APPENDIX
PROOF FOR LEMMA 1

In this appendix, we present our proof of Lemma 1, the key lemma in the proof of convergence of Algorithm 2. Before our formal proof, we firstly introduce some notations and preliminaries for matrix inequalities, which play an important role in our proof for Lemma 1.

Let A be an $n \times n$ matrix. The vector of eigenvalues of A is denoted by $\lambda(A) = (\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A))$, and they are ordered as $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A) \geq 0$. The vector of diagonal elements of A is denoted by $d(A) = (d_1(A), d_2(A), \ldots, d_n(A))$. For any A, there exists the singular value decomposition. The singular values are arranged in decreasing order and denoted by $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_n(A) \geq 0$.

It is clear that the singular values of A are the nonnegative square roots of the eigenvalues of the positive semidefinite matrix $A^\top A$, or equivalently, they are the eigenvalues of the positive semidefinite square root $(A^\top A)^{1/2}$, so that $\sigma_i(A) = [\lambda_i(A^\top A)]^{1/2} = \lambda_i[(A^\top A)^{1/2}], \quad (i = 1, \ldots, n)$.

We then introduce the theory of majorization, one of the most powerful techniques for deriving inequalities. Given a real vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we rearrange its components as $x_{[1]} \geq x_{[2]} \geq \ldots \geq x_{[n]}$. The definition of majorization is as follows. For $x, y \in \mathbb{R}^n$, if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \text{ for } k = 1, \ldots, n-1$$

and

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]},$$

then we say that x is *majorized* by y and denote $x \prec y$. If

$$\sum_{i=1}^{n} x_{[i]} \leq \sum_{i=1}^{n} y_{[i]},$$

we say that x is *weakly majorized* by y and denote $x \prec_w y$. We introduce some properties for majorization and weak majorization, which can be found in quite a wide range of literature, e.g. [34]. Interested readers can find detailed proofs in these references.

Lemma 2 (cf. [34], Ch. 1)**.** *Let $g(t)$ be an increasing and convex function. Let $g(x) := (g(x_1), g(x_2), \ldots, g(x_n))$ and $g(y) := (g(y_1), g(y_2), \ldots, g(y_n))$. Then, $x \prec_w y$ implies $g(x) \prec g(y)$.*

Theorem 3 (cf. [34], Ch. 9)**.** *If A is a Hermitian matrix (real symmetric for a real matrix A), then we have $d(A) \prec \lambda(A)$.*

Note that the singular values of A are the eigenvalues of the positive semidefinite matrix $A^\top A$. We then have:

Corollary 4. *If A is a real symmetric matrix, and we denote $|A|$ as the positive semidefinite square root of $A^\top A$, we have $d(|A|) \prec \lambda(|A|) = \sigma(A)$.*

Lemma 3 (cf. [34], Ch. 9)**.** *For any matrices A and B, we have $\sigma(AB) \prec_w \sigma(A) \circ \sigma(B)$, where $\circ$ denotes the Hadamard product (or entry-wise product).*

Lemma 4 (Abel's Lemma)**.** *For two sequences of real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, we have*

$$\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n-1} (a_i - a_{i+1}) \left(\sum_{j=1}^{i} b_i \right) + a_n \sum_{i=1}^{n} b_i.$$

Lemma 5. *If $x \prec y$ and $w = (w_1, w_2, \ldots, w_n)$, where $0 \leq w_1 \leq w_2 \leq \ldots \leq w_n$, we have*

$$\sum_{i=1}^{n} w_i x_i \geq \sum_{i=1}^{n} w_i y_i.$$

Proof. For any $1 \leq k < n$, we have

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i.$$

Then, since $w_k \leq w_{k+1}$, we have

$$(w_k - w_{k+1}) \sum_{i=1}^{k} x_i \geq (w_k - w_{k+1}) \sum_{i=1}^{k} y_i$$

In addition, for $k = n$, we have

$$w_n \sum_{i=1}^{n} x_i = w_n \sum_{i=1}^{n} y_i,$$

since $x \prec y$ implies that the summation of x_i and y_i are identical. Summing up all n inequalities above, we have

$$\begin{aligned} & \sum_{k=1}^{n-1} (w_k - w_{k+1}) \left(\sum_{i=1}^{k} x_i \right) + w_n \sum_{i=1}^{n} x_i \\ \geq \quad & \sum_{k=1}^{n-1} (w_k - w_{k+1}) \left(\sum_{i=1}^{k} y_i \right) + w_n \sum_{i=1}^{n} y_i. \end{aligned} \quad (18)$$

By applying the Abel's Lemma for both sides, we have

$$\sum_{k=1}^{n} w_k x_k \geq \sum_{k=1}^{n} w_k y_k,$$

which proves the lemma. ☐

Theorem 5 (cf. [34], Ch. 9)**.** *If A and B are two positive semidefinite matrices, then*

$$\mathrm{tr}(AB)^\alpha \quad \leq \quad \mathrm{tr}(A^\alpha B^\alpha), \quad \alpha > 1, \quad (19)$$

$$\mathrm{tr}(AB)^\alpha \quad \geq \quad \mathrm{tr}(A^\alpha B^\alpha), \quad 0 < \alpha \leq 1. \quad (20)$$

Now we are ready to prove our Lemma 1 as follows:

Proof of Lemma 1. The right inequality is a consequence of Lemma 3. To see this, let $g(t) = t^p$ for all $1 \leq p \leq 2$. For all $t \geq 0$, $g(t)$ is an increasing and convex funciton. Thus, we have $\sigma(AB) \prec_w \sigma(A) \circ \sigma(B)$ from Lemma 3, and from Lemma 2 we have $g(\sigma(AB)) \prec_w g(\sigma(A) \circ \sigma(B))$ and it implies that

$$\sum_{i=1}^{n} \sigma_i^p(AB) \leq \sum_{i=1}^{n} \sigma_i^p(A) \sigma_i^p(B).$$

So now we can focus on the left inequality. Here we denote $|A|$ as the positive semidefinite square root of $A^\top A$. Suppose the singular value decomposition of A is $U_A\Sigma_A V_A^\top$, and that of B is $U_B\Sigma_B V_B^\top$. By the unitary invariance of the singular values and the Schatten-p norm, we have

$$\begin{aligned}||AB||_p^p &= ||U_A\Sigma_A V_A^\top U_B\Sigma_B V_B^\top||_p^p \quad (21)\\ &= ||(\Sigma_A V_A^\top U_B)\Sigma_B||_p^p \quad (22)\\ &= ||A_1B_1||_p^p. \quad (23)\end{aligned}$$

Here we let $A_1 := \Sigma_A V_A^\top U_B$ and $B_1 := \Sigma_B$. Thus, without loss of generality, we can assume that B is diagonal. Then, from the definition of Schatten-p norm, we have

$$\begin{aligned}||AB||_p^p &= \operatorname{tr}(|AB|^p) = \operatorname{tr}\left(\sqrt{B^\top A^\top AB}\right)^p\\ &\geq \operatorname{tr}((B^\top)^{\frac{p}{2}}(A^\top A)^{\frac{p}{2}}B^{\frac{p}{2}}) \quad (24)\\ &= \operatorname{tr}((BB^\top)^{\frac{p}{2}}(A^\top A)^{\frac{p}{2}})\\ &= \operatorname{tr}(|B|^p(A^\top A)^{\frac{p}{2}})\\ &= \operatorname{tr}(|B|^p|A|^p) \quad (25)\end{aligned}$$

Here (24) is from (20) in Theorem 5, since $|B|$ is diagonal with all nonnegative entries, and $A^\top A$ is a real symmetric matrix. Since B is diagonal, $d(|B|)$ is just a permutation of its singular value vector $\sigma(B)$. Thus, we can simply rearrange the order of sum in (25) as

$$\begin{aligned}\operatorname{tr}(|B|^p|A|^p) &= \sum_{i=1}^{n} d_i(|B|)d_i(|A|) \quad (26)\\ &= \sum_{i=1}^{n} d_{[n-i+1]}(|B|)d_{\pi(i)}(|A|), \quad (27)\end{aligned}$$

where $\pi(\cdot)$ is a permutation indicating the order of the new summation, and $d_{[i]}(|B|) = \sigma_i(B)$. From Lemma 4, we can see that $d(|A|) \prec \sigma(A)$, and by Lemma 5, we finally have

$$\begin{aligned}||AB||_p^p &= \sum_{i=1}^{n} d^p_{[n-i+1]}(|B|)d^p_{\pi(i)}(|A|) \quad (28)\\ &= \sum_{i=1}^{n} \sigma^p_{n-i+1}(B)d^p_{\pi(i)}(|A|) \quad (29)\\ &\geq \sum_{i=1}^{n} \sigma^p_{n-i+1}(B)\sigma^p_i(A). \quad (30)\end{aligned}$$

□