Supplementary Material

Han Lu,[1] **Quanxue Gao,** [1*] **Qianqian Wang,** [1] **Ming Yang,** [2] **Wei Xia** [1]

[1] School of Telecommunications Engineering, Xidian University, Xi'an 710071, P.R.China.
[2] Mathematics department of the University of Evansville, Evansville, IN 47722 USA
luhan0@foxmail.com, {qxgao, qqwang}@xidian.edu.cn, {yangmingmath, xd.weixia}@gmail.com

Convergence Analysis of Algorithm 1 (see Algorithm 1 in our submisson)

Lemma 1 (Proposition 6.2 of (Lewis and Sendov 2005)) *Suppose $F : \mathbb{R}^{n_1\times n_2} \rightarrow \mathbb{R}$ is represented as $F(X) = f \circ \sigma(X)$, where $X \in \mathbb{R}^{n_1\times n_2}$ with SVD $X = U\mathrm{diag}(\sigma_1, \ldots, \sigma_n)V^{\mathrm{T}}$, $n = \min(n_1, n_2)$, and f is differentiable. The gradient of $F(X)$ at X is*

$$\frac{\partial F(X)}{\partial X} = U\mathrm{diag}(\theta)V^{\mathrm{T}}, \tag{1}$$

where $\theta = \frac{\partial f(y)}{\partial y}|_{y=\sigma(X)}$.

Theorem 1 *[Convergence Analysis of **Algorithm 1**] Let $P_k = \{\boldsymbol{\mathcal{Y}}_k, \boldsymbol{\mathcal{J}}_k, \boldsymbol{\mathcal{Q}}_k\}$, $1 \le k < \infty$ in (6) (see (6) in our submisson) be a sequence generated by **Algorithm 1**, then*

1. *P_k is bounded;*
2. *Any accumulation point of P_k is a stationary KKT point of (6) (see (6) in our submisson).*

Proof of the 1st part To minimize $\mathcal{J}$ at step $k+1$ in (17) (see (17) in our submisson), the optimal $\mathcal{J}_{k+1}$ needs to satisfy the first-order optimal condition $\lambda\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}_{k+1}\|^p_{\circledS p} + \mu_k(\boldsymbol{\mathcal{J}}_{k+1} - \boldsymbol{\mathcal{Y}}_{k+1} - \frac{1}{\mu_k}\boldsymbol{\mathcal{Q}}_k) = 0$.

Recall that when $0 < p < 1$, in order to overcome the singularity of $(|\eta|^p)' = p\eta/|\eta|^{2-p}$ near $\eta = 0$, we consider for $0 < \epsilon \ll 1$ the approximation

$$\partial|\eta|^p \approx \frac{p\eta}{\max\{\epsilon^{2-p}, |\eta|^{2-p}\}}.$$

Letting $\overline{\boldsymbol{\mathcal{J}}}^{(i)} = \overline{\boldsymbol{\mathcal{U}}}^{(i)}\mathrm{diag}\left(\sigma_j(\overline{\boldsymbol{\mathcal{J}}}^{(i)})\right)\overline{\boldsymbol{\mathcal{V}}}^{(i)\mathrm{H}}$, then it follows from Defn. 1 (see Defn. 1 in our submission) and Lemma 1 that

$$\frac{\partial\|\overline{\boldsymbol{\mathcal{J}}}^{(i)}\|^p_{\circledS p}}{\partial\overline{\boldsymbol{\mathcal{J}}}^{(i)}} = \overline{\boldsymbol{\mathcal{U}}}^{(i)}\mathrm{diag}\left(\frac{p\sigma_j(\overline{\boldsymbol{\mathcal{J}}}^{(i)})}{\max\{\epsilon^{2-p}, |\sigma_j(\overline{\boldsymbol{\mathcal{J}}}^{(i)})|^{2-p}\}}\right)\overline{\boldsymbol{\mathcal{V}}}^{(i)\mathrm{H}}.$$

And then one can obtain

$$\frac{p\sigma_j(\overline{\boldsymbol{\mathcal{J}}}^{(i)})}{\max\{\epsilon^{2-p}, |\sigma_j(\overline{\boldsymbol{\mathcal{J}}}^{(i)})|^{2-p}\}} \le \frac{p}{\epsilon^{1-p}}$$
$$\Longrightarrow \left\|\frac{\partial\|\overline{\boldsymbol{\mathcal{J}}}^{(i)}\|^p_{\circledS p}}{\partial\overline{\boldsymbol{\mathcal{J}}}^{(i)}}\right\|_F^2 \le \sum_{i=1}^{N}\frac{p^2}{\epsilon^{2(1-p)}}.$$

So $\frac{\partial\|\overline{\boldsymbol{\mathcal{J}}}\|^p_{\circledS p}}{\partial\overline{\boldsymbol{\mathcal{J}}}}$ is bounded.

Let us denote $\widetilde{\mathbf{F}}_V = \frac{1}{\sqrt{V}}\mathbf{F}_V$, $\mathbf{F}_V$ is the discrete Fourier transform matrix of size $V \times V$, $\mathbf{F}_V^{\mathrm{H}}$ denotes its conjugate transpose. For $\boldsymbol{\mathcal{J}} = \overline{\boldsymbol{\mathcal{J}}} \times_3 \widetilde{\mathbf{F}}_V$ and using the chain rule in matrix calculus, one can obtain that

$$\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}\|^p_{\circledS p} = \frac{\partial\|\boldsymbol{\mathcal{J}}\|^p_{\circledS p}}{\partial\overline{\boldsymbol{\mathcal{J}}}} \times_3 \widetilde{\mathbf{F}}_V^{\mathrm{H}}$$

is bounded.

And it follows that

$$\boldsymbol{\mathcal{Q}}_{k+1} = \boldsymbol{\mathcal{Q}}_k + \mu_k(\boldsymbol{\mathcal{Y}}_{k+1} - \boldsymbol{\mathcal{J}}_{k+1})$$
$$\Longrightarrow \lambda\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}_{k+1}\|^p_{\circledS p} = \boldsymbol{\mathcal{Q}}_{k+1},$$

$\{\boldsymbol{\mathcal{Q}}_{k+1}\}$ appears to be bounded.

Moreover, by using the updating rule

$$\boldsymbol{\mathcal{Q}}_k = \boldsymbol{\mathcal{Q}}_{k-1} + \mu_{k-1}(\boldsymbol{\mathcal{Y}}_k - \boldsymbol{\mathcal{J}}_k),$$

we can deduce

$$\begin{aligned}&\mathcal{L}_{\mu_k}(\boldsymbol{\mathcal{Y}}_{k+1}, \boldsymbol{\mathcal{J}}_{k+1}, \boldsymbol{\mathcal{Q}}_k) \le \mathcal{L}_{\mu_k}(\boldsymbol{\mathcal{Y}}_k, \boldsymbol{\mathcal{J}}_k, \boldsymbol{\mathcal{Q}}_k)\\ &= \mathcal{L}_{\mu_{k-1}}(\boldsymbol{\mathcal{Y}}_k, \boldsymbol{\mathcal{J}}_k; \boldsymbol{\mathcal{Q}}_{k-1})\\ &+ \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\|\boldsymbol{\mathcal{Q}}_k - \boldsymbol{\mathcal{Q}}_{k-1}\|_F^2 + \frac{\|\boldsymbol{\mathcal{Q}}_k\|_F^2}{2\mu_k} - \frac{\|\boldsymbol{\mathcal{Q}}_{k-1}\|_F^2}{2\mu_{k-1}}.\end{aligned} \tag{2}$$

Thus, summing two sides of (2) from $k = 1$ to n, we have

$$\begin{aligned}&\mathcal{L}_{\mu_n}(\boldsymbol{\mathcal{Y}}_{n+1}, \boldsymbol{\mathcal{J}}_{n+1}, \boldsymbol{\mathcal{Q}}_n) \le \mathcal{L}_{\mu_0}(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{J}}_1, \boldsymbol{\mathcal{Q}}_0)\\ &+\frac{\|\boldsymbol{\mathcal{Q}}_n\|_F^2}{2\mu_n} - \frac{\|\boldsymbol{\mathcal{Q}}_0\|_F^2}{2\mu_0} + \sum_{k=1}^{n}\left(\frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\|\boldsymbol{\mathcal{Q}}_k - \boldsymbol{\mathcal{Q}}_{k-1}\|_F^2\right).\end{aligned} \tag{3}$$

Observe that

$$\sum_{k=1}^{\infty}\frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2} < \infty,$$

*Corresponding author

we have the right-hand side of (3) is finite and thus $\mathcal{L}_{\mu_n}(\boldsymbol{\mathcal{Y}}_{n+1}, \boldsymbol{\mathcal{J}}_{n+1}, \boldsymbol{\mathcal{Q}}_n)$ is bounded. Notice

$$\mathcal{L}_{\kappa_n}(\boldsymbol{\mathcal{Y}}_{n+1}, \boldsymbol{\mathcal{J}}_{n+1}, \boldsymbol{\mathcal{Q}}_n) = \sum_{v=1}^{V} \operatorname{tr}(\mathbf{Y}_{n+1}^{(v)} \mathbf{D}^{(v)} \mathbf{Y}_{n+1}^{(v)\mathrm{T}}) + \lambda \|\boldsymbol{\mathcal{J}}_{n+1}\|_{\circledS p}^{p} + \frac{\mu_n}{2} \|\boldsymbol{\mathcal{Y}}_{n+1} - \boldsymbol{\mathcal{J}}_{n+1} + \frac{\boldsymbol{\mathcal{Q}}_n}{\mu_n}\|_F^2, \quad (4)$$

and each term of (4) is nonnegative, following from the boundedness of $\mathcal{L}_{\mu_n}(\boldsymbol{\mathcal{Y}}_{n+1}, \boldsymbol{\mathcal{J}}_{n+1}, \boldsymbol{\mathcal{Q}}_n)$, we can deduce each term of (4) is bounded. And $\|\boldsymbol{\mathcal{J}}_{n+1}\|_{\circledS p}^{p}$ being bounded implies that all singular values of $\boldsymbol{\mathcal{J}}_{n+1}$ are bounded and hence $\|\boldsymbol{\mathcal{J}}_{n+1}\|_F^2$ (the sum of squares of singular values) is bounded. Therefore, the sequence $\{\boldsymbol{\mathcal{J}}_k\}$ is bounded. Considering the updating rule

$$\boldsymbol{\mathcal{Q}}_k = \boldsymbol{\mathcal{Q}}_{k-1} + \mu_{k-1}(\boldsymbol{\mathcal{Y}}_k - \boldsymbol{\mathcal{J}}_k) \Longrightarrow \boldsymbol{\mathcal{Y}}_k = \frac{\boldsymbol{\mathcal{Q}}_k - \boldsymbol{\mathcal{Q}}_{k-1}}{\mu_{k-1}} + \boldsymbol{\mathcal{J}}_k,$$

since $\boldsymbol{\mathcal{Q}}_k, \boldsymbol{\mathcal{J}}_k$ are bounded, it is clear that $\{\boldsymbol{\mathcal{Y}}_k\}$ is bounded.

Proof of the 2nd part Since each infinite bounded sequence in $\mathbb{R}^n$ has a convergent subsequence (Bolzano-Weierstrass theorem), there exists at least one accumulation point of the sequence P_k. We denote one of the points $P^* = \{\boldsymbol{\mathcal{Y}}^*, \boldsymbol{\mathcal{J}}^*, \boldsymbol{\mathcal{Q}}^*\}$. Without loss of generality, we assume $\{\mathcal{P}_k\}_{k=1}^{+\infty}$ converge to P^*.

Note that from the updating rule for $\boldsymbol{\mathcal{Q}}$, we have

$$\boldsymbol{\mathcal{Q}}_{k+1} = \boldsymbol{\mathcal{Q}}_k + \mu_k(\boldsymbol{\mathcal{Y}}_k - \boldsymbol{\mathcal{J}}_k) \Longrightarrow \boldsymbol{\mathcal{J}}^* = \boldsymbol{\mathcal{Y}}^*.$$

In the $\boldsymbol{\mathcal{J}}$-subproblem, we have

$$\lambda \nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}_{k+1}\|_{\circledS p}^{p} = \boldsymbol{\mathcal{Q}}_{k+1} \Longrightarrow \boldsymbol{\mathcal{Q}}^* = \lambda \nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}^*\|_{\circledS p}^{p}.$$

In (8) (see (8) in our submission), we have

$$\alpha_v^r \partial \operatorname{tr}(\mathbf{Y}_{k+1}^{(v)} \mathbf{D}^{(v)} \mathbf{Y}_{k+1}^{(v)T}) - \mu_k(\mathbf{J}_{k+1}^{(v)} - \mathbf{Y}_{k+1}^{(v)} + \mathbf{Q}_k^{(v)}/\mu_k) = 0.$$

Now by the updating rule $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} + \mu(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{J}})$, we can see

$$\mathbf{Q}_{k+1}^{(v)} = 2\alpha_v^r \mathbf{Y}_{k+1}^{(v)} \mathbf{D}^{(v)} \Longrightarrow \mathbf{Q}^{(v)*} = 2\alpha_v^r \mathbf{Y}^{(v)*} \mathbf{D}^{(v)},$$

Therefore, one can see that the sequences $\boldsymbol{\mathcal{Y}}^*, \boldsymbol{\mathcal{J}}^*, \boldsymbol{\mathcal{Q}}^*$ satisfy the KKT conditions of the Lagrange function (8) (see (8) in our submisson). The proof is completed. It thus can be used to determine the stop conditions for Algorithm 1, which is $\|\boldsymbol{\mathcal{Y}}_k - \boldsymbol{\mathcal{J}}_k\|_F \leq \varepsilon$.

References

Lewis, A. S.; and Sendov, H. S. 2005. Nonsmooth analysis of singular values. Part I: Theory. *Set-Valued Analysis*, 13(3): 213–241.