7. Proof of Theorem 3.2

We prove Theorem 3.2 in this section. The high level roadmap of the proof is a standard one: by convex analysis, to show that M is the unique optimal solution to (1), it suffices to construct a *dual certificate* Y obeying certain optimality conditions. One of the conditions requires the spectral norm $\|Y\|$ to be small. Previous work bounds $\|Y\|$ by the the ℓ_∞ norm $\|Y'\|_\infty := \sum_{i,j} \left|Y'_{ij}\right|$ of a certain matrix Y', which gives rise to the standard and joint incoherence conditions involving uniform bounds by μ_0 and μ_{str}. Here, we derive a new bound using the weighted $\ell_{\infty,2}$ norm of Y', which is the maximum of the weighted row and column norms of Y'. These bounds lead to a tighter bound of $\|Y\|$ and hence less restrictive conditions for matrix completion.

We now turn to the details. To simplify the notion, we prove the results for square matrices ($n_1 = n_2 = n$). The results for non-square matrices are proved in exactly the same fashion. A few additional notations are needed. We use c and its derivatives (c', c_0, etc) for universal positive constants, which may differ from place to place. By *with high probability* (*w.h.p.*) we mean with probability at least $1 - c_1 n^{-c_2}$. The inner product between two matrices is given by $\langle Y, Z\rangle = \text{trace}(Y^\top Z)$. Recall that U and V are the left and right singular vectors of the underlying matrix M. We need several standard projection operators for matrices. The projections P_T and $P_{T^\perp}$ are given by

$$P_T(Z) := UU^\top Z + ZVV^\top - UU^\top VZZ^\top$$

and $P_{T^\perp}(Z) := Z - P_T(Z)$. $P_\Omega(Z)$ is the matrix with $(P_\Omega(Z))_{ij} = Z_{ij}$ if $(i,j) \in \Omega$ and zero otherwise, and $P_{\Omega^c}(Z) := Z - P_\Omega(Z)$. As usual, $\|z\|_2$ is the ℓ_2 norm of the vector z, and $\|Z\|_F$ and $\|Z\|$ are the Frobenius norm and spectral norm of the matrix Z, respectively. For a linear operator $\mathcal{A}$ on matrices, its operator norm is defined as $\|\mathcal{A}\|_{op} = \sup_{X\in\mathbb{R}^{n\times n}} \|\mathcal{A}(X)\|_F / \|X\|_F$. For each $1 \le i,j \le n$, we define the random variable $\delta_{ij} := \mathbb{I}((i,j)\in\Omega)$, where $\mathbb{I}(\cdot)$ is the indicator function. The matrix operator $R_\Omega : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^{n\times n}$ is defined as

$$R_\Omega(Z) = \sum_{i,j} \frac{1}{p_{ij}} \delta_{ij} \left\langle e_i e_j^\top, Z\right\rangle e_i e_j^\top. \tag{9}$$

Optimality Condition. Following our proof roadmap, we now state a sufficient condition for M to be the unique optimal solution to the optimization problem (1). This is the content of Proposition 7.1 below (proved in Section 7.1).

Proposition 7.1. *Suppose $p_{ij} \ge \frac{1}{n^{10}}$. The matrix M is the unique optimal solution to (1) if the following conditions hold.*

1. $\|P_T R_\Omega P_T - P_T\|_{op} \le \frac{1}{2}$.
2. *There exists a dual certificate $Y \in \mathbb{R}^{n\times n}$ which satisfies $P_\Omega(Y) = Y$ and*
   (a) $\left\|P_T(Y) - UV^\top\right\|_F \le \frac{1}{4n^5}$,
   (b) $\|P_{T^\perp}(Y)\| \le \frac{1}{2}$.

Validating the Optimality Condition. We begin by proving that Condition 1 in Proposition 7.1 is satisfied under the conditions of Theorem 3.2. This is done in the following lemma (proved in Section 7.2). The lemma shows that R_Ω is close to the identity operator on T.

Lemma 7.2. *If $p_{ij} \ge \min\{c_0 \frac{(\mu_i+\nu_j)r}{n} \log n, 1\}$ for all (i,j) and a sufficiently large c_0, then w.h.p.*

$$\|P_T R_\Omega P_T - P_T\|_{op} \le \frac{1}{2}. \tag{10}$$

Constructing the Dual Certificate. It remains to construct a matrix Y (the dual certificate) that satisfies the condition 2 in Proposition 7.1. We do this using the golfing scheme (Gross, 2011; Candès et al., 2011). Set $k_0 = 20\log n$. Suppose the set Ω of observed entries is generated from $\Omega = \bigcup_{k=1}^{k_0} \Omega_k$, where for each $k = 1, \ldots, k_0$ and matrix index (i,j), $\mathbb{P}[(i,j)\in\Omega_k] = q_{ij} := 1 - (1-p_{ij})^{1/k_0}$ independent of all others. Clearly this is equivalent to the original Bernoulli sampling model. Let $W_0 := 0$ and for $k = 1, \ldots, k_0$,

$$W_k := W_{k-1} + R_{\Omega_k} P_T(UV^\top - P_T W_{k-1}), \tag{11}$$

where the operator R_{Ω_k} is given by

$$R_{\Omega_k}(Z) = \sum_{i,j} \frac{1}{q_{ij}} \mathbb{I}\left((i,j) \in \Omega_k\right) \left\langle e_i e_j^\top, Z \right\rangle e_i e_j^\top.$$

The dual certificate is given $Y := W_{k_0}$. Clearly $P_\Omega(Y) = Y$ by construction. The proof of Theorem 3.2 is completed if we show that under the condition in theorem, Y satisfies Conditions 2(a) and 2(b) in Proposition 7.1 w.h.p.

Concentration Properties The key step in our proof is to show that Y satisfies Condition 2(b) in Proposition 7.1, i.e., we need to bound $\|P_{T^\perp}(Y)\|$. Here our proof departs from existing ones, as we establish concentration bounds on this quantity in terms of (an appropriately weighted version of) the $\ell_{\infty,2}$ norm, which we now define. The $\mu(\infty, 2)$-norm of a matrix $Z \in \mathbb{R}^{n\times n}$ is defined as

$$\|Z\|_{\mu(\infty,2)} := \max\left\{ \max_i \sqrt{\frac{n}{\mu_i r} \sum_b Z_{ib}^2}, \max_j \sqrt{\frac{n}{\nu_j r} \sum_a Z_{aj}^2} \right\},$$

which is the maximum of the weighted column and row norms of Z. We also need the $\mu(\infty)$-norm of Z, which is a weighted version of the matrix ℓ_∞ norm. This is given as

$$\|Z\|_{\mu(\infty)} := \max_{i,j} |Z_{ij}| \sqrt{\frac{n}{\mu_i r}} \sqrt{\frac{n}{\nu_j r}}.$$

which is the weighted entry-wise magnitude of Z. We now state three new lemmas concerning the concentration properties of these norms. The first lemma is crucial to our proof; it bounds the spectral norm of $(R_\Omega - I)\, Z$ in terms of the $\mu(\infty, 2)$ and $\mu(\infty)$ norms of Z. This obviates intermediate lemmas required previous approaches (Candès & Tao, 2010; Gross, 2011; Recht, 2009; Keshavan et al., 2010) which use the ℓ_∞ norm of Z.

Lemma 7.3. *Suppose Z is a fixed $n \times n$ matrix. For some universal constant $c > 1$, we have w.h.p.*

$$\|(R_\Omega - I)\, Z\| \le c \left(\max_{i,j} \left| \frac{Z_{ij}}{p_{ij}} \right| \log n + \sqrt{\max\left\{ \max_i \sum_{j=1}^n \frac{Z_{ij}^2}{p_{ij}}, \max_j \sum_{i=1}^n \frac{Z_{ij}^2}{p_{ij}} \right\} \log n} \right).$$

If $p_{ij} \ge \min\{c_0 \frac{(\mu_i+\nu_j)r}{n} \log n, 1\}, \forall (i,j)$, then we further have $\|(R_\Omega - I)\, Z\| \le \frac{c}{\sqrt{c_0}} \left(\|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty,2)} \right)$ w.h.p.

The next two lemmas further control the $\mu(\infty, 2)$ and $\mu(\infty)$ norms of a matrix after random projections.

Lemma 7.4. *Suppose Z is a fixed $n \times n$ matrix. If $p_{ij} \ge \min\{c_0 \frac{(\mu_i+\nu_j)r}{n} \log n, 1\}$ for all i, j and sufficiently large c_0, then w.h.p.*

$$\|(P_T R_\Omega - P_T)Z\|_{\mu(\infty,2)} \le \frac{1}{2} \left(\|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty,2)} \right)$$

Lemma 7.5. *Suppose Z is a fixed $n \times n$ matrix. If $p_{ij} \ge \min\{c_0 \frac{(\mu_i+\nu_j)r}{n} \log n, 1\}$ for all i, j and c_0 sufficiently large, then w.h.p.*

$$\|(P_T R_\Omega - P_T)\, Z\|_{\mu(\infty)} \le \frac{1}{2} \|Z\|_{\mu(\infty)}.$$

We prove Lemmas 7.3–7.5 in Section 7.2. Equipped with the three lemmas above, we are now ready to validate that Y satisfies Condition 2 in Proposition 7.1.

Validating Condition 2(a): Set $\Delta_k = UV^\top - P_T(W_k)$ for $k = 1, \ldots, k_0$. By definition of W_k, we have

$$\Delta_k = (P_T - P_T R_{\Omega_k} P_T)\, \Delta_{k-1}. \tag{12}$$

Note that Ω_k is independent of Δ_{k-1} and $q_{ij} \ge p_{ij}/k_0 \ge c_0'(\mu_i + \nu_j) r \log(n)/n$ under the condition in Theorem 3.2. Applying Lemma 7.2 with Ω replaced by Ω_k , we obtain that w.h.p.

$$\|\Delta_k\|_F \le \|P_T - P_T R_{\Omega_k} P_T\| \, \|\Delta_{k-1}\|_F \le \frac{1}{2} \|\Delta_{k-1}\|_F.$$

Applying the above inequality recursively with $k = k_0, k_0 - 1, \ldots, 1$ gives

$$\left\|P_T(Y) - UV^\top\right\|_F = \|\Delta_{k_0}\|_F \le \left(\frac{1}{2}\right)^{k_0} \left\|UV^\top\right\|_F \le \frac{1}{4n^6} \cdot \sqrt{r} \le \frac{1}{4n^5}.$$

Validating Condition 2(b): By definition, Y can be rewritten as $Y = \sum_{k=1}^{k_0} R_{\Omega_k} P_T \Delta_{k-1}$. It follows that

$$\|P_{T^\perp}(Y)\| = \left\| P_{T^\perp} \sum_{k=1}^{k_0} \left(R_{\Omega_k} P_T - P_T\right) \Delta_{k-1} \right\| \le \sum_{k=1}^{k_0} \left\| \left(R_{\Omega_k} - I\right) \Delta_{k-1} \right\|.$$

We apply Lemma 7.3 with Ω replaced by Ω_k to each summand in the last RHS to obtain w.h.p.

$$\|P_{T^\perp}(Y)\| \le \frac{c}{\sqrt{c_0}} \sum_{k=1}^{k_0} \|\Delta_{k-1}\|_{\mu(\infty)} + \frac{c}{\sqrt{c_0}} \sum_{k=1}^{k_0} \|\Delta_{k-1}\|_{\mu(\infty,2)}. \tag{13}$$

We bound each summand in the last RHS. Applying $(k-1)$ times (12) and Lemma 7.5 (with Ω replaced by Ω_k), we have w.h.p.

$$\|\Delta_{k-1}\|_{\mu(\infty)} = \left\| \left(P_T - P_T R_{\Omega_{k-1}} P_T\right) \Delta_{k-2} \right\|_{\mu(\infty)} \le \left(\frac{1}{2}\right)^{k-1} \left\|UV^\top\right\|_{\mu(\infty)}.$$

for each k. Similarly, repeatedly applying (12), Lemma 7.4 and the inequality we just proved above, we obtain w.h.p.

$$\|\Delta_{k-1}\|_{\mu(\infty,2)} \tag{14}$$

$$= \left\| \left(P_T - P_T R_{\Omega_{k-1}} P_T\right) \Delta_{k-2} \right\|_{\mu(\infty,2)} \tag{15}$$

$$\le \frac{1}{2} \|\Delta_{k-2}\|_{\mu(\infty)} + \frac{1}{2} \|\Delta_{k-2}\|_{\mu(\infty,2)} \tag{16}$$

$$\le \left(\frac{1}{2}\right)^{k-1} \left\|UV^\top\right\|_{\mu(\infty)} + \frac{1}{2} \|\Delta_{k-2}\|_{\mu(\infty,2)} \tag{17}$$

$$\le k \left(\frac{1}{2}\right)^{k-1} \left\|UV^\top\right\|_{\mu(\infty)} + \left(\frac{1}{2}\right)^{k-1} \|UV\|_{\mu(\infty,2)}. \tag{18}$$

It follows that w.h.p.

$$\|P_{T^\perp}(Y)\| \le \frac{c}{\sqrt{c_0}} \sum_{k=1}^{k_0} (k+1) \left(\frac{1}{2}\right)^{k-1} \left\|UV^\top\right\|_{\mu(\infty)} + \frac{c}{\sqrt{c_0}} \sum_{k=1}^{k_0} \left(\frac{1}{2}\right)^{k-1} \left\|UV^\top\right\|_{\mu(\infty,2)} \tag{19}$$

$$\le \frac{6c}{\sqrt{c_0}} \left\|UV^\top\right\|_{\mu(\infty)} + \frac{2c}{\sqrt{c_0}} \left\|UV^\top\right\|_{\mu(\infty,2)}. \tag{20}$$

Note that for all (i,j), we have $\left|\left(UV^\top\right)_{ij}\right| = \left|e_i^\top UV^\top e_j\right| \le \sqrt{\frac{\mu_i r}{n}} \sqrt{\frac{\nu_j r}{n}}$, $\left\|e_i^\top UV^\top\right\|_2 = \sqrt{\frac{\mu_i r}{n}}$ and $\left\|UV^\top e_j\right\|_2 = \sqrt{\frac{\nu_j r}{n}}$. Hence $\left\|UV^\top\right\|_{\mu(\infty)} \le 1$ and $\left\|UV^\top\right\|_{\mu(\infty,2)} = 1$. We conclude that

$$\|P_{T^\perp}(Y)\| \le \frac{6c}{\sqrt{c_0}} + \frac{2c}{\sqrt{c_0}} \le \frac{1}{2}$$

provided that the constant c_0 in Theorem 3.2 is sufficiently large. This completes the proof of Theorem 3.2.

7.1. Proof of Proposition 7.1

Proof. Consider any feasible solution X to (1) with $P_\Omega(X) = P_\Omega(M)$. Let G be an $n \times n$ matrix which satisfies $\|P_{T^\perp} G\| = 1$, and $\langle P_{T^\perp} G, P_{T^\perp}(X - M)\rangle = \|P_{T^\perp}(X - M)\|_*$. Such G always exists by duality between the nuclear norm and spectral norm. Because $UV^\top + P_{T^\perp} G$ is a sub-gradient of the function $f(Z) = \|Z\|_*$ at $Z = M$, we have

$$\|X\|_* - \|M\|_* \quad \ge \quad \left\langle UV^\top + P_{T^\perp} G, X - M \right\rangle. \tag{21}$$

But $\langle Y, X-M\rangle = \langle P_\Omega(Y), P_\Omega(X-M)\rangle = 0$ since $P_\Omega(Y)=Y$. It follows that

$$\begin{aligned}
\|X\|_* - \|M\|_* &\ge \left\langle UV^\top + P_{T^\perp}G - Y, X-M\right\rangle \\
&= \|P_{T^\perp}(X-M)\|_* + \left\langle UV^\top - P_T Y, X-M\right\rangle - \langle P_{T^\perp}Y, X-M\rangle \\
&\ge \|P_{T^\perp}(X-M)\|_* - \left\|UV^\top - P_T Y\right\|_F \|P_T(X-M)\|_F - \|P_{T^\perp}Y\| \, \|P_{T^\perp}(X-M)\|_* \\
&\ge \frac{1}{2}\|P_{T^\perp}(X-M)\|_* - \frac{1}{4n^5}\|P_T(X-M)\|_F ,
\end{aligned}$$

where in the last inequality we use conditions 1 and 2 in the proposition. Using Lemma 7.6 below, we obtain

$$\|X\|_* - \|M\|_* \ge \frac{1}{2}\|P_{T^\perp}(X-M)\|_* - \frac{1}{4n^5}\cdot\sqrt{2}n^5\|P_{T^\perp}(X-M)\|_* > \frac{1}{8}\|P_{T^\perp}(X-M)\|_* .$$

The RHS is strictly positive for all X with $P_\Omega(X-M)=0$ and $X \ne M$. Otherwise we must have $P_T(X-M) = X-M$ and $P_T P_\Omega P_T(X-M)=0$, contradicting the assumption $\|P_T R_\Omega P_T - P_T\|_{op} \le \frac{1}{2}$. This proves that M is the unique optimum. □

Lemma 7.6. *If $p_{ij} \ge \frac{1}{n^{10}}$ for all (i,j) and $\|P_T R_\Omega P_T - P_T\|_{op} \le \frac{1}{2}$, then we have*

$$\|P_T Z\|_F \le \sqrt{2}n^5\|P_{T^\perp}(Z)\|_* , \forall Z \in \{Z' : P_\Omega(Z')=0\}. \tag{22}$$

Proof. Define the operator $R_\Omega^{1/2} : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^{n\times n}$ by

$$R_\Omega^{1/2}(Z) := \sum_{i,j}\frac{1}{\sqrt{p_{ij}}}\delta_{ij}\left\langle e_i e_j^\top, Z\right\rangle e_i e_j^\top .$$

Note that $R_\Omega^{1/2}$ is self-adjoint and satisfies $R_\Omega^{1/2}R_\Omega^{1/2} = R_\Omega$. Hence we have

$$\begin{aligned}
\left\|R_\Omega^{1/2}P_T(Z)\right\|_F &= \sqrt{\langle P_T R_\Omega P_T Z, P_T Z\rangle} && (23)\\
&= \sqrt{\langle (P_T R_\Omega P_T - P_T)Z, P_T(Z)\rangle + \langle P_T(Z), P_T(Z)\rangle} && (24)\\
&\ge \sqrt{\|P_T(Z)\|_F^2 - \|P_T R_\Omega P_T - P_T\| \, \|P_T(Z)\|_F^2} && (25)\\
&\ge \frac{1}{\sqrt{2}}\|P_T(Z)\|_F , && (26)
\end{aligned}$$

where the last inequality follows from the assumption $\|P_T R_\Omega P_T - P_T\|_{op} \le \frac{1}{2}$. On the other hand, $P_\Omega(Z)=0$ implies $R_\Omega^{1/2}(Z)=0$ and thus

$$\left\|R_\Omega^{1/2}P_T(Z)\right\|_F = \left\|R_\Omega^{1/2}P_{T^\perp}(Z)\right\|_F \le \left(\max_{i,j}\frac{1}{\sqrt{p_{ij}}}\right)\|P_{T^\perp}(Z)\|_F \le n^5\|P_{T^\perp}(Z)\|_F .$$

Combining the last two display equations gives

$$\|P_T(Z)\|_F \le \sqrt{2}n^5\|P_{T^\perp}(Z)\|_F \le \sqrt{2}n^5\|P_{T^\perp}(Z)\|_* .$$

□

7.2. Proof of Technical Lemmas

We prove the four technical lemmas that are used in the proof of our main theorem. The proofs use the matrix Bernstein inequality given as Theorem 10.1 in Section 10. We also make frequent use of the following facts: for all i and j, we have $\max\left\{\frac{\mu_i r}{n}, \frac{\nu_j r}{n}\right\} \le 1$ and

$$\frac{(\mu_i+\nu_j)r}{n} \ge \left\|P_T(e_i e_j^\top)\right\|_F^2 . \tag{27}$$

We also use the shorthand $a \wedge b := \min\{a,b\}$.

7.2.1. PROOF OF LEMMA 7.2

For any matrix Z, we can write

$$(P_T R_\Omega P_T - P_T)(Z) = \sum_{i,j} \left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right) \left\langle e_i e_j^\top, P_T(Z) \right\rangle P_T(e_i e_j^\top) =: \sum_{i,j} \mathcal{S}_{ij}(Z).$$

Note that $\mathbb{E}\left[\mathcal{S}_{ij}\right] = 0$ and $\mathcal{S}_{ij}$'s are independent of each other. For all Z and (i,j), we have $\mathcal{S}_{ij} = 0$ if $p_{ij} = 1$. On the other hand, when $p_{ij} \geq c_0 \frac{(\mu_i + \nu_j) r \log n}{n}$, then it follows from (27) that

$$\left\| \mathcal{S}_{ij}(Z) \right\|_F \leq \frac{1}{p_{ij}} \left\| P_T(e_i e_j^\top) \right\|_F^2 \left\| Z \right\|_F \leq \max_{i,j} \left\{ \frac{1}{p_{ij}} \frac{(\mu_i + \nu_j) r}{n} \right\} \left\| Z \right\|_F \leq \frac{1}{c_0 \log n} \left\| Z \right\|_F.$$

Putting together, we have that $\|\mathcal{S}_{ij}\| \leq \frac{1}{c_0 \log n}$ under the condition of the lemma. On the other hand, we have

$$\begin{aligned} \left\| \sum_{i,j} \mathbb{E}\left[\mathcal{S}_{ij}^2(Z) \right] \right\|_F &= \left\| \sum_{i,j} \mathbb{E}\left[\left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right)^2 \left\langle e_i e_j^\top, P_T(Z) \right\rangle \left\langle e_i e_j^\top, P_T(e_i e_j^\top) \right\rangle P_T(e_i e_j^\top) \right] \right\|_F \\ &\leq \left(\max_{i,j} \frac{1 - p_{ij}}{p_{ij}} \left\| P_T(e_i e_j^\top) \right\|_F^2 \right) \left\| \sum_{i,j} \left\langle e_i e_j^\top, P_T(Z) \right\rangle P_T(e_i e_j^\top) \right\|_F \\ &\leq \max_{i,j} \left\{ \frac{1 - p_{ij}}{p_{ij}} \frac{(\mu_i + \nu_j) r}{n} \right\} \left\| P_T(Z) \right\|_F, \end{aligned}$$

This implies $\left\| \sum_{i,j} \mathbb{E}\left[\mathcal{S}_{ij}^2 \right] \right\| \leq \frac{1}{c_0 \log n}$ under the condition of the lemma. Applying the Matrix Bernstein inequality (Theorem 10.1), we obtain $\|P_T R_\Omega P_T - P_T\| = \left\| \sum_{i,j} \mathcal{S}_{ij} \right\| \leq \frac{1}{2}$ w.h.p. for sufficiently large c_0.

7.2.2. PROOF OF LEMMA 7.3

We can write $(R_\Omega - I)\, Z$ as the sum of independent matrices:

$$(R_\Omega - I)\, Z = \sum_{i,j} \left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right) Z_{ij} e_i e_j^\top =: \sum_{i,j} S_{ij}.$$

Note that $\mathbb{E}[S_{ij}] = 0$. For all (i,j), we have $S_{ij} = 0$ if $p_{ij} = 1$, and

$$\|S_{ij}\| \leq \frac{1}{p_{ij}} \left| Z_{ij} \right|.$$

Moreover,

$$\left\| \mathbb{E}\left[\sum_{i,j} S_{ij}^\top S_{ij} \right] \right\| = \left\| \sum_{i,j} Z_{ij}^2 e_i e_j^\top e_j e_i^\top \mathbb{E}\left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right)^2 \right\| = \max_i \sum_{j=1}^n \frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2.$$

The quantity $\left\| \mathbb{E}\left[\sum_{i,j} S_{ij} S_{ij}^\top \right] \right\|$ is bounded by $\max_j \sum_{i=1}^n (1 - p_{ij}) Z_{ij}^2 / p_{ij}$ in a similar way. The first part of the lemma then follows from the matrix Bernstein inequality (Theorem 10.1 in the Section 10). If $p_{ij} \geq 1 \wedge \frac{c_0(\mu_i + \nu_j) r \log n}{n} \geq 1 \wedge 2c_0 \sqrt{\frac{\mu_i r}{n} \cdot \frac{\nu_j r}{n}} \log n$, we have for all i and j: $\|S_{ij}\| \log n \leq (1 - \mathbb{I}(p_{ij} = 1)) \frac{1}{p_{ij}} |Z_{ij}| \log n \leq \frac{1}{c_0} \|Z\|_{\mu(\infty)}$, $\sum_{i=1}^n \frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2 \log n \leq \frac{1}{c_0} \|Z\|_{\mu(\infty,2)}^2$ and $\sum_{j=1}^n \frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2 \log n \leq \frac{1}{c_0} \|Z\|_{\mu(\infty,2)}^2$. The second part of the lemma follows again from applying the matrix Bernstein inequality.

7.2.3. Proof of Lemma 7.4

Let $X = (P_T R_\Omega - P_T)\, Z$. By definition we have $\|X\|_{\mu(\infty,2)} = \max_{a,b}\left\{\sqrt{\frac{n}{\mu_a r}}\,\|X_{a\cdot}\|_2\,,\sqrt{\frac{n}{\nu_b r}}\,\|X_{\cdot b}\|_2\right\}$, where $X_{a\cdot}$ and $X_{\cdot b}$ are the a-th row and b-th column of of X, respectively. We bound each term in the maximum. Observe that $\sqrt{\frac{n}{\nu_b r}}X_{\cdot b}$ can be written as the sum of independent column vectors:

$$\sqrt{\frac{n}{\nu_b r}}X_{\cdot b} = \sum_{i,j}\left(\frac{1}{p_{ij}}\delta_{ij} - 1\right) Z_{ij}\left(P_T(e_i e_j^\top)e_b\right)\sqrt{\frac{n}{\nu_b r}} =: \sum_{i,j} S_{ij},$$

where $\mathbb{E}\left[S_{ij}\right] = 0$. To control $\|S_{ij}\|_2$ and $\left\|\mathbb{E}\left[\sum_{i,j} S_{ij}^\top S_{ij}\right]\right\|$, we first need a bound for $\left\|P_T(e_i e_j^\top)e_b\right\|_2$. If $j = b$, we have

$$\left\|P_T(e_i e_j^\top)e_b\right\|_2 = \left\|UU^\top e_i + (I - UU^\top)e_i \left\|V^\top e_b\right\|_2^2\right\|_2 \le \sqrt{\frac{\mu_i r}{n}} + \sqrt{\frac{\nu_b r}{n}}, \tag{28}$$

where we use the triangle inequality and the definition of μ_i and ν_b . Similarly, if $j \ne b$, we have

$$\left\|P_T(e_i e_j^\top)e_b\right\|_2 = \left\|(I - UU^\top)e_i e_j^\top VV^\top e_b\right\|_2 \le \left|e_j^\top VV^\top e_b\right|. \tag{29}$$

Now note that $\|S_{ij}\|_2 \le (1 - \mathbb{I}(p_{ij} = 1))\,\frac{1}{p_{ij}}\,|Z_{ij}|\,\sqrt{\frac{n}{\nu_b r}}\,\left\|P_T(e_i e_j^\top)e_b\right\|_2$. Using the bounds (28) and (29), we obtain that for $j = b$,

$$\|S_{ij}\|_2 \le (1 - \mathbb{I}(p_{ij} = 1))\,\frac{1}{p_{ib}}\,|Z_{ib}|\sqrt{\frac{n}{\nu_b r}}\cdot\left(\sqrt{\frac{\mu_i r}{n}} + \frac{\nu_b r}{n}\right) \le \frac{2}{c_0\sqrt{\frac{\mu_i r \nu_b r}{n^2}}\log n}\,|Z_{ib}| \le \frac{2}{c_0 \log n}\,\|Z\|_{\mu(\infty)}\,,$$

where we use $p_{ib} \ge 1 \wedge \frac{c_0 \mu_i r \log n}{n}$ and $p_{ib} \ge 1 \wedge c_0\sqrt{\frac{\mu_i r}{n}\frac{\nu_b r}{n}}\log n$ in the second inequality. For $j \ne b$, we have

$$\|S_{ij}\|_2 \le (1 - \mathbb{I}(p_{ij} = 1))\,\frac{1}{p_{ij}}\,|Z_{ij}|\sqrt{\frac{n}{\nu_b r}}\cdot\sqrt{\frac{\nu_j r}{n}}\sqrt{\frac{\nu_b r}{n}} \le \frac{2}{c_0 \log n}\,\|Z\|_{\mu(\infty)}\,,$$

where we use $p_{ij} \ge 1 \wedge c_0\sqrt{\frac{\mu_i r}{n}\frac{\nu_j r}{n}}\log n$. We thus obtain $\|S_{ij}\|_2 \le \frac{2}{c_0 \log n}\,\|Z\|_{\mu(\infty)}$ for all (i,j).

On the other hand, note that

$$\begin{aligned}\left|\mathbb{E}\left[\textstyle\sum_{i,j} S_{ij}^\top S_{ij}\right]\right| &= \left|\textstyle\sum_{i,j}\mathbb{E}\left[\left(\frac{1}{p_{ij}}\delta_{ij} - 1\right)^2\right] Z_{ij}^2\left\|P_T(e_i e_j^\top)e_b\right\|_2^2\cdot\frac{n}{\nu_b r}\right| \\ &= \left(\textstyle\sum_{j=b,i} + \sum_{j\ne b,i}\right)\frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2\left\|P_T(e_i e_j^\top)e_b\right\|_2^2\cdot\frac{n}{\nu_b r}.\end{aligned}$$

Applying (28), we can bound the first sum by

$$\sum_{j=b,i} \le \sum_i \frac{1 - p_{ib}}{p_{ib}} Z_{ib}^2\cdot 2\left(\frac{\mu_i r}{n} + \frac{\nu_b r}{n}\right)\cdot\frac{n}{\nu_b r} \le \frac{2}{c_0 \log n}\frac{n}{\nu_b r}\,\|Z_{\cdot b}\|_2^2 \le \frac{2}{c_0 \log n}\,\|Z\|_{\mu(\infty,2)}^2\,,$$

where we use $p_{ib} \ge 1 \wedge \frac{c_0(\mu_i + \nu_b)r}{n}\log n$ in the second inequality. The second sum can be bounded using (29):

$$\begin{aligned}\sum_{j\ne b,i} &\le \sum_{j\ne b,i}\frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2\left|e_j^\top VV^\top e_b\right|^2\frac{n}{\nu_b r} \\ &= \frac{n}{\nu_b r}\sum_{j\ne b}\left|e_j^\top VV^\top e_b\right|^2\sum_i \frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2 \\ &\overset{(a)}{\le} \frac{n}{\nu_b r}\sum_{j\ne b}\left|e_j^\top VV^\top e_b\right|^2\left(\frac{1}{c_0 \log n}\sum_i Z_{ij}^2\frac{n}{\nu_j r}\right) \\ &\le \left(\frac{1}{c_0 \log n}\,\|Z\|_{\mu(\infty,2)}^2\right)\frac{n}{\nu_b r}\sum_{j\ne b}\left|e_j^\top VV^\top e_b\right|^2 \\ &\overset{(b)}{\le} \frac{1}{c_0 \log n}\,\|Z\|_{\mu(\infty,2)}^2\,,\end{aligned}$$

where we use $p_{ij} \geq 1 \wedge \frac{c_0 \nu_j r \log n}{n}$ in (a) and $\sum_{j \neq b} \left| e_j^\top V V^\top e_b \right|^2 \leq \left\| V V^\top e_b \right\|_2^2 \leq \frac{\nu_b r}{n}$ in (b). Combining the bounds for the two sums, we obtain $\left\| \mathbb{E} \left[\sum_{i,j} S_{ij}^\top S_{ij} \right] \right\| \leq \frac{3}{c_0 \log n} \|Z\|_{\mu(\infty,2)}^2$. We can bound $\left\| \mathbb{E} \left[\sum_{i,j} S_{ij} S_{ij}^\top \right] \right\|$ in a similar way. Applying the Matrix Bernstein inequality (Theorem 10.1) w.h.p.

$$\left\| \sqrt{\frac{n}{\nu_b r}} X_{\cdot b} \right\|_2 = \left\| \sum_{i,j} S_{ij} \right\|_2 \leq \frac{1}{2} \left(\|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty,2)} \right)$$

for c_0 sufficiently large. Similarly we can bound $\left\| \sqrt{\frac{n}{\mu_a r}} X_{a\cdot} \right\|_2$ by the same quantity. We take a union bound over all a and b to obtain the desired results.

7.3. Proof of Lemma 7.5

Fix a matrix index (a, b) and let $w_{ab} = \sqrt{\frac{\mu_a r}{n} \frac{\nu_b r}{n}}$. We can write

$$[(P_T R_\Omega - P_T) Z]_{ab} \sqrt{\frac{n}{\mu_a r}} \sqrt{\frac{n}{\nu_b r}} = \sum_{i,j} \left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right) Z_{ij} \left\langle e_i e_j^\top, P_T(e_a e_b^\top) \right\rangle \frac{1}{w_{ab}} =: \sum_{i,j} s_{ij},$$

which is the sum of independent zero-mean variables. We first compute the following bound:

$$\begin{aligned} & \left| \left\langle e_i e_j^\top, P_T(e_a e_b^\top) \right\rangle \right| \\ = & \left| e_i^\top U U^\top e_a e_b^\top e_j + e_i^\top (I - UU^\top) e_a e_b^\top V V^\top e_j \right| \\ = & \begin{cases} \left| e_a^\top U U^\top e_a + e_a^\top (I - UU^\top) e_a e_b^\top V V^\top e_b \right| \leq \frac{\mu_a r}{n} + \frac{\nu_b r}{n}, & i = a, j = b, \\ \left| e_a^\top (I - UU^\top) e_a e_b^\top V V^\top e_j \right| \leq \left| e_b^\top V V^\top e_j \right|, & i = a, j \neq b, \\ \left| e_i^\top U U^\top e_a e_b^\top (I - VV^\top) e_b \right| \leq \left| e_i^\top U U^\top e_a \right|, & i \neq a, j = b, \\ \left| e_i^\top U U^\top e_a e_b^\top V V^\top e_j \right| \leq \left| e_i^\top U U^\top e_a \right| \left| e_b^\top V V^\top e_j \right|, & i \neq a, j \neq b, \end{cases} \end{aligned} \tag{30}$$

where we use the fact that the matrices $I - UU^\top$ and $I - VV^\top$ have spectral norm at most 1. We proceed to bound $|s_{ij}|$. Note that

$$|s_{ij}| \leq (1 - \mathbb{I}(p_{ij} = 1)) \frac{1}{p_{ij}} |Z_{ij}| \left| \left\langle e_i e_j^\top, P_T(e_a e_b^\top) \right\rangle \right| \frac{1}{w_{ab}}.$$

We distinguish four cases. When $i = a$ and $j = b$, we use (30) and $p_{ab} \geq 1 \wedge \frac{c_0(\mu_a + \nu_b) r \log^2(n)}{n}$ to obtain $|s_{ij}| \leq |Z_{ij}| / (w_{ij} c_0 \log n) \leq \|Z\|_{\mu(\infty)} / (c_0 \log n)$. When $i = a$ and $j \neq b$, we apply (30) to get

$$|s_{ij}| \leq (1 - \mathbb{I}(p_{ij} = 1)) \frac{|Z_{aj}|}{p_{aj}} \cdot \sqrt{\frac{\nu_b r}{n} \frac{\nu_j r}{n}} \cdot \sqrt{\frac{n}{\mu_a r} \frac{n}{\nu_b r}} \overset{(a)}{\leq} |Z_{aj}| \cdot \sqrt{\frac{n}{\mu_a r} \frac{n}{\nu_j r}} \frac{1}{c_0 \log n} \leq \frac{\|Z\|_{\mu(\infty)}}{c_0 \log n},$$

where (a) follows from $p_{aj} \geq \min \left\{ c_0 \frac{\nu_j r \log n}{n}, 1 \right\}$. In a similar fashion, we can show that the same bound holds when $i \neq a$ and $j = b$. When $i \neq a$ and $j \neq b$, we use (30) to get

$$|s_{ij}| \leq (1 - \mathbb{I}(p_{ij} = 1)) \frac{|Z_{ij}|}{p_{ij}} \cdot \sqrt{\frac{\mu_i r}{n} \frac{\mu_a r}{n}} \sqrt{\frac{\nu_b r}{n} \frac{\nu_j r}{n}} \cdot \sqrt{\frac{n}{\mu_a r} \frac{n}{\nu_b r}} \overset{(b)}{\leq} |Z_{ij}| \cdot \sqrt{\frac{n}{\mu_i r} \frac{n}{\nu_j r}} \frac{1}{c_0 \log n} \leq \frac{\|Z\|_{\mu(\infty)}}{c_0 \log n},$$

where (b) follows from $p_{ij} \geq 1 \wedge c_0 \sqrt{\frac{\mu_i r}{n} \frac{\nu_j r}{n}} \log n$ and $\max \left\{ \sqrt{\frac{\mu_i r}{n}}, \sqrt{\frac{\nu_j r}{n}} \right\} \leq 1$. We conclude that $|s_{ij}| \leq \|Z\|_{\mu(\infty)} / (c_0 \log n)$ for all (i, j).

On the other hand, note that

$$\left| \mathbb{E} \left[\sum_{i,j} s_{ij}^2 \right] \right| = \sum_{i,j} \mathbb{E} \left[\left(\frac{1}{p_{ij}} \delta_{ij} - 1 \right)^2 \right] \frac{Z_{ij}^2}{w_{ab}^2} \left\langle e_i e_j^\top, P_T(e_a e_b^\top) \right\rangle^2 = \sum_{i=a, j=b} + \sum_{i=a, j \neq b} + \sum_{i \neq a, j=b} + \sum_{i \neq a, j \neq b}.$$

We bound each of the four sums. By (30) and $p_{ab} \geq 1 \wedge \frac{c_0(\mu_a+\nu_b) r \log n}{n} \geq 1 \wedge \frac{c_0(\mu_a+\nu_b)^2 r^2 \log n}{2n^2}$, we have

$$\sum_{i=a,j=b} \leq \frac{1-p_{ab}}{p_{ab} w_{ab}^2} Z_{ab}^2 \left(\frac{\mu_a r}{n} + \frac{\nu_b r}{n}\right)^2 \leq \frac{2 \left\|Z\right\|_{\mu(\infty)}^2}{c_0 \log n}.$$

By (30) and $p_{aj} w_{ab}^2 \geq w_{ab}^2 \wedge \left(c_0 w_{aj}^2 \frac{\nu_b r}{n} \log n\right)$, we have

$$\sum_{i=a,j\neq b} \leq \sum_{,j\neq b} \frac{1-p_{aj}}{p_{aj} w_{ab}^2} Z_{aj}^2 \left|e_b^\top V V^\top e_j\right| \leq \frac{\left\|Z\right\|_{\mu(\infty)}^2}{c_0 \log n} \cdot \frac{n}{\nu_b r} \sum_{j\neq b} \left|e_b^\top V V^\top e_j\right|,$$

which implies $\sum_{i=a,j\neq b} \leq \left\|Z\right\|_{\mu(\infty)}^2 / (c_0 \log n)$. Similarly we can bound $\sum_{i\neq a,j=b}$ by the same quantity. Finally, by (30) and $p_{ij} \geq 1 \wedge \left(c_0 \frac{\mu_i r}{n} \frac{\nu_j r}{n} \log n\right)$, we have

$$\sum_{i\neq a,j\neq b} \leq \frac{1}{w_{ab}^2} \sum_{i\neq a,j\neq b} \frac{(1-p_{ij}) Z_{ij}^2}{p_{ij}} \cdot \left|e_i^\top U U^\top e_a\right| \left|e_b^\top V V^\top e_j\right| \leq \frac{\left\|Z\right\|_{\mu(\infty)}^2}{c_0 \log n} \cdot \frac{1}{w_{ab}^2} \sum_{i\neq a} \left|e_i^\top U U^\top e_a\right| \sum_{j\neq b} \left|e_b^\top V V^\top e_j\right|,$$

which implies $\sum_{i\neq a,j\neq b} \leq \left\|Z\right\|_{\mu(\infty)}^2 / (c_0 \log n)$. Combining pieces, we obtain

$$\left|\mathbb{E}\left[\sum_{ij} s_{ij}^2\right]\right| \leq 5 \left\|Z\right\|_{\mu(\infty)}^2 / (c_0 \log n).$$

Applying the Bernstein inequality (Theorem 10.1), we conclude that

$$\left|\left[\left(P_T R_\Omega P_T - P_T\right) Z\right]_{ab} \sqrt{\frac{n}{\mu_a r}} \sqrt{\frac{n}{\nu_b r}}\right| = \left|\sum_{i,j} s_{ij}\right| \leq \frac{1}{2} \left\|Z\right\|_{\mu(\infty)}$$

w.h.p. for c_0 sufficiently large. The desired result follows from a union bound over all (a, b).

8. Proof of Remark 3.4

Recall the setting: for each row of M, we pick it and observe all its entries with probability p. We need a simple lemma. Let $J \subseteq [n]$ be the set of the indices of the row picked, and $P_J(Z)$ be the matrix that is obtained from Z by zeroing out the rows outside J. Recall that $U \Sigma V^\top$ is the SVD of M.

Lemma 8.1. *If* $\max_i \left\|U^\top e_j\right\| = \max_i \sqrt{\frac{\mu_i r}{n}} \leq \sqrt{\frac{\mu_0 r}{n}}$ *and* $p \geq c_0 \frac{\mu_0 r \log}{n}$ *for some universal constant* c_0*, then with high probability,*

$$\left\|U^\top P_J(U) - I_{r\times r}\right\| \leq \frac{1}{2},$$

where $I_{r\times r}$ *is the identity matrix in* $\mathbb{R}^{r\times r}$*.*

Proof. Let $\eta_j = \mathbb{I}(i \in J)$, where $\mathbb{I}(\cdot)$ is the indicator function. Note that

$$U^\top P_J(U) - I_{r\times r} = U^\top P_J(U) - U^\top U = S_{(i)} := \sum_{i=1}^{n} \left(\frac{1}{p} \eta_i - 1\right) U^\top e_i e_i^\top U.$$

Note that $\mathbb{E}\left[S_{(i)}\right]=0$, $\left\|S_{(i)}\right\| \leq \frac{1}{p}\left\|U^{\top}e_i\right\|_2^2 \leq \frac{\mu_0 r}{pn}$, and

$$\begin{aligned}
\left\|\mathbb{E}\left[\sum_{i=1}^{n} S_{(i)}S_{(i)}^{\top}\right]\right\| &= \left\|\mathbb{E}\left[\sum_{i=1}^{n} S_{(i)}^{\top}S_{(i)}\right]\right\| \\
&= \frac{1-p}{p}\left\|\sum_{i=1}^{n} U^{\top}e_ie_i^{\top}UU^{\top}e_ie_i^{\top}U\right\| \\
&= \frac{1-p}{p}\left\|U^{\top}\left(\sum_{i=1}^{n} e_ie_i^{\top}\left\|U^{\top}e_i\right\|_2^2\right)U\right\| \\
&\leq \frac{1}{p}\left\|\sum_{i=1}^{n} e_ie_i^{\top}\left\|U^{\top}e_i\right\|_2^2\right\| \\
&= \frac{1}{p}\max_i \left\|U^{\top}e_i\right\|_2^2 \leq \frac{\mu_0 r}{pn}.
\end{aligned}$$

It follows from the matrix Bernstein (Theorem 10.1) that

$$\left\|U^{\top}P_J(U)-I_{r\times r}\right\| \leq c\max\left\{\frac{\mu_0 r}{pn}\log n, \sqrt{\frac{\mu_0 r}{pn}\log n}\right\} \leq \frac{1}{2}$$

since c_0 in the statement of the lemma is sufficiently large. □

Note that $\left\|U^{\top}P_J(U)-I_{r\times r}\right\| \leq \frac{1}{2}$ implies that $U^{\top}P_J(U)$ is invertible, which further implies $P_J(U)\in\mathbb{R}^{n\times r}$ has rank-r. The rows picked are $P_J(M)=P_J(U)\Sigma V^{\top}$, which thus have full rank-r and their row space must be the same as the row space of M. We can then compute the local coherences $\{\nu_j\}$ from these rows, and sampled a set of entries of M according to the distribution

$$p_{ij}=\min\left\{c_0\frac{(\mu_0+\nu_j)r\log^2 n}{n}.\right\}$$

Applying Theorem 3.2, we are guaranteed to recover M exactly w.h.p. from these entries. Note that expectation of the total number of entries we have observed is

$$pn+\sum_{ij}p_{ij}=\Theta\left(\mu_0 r\log n+(\mu_0 rn+rn)\log^2 n\right)=\Theta\left(\mu_0 rn\log^2 n\right),$$

and by Hoeffding's inequality, the actual number of observations is at most two times the expectation w.h.p. for c_0 sufficiently large.

9. Proof of Theorem 5.1

Suppose the rank-r SVD of $\bar{M}$ is $\bar{U}\bar{\Sigma}\bar{V}^{\top}$; so $\bar{U}\bar{\Sigma}\bar{V}^{\top}=RMC=RU\Sigma V^{\top}C$. By definition, we have

$$\frac{\bar{\mu}_i r}{n}=\left\|P_{\tilde{U}}(e_i)\right\|_2^2,$$

where $P_{\tilde{U}}(\cdot)$ denotes the projection onto the column space of $\tilde{U}$, which is the same as the column space of RU. This projection has the explicit form

$$P_{\tilde{U}}(e_i)=RU\left(U^{\top}R^2U\right)^{-1}U^{\top}Re_i.$$

It follows that

$$\begin{aligned}
\frac{\bar{\mu}_i r}{n} &= \left\|RU\left(U^{\top}R^2U\right)^{-1}U^{\top}Re_i\right\|_2^2 \\
&= R_i^2 e_i^{\top}U\left(U^{\top}R^2U\right)^{-1}U^{\top}e_i \\
&\leq R_i^2\left[\sigma_r\left(RU\right)\right]^{-2}\left\|U^{\top}e_i\right\|_2^2 \\
&\leq R_i^2\frac{\mu_0 r}{n}\left[\sigma_r\left(RU\right)\right]^{-2}, \qquad (31)
\end{aligned}$$

where $\sigma_r(\cdot)$ denotes the r-th singular value and the last inequality follows from the standard incoherence assumption $\max_{i,j}\{\mu_i, \nu_j\} \le \mu_0$. We now bound $\sigma_r(RU)$. Since RU has rank r, we have

$$\sigma_r^2(RU) = \min_{\|x\|=1} \|RUx\|_2^2 = \min_{\|x\|=1} \sum_{i=1}^{n} R_i^2 \left|e_i^\top U x\right|^2.$$

If we let $z_i := \left|e_i^\top U x\right|^2$ for each $i \in [n]$, then z_i satisfies

$$\sum_{i=1}^{n} z_i = \|Ux\|_2^2 = \|x\|_2^2 = 1$$

and by the standard incoherence assumption,

$$z_i \le \left\|U^\top e_i\right\|_2^2 \|x\|_2^2 \le \frac{\mu_0 r}{n}.$$

Therefore, the value of the minimization above is lower-bounded by

$$\begin{aligned}
&\min_{z\in\mathbb{R}^n} \sum_{i=1}^{n} R_i^2 z_i \\
&\text{s.t.} \sum_{i=1}^{n} z_i = 1, \quad 0 \le z_i \le \frac{\mu_0 r}{n},\ i = 1,\ldots,n.
\end{aligned} \tag{32}$$

From the theory of linear programming, we know the minimum is achieved at an extreme point z^* of the feasible set. The extreme point z^* satisfies $z_i^* \ge 0, \forall i$ and n linear equalities

$$\begin{aligned}
\sum_{i=1}^{n} z_i^* &= 1 \\
z_i^* &= 0, \quad \text{for } i \in I_1 \\
z_i^* &= \frac{\mu_0 r}{n},\ \text{for } i \in I_2
\end{aligned}$$

for some index sets I_1 and I_2 such that $I_1 \cap I_2 = \varphi$,$|I_1| + |I_2| = n - 1$. It is easy to see that we must have $|I_2| = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor$. Since $R_1 \le R_2 \le \ldots \le R_n$, the minimizer z^* has the form

$$\begin{aligned}
z_i^* &= \frac{\mu_0 r}{n},\ i = 1,\ldots,\left\lfloor \frac{n}{\mu_0 r} \right\rfloor, \\
z_i^* &= 1 - \left\lfloor \frac{n}{\mu_0 r} \right\rfloor \cdot \frac{\mu_0 r}{n},\ i = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor + 1, \\
z_i^* &= 0,\ i = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor + 2,\ldots,n,
\end{aligned}$$

and the value of the minimization (32) is at least

$$\sum_{i=1}^{\lfloor n/(\mu_0 r)\rfloor} R_i^2 \frac{\mu_0 r}{n}.$$

This proves that $\sigma_r^2(RU) \ge \frac{\mu_0 r}{n} \sum_{i=1}^{\lfloor n/(\mu_0 r)\rfloor} R_i^2$. Combining with (31), we obtain that

$$\frac{\bar{\mu}_i r}{n} \le \frac{R_i^2}{\sum_{i'=1}^{\lfloor n/(\mu_0 r)\rfloor} R_{i'}^2}, \quad \frac{\bar{\nu}_j r}{n} \le \frac{C_j^2}{\sum_{j'=1}^{\lfloor n/(\mu_0 r)\rfloor} C_{j'}^2};$$

the proof for $\bar{\nu}_j$ is similar. Applying Theorem 3.2 to the equivalent problem (7) with the above bounds on $\bar{\mu}_i$ and $\bar{\nu}_j$ proves the theorem.

10. Matrix Bernstein Inequality

Theorem 10.1 ((Tropp, 2012)). *Let $X_1, \ldots, X_N \in \mathbb{R}^{n_1 \times n_2}$ be independent zero mean random matrices. Suppose*

$$\max\left\{\left\|\sum_{k=1}^{N} X_k X_k^\top\right\|, \left\|\sum_{k=1}^{N} X_k^\top X_k\right\|\right\} \leq \sigma^2 \tag{33}$$

and $\|X_k\| \leq B$ almost surely for all k. Then for any $c > 0$, we have

$$\left\|\sum_{k=1}^{N} X_k\right\| \leq 2\sqrt{c\sigma^2 \log(n_1 + n_2)} + cB \log(n_1 + n_2). \tag{34}$$

with probability at least $1 - (n_1 + n_2)^{-(c-1)}$.