Exact Linear Convergence Rate Analysis for Low-Rank Matrix Completion via Gradient Descent Appendix

Trung Vu and Raviv Raich
School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA
{*vutru, raich*}@*oregonstate.edu*

A. *Proof of Lemma 1*

Recall the gradient descent update in Algorithm 1:

$$\begin{aligned} \boldsymbol{X}^{k+1} &= \boldsymbol{X}^k - \eta \mathcal{P}_\Omega\big(\boldsymbol{X}^k {\boldsymbol{X}^k}^\top - \boldsymbol{M}\big)\boldsymbol{X}^k \\ &= (\boldsymbol{I}_n - \eta \mathcal{P}_\Omega(\boldsymbol{E}^k))\boldsymbol{X}^k. \end{aligned} \tag{1}$$

Substituting (1) into the definition of $\boldsymbol{E}^{k+1}$, we have

$$\begin{aligned} \boldsymbol{E}^{k+1} &= \boldsymbol{X}^{k+1} {\boldsymbol{X}^{k+1}}^\top - \boldsymbol{M} \\ &= \big(\boldsymbol{I}_n - \eta \mathcal{P}_\Omega(\boldsymbol{E}^k)\big)\boldsymbol{X}^k {\boldsymbol{X}^k}^\top \big(\boldsymbol{I}_n - \eta \mathcal{P}_\Omega(\boldsymbol{E}^k)\big)^\top - \boldsymbol{M}. \end{aligned}$$

From the fact that $\boldsymbol{E}^k$ is symmetric and Ω is a symmetric sampling, the last equation can be further expanded as

$$\begin{aligned} \boldsymbol{E}^{k+1} &= \boldsymbol{X}^k {\boldsymbol{X}^k}^\top - \eta \mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{X}^k {\boldsymbol{X}^k}^\top \\ &- \eta \boldsymbol{X}^k {\boldsymbol{X}^k}^\top \mathcal{P}_\Omega(\boldsymbol{E}^k) + \eta^2 \mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{X}^k {\boldsymbol{X}^k}^\top \mathcal{P}_\Omega(\boldsymbol{E}^k) - \boldsymbol{M}. \end{aligned} \tag{2}$$

Since $\boldsymbol{X}^k {\boldsymbol{X}^k}^\top = \boldsymbol{M} + \boldsymbol{E}^k$, (2) is equivalent to

$$\begin{aligned} \boldsymbol{E}^{k+1} = \boldsymbol{E}^k &- \eta\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{M} + \boldsymbol{M}\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) \\ &- \eta\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{E}^k + \boldsymbol{E}^k\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) \\ &+ \eta^2 \mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{M}\mathcal{P}_\Omega(\boldsymbol{E}^k) + \eta^2 \mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{E}^k\mathcal{P}_\Omega(\boldsymbol{E}^k). \end{aligned} \tag{3}$$

Note that $\|\mathcal{P}_\Omega(\boldsymbol{E}^k)\|_F \leq \|\boldsymbol{E}^k\|_F$. Hence, collecting terms that are of second order and higher, with respect to $\|\boldsymbol{E}^k\|_F$, on the RHS of (3) yields

$$\boldsymbol{E}^{k+1} = \boldsymbol{E}^k - \eta\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\boldsymbol{M} + \boldsymbol{M}\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) + \boldsymbol{\mathcal{O}}(\|\boldsymbol{E}^k\|_F^2).$$

Now by Definition 1, it is easy to verify that

$$\boldsymbol{S}\boldsymbol{S}^\top = \boldsymbol{I}_{n^2} \quad \text{and} \quad \mathrm{vec}\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) = \boldsymbol{S}^\top \boldsymbol{S} \boldsymbol{e}^k.$$

Using the property $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^\top \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{B})$, (5) can be vectorized as follows:

$$\begin{aligned} \boldsymbol{e}^{k+1} = \boldsymbol{e}^k &- \eta(\boldsymbol{M} \otimes \boldsymbol{I}_n)\,\mathrm{vec}\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) \\ &- \eta(\boldsymbol{I}_n \otimes \boldsymbol{M})\,\mathrm{vec}\big(\mathcal{P}_\Omega(\boldsymbol{E}^k)\big) + \boldsymbol{\mathcal{O}}(\|\boldsymbol{e}^k\|_2^2). \end{aligned}$$

The last equation can be reorganized as

$$\boldsymbol{e}^{k+1} = \Big(\boldsymbol{I}_{n^2} - \eta(\boldsymbol{M} \oplus \boldsymbol{M})(\boldsymbol{S}^\top \boldsymbol{S})\Big)\boldsymbol{e}^k + \boldsymbol{\mathcal{O}}(\|\boldsymbol{e}^k\|_2^2).$$

B. *Proof of Lemma 3*

($\Rightarrow$) Suppose $\boldsymbol{E} \in \mathcal{E}$. Then for (C1), i.e., $\boldsymbol{E}^\top = \boldsymbol{E}$, $\boldsymbol{E} = \boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{M}$ is symmetric since both $\boldsymbol{X}\boldsymbol{X}^\top$ and $\boldsymbol{M}$ are symmetric. For (C2), i.e., $\mathcal{P}_r(\boldsymbol{M}+\boldsymbol{E}) = \boldsymbol{M}+\boldsymbol{E}$, stems from the fact $\boldsymbol{M}+\boldsymbol{E} = \boldsymbol{X}\boldsymbol{X}^\top$ has rank no greater than r for $\boldsymbol{X} \in \mathbb{R}^{n \times r}$. Finally, for any $\boldsymbol{v} \in \mathbb{R}^n$, we have

$$\boldsymbol{v}^\top(\boldsymbol{M}+\boldsymbol{E})\boldsymbol{v} = \boldsymbol{v}^\top(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{v} = \|\boldsymbol{X}^\top \boldsymbol{v}\|_2^2 \geq 0.$$

($\Leftarrow$) From conditions (C1) and (C3), $\boldsymbol{M}+\boldsymbol{E}$ is a PSD matrix. In addition, $\mathcal{P}_r(\boldsymbol{M}+\boldsymbol{E}) = \boldsymbol{M}+\boldsymbol{E}$ implies $\boldsymbol{M}+\boldsymbol{E}$ must have rank no greater r. Since any PSD matrix $\boldsymbol{A}$ with rank less than or equal to r can be factorized as $\boldsymbol{A} = \boldsymbol{Y}\boldsymbol{Y}^\top$ for some $\boldsymbol{Y} \in \mathbb{R}^{n \times r}$, we conclude that $\boldsymbol{E} \in \mathcal{E}$.

C. *Proof of Lemma 4*

First, recall that any matrix $\boldsymbol{\Pi} \in \mathbb{R}^{n^2 \times n^2}$ is an orthogonal projection if and only if $\boldsymbol{\Pi}^2 = \boldsymbol{\Pi}$ and $\boldsymbol{\Pi} = \boldsymbol{\Pi}^\top$. Since $\boldsymbol{P}_{\boldsymbol{U}_\perp}^\top = \boldsymbol{P}_{\boldsymbol{U}_\perp}$, we have

$$\begin{aligned} \boldsymbol{P}_1^\top &= \big(\boldsymbol{I}_{n^2} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}\big)^\top \\ &= \boldsymbol{I}_{n^2}^\top - \boldsymbol{P}_{\boldsymbol{U}_\perp}^\top \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}^\top \\ &= \boldsymbol{I}_{n^2} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} = \boldsymbol{P}_1. \end{aligned}$$

In addition, since $\boldsymbol{P}_{\boldsymbol{U}_\perp}^2 = \boldsymbol{P}_{\boldsymbol{U}_\perp}$, we have

$$\begin{aligned} \boldsymbol{P}_1^2 &= (\boldsymbol{I}_{n^2} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp})(\boldsymbol{I}_{n^2} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp})^\top \\ &= \boldsymbol{I}_{n^2}^2 - 2\boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} + (\boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp})^2 \\ &= \boldsymbol{I}_{n^2} - 2\boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} + (\boldsymbol{P}_{\boldsymbol{U}_\perp}^2 \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp}^2) \\ &= \boldsymbol{I}_{n^2} - 2\boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} + \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} \\ &= \boldsymbol{I}_{n^2} - \boldsymbol{P}_{\boldsymbol{U}_\perp} \otimes \boldsymbol{P}_{\boldsymbol{U}_\perp} = \boldsymbol{P}_1. \end{aligned}$$

Second, using the fact that $\boldsymbol{T}_{n^2}^2 = \boldsymbol{I}_{n^2}$ and $\boldsymbol{T}_{n^2}$ is symmetric, we can derive similar result:

$$\boldsymbol{P}_2^\top = \left(\frac{\boldsymbol{I}_{n^2} + \boldsymbol{T}_{n^2}}{2}\right)^\top = \frac{\boldsymbol{I}_{n^2} + \boldsymbol{T}_{n^2}}{2} = \boldsymbol{P}_2,$$

and

$$\begin{aligned}\boldsymbol{P}_2^2 &= \frac{(\boldsymbol{I}_{n^2}+\boldsymbol{T}_{n^2})^2}{4}\\ &= \frac{\boldsymbol{I}_{n^2}+2\boldsymbol{T}_{n^2}+\boldsymbol{T}_{n^2}^2}{4}\\ &= \frac{2\boldsymbol{I}_{n^2}+2\boldsymbol{T}_{n^2}}{4}\\ &= \frac{\boldsymbol{I}_{n^2}+\boldsymbol{T}_{n^2}}{2} = \boldsymbol{P}_2.\end{aligned}$$

Third, we observe that $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are the vectorized version of the linear operators

$$\boldsymbol{\Pi}_1(\boldsymbol{E}) = \boldsymbol{E} - \boldsymbol{P}_{\boldsymbol{U}_\perp}\boldsymbol{E}\boldsymbol{P}_{\boldsymbol{U}_\perp}$$

and

$$\boldsymbol{\Pi}_2(\boldsymbol{E}) = \frac{1}{2}(\boldsymbol{E}+\boldsymbol{E}^\top),$$

respectively, for any $\boldsymbol{E}\in\mathbb{R}^{n\times n}$. Hence, in order to prove that $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ commute, it is sufficient to show that operators $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ commute. Indeed, we have

$$\begin{aligned}\boldsymbol{\Pi}_2\boldsymbol{\Pi}_1(\boldsymbol{E}) &= \frac{1}{2}\big((\boldsymbol{E}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\boldsymbol{E}\boldsymbol{P}_{\boldsymbol{U}_\perp}) + (\boldsymbol{E}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\boldsymbol{E}\boldsymbol{P}_{\boldsymbol{U}_\perp})^\top\big)\\ &= \frac{1}{2}(\boldsymbol{E}+\boldsymbol{E}^\top) - \boldsymbol{P}_{\boldsymbol{U}_\perp}\frac{1}{2}(\boldsymbol{E}+\boldsymbol{E}^\top)\boldsymbol{P}_{\boldsymbol{U}_\perp}\\ &= \boldsymbol{\Pi}_1\boldsymbol{\Pi}_2(\boldsymbol{E}).\end{aligned}$$

This implies $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ commute. Since $\boldsymbol{P}$ is the product of two commuting orthogonal projections, it is also an orthogonal projection.

Finally, let us restrict $\boldsymbol{E}$ to belong to $\mathcal{E}$ and denote $\boldsymbol{e} = \mathrm{vec}(\boldsymbol{E})$. Using Theorem 3 in [1], we have

$$\mathcal{P}_r(\boldsymbol{M}+\boldsymbol{E}) = \boldsymbol{M}+\boldsymbol{E}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\boldsymbol{E}\boldsymbol{P}_{\boldsymbol{U}_\perp} + \mathcal{O}(\|\boldsymbol{E}\|_F^2). \tag{4}$$

Since $\mathcal{P}_r(\boldsymbol{M}+\boldsymbol{E}) = \boldsymbol{M}+\boldsymbol{E}$, it follows from (4) that

$$\boldsymbol{P}_{\boldsymbol{U}_\perp}\boldsymbol{E}\boldsymbol{P}_{\boldsymbol{U}_\perp} = \mathcal{O}(\|\boldsymbol{E}\|_F^2).$$

Vectorizing the last equation, we obtain

$$(\boldsymbol{P}_{\boldsymbol{U}_\perp}\otimes\boldsymbol{P}_{\boldsymbol{U}_\perp})\boldsymbol{e} = \mathcal{O}(\|\boldsymbol{E}\|_F^2). \tag{5}$$

On the other hand, since $\boldsymbol{E}$ is symmetric,

$$\boldsymbol{e} = \boldsymbol{T}_{n^2}\boldsymbol{e} = \Big(\frac{\boldsymbol{I}_{n^2}+\boldsymbol{T}_{n^2}}{2}\Big)\boldsymbol{e}. \tag{6}$$

From (5) and (6), we have

$$\begin{aligned}\boldsymbol{e} &= (\boldsymbol{I}_{n^2}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\otimes\boldsymbol{P}_{\boldsymbol{U}_\perp})\boldsymbol{e} + \mathcal{O}(\|\boldsymbol{E}\|_F^2)\\ &= (\boldsymbol{I}_{n^2}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\otimes\boldsymbol{P}_{\boldsymbol{U}_\perp})\Big(\frac{\boldsymbol{I}_{n^2}+\boldsymbol{T}_{n^2}}{2}\Big)\boldsymbol{e} + \mathcal{O}(\|\boldsymbol{E}\|_F^2).\end{aligned} \tag{7}$$

Substituting

$$\boldsymbol{P} = \boldsymbol{P}_1\boldsymbol{P}_2 = (\boldsymbol{I}_{n^2}-\boldsymbol{P}_{\boldsymbol{U}_\perp}\otimes\boldsymbol{P}_{\boldsymbol{U}_\perp})\Big(\frac{\boldsymbol{I}_{n^2}+\boldsymbol{T}_{n^2}}{2}\Big)$$

into (7) completes our proof of the lemma.

D. Proof of Lemma 5

Let $\tilde{\mathcal{E}} = \{\mathrm{vec}(\boldsymbol{E}) \mid \boldsymbol{E}\in\mathcal{E}\}$. Recall that for any $\boldsymbol{e}\in\tilde{\mathcal{E}}$,

$$\boldsymbol{e} = \boldsymbol{P}\boldsymbol{e} + \boldsymbol{\mathcal{O}}(\|\boldsymbol{e}\|_2^2).$$

Therefore, by the triangle inequality, we obtain

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{e}\|_2 &= \|\boldsymbol{A}\big(\boldsymbol{P}\boldsymbol{e}+\boldsymbol{\mathcal{O}}(\|\boldsymbol{e}\|_2^2)\big)\|\\ &\le \|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2 + \|\boldsymbol{A}\boldsymbol{\mathcal{O}}(\|\boldsymbol{e}\|_2^2)\|_2.\end{aligned}$$

Since the second term on the RHS of the last inequality is $\mathcal{O}(\|\boldsymbol{e}\|_2^2)$, it is also $\mathcal{O}(\delta^2)$ for any $\boldsymbol{e}\in\tilde{\mathcal{E}}$ such that $\|\boldsymbol{e}\|_2\le\delta$. In other words,

$$\|\boldsymbol{A}\boldsymbol{e}\|_2 \le \|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2 + \mathcal{O}(\delta^2). \tag{8}$$

On the other hand, by the Pythagorean theorem, $\boldsymbol{A}\boldsymbol{e}$ is the sum of $\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}$ and its orthogonal counterpart. Thus,

$$\|\boldsymbol{A}\boldsymbol{e}\|_2 \ge \|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2. \tag{9}$$

From (8) and (9), we have

$$\frac{\|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2} \le \frac{\|\boldsymbol{A}\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2} \le \frac{\|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2} + \mathcal{O}(\delta). \tag{10}$$

Taking the limit of the supremum of (10) as $\delta\to0$ yields

$$\begin{aligned}\rho^{\mathcal{E}}(\boldsymbol{A}) &= \lim_{\delta\to0}\sup_{\substack{\boldsymbol{e}\in\tilde{\boldsymbol{e}}\\ \boldsymbol{e}\neq0\\ \|\boldsymbol{e}\|_2\le\delta}} \frac{\|\boldsymbol{A}\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2}\\ &= \lim_{\delta\to0}\sup_{\substack{\boldsymbol{e}\in\tilde{\boldsymbol{e}}\\ \boldsymbol{e}\neq0\\ \|\boldsymbol{e}\|_2\le\delta}} \frac{\|\boldsymbol{A}\boldsymbol{P}\boldsymbol{e}\|_2}{\|\boldsymbol{e}\|_2} = \rho^{\mathcal{E}}(\boldsymbol{A}\boldsymbol{P}).\end{aligned} \tag{11}$$

Now following similar argument in Lemma 6, we have

$$\begin{cases}\rho^{\mathcal{E}}(\boldsymbol{A}\boldsymbol{P}) = \rho(\boldsymbol{A}\boldsymbol{P}),\\ \rho^{\mathcal{E}}(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}) = \rho(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}).\end{cases} \tag{12}$$

Given (11) and (12), it remains to show that $\rho(\boldsymbol{A}\boldsymbol{P}) = \rho(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})$. Indeed, using Gelfand's formula [2], we have

$$\rho(\boldsymbol{A}\boldsymbol{P}) = \lim_{k\to\infty}\|(\boldsymbol{A}\boldsymbol{P})^k\|_2^{1/k}$$
$$\text{and } \rho(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}) = \lim_{k\to\infty}\|(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})^k\|_2^{1/k}.$$

Furthermore, by the property of operator norms

$$\|(\boldsymbol{A}\boldsymbol{P})^k\|_2 = \|\boldsymbol{A}(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})^{k-1}\|_2 \le \|\boldsymbol{A}\|_2\|(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})^{k-1}\|_2.$$

Thus,

$$\|(\boldsymbol{A}\boldsymbol{P})^k\|_2^{1/k} \le \|\boldsymbol{A}\|_2^{1/k}\Big(\|(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})^{k-1}\|_2^{1/(k-1)}\Big)^{(k-1)/k}.$$

Taking the limit of both sides of the last inequality as $k\to\infty$ yields $\rho(\boldsymbol{A}\boldsymbol{P})\le\rho(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})$. Similarly, since

$$\|(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})^k\|_2 = \|\boldsymbol{P}(\boldsymbol{A}\boldsymbol{P})^k\|_2 \le \|(\boldsymbol{A}\boldsymbol{P})^k\|_2,$$

we also obtain $\rho(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})\le\rho(\boldsymbol{A}\boldsymbol{P})$. This concludes our proof of the lemma.

E. Proof of Lemma 6

Without loss of generality, assume σ_1 is the eigenvalue with largest magnitude, i.e., $\rho(\boldsymbol{H}) = |\sigma_1| > 0$, and $\boldsymbol{q}_1$ is its corresponding eigenvector (with unit norm). Denote $\boldsymbol{G}$ the matrix such that $\mathrm{vec}(\boldsymbol{G}) = \delta \boldsymbol{q}_1$. First, we have $\|\boldsymbol{G}\|_F = \delta$. Second, since $\boldsymbol{H}\boldsymbol{q}_1 = \sigma_1 \boldsymbol{H}$, we have $\sigma_1 \mathrm{vec}(\boldsymbol{G}) = \boldsymbol{H}\,\mathrm{vec}(\boldsymbol{G})$. By the aforementioned eigendecomposition of $\boldsymbol{H}$, one can then verify that $\boldsymbol{G}$ is a solution of (13):

$$\frac{\|\boldsymbol{\Lambda_H} \boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G})\|_2}{\|\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G})\|_2} = |\sigma_1| = \rho(\boldsymbol{H}).$$

Third, recall that $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{H}$ and $\boldsymbol{P}_1\boldsymbol{P} = \boldsymbol{P}_1\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}$. Thus, $\boldsymbol{H} = \boldsymbol{P}_1\boldsymbol{H}$ and

$$\sigma_1 \mathrm{vec}(\boldsymbol{G}) = \boldsymbol{P}_1 \boldsymbol{H}\,\mathrm{vec}(\boldsymbol{G}) = \sigma_1 \boldsymbol{P}_1 \mathrm{vec}(\boldsymbol{G}).$$

Since $\boldsymbol{P}_1 = \boldsymbol{I}_{n^2} - \boldsymbol{P}_{U_\perp} \otimes \boldsymbol{P}_{U_\perp}$, $(\boldsymbol{P}_{U_\perp} \otimes \boldsymbol{P}_{U_\perp})\,\mathrm{vec}(\boldsymbol{G}) = \boldsymbol{0}$ or $\boldsymbol{P}_{U_\perp}\boldsymbol{G}\boldsymbol{P}_{U_\perp} = \boldsymbol{0}$. Finally, using the commutation of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, similar derivation also leads to $\boldsymbol{G}^\top = \boldsymbol{G}$. Since $\|\boldsymbol{E} - \boldsymbol{G}\|_F \le \|\boldsymbol{E} - \boldsymbol{F}\|_F + \|\boldsymbol{F} - \boldsymbol{G}\|_F$ for any matrices $\boldsymbol{E}, \boldsymbol{F}$, Lemmas 6 and 7 directly imply the existence of $\boldsymbol{E} \in \mathcal{E}$ such that $\|\boldsymbol{E} - \boldsymbol{G}\|_F = \mathcal{O}(\delta^2)$. We proceed to bound the distance

$$\rho(\boldsymbol{H}) - \rho^{\mathcal{E}}(\boldsymbol{H}) = \frac{\|\boldsymbol{\Lambda_H} \boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G})\|_2}{\|\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G})\|_2} - \frac{\|\boldsymbol{\Lambda_H} \boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{E})\|_2}{\|\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{E})\|_2}.$$

Denote $\tilde{\boldsymbol{e}} = \boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{E}) \in \tilde{\mathcal{E}}_{\boldsymbol{H}}$. Since $\boldsymbol{\Lambda_H}\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G}) = \sigma_1 \boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{G})$, we have

$$\boldsymbol{\Lambda_H} \tilde{\boldsymbol{e}} = \sigma_1 \tilde{\boldsymbol{e}} - (\sigma_1 \boldsymbol{I}_{n^2} - \boldsymbol{\Lambda_H})\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{E} - \boldsymbol{G}).$$

Thus, using the triangle inequality, we obtain

$$\|\boldsymbol{\Lambda_H}\tilde{\boldsymbol{e}}\|_2 \ge \|\sigma_1 \tilde{\boldsymbol{e}}\|_2 - \|(\sigma_1 \boldsymbol{I}_{n^2} - \boldsymbol{\Lambda_H})\boldsymbol{Q_H}^{-1} \mathrm{vec}(\boldsymbol{E} - \boldsymbol{G})\|_2.$$

The second term on the RHS can be upper-bounded by $\|\boldsymbol{E} - \boldsymbol{G}\|_F = \mathcal{O}(\delta^2)$. Therefore, $\|\boldsymbol{\Lambda_H}\tilde{\boldsymbol{e}}\|_2 = \rho(\boldsymbol{H})\|\tilde{\boldsymbol{e}}\|_2 - \mathcal{O}(\delta^2)$. Dividing both sides by $\|\tilde{\boldsymbol{e}}\|_2$ and reorganizing terms yield

$$\rho(\boldsymbol{H}) - \frac{\|\boldsymbol{\Lambda_H}\tilde{\boldsymbol{e}}\|_2}{\|\tilde{\boldsymbol{e}}\|_2} = \mathcal{O}(\delta).$$

The statement of the theorem follows immediately by the fact that $\|\boldsymbol{\Lambda_H}\tilde{\boldsymbol{e}}\|_2 / \|\tilde{\boldsymbol{e}}\|_2 \le \rho^{\mathcal{E}}(\boldsymbol{H}, \delta)$ for $\tilde{\boldsymbol{e}} \in \tilde{\mathcal{E}}_{\boldsymbol{H}}$.

F. Proof of Lemma 7

Denote $\boldsymbol{P_U} = \boldsymbol{U}\boldsymbol{U}^\top$, for any $\boldsymbol{v} \in \mathbb{R}^n$, we can decompose $\boldsymbol{v}$ into two orthogonal component:

$$\boldsymbol{v} = \boldsymbol{v_U} + \boldsymbol{v}_\perp,$$

where $\boldsymbol{v_U} = \boldsymbol{P_U}\boldsymbol{v}$ and $\boldsymbol{v}_\perp = \boldsymbol{P}_{U_\perp}\boldsymbol{v}$. Without loss of generality, assume that $\|\boldsymbol{v}\|_2 = \|\boldsymbol{v_U}\|_2^2 + \|\boldsymbol{v}_\perp\|_2^2 = 1$. Thus, we have

$$\begin{aligned}\boldsymbol{v}^\top(\boldsymbol{M} + \boldsymbol{G})\boldsymbol{v} &= (\boldsymbol{v_U} + \boldsymbol{v}_\perp)^\top(\boldsymbol{M} + \boldsymbol{G})(\boldsymbol{v_U} + \boldsymbol{v}_\perp) \\ &= \boldsymbol{v_U}^\top\boldsymbol{M}\boldsymbol{v_U} + \boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v_U} + \boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v}_\perp \\ &\quad + \boldsymbol{v}_\perp^\top\boldsymbol{G}\boldsymbol{v_U} + \boldsymbol{v}_\perp^\top\boldsymbol{G}\boldsymbol{v}_\perp, \end{aligned} \tag{13}$$

where the last equation stems from the fact that $\boldsymbol{M} = \mathcal{P}_{\boldsymbol{U}}\boldsymbol{M}\boldsymbol{P_U}$ and $\boldsymbol{P_U}\boldsymbol{P}_{U_\perp} = \boldsymbol{0}$. Since $\boldsymbol{P}_{U_\perp}\boldsymbol{G}\boldsymbol{P}_{U_\perp} = \boldsymbol{0}$, we have

$$\boldsymbol{v}_\perp^\top\boldsymbol{G}\boldsymbol{v}_\perp = \boldsymbol{v}^\top\boldsymbol{P}_{U_\perp}\boldsymbol{G}\boldsymbol{P}_{U_\perp}\boldsymbol{v} = 0.$$

Thus, (13) is equivalent to

$$\boldsymbol{v}^\top(\boldsymbol{M} + \boldsymbol{G})\boldsymbol{v} = \boldsymbol{v_U}^\top\boldsymbol{M}\boldsymbol{v_U} + \boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v_U} + 2\boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v}_\perp. \tag{14}$$

Now let us lower-bound each term on the RHS of (14) as follows. First, by the Rayleigh quotient, we have

$$\boldsymbol{v_U}^\top\boldsymbol{M}\boldsymbol{v_U} \ge \lambda_r\|\boldsymbol{v_U}\|_2^2, \tag{15}$$

and

$$\boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v_U} \ge \lambda_{\min}(\boldsymbol{G})\|\boldsymbol{v_U}\|_2^2 \ge -\|\boldsymbol{G}\|_F\|\boldsymbol{v_U}\|_2^2. \tag{16}$$

Next, by Cauchy-Schwarz inequality,

$$\boldsymbol{v_U}^\top\boldsymbol{G}\boldsymbol{v}_\perp \ge -\|\boldsymbol{G}\|_2\|\boldsymbol{v_U}\|_2\|\boldsymbol{v}_\perp\|_2 \ge -\|\boldsymbol{G}\|_F\|\boldsymbol{v_U}\|_2. \tag{17}$$

From (15), (16), and (17), we obtain

$$\boldsymbol{v}^\top(\boldsymbol{M} + \boldsymbol{G})\boldsymbol{v} \ge (\lambda_r - \|\boldsymbol{G}\|_F)\|\boldsymbol{v_U}\|_2^2 - 2\|\boldsymbol{G}\|_F\|\boldsymbol{v_U}\|_2. \tag{18}$$

Note that $\|\boldsymbol{G}\|_F = \delta$ and the quadratic $g(t) = (\lambda_r - \delta)t^2 - 2\delta t$ is minimized at

$$t_* = \frac{\delta}{\lambda_r - \delta}, \quad g(t_*) = -\frac{\delta^2}{\lambda_r - \delta}.$$

Combining this with (18) yields

$$\boldsymbol{v}^\top(\boldsymbol{M} + \boldsymbol{G})\boldsymbol{v} \ge -\frac{2}{\lambda_r}\delta^2,$$

for sufficiently small δ. Let $\boldsymbol{F} = \boldsymbol{G} + \frac{2}{\lambda_r}\delta^2\boldsymbol{I}_n$. Now we can easily verify that $\|\boldsymbol{F} - \boldsymbol{G}\|_F = \mathcal{O}(\delta^2)$ and $\boldsymbol{F} \in \mathcal{F}$.

G. Proof of Lemma 8

We shall show that the matrix $\boldsymbol{E} = \mathcal{P}_r(\boldsymbol{M} + \boldsymbol{F}) - \boldsymbol{M}$ belongs to $\mathcal{E}$ and satisfies

$$\|\boldsymbol{E} - \boldsymbol{F}\|_F = \mathcal{O}(\delta^2). \tag{19}$$

First, since $\boldsymbol{F} \in \mathcal{F}_\delta$, $\boldsymbol{M} + \boldsymbol{F}$ must be PSD. Thus, $\mathcal{P}_r(\boldsymbol{M} + \boldsymbol{F})$ is a PSD matrix of rank no greater than r and it admits a rank-r factorization $\mathcal{P}_r(\boldsymbol{M} + \boldsymbol{F}) = \boldsymbol{Z}\boldsymbol{Z}^\top$, for some $\boldsymbol{Z} \in \mathbb{R}^{n \times r}$. Therefore, by the definition of $\mathcal{E}$,

$$\boldsymbol{E} = \mathcal{P}_r(\boldsymbol{M} + \boldsymbol{F}) - \boldsymbol{M} = \boldsymbol{Z}\boldsymbol{Z}^\top - \boldsymbol{M} \in \mathcal{E}.$$

Next, using (4), we have

$$\begin{aligned}\boldsymbol{E} - \boldsymbol{F} &= \mathcal{P}_r(\boldsymbol{M} + \boldsymbol{F}) - \boldsymbol{M} - \boldsymbol{F} \\ &= \boldsymbol{P}_{U_\perp}\boldsymbol{F}\boldsymbol{P}_{U_\perp} + \boldsymbol{\mathcal{O}}(\|\boldsymbol{F}\|_F^2).\end{aligned}$$

Since $\boldsymbol{F} \in \mathcal{F}_\delta$ implies $\boldsymbol{P}_{U_\perp}\boldsymbol{F}\boldsymbol{P}_{U_\perp} = \boldsymbol{\mathcal{O}}(\|\boldsymbol{F}\|_F^2)$, we conclude that $\boldsymbol{E} - \boldsymbol{F} = \boldsymbol{\mathcal{O}}(\|\boldsymbol{F}\|_F^2)$.

References

[1] Trung Vu, Evgenia Chunikhina, and Raviv Raich, "Perturbation expansions and error bounds for the truncated singular value decomposition," *arXiv preprint arXiv:2009.07542*, 2020.

[2] Izrail Gelfand, "Normierte ringe," *Mathematics Sbornik*, vol. 9, no. 1, pp. 3–24, 1941.