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 $\ell_\infty$ norm $\|Y\|_\infty := \max_{i,j} |Y_{ij}|$ of a certain matrix $Y$, which gives rise to the standard and joint incoherence conditions involving uniform bounds by $\mu_0$ and $\mu_{0R}$. 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 $\langle Y, Z \rangle := \text{trace}(Y^T Z)$. 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

$$P_T(Z) := UU^T Z + ZZ^T V V^T - UU^T V V^T$$

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 $\ell_2$ 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 $A$ on matrices, its operator norm is defined as $\|A\|_{op} = \sup_{X \in \mathbb{R}^{n \times n}} \frac{\|A(X)\|_F}{\|X\|_F}$. For each $1 \leq i, j \leq n$, we define the random variable $\delta_{ij} := \mathbb{I}((i, j) \in \Omega)$, where $\mathbb{I}(\cdot)$ is the indicator function. The matrix operator $R_\Omega : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is defined as

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

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 $p_{ij} \geq \frac{1}{n^2}$. The matrix $M$ is the unique optimal solution to (1) if the following conditions hold.

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

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 $R_\Omega$ is close to the identity operator on $T$.

**Lemma 7.2.** If $p_{ij} \geq \min\{c_0 (\mu + \mu_{0R}) n \log n, 1\}$ for all $(i, j)$ and a sufficiently large $c_0$, then w.h.p.

$$\|P_T R_\Omega P_T - P_T\|_{op} \leq \frac{1}{2}.$$  

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 $\Omega$ 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 $(i, j)$, $P[(i, j) \in \Omega_k] = q_{ij} := 1 - (1 - p_{ij})^{1/k_0}$ independent of all others. Clearly this is equivalent to the original Bernoulli sampling model. Let $W_0 := 0$ and for $k = 1, \ldots, k_0$,

$$W_k := W_{k-1} + R_{\Omega_k} P_T(UV^T - P_T W_{k-1}).$$
where the operator $R_{\Omega_k}$ is given by
\[
R_{\Omega_k}(Z) = \sum_{i,j} \frac{1}{q_{ij}} \mathbb{1}((i,j) \in \Omega_k) \langle e_i e_j^\top, Z \rangle e_i e_j^\top.
\]

The dual certificate is given $Y := W_{k_0}$. Clearly $P_T(Y) = Y$ 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 $\|P_T(Y)\|$. 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
\[
\|Z\|_{\mu(\infty,2)} := \max_{i,j} \sqrt{\frac{n}{\mu_{ij}} \sum_b Z_{ib}^2} = \max_{i,j} \sqrt{\frac{n}{\mu_{ij}} \sum_a Z_{ai}^2},
\]
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 $\ell_\infty$ norm. This is given as
\[
\|Z\|_{\mu(\infty)} := \max_{i,j} |Z_{ij}| \sqrt{\frac{n}{\mu_{ij}}} \sqrt{\frac{n}{\mu_{ij}}}.
\]

**Lemma 7.3.** Suppose $Z$ is a fixed $n \times n$ matrix. For some universal constant $c > 1$, we have w.h.p.
\[
\left\| (R_{\Omega} - I) Z \right\| \leq c \left( \max_{i,j} \left| Z_{ij} \right| \log n + \sqrt{\max_{i,j} \left\{ \sum_{j=1}^n Z_{ij}^2, \sum_{i=1}^n Z_{ij}^2 \right\}} \log n \right).
\]

If $p_{ij} \geq \min \{c_0 \left( \frac{\mu_i + \mu_j}{n} \right) \log n, 1\}, \forall (i,j)$, then we further have $\| (R_{\Omega} - I) Z \| \leq \frac{c}{\sqrt{\log n}} \left( \|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty,2)} \right)$ 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 $p_{ij} \geq \min \{c_0 \left( \frac{\mu_i + \mu_j}{n} \right) \log n, 1\}$ for all $i, j$ and sufficiently large $c_0$, then w.h.p.
\[
\| (P_T R_{\Omega} - P_T) Z \|_{\mu(\infty,2)} \leq \frac{1}{2} \left( \|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty,2)} \right)
\]

**Lemma 7.5.** Suppose $Z$ is a fixed $n \times n$ matrix. If $p_{ij} \geq \min \{c_0 \left( \frac{\mu_i + \mu_j}{n} \right) \log n, 1\}$ for all $i, j$ and $c_0$ sufficiently large, then w.h.p.
\[
\| (P_T R_{\Omega} - P_T) Z \|_{\mu(\infty)} \leq \frac{1}{2} \|Z\|_{\mu(\infty)}.
\]

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
\[
\Delta_k = (P_T - P_T R_{\Omega_k} P_T) \Delta_{k-1}.
\]

Note that $\Omega_k$ is independent of $\Delta_{k-1}$ and $q_{ij} \geq p_{ij}/k_0 \geq c_0 (\mu_i + \mu_j) r \log(n)/n$ under the condition in Theorem 3.2. Applying Lemma 7.2 with $\Omega$ replaced by $\Omega_k$, we obtain that w.h.p.
\[
\|\Delta_k\|_F \leq \|P_T - P_T R_{\Omega_k} P_T\| \|\Delta_{k-1}\|_F \leq \frac{1}{2} \|\Delta_{k-1}\|_F.
\]
Applying the above inequality recursively with \(k = k_0, k_0 - 1, \ldots, 1\) gives
\[
\|P_T(Y) - UV^\top\|_F = \|\Delta_{k_0}\|_F \leq \left(\frac{1}{2}\right)^{k_0} \|UV^\top\|_F \leq \frac{1}{4\sqrt{n}} \cdot \sqrt{r} \leq \frac{1}{4\sqrt{n^5}}.
\]

**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 w.h.p.
\[
\|P_{T^\perp}(Y)\| = \left\| P_{T^\perp} \left( \sum_{k=1}^{k_0} (R_{\Omega_k} P_T - P_T) \Delta_{k-1} \right) \right\| \leq \sum_{k=1}^{k_0} \|(R_{\Omega_k} - I) \Delta_{k-1}\|.
\]

We bound each summand in the last RHS. Applying the above inequality recursively with \(\nu\) to each summand in the last RHS to obtain w.h.p.
\[
\|P_{T^\perp}(Y)\| \leq \frac{c}{\sqrt{n}} \sum_{k=1}^{k_0} \|\Delta_{k-1}\|_{\mu(\infty)} + \frac{c}{\sqrt{n}} \sum_{k=1}^{k_0} \|\Delta_{k-1}\|_{\mu(\infty,2)}.
\]

We bound each summand in the last RHS. Applying \((k - 1)\) times (12) and Lemma 7.5 (with \(\Omega\) replaced by \(\Omega_k\)), we have w.h.p.
\[
\|\Delta_{k-1}\|_{\mu(\infty)} = \|(P_T - P_T R_{\Omega_k-1} P_T) \Delta_{k-2}\|_{\mu(\infty)} \leq \left(\frac{1}{2}\right)^{k-1} \|UV^\top\|_{\mu(\infty)}.
\]

for each \(k\). Similarly, repeatedly applying (12), Lemma 7.4 and the inequality we just proved above, we obtain w.h.p.
\[
\|\Delta_{k-1}\|_{\mu(\infty,2)} \leq \left(\frac{1}{2}\right)^{k-1} \|UV^\top\|_{\mu(\infty)}.
\]

It follows that w.h.p.
\[
\|P_{T^\perp}(Y)\| \leq \frac{c}{\sqrt{n}} \sum_{k=1}^{k_0} (k + 1) \left(\frac{1}{2}\right)^{k-1} \|UV^\top\|_{\mu(\infty)} + \frac{c}{\sqrt{n}} \sum_{k=1}^{k_0} \left(\frac{1}{2}\right)^{k-1} \|UV^\top\|_{\mu(\infty,2)}.
\]

\[\leq \frac{6c}{\sqrt{n}} \|UV^\top\|_{\mu(\infty)} + \frac{2c}{\sqrt{n}} \|UV^\top\|_{\mu(\infty,2)}.
\]

Note that for all \((i, j)\), we have \(\left|(UV^\top)_{ij}\right| = |e_i^T UV^\top e_j| \leq \sqrt{\frac{\nu}{n}} \sqrt{\frac{r}{n}}, \|e_i^T UV^\top\|_2 = \sqrt{\frac{\nu}{n}}\) and \(\|UV^\top e_j\|_2 = \sqrt{\frac{\nu}{n}}\). Hence \(\|UV^\top\|_{\mu(\infty)} \leq 1\) and \(\|UV^\top\|_{\mu(\infty,2)} = 1\). We conclude that
\[
\|P_{T^\perp}(Y)\| \leq \frac{6c}{\sqrt{n}} + \frac{2c}{\sqrt{n}} \leq \frac{1}{2}
\]
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 \(P_3(X) = P_3(M)\). Let \(G\) be an \(n \times n\) matrix which satisfies \(\|P_{T^\perp} G\| = 1\), and \(\langle P_{T^\perp} G, P_{T^\perp} (X - M) \rangle = \|P_{T^\perp} (X - M)\|_s\). 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 \(f(Z) = \|Z\|_s\) at \(Z = M\), we have
\[
\|X\|_s - \|M\|_s \geq \langle UV^\top + P_{T^\perp} G, X - M \rangle.
\]
But \( \langle Y, X - M \rangle = \langle P_{\Omega}(Y), P_{\Omega}(X - M) \rangle = 0 \) since \( P_{\Omega}(Y) = Y \). It follows that
\[
\|X\|_* - \|M\|_* \geq \langle UV^T + P_{T^\perp} G - Y, X - M \rangle
= \|P_{T^\perp}(X - M)\|_* + \langle UV^T - P_{T^\perp} Y, X - M \rangle - \langle P_{T^\perp} Y, X - M \rangle
\geq \|P_{T^\perp}(X - M)\|_* - \|UV^T - P_{T^\perp} Y\|_F \|P_T(X - M)\|_F - \|P_{T^\perp} Y\|_F \|P_{T^\perp}(X - M)\|_*
\geq \frac{1}{2} \|P_{T^\perp}(X - M)\|_* - \frac{1}{4n^5} \|P_T(X - M)\|_F,
\]
where in the last inequality we use conditions 1 and 2 in the proposition. Using Lemma 7.6 below, we obtain
\[
\|X\|_* - \|M\|_* \geq \frac{1}{2} \|P_{T^\perp}(X - M)\|_* - \frac{1}{4n^5} \sqrt{2n^5} \|P_{T^\perp}(X - M)\|_* > \frac{1}{8} \|P_{T^\perp}(X - M)\|_*.
\]
The RHS is strictly positive for all \( X \) with \( P_{\Omega}(X - M) = 0 \) and \( X \neq M \). Otherwise we must have \( P_T(X - M) = X - M \) and \( P_T P_{\Omega} P_T(X - M) = 0 \), contradicting the assumption \( \|P_T P_{\Omega} P_T - P_T\|_{op} \leq \frac{1}{2} \). This proves that \( M \) is the unique optimum.

**Lemma 7.6.** If \( p_{ij} \geq \frac{1}{n^{10}} \) for all \( i, j \) and \( \|P_T P_{\Omega} P_T - P_T\|_{op} \leq \frac{1}{2} \), then we have
\[
\|P_T Z\|_F \leq \sqrt{2n^5} \|P_{T^\perp}(Z)\|_* \forall Z \in \{Z' : P_{\Omega}(Z') = 0\}. 
\]

**Proof.** Define the operator \( R_{\Omega}^{1/2} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n} \) by
\[
R_{\Omega}^{1/2}(Z) := \sum_{i,j} \frac{1}{\sqrt{p_{ij}}} \delta_{ij} \langle e_i e_j^T, Z \rangle e_i e_j^T.
\]
Note that \( R_{\Omega}^{1/2} \) is self-adjoint and satisfies \( R_{\Omega}^{1/2} R_{\Omega}^{1/2} = P_{\Omega} \). Hence we have
\[
\left\| R_{\Omega}^{1/2} P_T(Z) \right\|_F = \sqrt{\langle P_T R_{\Omega} P_T Z, P_T Z \rangle} \quad (23)
= \sqrt{\langle (P_T R_{\Omega} P_T - P_T) Z, P_T(Z) \rangle + \langle P_T(Z), P_T(Z) \rangle} \quad (24)
\geq \sqrt{\|P_T(Z)\|_F^2 - \|P_T R_{\Omega} P_T - P_T\|_F^2 \|P_T(Z)\|_F^2} \quad (25)
\geq \frac{1}{\sqrt{2}} \|P_T(Z)\|_F, \quad (26)
\]
where the last inequality follows from the assumption \( \|P_T R_{\Omega} P_T - P_T\|_{op} \leq \frac{1}{2} \). On the other hand, \( P_{\Omega}(Z) = 0 \) implies \( R_{\Omega}^{1/2}(Z) = 0 \) and thus
\[
\|R_{\Omega}^{1/2} P_T(Z)\|_F = \left\| R_{\Omega}^{1/2} P_{T^\perp}(Z) \right\|_F \leq \left( \max_{i,j} \frac{1}{\sqrt{p_{ij}}} \right) \|P_{T^\perp}(Z)\|_F \leq n^5 \|P_{T^\perp}(Z)\|_F.
\]
Combining the last two display equations gives
\[
\|P_T(Z)\|_F \leq \sqrt{2n^5} \|P_{T^\perp}(Z)\|_F \leq \sqrt{2n^5} \|P_{T^\perp}(Z)\|_*.
\]
\[\square\]

### 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
\[
\max \left\{ \frac{\mu_i}{n}, \frac{\nu_i}{n} \right\} \leq 1 \quad \text{and} \quad \frac{(\mu_i + \nu_j) r}{n} \geq \|P_T(e_i e_j^T)\|_F^2.
\]
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
\[
(P_T R_\Omega P_T - P_T)(Z) = \sum_{i,j} \left( \frac{1}{p_{ij}} \delta_{ij} - 1 \right) \langle e_i e_j^T, P_T(Z) \rangle P_T(e_i e_j^T) =: \sum_{i,j} S_{ij}(Z).
\]

Note that $\mathbb{E}[S_{ij}] = 0$ and $S_{ij}$'s are independent of each other. For all $Z$ and $(i,j)$, we have $S_{ij} = 0$ if $p_{ij} = 1$. On the other hand, when $p_{ij} \geq c_0 \left( \frac{\mu_i + \nu_j}{n} \right)^r \log n$, then it follows from (27) that
\[
\|S_{ij}(Z)\|_F \leq \frac{1}{p_{ij}} \|P_T(e_i e_j^T)\|_F^2 \|Z\|_F \leq \max_{i,j} \left\{ \frac{1}{p_{ij}} \left( \frac{\mu_i + \nu_j}{n} \right)^r \right\} \|Z\|_F \leq \frac{1}{c_0 \log n} \|Z\|_F.
\]

Putting together, we have that $\|S_{ij}\| \leq \frac{1}{c_0 \log n}$ under the condition of the lemma. On the other hand, we have
\[
\left\| \sum_{i,j} \mathbb{E} \left[ S_{ij}^2(Z) \right] \right\|_F \leq \frac{1}{c_0 \log n} \quad \text{under the condition of the lemma.}
\]

7.2.2. Proof of Lemma 7.3

We can write $(R_\Omega - I) Z$ as the sum of independent matrices:
\[
(R_\Omega - I) Z = \sum_{i,j} \left( \frac{1}{p_{ij}} \delta_{ij} - 1 \right) Z_{ij} e_i e_j^T =: \sum_{i,j} S_{ij}.
\]

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

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

The quantity $\left\| \mathbb{E} \left[ \sum_{i,j} S_{ij}^T S_{ij} \right] \right\|$ 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 $p_{ij} \geq 1 \land c_0 \left( \frac{\mu_i + \nu_j}{n} \right)^r \log n \geq 1 \land 2c_0 \sqrt{\frac{\|Z\|_F}{n}} \cdot \frac{\|Z\|_F}{n} \log n$, we have for all $i$ and $j$: $\|S_{ij}\| \log n \leq (1 - (p_{ij} = 1)) \frac{1}{p_{ij}} \|Z_{ij}\| \log n \leq \frac{1}{c_0} \|Z\|_F$ and $\sum_{i=1}^n \frac{1 - p_{ij}}{p_{ij}} Z_{ij}^2 \log n \leq \frac{1}{c_0} \|Z\|_2^2$. The second part of the lemma follows again from applying the matrix Bernstein inequality.
where we use

On the other hand, note that

where we use the triangle inequality and the definition of \( \mu_i \) and \( \nu_i \). Similarly, if \( j \neq b \), we have

Now note that \( \| S_{ij} \|_2 \leq (1 - (p_{ij} = 1)) \frac{1}{p_{ij}} |Z_{ij}| \sqrt{\frac{\mu_i r}{\nu_i r}} \left( \frac{\mu_i r}{n} + \frac{\nu_i r}{n} \right) \). Using the bounds (28) and (29), we obtain that for \( j = b \),

where we use \( p_{ib} \geq 1 \wedge \frac{c_0 \mu_i r \log n}{n} \) and \( p_{ib} \geq 1 \wedge c_0 \sqrt{\frac{\mu_i r \nu_i r}{n}} \log n \) in the second inequality. For \( j \neq b \), we have

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

On the other hand, note that

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

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

\[
\sum_{j \neq b} \leq \sum_{j \neq b} \leq \sum_{j \neq b} \frac{1}{p_{ij}} |Z_{ij}| \left( \frac{\mu_i r}{n} + \frac{\nu_i r}{n} \right) \left( \frac{\mu_i r}{n} \right) \frac{n}{\nu_i r} \| X_{ib} \|_2 \leq \frac{2}{c_0 \log n} \frac{n}{\nu_i r} \| X_{ib} \|_2 \leq \frac{2}{c_0 \log n} \| Z \|_{\mu(\infty, 2)}^2 ,
\]

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

\[
\left\| \sqrt{\frac{n}{\mu_a r}} X_a \right\|_2 = \left\| \sum_{i,j} S_{ij} \right\|_2 \leq \frac{1}{2} \left( \| Z \|_{\mu(\infty)} + \| Z \|_{\mu(\infty,2)} \right)
\]

for \( c_0 \) sufficiently large. Similarly we can bound \( \left\| \sqrt{\frac{n}{\mu_a r}} X_a \right\|_2 \) 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 \( w_{ab} = \sqrt{\frac{\nu_b r}{n}} \). We can write

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

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

\[
e_i^T U U^T e_a e_b^T e_j + e_i^T (I - U U^T) e_a e_b^T V V^T e_j \leq \frac{w_{ab}}{n} + \frac{\nu_b r}{n}, \quad i = a, j = b,
\]

\[
e_i^T (I - U U^T) e_a e_b^T e_j \leq \| e_b^T V V^T e_j \|, \quad i = a, j \neq b,
\]

\[
e_i^T U U^T e_a e_b^T e_j \leq \| e_a^T U U^T e_a \|, \quad i \neq a, j = b,
\]

\[
e_i^T U U^T e_a e_b^T V V^T e_j \leq \| e_b^T U U^T e_b \|, \quad i \neq a, j \neq b,
\]

where we use the fact that the matrices \( I - U U^T \) and \( I - V V^T \) have spectral norm at most 1. We proceed to bound \( |s_{ij}| \).

Note that

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

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

\[
|s_{ij}| \leq \frac{|Z_{ij}|}{(u_{ij} c_0 \log n)} \leq \frac{\left| \langle e_i e_j^T, P_T(e_a e_b^T) \rangle \right|}{c_0 \log n}.
\]

where (a) follows from \( p_{aj} \geq \min \left\{ c_0 \nu_b r \log n, 1 \right\} \). 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

\[
|s_{ij}| \leq (1 - I(p_{ij} = 1)) \frac{|Z_{ij}|}{p_{ij}} \cdot \sqrt{\frac{\mu_a r}{n} \frac{\nu_b r}{n}} \cdot \sqrt{\frac{n}{\mu_a r} \frac{n}{\nu_b r}} \leq \frac{1}{c_0 \log n} \cdot \sqrt{\frac{n}{\mu_a r} \frac{n}{\nu_b r}} \frac{1}{c_0 \log n} \leq \frac{\left| \langle e_i e_j^T, P_T(e_a e_b^T) \rangle \right|}{c_0 \log n},
\]

where (b) follows from \( p_{ij} \geq 1 \wedge c_0 \sqrt{\frac{\nu_b r r}{n} \log n} \) and max \( \left\{ \sqrt{\frac{\nu_b r}{n}}, \sqrt{\frac{\nu_a r r}{n}} \right\} \leq 1 \). We conclude that \( |s_{ij}| \leq \frac{\left| \langle e_i e_j^T, P_T(e_a e_b^T) \rangle \right|}{c_0 \log n} \) for all \((i, j)\).

On the other hand, note that

\[
E \left[ \sum_{i,j} s_{ij}^2 \right] = \sum_{i,j} E \left[ \left( \frac{1}{p_{ij}} \delta_{ij} - 1 \right)^2 \right] \frac{Z_{ij}^2}{w_{ab}^2} \left( e_i e_j^T, P_T(e_a e_b^T) \right)^2 = \sum_{i=a,j=b} + \sum_{i=a,j\neq b} + \sum_{i\neq a,j=b} + \sum_{i\neq a,j\neq b}.
\]
We bound each of the four sums. By (30) and \( p_{ab} \geq 1 \land \frac{c_0(\mu_n + \nu_0)^2 \log n}{n} \geq 1 \land \frac{c_0(\mu_n + \nu_0)^2 r^2 \log n}{2n^*} \), we have

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

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

\[
\sum_{i=a, j \neq b} \leq \sum_{j \neq b} \frac{1 - p_{aj}}{p_{aj} w_{ab}^2} Z_{aj}^2 |e_i^T VV^T e_j| \leq \frac{\|Z\|_{\mu(\infty)}^2}{c_0 \log n} \frac{n}{\mu_a r} \sum_{j \neq b} |e_i^T VV^T e_j|,
\]

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

\[
\sum_{i \neq a, j \neq b} \leq \frac{1}{w_{ab}^2} \sum_{i \neq a, j \neq b} \frac{(1 - p_{ij}) Z_{ij}^2}{p_{ij}} \cdot |e_i^T UU^T e_a| |e_b^T VV^T e_j| \leq \frac{\|Z\|_{\mu(\infty)}^2}{c_0 \log n} \frac{1}{w_{ab}^2} \sum_{i \neq a} |e_i^T UU^T e_a| \sum_{j \neq b} |e_b^T VV^T e_j|,
\]

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

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

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

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

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 \( U \Sigma V^T \) is the SVD of \( M \).

**Lemma 8.1.** If \( \max_i \|U^T e_i\| = \max_i \sqrt{\frac{\mu_a r}{n}} \leq \sqrt{\frac{\mu_a r}{n}} \) and \( p \geq c_0 \frac{\mu_a r \log n}{n} \) for some universal constant \( c_0 \), then with high probability,

\[
\|U^T P_J(U) - I_{r \times r}\| \leq \frac{1}{2},
\]

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

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

\[
U^T P_J(U) - I_{r \times r} = U^T P_J(U) - U^T U = S(i) := \sum_{i=1}^n \left( \frac{1}{p} \eta_i - 1 \right) U^T e_i e_i^T U.
\]
Note that $\mathbb{E} [S_{(i)}] = 0$, $\|S_{(i)}\| \leq \frac{1}{p} \|U^T e_i\|_2^2 \leq \frac{\mu_0}{pn}$, and

$$
\mathbb{E} \left[ \sum_{i=1}^{n} S_{(i)} S_{(i)}^T \right] = \left\| \sum_{i=1}^{n} S_{(i)} S_{(i)}^T \right\| = \frac{1 - p}{p} \left\| \sum_{i=1}^{n} U^T e_i e_i^T U U^T e_i e_i^T U \right\|
$$

$$
\leq \frac{1}{p} \left\| \sum_{i=1}^{n} e_i e_i^T \right\|_2 \left\| U U^T e_i e_i^T U \right\|
$$

It follows from the matrix Bernstein (Theorem 10.1) that

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

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

Note that $\left\| U^T P_j(U) - I_{r \times r} \right\| \leq \frac{1}{2}$ implies that $U^T P_j(U)$ is invertible, which further implies $P_j(U) \in \mathbb{R}^{n \times r}$ has rank-$r$. The rows picked are $P_j(M) = P_j(U) \Sigma V^T$, 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

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

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 $\bar{U} \Sigma \bar{V}^T$; so $\bar{U} \Sigma \bar{V}^T = RMC = RU \Sigma V^T C$. By definition, we have

$$
\frac{\mu_0 r}{n} = \|P_{\bar{U}}(e_i)\|_2^2,
$$

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

$$
P_{\bar{U}}(e_i) = RU \left( U^T R^2 U \right)^{-1} U^T R e_i.
$$

It follows that

$$
\frac{\mu_0 r}{n} = \left\| RU \left( U^T R^2 U \right)^{-1} U^T R e_i \right\|_2^2
$$

$$
= R_i^2 e_i^T U \left( U^T R^2 U \right)^{-1} U^T e_i
$$

$$
\leq R_i^2 \left( \sigma_r(RU) \right)^{-2} \left\| U^T e_i \right\|_2^2
$$

$$
\leq R_i^2 \frac{\mu_0 r}{n} \left( \sigma_r(RU) \right)^{-2}, \tag{31}
$$
where \( \sigma_r(\cdot) \) denotes the \( r \)-th singular value and the last inequality follows from the standard incoherence assumption 
\[ \max_{i,j} \{\mu_i, \nu_j\} \leq \mu_0. \] 
We now bound \( \sigma_r(RU) \). Since \( RU \) has rank \( r \), we have
\[
\sigma^2_r(RU) = \min_{\|x\|=1} \|RUx\|^2 = \min_{\|x\|=1} \sum_{i=1}^n R_i^2 |e_i^T U x|^2.
\]
If we let \( z_i := |e_i^T U x|^2 \) for each \( i \in [n] \), then \( z_i \) satisfies
\[
\sum_{i=1}^n z_i = \|U x\|^2 = \|x\|^2 = 1
\]
and by the standard incoherence assumption,
\[
z_i \leq \|U^T e_i\|^2 \|x\|^2 \leq \frac{\mu_0 r}{n}.
\]
Therefore, the value of the minimization above is lower-bounded by
\[
\min_{z \in \mathbb{R}^n} \sum_{i=1}^n R_i^2 z_i \quad \text{s.t.} \quad \sum_{i=1}^n z_i = 1, \quad 0 \leq z_i \leq \frac{\mu_0 r}{n}, \quad i = 1, \ldots, n.
\]
(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
\[
\sum_{i=1}^n z^*_i = 1
\]
\[
z^*_i = 0, \quad \text{for } i \in I_1
\]
\[
z^*_i = \frac{\mu_0 r}{n}, \quad \text{for } i \in I_2
\]
for some index sets \( I_1 \) and \( I_2 \) such that \( |I_1| + |I_2| = n - 1 \). It is easy to see that we must have \( |I_2| = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor \).

Since \( R_1 \leq R_2 \leq \ldots \leq R_n \), the minimizer \( z^* \) has the form
\[
z^*_i = \frac{\mu_0 r}{n}, \quad i = 1, \ldots, \left\lfloor \frac{n}{\mu_0 r} \right\rfloor,
\]
\[
z^*_i = 1 - \left\lfloor \frac{n}{\mu_0 r} \right\rfloor \frac{\mu_0 r}{n}, \quad i = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor + 1,
\]
\[
z^*_i = 0, \quad i = \left\lfloor \frac{n}{\mu_0 r} \right\rfloor + 2, \ldots, n,
\]
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 \( \sigma^2_r(RU) \geq \frac{\mu_0 r}{n} \sum_{i=1}^{\lfloor n/(\mu_0 r) \rfloor} R_i^2 \). Combining with (31), we obtain that
\[
\frac{\tilde{\mu}_1 r}{n} \leq \frac{R^2}{\sum_{i=1}^{\lfloor n/(\mu_0 r) \rfloor} R_i^2}, \quad \frac{\tilde{\nu}_1 r}{n} \leq \frac{C^2_j}{\sum_{j=1}^{\lfloor n/(\mu_0 r) \rfloor} C^2_j};
\]
the proof for \( \tilde{\nu}_j \) is similar. Applying Theorem 3.2 to the equivalent problem (7) with the above bounds on \( \tilde{\mu}_i \) and \( \tilde{\nu}_j \) 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

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

(33)

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

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

(34)

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