Supplements for "Tensor denoising and completion based on ordinal observations"

1 Additional results

1.1 Details of simulation

Here, the numerical rank is computed as the minimal rank for which the relative least-squares error is below 0.1, and $\mathcal{Z}$ is a rank-3 tensor with i.i.d. $N(0,1)$ entries in the (unnormalized) singular vectors. Reported ranks are averaged across 10 replicates of $\mathcal{Z}$, with standard errors given in error bars. Numerical values in both figures are obtained by running CP decomposition with random initialization.

Example 1 (Banded matrix)**.** Let $\boldsymbol{a} = (2^{-1}, 2^{-2}, \ldots, 2^{-d})^T \in \mathbb{R}^d$ and $\text{rev}(\boldsymbol{a}) = (2^{-d}, \ldots, 2^{-1})^T \in \mathbb{R}^d$. We consider the following matrix

$$\boldsymbol{M} = \boldsymbol{a} \otimes \boldsymbol{a} + \text{rev}(\boldsymbol{a}) \otimes \text{rev}(\boldsymbol{a}).$$

It is easy to see that the $\boldsymbol{M}$ is both symmetric and skew symmetric,

$$\boldsymbol{M}(i,j) = \boldsymbol{M}(j,i) = \boldsymbol{M}(d-i, d-j) = \boldsymbol{M}(d-j, d-i) = 2^{-d-1}\left(2^{j-i} + 2^{i-j}\right), \text{ for all } (i,j) \in [d]^2.$$

Furthermore, $\text{rank}(\boldsymbol{M}) = 2$, and $\text{sgn}(\boldsymbol{M} - \pi)$ is a banded matrix.

Example 2 (Addition examples that satisfying Proposition 2)**.** We provide a tensor example with $\text{rank}(\Theta) = d$ but $\text{srank}(\Theta) = 3$. Define $\Theta = \sum_{r=1}^d \boldsymbol{e}_r^{\otimes 2} \otimes \mathbf{1}_d^{\otimes (K-2)}$, where $\boldsymbol{e}_r = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is the r-th canonical basis in $\mathbb{R}^d$, and $\mathbf{1}_d \in \mathbb{R}^d$ is a vector with all entries 1. Based on the definition of Θ, we have

$$\text{rank}(\Theta) = \text{rank}(\boldsymbol{I}), \quad \text{srank}(\Theta) = \text{srank}(\boldsymbol{I}),$$

where $\boldsymbol{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. Therefore, it suffices to show that $\text{srank}(\boldsymbol{I}) = 3$. We now construct a rank-2 matrix $\boldsymbol{A}$ such that $\text{sgn}(\boldsymbol{A} - 1/2) = \text{sgn}(\boldsymbol{I})$. Define

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\frac{1}{2}\times 1 \\ 2^{-1} & -\frac{1}{2}\times 4^{-1} \\ \vdots & \vdots \\ 2^{-d+1} & -\frac{1}{2}\times 4^{-d+1} \end{bmatrix} \begin{bmatrix} 1 & 2 & \cdots & 2^{d-1} \\ 1 & 4 & \cdots & 4^{d-1} \end{bmatrix}.$$

It is easy to verify that $\boldsymbol{A}(i,j) = \frac{1}{2}$ if $i = j$, and $\boldsymbol{A}(i,j) < \frac{1}{2}$ otherwise. Therefore, $\text{sgn}(\boldsymbol{A} - 1/2) = \boldsymbol{I}$.

2 Additional numerical results

3 Proofs

Proof of Proposition 3. Based on the definition, the function $\text{Risk}(\cdot)$ relies only on the sign pattern of the tensor. Therefore, without loss of generality, we assume both tensors $\bar{\Theta}, \mathcal{Z} \in \{-1,1\}^{d_1\times\cdots\times d_K}$ are binary tensors. We evaluate the excess risk

$$\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta}) = \mathbb{E}_{\omega\sim\Pi} \underbrace{\mathbb{E}_{\mathcal{Y}(\omega)}\left\{|\mathcal{Y}(\omega) - \pi|\left[\left|\mathcal{Z}(\omega) - \text{sgn}(\bar{\mathcal{Y}}(\omega))\right| - \left|\bar{\Theta}(\omega) - \text{sgn}(\bar{\mathcal{Y}}(\omega))\right|\right]\right\}}_{=:I(\omega)}. \quad (1)$$

Denote $y = \mathcal{Y}(\omega)$, $z = \mathcal{Z}(\omega)$, $\bar{\theta} = \bar{\Theta}(\omega)$, and $\theta = \Theta(\omega)$. It follows from the expression of $I(\omega)$ that

$$\begin{aligned} I(\omega) &= \mathbb{E}_y\left[(y-\pi)(\bar{\theta}-z)\mathbb{1}(y \geq \pi) + (\pi - y)(z-\bar{\theta})\mathbb{1}(y<\pi)\right] \\ &= \mathbb{E}_y\left[(\bar{\theta}-z)(y-\pi)\right] \\ &= \left[\text{sgn}(\theta-\pi) - z\right](\theta-\pi) \\ &= |\text{sgn}(\theta-\pi)-z||\theta-\pi| \geq 0 \end{aligned} \tag{2}$$

where the third line uses the fact $\mathbb{E}y = \theta$ and $\bar{\theta} = \text{sgn}(\theta-\pi)$, and the last line uses the assumption $z \in \{-1, 1\}$. In particular, the equality is attained when $z = \text{sgn}(\theta - \pi)$ or $\theta = \pi$. Combining (2) with (1), we conclude

$$\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta}) = \mathbb{E}_{\omega\sim\Pi}|\text{sgn}(\Theta(\omega)-\pi) - \mathcal{Z}(\omega)||\Theta(\omega)-\pi| \geq 0,$$

for all $\mathcal{Z} \in \{-1,1\}^{d_1\times\cdots\times d_K}$. Therefore,

$$\text{Risk}(\bar{\Theta}) = \min\{\text{Risk}(\mathcal{Z})\colon \mathcal{Z} \in \mathbb{R}^{d_1\times\cdots\times d_K}\} \leq \min\{\text{Risk}(\mathcal{Z})\colon \text{rank}(\mathcal{Z}) \leq r\}.$$

Because $\text{srank}(\bar{\Theta}) \leq r$ by assumption, the last inequality becomes equality. The proof is complete. □

Proof. We verify two conditions.

1. Approximation error. For $\mathcal{Z}$ with $\text{rank}(\mathcal{Z}) \leq r$, we have $\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta}) = 0$ for all d.

2. Variance-to-mean relationship

$$\text{Var}_{\mathcal{Y},\Omega}[L(\mathcal{Z},\bar{\mathcal{Y}}_\pi) - L(\bar{\Theta},\mathcal{Y}_\pi)] \leq [\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta})]^{\alpha/(1+\alpha)} + \frac{1}{\rho(\pi,\mathcal{N})}[\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta})].$$

Apply Lemma 1 to the above condition, we obtain

$$\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta}) \leq t_n^{(\alpha+1)/(\alpha+2)} + \frac{1}{\rho(\pi,\mathcal{N})}t_n, \quad \text{where } t_n = \frac{Krd}{n}.$$

□

Lemma 1. *Because the classification rate is scale-free; Risk*$(\mathcal{Z})$ = *Risk*$(c\mathcal{Z})$ *for every* $c > 0$. *Therefore, without loss of generality, we solve the estimate subject to* $\|\mathcal{Z}\|_F \leq 1$,

$$\hat{\mathcal{Z}} = \mathop{\arg\min}_{\mathcal{Z}\colon \text{rank}(\mathcal{Z})\leq r, \|\mathcal{Z}\|_F\leq 1} L(\mathcal{Z},\bar{\mathcal{Y}}_\pi).$$

Write $|\Omega| = n$. *We have*

$$\mathbb{P}[\text{Risk}(\hat{\mathcal{Z}}) - \text{Risk}(\bar{\Theta}) \geq t_n] \leq \frac{7}{2}\exp(-Cnt_n).$$

The rate of convergence $t_n > 0$ *is determined by the solution to the following inequality,*

$$\frac{1}{t_n}\int_{t_n}^{\sqrt{t_n^\alpha+\rho^{-1}t_n}} \sqrt{\mathcal{H}_{[\ ]}(\varepsilon,\ \mathcal{F},\ \|\cdot\|_2)}d\varepsilon \leq n^{1/2},$$

where $\mathcal{F} = \{\mathcal{Z} \colon \operatorname{rank}(\mathcal{Z}) \leq r,\ \|\mathcal{Z}\|_F^2 \leq 1\}$ *and* $\rho = \rho(\pi, \mathcal{N})$*. By Lemma 2, we obtain*

$$t_n \asymp \left(\frac{Kdr}{n}\right)^{(\alpha+1)/(\alpha+2)} + \frac{1}{\rho^2(\pi, \mathcal{N})} \frac{Kdr}{n}.$$

Finally, we obtain

$$\mathbb{P}[\operatorname{Risk}(\hat{\mathcal{Z}}) - \operatorname{Risk}(\bar{\Theta}) \geq t_n] \leq \frac{7}{2} \exp(-Cd^{\frac{\alpha+1}{\alpha+2}} n^{\frac{1}{\alpha+2}}) \leq \frac{7}{2} \exp(-C\sqrt{d}),$$

where $C = C(k, r) > 0$ *is a constant independent of* d *and* n*.*

Lemma 2 (Bracketing number for bounded low rank tensor)**.**

$$\sqrt{\mathbb{E}_{\omega \sim \Pi} |\mathcal{Z}_1(\omega) - \mathcal{Z}_2(\omega)|^2} \leq \|\mathcal{Z}_1 - \mathcal{Z}_2\|_\infty \leq \|\mathcal{Z}_1 - \mathcal{Z}_2\|_F.$$

Therefore

$$\mathcal{H}_{[\ ]}(2\varepsilon, \mathcal{F}, \|\cdot\|_2) \leq \mathcal{H}(\varepsilon, \mathcal{F}, \|\cdot\|_F) \leq C(1 + Kdr) \log \frac{d}{\varepsilon},$$

where the covering number for low rank tensor is based on **??***.*