5. Appendix

We divide this section into five parts. In the first part we prove some common lemmas. In the second part we give the convergence guarantee for **PG-RMC** . In the third part we give another algorithm which has a sample complexity of $O(\mu^4 r^3 n \log^2 n \log \frac{\mu^2 r \sigma_1^*}{\epsilon})$ and prove its convergence guarantees. In the fourth part we prove a generalized form of lemma 1. In the fifth part we present some additional experiments.

For the sake of convenience in the following proofs, we will define some notations here.

We define $p = \frac{|\Omega_{k,t}|}{mn}$ and we consider the following equivalent update step for $L^{(t+1)}$ in the analysis:

$$
\begin{aligned}
L^{(t+1)} &:= \mathcal{P}_k(M^{(t)}) & M^{(t)} &:= L^* + H \\
H &:= E^{(t)} + \beta G & E^{(t)} &:= \widetilde{S}^* - \widetilde{S}^{(t)} \\
\widetilde{S}^{(t)} &:= \mathcal{HT}_\zeta \left(M - L^{(t)}\right) & G &:= \frac{1}{\beta}\left(\mathcal{I} - \frac{\mathcal{P}_{\Omega_{q,t}}}{p}\right) D \\
D &:= L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^* & \beta &:= \frac{2\sqrt{n}\left\|D\right\|_\infty}{\sqrt{p}}
\end{aligned}
$$

The singular values of L^* are denoted by $\sigma_1^*, \ldots, \sigma_r^*$ where $|\sigma_1^*| \geq \ldots \geq |\sigma_r^*|$ and we will let $\lambda_1, \ldots, \lambda_n$ denote the singular values of $M^{(t)}$ where $|\lambda_1| \geq \ldots \geq |\lambda_n|$.

5.1. Common Lemmas

We will begin by restating some lemmas from previous work that we will use in our proofs.

First, we restate Weyl's perturbation lemma from (Bhatia, 1997), a key tool in our analysis:
Lemma 2. *Suppose $B = A + E \in \mathbb{R}^{m\times n}$ matrix. Let $\lambda_1, \cdots, \lambda_k$ and $\sigma_1, \cdots, \sigma_k$ be the singular values of B and A respectively such that $\lambda_1 \geq \cdots \geq \lambda_k$ and $\sigma_1 \geq \cdots \geq \sigma_k$. Then:*

$$|\lambda_i - \sigma_i| \leq \left\|E\right\|_2 \;\forall\, i \in [k].$$

This lemma establishes a bound on the spectral norm of a sparse matrix.
Lemma 3. *Let $S \in \mathbb{R}^{m\times n}$ be a sparse matrix with row and column sparsity ρ. Then,*

$$\left\|S\right\|_2 \leq \rho \max\{m, n\} \left\|S\right\|_\infty$$

Proof. For *any* pair of unit vectors u and v, we have:

$$
\begin{aligned}
v^\top S u &= \sum_{1\leq i\leq m, 1\leq j\leq n} v_i u_j S_{ij} \leq \sum_{1\leq i\leq m, 1\leq j\leq n} |S_{ij}| \left(\frac{v_i^2 + u_j^2}{2}\right) \\
&\leq \frac{1}{2}\left(\sum_{1\leq i\leq m} v_i^2 \sum_{1\leq j\leq n} |S_{ij}| + \sum_{1\leq j\leq n} u_j^2 \sum_{1\leq i\leq m} |S_{ij}|\right) \leq \rho \max\{m, n\} \left\|S\right\|_\infty
\end{aligned}
$$

Lemma now follows by using $\|S\|_2 = \max_{u,v,\|u\|_2=1,\|v\|_2=1} u^T S v$. □

Now, we define a 0-mean random matrix with small higher moments values.
Definition 1 (Definition 7, (Jain & Netrapalli, 2015))**.** *H is a random matrix of size $m \times n$ with each of its entries drawn independently satisfying the following moment conditions:*

$$\mathbb{E}[h_{ij}] = 0, \quad |h_{ij}| < 1, \quad \mathbb{E}[|h_{ij}|^k] \leq \tfrac{1}{\max\{m,n\}},$$

for $i, j \in [n]$ and $2 \leq k \leq 2\log n$.

We now restate two useful lemmas from (Jain & Netrapalli, 2015):
Lemma 4 (Lemma 8 and 10 of (Jain & Netrapalli, 2015))**.** *We have the following two claims:*

- *Suppose H satisfies Definition 1. Then, w.p. $\geq 1 - 1/n^{10+\log\alpha}$, we have: $\left\|H\right\|_2 \leq 3\sqrt{\alpha}$.*

- *Let A be a $m \times n$ matrix with $n \geq m$. Suppose $\Omega \subseteq [m] \times [n]$ is obtained by sampling each element with probability $p \geq \frac{1}{4n}$. Then, the following matrix H satisfies Defintion 1:*

$$H := \frac{\sqrt{p}}{2\sqrt{n}\left\|A\right\|_\infty}\left(A - \frac{1}{p}\mathcal{P}_\Omega(A)\right).$$

Lemma 5 (Lemma 13, (Jain & Netrapalli, 2015))**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\sigma_1, \cdots, \sigma_n$ where $|\sigma_1| \geq \cdots \geq |\sigma_n|$. Let $B = A + C$ be a perturbation of A satisfying $\left\|C\right\|_2 \leq \frac{\sigma_k}{2}$ and let $\mathcal{P}_k(B) = U\Lambda U^\top$ be the rank-k projection of B. Then, Λ^{-1} exists and we have:*

1. $\left\|A - AU\Lambda^{-1}U^\top A\right\|_2 \leq |\sigma_{k+1}| + 5\left\|C\right\|_2$,
2. $\left\|AU\Lambda^{-a}U^\top A\right\|_2 \leq 4\left(\frac{|\sigma_k|}{2}\right)^{-a+2} \quad \forall a \geq 2.$

We now provide a lemma that bounds $\|\cdot\|_\infty$ norm of an incoherent matrix with its operator norm.

Lemma 6. *Let $A \in \mathbb{R}^{m\times n}$ be a rank r, μ-incoherent matrix. Then for any $C \in \mathbb{R}^{n\times m}$, we have:*

$$\left\|ACA\right\|_\infty \leq \frac{\mu^2 r}{\sqrt{mn}}\left\|ACA\right\|_2$$

Proof. Let $A = U\Sigma V^\top$. Then, $ACA = UU^\top ACAVV^\top$. The lemma now follows by using definition of incoherence with the fact that $\|U^\top ACAV\|_2 \leq \|ACA\|_2$. □

We now present a lemma that shows improvement in the error $\|L - L^*\|_\infty$ by using gradient descent on $L^{(t)}$.

Lemma 7. *Let L^*, Ω, $\widetilde{S}^*$ satisfy Assumptions 1,2,3 respectively. Also, let the following hold for the t-th inner-iteration of any stage q:*

1. $\left\|L^* - L^{(t)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{k+1} + \left(\frac{1}{2}\right)^z \sigma^*_k\right)$
2. $\left\|\widetilde{S}^* - \widetilde{S}^{(t)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{k+1} + \left(\frac{1}{2}\right)^z \sigma^*_k\right)$
3. $Supp(\widetilde{S}^{(t)}) \subseteq Supp(\widetilde{S}^*)$

where $z \geq -3$ and σ^*_k and σ^*_{k+1} are the k and $(k+1)^{th}$ singular values of L^*. Also, let $E_1 = \widetilde{S}^{(t)} - \widetilde{S}^*$ and $E_3 = \left(\mathcal{I} - \frac{\mathcal{P}_{\Omega_{q,t}}}{p}\right)\left(L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*\right)$ be the error terms defined also in (6)*. Then, the following holds w.p $\geq 1 - n^{-(10+\log\alpha)}$:*

$$\left\|E_1 + E_3\right\|_2 \leq \frac{1}{100}\left(\sigma^*_{k+1} + \left(\frac{1}{2}\right)^z \sigma^*_k\right) \tag{8}$$

Proof. Note from Lemma 4,

$$\frac{1}{\beta}E_3 = \frac{1}{\beta}\left(\mathcal{I} - \frac{\mathcal{P}_{\Omega_{q,t}}}{p}\right)\left(L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*\right),$$

satisfies Definition 1 with $\beta = \frac{2\sqrt{n}}{\sqrt{p}} \cdot \|L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*\|_\infty$.

We now bound the spectral norm of $E_1 + E_3$ as follows:

$$\begin{aligned}\left\|E_1 + E_3\right\|_2 &\leq \left\|E_1\right\|_2 + \beta\cdot\left\|\frac{1}{\beta}E_3\right\|_2 \overset{(\zeta_1)}{\leq} \rho n\left\|\widetilde{S}^{(t)} - \widetilde{S}^*\right\|_\infty + 3\beta\sqrt{\alpha},\\ &\overset{(\zeta_2)}{\leq} \frac{1}{200}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^z \sigma^*_{k_q}\right) + \frac{60\mu^2 r}{m}\sqrt{\frac{n}{p}}\sqrt{\alpha}\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^z\left|\sigma^*_{k_q}\right|\right),\\ &\overset{(\zeta_3)}{\leq} \frac{1}{100}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^z \sigma^*_{k_q}\right).\end{aligned}$$

where (ζ_1) follows from Lemma 3 and 4, (ζ_2) follows by our assumptions on ρ, $\left\|L^{(t)} - L^*\right\|_\infty$, $\left\|\widetilde{S}^{(t)} - \widetilde{S}^*\right\|_\infty$ and our assumption that $n = O(m)$ and (ζ_3) follows from our assumption on p. □

In the following lemma, we prove that the value of the threshold computed using $\sigma_k(M^{(t)}) = \sigma_k(L^* + E_1 + E_3)$, where E_1, E_3 are defined in (6), closely tracks the threshold that we would have gotten had we had access to the true eigenvalues of L^*, σ_k^*.

Lemma 8. *Let L^*, Ω, $\widetilde{S}^*$ satisfy Assumptions 1,2,3 respectively. Also, let the following hold for the t-th inner-iteration of any stage q:*

1. $\left\|L^* - L^{(t)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)$
2. $\left\|\widetilde{S}^* - \widetilde{S}^{(t)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)$
3. $Supp(\widetilde{S}^{(t)}) \subseteq Supp(\widetilde{S}^*)$

where $z \geq -3$ and σ_k^* and σ_{k+1}^* are the k and $(k+1)^{th}$ singular values of L^*. Also, let $E_1 = \widetilde{S}^{(t)} - \widetilde{S}^*$ and $E_3 = \left(\mathcal{I} - \frac{\mathcal{P}_{\Omega_{q,t}}}{p}\right)\left(L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*\right)$ be the error terms defined also in (6). Then, the following holds $\forall z > -3$ w.p $\geq 1 - n^{-(10+\log\alpha)}$:

$$\frac{7}{8}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^{z+1}\sigma_k^*\right) \leq \left(\lambda_{k+1} + \left(\frac{1}{2}\right)^{z+1}\lambda_k\right) \leq \frac{9}{8}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^{z+1}\sigma_k^*\right), \tag{9}$$

where $\lambda_k := \sigma_k(M^{(t)}) = \sigma_k(L^* + E_1 + E_3)$ and E_1, E_3 are defined in (6).

Proof. Using Weyl's inequality (Lemma 2), we have: : $|\lambda_k - \sigma_k^*| \leq \|E_1 + E_3\|_2$ and $\left|\lambda_{k+1} - \sigma_{k+1}^*\right| \leq \|E_1 + E_3\|_2$ We now proceed to prove the lemma as follows:

$$\begin{aligned}
\left|\lambda_{k+1} + \left(\frac{1}{2}\right)^{z+1}\lambda_k - \sigma_{k+1}^* - \left(\frac{1}{2}\right)^{z+1}\sigma_k^*\right| &\leq \left|\lambda_{k+1} - \sigma_{k+1}^*\right| + \left(\frac{1}{2}\right)^{z+1}|\lambda_k - \sigma_k^*|, \\
&\leq \|E_1 + E_3\|_2\left(1 + \left(\frac{1}{2}\right)^{z+1}\right) \overset{(\zeta)}{\leq} \frac{1}{100}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z\sigma_k^*\right)\left(1 + \left(\frac{1}{2}\right)^{z+1}\right), \\
&\leq \frac{1}{8}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^{z+1}\sigma_k^*\right),
\end{aligned}$$

where (ζ) follows from Lemma 7 and the last inequality follows from the assumption that $z \geq -3$. □

Next, we show that the projected gradient descent update (6) leads to a better estimate of L^*, i.e., we bound $\|L^{(t+1)} - L^*\|_\infty$. Under the assumptions of the below given Lemma, the proof follows arguments similar to (Netrapalli et al., 2014) with additional challenge that arises due to more involved error terms E_1, E_3.

Our proof proceeds by first symmetrizing our matrices by rectangular dilation. We first begin by noting some properties of symmetrized matrices used in the proof of the following lemma.

Remark 1. *Let A be a $m \times n$ dimensional matrix with singular value decomposition $U\Sigma V^\top$. We denote its symmetrized version by $A_s := \begin{bmatrix} 0 & A^\top \\ A & 0 \end{bmatrix}$. Then:*

1. *The eigenvalue decomposition of A_s is given by $A_s = U_s \Sigma_s U_s^\top$ where*

$$U_s := \frac{1}{\sqrt{2}}\begin{bmatrix} V & V \\ U & -U \end{bmatrix} \quad \Sigma_s := \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}$$

2. $\mathcal{P}_{2k}(A_s) = \begin{bmatrix} 0 & \mathcal{P}_k(A^\top) \\ \mathcal{P}_k(A) & 0 \end{bmatrix}$
3. *We have* $A_s^{2j} = \begin{bmatrix} (A^\top A)^j & 0 \\ 0 & (AA^\top)^j \end{bmatrix} \quad A_s^{2j+1} = \begin{bmatrix} 0 & (A^\top A)^j A^\top \\ (AA^\top)^j A & 0 \end{bmatrix}$

4. We have

$$U_s\Sigma_s^{-j}U_s^\top = \begin{bmatrix} V\Sigma^{-j}V^\top & 0 \\ 0 & U\Sigma^{-j}U^\top \end{bmatrix} \text{ when } j \text{ is even}$$

$$U_s\Sigma_s^{-j}U_s^\top = \begin{bmatrix} 0 & V\Sigma^{-j}U^\top \\ U\Sigma^{-j}V^\top & 0 \end{bmatrix} \text{ when } j \text{ is odd}$$

Lemma 9. *Let $L = P_k(L^* + H)$, where H is any perturbation matrix that satisfies the following:*

1. $\|H\|_2 \le \frac{\sigma_k^*}{4}$

2. $\forall i \in [n],\ a \le \lceil \frac{\log n}{2} \rceil$ *with* $\upsilon \le \frac{\sigma_k^*}{4}$

$$\left\|e_i^\top \left(H^\top H\right)^a V^*\right\|_2, \left\|e_i^\top \left(HH^\top\right)^a U^*\right\|_2 \le (\upsilon)^{2a}\mu\sqrt{\frac{r}{m}}$$
$$\left\|e_i^\top H^\top \left(HH^\top\right)^a U^*\right\|_2, \left\|e_i^\top H \left(H^\top H\right)^a V^*\right\|_2 \le (\upsilon)^{2a+1}\mu\sqrt{\frac{r}{m}}$$

where σ_k^* is the k^{th} singular value of L^*. Also, let L^* satisfy Assumption 1. Then, the following holds:

$$\|L - L^*\|_\infty \le \frac{\mu^2 r}{m}\left(\sigma_{k+1}^* + 20\|H\|_2 + 8\upsilon\right)$$

where μ and r are the rank and incoherence of the matrix L^* respectively.

Proof. Let L_s, H_s and L_s^* denote the symmetrized forms of L, H and L^* respectively. Now, we have:

$$L_s = \mathcal{P}_{2k}\left(L_s^* + H_s\right)$$

Let $l = m + n$. Let $\lambda_1, \cdots, \lambda_l$ be the eigenvalues of $M_s = L_s^* + H_s$ with $|\lambda_1| \ge |\lambda_2| \cdots \ge |\lambda_l|$. Let $u_1, u_2, \cdots, u_l$ be the corresponding eigenvectors of M_s. Using Lemma 2 along with the assumption on $\|H_s\|_2$, we have: $|\lambda_{2k}| \ge \frac{3\sigma_k^*}{4}$.

Let $U\Lambda V$ be the eigen vector decomposition of L. Let $U_s\Lambda_sU_s^\top$ to be the eigen vector decomposition of L_s. Then, using Remark 1 we have $\forall\ i \in [2k]$:

$$\left(L_s^* + H_s\right)u_i = \lambda_i u_i, \text{ i.e. } \left(I - \frac{H_s}{\lambda_i}\right)u_i = \frac{L_s^* u_i}{\lambda_i}.$$

As $|\lambda_{2k}| \ge \frac{3\sigma_k^*}{4}$ and $\|H_s\|_2 \le \frac{1}{4}\sigma_k^*$, we can apply the Taylor's series expansion to get the following expression for u_i:

$$u_i = \frac{1}{\lambda_i}\left(\sum_{j=0}^{\infty}\left(\frac{H_s}{\lambda_i}\right)^j\right)\frac{L_s^* u_i}{\lambda_i}.$$

That is,

$$L_s = \sum_{i=1}^{2k}\lambda_i u_i u_i^\top = \sum_{i=1}^{2k}\lambda_i^{-1}\sum_{0\le s,t<\infty}\left(\frac{H_s}{\lambda_i}\right)^s L_s^* u_i u_i^\top L_s^*\left(\frac{H_s}{\lambda_i}\right)^t,$$
$$= \sum_{0\le s,t<\infty}\sum_{i=1}^{2k}\lambda_i^{-(s+t+1)}H_s^s L_s^* u_i u_i^\top L_s^* H_s^t = \sum_{0\le s,t<\infty}H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)}U_s^\top L_s^* H_s^t.$$

Subtracting L_s^* on both sides and taking operator norm, we get:

$$\|L_s - L_s^*\|_\infty = \left\|U_s\Lambda_sU_s^\top - L_s^*\right\|_\infty = \left\|\sum_{0\le s,t<\infty}H_s^s L_s^* U_s\Lambda_s^{-(s+t+1)}U_s^\top L_s^* H_s^t - L_s^*\right\|_\infty,$$
$$= \left\|L_s^* U_s \Lambda_s^{-1}U_s^\top L_s^* - L_s^*\right\|_\infty + \sum_{1\le s+t<\infty}\left\|H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)}U_s^\top L_s^* H_s^t\right\|_\infty. \quad (10)$$

We separately bound the first and the second term of RHS. The first term can be bounded as follows:

$$\left\|L_s^* U_s \Lambda_s^{-1} U_s^\top L_s^* - L_s^*\right\|_\infty \overset{(\zeta_1)}{=} \left\| L_s^* \begin{bmatrix} 0 & V\Lambda^{-1}U^\top \\ U\Lambda^{-1}V^\top & 0 \end{bmatrix} L_s^* - L_s^* \right\|_\infty \tag{11}$$

$$\leq \left\|L^* V \Lambda^{-1} U^\top L^* - L^*\right\|_\infty \overset{(\zeta_2)}{\leq} \frac{\mu^2 r}{\sqrt{mn}} \left\|L^* V \Lambda^{-1} U^\top L^* - L^*\right\|_2 \overset{(\zeta_3)}{\leq} \frac{\mu^2 r}{\sqrt{mn}} \left(\left|\sigma_{k+1}^*\right| + 5 \left\|H\right\|_2 \right), \tag{12}$$

where (ζ_1) follows Remark 1, (ζ_2) from Lemma 6 and (ζ_3) follows from Claim 1 of Lemma 5 after symmetrization.

We now bound second term of RHS of (10) which we again split in two parts. We first bound the terms with $s+t > \log n$:

$$\begin{aligned} \left\| H_s^s L_s^* U_s \Lambda_s^{(s+t+1)} U_s^\top L_s^* H_s^t \right\|_\infty &\leq \left\| H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^* H_s^t \right\|_2 \overset{(\zeta_1)}{\leq} \left\|H_s\right\|_2^{s+t} 4 \left(\frac{2}{\sigma_k^*}\right)^{-(s+t-1)} \\ &\leq 4 \left\|H\right\|_2 \left(\left\|H\right\|_2 \frac{2}{\sigma_k^*} \right)^{(s+t-1)} \overset{(\zeta_2)}{\leq} 4 \left\|H\right\|_2 \left(\frac{1}{2}\right)^{(s+t-1)} \\ &\leq \frac{4}{n} \left\|H\right\|_2 \left(\frac{1}{2}\right)^{(s+t-1-\log n)} \leq 4 \frac{\mu^2 r}{m} \left\|H\right\|_2 \left(\frac{1}{2}\right)^{(s+t-1-\log n)}, \end{aligned} \tag{13}$$

where (ζ_1) follows from the second claim of Lemma 5 and noting that $\left\|H_s\right\|_2 = \left\|H\right\|_2$ and (ζ_2) follows from assumption on $\left\|H\right\|_2$ and using the fact that $s+t \geq \log n$.

Summing up over all terms with $s+t > \log n$, we get from (13) and (12):

$$\left\|L_s - L_s^*\right\|_\infty \leq \frac{\mu^2 r}{m} \left(\left|\sigma_{k+1}^*\right| + 20 \left\|H\right\|_2 \right) + \sum_{0 < s+t \leq \log n} \left\| H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^* H_s^t \right\|_\infty \tag{14}$$

where the first inequality follows because $m \leq n$.

Now, for terms corresponding to $1 \leq s+t \leq \log n$, we have:

$$\begin{aligned} \left\| H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^* H_s^t \right\|_\infty &= \max_{q_1 \in [m+n], q_2 \in [m+n]} \left| e_{q_1}^\top H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^* H_s^t e_{q_2} \right| \\ &\leq \left(\max_{q_1 \in [m+n]} \left\| e_{q_1}^\top H_s^s U_s^* \right\|_2 \right) \left\| \Sigma_s^* (U_s^*)^\top U_s \Lambda_s^{-(s+t+1)} U_s^\top U_s^* \Sigma_s^* \right\|_2 \left(\max_{q_2 \in [m+n]} \left\| e_{q_2}^\top H^t U_s^* \right\|_2 \right), \end{aligned} \tag{15}$$

We will now bound the terms, $\max_{q_1 \in [m+n]} \left\| e_{q_1}^\top H_s^s U_s^* \right\|_2$. Note from Remark 1.1 that $U_s^* = \frac{1}{\sqrt{2}} \begin{bmatrix} V^* & V^* \\ U^* & -U^* \end{bmatrix}$. Now, we have the following cases for H_s^s:

$$H_s^j = \begin{bmatrix} \left(H^\top H\right)^{\frac{s}{2}} & 0 \\ 0 & \left(HH^\top\right)^{\frac{s}{2}} \end{bmatrix} \text{ when } s \text{ is even} \quad H_s^j = \begin{bmatrix} 0 & H^\top \left(HH^\top\right)^{\lfloor \frac{s}{2} \rfloor} \\ H \left(H^\top H\right)^{\lfloor \frac{s}{2} \rfloor} & 0 \end{bmatrix} \text{ when } s \text{ is odd}$$

In these two cases, we have:

$$H_s^s U_s^* = \tfrac{1}{\sqrt{2}} \begin{bmatrix} \left(H^\top H\right)^{\frac{s}{2}} V^* & \left(H^\top H\right)^{\frac{s}{2}} V^* \\ \left(HH^\top\right)^{\frac{s}{2}} U^* & -\left(HH^\top\right)^{\frac{s}{2}} U^* \end{bmatrix} \quad H_s^s U_s^* = \tfrac{1}{\sqrt{2}} \begin{bmatrix} H^\top \left(HH^\top\right)^{\lfloor \frac{s}{2} \rfloor} U^* & -H^\top \left(HH^\top\right)^{\lfloor \frac{s}{2} \rfloor} U^* \\ H \left(H^\top H\right)^{\lfloor \frac{s}{2} \rfloor} V^* & H \left(H^\top H\right)^{\lfloor \frac{s}{2} \rfloor} V^* \end{bmatrix}$$

This leads to the following 4 cases for $\max_{q_1 \in [m+n]} \left\| e_{q_1}^\top H_s^s U_s^* \right\|_2$:

$$\begin{array}{lll} \text{for } s \text{ even} & \max_{q' \in [n]} \left\| e_{q'}^\top \left(H^\top H\right)^{\frac{s}{2}} V^* \right\|_2 & \max_{q' \in [m]} \left\| e_{q'}^\top \left(HH^\top\right)^{\frac{s}{2}} U^* \right\|_2 \\ \text{for } s \text{ odd} & \max_{q' \in [n]} \left\| e_{q'}^\top H^\top \left(HH^\top\right)^{\lfloor \frac{s}{2} \rfloor} U^* \right\|_2 & \max_{q' \in [m]} \left\| e_{q'}^\top H \left(H^\top H\right)^{\lfloor \frac{s}{2} \rfloor} V^* \right\|_2 \end{array}$$

We can now bound the terms in (15) as follows:

$$\begin{aligned}\left\|H_s^s L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^* H_s^t\right\|_\infty &\overset{(\zeta_1)}{\leq} \frac{\mu^2 r}{m} \upsilon^{s+t} \left\|L_s^* U_s \Lambda_s^{-(s+t+1)} U_s^\top L_s^*\right\|_2 \\ &\overset{(\zeta_2)}{\leq} \frac{4\mu^2 r}{m} \upsilon^{s+t} \left(\frac{2}{\sigma_k^*}\right)^{s+t-1} \leq \frac{4\mu^2 r}{m} \upsilon \left(\frac{1}{2}\right)^{s+t-1} \qquad (16)\end{aligned}$$

where (ζ_1) follows from the second assumption of the Lemma and the preceding argument and (ζ_2) follows from Claim 2 of Lemma 5 and the final step follows from our bound on υ.

Finally, note from the Remark 1 that $\|L_s^* - L_s\|_\infty = \left\|L^* - L^{(t+1)}\right\|_\infty$. Now, summing up (16) over all $1 \leq s+t \leq \log n$ and combining with (14), the lemma is proved. □

In the next lemma, we show that with the threshold chosen in the algorithm, we show an improvement in the estimation of $\widetilde{S}^*$ by $\widetilde{S}^{(t)}$.

Lemma 10. *In the t^{th} iterate of the q^{th} stage, assume the following holds:*

1. $\left\|L^* - L^{(t)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)$
2. $\frac{7}{8}\eta\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right) \leq \zeta^{(t)} \leq \frac{9}{8}\eta\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)$

where σ_k^* and σ_{k+1}^* are the k and $(k+1)^{th}$ singular values of L^*, λ_k and λ_{k+1} are the k and $(k+1)^{th}$ singular values of $M^{(t)}$ and, r and μ are the rank and incoherence of the $m \times n$ matrix L^* respectively. Then we have

1. *Supp* $\left(\widetilde{S}^{(t)}\right) \subseteq$ *Supp* $\left(\widetilde{S}^*\right)$
2. $\left\|\widetilde{S}^{(t)} - \widetilde{S}^*\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)$

Proof. We first prove the first claim of the lemma. Consider an index pair $(i,j) \notin Supp(\widetilde{S}^*)$.

$$\left|M_{ij} - L_{ij}^{(t)}\right| \leq \frac{2\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right) \overset{(\zeta_1)}{\leq} \frac{16\mu^2 r}{7m\eta}\zeta^{(t)} \overset{(\zeta_2)}{\leq} \zeta^{(t)}$$

where (ζ_1) follows from the second assumption of the lemma and (ζ_2) follows from our setting of $\eta = \frac{4\mu^2 r}{m}$. Hence, we do not threshold any entry that is not corrupted by $\widetilde{S}^*$.

Now, we prove the second claim of the lemma. Consider an index entry $(i,j) \in Supp(\widetilde{S}^*)$. Here, we consider two cases:

1. The entry $(i,j) \in Supp(\widetilde{S}^{(t)})$: Here the entry (i,j) is thresholded. We know that $L_{ij}^{(t)} + \widetilde{S}_{ij}^{(t)} = L_{ij}^* + \widetilde{S}_{ij}^*$ from which we get
$$\left|\widetilde{S}_{ij}^{(t)} - \widetilde{S}_{ij}^*\right| = \left|L_{ij}^* - L_{ij}^{(t)}\right| \leq \left\|L^* - L^{(t)}\right\|_\infty$$

2. The entry $(i,j) \notin Supp(\widetilde{S}^{(t)})$: Here the entry (i,j) is not thresholded. We know that $\left|L_{ij}^* + \widetilde{S}_{ij}^* - L_{ij}^{(t)}\right| \leq \zeta^{(t)}$ from which we get
$$\begin{aligned}\left|\widetilde{S}_{ij}^*\right| &\leq \zeta^{(t)} + \left|L_{ij}^* - L_{ij}^{(t)}\right| \\ &\overset{(\zeta_2)}{\leq} \frac{36\mu^2 r}{8m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right) + \frac{2\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right) \\ &\leq \frac{8\mu^2 r}{m}\left(\sigma_{k+1}^* + \left(\frac{1}{2}\right)^z \sigma_k^*\right)\end{aligned}$$
where (ζ_2) follows from the second assumption along with our setting of $\eta = 4\frac{\mu^2 r}{m}$.

The above two cases prove the second statement of the lemma. □

We will now prove Lemma 1

Proof of Lemma 1: Recall the definitions of $E_1 = \left(\widetilde{S}^* - \widetilde{S}^{(t)}\right)$, $E_2 = \left(L^{(t)} - L^*\right)$, $E_3 = \left(\mathcal{I} - \frac{\mathcal{P}_{\Omega_{q,t}}}{p}\right)(E_2 - E_1)$ and $\beta = 2\sqrt{\frac{n}{p}}\left\|E_2 - E_1\right\|_\infty$. Recall that $H := E_1 + E_3$ From Lemma 4, we have that $\frac{1}{\beta}E_3$ satisfies Definition 1. This implies that the matrix $\frac{1}{\beta}(E_1 + E_3)$ satisfies the conditions of Lemma 14. Now, we have $\forall 1 \leq a \leq \lceil \log n \rceil$ and $\forall i \in [n]$:

$$\left\|e_i(HH^\top)^a U^*\right\|_2 = \beta^{2a}\left\|e_i\left(\left(\frac{1}{\beta}H\right)\left(\frac{1}{\beta}H\right)^\top\right)^a U^*\right\|_2$$
$$\overset{(\zeta)}{\leq} \beta^{2a}\left(\frac{\rho n}{\beta}\left\|E_1\right\|_\infty + c\log n\right)^{2a}\mu\sqrt{\frac{r}{m}} \leq \mu\sqrt{\frac{r}{m}}\left(\rho n\left\|E_1\right\|_\infty + 2c\sqrt{\frac{n}{p}}\left(\left\|E_1 - E_2\right\|_\infty\right)\log n\right)^{2a}$$

where (ζ) follows from the application of Lemma 14 along with the incoherence assumption on U^*. The other statements of the lemma can be proved in a similar manner by invocations of the different claims of Lemma 14. □

5.2. Algorithm PG-RMC

Proof of Theorem 1: We know that $T \geq \log(\frac{20\mu^2 nr\sigma_1^*}{\epsilon})$. Consider the stage q reached at the termination of the algorithm. We know from Lemma 11 that:

1. $\left\|E^{(T)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{T-3}\sigma^*_{k_q}\right) \leq \frac{8\mu^2 r}{m}\sigma^*_{k_q+1} + \frac{\epsilon}{10n}$
2. $\left\|L^{(T)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{T-3}\left|\sigma^*_{k_q}\right|\right) \leq \frac{2\mu^2 r}{m}\sigma^*_{k_q+1} + \frac{\epsilon}{10n}$

Combining this with Lemmas 2 and 7, we get:

$$\left|\sigma_{k_q+1}(M^{(T)})\right| \geq \sigma^*_{k_q+1} - \frac{1}{100}\left(\sigma^*_{k_q+1} + \frac{m\epsilon}{10n\mu^2 r}\right) \tag{17}$$

When the while loop terminates, $\eta\sigma_{k_q+1}\left(M^{(T)}\right) < \frac{\epsilon}{2n}$, which from (17), implies that $\sigma^*_{k_q+1} < \frac{m\epsilon}{7n\mu^2 r}$. So we have:

$$\left\|L - L^*\right\|_\infty = \left\|L^{(T)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}\sigma^*_{k_q+1} + \frac{\epsilon}{10n} \leq \frac{\epsilon}{2n}.$$

We will now bound the number of iterations required for the **PG-RMC** to converge.

From claim 2 of Lemma 12, we have $\sigma^*_{k_q+1} \leq \frac{17}{32}\sigma^*_{k_{q-1}+1} \quad \forall q \geq 1$. By recursively applying this inequality, we get $\sigma^*_{k_q+1} \leq \left(\frac{17}{32}\right)^q\sigma^*_1$. We know that when the algorithm terminates, $\sigma^*_{k_q+1} < \frac{\epsilon}{7\mu^2 r}$. Since, $\left(\frac{17}{32}\right)^q\sigma^*_1$ is an upper bound for $\sigma^*_{k_q+1}$, an upper bound for the number of iterations is $5\log\left(\frac{7\mu^2 r\sigma^*_1}{\epsilon}\right)$. Also, note that an upper bound to this quantity is used to partition the samples provided to the algorithm. This happens with probability $\geq 1 - T^2 n^{-(10+\log\alpha)} \geq 1 - n^{-\log\alpha}$. This concludes the proof. □

In the following lemma, we show that we make progress simultaneously in the estimation of both $\widetilde{S}^*$ and L^* by $\widetilde{S}^{(t)}$ and $L^{(t)}$. We make use of Lemmas 9 and 10 to show progress in the estimation of one affects the other alternatively. We also emphasize the roles of the following quantities in enabling us to prove our convergence result:

1. $\left\|H\right\|_2$ - We use Lemma 7 to bound this quantity
2. The analysis of the following 4 quantities is crucial to obtaining error bounds in $\|\|_\infty$ norm

$$\text{for } j \text{ even} \quad \max_{q'\in[n]}\left\|e_{q'}^\top\left(H^\top H\right)^{\frac{j}{2}}V^*\right\|_2 \quad \max_{q'\in[m]}\left\|e_{q'}^\top\left(HH^\top\right)^{\frac{j}{2}}U^*\right\|_2$$
$$\text{for } j \text{ odd} \quad \max_{q'\in[n]}\left\|e_{q'}^\top H^\top\left(HH^\top\right)^{\lfloor\frac{j}{2}\rfloor}U^*\right\|_2 \quad \max_{q'\in[m]}\left\|e_{q'}^\top H\left(H^\top H\right)^{\lfloor\frac{j}{2}\rfloor}V^*\right\|_2$$

We use Lemma 1 to bound this quantity.

Lemma 11. *Let L^*, Ω, $\widetilde{S}^*$ and $\widetilde{S}^{(t)}$ satisfy Assumptions 1,2,3 respectively. Then, in the t^{th} iteration of the q^{th} stage of Algorithm 1, $\widetilde{S}^{(t)}$ and $L^{(t)}$ satisfy:*

$$\left\|\widetilde{S}^{(t)} - \widetilde{S}^*\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-3}\left|\sigma^*_{k_q}\right|\right),$$

$$\textit{Supp}\left(\widetilde{S}^{(t)}\right) \subseteq \textit{Supp}\left(\widetilde{S}^*\right),$$

$$\frac{7}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-2}\left|\sigma^*_{k_q}\right|\right) \leq \zeta^{(t+1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-2}\left|\sigma^*_{k_q}\right|\right) \textit{ and}$$

$$\left\|L^{(t)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-3}\left|\sigma^*_{k_q}\right|\right).$$

with probability $\geq 1 - ((q-1)T + t - 1)n^{-(10+\log\alpha)}$ where T is the number of iterations in the inner loop.

Proof. We prove the lemma by induction on both q and t. Recall that $E^{(t)} = \widetilde{S}^* - \widetilde{S}^{(t)}$

Base Case: $q = 1$ **and** $t = 0$
We begin by first proving an upper bound on $\|L^*\|_\infty$. We do this as follows:

$$\left|L^*_{ij}\right| = \left|\sum_{k=1}^r \sigma^*_k u^*_{ik} v^*_{jk}\right| \leq \sum_{k=1}^r \sigma^*_k \left|u^*_{ik} v^*_{jk}\right| \leq \sigma^*_1 \sum_{k=1}^r \left|u^*_{ik} v^*_{jk}\right| \leq \frac{\mu^2 r}{\sqrt{mn}}\sigma^*_1$$

where the last inequality follows from Cauchy-Schwartz and the incoherence of U^*. This directly proves the third claim of the lemma for the base case. Recall, that $\zeta^{(0)} = \eta\sigma^*_1$. We now have from the thresholding step and the incoherence assumption on L^*:

1. $\left\|E^{(0)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_2 + 2\sigma^*_1\right) \overset{(\zeta)}{\leq} \frac{8\mu^2 r}{m}\left(8\sigma^*_{k_1}\right)$, and
2. Supp $\left(\widetilde{S}^{(t)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$.

where (ζ) follows from Lemma 12.

Finally, from Lemma 8, we have:

$$\frac{7}{8}\eta\left(\sigma^*_{k_1+1} + 4\sigma^*_{k_1}\right) \leq \zeta^{(1)} = \eta\left(\sigma_{k_1+1}(M^{(t)}) + 4\sigma_{k_1}(M^{(t)})\right) \leq \frac{9}{8}\eta\left(\sigma^*_{k_1+1} + 4\sigma^*_{k_1}\right)$$

So the base case of induction is satisfied.

Induction over t
We first prove the inductive step over t (for a fixed q). By inductive hypothesis we assume that:

a) $\left\|E^{(t)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-3}\sigma^*_{k_q}\right)$

b) Supp $\left(\widetilde{S}^{(t)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$.

c) $\left\|L^* - L^{(t)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-3}\sigma^*_{k_q}\right)$

d) $\frac{7}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-2}\left|\sigma^*_{k_q}\right|\right) \leq \zeta^{(t+1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-2}\left|\sigma^*_{k_q}\right|\right)$

with probability $1 - ((q-1)T + t - 1)n^{-(10+\log\alpha)}$. Then by Lemma 9, we have:

$$\left\|L^{(t+1)} - L^*\right\|_\infty \leq \frac{\mu^2 r}{m}\left(\sigma^*_{k_q+1} + 20\left\|H\right\|_2 + 8\upsilon\right) \tag{18}$$

From Lemma 1, we have:

$$v \le \rho n \left\| E^{(t)} \right\|_\infty + 8\beta\alpha \log n \overset{(\zeta_1)}{\le} \frac{1}{100}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-3} \sigma^*_{k_q}\right) + 8\beta\alpha\log n \overset{(\zeta_2)}{\le} \frac{1}{50}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-3}\sigma^*_{k_q}\right) \quad (19)$$

where (ζ_1) follows from our assumptions on ρ and our inductive hypothesis on $\left\|E^{(t)}\right\|_\infty$ and (ζ_2) follows from our assumption on p and by noticing that $\|D\|_\infty \le \left\|E^{(t)}\right\|_\infty + \left\|L^* - L^{(t)}\right\|_\infty$. Recall that $D = L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*$.

From Lemma 7:

$$\|H\|_2 \le \frac{1}{100}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-3}\sigma^*_{k_q}\right) \quad (20)$$

with probability $\ge 1 - n^{-(10+\log\alpha)}$. From Equations (20), (19) and (18), we have:

$$\left\|L^* - L^{(t+1)}\right\|_\infty \le \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-2}\sigma^*_{k_q}\right)$$

which by union bound holds with probability $\ge 1 - ((q-1)T + t)n^{-(10+\log\alpha)}$. Hence, using Lemma 10 and our inductive hypothesis on $\zeta^{(t+1)}$ we have:

1. $\left\|E^{(t+1)}\right\|_\infty \le \frac{8\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{t-2}\sigma^*_{k_q}\right)$
2. Supp $\left(\widetilde{S}^{(t+1)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$.

which also holds with probability $\ge 1 - ((q-1)T + t)n^{-(10+\log\alpha)}$. This concludes the proof for induction over t.

Finally, from Lemma 8 and our bounds on $\left\|E^{(t+1)}\right\|_\infty$ and $\left\|L^* - L^{(t+1)}\right\|_\infty$, we have:

$$\frac{7}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-1}\left|\sigma^*_{k_q}\right|\right) \le \zeta^{(t+2)} \le \frac{9}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{t-1}\left|\sigma^*_{k_q}\right|\right)$$

Induction Over Stages q

We now prove the induction over q. Suppose the hypothesis holds for stage q. At the end of stage q, we have:

1. $\left\|E^{(T)}\right\|_\infty \le \frac{8\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{T-3}\sigma^*_{k_q}\right) \le \frac{8\mu^2 r\sigma^*_{k_q+1}}{m} + \frac{\epsilon}{10n}$, and
2. Supp $\left(\widetilde{S}^{(T)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$.
3. $\frac{7}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{T-2}\left|\sigma^*_{k_q}\right|\right) \le \zeta^{(T+1)} \le \frac{9}{8}\eta\left(\left|\sigma^*_{k_q+1}\right| + \left(\frac{1}{2}\right)^{T-2}\left|\sigma^*_{k_q}\right|\right)$

with probability $\ge 1 - (qT-1)n^{-(10+\log\alpha)}$. From Lemmas 2 and 7, we get:

$$\left|\sigma_{k_q+1}\left(M^{(T)}\right) - \sigma^*_{k_q+1}\right| \le \|H\|_2 \le \frac{1}{100}\left(\sigma^*_{k_q+1} + \frac{m\epsilon}{10n\mu^2 r}\right) \quad (21)$$

with probability $1 - n^{-(10+\log\alpha)}$. We know that $\eta\sigma_{k_q+1}\left(M^{(t)}\right) \ge \frac{\epsilon}{2n}$ which with (21) implies that $\left|\sigma^*_{k_q+1}\right| > \frac{m\epsilon}{10n\mu^2 r}$.

$$\begin{aligned}\left\|L^{(T+1)} - L^*\right\|_\infty &\le \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \left(\frac{1}{2}\right)^{T-2}\sigma^*_{k_q}\right) \le \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \frac{m\epsilon}{20n\mu^2 rn}\right)\\ &\le \frac{2\mu^2 r}{m}\left(\sigma^*_{k_q+1} + \frac{\sigma^*_{k_q+1}}{2}\right) \le \frac{2\mu^2 r}{m}\left(2\sigma^*_{k_q+1}\right) \overset{(\zeta_4)}{\le} \frac{2\mu^2 r}{m}\left(8\sigma^*_{k_{q+1}}\right)\end{aligned}$$

where (ζ_4) follows from Lemma 12. By union bound this holds with probability $\geq 1 - qTn^{-(10+\log\alpha)}$.

Now, from Lemma 10 and the inductive hypothesis on ζ^{T+1}, we have through a similar series of arguments as above:

1. $\left\|E^{(0)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(8\sigma^*_{k_{q+1}}\right)$
2. Supp $\left(\widetilde{S}^{(0)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$

which holds with probability $\geq 1 - qTn^{-(10+\log\alpha)}$.

Recall, now that $L^{(0)} = L^{(T+1)}$. Finally, from Lemma 8 and our bounds on $\left\|E^{(0)}\right\|_\infty$ and $\left\|L^{(0)} - L^*\right\|_\infty$, we have:

$$\frac{7}{8}\eta\left(\left|\sigma^*_{k_{q+1}+1}\right| + 4\left|\sigma^*_{k_{q+1}}\right|\right) \leq \zeta^{(1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{k_{q+1}+1}\right| + 4\left|\sigma^*_{k_{q+1}}\right|\right)$$

□

Lemma 12. *Suppose at the beginning of the q^{th} stage of algorithm 1:*

1. $\left\|L^* - L^{(0)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(2\sigma^*_{k_{q-1}+1}\right)$

2. $\left\|E^{(0)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(2\sigma^*_{k_{q-1}+1}\right)$

Then, the following hold:

1. $\sigma^*_{k_q} \geq \frac{15}{32}\sigma^*_{k_{q-1}+1}$

2. $\sigma^*_{k_q+1} \leq \frac{17}{32}\sigma^*_{k_{q-1}+1}$

with probability $\geq 1 - n^{-(10+\log\alpha)}$

Proof. We know that:

$$\lambda_{k_q} \leq \sigma^*_{k_q} + \|H\|_2, \qquad \lambda_{k_{q-1}+1} \geq \sigma^*_{k_{q-1}+1} - \|H\|_2, \qquad \lambda_{k_q} \geq \frac{\lambda_{k_{q-1}+1}}{2}$$

Combining the three inequalities, we get:

$$\sigma^*_{k_q} \geq \frac{\sigma^*_{k_{q-1}+1} - 3\|H\|_2}{2}$$

Applying Lemma 7, we get the first claim of the lemma.

Similar to the first claim, we have:

$$\lambda_{k_q+1} \geq \sigma^*_{k_q+1} - \|H\|_2, \qquad \lambda_{k_{q-1}+1} \leq \sigma^*_{k_{q-1}+1} + \|H\|_2, \qquad \lambda_{k_q+1} \leq \frac{\lambda_{k_{q-1}+1}}{2}$$

Again, combining the three inequalities, we get:

$$\sigma^*_{k_q+1} \leq \frac{\sigma^*_{k_{q-1}+1} + 3\|H\|_2}{2}$$

Another application of Lemma 7 gives the second claim. □

Algorithm 3 $\widehat{L} = \textbf{R-RMC}(\Omega, \mathcal{P}_\Omega(M), \epsilon, r, \eta, \sigma)$: Non-convex Robust Matrix Completion

Input: Observed entries Ω, Matrix $\mathcal{P}_\Omega(M) \in \mathbb{R}^{m\times n}$, convergence criterion ϵ, target rank r, thresholding parameter η, upper bound on σ_1^* σ
$T \leftarrow 10 \log \frac{20\mu^2 nr\sigma}{\epsilon}$ */*Number of inner iterations*/*
Partition Ω into $rT+1$ subsets $\{\Omega_0\} \cup \{\Omega_{q,t} : q \in [r], t \in [T]\}$ using 2
$L^{(0)} = 0$, $\zeta^{(0)} \leftarrow \eta\sigma$
$M^{(0)} \leftarrow \frac{mn}{|\Omega_0|}\mathcal{P}_{\Omega_0}(M - \mathcal{HT}_\zeta(M))$
$q \leftarrow 0$
while $\sigma_{q+1}(M^{(0)}) > \frac{\epsilon}{2\eta m}$ **do**
  $q \leftarrow q+1$
  for Iteration $t = 0$ to $t = T$ **do**
    $S^{(t)} = H_\zeta(\mathcal{P}_{\Omega_{q,t}}(M - L^{(t)}))$ */*Projection onto set of sparse matrices*/*
    $M^{(t)} = L^{(t)} - \frac{mn}{|\Omega_{q,t}|}\mathcal{P}_{\Omega_{q,t}}(L^{(t)} + S^{(t)} - M)$ */*Gradient Descent Update*/*
    $L^{(t+1)} = P_q(M^{(t)})$ */*Projected Gradient Descent step*/*
    Set threshold $\zeta^{(t+1)} \leftarrow \eta\left(\sigma_{q+1}(M^{(t)}) + \left(\frac{1}{2}\right)^t \sigma_q(M^{(t)})\right)$
  end for
  $S^{(0)} = S^{(T)}, L^{(0)} = L^{(T+1)}, M^{(0)} = M^{(T)}, \zeta^{(0)} = \zeta^{(T+1)}$
end while
Return: $L^{(T+1)}$

5.3. Algorithm R-RMC

Proof of Theorem 2: We know that $T \geq \log(\frac{20\mu^2 nr\sigma_1^*}{\epsilon})$.

Consider the stage q reached at the termination of the algorithm. We know from Lemma 13 that:

1. $\left\|E^{(T)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma_{q+1}^* + \left(\frac{1}{2}\right)^{T-1}\sigma_q^*\right) \leq \frac{8\mu^2 r}{m}\sigma_{q+1}^* + \frac{\epsilon}{10n}$

2. $\left\|L^{(T)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma_{q+1}^* + \left(\frac{1}{2}\right)^{T-1}\sigma_q^*\right) \leq \frac{2\mu^2 r}{m}\sigma_{q+1}^* + \frac{\epsilon}{10n}$

Combining this with Lemmas 2 and 7, we get:

$$\sigma_{q+1}(M^{(T)}) \geq \sigma_{q+1}^* - \frac{1}{100}\left(\sigma_{q+1}^* + \frac{m\epsilon}{10n\mu^2 r}\right) \tag{22}$$

When the while loop terminates, $\eta\sigma_{q+1}\left(M^{(T)}\right) < \frac{\epsilon}{2n}$, which from (22), implies that $\sigma_{q+1}^* < \frac{m\epsilon}{7n\mu^2 r}$. So we have:

$$\|L - L^*\|_\infty = \left\|L^{(T)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}|\sigma_{k_q+1}^*| + \frac{\epsilon}{10n} \leq \frac{\epsilon}{2n}.$$

□

As in the case of the proof of Theorem 1, the following lemma shows that we simultaneously make progress in both the estimation of L^* and $\widetilde{S}^*$ by $L^{(t)}$ and $\widetilde{S}^{(t)}$ respectively. Similar to Lemma 11, we make use of Lemmas 10 and 9 to show how improvement in estimation of one of the quantities affects the other and the other five terms, $\|H\|_2$, $\max_{q'\in[n]}\left\|e_{q'}^\top\left(H^\top H\right)^j V^*\right\|_2$, $\max_{q'\in[m]}\left\|e_{q'}^\top\left(HH^\top\right)^j U^*\right\|_2$, $\max_{q'\in[n]}\left\|e_{q'}^\top H^\top\left(HH^\top\right)^j U^*\right\|_2$ and $\max_{q'\in[m]}\left\|e_{q'}^\top H\left(H^\top H\right)^j V^*\right\|_2$ are analyzed the same way:

Lemma 13. *Let L^*, Ω, $\widetilde{S}^*$ and $\widetilde{S}^{(t)}$ satisfy Assumptions 1,2,3 respectively. Then, in the t^{th} iteration of the q^{th} stage of*

Algorithm 3, $\widetilde{S}^{(t)}$ and $L^{(t)}$ satisfy:

$$\left\|\widetilde{S}^{(t)} - \widetilde{S}^*\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t-1}\sigma^*_q\right),$$
$$\textit{Supp}\left(\widetilde{S}^{(t)}\right) \subseteq \textit{Supp}\left(\widetilde{S}^*\right),$$
$$\frac{7}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^t\left|\sigma^*_q\right|\right) \leq \zeta^{(t+1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^t\left|\sigma^*_q\right|\right) \textit{ and}$$
$$\left\|L^{(t)} - L^*\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t-1}\sigma^*_q\right).$$

with probability $\geq 1 - ((q-1)T + t - 1)n^{-(10+\log\alpha)}$ where T is the number of iterations in the inner loop.

Proof. We prove the lemma by induction on both q and t.

Base Case: $q = 1$ **and** $t = 0$
We begin by first proving an upper bound on $\|L^*\|_\infty$. We do this as follows:

$$\left|L^*_{ij}\right| = \left|\sum_{k=1}^{r}\sigma^*_k u^*_{ik} v^*_{jk}\right| \leq \sum_{k=1}^{r}\left|\sigma^*_k u^*_{ik} v^*_{jk}\right| \leq \sigma^*_1\sum_{k=1}^{r}\left|u^*_{ik} v^*_{jk}\right| \leq \frac{\mu^2 r}{m}\sigma^*_1$$

where the last inequality follows from Cauchy-Schwartz and the incoherence of U^*. This directly proves the third claim of the lemma for the base case. Recall that $\zeta^{(0)} = \eta\sigma^*_1$. We also note that due to the thresholding step and the incoherence assumption on L^*, we have:

1. $\left\|E^{(0)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_2 + 2\sigma^*_1\right)$
2. $\text{Supp}\left(\widetilde{S}^{(t)}\right) \subseteq \text{Supp}\left(\widetilde{S}^*\right)$.

From Lemma 8 and our bounds on $E^{(0)}$ and $\left\|L^{(1)} - L^*\right\|_\infty$, we have:

$$\frac{7}{8}\eta\left(|\sigma^*_2| + |\sigma^*_1|\right) \leq \zeta^{(1)} \leq \frac{9}{8}\eta\left(|\sigma^*_2| + |\sigma^*_1|\right)$$

So the base case of induction is satisfied.

Induction over t
We first prove the inductive step over t (for a fixed q). By inductive hypothesis we assume that:

a) $\left\|E^{(t)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(|\sigma^*_{q+1}| + \left(\frac{1}{2}\right)^{t-1}|\sigma^*_q|\right)$

b) $\text{Supp}\left(\widetilde{S}^{(t)}\right) \subseteq \text{Supp}\left(\widetilde{S}^*\right)$.

c) $\left\|L^* - L^{(t)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(|\sigma^*_{q+1}| + \left(\frac{1}{2}\right)^{t-1}|\sigma^*_q|\right)$

d) $\frac{7}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^t\left|\sigma^*_q\right|\right) \leq \zeta^{(t+1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^t\left|\sigma^*_q\right|\right)$

with probability $1 - ((q-1)T + t - 1)n^{-(10+\log\alpha)}$.

Then by Lemma 9, we have:

$$\left\|L^{(t+1)} - L^*\right\|_\infty \leq \frac{\mu^2 r}{m}\left(|\sigma^*_{k_q+1}| + 20\|H\|_2 + 8\upsilon\right) \tag{23}$$

From Lemma 1, we have:

$$\upsilon \leq \rho n \left\|E^{(t)}\right\|_\infty + 8\beta\alpha\log n \overset{(\zeta_1)}{\leq} \frac{1}{100}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t-1}\sigma^*_q\right) + 8\beta\alpha\log n \overset{(\zeta_2)}{\leq} \frac{1}{50}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t-1}\sigma^*_q\right) \tag{24}$$

where (ζ_1) follows from our assumptions on ρ and our inductive hypothesis on $\left\|E^{(t)}\right\|_\infty$ and (ζ_2) follows from our assumption on p and by noticing that $\|D\|_\infty \leq \left\|E^{(t)}\right\|_\infty + \left\|L^* - L^{(t)}\right\|_\infty$. Recall that $D = L^{(t)} - L^* + \widetilde{S}^{(t)} - \widetilde{S}^*$.

From Lemma 7:

$$\|H\|_2 \leq \frac{1}{100}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t-1}\sigma^*_q\right) \tag{25}$$

with probability $\geq 1 - n^{-(10+\log\alpha)}$. From Equations (25), (24) and (23), we have:

$$\left\|L^* - L^{(t+1)}\right\|_\infty \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t}\sigma^*_q\right)$$

which by union bound holds with probability $\geq 1 - ((q-1)T + t)n^{-(10+\log\alpha)}$. Hence, using Lemma 10 and the inductive hypothesis on $\zeta^{(t+1)}$ we have:

1. $\left\|E^{(t+1)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{t}\sigma^*_q\right)$
2. Supp $\left(\widetilde{S}^{(t+1)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$.

which also holds with probability $\geq 1 - ((q-1)T + t)n^{-(10+\log\alpha)}$. This concludes the proof for induction over t.

Finally, using Lemma 8 and our bounds on $\left\|E^{(t+1)}\right\|_\infty$ and $\left\|L^{(t+1)} - L^*\right\|_\infty$, we have:

$$\frac{7}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^{t+1}\left|\sigma^*_q\right|\right) \leq \zeta^{(t+2)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^{t+1}\left|\sigma^*_q\right|\right)$$

Induction Over Stages q

We now prove the induction over q. Suppose the hypothesis holds for stage q. At the end of stage q, we have:

1. $\left\|E^{(T)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{T-1}\sigma^*_q\right) \leq \frac{8\mu^2 r\sigma^*_{q+1}}{m} + \frac{\epsilon}{10n}$
2. Supp $\left(\widetilde{S}^{(T)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$
3. $\frac{7}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^{T}\left|\sigma^*_q\right|\right) \leq \zeta^{(T+1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{q+1}\right| + \left(\frac{1}{2}\right)^{T}\left|\sigma^*_q\right|\right)$

with probability $\geq 1 - (qT - 1)n^{-(10+\log\alpha)}$.

From Lemmas 2 and 7 we get:

$$\left|\sigma_{q+1}\left(M^{(T)}\right) - \sigma^*_{q+1}\right| \leq \|H\|_2 \leq \frac{1}{100}\left(\sigma^*_{q+1} + \frac{m\epsilon}{10n\mu^2 r}\right) \tag{26}$$

with probability $1 - n^{-(10+\log\alpha)}$. We know that $\eta\sigma_{q+1}\left(M^{(t)}\right) \geq \frac{\epsilon}{2n}$ which with (26) implies that $\sigma^*_{q+1} > \frac{m\epsilon}{10n\mu^2 r}$.

$$\begin{aligned}\left\|L^{(T+1)} - L^*\right\|_\infty &\leq \frac{2\mu^2 r}{m}\left(\sigma^*_{q+1} + \left(\frac{1}{2}\right)^{T}\sigma^*_q\right) \leq \frac{2\mu^2 r}{m}\left(\sigma^*_{q+1} + \frac{m\epsilon}{20\mu^2 rn}\right)\\ &\leq \frac{2\mu^2 r}{m}\left(\sigma^*_{q+1} + \frac{\sigma^*_{q+1}}{2}\right) \leq \frac{2\mu^2 r}{m}\left(2\sigma^*_{q+1}\right)\end{aligned}$$

By union bound this holds with probability $\geq 1 - qTn^{-(10+\log\alpha)}$.

Now, from Lemma 10 and our inductive hypothesis on $\zeta^{(T+1)}$, we have through a similar series of arguments as above:

1. $\left\|E^{(0)}\right\|_\infty \leq \frac{8\mu^2 r}{m}\left(2\sigma^*_{q+1}\right)$
2. Supp $\left(\widetilde{S}^{(0)}\right) \subseteq$ Supp $\left(\widetilde{S}^*\right)$

which holds with probability $\geq 1 - qTn^{-(10+\log\alpha)}$.

Recall, now that $L^{(0)} = L^{(T+1)}$. Finally, from Lemma 8 and our bounds on $\left\|E^{(0)}\right\|_\infty$ and $\left\|L^{(0)} - L^*\right\|_\infty$, we have:

$$\frac{7}{8}\eta\left(\left|\sigma^*_{q+2}\right| + \left|\sigma^*_{q+1}\right|\right) \leq \ \zeta^{(1)} \leq \frac{9}{8}\eta\left(\left|\sigma^*_{q+2}\right| + \left|\sigma^*_{q+1}\right|\right)$$

□

5.4. Proof of a generalized form of Lemma 1

Lemma 14. *Suppose $H = H_1 + H_2$ and $H \in \mathbb{R}^{m\times n}$ where H_1 satisfies Definition 1 (Definition 7 from (Jain & Netrapalli, 2015)) and H_2 is a matrix with column and row sparsity ρ. Let U be a matrix with rows denoted as $u_1, \dots, u_m$ and let V be a matrix with rows denoted as $v_1, \dots, v_n$. Let e_q be the q^{th} vector from standard basis. Let $\tau = \max\{\max_{i\in[m]}\|u_i\|, \max_{i\in[n]}\|v_i\|\}$. Then, for $0 \leq a \leq \log n$:*

$$\max_{q\in[n]}\left\|e_q^\top\left(H^\top H\right)^a V\right\|_2, \max_{q\in[m]}\left\|e_q^\top\left(HH^\top\right)^a U\right\|_2 \leq \left(\rho n\left\|H_2\right\|_\infty + c\log n\right)^{2a}\tau$$
$$\max_{q\in[n]}\left\|e_q^\top H^\top\left(HH^\top\right)^a U\right\|_2, \max_{q\in[m]}\left\|e_q^\top H\left(H^\top H\right)^a V\right\|_2 \leq \left(\rho n\left\|H_2\right\|_\infty + c\log n\right)^{2a+1}\tau$$

with probability $n^{-2\log\frac{c}{4}+4}$.

Proof. Similar to (Jain & Netrapalli, 2015), we will prove the statement for $q = 1$ and it can be proved for $q \in [n]$ by taking a union bound over all q. For the sake of brevity, we will prove only the inequality:

$$\max_{q\in[n]}\left\|e_q^\top\left(H^\top H\right)^a V\right\|_2 \leq \left(\rho n\left\|H_2\right\|_\infty + c\log n\right)^{2a}\tau$$

The rest of the lemma follows by applying similar arguments to the appropriate quantities.

Let $\omega : [2a] \to \{1, 2\}$ be a function used to index a single term in the expansion of $(H^\top H)^a$. We express the term as follows:

$$(H^\top H)^a = \sum_\omega \prod_{i=1}^a H^\top_{\omega(2i-1)} H_{\omega(2i)}$$

We will now fix one such term ω and then bound the length of the following random vector:

$$v_\omega = e_1^\top \prod_{i=1}^a (H^\top_{\omega(2i-1)} H_{\omega(2i)}) V$$

Let α be used to denote a tuple (i, j) of integers used to index entries in a matrix. Let $T(i)$ be used to denote the parity function computed on i, i.e, 0 if i is divisible by 2 and 1 otherwise. This function indicates if the matrix in the expansion is transposed or not. We now introduce $B^{p,q}_{(i,j),(k,l)}$, $p \in \{1, 2\}$, $q \in \{0, 1\}$ and $A^p_{(i,j)}$, $p \in \{1, 2\}$ which are defined as follows:

$$A^{p}_{(i,j)} := \delta_{i,1}(\delta_{p,1} + \delta_{p,2}\mathbb{1}_{\{(i,j)\in Supp(H_2)\}})$$

$$B^{p,q}_{(i,j),(k,l)} := (\delta_{q,1}\delta_{j,l} + \delta_{q,0}\delta_{i,k})(\delta_{p,1} + \delta_{p,2}\mathbb{1}_{\{(k,l)\in Supp(H_2)\}})$$

where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise. We will subsequently write the random vector v_ω in terms of the individual entries of the matrices. The role of $B^{p,q}_{(i,j),(k,l)}$ and $A^{p}_{(i,j)}$ is to ensure consistency in the terms used to describe v_ω. We will use $h_{i,\alpha}$ to refer to $(H_i)_\alpha$.

With this notation in hand, we are ready to describe v_ω.

$$v_\omega = \sum_{\substack{\alpha_1,\ldots,\alpha_{2a}\\ \alpha_1(1)=1}} A^{\omega(1)}_{\alpha_1} B^{\omega(2),T(2)}_{\alpha_1\alpha_2} \ldots B^{\omega(2a),T(2a)}_{\alpha_{2a-1}\alpha_{2a}} h_{\omega(1),\alpha_1} \cdots h_{\omega(2a),\alpha_{2a}} v_{\alpha_{2a}(2)}$$

We now write the squared length of v_ω as follows:

$$X_\omega = \sum_{\substack{\alpha_1,\ldots,\alpha_{2a},\alpha'_1,\ldots,\alpha'_{2a}\\ \alpha_1(1)=1,\alpha'_1(1)=1}} A^{\omega(1)}_{\alpha_1} B^{\omega(2),T(2)}_{\alpha_1\alpha_2} \ldots B^{\omega(2a),T(2a)}_{\alpha_{2a-1}\alpha_{2a}} h_{\omega(1),\alpha_1} \cdots h_{\omega(2a),\alpha_{2a}} \\ A^{\omega(1)}_{\alpha_1} B^{\omega(2),T(2)}_{\alpha'_1\alpha'_2} \ldots B^{\omega(2a),T(2a)}_{\alpha'_{2a-1}\alpha'_{2a}} h_{\omega(1),\alpha'_1} \cdots h_{\omega(2a),\alpha'_{2a}} \langle v_{\alpha_{2a}(2)}, v_{\alpha'_{2a}(2)}\rangle$$

We can see from the above equations that the entries used to represent v_ω are defined with respect to paths in a bipartite graph. In the following, we introduce notations to represent entire paths rather than just individual edges:

Let $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{2a})$ and

$$\boldsymbol{\zeta_\alpha} := A^{\omega(1)}_{\alpha_1} B^{\omega(2),T(2)}_{\alpha_1\alpha_2} \ldots B^{\omega(2a),T(2a)}_{\alpha_{2a-1}\alpha_{2a}} h_{\omega(1),\alpha_1} \ldots h_{\omega(2a),\alpha_{2a}}$$

Now, we can write:

$$X_\omega = \sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\alpha'}\\ \alpha_1(1)=\alpha'_1(1)=1}} \zeta_{\boldsymbol{\alpha}}\zeta_{\boldsymbol{\alpha'}} \langle v_{\alpha_{2a}(2)}, v_{\alpha'_{2a}(2)}\rangle$$

Calculating the k^{th} moment expansion of X_ω for some number k, we obtain:

$$\mathbb{E}[X^k_\omega] = \sum_{\boldsymbol{\alpha^1},\ldots,\boldsymbol{\alpha^{2k}}} \mathbb{E}[\zeta_{\boldsymbol{\alpha^1}} \ldots \zeta_{\boldsymbol{\alpha^{2k}}} \langle v_{\alpha^1_{2a}(2)}, v_{\alpha^2_{2a}(2)}\rangle \ldots \langle v_{\alpha^{2k-1}_{2a}(2)}, v_{\alpha^{2k}_{2a}(2)}\rangle] \tag{27}$$

We now show how to bound the above moment effectively. Notice that the moment is defined with respect to a collection of $2k$ paths. We denote this collection by $\Delta := (\boldsymbol{\alpha^1}, \ldots, \boldsymbol{\alpha^{2k}})$. For each such collection, we define a partition $\Gamma(\Delta)$ of the index set $\{(s,l) : s \in [2k], l \in [2a]\}$ where (s,l) and (s',l') are in the same equivalence class if $\omega(l) = \omega(l') = 1$ and $\alpha^s_l = \alpha^{s'}_{l'}$. Additionally, each (s,l) such that $\omega(l) = 2$ is in a separate equivalence class.

We bound the expression in (27) by partitioning all possible collections of $2k$ paths based on the partitions defined by them in the above manner. We then proceed to bound the contribution of any one specific path to (27) following a particular partition Γ, the number of paths satisfying that particular partition and finally, the total number of partitions. Consider a partition Γ with non-zero contribution to the k^{th} moment. Since, H_1 is a matrix with 0 mean, any equivalence class of Γ containing an index (s,l) such that $\omega(l) = 1$ contains at least two elements (Otherwise, for any Δ satisfying Γ has 0 contribution to the k^{th} as the element in the singleton equivalence class has mean 0).

We proceed to bound (27) by taking absolute values:

$$\mathbb{E}[X^k_\omega] \le \sum_{\boldsymbol{\alpha^1},\ldots,\boldsymbol{\alpha^{2k}}} \mathbb{E}[|\zeta_{\boldsymbol{\alpha^1}}| \ldots |\zeta_{\boldsymbol{\alpha^{2k}}}| |\langle v_{\alpha^1_{2a}(2)}, v_{\alpha^2_{2a}(2)}\rangle| \ldots |\langle v_{\alpha^{2k-1}_{2a}(2)}, v_{\alpha^{2k}_{2a}(2)}\rangle|] \tag{28}$$

We now fix one particular partition and bound the contribution to (28) of all collections of paths Δ that correspond to a valid partition Γ.

We construct from Γ a directed multigraph G. The equivalence classes of Γ form the vertex set of G, $V(G)$. There are 4 kinds of edges in G where each type is indexed by a tuple (p,q) where $p \in \{1,2\}$, $q \in \{0,1\}$. We denote the edge sets corresponding to these 4 edge types by $E_{(1,0)}$, $E_{(1,1)}$, $E_{(2,0)}$ and $E_{(2,1)}$ respectively. An edge of type (p,q) exists from equivalence class γ_1 to equivalence class γ_2 if there exists $(s,l) \in \gamma_1$ and $(s',l') \in \gamma_2$ such that $l' = l+1$, $s = s'$, $\omega(s') = p$ and $T(l') = q$.

The summation in 28 can be written as follows:

$$\begin{aligned}
&\mathbb{E}[|\zeta_{\boldsymbol{\alpha}^1}| \dots |\zeta_{\boldsymbol{\alpha}^{2k}}| \left|\langle v_{\alpha^1_{2a}(2)}, v_{\alpha^2_{2a}(2)}\rangle\right| \dots \left|\langle v_{\alpha^{2k-1}_{2a}(2)}, v_{\alpha^{2k}_{2a}(2)}\rangle\right|] \\
&\leq \tau^{2k} \left(\prod_{s=1}^{2k} A^{\omega(1)}_{\alpha^s_1} \prod_{l=1}^{2a-1} B^{\omega(l+1),T(l+1)}_{\alpha^s_l,\alpha^s_{l+1}}\right) \mathbb{E}\left[\left(\prod_{s=1}^{2k}\prod_{l=1}^{2a} \left|h_{\omega(l),\alpha^s_l}\right|\right)\right] \\
&\overset{(\zeta_1)}{\leq} \tau^{2k} \left(\prod_{s=1}^{2k} A^{\omega(1)}_{\alpha^s_1} \prod_{l=1}^{2a-1} B^{\omega(l+1),T(l+1)}_{\alpha^s_l,\alpha^s_{l+1}}\right) \prod_{\gamma\in V_1(G)} \frac{1}{n} \prod_{\gamma \in V_2(G)} \|H_2\|_\infty \\
&= \frac{\tau^{2k}\|H_2\|_\infty^{w_2}}{n^{w_1}} \left(\prod_{s=1}^{2k} A^{\omega(1)}_{\alpha^s_1} \prod_{l=1}^{2a-1} B^{\omega(l+1),T(l+1)}_{\alpha^s_l,\alpha^s_{l+1}}\right)
\end{aligned}$$

where (ζ_1) follows from the moment conditions on H_1. $V_1(G)$ and $V_2(G)$ are the vertices in the graph corresponding to tuples (i,j) such that $\omega(j)=1$ and $\omega(j)=2$ respectively and $w_1 = |V_1(G)|$, $w_2 = |V_2(G)|$.

We first consider an equivalence class γ_1 such that there exists an index $(s,l) \in \gamma_1$ and $l=1$. We form a spanning tree T_1 of all the nodes reachable from γ_1 with γ_1 as root. We then remove the nodes $V(T_1)$ from the graph G and repeat this procedure until we obtain a set of l trees $T_1, \dots, T_l$ with roots $\gamma_1, \dots, \gamma_l$ such that $\bigcup_{i=1}^{l} V(G_i) = V(G)$. This happens because every node is reachable from some equivalence class which contains an index of the form $(s,1)$. Also, each of these trees T_i, $\forall\, i \in [l]$ is disjoint in their vertex sets. Given this decomposition, we can factorize the above product as follows:

$$\begin{aligned}
\mathbb{E}[X^k_\omega|\Gamma] \leq \frac{\tau^{2k}\|H_2\|_\infty^{w_2}}{n^{w_1}} \prod_{j=1}^{l} \sum_{\alpha_\gamma,\gamma\in T_j} A^{\omega(1)}_{\alpha_{\gamma_j}} \prod_{\{\gamma,\gamma'\}\in E_{(1,0)}(T_j)} B^{1,0}_{\alpha_\gamma\alpha_{\gamma'}} \\
\prod_{\{\gamma,\gamma'\}\in E_{(1,1)}(T_j)} B^{1,1}_{\alpha_\gamma\alpha_{\gamma'}} \prod_{\{\gamma,\gamma'\}\in E_{(2,0)}(T_j)} B^{2,0}_{\alpha_\gamma\alpha_{\gamma'}} \prod_{\{\gamma,\gamma'\}\in E_{(2,1)}(T_j)} B^{2,1}_{\alpha_\gamma\alpha_{\gamma'}} \qquad (29)
\end{aligned}$$

where the inner sum is over all possible assignments to the elements in the equivalence classes of tree T_j.

For a single connected component, we can compute the summation bottom up from the leaves. First, notice that as each $B^{i,j}_{\alpha_\gamma,\alpha_{\gamma'}}$ is bounded by 1:

$$\begin{aligned}
&\sum_{\alpha_{\gamma'}} B^{2,1}_{\alpha_\gamma\alpha_{\gamma'}} \leq \rho n \quad &&\sum_{\alpha_{\gamma'}} B^{2,0}_{\alpha_\gamma\alpha_{\gamma'}} \leq \rho n \\
&\sum_{\alpha_{\gamma'}} B^{1,1}_{\alpha_\gamma\alpha_{\gamma'}} = n \quad &&\sum_{\alpha_{\gamma'}} B^{1,0}_{\alpha_\gamma\alpha_{\gamma'}} = n
\end{aligned}$$

Where the first two follow from the sparsity of H_2. Every node in the tree T_j with the exception of the root has a single incoming edge. For the root, γ_j, we have:

$$\sum_{\alpha_1} A_{\alpha_1}^{\omega(1)} \leq \rho n \text{ for } \omega(1) = 2 \quad \sum_{\alpha_1} A_{\alpha_1}^{\omega(1)} = n \text{ for } \omega(1) = 1$$

From the above two observations, we have:

$$\sum_{\alpha_1,\ldots,\alpha_{v_j}} A_{\alpha_1}^{\omega(1)} \prod_{\{\gamma,\gamma'\}\in E_{(1,0)}(T_j)} B^{1,0}_{\alpha_\gamma\alpha_{\gamma'}} \prod_{\{\gamma,\gamma'\}\in E_{(1,1)}(T_j)} B^{1,1}_{\alpha_\gamma\alpha_{\gamma'}} \prod_{\{\gamma,\gamma'\}\in E_{(2,0)}(T_j)} B^{2,0}_{\alpha_\gamma\alpha_{\gamma'}} \prod_{\{\gamma,\gamma'\}\in E_{(2,1)}(T_j)} B^{2,1}_{\alpha_\gamma\alpha_{\gamma'}} \leq (\rho n)^{w_{2,j}} n^{w_{1,j}}$$

where $w_{k,j}$ represents the number of vertices in the j^{th} component which contain tuples (y,z) such that $\omega(z) = k$ for $k \in \{1,2\}$.

Plugging the above in (29) gives us

$$\mathbb{E}[X_\omega^k(\Gamma)] \leq \frac{\tau^{2k} \left\|H_2\right\|_\infty^{w_2}}{n^{w_1}} (\rho n)^{\sum_j w_{2,j}} n^{\sum_j w_{1,j}} = \tau^{2k} \left\|H_2\right\|_\infty^{w_2} (\rho n)^{w_2}$$

Let a_1 and a_2 be defined as $|\{i : \omega(i) = 1\}|$ and $|\{i : \omega(i) = 2\}|$ respectively (Note that $w_2 = 2a_2 k$). Summing up over all possible partitions (there are at most $(2a_1 k)^{2a_1 k}$ of them), we get our final bound on $\mathbb{E}\left[\hat{X}_\omega^k\right]$ as $\tau^{2k}(\rho n \left\|H_2\right\|_\infty)^{2a_2 k}(2a_1 k)^{2a_1 k}$.

Now, we bound the probability that $\hat{X}_\omega$ is too large. Choosing $k = \left\lceil \frac{\log n}{a_1} \right\rceil$ and applying the k^{th} moment Markov inequality, we obtain:

$$\begin{aligned}
\mathbf{Pr}\left[\left|\hat{X}_\omega\right| > (c\log n)^{2a_1}\tau^2(\rho n \left\|H_2\right\|_\infty)^{2a_2}\right] &\leq \mathbb{E}\left[\left|\hat{X}_\omega\right|^k\right] \left(\frac{1}{(c\log n)^{2a_1}\tau^2(\rho n\left\|H_2\right\|_\infty)^{2a_2}}\right)^k \\
&\leq \left(\frac{2ka_1}{c\log n}\right)^{2ka_1} \\
&\leq n^{-2\log\frac{c}{4}}
\end{aligned}$$

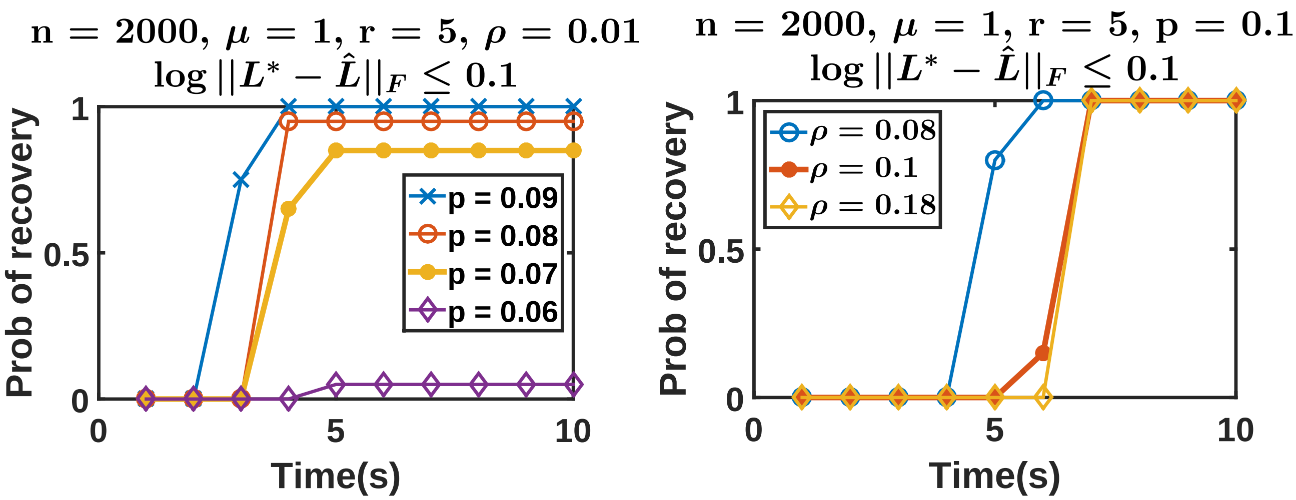
Taking a union bound over all the 2^{2a} possible ω, over values of a from 1 to $\log n$ and over the n values of q, and summing up the high probability bound over all possible values of ω, we get the required result. $\square$

5.5. Additional Experimental Results

We detail some additional experiments performed with Algorithm 1 in this section. The experiments were performed on synthetic data and real world data sets.

Synthetic data. We generate a random matrix $M \in \mathbb{R}^{2000\times 2000}$ in the same way as described in Section 4. In these experiments our aim is to analyze the behavior of the algorithm in extremal cases. We consider two of such cases : 1) sampling probability is very low (Figure 3 (a)), 2) number of corruptions is very large (Figure 3 (b)). In the first case, we see that the we get a reasonably good probability of recovery (~ 0.8) even with very low sampling probability (0.07). In the second case, we observe that the time taken to recover seems almost independent of the number of corruptions as long as they are below a certain threshold. In our experiments we saw that on increasing the ρ to 0.2 the probability of recovery went to 0. To compute the probability of recovery we ran the experiment 20 times and counted the number of successful runs.

Foreground-background separation. We present results for one more real world data set in this section. We applied our **PG-RMC** method (with varying p) to the Escalator video. Figure 4 (a) shows one frame from the video. Figure 4 (b) shows the extracted background from the video by using our method (**PG-RMC** , Algorithm 1) with probability of sampling $p = 0.05$. Figure 4 (c) compares objective function value for different p values.



(a) (b)

Figure 3: We run the **PG-RMC** algorithm with extremal values of sampling probability and fraction of corruptions, and record the probability with which we recover the original matrix, (a) : time vs probability of recovery for very small values of sampling probability, (b) : time vs probability of recovery for large number of corruptions (ρn^2)

(a) (b) (c)

Figure 4: **PG-RMC** on Escalator video. (a): a video frame (b): an extracted background frame (c): time vs error for different sampling probabilities; **PG-RMC** takes 7.3s while St-NcRPCA takes 52.9s