Supplementary Material On Local Linear Convergence Rate of Iterative Hard Thresholding for Matrix Completion

Trung Vu, *Graduate Student Member, IEEE*, Evgenia Chunikhina, and Raviv Raich, *Senior Member, IEEE*

I. Comparison to prior results

In our main theorem, the rate of convergence depends on

$$\boldsymbol{H} = \boldsymbol{S}_{\bar{\Omega}}^\top (\boldsymbol{P}_{\boldsymbol{V}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}) \boldsymbol{S}_{\bar{\Omega}} \in \mathbb{R}^{(n_1 n_2 - s) \times (n_1 n_2 - s)},$$

where $\boldsymbol{S}_{\bar{\Omega}} \in \mathbb{R}^{n_1 n_2 \times (n_1 n_2 - s)}$ is the selection matrix corresponding to the complement set $\bar{\Omega}$. $P_{\boldsymbol{U}_\perp}$ and $P_{\boldsymbol{V}_\perp}$ are the projections onto the left and right null spaces of $\boldsymbol{M}$. Viewing $\boldsymbol{H}$ as a function of $\boldsymbol{M}$ and Ω, let us consider the set

$$\mathcal{S} = \{(\boldsymbol{X}, \Omega) \mid \boldsymbol{H}(\boldsymbol{X}, \Omega) \text{ is full rank}\}.$$

In the following, we show that our proposed set $\mathcal{S}$ contains the set of incoherent matrices and uniform sampling patterns. In other words, if $\boldsymbol{M}$ is incoherent and Ω is a uniform sampling, then $(\boldsymbol{M}, \Omega) \in \mathcal{S}$ *w.h.p.* First, we highlight the fact that the invertibility of $\boldsymbol{H}$ is related to the injectivity of the sampling operator restricted to $T_{\boldsymbol{M}}(\mathcal{M}_{\leq r})$ - the tangent space T to $\mathcal{M}_{\leq r} = \{\boldsymbol{X} \in R^{n_1 \times n_2} \mid \operatorname{rank}(\boldsymbol{X}) \leq r\}$ at $\boldsymbol{M}$. In particular, recall that this operator is of the form $\mathcal{A}_{\Omega T} = \mathcal{P}_\Omega \mathcal{P}_T$, where $\mathcal{P}_\Omega : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ is the orthogonal projector onto the indices in Ω and $\mathcal{P}_T : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ is the orthogonal projection onto T (see [1]-Eqn. 3.5)

$$\mathcal{P}_T(\boldsymbol{X}) = \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{X} + \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{V}} - \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{V}} = \boldsymbol{X} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{V}_\perp},$$

for all $\boldsymbol{X} \in \mathbb{R}^{m \times n}$. Using vectorization, one can show that $\operatorname{vec}(\mathcal{P}_\Omega(\boldsymbol{X})) = \boldsymbol{S}_\Omega \boldsymbol{S}_\Omega^\top \operatorname{vec}(\boldsymbol{X}) = (\boldsymbol{I}_{n_1 n_2} - \boldsymbol{S}_{\bar{\Omega}} \boldsymbol{S}_{\bar{\Omega}}^\top) \operatorname{vec}(\boldsymbol{X})$ and $\operatorname{vec}(\mathcal{P}_T(\boldsymbol{X})) = (\boldsymbol{I}_{n_1 n_2} - \boldsymbol{P}_{\boldsymbol{V}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}) \operatorname{vec}(\boldsymbol{X}) = \boldsymbol{Q}_\perp \boldsymbol{Q}_\perp^\top \operatorname{vec}(\boldsymbol{X})$, where $\boldsymbol{Q}_\perp \in \mathbb{R}^{n_1 n_2 \times r(n_1 + n_2 - r)}$ is the basis of $T_{\boldsymbol{M}}(\mathcal{M}_{\leq r})$, i.e., $\boldsymbol{Q}_\perp \boldsymbol{Q}_\perp^\top = \boldsymbol{I}_{n_1 n_2} - \boldsymbol{P}_{\boldsymbol{V}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}$. Therefore, the eigenvalues of the operator $\mathcal{A}_{\Omega T}^* \mathcal{A}_{\Omega T} = \boldsymbol{P}_T \boldsymbol{P}_\Omega \boldsymbol{P}_T : \mathbb{R}^{n_1 n_2} \to \mathbb{R}^{n_1 n_2}$ restricted to $T_{\boldsymbol{M}}(\mathcal{M}_{\leq r})$ are the same as those of the $r(n_1 + n_2 - r) \times r(n_1 + n_2 - r)$ matrix

$$\begin{aligned} \hat{\boldsymbol{H}} &= \boldsymbol{Q}_\perp^\top (\boldsymbol{I}_{n_1 n_2} - \boldsymbol{S}_{\bar{\Omega}} \boldsymbol{S}_{\bar{\Omega}}^\top) \boldsymbol{Q}_\perp \\ &= \boldsymbol{I}_{r(n_1 + n_2 - r)} - \boldsymbol{Q}_\perp^\top \boldsymbol{S}_{\bar{\Omega}} \boldsymbol{S}_{\bar{\Omega}}^\top \boldsymbol{Q}_\perp. \end{aligned}$$

Now representing $\boldsymbol{H} = \boldsymbol{S}_{\bar{\Omega}}^\top (\boldsymbol{I}_{r(n_1 + n_2 - r)} - \boldsymbol{Q}_\perp \boldsymbol{Q}_\perp^\top) \boldsymbol{S}_{\bar{\Omega}} = \boldsymbol{I}_{n_1 n_2 - s} - \boldsymbol{S}_{\bar{\Omega}}^\top \boldsymbol{Q}_\perp \boldsymbol{Q}_\perp^\top \boldsymbol{S}_{\bar{\Omega}}$, it can be showed that $\boldsymbol{H}$ and $\hat{\boldsymbol{H}}$ share the same set of eigenvalues except those at 1. Equivalently, the injectivity of $\mathcal{A}_{\Omega T}$ restricted to $T_{\boldsymbol{M}}(\mathcal{M}_{\leq r})$ implies the invertibility of $\boldsymbol{H}$. Second, we recall the so-called result from Candes and Recht that the operator $\mathcal{A}_{\Omega T}$ is *most likely* injective when restricted to $T_{\boldsymbol{M}}(\mathcal{M}_{\leq r})$. Specifically, Eqn. (4.11) in [1] states that if Ω is sampled according to the Bernoulli model with probability $p \approx s/n_1 n_2$ and the solution $\boldsymbol{M}$ is a rank-r matrix satisfying μ-coherent property, then for all $\boldsymbol{Y} \in \mathbb{R}^{n_1 \times n_2}$:

$$\begin{aligned} (1 - \tau) p \left\| \boldsymbol{P}_T(\boldsymbol{Y}) \right\|_F &\leq \left\| (\boldsymbol{P}_T \boldsymbol{P}_\Omega \boldsymbol{P}_T)(\boldsymbol{Y}) \right\|_F \\ &\leq (1 + \tau) p \left\| \boldsymbol{P}_T(\boldsymbol{Y}) \right\|_F, \qquad w.h.p., \end{aligned}$$

where τ is an arbitrarily small constant such that $C_R \sqrt{\frac{\mu n r \log n}{s}} \leq \tau < 1$, [1] for $n = \max(n_1, n_2)$. Finally, translating this into our context, we can show that under the same assumptions (uniform sampling and incoherence property) and *w.h.p.*, the matrix $\boldsymbol{H}$ is full rank with the property

$$\left\| \boldsymbol{H} \boldsymbol{x} \right\| \geq (1 - \tau) p \left\| \boldsymbol{x} \right\|, \quad \forall \boldsymbol{x} \in \mathbb{R}^{n_1 n_2 - s}.$$

This implies $\lambda_{\min}(\boldsymbol{H}) \geq (1 - \tau) p > 0$. We conclude that if $\boldsymbol{M}$ is incoherence and Ω is a uniform sampling, then $(\boldsymbol{X}, \Omega) \in \mathcal{S}$ *w.h.p.* Beyond these traditional assumptions, the definition of $\mathcal{S}$ allows us to identify other cases that can guarantee linear convergence (e.g., in deterministic settings of Ω and various structures of $\boldsymbol{X}$ that does not satisfy incoherence property).

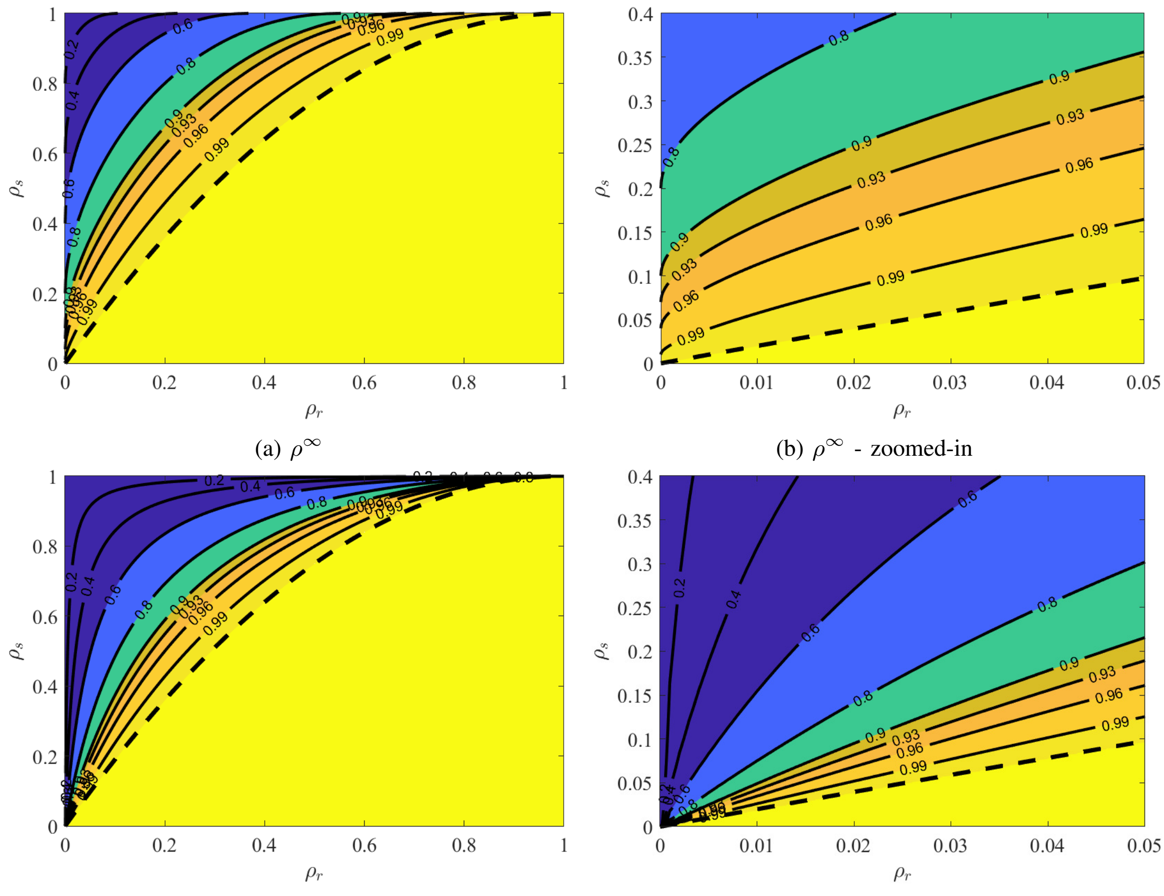
II. Convergence of IHT with the Optimal Step Size for Large-Scale Matrix Completion

It is noteworthy that the exact expression of the convergence rate provides more insights into the asymptotic behavior of PGD that can be **independent of the local structure** of the problem. As it is studied in Section IV of the original manuscript, our result extends outside the fixed (low) rank regime considered in existing works and offer a way to evaluate the behavior of IHTSVD under more challenging conditions. In particular, when $\boldsymbol{U}$ and $\boldsymbol{V}$ are selected at random (e.g., from the Haar ensemble) with $r \sim O(\min\{n_1, n_2\})$ and $s \sim O(n_1 n_2)$, we show that the convergence rate approaches a limit which is independent of the actual matrix $\boldsymbol{X}$ and only depends on its dimensions (n_1, n_2), its rank r, and the sampling rate s: $\rho_\eta^\infty \to \max\{|1 - \eta\big(\sqrt{(1 - \rho_r)^2 \rho_s} + \sqrt{\rho_r (2 - \rho_r)(1 - \rho_s)}\big)^2|, |1 -$

[1]In [1], C_R is some absolute constant that is independent of the problem parameters and the authors pick $\tau = 1/2$.

Trung Vu and Raviv Raich are with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA. (e-mails: vutru@oregonstate.edu and raich@oregonstate.edu).

Evgenia Chunikhina is with Department of Mathematics and Computer Science, Pacific University, Forest Grove, OR 97116, USA. (e-mails:chunikhina@pacificu.edu).



(a) ρ^{∞} (b) ρ^{∞} - zoomed-in

(c) ρ^{∞}_{opt} (d) ρ^{∞}_{opt} - zoomed-in

Fig. 1: Contour plots of ρ_1^{∞} and ρ_{opt}^{∞} as 2-D functions of ρ_r and ρ_s given by (1) and (2), respectively. **(a) and (c):** the entire feasible range $\rho_r \in (0,1]$ and $\rho_s \in (0,1]$; **(b) and (d):** zoomed-in version of (a) and (c) near the bottom-left corner, respectively. The isoline at which $\rho_1^{\infty} = \rho_{opt}^{\infty} = 1$ is represented by the dashed line, corresponding to the case $\rho_s = 1-(1-\rho_r)^2$. The yellow region below this isoline corresponds to the under-determined setting $\rho_s < 1-(1-\rho_r)^2$. The common setting considered in the literature (e.g., [2]–[4] is the zoomed-in region where $\rho_r \ll \rho_s \ll 1$. On the other hand, our local convergence analysis covers the entire region in which the rank ratio and the sampling rate are not necessarily small.

$\eta\big(\sqrt{(1-\rho_r)^2\rho_s} - \sqrt{\rho_r(2-\rho_r)(1-\rho_s)}\big)^2|\}$. When $\eta = 1$, we have

$$\rho_1^{\infty} = 1 - \Big(\sqrt{(1-\rho_r)^2\rho_s} - \sqrt{\rho_r(2-\rho_r)(1-\rho_s)}\Big)^2. \quad (1)$$

In addition, the optimal step size selected using this strategy and the corresponding optimal rate are given by

$$\eta_{opt}^{\infty} = \frac{1}{(1-\rho_r)^2\rho_s + \rho_r(2-\rho_r)(1-\rho_s)},$$
$$\rho_{opt}^{\infty} = \frac{2\sqrt{(1-\rho_r)^2\rho_s\rho_r(2-\rho_r)(1-\rho_s)}}{(1-\rho_r)^2\rho_s + \rho_r(2-\rho_r)(1-\rho_s)}. \quad (2)$$

Figure 1 demonstrates the rate of convergence in various setting of ρ_r and ρ_s. Note that if we evaluate this step-size choice under the regime suggested in [2], [3], i.e., $\lim_{\min\{n_1,n_2\}\to\infty} \rho_s = 0$ and $\lim_{m\to\infty} \rho_r/\rho_s^2 = 0$, then

$$\eta_{opt} = \frac{1}{\rho_s}\Big(1 + o(\rho_s)\Big), \quad \rho_{opt} = 2\sqrt{\frac{2\rho_r}{\rho_s}}\big(1 + o(\rho_s)\big). \quad (3)$$

Comparing this with the step size $1/\rho_s$ selected in [3], [4], this provides the insight that the step size used in the approach of [3] not only guarantees linear convergence but also is optimal and cannot be improved upon. Notwithstanding, our local convergence analysis offers more precise estimate of the convergence rate compared to the 0.5 upper bound in prior works. In particular, in the aforementioned regime ($\rho_r \ll \rho_s$), our estimate of the rate ρ_{opt} approaches 0, which is much faster than the upper bound 0.5 (see Fig. 2).

III. Details of Example 1

A. The first case

Using the same argument as in Lemma 5.3 in [5], we can replace the complex matrix in (22) by a real PSD matrix and prove the following lemma:

Lemma 1. *Let* $\boldsymbol{a} = [a_1,\ldots,a_{qn}]^{\top}$ *is a random vector with* $i.i.d$ *entries, where* $a_i \sim \mathcal{N}(0,1/n)$. *Then for any sequence of* $qn \times qn$ *PSD matrices* $\boldsymbol{M}_{qn}$ *with uniformly bounded spectral norms* $\|\boldsymbol{M}_{qn}\|_2$, *we have*

$$\Big(\boldsymbol{a}^{\top}\boldsymbol{M}_{qn}\boldsymbol{a} - \frac{1}{n}\mathrm{tr}(\boldsymbol{M}_{qn})\Big) \xrightarrow{p} 0 \text{ as } n\to\infty.$$

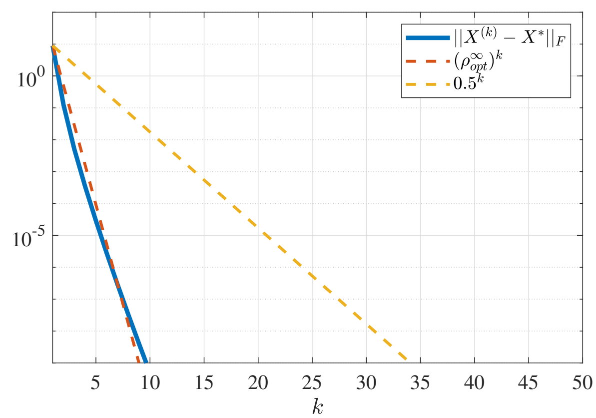


Fig. 2: Convergence of IHT with step size $\eta = n_1 n_2/s$ under the setting $\rho_s = .2$ and $\rho_r = 0.0001$. With the matrix dimension being 10000, the difference between $\eta = n_1 n_2/s$ and the optimal step size η_{opt} given in (2) is as small as 0.003. The blue solid line represents the error through IHT iterations. The red and yellow dashed lines represent the exponential decrease at rates $\rho_{opt}^{\infty} = 0.056$ given in (2) and 0.5 given in [4], respectively. Our estimate of the rate is tighter than the 0.5 global upper-bound in this asymptotic regime.

Proof. To simplify our notation, let us denote the (i,j)-th entry of $\boldsymbol{M}_{qn}$ by M_{ij} and δ_{ij} is the indicator of the event $i = j$. We follow a similar derivation as in [6]. Since a_i are $i.i.d$ normally distributed, we have

$$\mathbb{E}[a_i] = 0, \quad \mathbb{E}[a_i a_j] = \delta_{ij}\frac{1}{n},$$
$$\mathbb{E}[a_i a_j a_k a_l] = (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\frac{1}{n^2}, \tag{4}$$

for any indices $1 \leq i,j,k,l \leq n$. In order to prove $\big(\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a} - \frac{1}{n}\operatorname{tr}(\boldsymbol{M}_{qn})\big) \xrightarrow{\text{p}} 0$, it is sufficient to show that

$$\begin{cases} \mathbb{E}[\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a}] = \frac{1}{n}\operatorname{tr}(\boldsymbol{M}_{qn}), \\ \operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a}) \to 0 \text{ as } n \to \infty. \end{cases}$$

First, by the linearity of expectation, we have

$$\mathbb{E}[\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a}] = \mathbb{E}\Big[\sum_{i,j} M_{ij}a_i a_j\Big] = \sum_{i,j} M_{ij}\mathbb{E}[a_i a_j]$$
$$= \sum_{i,j} M_{ij}\delta_{ij}\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{qn} M_{ii} = \frac{1}{n}\operatorname{tr}(\boldsymbol{M}_{qn}). \tag{5}$$

Second, by rewriting the variance of the summation $\sum_{i,j} M_{ij}a_i a_j$ in terms of the sum of covariances, we obtain

$$\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a}) = \operatorname{Var}\Big(\sum_{i,j} M_{ij}a_i a_j\Big)$$
$$= \sum_{i,j,k,l} \operatorname{Cov}(M_{ij}a_i a_j, M_{kl}a_k a_l). \tag{6}$$

Using the formula

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \tag{7}$$

and the linearity of expectation, (6) can be represented as

$$\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a})$$
$$= \sum_{i,j,k,l} M_{ij}M_{kl}\Big(\mathbb{E}[a_i a_j a_k a_l] - \mathbb{E}[a_i a_j]\mathbb{E}[a_k a_l]\Big)$$
$$= \sum_{i,j,k,l} M_{ij}M_{kl}\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big)\frac{1}{n^2}$$
$$= \frac{2}{n^2}\sum_{i,j} M_{ij}^2 = \frac{2}{n^2}\|\boldsymbol{M}_{qn}\|_F^2. \tag{8}$$

Since $\boldsymbol{M}_{qn}$ is PSD and has bounded spectral norm, all of its eigenvalues are bounded by $0 \leq \lambda_i(\boldsymbol{M}_{qn}) \leq C$, for some constant C, and hence,

$$\|\boldsymbol{M}_{qn}\|_F^2 = \sum_{i=1}^{qn} \lambda_i^2(\boldsymbol{M}_{qn}) \leq qnC^2.$$

Thus, substituting back into (8) yields

$$\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{qn}\boldsymbol{a}) \leq \frac{2}{n^2}qnC^2 \to 0 \text{ as } n \to \infty.$$

This completes our proof of the lemma. □

B. The second case

Similarly, we consider the following lemma:

Lemma 2. *Let $\boldsymbol{b} = [b_1, \ldots, b_{qn}]$ and $\boldsymbol{c} = [c_1, \ldots, c_{qn}]$ are random vectors with $i.i.d$ entries, where $b_i, c_j \sim \mathcal{N}(0, 1/n)$. Denote $m = n^2$, $k = q^2$ and $\boldsymbol{a} = \boldsymbol{b} \otimes \boldsymbol{c}$. Then for any sequence of $km \times km$ PSD matrices $\boldsymbol{M}_{km}$ with uniformly bounded spectral norms $\|\boldsymbol{M}_{km}\|_2$, we have*

$$\Big(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a} - \frac{1}{m}\operatorname{tr}(\boldsymbol{M}_{km})\Big) \xrightarrow{\text{p}} 0 \text{ as } n \to \infty.$$

Proof. Denote $\boldsymbol{M}_{[ij]}$ is the (i,j)-th $qn \times qn$ block of $\boldsymbol{M}_{km}$. Then it is straightforward to verify that

$$\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a} = \sum_{i,j} b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j.$$

In order to prove $\big(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a} - \frac{1}{m}\operatorname{tr}(\boldsymbol{M}_{km})\big) \xrightarrow{\text{p}} 0$, it is sufficient to show that

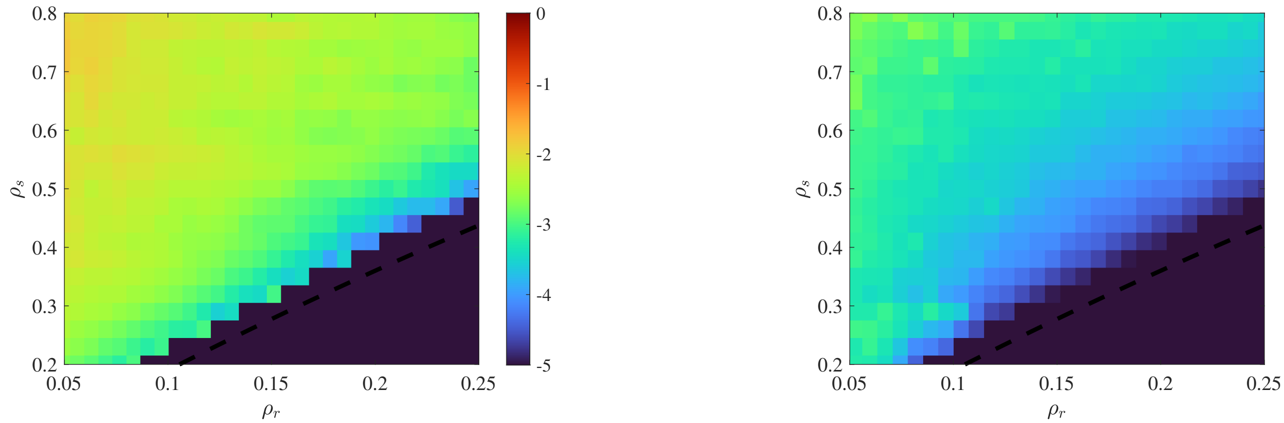
$$\begin{cases} \mathbb{E}[\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}] = \frac{1}{m}\operatorname{tr}(\boldsymbol{M}_{km}), \\ \operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}) \to 0 \text{ as } n \to \infty. \end{cases}$$

First, we use the linearity of expectation to obtain $\mathbb{E}[\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}] = \mathbb{E}\Big[\sum_{i,j} b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j\Big] = \sum_{i,j}\mathbb{E}[b_i b_j]\mathbb{E}[\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c}]$. From (5) and Lemma 1, the last equation is equivalent to $\mathbb{E}[\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}] = \sum_{i,j}\delta_{ij}\frac{1}{n}\cdot\frac{1}{n}\operatorname{tr}(\boldsymbol{M}_{[ij]}) = \frac{1}{m}\operatorname{tr}(\boldsymbol{M}_{km})$. Second, we have

$$\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}) = \operatorname{Var}\Big(\sum_{i,j} b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j\Big)$$
$$= \sum_{i,j,k,l}\operatorname{Cov}\big(b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j, b_k(\boldsymbol{c}^\top \boldsymbol{M}_{[kl]}\boldsymbol{c})b_l\big). \tag{9}$$

From (7), each covariance on the RHS of (9) can be represented as

$$\operatorname{Cov}\big(b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j, b_k(\boldsymbol{c}^\top \boldsymbol{M}_{[kl]}\boldsymbol{c})b_l\big)$$
$$= \mathbb{E}[b_i b_j b_k b_l]\cdot\mathbb{E}[\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c}\cdot\boldsymbol{c}^\top \boldsymbol{M}_{[kl]}\boldsymbol{c}]$$
$$- \mathbb{E}[b_i b_j]\cdot\mathbb{E}[b_k b_l]\cdot\mathbb{E}[\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c}]\cdot\mathbb{E}[\boldsymbol{c}^\top \boldsymbol{M}_{[kl]}\boldsymbol{c}]. \tag{10}$$



(a) $n_1 = 500, n_2 = 400$ (b) $n_1 = 1200, n_2 = 1000$

Fig. 3: The coefficient of variation (on a log10 scale) of the empirical rate shown in Fig. 5-(a) and (b), respectively. In each plot, the black dashed line corresponds to the boundary line $1-\rho_s = (1-\rho_r)^2$ and the black region on the bottom-right corner corresponds to the settings where no linear convergence is observed (i.e., the empirical rate is set to 1). The darker color in the right plot demonstrates the increasing concentration of the empirical rate as a random variable when the dimensions grow larger. It is also interesting to note that the variability in relation to the mean decreases as it approaches the boundary line (i.e., from the top-left corner to the bottom-right corner).

Lemma 3. *Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be matrices in $\mathbb{R}^{qn\times qn}$. Then*

$$\mathbb{E}[\boldsymbol{c}^\top \boldsymbol{P}\boldsymbol{c}\cdot \boldsymbol{c}^\top \boldsymbol{Q}\boldsymbol{c}] = \frac{\operatorname{tr}(\boldsymbol{P})\operatorname{tr}(\boldsymbol{Q}) + \operatorname{tr}(\boldsymbol{P}\boldsymbol{Q}^\top) + \operatorname{tr}(\boldsymbol{P}\boldsymbol{Q})}{n^2}.$$

The proof of Lemma 3 is straightforward from (4). From Lemma 3 and (4), we can simplify (10) as

$$\begin{aligned}
&\operatorname{Cov}\big(b_i(\boldsymbol{c}^\top \boldsymbol{M}_{[ij]}\boldsymbol{c})b_j, b_k(\boldsymbol{c}^\top \boldsymbol{M}_{[kl]}\boldsymbol{c})b_l\big)\\
&\quad = \frac{1}{n^4}\Big(\operatorname{tr}(\boldsymbol{M}_{[ij]}\boldsymbol{M}_{[kl]}) + \operatorname{tr}(\boldsymbol{M}_{[ij]}\boldsymbol{M}_{[kl]}^\top)\\
&\qquad + \operatorname{tr}^2(\boldsymbol{M}_{[ij]}) + \operatorname{tr}(\boldsymbol{M}_{[ij]}^2) + \operatorname{tr}(\boldsymbol{M}_{[ij]}\boldsymbol{M}_{[ij]}^\top)\\
&\qquad + \operatorname{tr}(\boldsymbol{M}_{[ij]})\operatorname{tr}(\boldsymbol{M}_{[ij]}^\top) + \operatorname{tr}(\boldsymbol{M}_{[ij]}^2)\Big).
\end{aligned}$$

Substituting the last equation back into (9) yields

$$\begin{aligned}
\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}) &= \frac{2}{n^4}\Big(\sum_{i,j}\operatorname{tr}^2(\boldsymbol{M}_{[ij]}) + \sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ii]}\boldsymbol{M}_{[jj]})\\
&\quad + \sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ij]}^\top \boldsymbol{M}_{[jj]}) + \sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ij]}^2)\Big). \qquad (11)
\end{aligned}$$

Next, we bound each term on the RHS of (11). To that end, we utilize the following lemma:

Lemma 4. *For any matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$, it holds that*

1) $\|\boldsymbol{A}\|_F \le \sqrt{n}\,\|\boldsymbol{A}\|_2$,
2) $\operatorname{tr}^2(\boldsymbol{A}) \le n\,\|\boldsymbol{A}\|_F^2$,
3) $\operatorname{tr}(\boldsymbol{A}^\top \boldsymbol{B}) \le \|\boldsymbol{A}\|_F\,\|\boldsymbol{B}\|_F \le n\,\|\boldsymbol{A}\|_2\,\|\boldsymbol{B}\|_2$,
4) $\operatorname{tr}(\boldsymbol{A}^2) \le \|\boldsymbol{A}\|_F^2 = \operatorname{tr}(\boldsymbol{A}^\top \boldsymbol{A})$.

The proof of Lemma 4 can be found in [7] - Chapter 5. Applying Lemma 4 with the blocks of size $qn \times qn$, we obtain

$$\begin{aligned}
\sum_{i,j}\operatorname{tr}^2(\boldsymbol{M}_{[ij]}) &\le \sum_{i,j} qn\,\|\boldsymbol{M}_{[ij]}\|_F^2 = qn\,\|\boldsymbol{M}\|_F^2\\
&\le (qn)^3\,\|\boldsymbol{M}\|_2 \le C(qn)^3,
\end{aligned}$$

$$\begin{aligned}
\sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ii]}\boldsymbol{M}_{[jj]}) &\le \sum_{i,j} qn\,\|\boldsymbol{M}_{[ii]}\|_2\,\|\boldsymbol{M}_{[jj]}\|_2\\
&\le \sum_{i,j} qn\,\|\boldsymbol{M}\|_2\,\|\boldsymbol{M}\|_2 = C^2(qn)^3,
\end{aligned}$$

$$\sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ij]}^\top \boldsymbol{M}_{[jj]}) = \sum_{i,j}\|\boldsymbol{M}_{[ij]}\|_F^2 = \|\boldsymbol{M}\|_F^2 \le C(qn)^2,$$

$$\sum_{i,j}\operatorname{tr}(\boldsymbol{M}_{[ij]}^2) \le \sum_{i,j}\|\boldsymbol{M}_{[ij]}\|_F^2 = \|\boldsymbol{M}\|_F^2 \le C(qn)^2.$$

Therefore, (11) can be bounded as

$$\operatorname{Var}(\boldsymbol{a}^\top \boldsymbol{M}_{km}\boldsymbol{a}) \le \frac{2}{n^4}(C(qn)^3 + C^2(qn)^3 + 2C(qn)^2).$$

The conclusion of the lemma follows by the fact that the RHS of the last equation which approaches 0 as $n \to \infty$. □

REFERENCES

[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, p. 717, 2009.
[2] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 937–945.
[3] P. Jain and P. Netrapalli, "Fast exact matrix completion with finite samples," in *Proc. Conf. Learn. Theory*, 2015, pp. 1007–1034.
[4] L. Ding and Y. Chen, "Leave-one-out approach for matrix completion: Primal and dual analysis," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 7274–7301, 2020.
[5] P. Yaskov, "The universality principle for spectral distributions of sample covariance matrices," *arXiv preprint arXiv:1410.5190*, 2014.
[6] R. Raich and J. Kim, "On the eigenvalue distribution of column subsampled semi-unitary matrices," in *Proc. IEEE Stat. Signal Process. Workshop*. IEEE, 2016, pp. 1–5.
[7] C. D. Meyer, *Matrix analysis and applied linear algebra*. Siam, 2000, vol. 71.