# Appendix

## Proof of Theorem 1

First, we prove the order of singular values is preserved in a neighborhood of the rank-$r$ matrix $M$. Using Weyl's theorem, we have

$$|\sigma_i(M+\Delta) - \sigma_i| \leq \|\Delta\|_F\,, \qquad \text{for } 1 \leq i \leq n.$$

For any $i$ such that $\sigma_i > \sigma_{i+1}$: since $\|\Delta\|_F < \frac{\epsilon}{2} \leq \frac{\sigma_i - \sigma_{i+1}}{2}$, the following inequality holds

$$\sigma_{i+1}(M+\Delta) < \sigma_{i+1} + \frac{\sigma_i - \sigma_{i+1}}{2} = \sigma_i - \frac{\sigma_i - \sigma_{i+1}}{2} < \sigma_i(M+\Delta).$$

Thus, the order of singular values is preserved. Moreover, since $\sigma_r(M+\Delta) - \sigma_{r+1}(M+\Delta) > 0$, the top $r$ singular value components are unique and consequently $\mathcal{P}_r(M+\Delta)$ is unique.

Let $M = \sum_{i=1}^r \sigma_i u_i v_i^T$ be the rank-$r$ matrix of interest. From matrix perturbation theory [1], we can describe the decomposition of the perturbed matrix

$$M + \Delta = \sum_{i=1}^{r} (\sigma_i + \delta_i)(u_i + \delta u_i)(v_i + \delta v_i)^T + \sum_{i=r+1}^{n} \delta_i (u_i + \delta u_i)(v_i + \delta v_i)^T \tag{1}$$

where $\delta_i, \delta u_i$, and $\delta v_i$ have norms **in the order of** $O(\|\Delta\|_F)$. Since the top-$r$ singular values of $M$ are preserved under perturbation, we have $\mathcal{P}_r(M+\Delta) = \sum_{i=1}^r (\sigma_i + \delta_i)(u_i + \delta u_i)(v_i + \delta v_i)^T$ and (1) can be reorganized as

$$\mathcal{P}_r(M+\Delta) - M = \Delta - \sum_{i=r+1}^{n} \delta_i (u_i + \delta u_i)(v_i + \delta v_i)^T = \Delta - \sum_{i=r+1}^{n} u_i \delta_i v_i^T + O(\|\Delta\|_F^2). \tag{2}$$

Further, substituting $M = \sum_{i=1}^r \sigma_i u_i v_i^T$ into (1) yields

$$\Delta = \sum_{i=1}^{n} \left( \delta_i u_i v_i^T + \sigma_i \delta u_i v_i^T + \sigma_i u_i {\delta v_i}^T \right) + O(\|\Delta\|_F^2).$$

Then using the orthogonality of $u_i, v_i$, we can obtain

$$u_i^T \Delta v_i = \delta_i + \sigma_i (u_i^T \delta u_i + {\delta v_i}^T v_i) + O(\|\Delta\|_F^2), \tag{3}$$

$$u_i^T \Delta v_j = O(\|\Delta\|_F^2), \tag{4}$$

The second term on the RHS can be computed as follows

$$I = \sum_{i=1}^{n} (u_i + \delta u_i)(u_i + \delta u_i)^T \qquad \Rightarrow 1 = u_i^T u_i = 1 + u_i^T \delta u_i + {\delta u_i}^T u_i + O(\|\Delta\|_F^2) \qquad \Rightarrow u_i^T \delta u_i = O(\|\Delta\|_F^2)$$

Similarly, we also have $v_i^T \delta v_i = O(\|\Delta\|_F^2)$. Substituting back to (3), we get $\delta_i = u_i^T \Delta v_i + O(\|\Delta\|_F^2)$. Thus, (2) can be rewritten as

$$\mathcal{P}_r(M+\Delta) - M = \Delta - \sum_{i=r+1}^{n} u_i u_i^T \Delta v_i v_i^T + O(\|\Delta\|_F^2) = \Delta - U_2 U_2^T \Delta V_2 V_2^T + O(\|\Delta\|_F^2)$$

where the last equation stems from (4).

## Proof of Theorem 3

The error matrix can be represented as follows:

$$\begin{aligned} E^{(k)} = Y^{(k)} - M &= \mathcal{P}_{M,\mathcal{S}}\Big(X^{(k)} + \beta(X^{(k)} - X^{(k-1)})\Big) - M \\ &= [(1+\beta)(X^{(k)} - M) - \beta(X^{(k-1)} - M)]_{\mathcal{S}^c} \\ &= (1+\beta)[\mathcal{P}_r(Y^{(k-1)}) - M]_{\mathcal{S}^c} - \beta[\mathcal{P}_r(Y^{(k-2)}) - M]_{\mathcal{S}^c}. \end{aligned}$$

Using a vertorized version of Theorem 1, we can reformulate the above equation as

$$e^{(k)} = (1+\beta)(I_d - H)e^{(k-1)} - \beta(I_d - H)e^{(k-2)} + (1+\beta)q(e^{(k-1)}) - \beta q(e^{(k-2)}).$$

where $d = mn - s$, $e^{(k)} = S_c \operatorname{vec}(E^{(k)})$, $H = S_c(V_2 \otimes U_2)(V_2 \otimes U_2)^T S_c^T$ and $q(S_c \operatorname{vec}(\Delta)) = S_c \operatorname{vec}(Q(\Delta))$. By stacking $e^{(k)}$ and $e^{(k-1)}$ together, the recursion can be rewritten as follows

$$\begin{bmatrix} e^{(k)} \\ e^{(k-1)} \end{bmatrix} = \underbrace{\begin{bmatrix} (1+\beta)(I_d - H) & -\beta(I_d - H) \\ I_d & \mathbf{0} \end{bmatrix}}_{T} \begin{bmatrix} e^{(k-1)} \\ e^{(k-2)} \end{bmatrix} + \begin{bmatrix} (1+\beta)q(e^{(k-1)}) - \beta q(e^{(k-2)}) \\ 0 \end{bmatrix}.$$

Now, using Lemma 10 in [2], we obtain the upper bound

$$\left\| \begin{bmatrix} e^{(k)} \\ e^{(k-1)} \end{bmatrix} \right\|_2 \leq \big(\rho(T) + o(1)\big)^{k-1} \left\| \begin{bmatrix} e^{(1)} \\ e^{(0)} \end{bmatrix} \right\|_2$$

where $\rho(T)$ is the spectral radius of $T$ and is equal to the maximum magnitude of any eigenvalue of $T$.

We compute $\rho(T)$ as follows. Since $H$ is a real symmetric in $\mathbb{R}^{d \times d}$, let $H = U \Lambda U^T$ be the eigenvalue decomposition of $H$, where $U$ is a unitary matrix and $\Lambda$ is a diagonal matrix whose entries are the eigenvalues of $H$:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d = \sigma^2.$$

Define the permutation $\pi$ as

$$\pi(j) = \begin{cases} 2j - 1 \text{ if } j \leq d, \\ 2j - 2d \text{ otherwise.} \end{cases}$$

Denote the permutation matrix associated with $\pi$ by $P_\pi$. Then, $T$ can be shown to be similar to a block diagonal matrix

$$T \sim P_\pi \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^T \begin{bmatrix} (1+\beta)(I_d - H) & -\beta(I_d - H) \\ I_d & \mathbf{0} \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} P_\pi^T = \begin{bmatrix} T_1 & 0 & \ldots & 0 \\ 0 & T_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & T_d \end{bmatrix}$$

where each $2 \times 2$ block $T_j$ is of the form

$$\begin{bmatrix} (1+\beta)(1 - \lambda_j) & -\beta(1 - \lambda_j) \\ 1 & 0 \end{bmatrix}$$

for $j = 1, \ldots, mn$. Thus, the eigenvalues of $T$ are also the eigenvalues of all blocks $T_j$. Finding optimal step size $\beta$ is equivalent to solving the following problem

$$\min_\beta \max_r |r| \quad \text{such that } r^2 - (1+\beta)(1-\lambda_j)r + \beta(1-\lambda_j) = 0, \quad \text{for some } j \in \{1, \ldots, d\}.$$

Since $H$ is a semi-unitary matrix, we have $\lambda_j \leq 1$ for all $j$. Each quadratic equation has three cases:

1. If $\Delta = (1-\lambda_j)\left((1-\lambda_j)(1+\beta)^2 - 4\beta\right) = 0$ or $\beta = \beta_j^* = \frac{1-\sqrt{\lambda_j}}{1+\sqrt{\lambda_j}}$, then there are two real repeat roots $r_{j1} = r_{j2} = \sqrt{\beta(1-\lambda_j)}$.
2. If $\Delta > 0$ or $\beta < \beta_j^*$, then there are two real distinct roots $r_{j1}, r_{j2}$. The convergence rate depends on $\max\{|r_{j1}|, |r_{j2}|\}$, which is greater than $\sqrt{|\beta(1-\lambda_j)|}$.
3. If $\Delta < 0$ or $\beta > \beta_j^*$, then there are two conjugate complex roots satisfying $|r_{j2}| = |r_{j2}| = \sqrt{\beta(1-\lambda_j)}$.

In any case, we have $\rho(T) = \max_j |r_j| \geq \sqrt{|\beta(1-\lambda_d)|}$. The equality holds when setting $\beta = \frac{1-\sqrt{\lambda_d}}{1+\sqrt{\lambda_d}}$.

# References


[1] F. Li and R.-J. Vaccaro, "Unified analysis for DOA estimation algorithms in array signal processing," *Signal Processing*, vol. 22, pp. 147–169, 1991.

[2] B. Polyak, "Some methods of speeding up the convergence of iteration methods," in *Ussr Computational Mathematics and Mathematical Physics*, 1964, vol. 4, pp. 1–17.