Pseudocode of Alternating Minimizations for Low-rank Matrix Recovery*

Qiuwei Li[☆] Zhihui Zhu[☆] Gongguo Tang

1 Observation Model

Assume we are given a linear observation of the low-rank ground-truth matrix $\mathbf{M}^\star$:

$$\mathbf{y} = \mathcal{A}(\mathbf{M}^\star) \doteq \begin{bmatrix} \mathbf{a}_1^\top \operatorname{vec}(\mathbf{M}^\star) \\ \mathbf{a}_2^\top \operatorname{vec}(\mathbf{M}^\star) \\ \vdots \\ \mathbf{a}_p^\top \operatorname{vec}(\mathbf{M}^\star) \end{bmatrix} = \mathbf{A} \operatorname{vec}(\mathbf{M}^\star), \tag{1}$$

where $\{\mathbf{a}_i\}_{i\in[p]}$ are p predetermined sampling vectors in $\mathbb{R}^{nm\times 1}$ and $\mathbf{A} \doteq [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_p] \in \mathbb{R}^{nm\times p}$.

Example 1 (Matrix Sensing)**.** *When the p predetermined sampling vectors $\{\mathbf{a}_i\}_{i\in[p]}$ are random vectors (e.g., Gaussian vectors), the linear observation model* (1) *covers the matrix sensing problem.*

Example 2 (Matrix Completion)**.** *When the p predetermined sampling vectors $\{\mathbf{a}_i\}_{i\in[p]}$ are a subset of the vectorized canonical basis vectors in $\mathbb{R}^{n\times m}$, the linear observation model* (1) *covers the matrix completion problem.*

2 Optimization Method

To recover the low-rank ground-truth matrix $\mathbf{M}^\star$ from $\mathbf{y}$, one natural way is solving the following rank-constrained minimization problem:

$$\operatorname*{minimize}_{\mathbf{M}\in\mathbb{R}^{n\times m}} \|\mathbf{y} - \mathcal{A}(\mathbf{M})\|_2^2 + \lambda\|\mathbf{M}\|_* \ \text{ subject to } \ \operatorname{rank}(\mathbf{M}) \le r. \tag{2}$$

Since rank-constraint is not easy to handle, one elegant approach is via the Burer-Monteiro Factorization (BMF) method, which explicitly enforces the rank-constraint by factorizing the low-rank matrix variable $\mathbf{M}$ into two smaller matrices $\mathbf{X}\mathbf{Y}^\top$ with $\mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}$, and then focuses on the new nonconvex formulation

$$\operatorname*{minimize}_{\mathbf{X}\in\mathbb{R}^{n\times r},\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathcal{A}(\mathbf{X}\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\left(\|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2\right) \tag{3}$$

Despite the nonconvexity of problem (3), it is well-known that when the original convex objective function in (2) is well-conditioned, problem (3) has a benign landscape where all the second-order stationary points are globally optimal. Therefore, even simple iterative algorithms (e.g., gradient descent, stochastic gradient descent, and now (proximal) alternating minimizations) can used to solve the nonconvex problem (3) to global optimality.

*This short note provides pseudocode of alternating minimizations for low-rank matrix recovery, to complement the work "*Qiuwei Li, Zhihui Zhu, and Gongguo Tang. Alternating Minimizations Converge to Second-Order Optimal Solutions. In International Conference on Machine Learning, 2019.*"

3 Pseudocode of of Alternating Minimizations

Algorithm 1 Pseudocode of Alternating Minimization for (3)

1: **Initialization:** $\mathbf{Y}_0 \in \mathbb{R}^{m\times r}$.
2: **Recursion:** For k=0,1,2,3 ..., recursively compute $(\mathbf{X}_k, \mathbf{Y}_k)$ as follows

$$\begin{aligned}
\mathbf{X}_{k+1} &= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathcal{A}(\mathbf{X}\mathbf{Y}_k^\top)\|_2^2 + \frac{\lambda}{2}\|\mathbf{X}\|_F^2;\\
&= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathbf{A}^\top \operatorname{vec}(\mathbf{X}\mathbf{Y}_k^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{X})\|_2^2;\\
&= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathbf{A}^\top (\mathbf{Y}_k\otimes \mathbf{I}_n)\operatorname{vec}(\mathbf{X})\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{X})\|_2^2;\\
&= \operatorname{Mat}\left(\left(2\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big]^\top\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big] + \lambda\mathbf{I}\right)^{-1}\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big]^\top 2\mathbf{y}\right)\\
\mathbf{Y}_{k+1} &= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathcal{A}(\mathbf{X}_{k+1}\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\mathbf{Y}\|_F^2\\
&= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathbf{A}^\top \operatorname{vec}(\mathbf{X}_{k+1}\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{Y})\|_2^2\\
&= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathbf{A}^\top (\mathbf{I}_m\otimes\mathbf{X}_{k+1})\operatorname{vec}(\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{Y}^\top)\|_2^2\\
&= \operatorname{Mat}\left(\left(2\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big]^\top\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big] + \lambda\mathbf{I}\right)^{-1}\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big]^\top 2\mathbf{y}\right)^\top
\end{aligned}$$

Algorithm 2 Pseudocode of Proximal Alternating Minimization for (3)

1: **Initialization:** $(\mathbf{X}_0, \mathbf{Y}_0) \in \mathbb{R}^{n\times r}\times\mathbb{R}^{m\times r}$.
2: **Input:** $\beta > 8\|\mathbf{A}\|^2 f(\mathbf{X}_0,\mathbf{Y}_0)/\lambda + 4\|\mathbf{A}\|\sqrt{f(\mathbf{X}_0,\mathbf{Y}_0)} + \lambda$, where $f(\mathbf{X},\mathbf{Y})$ is the objective function of (3).
3: **Recursion:** For k=0,1,2,3 ..., recursively compute $(\mathbf{X}_k, \mathbf{Y}_k)$ as follows

$$\begin{aligned}
\mathbf{X}_{k+1} &= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathcal{A}(\mathbf{X}\mathbf{Y}_k^\top)\|_2^2 + \frac{\lambda}{2}\|\mathbf{X}\|_F^2 + \frac{\beta}{2}\|\mathbf{X}-\mathbf{X}_k\|_F^2;\\
&= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathbf{A}^\top \operatorname{vec}(\mathbf{X}\mathbf{Y}_k^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{X})\|_2^2;\\
&= \arg\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \|\mathbf{y} - \mathbf{A}^\top (\mathbf{Y}_k\otimes \mathbf{I}_n)\operatorname{vec}(\mathbf{X})\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{X})\|_2^2 + \frac{\beta}{2}\|\operatorname{vec}(\mathbf{X}) - \operatorname{vec}(\mathbf{X}_k)\|_2^2;\\
&= \operatorname{Mat}\left(\left(2\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big]^\top\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big] + \lambda\mathbf{I} + \beta\mathbf{I}\right)^{-1}\left(\Big[\mathbf{A}(\mathbf{Y}_k\otimes\mathbf{I}_n)\Big]^\top 2\mathbf{y} + \beta\operatorname{vec}(\mathbf{X}_k)\right)\right)\\
\mathbf{Y}_{k+1} &= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathcal{A}(\mathbf{X}_{k+1}\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\mathbf{Y}\|_F^2 + \frac{\beta}{2}\|\mathbf{Y}-\mathbf{Y}_k\|_F^2\\
&= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathbf{A}^\top \operatorname{vec}(\mathbf{X}_{k+1}\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{Y}^\top)\|_2^2 + \frac{\beta}{2}\|\operatorname{vec}(\mathbf{Y}^\top) - \operatorname{vec}(\mathbf{Y}_k^\top)\|_2^2\\
&= \arg\min_{\mathbf{Y}\in\mathbb{R}^{m\times r}} \|\mathbf{y} - \mathbf{A}^\top (\mathbf{I}_m\otimes\mathbf{X}_{k+1})\operatorname{vec}(\mathbf{Y}^\top)\|_2^2 + \frac{\lambda}{2}\|\operatorname{vec}(\mathbf{Y}^\top)\|_2^2 + \frac{\beta}{2}\|\operatorname{vec}(\mathbf{Y}^\top) - \operatorname{vec}(\mathbf{Y}_k^\top)\|_2^2\\
&= \operatorname{Mat}\left(\left(2\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big]^\top\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big] + \lambda\mathbf{I} + \beta\mathbf{I}\right)^{-1}\left(\Big[\mathbf{A}^\top(\mathbf{I}_m\otimes\mathbf{X}_{k+1})\Big]^\top 2\mathbf{y} + \beta\operatorname{vec}(\mathbf{Y}_k^\top)\right)\right)^\top
\end{aligned}$$