Supplementary Material for "Intrinsic Tensor Sparsity Regularization and Its Applications to Tensor Recvery"

Qi Xie, Qian Zhao, Deyu Meng and Zongben Xu

Abstract—In this supplementary material, we provide the proofs to Theorems 1 presented in the maintext. We also present more clarifications on the parameter settings in our experiment.

Index Terms—proofs to Theorems, parameter settings.

1 Proof to Theorem 1

In this section we propose the proof of Theorem 1.

Theorem 1. *For sequences* $\{\mathcal{S}^{(l)}\}$, $\{\mathcal{M}_k^{(l)}\}$ *and* $\{U_k^{(l)}\}$, $k = 1, 2, \cdots, N$, *generated by Algorithm 1 and* $\mathcal{X}^{(l)} = \mathcal{S}^{(l)} \times_1 U_1^{(l)} \times_2 \cdots \times_N U_N^{(l)}$, $\{\mathcal{M}_k^{(l)}\}$ *and* $\{\mathcal{X}^{(l)}\}$ *satisfy:*

$$\begin{aligned} \left\|\mathcal{X}^{(l)} - \mathcal{M}_k^{(l)}\right\|_F &\le O\left(\mu^{(0)}\rho^{-l/2}\right), \\ \left\|\mathcal{X}^{(l+1)} - \mathcal{X}^{(l)}\right\|_F &\le O\left(\mu^{(0)}\rho^{-l/2}\right). \end{aligned} \tag{1}$$

Proof. We first proof that when $\lim_{l\to\infty} a^{(l)} = 0$,

$$|x - \mathrm{D}_{a^{(l)},\varepsilon}(x)| \le O((a^{(l)})^{\frac{1}{2}}). \tag{2}$$

Based on the definition of $D_{a^{(l)},\varepsilon}$ defined by (**??**), we can obtain that if $|x| < 2\sqrt{a^{(l)}} - \varepsilon$ we have

$$|x - \mathrm{D}_{a^{(l)},\varepsilon}(x)| = |x| < 2\sqrt{a^{(l)}} - \varepsilon \le O((a^{(l)})^{\frac{1}{2}}), \tag{3}$$

and if $|x| \ge 2\sqrt{a^{(l)}} - \varepsilon$, we have

$$\begin{aligned} |x - \mathrm{D}_{a^{(l)},\varepsilon}(x)| &= \frac{|x| + \varepsilon - \sqrt{(|x|+\varepsilon)^2 - 4a^{(l)}}}{2} \\ &= a^{(l)} + O((a^{(l)})^2) \\ &= O((a^{(l)})^1) \\ &\le O((a^{(l)})^{\frac{1}{2}}). \end{aligned}$$

This prove (2).

For $\forall k = 1,\cdots,N$, denote by $V_1\Sigma V_2^T$ as the SVD of $\mathrm{unfold}_k\left(\mathcal{X}^{(l)} + \frac{1}{\mu^{(l)}}\mathcal{P}_k^{(l)}\right)$ in the l^{th} iteration. Based on the updating equation in Algorithm 1, we have

$$\mathcal{M}_k^{(l)} = \mathrm{fold}_k\left(V_1\Lambda_k V_2^T\right),$$

where $\Lambda = \mathrm{diag}\left(\mathrm{D}_{a_k,\varepsilon}(\sigma_1),\cdots,\mathrm{D}_{a_k,\varepsilon}(\sigma_n)\right)$, σ_i is the i^{th} diagonal element of Σ and

$$a_k = \left(\frac{\lambda}{\mu}\prod_{j\neq k} P_{ls}^*\left(M_{j(j)^{(l)}}\right)\right) = O\left(\left(\mu^{(l)}\right)^{-1}\right), \tag{4}$$

where $\mu^{(l)} \to \infty$. Thus, combining the multiplier updating rule of ADMM and (2), we can obtain:

$$\begin{aligned} &\left\|\mathcal{P}_k^{(l+1)}\right\|_F = \left\|\mathcal{P}_k^{(l)} + \mu^{(l)}(\mathcal{X}^{(l)} - \mathcal{M}_k^{(l)})\right\|_F \\ &= \mu^{(l)}\left\|V_1(\Sigma - \Lambda)V_2^T\right\|_F \\ &= \mu^{(l)}\left\|\mathrm{diag}\left(\sigma_1 - \mathrm{D}_{a_k,\varepsilon}(\sigma_1),\cdots,\sigma_n - \mathrm{D}_{a_k,\varepsilon}(\sigma_n)\right)\right\|_F \\ &\le \mu^{(l)} O((a_k)^{\frac{1}{2}}) \\ &= O\left(\left(\mu^{(l)}\right)^{\frac{1}{2}}\right). \end{aligned} \tag{5}$$

By using (5), we can obtain

$$\begin{aligned} &\left\|\mathcal{X}^{(l)} - \mathcal{M}_k^{(l)}\right\|_F = \frac{1}{\mu^{(l)}}\left\|\mathcal{P}_k^{(l+1)} - \mathcal{P}_k^{(l)}\right\|_F \\ &\le \frac{1}{\mu^{(l)}}\left(\left\|\mathcal{P}_k^{(l+1)}\right\|_F + \left\|\mathcal{P}_k^{(l)}\right\|_F\right) \\ &= O\left(\left(\mu^{(l)}\right)^{-\frac{1}{2}}\right). \end{aligned} \tag{6}$$

This proves the first term in (1) in Theorem 3.

Denote $\mathcal{O}^{(l)} = \frac{\beta\mathcal{Y} + \sum_j(\mu^{(l)}\mathcal{M}_j^{(l)} - \mathcal{P}_j^{(l)})}{\beta + N\mu^{(l)}}$, and then we have

$$\begin{aligned} &\left\|\mathcal{O}^{(l)} - \mathcal{X}^{(l)}\right\|_F = \left\|\frac{\beta\mathcal{Y} + \sum_j(\mu^{(l)}\mathcal{M}_j^{(l)} - \mathcal{P}_j^{(l)})}{\beta + N\mu^{(l)}} - \mathcal{X}^{(l)}\right\|_F \\ &= \left\|\frac{\beta(\mathcal{Y} - \mathcal{X}^{(l)})}{\beta + N\mu^{(l)}}\right\|_F + \left\|\frac{\sum_j(\mu^{(l)}(\mathcal{M}_j^{(l)} - \mathcal{X}^{(l)}) - \mathcal{P}_j^{(l)})}{\beta + N\mu^{(l)}}\right\|_F \\ &\le O\left(\left(\mu^{(l)}\right)^{-1}\right) + \frac{1}{\beta + N\mu^{(l)}}\sum_j\left\|\mathcal{P}_j^{(l+1)}\right\|_F \\ &\le O\left(\left(\mu^{(l)}\right)^{-\frac{1}{2}}\right). \end{aligned} \tag{7}$$

Q.Xie, Q. Zhao, D. Meng and Z.Xu are with School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Shaanxi, P.R.China.
Email: xq.liwu@stu.xjtu.edu.cn, {qzhao, dymeng, zbxu}@mail.xjtu.edu.cn

Denote $\mathcal{Q} = \mathcal{O}^{(l)} \times_1 U_1^{(l)T}, \cdots, \times_N U_N^{(l)T}$, $b^{(l)} = \frac{1}{\beta + N\mu^{(l)}}$, and we have

$$\begin{aligned}
&\left\| \mathcal{S}^{(l+1)} \times_1 U_1^{(l)}, \cdots, \times_N U_N^{(l)} - \mathcal{O}^{(l)} \right\|_F = \left\| \mathcal{S}^{(l+1)} - \mathcal{Q} \right\|_F \\
&= \left(\sum_{i_1,\cdots,i_N} (q_{i_1,\cdots,i_N} - \mathrm{D}_{b^{(l)},\varepsilon}(q_{i_1,\cdots,i_N})^2 \right)^{\frac{1}{2}} \\
&\leq \left(\sum_{i_1,\cdots,i_N} (O((\mu^{(l)})^{-\frac{1}{2}}))^2 \right)^{\frac{1}{2}} \\
&= O\left(\left(\mu^{(l)} \right)^{-\frac{1}{2}} \right).
\end{aligned} \tag{8}$$

$\forall n = 1, \cdots, N$, update $U_n^{(l+1)}$ by solving

$$\min_U \left\| \mathcal{S} \times_1 U_1^{(l+1)}, \cdots, U_{n-1}^{(l+1)} \times_n U \times_{n+1} U_{n+1}^{(l)}, \cdots, U_N^{(l)} - \mathcal{O} \right\|_F^2,$$

and then

$$\begin{aligned}
&\left\| \mathcal{X}^{(l+1)} - \mathcal{O}^{(l)} \right\|_F \\
&= \left\| \mathcal{S}^{(l+1)} \times_1 U_1^{(l+1)}, \cdots, \times_N U_N^{(l+1)} - \mathcal{O}^{(l)} \right\|_F \\
&\leq \left\| \mathcal{S}^{(l+1)} \times_1 U_1^{(l)}, \cdots, \times_N U_N^{(l)} - \mathcal{O}^{(l)} \right\|_F \\
&\leq O\left(\left(\mu^{(l)} \right)^{-\frac{1}{2}} \right).
\end{aligned} \tag{9}$$

Based on (7)and(9), we can obtain

$$\begin{aligned}
\left\| \mathcal{X}^{(l+1)} - \mathcal{X}^{(l)} \right\|_F &\leq \left\| \mathcal{X}^{(l+1)} - \mathcal{O}^{(l)} \right\|_F + \left\| \mathcal{O}^{(l)} - \mathcal{X}^{(l)} \right\|_F \\
&= O\left(\left(\mu^{(l)} \right)^{-\frac{1}{2}} \right).
\end{aligned}$$

This proves the second term (1) in Theorem 3.■

2 PARAMETER SETTING DETAILS

2.1 Discussion on β setting

In our experiments we set $\beta = cv^{-1}$, and in this section, we will provide clarifications on this parameter setting strategy. Let's first consider the following matrix-based proximal problem:

$$\hat{X} = \arg \min_{X \in \mathbb{R}^{m \times n}} \frac{1}{\beta} \log(|d_i| + \varepsilon) + \frac{1}{2} \|X - Y\|_F^2, \tag{10}$$

where d_i is the i-th singular value of X. Based Theorem 2, we can define $\hat{X} = U\hat{D}V^T$ as the SVD of $\hat{X}$, and $Y = U\Sigma V^T$ as the SVD of Y, where $\hat{D} = \mathrm{diag}(\hat{d}_1, \hat{d}_2, ..., \hat{d}_n)$ and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_n)$. Approximately let $\varepsilon = 0$, and then we have

$$\begin{aligned}
&\langle \hat{X}, Y - \hat{X} \rangle = Tr(\hat{X}^T(Y - \hat{X})) \\
&= Tr\left(\left(U\hat{D}V^T \right)^T U \left(\Sigma - \hat{D} \right) V^T \right) = \sum_{i=1}^r \hat{d}_i(\sigma_i - \hat{d}_i) \\
&= \sum_{i=1}^r \left(\frac{\sigma_i + \sqrt{\sigma^2 - 4\beta^{-1}}}{2} \right) \left(\frac{\sigma_i - \sqrt{\sigma^2 - 4\beta^{-1}}}{2} \right) \\
&= \sum_{i=1}^r \frac{\sigma_i^2 - (\sigma_i^2 - 4\beta^{-1})}{4} = r\beta^{-1}.
\end{aligned}$$

Thus, we can obtain $\beta = \frac{r}{\langle \hat{X}, Y - \hat{X} \rangle}$. Since $\frac{\langle \hat{X}, Y - \hat{X} \rangle}{mn}$ is with the same order of magnitude as

$$E\{\hat{X}\}E\{Y - \hat{X}\} = E\{Y\}E\{Y - \hat{X}\} \approx E\{Y\}E\{Y - X\},$$

where X define the groundtruth, and $E\{A\}$ define the mean value of all elements in a matrix A. Since $E\{Y - X\}$ is with the same order of magnitude as $v = \sqrt{E(\{Y - X\}^2)}$, we have $\beta \approx \hat{c}\frac{r}{mnE\{Y\}}v^{-1}$. This derivation can be easily generalized to tensor case in a similar fashion, and thus we can use $\beta = cv^{-1}$ to set β.