7 Appendix

7.1 Preliminaries

We first revisit some basic properties of defined linear operators and projections. Recall that $H_0 = U\Sigma V^T$ is the reduced SVD of H_0, and the space T is defined as:

$$T := \{UA^T + BV^T \mid A, B \in \mathbb{R}^{d\times r}\},$$

and $\mathcal{P}_T$ is the orthogonal projection onto T. It is known that any subgradient of $\|H_0\|_*$ has the form $UV^T + W$, where $\mathcal{P}_T W = 0, \|W\| \le 1$.

Similarly, we have defined Ω to be the set of matrices whose entries supported as the same as S_0, and $\mathcal{P}_\Omega$ is the orthogonal projection onto Ω. It is also known that any subgradient of $\|S_0\|_1$ takes the form $\text{sgn}(S_0) + F$, where $\mathcal{P}_\Omega F = 0, \|F\|_\infty \le 1$.

Under the incoherence assumptions, we also introduce a norm inequality on rank-1 matrices which we will use frequently in the proof. Given any matrix with the form $\mathbf{x}_i\mathbf{y}_j^T \in \mathbb{R}^{d\times d}$, we have

$$\begin{aligned}\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F^2 &= \langle \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T), \mathbf{x}_i\mathbf{y}_j^T\rangle \\ &\le \|U^T\mathbf{x}_i\|_2^2\|\mathbf{y}_j\|_2^2 + \|V^T\mathbf{y}_j\|_2^2\|\mathbf{x}_i\|_2^2 \\ &\le \frac{2\mu_0\mu_1 rd}{n^2}. \end{aligned} \tag{10}$$

In particular, if we let p be any probability that satisfies:

$$p \ge 2C\epsilon^{-2}\frac{\mu_0\mu_1 rd\log d}{n^2} \tag{11}$$

with a numerical constant $C > 0$, then the inequality becomes:

$$\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F^2 \le \epsilon^2\frac{p}{C\log d}. \tag{12}$$

7.2 Proof of Lemma 1

Here, we provide a proof of dual certification lemma (Lemma 1).

Proof. Consider any feasible perturbation $(H_0 + \Delta, S_0 - X\Delta Y^T)$ from the claimed optimum. We will prove the lemma by showing that such perturbed pair increases the objective (3) unless $\Delta = 0$. Let $UV^T + W_0$ be any subgradient of $\|H_0\|_*$ and $\text{sgn}(S_0) + F_0$ be any subgradient of $\|S_0\|_1$, then by the definition of subgradient, $W_0 \in T^\perp, \|W_0\|_2 \le 1$, $F_0 \in \Omega^\perp, \|F_0\|_\infty \le 1$, and

$$\begin{aligned}&\|H_0+\Delta\|_* + \lambda\|S_0 - X\Delta Y^T\|_1 \\ \ge&\|H_0\|_* + \lambda\|S_0\|_1 + \langle UV^T + W_0, \Delta\rangle \\ &- \lambda\langle \text{sgn}(S_0) + F_0, X\Delta Y^T\rangle.\end{aligned}$$

Select W_0 and F_0 such that $\langle W_0, \Delta\rangle = \|\mathcal{P}_{T^\perp}\Delta\|_*$ and $\langle F_0, X\Delta Y^T\rangle = -\|\mathcal{P}_{\Omega^\perp}(X\Delta Y^T)\|_1$ [5], then we have:

$$\begin{aligned}&\|H_0+\Delta\|_* + \lambda\|S_0 - X\Delta Y^T\|_1 \\ \ge&\|H_0\|_* + \|S_0\|_1 + \|\mathcal{P}_{T^\perp}\Delta\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}(X\Delta Y^T)\|_1 \\ &+ \langle UV^T, \Delta\rangle - \lambda\langle\text{sgn}(S_0), X\Delta Y^T\rangle. \end{aligned}\tag{13}$$

Now, since $\langle UV^T, \Delta\rangle = \langle \mathcal{T}_L(UV^T), \mathcal{T}_L\Delta\rangle$, we can bound the inner product terms by:

$$\begin{aligned}&|\langle UV^T, \Delta\rangle - \lambda\langle\text{sgn}(S_0), X\Delta Y^T\rangle| \\ =&|\langle \mathcal{T}_L(UV^T) - \lambda\text{sgn}(S_0), X\Delta Y^T\rangle| \\ \le&|\langle \mathcal{T}_L(W), X\Delta Y^T\rangle| + |\langle M, X\Delta Y^T\rangle| \\ &+ |\lambda\langle F, X\Delta Y^T\rangle| + |\lambda\langle \mathcal{P}_\Omega D, X\Delta Y^T\rangle| \\ \le&\frac{1}{2}\|\mathcal{P}_{T^\perp}\Delta\|_* + \frac{\lambda}{2}\|\mathcal{P}_{\Omega^\perp}X\Delta Y^T\|_1 + \frac{\lambda}{4}\|\mathcal{P}_\Omega X\Delta Y^T\|_F,\end{aligned}$$

where in the third inequality we use the fact that $\langle M, X\Delta Y^T\rangle = \langle M, \mathcal{P}_Q(X\Delta Y^T)\rangle = 0$. Thus, equation (13) can be reduced to:

$$\begin{aligned}&\|H_0+\Delta\|_* + \lambda\|S_0 - X\Delta Y^T\|_1 \\ \ge&\|H_0\|_* + \|S_0\|_1 + \frac{1}{2}(\|\mathcal{P}_{T^\perp}\Delta\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}(X\Delta Y^T)\|_1) \\ &- \frac{\lambda}{4}\|\mathcal{P}_\Omega X\Delta Y^T\|_F. \end{aligned}\tag{14}$$

We can further bound the term $\|\mathcal{P}_\Omega X\Delta Y^T\|_F$ by:

$$\begin{aligned}&\|\mathcal{P}_\Omega X\Delta Y^T\|_F \\ \le&\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\Delta\|_F + \|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_{T^\perp}\Delta\|_F \\ \le&\frac{1}{2}\|\Delta\|_F + \|\mathcal{P}_{T^\perp}\Delta\|_F \\ \le&\frac{1}{2}(\|\mathcal{P}_\Omega\mathcal{T}_L\Delta\|_F + \|\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\Delta\|_F) + \|\mathcal{P}_{T^\perp}\Delta\|_F.\end{aligned}$$

By definition, $\mathcal{P}_\Omega\mathcal{T}_L\Delta = \mathcal{P}_\Omega = X\Delta Y^T$, so

$$\begin{aligned}\|\mathcal{P}_\Omega\mathcal{T}_L\Delta\|_F &\le \|\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\Delta\|_F + 2\|\mathcal{P}_{T^\perp}\Delta\|_F \\ &\le \|\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\Delta\|_1 + 2\|\mathcal{P}_{T^\perp}\Delta\|_*.\end{aligned}$$

Therefore, equation (14) becomes:

$$\begin{aligned}&\|H_0+\Delta\|_* + \lambda\|S_0 - X\Delta Y^T\|_1 \\ \ge&\|H_0\|_* + \|S_0\|_1 \\ &+ \frac{1}{2}\left((1-\lambda)\|\mathcal{P}_{T^\perp}\Delta\|_* + \frac{\lambda}{2}\|\mathcal{P}_{\Omega^\perp}(X\Delta Y^T)\|_1\right). \end{aligned}\tag{15}$$

However, by assumption, $\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| \le \frac{1}{2} < 1$ implies that $\mathcal{T}_L(T)\cap\Omega = \{0\}$. Therefore, for any $\Delta \ne 0$, if $\Delta \notin T$ then $\|\mathcal{P}_{T^\perp}\Delta\| > 0$, and if $\Delta \in T$ then $\|\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\Delta\|_1 > 0$. Thus the LHS of (15) will be strictly larger than RHS unless $\Delta = 0$, which concludes the proof. □

7.3 Preliminary Lemmas

We need several lemmas to prove the validity of constructed dual certificates introduced in Section 4.3. For fol-

[5] Such W_0 and F_0 exist. See Candès et al. (2011) for an example of such matrices.

lowing lemmas, when we say the equation holds with large probability, we mean that the event will hold with probability at least $1 - O(d^{-10})$.

Most of the probability bounds in our results are from the Bernstein inequality stated as below.

Proposition 1 (Noncommutative Matrix Bernstein Inequality (Recht, 2011))**.** *Let $X_1 \cdots X_k$ be k independent, zero-mean random matrices where each $X_i \in \mathbb{R}^{n_1 \times n_2}$. Suppose for each X_i, $\|X_i\| \leq R$, and the norm of the sum of covariance matrices is bounded by:*

$$\max\left\{\|\sum_{i=1}^{k} \mathbb{E}\left[X_i X_i^T\right]\|, \|\sum_{i=1}^{k} \mathbb{E}\left[X_i^T X_i\right]\|\right\} \leq \sigma^2.$$

Then for any $t > 0$:

$$\Pr(\|\sum_{i=1}^{k} X_i\| > t) \leq (n_1 + n_2) \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

We begin with a core lemma which generalizes the result of Theorem 4.1 in Candès & Recht (2012).

Lemma 4. *Suppose $\Omega_0 \sim Ber(\rho)$. Then with large probability,*

$$\|\mathcal{P}_T - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L\mathcal{P}_T\| \leq \epsilon$$

provided that $\rho \geq C_0\epsilon^{-2}(2\mu_0\mu_1 rd\log d)/n^2$ with some constant $C_0 > 0$.

Proof. First we decompose the matrix $(\mathcal{P}_T - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L\mathcal{P}_T)Z$ as:

$$\begin{aligned}
&(\mathcal{P}_T - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L\mathcal{P}_T)Z\\
=&(\mathcal{P}_T\mathcal{T}_S(\mathcal{I} - \rho^{-1}\mathcal{P}_{\Omega_0})\mathcal{T}_L\mathcal{P}_T)Z\\
=&\sum_{(i,j)}(1-\rho^{-1}\delta_{ij})\langle Z, \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\rangle\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T).
\end{aligned}$$

This yields us to define a linear operator $\mathcal{S}_{ij}$ as:

$$\mathcal{S}_{ij}(Z) = (1-\rho^{-1}\delta_{ij})\langle Z, \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\rangle\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T),$$

which maps any $Z \in \mathbb{R}^{d\times d}$ to $\mathbb{R}^{d\times d}$. The operator is symmetric, zero in expectation (i.e., $\mathbb{E}\left[\mathcal{S}_{ij}(Z)\right] = 0$) and its operator norm, by definition, is bounded by:

$$\sup_{Z\neq 0}\frac{\|\mathcal{S}_{ij}(Z)\|_F}{\|Z\|_F}.$$

Thus, the original operator $\mathcal{P}_T - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L\mathcal{P}_T$ can be viewed as a sum of independent, zero-mean operators $\mathcal{S}_{ij}$, where each operator has a bounded operator norm as:

$$\begin{aligned}
\|\mathcal{S}_{ij}(Z)\|_F &\leq \rho^{-1}|\langle Z, \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\rangle|\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F\\
&\leq \rho^{-1}\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F^2\|Z\|_F\\
&\leq \frac{\epsilon^2}{C_0\log d}\|Z\|_F,
\end{aligned}$$

where the last line is derived by applying (12). Also, we can bound the quantity $\|\sum_{(i,j)}\mathbb{E}\left[S_{ij}^2\right]\|$ similarly. Since

$$\begin{aligned}
&\|\sum_{(i,j)}\mathbb{E}\left[\mathcal{S}_{ij}^2(Z)\right]\|_F\\
=&\Bigg\|\sum_{(i,j)}\mathbb{E}\left[(1-\rho^{-1}\delta_{ij})^2\right]\langle Z, \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\rangle\\
&\qquad\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F^2\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\Bigg\|_F,
\end{aligned}$$

and $\mathbb{E}\left[(1-\rho^{-1}\delta_{ij})^2\right] = (1-\rho)/\rho \leq 1/\rho$, therefore,

$$\begin{aligned}
&\|\sum_{(i,j)}\mathbb{E}\left[\mathcal{S}_{ij}^2(Z)\right]\|_F\\
\leq&\frac{\epsilon^2}{C_0\log d}\|\mathcal{P}_T\sum_{(i,j)}\langle\mathcal{P}_T Z, \mathbf{x}_i\mathbf{y}_j^T\rangle\mathbf{x}_i\mathbf{y}_j^T\|_F\\
=&\frac{\epsilon^2}{C_0\log d}\|\mathcal{P}_T\mathcal{T}_S\mathcal{T}_L\mathcal{P}_T(Z)\|_F\\
\leq&\frac{\epsilon^2}{C_0\log d}\|Z\|_F
\end{aligned}$$

With above bounds, the claim follows by applying matrix Bernstein inequality. □

An important fact from this lemma is that it implies $\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\|$ will not be too large provided that $|\Omega|$ is not extremely large. More formally, we can prove the following Lemma:

Lemma 5. *Suppose $\Omega \sim Ber(\rho)$ where $1-\rho \geq C_0\epsilon^{-2}(2\mu_0\mu_1 rd\log d)/n^2$. Then with high probability, we have $\|\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\| \leq \sqrt{\rho+\epsilon}$.*

Proof. Suppose $1-\rho \geq C_0\epsilon^{-2}(2\mu_0\mu_1 rd\log d)/n^2$, then from Lemma 4, we know that with high probability,

$$\|\mathcal{P}_T - (1-\rho)^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\mathcal{P}_T\| \leq \epsilon.$$

Now, by the fact that $\mathcal{P}_{\Omega^\perp} = \mathcal{I} - \mathcal{P}_\Omega$, we can rewrite the operator as:

$$\begin{aligned}
&\mathcal{P}_T - (1-\rho)^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega^\perp}\mathcal{T}_L\mathcal{P}_T\\
=&(1-\rho)^{-1}(\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T - \rho\mathcal{P}_T),
\end{aligned}$$

from which we can conclude that

$$\|\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| \leq \epsilon(1-\rho) + \rho\|\mathcal{P}_T\| = \rho + \epsilon(1-\rho)$$

by the triangle inequality. The claim is thus proved by the fact that $\|\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| \leq \|\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\|^2$. □

Lemma 4 implies that if $Z \in T$, then its Frobenius norm will decrease sufficiently large after applying the operator $\mathcal{I} - \mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L$. The next lemma says that, after applying such operator, its "$\mathcal{T}_L$ infinity norm" will also decrease sufficiently large.

Lemma 6. *Suppose $\Omega_0 \sim Ber(\rho)$ and $Z \in T$. Then with large probability,*

$$\|\mathcal{T}_L(Z - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L Z)\|_\infty \leq \epsilon\|\mathcal{T}_L Z\|_\infty$$

provided that $\rho \geq C_0\epsilon^{-2}(2\mu_0\mu_1 rd\log d)/n^2$ with some constant $C_0 > 0$.

Proof. Let $K = \mathcal{T}_L(Z - \rho^{-1}\mathcal{P}_T\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L Z)$. Observe that any element K_{ab} can be represented as a sum of independent variables, i.e. $K_{ab} = \sum_{(i,j)} s_{ij}$, where s_{ij} is defined as:

$$s_{ij} = (1-\rho^{-1}\delta_{ij})\langle Z, \mathbf{x}_i\mathbf{y}_j^T\rangle\langle \mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T), \mathbf{x}_a\mathbf{y}_b^T\rangle.$$

Again, each s_{ij} has zero mean ($\mathbb{E}\big[s_{ij}\big] = 0$), and each $|s_{ij}|$ can be bounded by:

$$\begin{aligned}|s_{ij}| &\leq \rho^{-1}|\mathbf{x}_i^T Z\mathbf{y}_j|\|\mathcal{P}_T(\mathbf{x}_i\mathbf{y}_j^T)\|_F\|\mathcal{P}_T(\mathbf{x}_a\mathbf{y}_b^T)\|_F \\ &\leq \frac{\epsilon^2}{C_0\log d}\|\mathcal{T}_L Z\|_\infty.\end{aligned}$$

Also, $|\sum_{(i,j)}\mathbb{E}\big[s_{ij}^2\big]|$ can be bounded by:

$$\begin{aligned}|\sum_{(i,j)}\mathbb{E}\big[s_{ij}^2\big]| &\leq |\sum_{(i,j)}\rho^{-1}(\mathbf{x}_i^T Z\mathbf{y}_j)^2\langle\mathbf{x}_i\mathbf{y}_j^T, \mathcal{P}_T(\mathbf{x}_a\mathbf{y}_b^T)\rangle^2| \\ &\leq \rho^{-1}\|\mathcal{T}_L Z\|_\infty^2|\sum_{(i,j)}\langle\mathbf{x}_i\mathbf{y}_j^T, \mathcal{P}_T(\mathbf{x}_a\mathbf{y}_b^T)\rangle^2| \\ &\leq \rho^{-1}\|\mathcal{T}_L Z\|_\infty^2\|\mathcal{T}_L\mathcal{P}_T(\mathbf{x}_a\mathbf{y}_b^T)\|_F^2 \\ &\leq \frac{\epsilon^2}{C_0\log d}\|\mathcal{T}_L Z\|_\infty^2.\end{aligned}$$

Note that in both bounds we apply the inequality (12) because ρ obeys (11). Therefore, by Bernstein inequality, we have:

$$\Pr(|K_{ab}| > \epsilon\|\mathcal{T}_L Z\|_\infty) \leq 2\exp\left(-\frac{3}{8}C_0\log d\right),$$

and the claim is proved by applying an union bound. ◻

Lemma 7. *For any fixed matrix $Z \in \mathbb{R}^{d\times d}$, with large probability,*

$$\|(I - \rho^{-1}\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L)Z\| \leq C_0'\sqrt{\frac{d\log d}{\rho}}\|\mathcal{T}_L Z\|_\infty$$

with some constant $C_0' > 0$, provided that $\rho \geq C_0\mu_1^2 d\log d/n^2$ with some constant $C_0 > 0$.

Proof. Again we can decompose the matrix $(I - \rho^{-1}\mathcal{T}_S\mathcal{P}_{\Omega_0}\mathcal{T}_L)Z$ as $\sum_{(i,j)} S_{ij}$, where S_{ij} is defined as:

$$S_{ij} = (1-\rho^{-1}\delta_{ij})\langle Z, \mathbf{x}_i\mathbf{y}_j^T\rangle\mathbf{x}_i\mathbf{y}_j^T.$$

Each S_{ij} is independent with zero means (i.e. $\mathbb{E}\big[S_{ij}\big] = 0$). Furthermore, we can bound $\|S_{ij}\|$ by:

$$\|S_{ij}\| \leq \rho^{-1}|\mathbf{x}_i^T Z\mathbf{y}_j|\|\mathbf{x}_i\|_2\|\mathbf{y}_j\|_2 \leq \rho^{-1}\frac{\mu_1 d}{n}\|\mathcal{T}_L Z\|_\infty,$$

and the term $\|\sum_{(i,j)}\mathbb{E}\big[S_{ij}^T S_{ij}\big]\|$ can be bounded by:

$$\begin{aligned}&\|\sum_{(i,j)}\mathbb{E}\big[S_{ij}^T S_{ij}\big]\| \\ =&\|\sum_{(i,j)}\mathbb{E}\big[(1-\rho^{-1}\delta_{ij})^2\big](\mathbf{x}_i^T X\mathbf{y}_j)^2\mathbf{y}_j\mathbf{x}_i^T\mathbf{x}_i\mathbf{y}_j^T\| \\ \leq&\rho^{-1}\|\mathcal{T}_L Z\|_\infty^2\|\sum_i\|\mathbf{x}_i\|_2^2\sum_j\mathbf{y}_j^T\mathbf{y}_j\| \\ =&\rho^{-1}d\|\mathcal{T}_L Z\|_\infty^2\|Y^T Y\| \\ =&\rho^{-1}d\|\mathcal{T}_L Z\|_\infty^2.\end{aligned}$$

Same bound on $\|\sum_{(i,j)}\mathbb{E}\big[S_{ij}S_{ij}^T\big]\|$ can be derived similarly. Thus, the lemma follows by applying matrix Bernstein inequality. ◻

Equipped with the above lemmas, now we are able to prove Lemma 2. For convenience, we will take $\epsilon \leq e^{-1}$ in the proof.

7.4 Proof of Lemma 2

proof of 2a. Recall that by the definition of Y_j and Z_j, $Y_{j_0} = \sum_j q^{-1}\mathcal{T}_S\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1}$. Thus,

$$\begin{aligned}\|W^L\| &= \|\mathcal{P}_{T^\perp}Y_{j_0}\| \\ &\leq \sum_j\|q^{-1}\mathcal{P}_{T^\perp}\mathcal{T}_S\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1}\| \\ &= \sum_j\|\mathcal{P}_{T^\perp}(q^{-1}\mathcal{T}_S\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1} - Z_{j-1})\| \\ &\leq \sum_j\|q^{-1}\mathcal{T}_S\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1} - Z_{j-1}\|,\end{aligned}$$

where the second equality comes from $\mathcal{P}_{T^\perp}Z_{j-1} = 0$. As q is chosen to obey (11), we can apply Lemma 7 so that:

$$\begin{aligned}\|W^L\| &\leq C_0'\sqrt{\frac{d\log d}{q}}\sum_j\|\mathcal{T}_L Z_{j-1}\|_\infty \\ &\leq C_0'\sqrt{\frac{d\log d}{q}}\sum_j\epsilon^{j-1}\|\mathcal{T}_L(UV^T)\|_\infty \\ &\leq C_0'(1-\epsilon)^{-1}\sqrt{\frac{d\log d}{q}}\frac{\sqrt{\mu_0 r}}{n}.\end{aligned}$$

From here we can conclude that

$$\|W^L\| \leq C'\epsilon \leq \frac{1}{4}$$

for some universal constant C', by choosing a small enough ϵ. ◻

proof of 2b. We have

$$\begin{aligned}&\mathcal{P}_\Omega\mathcal{T}_L(UV^T + W^L) + \mathcal{P}_\Omega M^L \\ =&\mathcal{P}_\Omega\mathcal{T}_L(UV^T - \mathcal{P}_T Y_{j_0}) + \mathcal{P}_\Omega\mathcal{T}_L Y_{j_0} + \mathcal{P}_\Omega M^L \\ =&\mathcal{P}_\Omega\mathcal{T}_L(Z_{j_0}) + \mathcal{P}_\Omega(\mathcal{T}_L Y_{j_0} + M^L) \\ =&\mathcal{P}_\Omega\mathcal{T}_L(Z_{j_0}),\end{aligned}$$

where the last equation holds because:

$$\begin{aligned}&\mathcal{T}_L Y_{j_0} + M^L\\ =&\sum_j q^{-1}\mathcal{T}_L\mathcal{T}_S\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1} + \mathcal{P}_{Q^\perp}\sum_j q^{-1}\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1}\\ =&\sum_j q^{-1}\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1} \qquad (16)\end{aligned}$$

is a matrix only supported on Ω^C. Now, by applying Lemma 4, we have

$$\begin{aligned}\|\mathcal{P}_\Omega\mathcal{T}_L Z_{j_0}\|_F \le& \|\mathcal{T}_L(Z_{j_0})\|_F = \|Z_{j_0}\|_F\\ \le& \epsilon^{j_0}\|UV^T\|_F = \epsilon^{j_0}\sqrt{r}.\end{aligned}$$

Since $\epsilon \le e^{-1}$ and $j_0 \ge 2\log n$, the above quantity is less than $\lambda/4$. □

proof of 2c. By construction, $\mathcal{T}_L(UV^T + W^L) + M^L = \mathcal{T}_L Z_{j_0} + \mathcal{T}_L Y_{j_0} + M^L$. From part b we have $\|\mathcal{T}_L Z_{j_0}\|_\infty \le \|\mathcal{T}_L Z_{j_0}\|_F \le \lambda/4$, and the matrix $\mathcal{T}_L Y_{j_0} + M^L$ is supported on Ω^C. Thus, the claim is proved if we can show:

$$\|\mathcal{T}_L Y_{j_0} + M^L\|_\infty \le \frac{\lambda}{8}.$$

Using (16), we have:

$$\begin{aligned}\|\mathcal{T}_L Y_{j_0} + M^L\|_\infty \le& q^{-1}\sum_j \|\mathcal{P}_{\Omega_j}\mathcal{T}_L Z_{j-1}\|_\infty\\ \le& q^{-1}\sum_j \|\mathcal{T}_L Z_{j-1}\|_\infty\\ \le& q^{-1}\sum_j \epsilon^{j-1}\|\mathcal{T}_L(UV^T)\|_\infty\\ \le& q^{-1}(1-\epsilon)^{-1}\frac{\sqrt{\mu_0 r}}{n}.\end{aligned}$$

For q obeys (11), we have:

$$\|\mathcal{T}_L Y_{j_0} + M^L\|_\infty \le C\epsilon^2\sqrt{\frac{n^2}{\mu_0^2\mu_1 r d^2(\log d)^2}},$$

which will be smaller than $\lambda/8$ if:

$$\epsilon \le C'\left(\frac{\mu_0^2\mu_1 r d^2(\log d)^2}{n^3}\right)^{1/4}.$$

□

In summary, the proof above shows that 2a ∼2c hold if q is chosen to obey (11) and ϵ is chosen to be sufficiently small. As we fix a $j_0 \ge \lceil 2\log n\rceil$ and a small enough ϵ, a well-defined q can always be set to obey $1 > q \ge 2C\epsilon^{-2}(\mu_0\mu_1 r d\log d)/n^2$. This concludes the proof.

7.5 Proof of Lemma 3

For convenience, define $E = \text{sgn}(S_0)$ whose sign is randomly distributed as:

$$E_{ij} := \begin{cases}1, & \text{w.p. } \rho/2\\ 0, & \text{w.p. } 1\text{-}\rho\\ -1, & \text{w.p. } \rho/2\end{cases}$$

In the following two parts of proof, we will focus on the event $\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| < \sigma$. Notice that by Lemma 5, for any $\sigma > 0$, the event holds with large probability given a small enough ρ.

Proof of 3a. By construction, we have:

$$\begin{aligned}W^S =& \lambda\mathcal{P}_{T^\perp}\mathcal{T}_S(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}E\\ =& \mathcal{P}_{T^\perp}\mathcal{T}_S K^{(1)} + \mathcal{P}_{T^\perp}\mathcal{T}_S K^{(2)}, \qquad (17)\end{aligned}$$

where $K^{(1)}, K^{(2)}$ is defined by

$$\begin{aligned}K^{(1)} =& \lambda E,\\ K^{(2)} =& \lambda\sum_{k\ge1}(\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^k E.\end{aligned}$$

We first bound the first term of (17). Since $\|\mathcal{P}_{T^\perp}\mathcal{T}_S K^{(1)}\| \le \|K^{(1)}\| \le \|\lambda E\|$, thus, using the argument in both Vershynin (2010); Candès et al. (2011), with high probability,

$$\|E\| \le 4\sqrt{n\rho}.$$

As $\lambda = 1/\sqrt{n}$, it implies:

$$\|\mathcal{P}_{T^\perp}\mathcal{T}_S K^{(1)}\| \le \|\lambda E\| \le 4\sqrt{\rho}. \qquad (18)$$

Now consider the second term $\|\mathcal{P}_{T^\perp}\mathcal{T}_S K^{(2)}\|$. For convenience, set the operator $\mathcal{R} := \sum_{k\ge1}(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^k$. Then, $\|\mathcal{P}_{T^\perp}\mathcal{T}_S K^{(2)}\| \le \|K^{(2)}\| \le \|\lambda\mathcal{R}(E)\|$, and a standard covering argument could bound this operator norm. By Lemma 5.2 in Vershynin (2010), There exists a $1/2-$net N for a hypersphere S^{n-1} with its size $\le 5^n$. Then, From Lemma 5.3 in Vershynin (2010), we have:

$$\|\mathcal{R}(E)\| = \sup_{\mathbf{a},\mathbf{b}\in S^{n-1}}\langle\mathbf{a},\mathcal{R}(E)\mathbf{b}\rangle \le 4\sup_{\mathbf{a},\mathbf{b}\in N}\langle\mathbf{a},\mathcal{R}(E)\mathbf{b}\rangle.$$

Thus, consider any arbitrary pair $(\mathbf{a},\mathbf{b}) \in N\times N$ with $\|\mathbf{a}\|_2 = \|\mathbf{b}\|_2 = 1$, we can define a random variable $S(\mathbf{a},\mathbf{b})$ as:

$$S(\mathbf{a},\mathbf{b}) = \langle\mathbf{a},\mathcal{R}(E)\mathbf{b}\rangle = \langle\mathcal{R}(\mathbf{a}\mathbf{b}^T), E\rangle$$

by the fact that $\mathcal{R}$ is self-adjoint. Moreover, observe that given position of Ω is fixed, only random part of E is its sign and since the distribution is i.i.d. symmetric, we could apply Hoeffding's inequality to bound the probability that:

$$\Pr(|S(\mathbf{a},\mathbf{b})| > t) \le 2\exp\left(-\frac{2t^2}{\|\mathcal{R}(\mathbf{a}\mathbf{b}^T)\|_F^2}\right).$$

Note that by definition of operator 2-norm, $\|\mathcal{R}\| = \sup_{\hat{\mathbf{a}},\hat{\mathbf{b}}}\|\mathcal{R}(\hat{\mathbf{a}}\hat{\mathbf{b}}^T)\|_F/\|\hat{\mathbf{a}}\hat{\mathbf{b}}^T\|_F \ge \|\mathcal{R}(\mathbf{a}\mathbf{b}^T)\|_F$. Therefore, by an union bound:

$$\Pr(\sup_{\mathbf{a},\mathbf{b}\in N}|S(\mathbf{a},\mathbf{b})| > t) \le 2|N|^2\exp\left(-\frac{2t^2}{\|\mathcal{R}\|^2}\right),$$

which leads to:

$$\Pr(\|\mathcal{R}(E)\| > t) \le 2|N|^2\exp\left(-\frac{t^2}{8\|\mathcal{R}\|^2}\right).$$

Furthermore, on the event $\|\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega\| \le \sigma$, we can bound the operator norm by:

$$\|\mathcal{R}\| \leq \sum_{k\geq 1} \sigma^{2k} = \frac{\sigma^2}{1-\sigma^2}.$$

Putting all together, we can upper bound the second term of (17) by:

$$\Pr(\lambda\|\mathcal{R}(E)\| > t) \leq 2 \times 5^{2n} \exp\left(\frac{\gamma^2 t^2}{2\lambda^2}\right) + \Pr(\|\mathcal{P}_\Omega \mathcal{T}_L \mathcal{P}_T\| > \sigma)$$

where $\gamma = (1-\sigma^2)/2\sigma^2$. Thus, combining this bound with (18), and set $\lambda = 1/\sqrt{n}$, we can conclude $\|W^S\| \leq \frac{1}{4}$ with high probability if ρ (and thus σ) is sufficiently small. □

Proof of 3b. Let K be the matrix $\mathcal{P}_{\Omega^\perp}\mathcal{T}_L W^S + \mathcal{P}_{\Omega^\perp} M^S$, and our goal is to bound $\|K\|_\infty$. We first note that

$$\begin{aligned} K &= \mathcal{P}_{\Omega^\perp}\mathcal{T}_L W^S + \mathcal{P}_{\Omega^\perp} M^S \\ &= \lambda \mathcal{P}_{\Omega^\perp}(\mathcal{P}_Q - \mathcal{T}_L\mathcal{P}_T\mathcal{T}_S + \mathcal{P}_{Q^\perp})(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}E \\ &= -\lambda \mathcal{P}_{\Omega^\perp}\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}E. \end{aligned}$$

Consider any $K_{ij} \neq 0$. It must be in support of Ω^C and the element can be expressed as:

$$K_{ij} = \langle K, \mathbf{e}_i\mathbf{e}_j^T\rangle = \lambda\langle S(i,j), E\rangle,$$

where $S(i,j)$ is an $n \times n$ matrix defined by:

$$S(i,j) = (\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S(\mathbf{e}_i\mathbf{e}_j^T).$$

Now, conditional on Ω, the sign of E is i.i.d. symmetric and again, by Hoeffding's inequality, each K_{ij} could be bounded by:

$$\Pr(|K_{ij}| > t\lambda) \leq 2\exp\left(-\frac{2t^2}{\|S(i,j)\|_F^2}\right),$$

and thus, by an union bound, we have:

$$\Pr(\max_{(i,j)} |K_{ij}| > t\lambda) \leq 2n^2 \exp\left(-\frac{2t^2}{\max_{(i,j)} \|S(i,j)\|_F^2}\right).$$

Furthermore, since (10) holds, we have:

$$\begin{aligned} &\|S(i,j)\|_F \\ \leq &\|(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}\|\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\|\frac{\sqrt{2\mu_0\mu_1 rd}}{n}. \end{aligned}$$

In addition, on the event $\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| \leq \sigma$, we can also bound $\|(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\mathcal{T}_S\mathcal{P}_\Omega)^{-1}\| \leq 1/(1-\sigma^2)$, and therefore,

$$\Pr(\|K\|_\infty > t\lambda) \leq 2n^2 \exp\left(-\frac{n^2\gamma^2 t^2}{\mu_0\mu_1 rd}\right) + \Pr(\|\mathcal{P}_\Omega\mathcal{T}_L\mathcal{P}_T\| > \sigma),$$

where $\gamma = (1-\sigma^2)/\sigma$. The Lemma is thus proved provided that $r \leq \rho_r(\mu_0\mu_1)^{-1}n^2/(d\log n)$ with some small enough ρ_r.

□