# CONCENTRATION OF DISTANCES IN WIGNER MATRICES

HOI H. NGUYEN

Abstract. It is well-known that distances in random iid matrices are highly concentrated around their mean. In this note we extend this concentration phenomenon to Wigner matrices. Exponential bounds for the lower tail are also included.

## 1. Introduction

Let $\xi$ be a real random variable of mean zero, variance one, and there exists a parameter $K_0 > 0$ such that for all $t$

$$\mathbf{P}(|\xi| \geq t) = O(\exp(-t^2/K_0)).$$

Let $A = (a_{ij})$ be a random matrix of size $N$ by $N$ where $a_{ij}$ are iid copies of $\xi$. For convenience, we denote by $\mathbf{r}_i(A) = (a_{i1}, \dots, a_{in})$ the $i$-th row vector of $A$.

For a given $1 \leq n \leq N-1$ let $B$ be the submatrix of $A$ formed by $\mathbf{r}_2(A), \dots, \mathbf{r}_{n+1}(A)$ and let $H \subset \mathbf{R}^N$ be the subspace generated by these row vectors. The following concentration result is well known (see for instance [16, Lemma 43], [13, Corollary 2.1.19] or [10, Corollary 3.1]).

**Theorem 1.1.** *With $m = N - n$, we have*

$$\mathbf{P}\left(|\mathrm{dist}(\mathbf{r}_1, H) - \sqrt{m}| \geq t\right) \leq \exp(-t^2/K_0^4).$$

One can justify Theorem 1.1 by applying concentration results of Talagrand or of Hanson-Wright. But all of these methods heavily rely on the fact that $\mathbf{r}_1$ is independent from $H$. In fact, Theorem 1.1 holds as long as $H$ is any deterministic non-degenerate subspace.

As Theorem 1.1 has found many applications (see for instance [6, 14, 15, 16]) and as concentration is useful in Probability in general, it is natural to ask if Theorem 1.1 (or its variant) continues to hold when $\mathbf{r}_1$ and $H$ are correlated. We will address this issue for one of the simplest models, the symmetric Wigner ensembles. In our matrix model $A = (a_{ij})$, the upper diagonal entries $a_{ij}, i \leq j$ are iid copies of $\xi$.

**Theorem 1.2** (Main result, concentration of distance). *With the assumption as above, there exists a sufficiently small positive constant $\kappa$ depending on the subgaussian parameter $K_0$ of $\xi$ such that*

The author is supported by NSF grants DMS-1358648, DMS-1128155 and CCF-1412958. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NSF.

*the following holds for any $m \geq \log^{\kappa^{-1}} N$*

$$\mathbf{P}\left(|\mathrm{dist}(\mathbf{r}_1, H) - \sqrt{m}| \geq t\right) = O\Big(N^{-\omega(1)} + \exp(-\kappa t)\Big).$$

Note that our bound is weaker than Theorem 1.1 mainly due to the use of other spectral concentration results (Lemma 3.3 and Theorem A.1). Theorem 1.2 also implies identical control for the distance from $\mathbf{r}_1$ to other $n$ rows (not necessarily consecutive). We will be also giving tail bounds throughout the proof (see for instance Theorem 5.6 and 6.1).

As we have mentioned, the main bottleneck of our problem is that $\mathbf{x}$ and $H$ are not independent. Roughly speaking, one might guess that the distance of $\mathbf{r}_1 = (a_{11}, \dots, a_{1n})$ to $H$ is close to the distance from the truncated vector $(a_{12}, \dots, a_{1n})$ to the subspace in $\mathbf{R}^{N-1}$ generated by the corresponding truncated row vectors of $B$. The main problem, however, is that the "error term" of this approximation involves a few non-standard statistics of the matrix $P$ which is obtained from $B$ by removing its first column. Notably, we have to resolve the following two obstacles

(1) the operator norm $\|(PP^T)^{-1}\|_2$ must be under cotroll;

(2) the sum $\sum_{1\leq i\leq n}((PP^T)^{-1}P)_{ii}$ must be small;

For bounding $\|(PP^T)^{-1}\|_2$ in (1), as it is the reciprocal of the least singular value $\sigma_n(P)$ of $P$, we need to find an efficient lower bound for $\sigma_n(P)$.

When $n = N-1$, $P$ can be viewed as a random square symmetric matrix of size $N-1$ (and so let's pass to $A$). It is known via the work of [4] and [17] that in this case $A$ is non-singular with very high probability. More quantitatively, the result of Vershynin in [17] shows the following.

**Theorem 1.3.** *There exists a positive constant $\kappa'$ such that the following holds. Let $\sigma_N(A)$ be the least singular value of $A$, then for any $\varepsilon < 1$,*

$$\mathbf{P}(\sigma_N(A) \leq \kappa' \varepsilon N^{-1/2}) = O(\varepsilon^{1/8} + e^{-n^{\kappa'}}).$$

It is conjectured that the bound can be replaced by $O(\varepsilon + e^{-\Theta(n)})$, but in any case these bounds show that we cannot hope for a good control on $\sigma_N^{-1}$. The situation becomes better when we truncate $A$ as the matrix becomes less singular. In fact this has been observed for quite a long time (see for instance [1]). Here we will show the following variant of a recent result by Rudelson and Vershynin from [9].

**Theorem 1.4.** *There exist positive constants $\delta, \kappa'$ such that the following holds*

$$\mathbf{P}(\sigma_n(B) \leq \kappa' \varepsilon m N^{-1/2}) \leq \varepsilon^{\delta m} + e^{-n^{\kappa'}}.$$

Thus, if $m$ grows to infinity with $N$, this theorem implies that $\|(PP^T)^{-1}\|_2$ is well under control with very high probability. We will discuss a detailed treatment for Theorem 1.4 starting from Section 4 by modifying the approach by Rudelson-Vershynin from [9] and by Vershynin from [17].

For the task of controlling $T = \sum_{1\leq i\leq n}((PP^T)^{-1}P)_{ii}$ in (2), which is the main contribution of our note, there are two main steps. In the first step we show that with high probability $((PP^T)^{-1}P)_{ii}$ is close to $-\frac{\sum_{1\le j\le n-1}((R_iR_i^T)^{-1}R_i)_j}{D_i}$, where $D_i \asymp N$ and $R_i$ is the matrix obtained from $P$ by removing its $i$-th row and column. On the other hand, in the second step we show that $T$ is close to $\sum_{1\le j\le n-1}((R_iR_i^T)^{-1}R_i)_j$ for any $1 \le i \le n$ with high probability. These two results will then imply that $T$ is close to zero. Our detailed analysis will be presented in Section 2 and Section 3.

In short, our proof of Theorem 1.2 uses both spectral concentration and anti-concentration coupled with linear algebra identities. Given the simplicity of the statement and of its non-Hermitian counterpart, perhaps it is natural to seek for more direct proof.

Finally, to complete our discussion, we give here an application of Theorem 1.2 on the delocalization of the normal vectors in random symmetric matrices.

**Corollary 1.5.** *Let $\mathbf{x}$ be the normal vector of the subspace generated by $\mathbf{r}_2(A), \dots, \mathbf{r}_N(A)$. Then with overwhelming probability*

$$\|\mathbf{x}\|_\infty = O(\frac{\log^{3/2} n}{\sqrt{n}}).$$

This is an analog of a result from [5] where $A$ is non-symmetric matrix, see also [11]. However, the method in these papers do not seem to extend to our symmetric model.

For terminology, we will use both row and column vectors frequently in this note, here $\mathbf{r}_i(A)$ and $\mathbf{c}_j(A)$ stand for the $i$-th row and $j$-th column of $A$ respectively. Throughout this paper, we regard $N$ as an asymptotic parameter going to infinity. We write $X = O(Y), X \ll Y$, or $Y \gg X$ to denote the claim that $|X| \le CY$ for some fixed $C$; this fixed quantity $C$ is allowed to depend on other fixed quantities such as the sub-gaussian parameter $K_0$ of $\xi$ unless explicitly declared otherwise. We also use $o(Y)$ to denote any quantity bounded in magnitude by $c(N)Y$ for some $c(N)$ that goes to zero as $N \to \infty$.

## 2. Proof of Theorem 1.2: main ingredients

Let $\mathbf{x} = \mathbf{r}_1(A) = (a_{11}, \dots, a_{1N})$ be the first row vector of $A$, and $\mathbf{y}_i := \mathbf{r}_{i+1} = (a_{(i+1)1}, \dots, a_{(i+1)N})$ be the $i+1$-th row vector of $A$ for $1 \le i \le n$. Recall that $B$ is the matrix of size $n \times N$ generated by $\mathbf{y}_i, 1 \le i \le n$. We can represent $\mathrm{dist}(\mathbf{r}_1, H) = \mathrm{dist}(\mathbf{x}, H)$ by the following formula

$$\mathrm{dist}^2(\mathbf{x}, H) = \mathbf{x}(I_N - B^T(BB^T)^{-1}B)\mathbf{x}^T. \tag{1}$$

**Lemma 2.1.** *We have*

$$\mathrm{dist}^2(\mathbf{x}, H) = \mathbf{x}^{[1]}(I_{N-1} - P^T(PP^T)^{-1}P)\mathbf{x}^{[1]^T} + \frac{[a_{11} - \mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]^T}]^2}{1 + \mathbf{z}^T(PP^T)^{-1}\mathbf{z}},$$

*where $P$ be the matrix received from $B$ by removing its first column $\mathbf{z} := (a_{21}, \dots, a_{(n+1)1})^T$, and $\mathbf{x}^{[1]}$ is the truncated row vector of $\mathbf{x}$, $\mathbf{x}^{[1]} = (a_{12}, \dots, a_{1N})$, $B = \begin{pmatrix} a_{11} & \mathbf{x}^{[1]} \\ \mathbf{z} & P \end{pmatrix}$.*

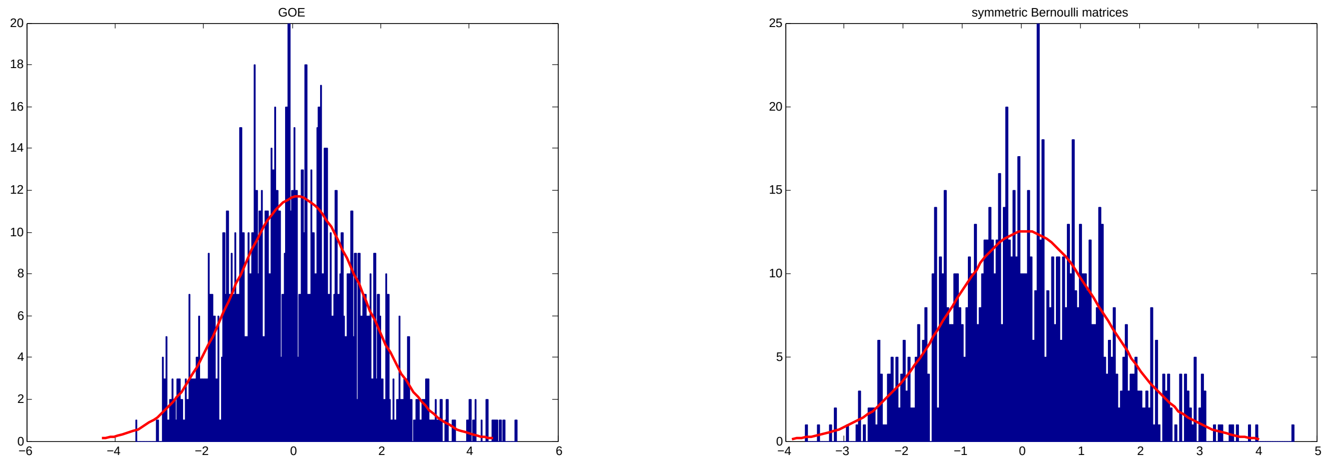


FIGURE 1. We sampled 1000 random matrices of size 1000 from GOE and symmetric Bernoulli ensembles. The histogram represents $(\mathrm{dist}^2 - m)/\sqrt{m}$, where the distance is measured from the first row to the subspace generated by the next 900 rows.

*Proof.* (of Lemma 2.1) We first have

$$Q := (BB^T)^{-1} = (PP^T + \mathbf{z}\mathbf{z}^T)^{-1} = (PP^T)^{-1} - \frac{(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}}{1 + \mathbf{z}^T(PP^T)^{-1}\mathbf{z}}. \tag{2}$$

Thus $B^T(BB^T)^{-1}B = B^TQB$, which has the form

$$B^TQB = \begin{pmatrix} \mathbf{z}^TQ\mathbf{z} & \mathbf{z}^TQP \\ P^TQ\mathbf{z} & P^TQP \end{pmatrix}.$$

Hence,

$$\mathbf{x}(I_N - B^T(BB^T)^{-1}B)\mathbf{x}^T = \mathbf{x}\mathbf{x}^T - a_{11}^2\mathbf{z}^TQ\mathbf{z} - 2a_{11}\mathbf{z}^TQP{\mathbf{x}^{[1]}}^T - \mathbf{x}^{[1]}P^TQP{\mathbf{x}^{[1]}}^T.$$

Using again (2), we obtain the followings

$$\mathbf{z}^T Q\mathbf{z} = \mathbf{z}^T(PP^T)^{-1}\mathbf{z} - \frac{(\mathbf{z}^T(PP^T)^{-1}\mathbf{z})^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} = \frac{\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}};$$

as well as

$$\begin{aligned}
\mathbf{z}^T QP{\mathbf{x}^{[1]}}^T &= \mathbf{z}^T[(PP^T)^{-1} - \frac{(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}]P{\mathbf{x}^{[1]}}^T \\
&= \mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]} - \frac{\mathbf{z}^T(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} \\
&= \frac{\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}};
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{x}^{[1]}P^T QP{\mathbf{x}^{[1]}}^T &= \mathbf{x}^{[1]}P^T[(PP^T)^{-1} - \frac{(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}]P{\mathbf{x}^{[1]}}^T \\
&= \mathbf{x}^{[1]}P^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T - \frac{\mathbf{x}^{[1]}P^T(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}.
\end{aligned}$$

Putting together, one obtains the following

$$\begin{aligned}
\mathrm{dist}^2(\mathbf{x}, H) &= \mathbf{x}(I_N - B^T(BB^T)^{-1}B)\mathbf{x}^T \\
&= \mathbf{x}^{[1]}(I_{N-1} - P^T(PP^T)^{-1}P){\mathbf{x}^{[1]}}^T + a_{11}^2 - a_{11}^2\frac{\mathbf{z}^T(PP^T)^{-1}\mathbf{x}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} \\
&- 2a_{11}\frac{\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} - \frac{\mathbf{x}^{[1]}P^T(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} \\
&= \mathbf{x}^{[1]}(I_{N-1} - P^T(PP^T)^{-1}P){\mathbf{x}^{[1]}}^T \\
&+ \frac{a_{11}^2 - 2a_{11}\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T + \mathbf{x}^{[1]}P^T(PP^T)^{-1}\mathbf{z}\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} \\
&= \mathbf{x}^{[1]}(I_{N-1} - P^T(PP^T)^{-1}P){\mathbf{x}^{[1]}}^T + \frac{[a_{11} - \mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T]^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}},
\end{aligned}$$

proving the lemma. □

Note that the term $\mathbf{x}^{[1]}(I_{N-1} - P^T(PP^T)^{-1}P){\mathbf{x}^{[1]}}^T$ in Lemma 2.1 is just the square distance from $\mathbf{x}^{[1]}$ to the subspace generated by the rows of $P$ in $\mathbf{R}^{N-1}$. The key difference here is that $\mathbf{x}^{[1]}$ is now independent of $P$, so we can apply Theorem 1.1, noting that by Theorem 1.3 the co-dimension of this subspace in $\mathbf{R}^{N-1}$ is $N-1-n=m-1$ with probability at least $\exp(-n^{\kappa_2})$.

**Theorem 2.2.** *Assume that the entries of $A$ have subgaussian parameter $K_0>0$. Then*

$$\mathbf{P}\left(|\sqrt{{\mathbf{x}^{[1]}}^T(I_{N-1}-P^T(PP^T)^{-1}P)\mathbf{x}^{[1]}}-\sqrt{m-1}|\ge \lambda\right)\le \exp(-\lambda^2/K_0)+\exp(-n^{\kappa_2}).$$

This confirms the lower tail bound in Theorem 1.2. To prove the upper tail, we need to study the remaining term $\frac{[x_1-\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T]^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}$ in Lemma 2.1.

**Theorem 2.3** (Bound on the error term, key lemma)**.** *We have*

$$\mathbf{P}(\frac{[x_1-\mathbf{z}^T(PP^T)^{-1}P{\mathbf{x}^{[1]}}^T]^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}\ge t)=O\Big(N^{-\omega(1)}+\exp(-\kappa t)\Big).$$

It is clear that Theorem 2.2 and Theorem 2.3 together imply Theorem 1.2. It remains to justify Theorem 2.3.

## 3. Proof of Theorem 2.3

We will view $P$ as the matrix of the first $n$ rows of a symmetric matrix $A$ of size $N\times N$. We will first need a standard deviation lemma (see for instance [10]).

**Lemma 3.1** (Hanson-Wright inequality)**.** *There exists a constant $C=C(K_0)$ depending on the sub-gaussian moment such that the following holds.*

*(i) Let $A$ be a fixed $M\times M$ matrix. Consider a random vector $\mathbf{x}=(x_1,\dots,x_M)$ where the entries are i.i.d. sub-gaussian of mean zero and variance one. Then*

$$\mathbf{P}(|\mathbf{x}^TA\mathbf{x}-\mathbf{E}\mathbf{x}^TA\mathbf{x}|>t)\le 2\exp(-C\min(\frac{t^2}{\|A\|_{HS}^2},\frac{t}{\|A\|_2})).$$

*In particularly, for any $\alpha>0$*

$$\mathbf{P}\left(|\mathbf{x}^TA\mathbf{x}-\mathbf{E}\mathbf{x}^TA\mathbf{x}|>t\|A\|_{HS}\right)\le \exp(-Ct).$$

*(ii) Let $A$ be a fixed $N\times M$ matrix. Consider a random vector $\mathbf{x}=(x_1,\dots,x_M)$ where the entries are i.i.d. sub-gaussian of mean zero and variance one. Then*

$$\mathbf{P}\left(|\|A\mathbf{x}\|_2-\|A\|_{HS}|>t\|A\|_2\right)\le \exp(-Ct^2).$$

Here the Hilbert-Schmidt norm of $A$ is defined as

$$\|A\|_{HS}=\sqrt{\sum_{i,j}a_{ij}^2}.$$

By the first point of Lemma 3.1,

$$\mathbf{P}\left(|\mathbf{z}^T(PP^T)^{-1}\mathbf{z}-\mathrm{tr}(PP^T)^{-1}|>t\|(PP^T)^{-1}\|_{HS}\right)\le\exp(-Ct) \tag{3}$$

and

$$\mathbf{P}\left(|\mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]^T}-\sum_i((PP^T)^{-1}P)_{ii}|\ge t\|(PP^T)^{-1}P\|_{HS})\right)\le\exp(-Ct). \tag{4}$$

3.2. **Treatment for $\mathbf{z}(PP^T)^{-1}\mathbf{z}^T$.** To estimate (3), we need to consider the stable rank of the matrix $(PP^T)^{-1}$. By Theorem 1.4 (where $B$ is replaced by $P$) and by definition, we have the following important estimate with very high probability

$$\mathbf{P}(\sigma_n^{-2}(P)\le(\kappa'\varepsilon)^{-2}N/m^2)\ge 1-\varepsilon^{-\delta m}-\exp(-n^{\kappa'}). \tag{5}$$

We next introduce another useful ingredient whose proof is deferred to Appendix A.

**Lemma 3.3.** *With probability at least $1-N^{-\omega(1)}$, for any $\log^{\Theta(1)}N\ll m\ll N$, the interval $[x_0+C_1m/N^{1/2},x_0+C_2m/N^{1/2}],x_0>0$ inside the bulk contains $\Theta(m)$ singular values of $P$.*

For convenience, we will set

$$p(m,N):=N^{-\omega(1)}+\exp(-\delta m)+\exp(-n^{\kappa'}. \tag{6}$$

We deduce the following asymptotic behavior of the trace of $(PP^T)^{-1}$.

**Corollary 3.4.** *With probability at least $1-p(m,N)$,*

$$\|(PP^T)^{-1}P\|_{HS}^2=\mathrm{tr}((PP^T)^{-1})\asymp N/m.$$

*Proof.* (of Corollary 3.4) For the lower bound, let $\sigma_i,\dots,\sigma_j$ be the $\Theta(m)$ singular values of $P$ lying in the interval $[m/N^{1/2},Cm/N^{1/2}]$. Then

$$\mathrm{tr}((PP^T)^{-1})\ge\sum_{i\le k\le j}\lambda_i^{-2}\gg mN/m^2=N/m.$$

For the upper bound, first of all by (5), it suffices to assume that all of the singular values of $P$ are at least $cm/N^{1/2}$ for some sufficiently small $c$.

First, it is clear that the inverval $[cm/N^{1/2},C_1m/N^{1/2}]$ contains $O(m)$ singular values of $P$. For the remaining interval $[C_1m/N^{1/2},\Theta(N/N^{1/2})]$ we divide it into intervals $I_k$ of length $(C_2-C_1)m/N^{1/2}$ each with $1\le k\le\Theta(N/m)$. By Lemma 3.3 the number of singular vectors in each interval is proportional to $m$. As such

$$\mathrm{tr}((PP^T)^{-1}) = \sum_i \sigma_i^{-2} = \sum_k \sum_{i \in I_k} \lambda_i^{-2} \ll \sum_k mN/k^2m^2 \ll N/m.$$

□

By Corollary 3.4 and (5),

$$\mathbf{P}(\frac{\mathrm{tr}(PP^T)^{-1}}{\|(PP^T)^{-1}\|_2} \gg m) \ge 1 - p(m,N). \tag{7}$$

Notice that

$$\begin{aligned}\|(PP^T)^{-1}\|_{HS} = \sqrt{\sum_i \lambda_i^2((PP^T)^{-1})} &\le \sqrt{\|(PP^T)^{-1}\|_2 \sum_i \lambda_i((PP^T)^{-1})} \\ &= \sqrt{\|(PP^T)^{-1}\|_2}\sqrt{\mathrm{tr}((PP^T)^{-1})}.\end{aligned}$$

As a consequence, with probability at least $1 - p(m,N)$

$$\frac{\mathrm{tr}((PP^T)^{-1})}{\|(PP^T)^{-1}\|_{HS}} \ge \frac{\sqrt{\mathrm{tr}((PP^T)^{-1})}}{\sqrt{\|(PP^T)^{-1}\|_2}} \gg m^{1/2}, \tag{8}$$

where we used (7) in the last estimate.

Combining (3), (7) and (8), we have learned that

**Lemma 3.5.** *For any $t > 0$,*

$$\mathbf{P}\Big(|\mathbf{z}^T(PP^T)^{-1}\mathbf{z} - \mathrm{tr}((PP^T)^{-1})| \ge \frac{t}{m^{1/2}}\mathrm{tr}((PP^T)^{-1})\Big) \le p(m,N) + \exp(-Ct).$$

*Consequently, with probability at least $1 - (p(m,N) + \exp(-c\sqrt{m}))$*

$$\mathbf{z}^T(PP^T)^{-1}\mathbf{z} \asymp \|\mathrm{tr}((PP^T)^{-1}) \asymp \frac{N}{m}.$$

3.6. **Treatment for $\mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]}$.** To show this quantity small, we will divide the treatment into two main steps.

**Step 1: Set-up.** We will present here the calculation for $((PP^T)^{-1}P)_{11}$, the formula for other $((PP^T)^{-1}P)_{ii}$ follows the same line. For simplicity, we will drop the super index [1] in all $\mathbf{y}_i$ (as we can view $P$ as the submatrix of the first $n$ rows of $A$). One first interprets

$$((PP^T)^{-1}P)_{11} = \langle \mathbf{r}_1((PP^T)^{-1}), \mathbf{c}_1(P)\rangle.$$

Because the matrix $PP^T$ depends on $\mathbf{c}_1(P)$, we need to separate the dependences. First of all, we write $PP^T$ as a rank-one perturbation $PP^T = QQ^T + \mathbf{c}_1(P)\mathbf{c}_1^T(P)$, where $Q$ is the matrix obtained from $P$ by removing its first column. As such

$$\begin{aligned}(PP^T)^{-1} &= (QQ^T + \mathbf{c}_1(P)\mathbf{c}_1^T(P))^{-1}\\ &= (QQ^T)^{-1} - \frac{[(QQ^T)^{-1}\mathbf{c}_1(P)][(\mathbf{c}_1(P))^T((QQ)^T)^{-1}]}{1+(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P)}.\end{aligned}$$

It follows that

$$\begin{aligned}(PP^T)^{-1}\mathbf{c}_1(P) &= (QQ^T)^{-1}\mathbf{c}_1(P) - \frac{1}{1+(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P)}[(QQ^T)^{-1}\mathbf{c}_1(P)][(\mathbf{c}_1(P))^TQQ^T)^{-1}]\mathbf{c}_1(P)\\ &= \frac{1}{1+(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P)}(QQ^T)^{-1}\mathbf{c}_1(P). \qquad (9)\end{aligned}$$

In particularly,

$$((PP^T)^{-1}P)_{11} = \langle \mathbf{r}_1((PP^T)^{-1}), \mathbf{c}_1(P)\rangle = \frac{1}{1+(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P)}\langle \mathbf{r}_1((QQ^T)^{-1}), \mathbf{c}_1(P)\rangle.$$

Now the matrix $QQ^T$ still depends on $\mathbf{c}_1(P)$, so we are going to separate independence once more, this time using Schur complement.

**Fact 3.7.** *Let $M = \begin{pmatrix} X & Y \\ Y^T & Z\end{pmatrix}$. Assuming invertibility whenever necessary, we have*

$$M^{-1} = \begin{pmatrix} (X-YZ^{-1}Y^T)^{-1} & -X^{-1}Y(Z-Y^TX^{-1}Y)^{-1} \\ (-X^{-1}Y(Z-Y^TX^{-1}Y)^{-1})^T & (Z-Y^TX^{-1}Y)^{-1}\end{pmatrix}.$$

Note that $QQ^T$ can be written as $\begin{pmatrix} \mathbf{y}_1^{[1]}(\mathbf{y}_1^{[1]})^T & \mathbf{y}_1^{[1]}R^T \\ R(\mathbf{y}_1^{[1]})^T & RR^T\end{pmatrix}$, where $R$ is the matrix obtained from $Q$ by removing its first row $\mathbf{y}_1^{[1]}$. From now on, for short we set

$$x^2 := \mathbf{y}_1^{[1]}(\mathbf{y}_1^{[1]})^T$$

and

$$d^2 := x^2 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}R(\mathbf{y}_1^{[1]})^T.$$

To begin with, the top left corner is

$$(X - YZ^{-1}Y^T)^{-1} = (x^2 - \mathbf{y}_1^{[1]} R^T (RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T)^{-1} = d^{-2}.$$

Next, the bottom right Schur complement then can be calculated as

$$(Z - Y^T X^{-1} Y)^{-1} = [RR^T - \frac{1}{x^2} R(\mathbf{y}_1^{[1]})^T \mathbf{y}_1^{[1]} R^T]^{-1}.$$

Note that this again can be considered as rank-one perturbation,

$$\begin{aligned}
[RR^T - \frac{1}{x^2} R(\mathbf{y}_1^{[1]})^T \mathbf{y}_1^{[1]} R^T]^{-1} &= (RR^T)^{-1} + \frac{1}{x^2} \frac{[(RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T][\mathbf{y}_1^{[1]} R^T (RR^T)^{-1}]}{1 - \frac{1}{x^2}\mathbf{y}_1^{[1]} R^T (RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T} \\
&= (RR^T)^{-1} + \frac{[(RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T][\mathbf{y}_1^{[1]} R^T (RR^T)^{-1}]}{d^2}.
\end{aligned}$$

Similarly, the top right Schur complement is

$$\begin{aligned}
-X^{-1}Y(Z - Y^T X^{-1} Y)^{-1} &= -\frac{1}{x^2}\mathbf{y}_1^{[1]} R^T \left[(RR^T)^{-1} + \frac{[(RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T][\mathbf{y}_1^{[1]} R^T (RR^T)^{-1}]}{d^2}\right] \\
&= -\frac{1}{x^2}\mathbf{y}_1^{[1]} R^T (RR^T)^{-1} + (\frac{1}{x^2} - \frac{1}{d^2})\mathbf{y}_1^{[1]} R^T (RR^T)^{-1} \\
&= -d^{-2}\mathbf{y}_1^{[1]} R^T (RR^T)^{-1}.
\end{aligned}$$

Putting together,

$$(QQ^T)^{-1} = \begin{pmatrix} d^{-2} & -d^{-2}\mathbf{y}_1^{[1]} R^T (RR^T)^{-1} \\ -d^{-2}(RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T & (RR^T)^{-1} + d^{-2}[(RR^T)^{-1} R(\mathbf{y}_1^{[1]})^T][\mathbf{y}_1^{[1]} R^T (RR^T)^{-1}] \end{pmatrix},$$

where we note that the involved matrices are invertible by Theorem 1.3 with extremely large probability.

It follows that

$$\langle \mathbf{r}_1((QQ^T)^{-1}), \mathbf{c}_1(P)\rangle = d^{-2}[\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]} R^T (RR^T)^{-1} \mathbf{c}_1(P)^{[1]}].$$

Also, the denominator of (9) can be written as

$$\begin{aligned}(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P) &= d^{-2}[(\mathbf{c}_1(P)_1)^2 - 2\mathbf{c}_1(P)_1(\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}R(\mathbf{y}_1^{[1]})] \\ &+ (\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]} \\ &+ d^{-2}[(\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}R(\mathbf{y}_1^{[1]})^T][\mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}] \\ &= (\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]} + d^{-2}[\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}]^2.\end{aligned}$$

Combining the formulas, we arrive at the following.

**Lemma 3.8.** *We have*

$$((PP^T)^{-1}P)_{11} = \frac{\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}}{d^2(1 + (\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}) + (\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]})^2}.$$

To speculate further, note that by concentration, $d^2 = x^2 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}R(\mathbf{y}_1^{[1]})^T$ is just the distance from $\mathbf{y}_1^{[1]}$ to the subspace generated by the rows of $R$, and thus is well concentrated around $(N-1)-(n-1)=m$ by Theorem 1.1, that is with probability at least $\exp(-cm)$

$$d^2 \gg m. \tag{10}$$

**Corollary 3.9.** *With probability at least* $1-(p(m,N)+\exp(-c\sqrt{m}))$,

$$|((PP^T)^{-1}P)_{11}| = O(\frac{1}{\sqrt{N}}).$$

*Consequently,*

$$\sum_i (PP^T)^{-1}P)_{ii} = O(\frac{n}{\sqrt{N}}) = O(\sqrt{N}).$$

*Proof.* (of Corollary 3.10) By Cauchy-Schwarz,

$$\begin{aligned}&\frac{|\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}|}{d^2(1 + (\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}) + (\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]})^2} \\ &\le \frac{1}{2\sqrt{d^2(1 + (\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]})}} = O(\frac{1}{\sqrt{N}}),\end{aligned}$$

where in the last estimate we used $d^2 \gg m$ and Lemma 3.5. □

We also deduce the following consequence.

**Corollary 3.10.** *With probability at least* $1 - p(m,N) - \exp(-c\sqrt{m})$

$$|\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]} R^T (RR^T)^{-1} \mathbf{c}_1(P)^{[1]}| \ll \sqrt{N}.$$

*Proof.* (of Corollary 3.10) First, by Lemma 3.1, with probability at least $1 - \exp(-c\sqrt{m})$

$$|\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]} R^T (RR^T)^{-1} \mathbf{c}_1(P)^{[1]} - \sum_i (R^T(RR^T)^{-1})_{ii}| \le \sqrt{m} \|R^T(RR^T)^{-1}\|_{HS}.$$

By Corollary 3.9, with probability at least $1 - p(m,N) - \exp(c\sqrt{m})$

$$|\sum_i (R^T(RR^T)^{-1})_{ii}| = O(\sqrt{N})$$

and by Corollary 3.4, with probability at least $1 - p(m,N)$

$$\|R^T(RR^T)^{-1}\|_{HS}^2 = \mathrm{tr}((RR^T)^{-1}) = O(\frac{N}{m}).$$

□

We remark that Lemma 3.5 and Corollary 3.10 allows us to conclude that $\frac{[x_1 - \mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]^T}]^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}}$ has order $m$ with high probability, but this is not strong enough for Theorem 2.3. We will improve this in the next phase of the proof.

**Step II: Comparison.** In this step we show that $\sum_{1\le i\le n}((PP^T)^{-1}P)_{ii}$ is close to $\sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii}$.

Recall that

$$\begin{aligned}(PP^T)^{-1} &= (QQ^T + \mathbf{w}\mathbf{w}^T)^{-1}\\ &= (QQ^T)^{-1} - \frac{1}{1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w}}[(QQ^T)^{-1}\mathbf{w}][\mathbf{w}^T(QQ^T)^{-1}]\\ &:= (QQ^T)^{-1} - Q'\end{aligned}$$

where for convenience, we denote the second matrix by $Q'$. Note furthermore that

$$(QQ^T)^{-1} = \begin{pmatrix} d^{-2} & -d^{-2}\mathbf{y}R^T(RR^T)^{-1} \\ -d^{-2}(RR^T)^{-1}R\mathbf{y}^T & (RR^T)^{-1} + d^{-2}[(RR^T)^{-1}R\mathbf{y}^T][\mathbf{y}R^T(RR^T)^{-1}] \end{pmatrix}$$

and

$$P = (\mathbf{w} \quad Q) = \begin{pmatrix} x_0 & \mathbf{y} \\ \mathbf{z} & R \end{pmatrix}$$

where for short we denote $\mathbf{w} = \mathbf{c}_1(P) = (x_0, \mathbf{z})^T$ and $\mathbf{y} = \mathbf{y}^{[1]}$ and also recall that

$$d^2 = \mathbf{y}\mathbf{y}^T - \mathbf{y}R^T(RR^T)^{-1}R\mathbf{y}^T.$$

We now compute $\sum_{2\le i\le n}((PP^T)^{-1}P)_{ii}$. To do this, we start from $(QQ^T)^{-1}P$ and eliminate its first row and column to obtain a matrix $M$ of size $(n-1)\times(N-1)$

$$\begin{aligned} M_1 &= -d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y} + \Big((RR^T)^{-1} + d^{-2}[(RR^T)^{-1}R\mathbf{y}^T][\mathbf{y}R^T(RR^T)^{-1}]\Big)R \\ &= (RR^T)^{-1}R - d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I - R^T(RR^T)^{-1}R) \\ &= (RR^T)^{-1}R - M_1', \end{aligned} \tag{11}$$

with

$$M_1' := d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I - R^T(RR^T)^{-1}R). \tag{12}$$

Next, for the contribution of $Q'P$ (after the elimination of its first row and column), we need to compute the vector $(QQ^T)^{-1}\mathbf{w}$.

$$\begin{aligned} (QQ^T)^{-1}\mathbf{w} &= \begin{pmatrix} d^{-2}(x_0 - \mathbf{y}R^T(RR^T)^{-1}\mathbf{z}) \\ -x_0 d^{-2}(RR^T)^{-1}R\mathbf{y}^T + \Big((RR^T)^{-1} + d^{-2}[(RR^T)^{-1}R\mathbf{y}^T][\mathbf{y}R^T(RR^T)^{-1}]\Big)\mathbf{z} \end{pmatrix} \\ &= \begin{pmatrix} d^{-2}(x_0 - \mathbf{y}R^T(RR^T)^{-1}\mathbf{z}) \\ -d^{-2}(x_0 - \mathbf{y}R^T(RR^T)^{-1}\mathbf{z})(RR^T)^{-1}R\mathbf{y}^T + (RR^T)^{-1}\mathbf{z} \end{pmatrix}. \\ &= \begin{pmatrix} a \\ -a(RR^T)^{-1}R\mathbf{y}^T + (RR^T)^{-1}\mathbf{z} \end{pmatrix}, \end{aligned}$$

with

$$a := d^{-2}(x_0 - \mathbf{y}R^T(RR^T)^{-1}\mathbf{z}).$$

As a result, the bottom left submatrix of $(QQ^T)^{-1}\mathbf{w}\mathbf{w}^T(QQ^T)^{-1}$ is the vector $-a^2(RR^T)^{-1}R\mathbf{y}^T + a(RR^T)^{-1}\mathbf{z}$ and the bottom right submatrix is the matrix

$$a^2(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}R^T(RR^T)^{-1} + (RR^T)^{-1}\mathbf{z}\mathbf{z}^T(RR^T)^{-1} - a(RR^T)^{-1}R\mathbf{y}^T\mathbf{z}^T(RR^T)^{-1} - a(RR^T)^{-1}\mathbf{z}\mathbf{y}R^T(RR^T)^{-1}.$$

It follows that the matrix $M_2$ obtained by eliminating the first row and column of $Q'P$ can be written as

$$
\begin{aligned}
M_2 &= \frac{1}{1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w}}\Big[-a^2(RR^T)^{-1}R\mathbf{y}^T\mathbf{y} + a(RR^T)^{-1}\mathbf{z}\mathbf{y} \\
&\quad + a^2(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}R^T(RR^T)^{-1}R + (RR^T)^{-1}\mathbf{z}\mathbf{z}^T(RR^T)^{-1}R - a(RR^T)^{-1}[R\mathbf{y}^T\mathbf{z}^T + \mathbf{z}\mathbf{y}R^T](RR^T)^{-1}R\Big] \\
&= \frac{1}{1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w}}\Big[a^2(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(R^T(RR^T)^{-1}R - I) + a(RR^T)^{-1}\mathbf{z}\mathbf{y} \\
&\quad + (RR^T)^{-1}\mathbf{z}\mathbf{z}^T(RR^T)^{-1}R - a(RR^T)^{-1}[R\mathbf{y}^T\mathbf{z}^T + \mathbf{z}\mathbf{y}R^T](RR^T)^{-1}R\Big]. \qquad (13)
\end{aligned}
$$

In summary,

$$\sum_{2\le i\le n}((PP^T)^{-1}P)_{ii} = \sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii} - \sum_{1\le i\le n-1}(M_1'+M_2)_{ii}.$$

In what follows, by using the formulas for $M_1', M_2$ from (11), (13) we show that $\sum_{1\le i\le n-1}(M_1' + M_2)_{ii}$ is negligible.

We will try to simplify the formulae a bit. First,

$$
\begin{aligned}
1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w} &= 1+(\mathbf{c}_1(P))^T(QQ^T)^{-1}\mathbf{c}_1(P) = 1+(\mathbf{c}_1(P)^{[1]})^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]} \\
&\quad + \frac{1}{x^2-\mathbf{y}_1^{[1]}R^T(RR^T)^{-1}R(\mathbf{y}_1^{[1]})^T}[\mathbf{c}_1(P)_1 - \mathbf{y}_1^{[1]}R^T(RR^T)^{-1}\mathbf{c}_1(P)^{[1]}]^2 \\
&= 1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z} + d^{-2}[x_0 - \mathbf{y}R^T(RR^T)^{-1}\mathbf{z}]^2. \qquad (14)
\end{aligned}
$$

Thus, by Lemma 3.5, with probability at least $1-(p(m,N)+\exp(-c\sqrt{m}))$,

$$\mathbf{w}^T(QQ^T)^{-1}\mathbf{w} \asymp \mathrm{tr}((RR^T)^{-1}) \asymp \frac{N}{m}. \qquad (15)$$

Also from (14) and (12),

$$
\begin{aligned}
(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})M_1' &= [d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)]\times \\
&\quad \times\Big(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z} + d^{-2}[x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}]^2\Big) \\
&= d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})+ \\
&\quad + a^2(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R).
\end{aligned}
$$

Hence the normalized matrix $(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})(M_1'+M_2)$ can be expressed as

$$\begin{aligned}(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})(M_1'+M_2) &= d^{-2}(RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})+a(RR^T)^{-1}\mathbf{zy}\\ &+(RR^T)^{-1}\mathbf{zz}^T(RR^T)^{-1}R-a(RR^T)^{-1}[R\mathbf{y}^T\mathbf{z}^T+\mathbf{zy}R^T](RR^T)^{-1}R\\ &=((RR^T)^{-1})\Big(d^{-2}(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)+a\mathbf{zy}\\ &+\mathbf{zz}^T(RR^T)^{-1}R-a[R\mathbf{y}^T\mathbf{z}^T+\mathbf{zy}R^T](RR^T)^{-1}R\Big)\\ &:=(RR^T)^{-1}S.\end{aligned}$$

We can write the second matrix $S$ as $\sum_1+\sum_2$ where

$$\begin{aligned}\sum_1 &:= d^{-2}(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)+a\mathbf{zy}-a\mathbf{zy}R^T(RR^T)^{-1}R\\ &=d^{-2}(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)+a\mathbf{zy}(I-R^T(RR^T)^{-1}R)\\ &=d^{-2}\Big(\big[(x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z})\mathbf{z}+(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})R\mathbf{y}^T\big]\big[\mathbf{y}(I-R^T(RR^T)^{-1}R)\big]\Big),\end{aligned}$$

and

$$\begin{aligned}\sum_2 &:= \mathbf{zz}^T(RR^T)^{-1}R-aR\mathbf{y}^T\mathbf{z}^T(RR^T)^{-1}R=(\mathbf{z}-aR\mathbf{y}^T)[\mathbf{z}^T(RR^T)^{-1}R]\\ &=d^{-2}\Big(\big[(\mathbf{y}(I-R^T(RR^T)^{-1}R)\mathbf{y}^T)\mathbf{z}-(x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}))R\mathbf{y}^T\big]\big[\mathbf{z}^T(RR^T)^{-1}R\big]\Big).\end{aligned}$$

Hence

$$\begin{aligned}&(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})(M_1'+M_2)=((RR^T)^{-1})S=(RR^T)^{-1}(\sum_1+\sum_2)\\ &=d^{-2}\big[(x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z})(RR^T)^{-1}\mathbf{z}+(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})(RR^T)^{-1}R\mathbf{y}^T\big]\big[\mathbf{y}(I-R^T(RR^T)^{-1}R)\big]\\ &+d^{-2}\big[(\mathbf{y}(I-R^T(RR^T)^{-1}R)\mathbf{y}^T)(RR^T)^{-1}\mathbf{z}-(x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}))(RR^T)^{-1}R\mathbf{y}^T\big]\big[\mathbf{z}^T(RR^T)^{-1}R\big].\end{aligned}$$

By the triangle inequality,

$$
\begin{aligned}
&|\sum_{1\le i\le n-1}(M_1'+M_2)_{ii}|\le\\
&\le \frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}|x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}||\sum_{1\le i\le n-1}\big((RR^T)^{-1}\mathbf{z}\mathbf{y}(I-R^T(RR^T)^{-1}R)\big)_{ii}|\\
&+\frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})|\sum_{1\le i\le n-1}\big((RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)\big)_{ii}|\\
&+\frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}(\mathbf{y}(I-R^T(RR^T)^{-1}R)\mathbf{y}^T)|\sum_{1\le i\le n-1}\big((RR^T)^{-1}\mathbf{z}\mathbf{z}^T(RR^T)^{-1}R\big)_{ii}|\\
&-\frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}|x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}||\sum_{1\le i\le n-1}\big((RR^T)^{-1}R\mathbf{y}^T\mathbf{z}^T(RR^T)^{-1}R\big)_{ii}|\\
&:=E_1+E_2+E_3+E_4. \qquad (16)
\end{aligned}
$$

To complete our estimates for $E_1, E_2, E_3, E_4$, we recall from Corollary 3.4, Corollary 3.10, from (7), (10) and (15) that with probability at least $1-(p(m,N)+\exp(c\sqrt{m}))$

- $$m \ll d^2, \qquad (17)$$
- $$|x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}| \ll \sqrt{N}, \qquad (18)$$
- $$m^{1/2}\|(RR^T)^{-1}\|_{HS} \ll \mathrm{tr}((RR^T)^{-1}). \qquad (19)$$

We will also need an elementary claim.

**Fact 3.11.** *Assume that* $\mathbf{a}=(a_1,\dots,a_{m_1})\in\mathbf{R}^{m_1}$ *and* $\mathbf{b}=(b_1,\dots,b_{m_2})\in\mathbf{R}^{m_2}$ *are column vectors with* $m_1\le m_2$*. Then*

$$\sum_{1\le i\le m_1}(\mathbf{a}\mathbf{b}^T)_{ii}=\sum_{1\le i\le m_1}a_ib_i\le \|\mathbf{a}\|_2\|\mathbf{b}\|_2.$$

Using Fact 3.11 and (2) of Theorem 3.1, we have

$$
\begin{aligned}
|E_1| &= \frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}|x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}||\sum_{1\le i\le n-1}\big((RR^T)^{-1}\mathbf{z}\mathbf{y}(I-R^T(RR^T)^{-1}R)\big)_{ii}|\\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\sqrt{N}\|(RR^T)^{-1}\|_{HS}\|(I-R^T(RR^T)^{-1}R)\|_{HS}\\
&= (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\sqrt{N}\|(RR^T)^{-1}\|_{HS}\sqrt{\mathrm{tr}((I-R^T(RR^T)^{-1}R)^2)}\\
&= (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\sqrt{N}\|(RR^T)^{-1}\|_{HS}\sqrt{\mathrm{tr}(I-R^T(RR^T)^{-1}R)}\\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\sqrt{N}\sqrt{m}\|(RR^T)^{-1}\|_{HS}\ll \frac{\sqrt{N}}{m}.
\end{aligned}
$$

Similarly,

$$\begin{aligned}
E_2 =& \frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}(1+\mathbf{z}^T(RR^T)^{-1}\mathbf{z})|\sum_{1\le i\le n-1}\big((RR^T)^{-1}R\mathbf{y}^T\mathbf{y}(I-R^T(RR^T)^{-1}R)\big)_{ii}| \\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\mathrm{tr}((RR^T)^{-1})\|(RR^T)^{-1}R\|_{HS}\|(I-R^T(RR^T)^{-1}R)\|_{HS} \\
&\ll m^{-1}\sqrt{\frac{N}{m}}\sqrt{m} = \frac{\sqrt{N}}{m};
\end{aligned}$$

and

$$\begin{aligned}
E_3 =& \frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}(\mathbf{y}(I-R^T(RR^T)^{-1}R)\mathbf{y}^T)|\sum_{1\le i\le n-1}\big((RR^T)^{-1}\mathbf{z}\mathbf{z}^T(RR^T)^{-1}R\big)_{ii}| \\
&\le (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}m\|(RR^T)^{-1}\|_{HS}\|(RR^T)^{-1}R)\|_{HS} \\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{-1/2}\|(RR^T)^{-1}\|_{HS} \ll (\mathrm{tr}((RR^T)^{-1}))^{-1/2}\mathrm{tr}((RR^T)^{-1})m^{-1/2} \\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{1/2}m^{-1/2} \ll \frac{\sqrt{N}}{m}.
\end{aligned}$$

Lastly,

$$\begin{aligned}
|E_4| =& \frac{1}{d^2(1+\mathbf{w}^T(QQ^T)^{-1}\mathbf{w})}|x_0-\mathbf{y}R^T(RR^T)^{-1}\mathbf{z}||\sum_{1\le i\le n-1}\big((RR^T)^{-1}R\mathbf{y}^T\mathbf{z}^T(RR^T)^{-1}R\big)_{ii}| \\
&\ll (\mathrm{tr}((RR^T)^{-1}))^{-1}m^{-1}\sqrt{N}\|(RR^T)^{-1}R\|_{HS}^2 \\
&\ll \frac{\sqrt{N}}{m}.
\end{aligned}$$

We sum up below

$$\begin{aligned}
&|\sum_{1\le i\le n}((PP^T)^{-1}P)_{ii} - \sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii}| \\
&\le |((PP^T)^{-1}P)_{11}| + |\sum_{2\le i\le n}((PP^T)^{-1}P)_{ii} - \sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii}| \\
&\le |((PP^T)^{-1}P)_{11}| + |E_1| + E_2 + E_3 + |E_4| \\
&\ll \frac{1}{\sqrt{N}} + \frac{\sqrt{N}}{m} \ll \frac{\sqrt{N}}{m}.
\end{aligned}$$

**Lemma 3.12.** *With probability at least* $1-(p(m,N)+\exp(-c\sqrt{m}))$ *we have*

$$\mathcal{E}_1 := |\sum_{1\le i\le n}((PP^T)^{-1}P)_{ii} - \sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii}| \ll \frac{\sqrt{N}}{m}.$$

Here it is emphasized that the index 1 of $\mathcal{E}_1$ shows the error of comparison between $\sum_{1\le i\le n}((PP^T)^{-1}P)_{ii}$ and $\sum_{1\le i\le n-1}((RR^T)^{-1}R)_{ii}$ where $R$ is obtain by removing the *first* row and column of $P$. If we remove its $k$-th row and column instead, then we use $\mathcal{E}_k$ to denote the difference.

3.13. **Putting things together.** Set

$$T := \sum_{1\le i\le n} ((PP^T)^{-1}P)_{ii}.$$

By Lemma 3.8 and Lemma 3.1, for each $1 \le i \le n$

$$\mathbf{P}\Big(((PP^T)^{-1}P)_{ii} \le \frac{(\mathcal{E}_i - T) + t\|(R_iR_i^T)^{-1}R_i\|_{HS}}{D_i}\Big) \ge 1 - (p(m,N) + \exp(-c\sqrt{m})) - \exp(-Ct), \tag{20}$$

where the denominator

$$\begin{aligned} D_i = &\Big[x^2 - \mathbf{y}_i^{[1]} R^T (R_iR_i^T)^{-1} R_i (\mathbf{y}_i^{[1]})^T\Big]\Big[1 + (\mathbf{c}_i(P)^{[1]})^T (R_iR_i^T)^{-1}\mathbf{c}_i(P)^{[1]}\Big] \\ &+ [\mathbf{c}_i(P)_1 - \mathbf{y}_i^{[1]} R_i^T (R_iR_I^T)^{-1}\mathbf{c}_i(P)^{[1]}]^2 \gg N, \end{aligned}$$

with probability at least $1 - (p(m,N) + \exp(-c\sqrt{m}))$.

Rewriting (20) and summing over $i$, with probability at least $1 - n(p(m,N) + \exp(-c\sqrt{m}) + \exp(-Ct))$,

$$\begin{aligned} |T + T(\sum_i \frac{1}{D_i})| &\le \sum_i \frac{\mathcal{E}_i + t\|(R_iR_i^T)^{-1}R_i\|_{HS})}{D_i} \\ &\ll N\frac{(t+\frac{1}{\sqrt{m}})\sqrt{\frac{N}{m}}}{N} \\ &\ll t\sqrt{\frac{N}{m}}. \end{aligned}$$

We summarize into a lemma as follows.

**Lemma 3.14.** *With probability at least* $1 - n(p(m,N) + \exp(-c\sqrt{m}) + \exp(-Ct))$,

$$T = |\sum_{1\le i\le n} ((PP^T)^{-1}P)_{ii}| \ll (t+1)\sqrt{\frac{N}{m}}.$$

We now complete the proof of Theorem 2.3. By Hanson-Wright estimates, and by Lemma 3.14, with probability at least $1-n(p(m,N)+\exp(-c\sqrt{m})+2\exp(-Ct))$

$$\frac{[x_1-\mathbf{z}^T(PP^T)^{-1}P\mathbf{x}^{[1]^T}]^2}{1+\mathbf{z}^T(PP^T)^{-1}\mathbf{z}} \le \frac{\left[|\sum_{1\le i\le n}((PP^T)^{-1}P)_{ii}|+t\|(PP^T)^{-1}P\|_{HS}\right]^2}{\mathrm{tr}(PP^T)}$$
$$\ll \frac{(t\sqrt{\frac{N}{m}})^2}{N/m}$$
$$\ll t^2.$$

Our proof is then complete by choosing $\kappa$ (stated in Theorem 2.3) to be any constant smaller than $\delta,\kappa',c$.

## 4. Proof of Theorem 1.4: sketch

In this section we sketch the idea to prove Theorem 1.4, details of the proof will be presented in later sections. Roughly speaking, we will follow the treatment by Rudelson and Vershynin from [9] and by Vershynin from [17] with some modifications. We also refer the reader to a more recent paper by Rudelson and Vershynin [12] for similar treatments.

For convenience, most of the constants used in our subsequent treatments, if not specified, are locally restricted.

We first need some preparations, for a cosmetic reason, let us view $B$ as a *column* matrix of size $N$ by $n$ of the last $n$ columns of $A$ from now on,

$$B=\begin{pmatrix}\mathbf{c}_{m+1}(A) & \dots & \mathbf{c}_N(A)\end{pmatrix}.$$

Let $c_0,c_1\in(0,1)$ be two numbers. We will choose their values later as small constants that depend only on the subgaussian parameter $K_0$.

**Definition 4.1.** A vector $\mathbf{x}\in\mathbf{R}^n$ is called *sparse* if $|\mathbf{supp}(\mathbf{x})|\le c_0n$. A vector $\mathbf{x}\in S^{n-1}$ is called *compressible* if $\mathbf{x}$ is within Euclidean distance $c_1$ from the set of all sparse vectors. A vector $\mathbf{x}\in S^{n-1}$ is called *incompressible* if it is not compressible.

The sets of compressible and incompressible vectors in $S^{n-1}$ will be denoted by $\mathbf{Comp}(c_0,c_1)$ and $\mathbf{Incomp}(c_0,c_1)$ respectively.

Given a vector random variable $\mathbf{x}$ and a radius $r$, we define the *Levy concentration* of $\mathbf{x}$ (or the *small ball probability* with radius $r$) to be

$$\mathcal{L}(\mathbf{x},r):=\sup_{\mathbf{u}}\mathbf{P}(\|\mathbf{x}-\mathbf{u}\|_2\le r).$$

In order to prove Theorem 1.4, we decopose $S^{n-1}$ into compressible and incompressible vectors for some appropriately chosen parameter $c_0$ and $c_1$. Let $\mathcal{E}_K$ be the event that

$$\mathcal{E}_K = \{\|B\|_2 \le 3K\sqrt{N}\}.$$

$$\mathbf{P}\big(\min_{\mathbf{x}\in S^{n-1}} \|B^T\mathbf{x}\|_2 \le \varepsilon(\sqrt{N}-\sqrt{n}) \cap \mathcal{E}_K\big) \le \mathbf{P}\big(\min_{\mathbf{x}\in \mathbf{Comp}(c_0,c_1)} \|B^T\mathbf{x}\|_2 \le \varepsilon(\sqrt{N}-\sqrt{n}) \cap \mathcal{E}_K\big)$$
$$+\mathbf{P}\big(\min_{\mathbf{x}\in \mathbf{Incomp}(c_0,c_1)} \|B^T\mathbf{x}\|_2 \le \varepsilon(\sqrt{N}-\sqrt{n}) \cap \mathcal{E}_K\big).$$

In this section we only treat with the compressible vectors (by giving a stronger bound).

First of all, we bound for a fixed vector $\mathbf{x}$. The following follows from the mentioned work by Vershynin.

**Lemma 4.2.** [17, Proposition 4.1] *For every vector $\mathbf{x} \in S^{n-1}$ one has*
$$\mathcal{L}(B^T\mathbf{x}, c\sqrt{N}) = \sup_{\mathbf{u}} \mathbf{P}(\|B^T\mathbf{x}-\mathbf{u}\|_2 \le c\sqrt{N}) \le \exp(-cn).$$

*Proof.* (of Lemma 4.2) we first decompose the set $[N]$ into two sets $\{1,\dots,n/2\}\cup\{n/2+1,\dots,N\}$. This induces the decomposition of $B$ and $\mathbf{x}=(\mathbf{y},\mathbf{z}), \mathbf{u}=(\mathbf{v},\mathbf{w})$ accordingly

$$B = \begin{pmatrix} D & G \\ G' & E \end{pmatrix}.$$

Thus

$$\|B\mathbf{x}-\mathbf{u}\|_2^2 = \|D\mathbf{y}+G\mathbf{z}-\mathbf{v}\|_2^2 + \|G'\mathbf{y}+E\mathbf{z}-\mathbf{w}\|_2^2.$$

Using the fact that the matrix $G$ has independent entries,

$$\|D\mathbf{y}+G\mathbf{z}-\mathbf{v}\|_2^2 = \sum_{i=1}^{n/2}(\langle \mathbf{r}_i(G),\mathbf{z}\rangle - d_i)^2,$$

where $\mathbf{r}_i(G)$ is the rows of $G$ and $d_i$ denote the coordinates of the fixed vector $D\mathbf{y}-\mathbf{v}$. As $\langle \mathbf{r}_i(G),\mathbf{z}\rangle$ is a linear form in the random variables,

$$\mathcal{L}(\langle \mathbf{r}_i(G), \mathbf{z}/\|\mathbf{z}\|_2\rangle, 1/2) \le c_3 \in (0,1).$$

Using tenzorization trick (see for instance [8, Lemma 2.2]), we thus conclude

$$\mathcal{L}(G\mathbf{z}, c_2\|\mathbf{z}\|_2\sqrt{n/2}) \le c_0^{n/2}.$$

This implies that

$$\mathbf{P}(\|D\mathbf{y}+G\mathbf{z}-\mathbf{v}\|_2 \le c_2\|\mathbf{z}\|_2\sqrt{n/2}) \le c_3^{n/2}.$$

Argue similarly,

$$\mathbf{P}(\|G'\mathbf{y}+E\mathbf{z}-\mathbf{w}\|_2 \le c_2\|\mathbf{y}\|_2\sqrt{N-n/2}) \le c_3^{N-n/2}.$$

The proof is complete by noting that $\|\mathbf{y}\|_2^2+\|\mathbf{z}\|_2^2=\|\mathbf{x}\|_2^2=1$.

□

Now we bound uniformly over all compressible vectors.

**Lemma 4.3.** [17, Proposition 4.2] *We have*

$$\mathbf{P}\Big\{\inf_{\mathbf{x}/\|\mathbf{x}\|_2\in \mathbf{Comp}(c_0,c_1)} \|B\mathbf{x}-\mathbf{u}\|_2 \le c\sqrt{N}\|\mathbf{x}\|_2 \cap \mathcal{E}_K\Big\} \le e^{-cn/2}.$$

*Proof.* (of Lemma 4.3) Following the proof of [17, Proposition 4.2], there exists a $(2c_1)$-net $\mathcal{N}$ of the set $\mathbf{Comp}(c_0,c_1)$ such that

$$\mathcal{N} \le (9/c_0c_1)^{c_0n}.$$

It is not hard to show that if there exist $\mathbf{x}/\|\mathbf{x}\|_2 \in \mathbf{Comp}(c_0,c_1)$ with $\|B\mathbf{x}-\mathbf{u}\|_2 \le c\sqrt{N}\|\mathbf{x}\|_2$ and assuming $\mathcal{E}_K$, then there exists $\mathbf{x}_0 \in \mathcal{N}$ such that

$$\|B\mathbf{x}_0-\mathbf{v}_0\|_2 \le c\sqrt{N}$$

for some $\mathbf{v}_0$.

The proof is complete by noting that

$$(9/c_0c_1)^{c_0n}\frac{20}{c_1}2\exp(-cn) \le \exp(-cn/2),$$

by choosing $c_0$ small enough depending on $c_1$ and $c$. □

## 5. Proof of Theorem 1.4: treatment for incompressible vectors

Let $\mathbf{x}_1,\dots,\mathbf{x}_n \in \mathbf{R}^N$ denote the columns of the matrix $B$. Given a subset $J \subset [n]^d$, where $d=\delta m=\delta(N-n)$ for some sufficiently small $\delta$ to be chosen, we consider the subspace

$$H_J := span(\mathbf{x}_i)_{i\in J}.$$

Define

$$\mathbf{spread}_J := \Big\{\mathbf{y}\in S(\mathbf{R}^J) : K_1/\sqrt{d}\le |y_k|\le K_2/\sqrt{d}, k\in J\Big\}.$$

In what follows $J$ is a subset chosen randomly uniformly among the subsets of cardinality $d$ in $[n]$ and $P_J$ is the projection onto the coordinates indexed by $J$.

**Lemma 5.1.** [9, Lemma 6.1] *For every $c_0, c_1 \in (0,1)$, there exist $K_1, K_2, c > 0$ which depend only on $c_0, c_1$ such that the following holds. For every $\mathbf{x}\in \mathbf{Incomp}(c_0, c_1)$, the event*

$$E(\mathbf{x}) := \Big\{P_J\mathbf{x}/\|P_J\mathbf{x}\|_2 \in \mathbf{spread}_J \text{ and } c_1\sqrt{d}/\sqrt{2N}\le \|P_J\mathbf{x}\|_2 \le \sqrt{d}/\sqrt{c_0N}\Big\}$$

*satisfies*

$$\mathbf{P}_J(E(\mathbf{x})) > c^d.$$

Allow us to insert here a short proof of this fact. We will need the following simple property of incompressible vectors.

**Claim 5.2.** [9, Lemma 2.5] *Let $\mathbf{x}\in\mathbf{Incomp}(c_0, c_1)$. Then there exists a set $\sigma=\sigma(\mathbf{x})\subset[n]$ of cardinality $|\sigma|\ge c_0c_1^2n/2$ such that*

$$c_1/\sqrt{2n}\le |x_k|\le 1/\sqrt{c_0n}, k\in\sigma.$$

*Proof.* (of Lemma 5.1) Let $\sigma\subset[n]$ be the subset from Claim 5.2. Then as $d\le m\ll|\sigma|$,

$$\mathbf{P}_J(J\subset\sigma) = \binom{|\sigma|}{d}\Big/\binom{n}{d} > (c_0c_1^2/2e)^d := c^d.$$

It is clear that if $J\subset\sigma$ then we obtain the two-sided bound for $P_J\mathbf{x}$ with $K_1 = c_1\sqrt{c_0/2}, K_2 = 1/K_1, c = c_0c_1^2/2e$.

□

We now pass our estimate to $\mathbf{spread}_J$.

**Lemma 5.3.** *Let $c_0, c_1\in(0,1)$. There exist $C, c>0$ which depend only on $c_0, c_1$ such that the following holds. Then for any $\varepsilon>0$*

$$\mathbf{P}\Big(\inf_{\mathbf{x}\in\mathbf{Incomp}(c_0,c_1)}\|B\mathbf{x}\|_2 < c\varepsilon\sqrt{\frac{d}{n}}\Big) \le C^d \max_{J\in[n]^d}\mathbf{P}\Big(\inf_{\mathbf{z}\in\mathbf{spread}_J}\mathrm{dist}(B\mathbf{z}, H_{J^c})<\varepsilon\Big),$$

Where $H_{J^c}$ is the subspace generated by the columns of $B$ indexed by $J^c$.

We remark that there is a slight difference between this result and Lemma [9, Lemma 6.2] in that we take the supremum over all choices of $J$, as in this case the distance estimate for each $J$ is not identical.

Note that the following proof gives $K_1 = c_1\sqrt{c_0/2}, K_2 = 1/K_1, c = c_2/\sqrt{2}, C = 2e/c_0c_1^2$.

*Proof.* (of Lemma 5.3) We follow the proof of [9, Lemma 6.2]. Let $\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)$. For every $J$ we have

$$\|B\mathbf{x}\|_2 \geq \operatorname{dist}(B\mathbf{x}, H_{J^c}) = \operatorname{dist}(BP_J\mathbf{x}, H_{J^c}).$$

Condition on $E(\mathbf{x})$ of Lemma 5.1, we have $\mathbf{z} = P_J\mathbf{x}/\|\mathbf{P}_J\mathbf{x}\|_2 \in \mathbf{spread}_J$, and

$$\|B\mathbf{x}\|_2 \geq \|P_J\mathbf{x}\|_2 \times \inf_{\mathbf{z}\in\mathbf{spread}_J} \operatorname{dist}(B\mathbf{z}, H_{J^c}) = \|P_J\mathbf{x}\|_2 D(B, J)$$

with

$$D(B, J) := \inf_{\mathbf{z}\in\mathbf{spread}_J} \operatorname{dist}(B\mathbf{z}, H_{J^c}).$$

Thus on $E(\mathbf{x})$,

$$\|B\mathbf{x}\|_2 \geq (c\sqrt{d/N})D(B, J).$$

Define the event

$$\mathcal{F} := \{B : P_J(D(B, J) \geq \varepsilon) > 1 - c^d\}. \tag{21}$$

Markov's inequality then implies that

$$\begin{aligned}\mathbf{P}_B(\mathcal{F}^c) \leq c^{-d}\mathbf{E}_B\mathbf{P}_J(D(B, J) < \varepsilon) &\leq c^{-d}\mathbf{E}_J\mathbf{P}_B(D(B, J) < \varepsilon)\\ &\leq c^{-d}\max_J P_B(D(B, J) < \varepsilon).\end{aligned}$$

Fix any realization of $B$ for which $\mathcal{F}$ holds, and fix any $\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)$. Then

$$P_J(D(B, J) \geq \varepsilon) + \mathbf{P}_J(E(\mathbf{x})) \geq (1 - c^d) + c^d > 1.$$

Thus for any $\mathbf{x}$ there exists $J$ such that $E(\mathbf{x})$ and $D(B, J) \geq \varepsilon$. We then conclude from (21) that for any $B$ for which $\mathcal{F}$ holds,

$$\inf_{\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)} \|B\mathbf{x}\|_2 \geq \varepsilon c \sqrt{d/n},$$

completing the proof. □

By Lemma 5.3, we need to study $\mathbf{P}(\inf_{\mathbf{z} \in \mathbf{spread}_J} \mathrm{dist}(B\mathbf{z}, H_{J^c}) < \varepsilon)$ for any fixed $J \subset [n]^d$. From now on we assume that $J = \{N - d + 1, \dots, N\}$ as the other cases can be treated similarly. We restate the problem below.

**Theorem 5.4.** *Let $B$ be the matrix of the last $n$ columns of $A$, and let $J = \{N - d + 1, \dots, N\}$. Then*

$$\mathbf{P}\Big(\inf_{\mathbf{z} \in \mathbf{spread}_J} \mathrm{dist}(B\mathbf{z}, H_{J^c}) \leq \varepsilon\Big) \leq \varepsilon^{cm} + O(\exp(-n^{\varepsilon_0})),$$

*for some absolute constanst $c, \varepsilon_0 > 0$.*

Notice that $H_{J^c}$ has co-dimension $N - (n - d) = m + d$ in $\mathbf{R}^N$, thus $\varepsilon^{\Theta(m)}$ is expected in the RHS in Theorem 5.4.

As there is still dependence between $B\mathbf{z}$ and $H_{J^c}$, we will delete the last $m$ rows from $B$ to arrive at a matrix $B'$ of size $N - d$ by $n$.

That is

$$B = \begin{pmatrix} \mathbf{c}_{m+1}(A) & \dots & \mathbf{c}_N(A) \end{pmatrix} = \begin{pmatrix} B' \\ \mathbf{r}_{N-d+1} \\ \dots \\ \mathbf{r}_N \end{pmatrix}$$

and

$$B' = \begin{pmatrix} a_{1(m+1)} & a_{1(m+2)} & \dots & a_{1N} \\ a_{2(m+1)} & a_{1(m+2)} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{(N-d)(m+1)} & a_{(N-d)(m+2)} & \dots & a_{(N-d)N} \end{pmatrix}.$$

We now ignore the contribution of distances from the last $d$ rows.

**Claim 5.5.** *Assume that $B$ is as in Theorem 5.6, and $B'$ is obtained from $B$ by deleting its last $d$ rows. Then for any $\mathbf{z} \in \mathbf{spread}_J$,*

$$\mathrm{dist}(B\mathbf{z}, H_{J^c}) \geq \mathrm{dist}(B'\mathbf{z}, H_{J^c}(B')),$$

*where $H_{J^c}(B')$ is the subspace spanned by the first $n - d$ columns of $B'$.*

By Claim 5.5, to prove Theorem 5.4, it suffices to show

$$\mathbf{P}\Big(\inf_{\mathbf{z}\in\mathbf{spread}_J} \operatorname{dist}(B'\mathbf{z}, H_{J^c}(B')) \le \varepsilon\Big) \le \varepsilon^{cm} + O(\exp(-n^{\varepsilon_0})).$$

Note that the matrix $B''$ generated by the columns of $B'$ indexed from $J^c$ has size $(N-d)\times(n-d)$, and thus $H_{J^c}(B')$ has co-dimension $(N-d)-(n-d) = m$ in $\mathbf{R}^{N-d}$. Also

$$B'' := \left( \begin{array}{c} D \\ G \end{array} \right),$$

where $G$ is a random symmetric matrix (inherited from $A$) of size $N-m-d = n-d$, and $D$ is a matrix of size $m\times(n-d)$,

$$B'' = \begin{pmatrix} a_{1(m+1)} & a_{1(m+2)} & \dots & a_{1(N-d)} \\ a_{2(m+1)} & a_{1(m+2)} & \dots & a_{2(N-d)} \\ \dots & \dots & \dots & \dots \\ & & & \\ a_{(m+1)(m+1)} & a_{1(m+2)} & \dots & a_{(m+1)(N-d)} \\ \dots & \dots & \dots & \dots \\ a_{(N-d)(m+1)} & a_{(N-d)(m+2)} & \dots & a_{(N-d)(N-d)} \end{pmatrix}.$$

As for any fixed $\mathbf{z}\in\mathbf{spread}_J$, the vector $B'\mathbf{z}$ is independent of $H_{J^c}(B')$, and that the entries are iid copies of a random variable having the same subgaussian property as our original setting. Our next task is to prove another variant of the distance problem in contrast to Theorem 1.2.

**Theorem 5.6** (Distance problem, lower bound)**.** *Let $B'$ be as above. Let $\mathbf{x} = (x_1, \dots, x_{N-d})$ be a random vector where $x_i$ are iid copies of a subgaussian random variable of mean zero and variance one and are independent of $H_{J^c}(B')$), then*

$$\mathbf{P}\Big(\operatorname{dist}(\mathbf{x}, H_{J^c}(B')) \le \varepsilon\sqrt{m}\Big) \le \varepsilon^{cm} + O(\exp(-n^{\varepsilon_0})),$$

*for some absolute positive constants $c, \varepsilon_0$.*

We remark that the upper bound of distance is now $\varepsilon\sqrt{m}$ as we do not normalize $\mathbf{x}$. Notice that Theorem 5.6 is equivalent with

$$\mathbf{P}\Big(\operatorname{dist}(\mathbf{x}, H_{J^c}(B')) \le t\sqrt{d}\Big) \le (t/\sqrt{\delta})^{(c\delta^{-1})d} + O(\exp(-n^{\varepsilon_0})), \quad (22)$$

where

$$t := \varepsilon\sqrt{\delta^{-1}}.$$

We will prove Theorem 5.6 in Sections 6 and 7. Assuming it for now, we can pass back to $\mathbf{P}(\inf_{\mathbf{z}\in\mathbf{spread}_J}\mathrm{dist}(B\mathbf{z}, H_{J^c}) < \varepsilon)$ to complete the proof of Theorem 5.4. First of all, for short let $P$ be the projection onto $(H_{J^c}(B'))^{\perp}$, and let $W$ be the random matrix $W = PB'|_{\mathbf{R}^J}$. Notice that for $\mathbf{z} \in \mathbf{spread}_J$,

$$\mathrm{dist}(B'\mathbf{z}, (H_{J^c}(B'))^{\perp}) = W\mathbf{z}.$$

By Theorem 1.1

$$\mathbf{P}(\|W\|_2 \geq K\sqrt{d}) \leq \exp(-CK^2 d) \tag{23}$$

for any $K$ sufficiently large.

In what follows we will choose $K = C_0$ sufficiently large so that the RHS $\exp(-cK^2d)$ of (23) is much smaller than $C^{-m}$ from Lemma 5.3, for instance one can take

$$C_0 = (c^{-1})\log Cm/d = (c^{-1})(\log C)\delta^{-1}. \tag{24}$$

**Lemma 5.7.** [9, Lemma 7.4] *Assume that $W$ is the projection $W = PB|_{\mathbf{R}^J}$, then for an $t \geq \exp(-N/d)$ we have*

$$\mathbf{P}\Big(\inf_{\mathbf{z}\in\mathbf{spread}_J}\|Wz\|_2 < t\sqrt{d} \text{ and } \|W\|_2 \leq K_0\sqrt{d}\Big) \leq (Ct)^{cm} + \exp(-K_0^2 d).$$

*Proof.* (of Lemma 5.7) Let $s = t/K_0$. There is an $s$-net $\mathcal{N}$ of $\mathbf{spread}_J \subset S(\mathbf{R}^J)$ of cardinality

$$|\mathcal{N}| \leq 2d(1+2/s)^{d-1} \leq 2d(3K_0/t)^{d-1}.$$

Consider the event

$$\mathcal{E} := \{\inf \|Wz\|_2 < 2t\sqrt{d}\}.$$

Taking the union bound, and using the equivalent form (22) of Theorem 5.6, we have

$$\begin{aligned}\mathbf{P}(\mathcal{E}) \leq |\mathcal{N}|\max_{\mathbf{z}\in\mathcal{N}}\mathbf{P}(\|Wz\|_2 \leq 2t\sqrt{d}) &\leq 2d(3K_0/t)^{d-1}\Big[(2t/\sqrt{\delta})^{(c\delta^{-1})d} + O(\exp(-n^{\varepsilon_0}))\Big]\\ &\leq (Ct)^{cm/2},\end{aligned}$$

provided that $\delta$ was chosen so that $c\delta^{-1} \gg 1$, and the constant $C$ depends only $\delta, K_0$, and $\exp(-n^{\varepsilon_0}) \leq t^d$.

Now suppose the event stated in Lemma 5.7 holds, i.e. there exists $\mathbf{z}' \in \mathbf{spread}_J$ such that

$$\|W\mathbf{z}'\|_2 \le t\sqrt{d} \text{ and } \|W\|_2 \le K_0\sqrt{d}.$$

Choose $\mathbf{z} \in \mathcal{N}$ such that $\|\mathbf{z}-\mathbf{z}'\|_2 \le \varepsilon$. Then by the triangle inequality,

$$\|W\mathbf{z}\|_2 \le \|W\mathbf{z}'\|_2 + \|W\|_2\|\mathbf{z}-\mathbf{z}'\|_2 \le t\sqrt{d} + K_0\sqrt{d}(t/K_0) \le 2t\sqrt{d}.$$

□

## 6. Proof of Theorem 5.6: preparation

Without loss of generality, we restate the result below by changing $n$ to $N$ and $d$ to $m$.

**Theorem 6.1** (Distance problem, again)**.** *Let $B$ be the matrix obtained from $A$ by removing its first $m$ columns. Let $H$ be the subspace generated by the columns of $B$, and let $\mathbf{x} = (x_1, \dots, x_N)$ be a random vector independent of $B$ whose entries are iid copies of a subgaussian random variable of mean zero and variance one, then*

$$\mathbf{P}(\mathrm{dist}(\mathbf{x}, H) \le \varepsilon\sqrt{m}) \le \varepsilon^{\delta m} + O(\exp(-N^{\varepsilon_0})),$$

*for some absolute constants $\delta, \varepsilon_0 > 0$.*

After discovering Theorem 6.1, the current author had found that this result is essentially [12, Theorem 8.1]. As the proof here look a bit shorter (mainly thanks to the simplicity of our model), we decide to sketch here for the sake of completion.

Recall that for any random variable $S$, then the Levy concentration of radius $r$ (or small ball probability of radius $r$) is defined by

$$\mathcal{L}(S, r) = \sup_s \mathbf{P}(|S - s| \le r).$$

6.2. **The least common denominator.** Let $\mathbf{x} = (x_1, \dots, x_N)$. Rudelson and Vershynin [9] defined the essential *least common denominator* (**LCD**) of $\mathbf{x} \in \mathbf{R}^N$ as follows. Fix parameters $\alpha$ and $\gamma$, where $\gamma \in (0,1)$, and define

$$\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}) := \inf\Big\{\theta > 0 : \mathrm{dist}(\theta\mathbf{x}, \mathbf{Z}^N) < \min(\gamma\|\theta\mathbf{x}\|_2, \alpha)\Big\}.$$

We remark that for convenience we do not require $\|\mathbf{x}\|_2$ to be larger than 1, and it follows from the definition that for any $\delta > 0$,

$$\mathbf{LCD}_{\alpha,\gamma}(\delta\mathbf{x}) \le \delta^{-1}\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}).$$

**Theorem 6.3.** [9] *Consider a vector* $\mathbf{x} \in \mathbf{R}^N$ *which satisfies* $\|\mathbf{x}\|_2 \geq 1$. *Then, for every* $\alpha > 0$ *and* $\gamma \in (0,1)$, *and for*

$$\varepsilon \geq \frac{1}{\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x})},$$

*we have*

$$\mathcal{L}(S, \varepsilon) \leq C_0(\frac{\varepsilon}{\gamma} + e^{-2\alpha^2}),$$

*where* $C_0$ *is an absolute constant depending on the sub-gaussian parameter of* $\xi$.

The definition of the essential least common denominator above can be extended naturally to higher dimensions. To this end, consider $d$ vectors $\mathbf{x}_1 = (x_{11}, \dots, x_{1N}), \dots, \mathbf{x}_m = (x_{m1}, \dots, x_{mN}) \in \mathbf{R}^N$. Define $\mathbf{y}_1 = (x_{11}, \dots, x_{m1}), \dots, \mathbf{y}_n = (x_{1N}, \dots, x_{mN})$ be the corresponding vectors in $\mathbf{R}^m$. Then we define, for $\alpha > 0$ and $\gamma \in (0,1)$,

$$\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}_1, \dots, \mathbf{x}_m)$$

$$:= \inf\Big\{\|\Theta\|_2 : \ \Theta \in \mathbf{R}^m, \mathrm{dist}((\langle\Theta, \mathbf{y}_1\rangle, \dots, \langle\Theta, \mathbf{y}_N\rangle), \mathbf{Z}^N) < \min(\gamma\|(\langle\Theta, \mathbf{y}_1\rangle, \dots, \langle\Theta, \mathbf{y}_N\rangle)\|_2, \alpha)\Big\}.$$

The following generalization of Theorem 6.3 gives a bound on the small ball probability for the random sum $S = \sum_{i=1}^N a_i \mathbf{y}_i$, where $a_i$ are iid copies of $\xi$, in terms of the additive structure of the coefficient sequence $\mathbf{x}_i$.

**Theorem 6.4** (Diophatine approximation, multi-dimensional case)**.** [9] *Consider d vectors* $\mathbf{x}_1, \dots, \mathbf{x}_m$ *in* $\mathbf{R}^N$ *which satisfies*

$$\sum_{i=1}^N \langle \mathbf{y}_i, \Theta\rangle^2 \geq \|\Theta\|_2^2 \qquad \text{for every } \Theta \in \mathbf{R}^m, \tag{25}$$

*where* $\mathbf{y}_i = (x_{i1}, \dots, x_{im})$. *Let* $\xi$ *be a random variable such that* $\sup_a \mathbf{P}(\xi \in B(a,1)) \leq 1 - b$ *for some* $b > 0$ *and* $a_1, \dots, a_N$ *be iid copies of* $\xi$. *Then, for every* $\alpha > 0$ *and* $\gamma \in (0,1)$, *and for*

$$\varepsilon \geq \frac{\sqrt{m}}{\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}_1, \dots, \mathbf{x}_m)},$$

*we have*

$$\mathcal{L}(S, \varepsilon\sqrt{m}) \leq \Big(\frac{C\varepsilon}{\gamma\sqrt{b}}\Big)^m + C^m e^{-2b\alpha^2}.$$

We next introduce the definition of LCD of a subspace.

**Definition 6.5.** Let $H \subset \mathbf{R}^N$ be a subspace. Then then LCD of $H$ is defined to be

$$\mathbf{LCD}_{\alpha,\gamma}(H) := \inf_{\mathbf{y}_0 \in H, \|\mathbf{y}_0\|_2 = 1} \mathbf{LCD}_{\alpha,\gamma}(\mathbf{y}_0).$$

In what follows we prove some useful results regarding this $LCD$.

**Lemma 6.6.** *Assume that* $\|\mathbf{x}_1\|_2 = \cdots = \|\mathbf{x}_m\|_2 = 1$. *Let* $H \subset \mathbf{R}^N$ *be the subspace generated by* $\mathbf{x}_1, \dots, \mathbf{x}_m$. *Then*

$$\sqrt{m}\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}_1, \dots, \mathbf{x}_m) \geq \mathbf{LCD}_{\alpha,\gamma}(H).$$

*Proof.* (of Lemma 6.6) Assume that

$$\mathrm{dist}((\langle \Theta, \mathbf{y}_1 \rangle, \dots, \langle \Theta, \mathbf{y}_N \rangle), \mathbf{Z}^N) < \min(\gamma \|(\langle \Theta, \mathbf{y}_1 \rangle, \dots, \langle \Theta, \mathbf{y}_N \rangle)\|_2, \alpha).$$

Set $\mathbf{y}_0 := \frac{1}{t}(\theta_1 \mathbf{x}_1 + \cdots + \theta_m \mathbf{x}_m)$ where $t$ is chosen so that $\|\mathbf{y}_0\|_2 = 1$. By definition

$$\begin{aligned} \mathrm{dist}(t\mathbf{y}_0, \mathbf{Z}^N) &= \mathrm{dist}((\langle \Theta, \mathbf{y}_1 \rangle, \dots, \langle \Theta, \mathbf{y}_N \rangle), \mathbf{Z}^N) \\ &< \min(\gamma \|(\langle \Theta, \mathbf{y}_1 \rangle, \dots, \langle \Theta, \mathbf{y}_N \rangle)\|_2, \alpha) \\ &= \min(\gamma \|t\mathbf{y}_0\|_2, \alpha). \end{aligned}$$

On the other hand, as $\|\mathbf{x}_i\|_2 = 1$, one has

$$\|\theta_1 \mathbf{x}_1 + \cdots + \theta_m \mathbf{x}_m\|_2 \leq |\theta_1| + \cdots + |\theta_m| \leq \sqrt{m}\|\Theta\|_2.$$

So,

$$t \leq \sqrt{m}\|\Theta\|_2.$$

Hence,

$$\mathbf{LCD}_{\alpha,\gamma}(\mathbf{y}_0) \leq \sqrt{m}\mathbf{LCD}_{\alpha,\gamma}(\mathbf{x}_1, \dots, \mathbf{x}_m).$$

□

**Corollary 6.7.** *Let* $H \subset \mathbf{R}^N$ *be a subspace of co-dimension* $m$ *such that* $\mathbf{LCD}(H^\perp) \geq D$ *for some* $D$. *Let* $\mathbf{a} = (a_1, \dots, a_N)$ *be a random vector where* $a_i$ *are iid copies of* $\xi$. *Then for any* $\varepsilon \geq m/D$

$$\mathbf{P}(\mathrm{dist}(\mathbf{a}, H) \leq \varepsilon\sqrt{m}) \leq \Big(\frac{C\varepsilon}{\gamma\sqrt{b}}\Big)^m + C^m e^{-2b\alpha^2}.$$

*Proof.* (of Corollary 6.7) Let $\mathbf{e}_1, \dots, \mathbf{e}_m$ be an orthogonal basis of $H^\perp$ and let $M$ be the matrix of size $m \times N$ generated by these vectors. By Lemma 6.6,

$$\mathbf{LCD}_{\alpha,\gamma}(\mathbf{e}_1, \dots, \mathbf{e}_m) \geq D/\sqrt{m}.$$

Also, by definition

$$\mathrm{dist}(\mathbf{a}, H) = \|M\mathbf{a}\|_2.$$

Thus by Theorem 6.4, for $\varepsilon \geq \frac{m}{D}$, we have

$$\mathcal{L}(M\mathbf{a}, \varepsilon\sqrt{m}) \leq \Big(\frac{C\varepsilon}{\gamma\sqrt{b}}\Big)^m + C^m e^{-2b\alpha^2}.$$

□

Now we discuss another variant of arithmetic structure which will be useful for matrices of correlated entries.

6.8. **Regularized LCD.** Let $\mathbf{x} = (x_1, \dots, x_N)$ be a unit vector. Let $c_*, c_0, c_1$ be given constants. We assign a subset $\mathbf{spread}(\mathbf{x})$ so that for all $k \in \mathbf{spread}(\mathbf{x})$,

$$\frac{c_0}{\sqrt{N}} \leq |x_k| \leq \frac{c_1}{\sqrt{N}}.$$

Following Vershynin [17] (see also [7]), we define another variant of LCD as follows.

**Definition 6.9** (Regularized LCD)**.** Let $\lambda \in (0, c_*)$. We define the *regularized LCD* of a vector $\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)$ as

$$\widehat{\mathbf{LCD}}_{\alpha,\gamma}(\mathbf{x}, \lambda) = \max\Big\{\mathbf{LCD}_{\alpha,\gamma}\big(\mathbf{x}_I/\|x_I\|_2\big) : I \subseteq \mathbf{spread}(\mathbf{x}), |I| = \lceil \lambda N \rceil\Big\}.$$

We will denote by $I(\mathbf{x})$ the maximizing set $I$ in this definition.

Note that in our later application $\lambda$ can be chosen within $n^{-\lambda_0} \leq \lambda \leq \lambda_0$ for some sufficiently small constant $\lambda_0$.

From the definition, it is clear that if $\mathbf{LCD}(\mathbf{x})$ small then so is $\widehat{\mathbf{LCD}}(\mathbf{x})$ (with slightly different parameter).

**Lemma 6.10.** *For any $x \in S^{N-1}$ and any $0 < \gamma < c_1\sqrt{\lambda}/2$, we have*

$$\widehat{\mathbf{LCD}}_{\alpha,\gamma(c_1\sqrt{\lambda}/2)^{-1}}(\mathbf{x}, \lambda) \leq \frac{1}{c_0}\sqrt{\lambda}\mathbf{LCD}_{\alpha,\gamma}(x).$$

*Consequently, for any $0 < \gamma < 1$*

$$\widehat{\mathbf{LCD}}_{\kappa,\gamma}(x,\alpha) \le \frac{1}{c_0}\sqrt{\alpha}\mathbf{LCD}_{\kappa,\gamma(c_1\sqrt{\alpha}/2)}(x).$$

*Proof.* (of Lemma 6.10) See [7, Lemma 5.7]. □

We now introduce a result connecting the small ball probability with the regularized LCD.

**Lemma 6.11.** *Assume that*

$$\varepsilon \ge \frac{1}{c_1}\sqrt{\lambda}(\widehat{\mathbf{LCD}}_{\alpha,\gamma}(\mathbf{x},\lambda))^{-1}.$$

*Then we have*

$$\mathcal{L}(S,\varepsilon) = O\left(\frac{\varepsilon}{\gamma c_1\sqrt{\lambda}} + e^{-\Theta(\alpha^2)}\right).$$

*Proof.* (of Lemma 6.11) See for instance [7, Lemma 5.8]. □

## 7. Estimating additive structure and completing the proof of Theorem 6.1

Again, we will be following [9, 17] with modifications. A major part of this treatment can also be found in [7, Appendix B] but allow us to recast here for completion.

We first show that with high probability $H^\perp$ does not contain any compressible vector, where we recall that $H$ is spanned by the column vectors of $B$.

**Theorem 7.1** (Incompressible of subspace)**.** *Consider the event $\mathcal{E}_1$,*

$$\mathcal{E}_1 := \{H^\perp \cap \mathbf{Comp}(c_0, c_1) = \emptyset\}.$$

*We then have*

$$\mathbf{P}(\mathcal{E}_1^c) \le \exp(-cn).$$

The treatment is similar to Section 4 except the fact that we are working with $B^T$ and vectors in $\mathbf{R}^N$. We start with a version of Lemma 4.2.

**Lemma 7.2.** *For every $c_0$-sparse vector $\mathbf{x} \in S^{N-1}$ one has*

$$\mathcal{L}(B^T\mathbf{x}, c\sqrt{N}) = \sup_{\mathbf{u}} \mathbf{P}(\|B^T\mathbf{x} - \mathbf{u}\|_2) \le \exp(-cN).$$

*Proof.* (of Lemma 7.2) Without loss of generality, assume that the last $(1-c_0)N$ components of $\mathbf{x}$ are all zero. What remains is similar to the proof of Lemma 4.2. □

*Proof.* (of Theorem 7.1) First of all, there exists a $(2c_1)$-net $\mathcal{N}$ of sparse vectors only of the set $\mathbf{Comp}(c_0, c_1)$ such that

$$|\mathcal{N}| \le (9/c_0c_1)^{c_0 N}.$$

It is not hard to show that if there exist $\mathbf{x} \in \mathbf{Comp}(c_0, c_1)$ with $\|B^T\mathbf{x} - \mathbf{u}\|_2 \le cN^{1/2}\|\mathbf{x}\|_2$ and assuming $\mathcal{E}_K$, then there exists $\mathbf{x}_0 \in \mathcal{N}$ such that

$$\|B^T\mathbf{x}_0 - \mathbf{v}_0\|_2 \le c\sqrt{N}$$

for some $\mathbf{v}_0$.

This leaves us to estimate the probability $\mathbf{P}(\|B^T\mathbf{x}_0 - \mathbf{v}_0\|_2 \le c\sqrt{N})$ for each individual sparse vector $\mathbf{x}_0$, and for this it suffices to apply Lemma 7.2. □

The main goal of this section is to verify the following result.

**Theorem 7.3** (Structure theorem)**.** *Consider the event* $\mathcal{E}_2$

$$\mathcal{E}_2 := \{\forall \mathbf{y}_0 \in H^\perp : \widehat{\mathbf{LCD}}_{\alpha,c}(\mathbf{y}_0, \lambda) \ge N^{c/\lambda}\}.$$

*We then have*

$$\mathbf{P}(\mathcal{E}_2^c) \le \exp(-cN).$$

Notice that in this result, $c = \gamma(c_0\sqrt{\lambda})^{-1}$. Assume Theorem 7.3 for the moment, we provide a proof of our distance theorem.

*Proof.* (of Theorem 6.1) Within $\mathcal{E}_2$, $\widehat{\mathbf{LCD}}(\mathbf{y}_0)$ is extremely large, and so $\mathbf{LCD}(H^\perp)$ is also large because of Lemma 6.6 (where the factor $\sqrt{m}$ is absorbed into $N^{c/\lambda}$). We then apply Theorem 6.4 to complete the proof. □

7.4. **Proof of Theorem 7.3.** (See also [17] and [7, Appendix B]). The first step is to show that the set of vectors of small $\widehat{\mathbf{LCD}}$ accepts a net of considerable size.

**Lemma 7.5.** *Let* $\lambda \in (c/N, c_*)$*. For every* $D \ge 1$*, the subset* $\{\mathbf{x} \in \mathbf{Incomp}(c_0, c_1) : \widehat{\mathbf{LCD}}_{\alpha,c}(\mathbf{x}, \lambda) \le D\}$ *has a* $\alpha/D\sqrt{\lambda}$*-net* $\mathcal{N}$ *of size*

$$|\mathcal{N}| \le [CD/(\lambda N)^c]^N D^{1/\lambda}.$$

**Definition 7.6.** Let $D_0 \geq \gamma_0\sqrt{N}$. Define $S_{D_0}$ as

$$S_{D_0} := \{\mathbf{x} \in \mathbf{Incomp} : D_0 \leq \mathbf{LCD}_{\alpha,c}(\mathbf{x}) \leq 2D_0\},$$

where $\gamma_0$ is a constant.

**Lemma 7.7.** [9, Lemma 4.7] *There exists a $(4\alpha/D_0)$-net of $S_{D_0}$ of cardinality at most $(C_0D_0/\sqrt{N})^N$.*

One can in fact obtain a more general form as follows .

**Lemma 7.8.** *Let $c \in (0,1)$ and $D \geq D_0 \geq c\sqrt{N}$. Then the set $S_{D_0}$ has a $(4\alpha/D)$-net of cardinality at most $(C_0D/\sqrt{N})^N$.*

*Proof.* (of Lemma 7.8) First, by the lemma above one can cover $S_{D_0}$ by $(C_0D_0/\sqrt{N})^N$ balls of radius $4\alpha/D_0$. We then cover these balls by smaller balls of radius $4\alpha/D$, the number of such small balls is at most $(5D/D_0)^N$. Thus in total there are at most $(20C_0D/\sqrt{N})^N$ balls in total. □

Now we put the nets together over dyadic intervals.

**Lemma 7.9.** *Let $c \in (0,1)$ and $D \geq c\sqrt{N}$. Then the set $\{X \in \mathbf{Incomp}(c_0, c_1) : c\sqrt{N} \leq \mathbf{LCD}_{\alpha,c}(X) \leq D\}$ has a $(4\alpha/D)$-net of cardinality at most $(C_0D/\sqrt{N})^N \log_2 D$.*

Notice that in the above lemmas, $\|\mathbf{x}\|_2 \geq 1$ was assumed implicitly. Using the trivial bound $\log_2 D(D/\alpha) \leq D^2$, we arrive at

**Lemma 7.10.** *Let $c \in (0,1)$ and $D \geq c\sqrt{N}$. Then the set $\{\mathbf{x} \in \mathbf{Incomp}(c_0, c_1) : c\sqrt{N} \leq \mathbf{LCD}_{\alpha,c}(Bx/\|\mathbf{x}\|_2) \leq D\}$ has a $(4\alpha/D)$-net of cardinality at most $(C_0D/\sqrt{N})^N D^2$.*

*Proof.* (of Lemma 7.5) Write $\mathbf{x} = \mathbf{x}_{I_0} \cup \mathbf{spread}(\mathbf{x})$, where $\mathbf{spread}(\mathbf{x}) = I_1 \cup \cdots \cup I_{k_0} \cup J$ such that $|I_k| = \lambda N$ and $|J| \leq \lambda N$.

Notice that we trivially have

$$|\mathbf{spread}(x)| \geq |I_1 \cup \cdots \cup I_{k_0}| = k_0\lceil \alpha n\rceil \geq |\mathbf{spread}(x)| - \alpha n \geq c' n/2.$$

Thus we have

$$\frac{c'}{2\alpha} \leq k_0 \leq \frac{2c'}{\alpha}.$$

In the next step, we will construct nets for each $x_{I_j}$. For $x_{I_0}$, we construct trivially a $(1/D)$-net $\mathcal{N}_0$ of size

$$|\mathcal{N}_0| \leq (3D)^{|I_0|}.$$

For each $I_k$, as

$$\mathbf{LCD}_{\kappa,\gamma}(x_{I_k}/\|x_{I_k}\|) \le \widehat{\mathbf{LCD}}_{\kappa,\gamma}(x) \le D,$$

by Lemma 7.10 (where the condition $\mathbf{LCD}_{\kappa,\gamma}(x_{I_k}/\|x_{I_k}\|) \gg \sqrt{|I_k|}$ follows the standard Littlewood-Offord estimate because the entries of $x_{I_k}/\|x_{I_k}$ are all of order $\sqrt{\alpha n}$ while $\kappa = o(\sqrt{\alpha n})$), one obtains a $(2\kappa/D)$-net $\mathcal{N}_k$ of size

$$|\mathcal{N}_k| \le \left(\frac{C_0 D}{\sqrt{|I_k|}}\right)^{|I_k|} D^2.$$

Combining the nets together, as $x = (x_{I_0}, x_{I_1}, \dots, x_{I_{k_0}}, x_J)$ can be approximated by $y = (y_{I_0}, y_{I_1}, \dots, y_{I_{k_0}}, y_J)$ with $\|x_{I_j} - y_{I_j}\| \le \frac{2\kappa}{D}$, we have

$$\|x - y\| \le \sqrt{k_0 + 1}\frac{2\kappa}{D} \ll \frac{\kappa}{\sqrt{\alpha} D}.$$

As such, we have obtain a $\beta$-net $\mathcal{N}$, where $\beta = O(\frac{\kappa}{\sqrt{\alpha} D})$, of size

$$|\mathcal{N}| \le 2^n |\mathcal{N}_0||\mathcal{N}_1| \dots |\mathcal{N}_{k_0}| \le 2^n (3D)^{|I_0|} \prod_{k=1}^{k_0} \left(\frac{CD}{\sqrt{|I_k|}}\right)^{|I_k|} D^2.$$

This can be simplified to

$$|\mathcal{N}| \le \frac{(CD)^n}{\sqrt{\alpha n}^{c'n/2}} D^{O(1/\alpha)}.$$

□

Now we complete the proof of Theorem 7.3 owing to Lemma 7.5 and the following bound for any fixed $\mathbf{x}$.

**Lemma 7.11.** [17, Proposition 6.11] *Let* $\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)$ *and* $\lambda \in (0, c_*)$. *Then for any* $\varepsilon > 1/\widehat{\mathbf{LCD}}_{\alpha,\gamma}(\mathbf{x})$ *one has*

$$\mathcal{L}(B^T \mathbf{x}, \varepsilon\sqrt{N}) \le (\frac{\varepsilon}{\gamma\sqrt{\lambda}} + \exp(-\alpha^2))^{N - \lambda N}.$$

*Proof.* (of Theorem 7.3) Assume that $D \le N^{c/\gamma}$. Then with $\beta = \alpha/(D\lambda) \ge 1/D$, by a union bound

$$\mathbf{P}\Big(\exists \mathbf{y}_0 \in H \subset S_D, \|B^T \mathbf{y}_0 - u\|_2 \le \beta\sqrt{N}\Big) \le \Big(\frac{\varepsilon}{\gamma\sqrt{\lambda}} + \exp(-\alpha^2)\Big)^{N - \lambda N} \times \frac{(CD)^N}{\sqrt{\lambda n}^{c_* N/2}} D^{2/\lambda} = N^{-cN}.$$

This completes the proof of our theorem. □

## 8. Application: proof of Corollary 1.5

For short, denote by $B$ the $(N-1)\times N$ matrix generated by $\mathbf{r}_2(A),\dots,\mathbf{r}_N(A)$. We will follow the approach of [11, 5]. Let $I=\{i_1,\dots,i_{m-1}\}$ be any subset of size $m-1$ of $\{2,\dots,N\}$, and let $H$ be the subspace generated by the remaining columns of $B$. Let $P_H$ be the projection from $\mathbf{R}^{N-1}$ onto the orthogonal complement $H^\perp$ of $H$. For now we view $P_H$ as an idempotent matrix of size $(N-1)\times(N-1)$, $P_H^2=P_H$. It is known (see for instance [17, 4]) that with probability $1-\exp(-N^c)$ we have $\dim(H^\perp)=m-1$. So without loss of generality we assume $\mathrm{tr}(P_H)=m-1$.

Recall that by definition,

$$x_1\mathbf{c}_1(B)+x_{i_1}\mathbf{c}_{i_1}(B)+\dots+x_{i_{m-1}}\mathbf{c}_{i_{m-1}}(B)+\sum_{i\notin\{1,i_1,\dots,i_{m-1}\}}x_i\mathbf{c}_i(B)=0. \qquad (26)$$

Thus, projecting onto $H^\perp$ would then yield

$$x_1P_H\mathbf{c}_1(B)+x_{i_1}P_H\mathbf{c}_{i_1}(B)+\dots+x_{i_{m-1}}P_H\mathbf{c}_{i_{m-1}}(B)=0.$$

It follows that

$$|x_1|\|P_H(\mathbf{c}_1(B))\|_2=\|x_{i_1}P_H(\mathbf{c}_{i_1}(B))+\dots+x_{i_{m-1}}P_H(\mathbf{c}_{i_{m-1}}(B))\|_2. \qquad (27)$$

Now if $m=C\log n$ with sufficiently large $C$, then by Theorem 1.2 the following holds with overwhelming probability (that is greater than $1-O(n^{-C})$ for any given $C$)

$$\|P_H\mathbf{c}_1(B)\|_2\asymp\sqrt{m} \text{ and } \|P_H\mathbf{c}_{i_j}(B)\|_2\asymp\sqrt{m}, 1\le j\le m-1;$$

and hence trivially

$$|\mathbf{c}_{i_{j_1}}^T(B)P_H\mathbf{c}_{i_{j_2}}(B)|\ll m, j_1\neq j_2.$$

Let $\mathcal{E}_I$ be this event, on which by Cauchy-Schwarz we can bound the square of the RHS of (27) by

$$\|x_{i_1}P_H(\mathbf{c}_{i_1})+\dots+x_{i_{m-1}}P_H(\mathbf{c}_{i_{m-1}})\|_2^2\ll m(\sum_{j=1}^{m-1}x_{i_j}^2)+m^2(\sum_{j=1}^{m-1}x_{i_j}^2).$$

Thus we obtain

$$|x_1| \ll m^{1/2}(\sum_{j=1}^{m-1} x_{i_j}^2)^{1/2}. \tag{28}$$

Now let $I_1, \dots, I_{\frac{n-1}{m-1}}$ be any partition of $\{2, \dots, n\}$ into subsets of size $m-1$ each (where for simplicity we assume $m-1|n-1$). Set

$$\mathcal{E} := \wedge_{1 \le j \le \frac{n-1}{m-1}} \mathcal{E}_{I_j}.$$

By a union bound, $\mathcal{E}$ holds with overwhelming probability. Furthermore, it follows from (28) that on $\mathcal{E}$,

$$|x_1| \le \min\Big\{m^{1/2}(\sum_{i \in I_j} x_i^2)^{1/2}, 1 \le j \le \frac{n-1}{m-1}\Big\}.$$

But as $\sum_j \sum_{i \in I_j} x_i^2 = 1 - x_1^2 < 1$, by the pigeon-hole principle

$$\min\{\sum_{i \in I_j} x_i^2, 1 \le j \le \frac{n-1}{m-1}\} \le \frac{m-1}{n-1}.$$

Thus conditioned on $\mathcal{E}$,

$$|x_1| \le m^{1/2}\sqrt{\frac{m-1}{n-1}} = O(\frac{(\log n)^{3/2}}{\sqrt{n}}).$$

The claim then follows by Bayes' identity.

## Appendix A. Proof of Lemma 3.3

We can rely on the powerful concentration result of eigenvalues inside the bulk for random Wigner matrices from [2], [16] or [3].

**Theorem A.1.** *Let $A$ be a random Wigner matrix as in Theorem 1.1. Let $\varepsilon, \delta$ be given positive constants. Then there exists a positive constant $\kappa$ such that the following holds with probability at least $1 - n^{-\omega}$: let $I$ be any interval of length $\log^{\kappa^{-1}} N/\sqrt{N}$ inside $[0, 2-\varepsilon]$, then the number $N_I$ of eigenvalues $\lambda_i$ with modulus $|\lambda_i| \in I$ is well concentrated*

$$|N_I - \int_{x \in I} \rho_{qc}(x)dx| \le \delta\sqrt{N}I,$$

*where $\rho_{qc}$ is the quarter-circle density.*

As a consequence, with probability at least $1 - N^{-\omega(1)}$, for any $\log^{\kappa'} N \ll m \ll N$, any interval $[x_0 + C_1 m/N^{1/2}, x_0 + C_2 m/N^{1/2}], x_0 \geq 0$ inside the bulk contains at least $2m$ and at most $C'm$ singular eigenvalues of $A$, where $C_1, C_2, C'$ depend on $\delta, \varepsilon, K_0$. Lemma 3.3 then can be obtained by iterating the interlacing law for singular values.

## References


[1] G. Bennett, L. E. Dor, V. Goodman, W. B. Johnson and C. M. Newman, On uncomplemented subspaces of $L_p, 1 < p < 2$, *Israel J. Math.* **26** (1977), 178-187.

[2] L. Erdős, B. Schlein and H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices, *International Mathematics Research Notices*, 2010, no. 3, 436-479.

[3] L. Erdős, H.-T. Yau and J. Yin, Rigidity of Eigenvalues of Generalized Wigner Matrices, *Advances in Mathematics*, **229** (2012), no. 3, 1435-1515.

[4] H. Nguyen, Inverse Littlewood-Offord problems and the singularity of random symmetric matrices, *Duke Mathematics Journal*, **161**, 4 (2012), 545-586.

[5] H. Nguyen and V. Vu, Random non-Hermitian matrices: normality of vectors, in preparation.

[6] H. Nguyen and V. Vu, Random matrices: law of the determinant, *Annals of Probability*, **42** (2014), no. 1, 146-167.

[7] H. Nguyen, V. Vu and T Tao, Random matrices: tail bounds for gaps between eigenvalues, to appear in Probability Theory and Related Fields, `arxiv.org/abs/1504.00396` .

[8] M. Rudelson and R. Vershynin, The Littlewood-Offord Problem and invertibility of random matrices, *Advances in Mathematics*, **218** (2008), 600-633.

[9] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, *Communications on Pure and Applied Mathematics*, **62** (2009), 1707-1739.

[10] M. Rudelson and R. Vershynin, Hanson-Wright inequality and sub-gaussian concentration, *Electronic Communications in Probability*, **18** (2013), 1-9.

[11] M. Rudelson and R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, *Duke Mathematical Journal*, to appear, `arXiv:1306.2887`.

[12] M. Rudelson and R. Vershynin, No-gaps delocalization for general random matrices, `http://arxiv.org/abs/1506.04012`.

[13] T. Tao, Topics in random matrix theory, *Graduate Studies in Mathematics*, **132**, American Mathematical Society, Providence, RI, 2012.

[14] T. Tao and V. Vu and appendix by M. Krishnapur, Random matrices: universality of ESDs and the circular law, *Annals of Probability*, **38**, (2010), no. 5, 2023-2065.

[15] T. Tao and V. Vu, Random Matrices: the Distribution of the Smallest Singular Values, *Geometric and Functional Analysis*, **20** (2010), no. 1, 260-297.

[16] T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics, *Acta Mathematica*, **206** (2011), 127-204.

[17] R. Vershynin, Invertibility of symmetric random matrices, *Random Structures & Algorithms*, **44** (2014), no. 2, 135-182.



Department of Mathematics, The Ohio State University, Columbus OH 43210

*E-mail address*: `nguyen.1261@math.osu.edu`