $\frac{n}{K^2} \geq \frac{128}{\alpha^2} \log(16/(\alpha\delta))$, $K \geq \log(\frac{8}{\alpha}) + 8 \log(16/(\alpha\delta)) + \frac{128}{\alpha^2} \log(16/(\alpha\delta))$, $L > eK$, and $n \geq 2/\gamma$, then we have

$$\begin{aligned} & \mathbb{P}(\text{card}(\{l \in [K] : \frac{\nu}{\sqrt{2n}} \leq |W_l| \leq \sqrt{\frac{2}{\alpha n}}\}) \geq \frac{\alpha}{32}K) \\ &= \mathbb{P}(\{\omega : (\Xi_1(\omega), \Xi_2(\omega)) \in \mathcal{E}\}) \\ &\geq 1 - \frac{1}{4}\delta\alpha \end{aligned}$$

We will verify these requirements below. Assuming that they hold for the moment, we have

$$\begin{aligned} & \mathbb{P}(|\langle Y_n, Z \rangle| < \frac{t}{\sqrt{n}}) \\ &\leq \mathbb{E} \mathbb{1}_{\{|\langle Y_n, Z \rangle| < \frac{t}{\sqrt{n}}\}}(\omega) \mathbb{1}_{\{(\Xi_1(\omega), \Xi_2(\omega)) \in \mathcal{E}\}}(\omega) + \mathbb{P}((\Xi_1(\omega), \Xi_2(\omega)) \in \mathcal{E}^C) \\ &= \mathbb{E}_{\Xi_1, \Xi_2} \mathbb{E}_{\Xi_3} \mathbb{1}_{\{|\langle Y_n, Z \rangle| < \frac{t}{\sqrt{n}}\}}(\omega) \mathbb{1}_{\{(\Xi_1(\omega), \Xi_2(\omega)) \in \mathcal{E}\}}(\omega) + \frac{1}{4}\delta\alpha \\ &= \mathbb{E}_{\Xi_1, \Xi_2} [\mathbb{1}_{\mathcal{E}}(\Xi_1, \Xi_2) \mathbb{E}_{\Xi_3} \mathbb{1}_{\{|\langle Y_n, Z \rangle| < \frac{t}{\sqrt{n}}\}}(\omega)] + \frac{1}{4}\delta\alpha \\ &= \mathbb{E}_{\Xi_1, \Xi_2} [\mathbb{1}_{\mathcal{E}}(\Xi_1, \Xi_2) \mathbb{P}_{\Xi_3}(|\langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}| < \frac{t}{\sqrt{n}})] + \frac{1}{4}\delta\alpha \end{aligned}$$

For fixed $(\xi_1, \xi_2) \in \mathcal{E}$, define

$$\Theta = \left\{ l \in [K] : \frac{\nu}{\sqrt{2n}} \leq |Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}| \leq \sqrt{\frac{2}{\alpha n}} \right\}.$$

Then $|\Theta| \geq \alpha K/32$. After conditioning further on the signs $(X_{l,n})_{l \notin \Theta}$, the contribution of the indices outside Θ is a deterministic shift and the signs $(X_{l,n})_{l \in \Theta}$ remain independent. Therefore, for fixed $(\xi_1, \xi_2) \in \mathcal{E}$, we have

$$\begin{aligned} & \mathbb{P}_{\Xi_3} \left(\left| \langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}} \right| < \frac{t}{\sqrt{n}} \right) \\ &\leq \mathcal{L}_{(X_{l,n})_{l \in \Theta}} \left(\frac{\varepsilon}{L} \sum_{l \in \Theta} X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}, \frac{t}{\sqrt{n}} \right) \\ &\leq \frac{c_{7.7.1} L t}{\varepsilon \nu \sqrt{\frac{\alpha}{32} K}}, \end{aligned}$$

where the last line follows from Lemma 7.7 for $t \geq \varepsilon \nu / L$.

Therefore, by iterated expectation, we have

$$\begin{aligned} & \mathbb{E}_{\Xi_1, \Xi_2} [\mathbb{1}_{\mathcal{E}}(\Xi_1, \Xi_2) \mathbb{P}_{\Xi_3}(|\langle \text{Col}_n(A + \varepsilon R_1), Z \rangle + \frac{\varepsilon}{L} \sum_{l=1}^K X_{l,n} Z_{j(n,l)} \mathbf{1}_{\{j(n,l) \in I_{\text{light},n}\}}| < \frac{t}{\sqrt{n}})] \\ &\leq \mathbb{E}_{\Xi_1, \Xi_2} \left[\mathbb{1}_{\mathcal{E}}(\Xi_1, \Xi_2) \frac{c_{7.7.1} L t}{\varepsilon \nu \sqrt{\frac{\alpha}{32} K}} \right] \\ &\leq \frac{c_{7.7.1} L t}{\varepsilon \nu \sqrt{\frac{\alpha}{32} K}}. \end{aligned}$$

and therefore we have

$$\mathbb{P}\left(|\langle Y_n, Z \rangle| < \frac{t}{\sqrt{n}}\right) \leq \frac{c_{7.7.1}Lt}{\varepsilon \nu \sqrt{\frac{\alpha}{32}K}} + \frac{1}{4}\delta\alpha$$

for $t \geq \varepsilon \nu / L$.

If we also require

$$\frac{c_{7.7.1}Lt}{\varepsilon \nu \sqrt{\frac{\alpha}{32}K}} \leq \frac{1}{4}\delta\alpha,$$

then the same argument, with column k in place of column n , gives the same bound for every $k \in [n]$. Applying Lemma 7.1, we have

$$\begin{aligned} \mathbb{P}\left(\inf_{x \in \text{Incomp}(\alpha, \nu)} \|Mx\|_2 < t\nu n^{-1}\right) &\leq \frac{1}{\alpha n} \sum_{k=1}^n \mathbb{P}\left(\text{dist}(Y_k, H_k) < tn^{-1/2}\right) \\ &\leq \frac{1}{\alpha n} n \left(\frac{1}{4}\delta\alpha + \frac{1}{4}\delta\alpha\right) \\ &= \frac{1}{2}\delta. \end{aligned}$$

In summary, the requirements are

1. $L > eK$.
2. $K \geq \log\left(\frac{8}{\alpha}\right) + 8\log\left(\frac{16}{\alpha\delta}\right) + \frac{128}{\alpha^2}\log\left(\frac{16}{\alpha\delta}\right)$.
3. $n \geq \frac{1}{c_{7.2.1}\gamma}\log\left(\frac{16}{\alpha\delta}\right)$ and $n \geq 2/\gamma$.
4. $\frac{n}{K^2} \geq \frac{128}{\alpha^2}\log\left(\frac{16}{\alpha\delta}\right)$.
5. $\frac{c_{7.7.1}Lt}{\varepsilon \nu \sqrt{\frac{\alpha}{32}K}} \leq \frac{1}{4}\delta\alpha$.
6. $t \geq \varepsilon \nu / L$.

To satisfy all the requirements, we choose

$$K = \left\lceil \frac{c_1}{\delta^2\alpha^3} \right\rceil, \quad L = \lceil 2eK \rceil, \quad t = \frac{\varepsilon \nu}{L},$$

where $c_1 > 0$ is a sufficiently large absolute constant, to be chosen below. Note that the requirement (6) is already satisfied by the choice of t .

We claim that with this choice, if

$$n \geq \frac{1}{c_{7.2.1}\gamma}\log\left(\frac{16}{\alpha\delta}\right), \quad n \geq \frac{2}{\gamma}, \quad \text{and} \quad n \geq \frac{8192c_1^2}{\delta^5\alpha^9},$$

then all the requirements are satisfied. Since $L = \lceil 2eK \rceil$, we have $L > eK$. Therefore (1) holds.

We next verify the requirement (2). Since $0 < \alpha \leq 1$ and $0 < \delta < 1$, we have

$$\log\left(\frac{16}{\alpha\delta}\right) \leq \frac{16}{\alpha\delta} \leq \frac{16}{\alpha^3\delta^2}.$$

which implies

$$8\log\left(\frac{16}{\alpha\delta}\right) \leq \frac{128}{\alpha^3\delta^2}$$

and

$$\frac{128}{\alpha^2} \log\left(\frac{16}{\alpha\delta}\right) \leq \frac{128}{\alpha^2} \cdot \frac{16}{\alpha\delta} \leq \frac{2048}{\alpha^3\delta^2}.$$

Also, since $0 < \alpha, \delta \leq 1$, we have

$$\log\left(\frac{8}{\alpha}\right) \leq \frac{8}{\alpha} \leq \frac{64}{\alpha^3\delta^2}.$$

Combining these bounds, we obtain

$$\begin{aligned} \log\left(\frac{8}{\alpha}\right) + 8 \log\left(\frac{16}{\alpha\delta}\right) + \frac{128}{\alpha^2} \log\left(\frac{16}{\alpha\delta}\right) &\leq \frac{64}{\delta^2\alpha^3} + \frac{128}{\delta^2\alpha^3} + \frac{2048}{\delta^2\alpha^3} \\ &= \frac{2240}{\delta^2\alpha^3}. \end{aligned}$$

If $c_1 \geq 2240$, then we have

$$K \geq \log\left(\frac{8}{\alpha}\right) + 8 \log\left(\frac{16}{\alpha\delta}\right) + \frac{128}{\alpha^2} \log\left(\frac{16}{\alpha\delta}\right)$$

and therefore requirement (2) is satisfied.

Next if $c_1 \geq 512c_{7.7,1}^2$, then we have

$$\begin{aligned} \frac{c_{7.7,1}Lt}{\varepsilon\nu\sqrt{\frac{\alpha}{32}K}} &= \frac{c_{7.7,1}}{\sqrt{\frac{\alpha}{32}K}} \\ &\leq \frac{c_{7.7,1}\sqrt{32}}{\sqrt{c_1}}\alpha\delta \\ &\leq \frac{\alpha\delta}{4} \end{aligned}$$

so the requirement (5) is satisfied.

To guarantee requirement (4), it is enough to impose the condition

$$\frac{n}{K^2} \geq \frac{128}{\alpha^2} \cdot \frac{16}{\alpha\delta} = \frac{2048}{\alpha^3\delta},$$

because $\log(16/(\alpha\delta)) \leq 16/(\alpha\delta)$. Since

$$K = \left\lceil \frac{c_1}{\delta^2\alpha^3} \right\rceil,$$

and

$$\frac{c_1}{\delta^2\alpha^3} \geq 1,$$

we have

$$K \leq \frac{2c_1}{\delta^2\alpha^3}.$$

Therefore it is sufficient to require

$$n \geq \frac{2048}{\alpha^3\delta} \left(\frac{2c_1}{\delta^2\alpha^3} \right)^2 = \frac{8192c_1^2}{\delta^5\alpha^9}.$$

This same bound also guarantees $K \leq n$, so the random K -subsets of $[n]$ are well-defined.

Therefore, if

$$n \geq \max \left\{ \frac{1}{c_{7.2,1}\gamma} \log\left(\frac{16}{\alpha\delta}\right), \frac{2}{\gamma}, \frac{8192c_1^2}{\delta^5\alpha^9} \right\},$$

and if we choose

$$K = \left\lceil \frac{c_1}{\delta^2\alpha^3} \right\rceil, \quad L = \lceil 2eK \rceil, \quad t = \frac{\varepsilon\nu}{L},$$

with $c_1 \geq \max\{2240, 512c_{7.7,1}^2, 1\}$, then all the required constraints are satisfied.

Next, we eliminate L and obtain a lower bound for our choice of t in terms of $\varepsilon, \delta, \nu, \alpha$. Since $0 < \delta, \alpha < 1$, we have

$$\frac{c_1}{\delta^2\alpha^3} \geq 1,$$

and therefore

$$L = \lceil 2eK \rceil \leq \frac{8ec_1}{\delta^2\alpha^3}$$

which implies

$$\frac{1}{L} \geq \frac{\delta^2\alpha^3}{8ec_1}.$$

Substituting into the formula for t , we obtain

$$\begin{aligned} t &= \frac{\varepsilon\nu}{L} \\ &\geq \frac{\varepsilon\nu\delta^2\alpha^3}{8ec_1} \end{aligned}$$

which implies

$$t \geq c_2 \varepsilon \delta^2 \nu \alpha^3$$

where $c_2 = \frac{1}{8ec_1}$.

In summary, we have

$$\begin{aligned} t\nu &\geq c_2 \varepsilon \delta^2 \nu^2 \alpha^3 \\ &= \frac{c_{6.1.2}^2}{16ec_1} \frac{\beta^2\gamma}{\rho^2} \varepsilon^3 \delta^2 \alpha^3. \end{aligned}$$

□

8 Invertibility on the whole sphere and condition number Combining the results on invertibility for compressible and incompressible vectors, we can obtain the following conclusion about invertibility on the whole sphere.

THEOREM 8.1 (Bound for smallest singular value). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon, \delta < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Then there exist constants $c_{8.1.1}, c_{8.1.2}, c_{8.1.3}, c_{8.1.4}, c_{8.1.5}$ such that for any*

$$\alpha < \frac{\gamma/2}{c_{8.1.1} \log\left(\frac{c_{8.1.1}\rho}{\beta(\gamma/2)^{3/2}}\right)}, \quad \nu = c_{8.1.2} \varepsilon \frac{\beta\sqrt{\gamma/2}}{\rho},$$

$$K = \left\lceil \frac{c_{8.1.3}}{\delta^2\alpha^3} \right\rceil, \quad L = \lceil 2eK \rceil,$$

and

$$n \geq c_{8.1.4} \max\left\{\frac{1}{\delta^5\alpha^9}, \frac{1}{\gamma} \log\left(\frac{16}{\alpha\delta}\right), \frac{2}{\gamma}\right\},$$

we have

$$\mathbb{P}\left(s_n(M) \geq c_{8.1.5} \frac{\beta^2\gamma}{\rho^2} \varepsilon^3 \delta^2 \alpha^3 n^{-1}\right) \geq 1 - \delta$$

Proof. This result now follows directly from combining Theorem 6.1 and Theorem 7.8. □

THEOREM 8.2 (Bound for condition number). *Let A be an $n \times n$ deterministic matrix such that $\|A\| = 1$. Let $0 < \varepsilon, \delta < 1$. Let $0 < \alpha, \beta, \gamma < 1$ and $\rho \geq 1$. Let V be an n dimensional pattern matrix with parameters $\alpha, \beta, \gamma, \rho$. Let K and L with $L > K$ be positive integers. Let $(R_1, \{J_i\}_{i \in [n]}, \bar{H}, H, \{X_{l,i}\}_{i \in [n], l \in [K]}, \bar{R}_2, R_2)$ be a n -dimensional oblivious perturbation model with pattern matrix V and parameters ρ, K , and L . Let $M = A + \varepsilon(R_1 + R_2)$. Then there exist constants $c_{8.2.1}, c_{8.2.2}, c_{8.2.3}, c_{8.2.4}, c_{8.2.5}$ such that for any*

$$\alpha < \frac{\gamma/2}{c_{8.2.1} \log\left(\frac{c_{8.2.1}\rho}{\beta(\gamma/2)^{3/2}}\right)}, \quad \nu = c_{8.2.2} \varepsilon \frac{\beta\sqrt{\gamma/2}}{\rho},$$

$$K = \left\lceil \frac{c_{8.2.3}}{\delta^2 \alpha^3} \right\rceil, \quad L = \lceil 2eK \rceil,$$

and

$$n \geq c_{8.2.4} \max \left\{ \frac{1}{\delta^5 \alpha^9}, \frac{1}{\gamma} \log \left(\frac{16}{\alpha \delta} \right), \frac{2}{\gamma} \right\},$$

we have

$$\mathbb{P} \left(\kappa(M) \leq \frac{c_{8.2.5} n \rho^2}{\beta^2 \gamma \varepsilon^3 \delta^2 \alpha^3} \right) \geq 1 - \delta$$

Proof. Recall that we wish to show that

$$\mathbb{P}(\kappa(M) \leq C(\varepsilon, \delta)n) \geq 1 - \delta.$$

where $M = A + \varepsilon(R_1 + R_2)$ for a deterministic matrix A with $\|A\| \leq 1$.

By Remark 4.4, we have $s_1(M) \leq \|A\| + \varepsilon\|R_1\| + \varepsilon\|R_2\| \leq 3$, so it is sufficient to give a lower bound for

$$s_n(M) = \inf_{x \in \mathbb{S}^{n-1}} \|Mx\|_2$$

□

which follows from Theorem 8.1.

Proof of Theorem 2.1. The claim about the smallest singular value follows directly from Theorem 8.1, and we can assume $\|R\| \leq 1$ by working with $\varepsilon/2$ instead of ε . More precisely, if we set $R = \frac{R_1 + R_2}{2}$ then $\|R\| \leq 1$ by Remark 4.4 and then $A + \varepsilon R = A + (\varepsilon/2)(R_1 + R_2)$. So we only need to calculate the number of random bits needed to generate the corresponding random perturbation model.

To calculate the number of random bits needed to construct R_2 , recall that

$$R_1 = \frac{1}{\rho\sqrt{n}} \cdot D_1 V D_2$$

for $D_1 = \text{diag}(\eta_{1,1}, \dots, \eta_{1,n})$ and $D_2 = \text{diag}(\eta_{2,1}, \dots, \eta_{2,n})$ where $\eta_{1,1}, \dots, \eta_{1,n}$ and $\eta_{2,1}, \dots, \eta_{2,n}$ are i.i.d. Rademacher random variables, and V is a deterministic pattern matrix. The number of random bits to generate D_1 and D_2 are $O(n)$. Also, by Corollary 5.3, the number of random bits needed to generate V is $O((\log n)^2)$. So, in summary, the number of random bits needed to generate R_1 is $O(n)$.

To calculate the number of random bits needed to construct R_2 , let $\mathcal{Z}$ be the set of injections from $[k]$ to $[n]$, and we observe that there is a bijection Φ from $\mathcal{X}_{n,k} := [n] \times [n-1] \times \dots \times [n-k+1]$ to $\mathcal{Z}$. Let $U = (U_1, \dots, U_k)$ be a random variable with values in $\mathcal{X}_{n,k}$ such that the coordinates are independent and satisfy

$$U_1 \sim \text{Uniform}([n]), \quad U_2 \sim \text{Uniform}([n-1]), \quad \dots, \quad U_k \sim \text{Uniform}([n-k+1]).$$

Therefore, by [10, Theorem 5.11.2], generating each U_i takes $O(\log(n))$ random bits. Define

$$T := \Phi(U).$$

Let

$$\mathcal{S} = (s_1, \dots, s_n)$$

be a finite ordered sequence with pairwise distinct entries, and define

$$J = \text{Range}(s \circ T) = \{s_{T(1)}, \dots, s_{T(k)}\}.$$

Then by Definition 3.9, the random sequence T is a random sampling of an ordered k -element sequence uniformly from $[n]$ without replacement, and therefore, by Definition 3.13, the random set J is a random sampling of an unordered k -element subset uniformly from $\mathcal{J} = \{s_1, \dots, s_n\}$ without replacement. As a result, we can generate J from $O(K \log n)$ random bits. Since $K = \left\lceil \frac{c_{8.2.3}}{\delta^2 \alpha^3} \right\rceil = O\left(\frac{1}{\delta^2}\right)$, the total number of random bits need to generate R_2 is $O\left(\frac{n \log n}{\delta^2}\right)$. □

Acknowledgments. The authors used ChatGPT for literature searches, drafting initial TikZ codes for figures based on their specific instructions and hand-drawn sketches, limited text editing on author written materials, and checking the full manuscript for potential errors. The authors assume responsibility for all content.

References

- [1] N. AMSEL, Y. BAUMANN, P. BECKMAN, P. BÜRGISSER, C. CAMAÑO, T. CHEN, E. CHOW, A. DAMLE, M. DEREZINSKI, M. EMBREE, ET AL., *Linear systems and eigenvalue problems: Open questions from a simons workshop*, arXiv preprint arXiv:2602.05394, (2026).
- [2] D. F. ANDERSON, T. SEPPÄLÄINEN, AND B. VALKÓ, *Introduction to probability*, Cambridge University Press, 2017.
- [3] J. BANKS, J. GARZA-VARGAS, A. KULKARNI, AND N. SRIVASTAVA, *Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time*, Foundations of computational mathematics, 23 (2023), pp. 1959–2047.
- [4] A. BASAK AND M. RUDELSON, *Invertibility of sparse non-hermitian matrices*, Advances in Mathematics, 310 (2017), pp. 426–483, <https://doi.org/https://doi.org/10.1016/j.aim.2017.02.009>.
- [5] A. BASAK AND M. RUDELSON, *The circular law for sparse non-hermitian matrices*, The Annals of Probability, 47 (2019), pp. 2359–2416.
- [6] S. BOUCHERON, G. LUGOSI, AND P. MASSART, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.
- [7] P. BÜRGISSER AND F. CUCKER, *Condition: The geometry of numerical algorithms*, vol. 349, Springer Science & Business Media, 2013.
- [8] T. CHEN, E. N. EPPERLY, R. A. MEYER, C. MUSCO, AND A. RAO, *Does block size matter in randomized block krylov low-rank approximation?*, in Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2026, pp. 1026–1046.
- [9] D. CIFUENTES AND A. MOITRA, *Polynomial time guarantees for the burer-monteiro method*, Advances in Neural Information Processing Systems, 35 (2022), pp. 23923–23935.
- [10] T. M. COVER AND J. A. THOMAS, *Elements of Information Theory*, John Wiley & Sons, 2 ed., 2006.
- [11] C. DASKALAKIS, N. GOLOWICH, N. HAGHTALAB, AND A. SHETTY, *Smooth nash equilibria: Algorithms and complexity*, in 15th Innovations in Theoretical Computer Science Conference (ITCS 2024), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024, pp. 37–1.
- [12] E. B. DAVIES, *Approximate diagonalization*, SIAM journal on matrix analysis and applications, 29 (2008), pp. 1051–1064.
- [13] M. DEREZIŃSKI, Y. NAKATSUKASA, AND E. REBROVA, *Towards universal convergence of backward error in linear system solvers*, arXiv preprint arXiv:2604.16075, (2026).
- [14] D. L. DONOHO AND P. B. STARK, *Uncertainty principles and signal recovery*, SIAM Journal on Applied Mathematics, 49 (1989), pp. 906–931.
- [15] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, PA, USA, 1997.
- [16] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, USA, second ed., 2002.
- [17] W. HOEFFDING, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, 58 (1963), pp. 13–30, <https://doi.org/10.1080/01621459.1963.10500830>.
- [18] A. T. KALAI, A. SAMORODNITSKY, AND S.-H. TENG, *Learning and smoothed analysis*, in 2009 50th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2009, pp. 395–404.

- [19] A. E. LITVAK, A. PAJOR, M. RUDELSON, AND N. TOMCZAK-JAEGERMANN, *Smallest singular value of random matrices and geometry of random polytopes*, *Advances in Mathematics*, 195 (2005), pp. 491–523.
- [20] S. LOVETT AND S. SODIN, *Almost euclidean sections of the N -dimensional cross-polytope using $O(n)$ random bits*, *Communications in Contemporary Mathematics*, 10 (2008), pp. 477–489, <https://doi.org/10.1142/S0219199708002879>.
- [21] R. MEYER, C. MUSCO, AND C. MUSCO, *On the unreasonable effectiveness of single vector krylov methods for low-rank approximation*, in *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SIAM, 2024, pp. 811–845.
- [22] N. MINAMI, *Local fluctuation of the spectrum of a multidimensional anderson tight binding model*, *Communications in mathematical physics*, 177 (1996), pp. 709–725.
- [23] R. PENG AND S. S. VEMPALA, *Solving sparse linear systems faster than matrix multiplication*, *Journal of the ACM*, 72 (2025), pp. 1–72.
- [24] A. RAKHLIN, K. SRIDHARAN, AND A. TEWARI, *Online learning: Stochastic, constrained, and smoothed adversaries*, *Advances in neural information processing systems*, 24 (2011).
- [25] J. RENEGAR, *Incorporating condition measures into the complexity theory of linear programming*, *SIAM Journal on Optimization*, 5 (1995), pp. 506–524.
- [26] B. A. ROGOZIN, *On the increase of dispersion of sums of independent random variables*, *Theory of Probability & Its Applications*, 6 (1961), pp. 97–99, <https://doi.org/10.1137/1106010>, <https://doi.org/10.1137/1106010>, <https://arxiv.org/abs/https://doi.org/10.1137/1106010>.
- [27] M. RUDELSON AND K. TIKHOMIROV, *The sparse circular law under minimal assumptions*, *Geometric and Functional Analysis*, 29 (2019), pp. 561–637.
- [28] M. RUDELSON AND R. VERSHYNIN, *The littlewood–offord problem and invertibility of random matrices*, *Advances in Mathematics*, 218 (2008), pp. 600–633.
- [29] Y. SAAD, *Iterative methods for sparse linear systems*, SIAM, 2003.
- [30] A. SANKAR, D. A. SPIELMAN, AND S.-H. TENG, *Smoothed analysis of the condition numbers and growth factors of matrices*, *SIAM Journal on Matrix Analysis and Applications*, 28 (2006), pp. 446–476.
- [31] R. SHAH AND S. SILWAL, *Smoothed analysis of the condition number under low-rank perturbations*, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, pp. 40–1.
- [32] R. SHAH, N. SRIVASTAVA, AND E. ZENG, *Sparse pseudospectral shattering*, *arXiv preprint arXiv:2411.19926*, (2024).
- [33] D. A. SPIELMAN AND S.-H. TENG, *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*, *Journal of the ACM (JACM)*, 51 (2004), pp. 385–463.
- [34] T. TAO, *An uncertainty principle for cyclic groups of prime order*, *Mathematical Research Letters*, 12 (2005), pp. 121–127.
- [35] T. TAO AND V. VU, *Random matrices: the circular law*, *Communications in Contemporary Mathematics*, 10 (2008), pp. 261–307.
- [36] S. P. VADHAN, *Pseudorandomness*, *Foundations and Trends® in Theoretical Computer Science*, 7 (2012), pp. 1–336, <https://doi.org/10.1561/04000000010>.
- [37] R. VERSHYNIN, *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press, 2018.

- [38] J. VON NEUMANN AND H. GOLDSTINE, *Numerical inverting of matrices of high order*, Bulletin of the American Mathematical Society, 53 (1947), pp. 1021–1099.
- [39] J. H. WILKINSON, *Error analysis of floating-point computation*, Numerische Mathematik, 2 (1960), pp. 319–340.
- [40] J. H. WILKINSON, *Error analysis of direct methods of matrix inversion*, Journal of the ACM (JACM), 8 (1961), pp. 281–330.