Robust Low-Rank Tensor Minimization via a New Tensor Spectral k-Support Norm: Supplementary Material

Jian Lou and Yiu-ming Cheung, *Fellow, IEEE*

Appendix A
Proofs for Section III

A1. Proof of Proposition 8

Proof. Define the indicator function of set $\mathcal{C}$ as follows,

$$\mathcal{X}_{\mathcal{C}}(\mathcal{A}) = \begin{cases} 0, & \mathcal{A} \in \mathcal{C} \\ \infty, & \mathcal{A} \notin \mathcal{C}. \end{cases} \tag{1}$$

We prove the Proposition by computing the Fenchel bi-conjugate of the indicator function of set $\mathcal{C}_k^{(sp)}$. Recall the block diagonal matrix $\mathbf{A}_{\mathcal{F}}$. $\mathcal{X}_{\mathcal{C}_k^{(sp)}}(\mathcal{A})$ can be equivalently interpreted as,

$$\begin{aligned} \mathcal{C}_k^{(sp)} = \Big\{ & \mathbf{A}_{\mathcal{F}} : \|\mathbf{A}_{\mathcal{F}}\|_2 = \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_\infty \leq \alpha, \\ & \sum_{i=1}^{n_3} rank(\mathbf{A}_{\mathcal{F}}^{(i)}) = rank(\mathbf{A}_{\mathcal{F}}) = \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_0 \leq k \Big\}. \end{aligned} \tag{2}$$

We first compute the Fenchel conjugate of $\mathcal{X}_{\mathcal{C}_k^{(sp)}}(\mathcal{A})$ by

$$\begin{aligned} \mathcal{X}^*_{\mathcal{C}_k^{(\infty)}}(\mathcal{B}) &= \sup_{\mathcal{A}} \langle \mathcal{A}, \mathcal{B} \rangle - \mathcal{X}_{\mathcal{C}_k^{(sp)}}(\mathcal{A}) \\ &= \sup_{\mathbf{A}_{\mathcal{F}}} \frac{1}{\alpha} \langle \mathbf{A}_{\mathcal{F}}, \mathbf{B}_{\mathcal{F}} \rangle - \mathcal{X}_{\mathcal{C}_k^{(sp)}}(\mathbf{A}_{\mathcal{F}}) \\ &\overset{(i)}{=} \sup_{\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}} \langle \frac{1}{\alpha} \boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}, \boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}} \rangle - \mathcal{X}_{\mathcal{C}_k^{(\infty)}}(\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}) \\ &= \|(\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}})_{[1:k]}\|_1, \end{aligned} \tag{3}$$

where the equality (i) is by Von Neumann's trace inequality with $\mathbf{B}_{\mathcal{F}}$ sharing the same unitary matrices $\mathbf{U}_{\mathcal{F}}$ and $\mathbf{V}_{\mathcal{F}}$ with $\mathbf{A}_{\mathcal{F}}$; the last equality is by the property of the ℓ_1-norm and eq.(2) (please note that eq.(2) also amounts to $\mathcal{C}_k^{(\infty)} = \{\frac{1}{\alpha}\boldsymbol{\sigma}_{\mathcal{A}_{\mathcal{F}}} : \|\frac{1}{\alpha}\boldsymbol{\sigma}_{\mathcal{A}_{\mathcal{F}}}\|_\infty \leq 1, \|\boldsymbol{\sigma}_{\mathcal{A}_{\mathcal{F}}}\|_0 \leq k\}$). Then, the Fenchel bi-conjugate is computed by the Fenchel conjugate of $\mathcal{X}^*_{\mathcal{C}_k^{(sp)}}(\mathcal{B})$ as follows,

$$\begin{aligned} \mathcal{X}^{**}_{\mathcal{C}_k^{(sp)}}(\mathcal{A}) &= \sup_{\mathcal{A}} \langle \mathcal{B}, \mathcal{A} \rangle - \mathcal{X}^*_{\mathcal{C}_k^{(sp)}}(\mathcal{B}) \\ &= \sup_{\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}}} \frac{1}{\alpha} \langle \boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}}, \boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}} \rangle - \|(\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}})_{[1:k]}\|_1 \\ &= \begin{cases} 0, & \|\frac{1}{\alpha}\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_\infty \leq 1, \frac{1}{k}\|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_1 \leq 1 \\ \infty, & \text{otherwise} \end{cases} \\ &= \mathcal{X}_{conv(\mathcal{C}_k^{(sp)})}, \end{aligned} \tag{4}$$

which indicates that the convex envelop of the sum of tubal rank on the α-scaled tensor spectral norm ball is the general α-tensor nuclear norm. In addition, by substituting $\alpha = 1$ or n_3 into the above proof process, one can immediately recover the relaxation of 1-TNN and n_3-TNN. ☐

A2. Proof of Proposition 10

Proof. We prove the Proposition by computing the Fenchel bi-conjugate of the indicator function of the set $\mathcal{C}_k^{(Fro)}$. The set $\mathcal{C}_k^{(Fro)}$ can be equivalently interpreted as,

$$\begin{aligned} \mathcal{C}_k^{(Fro)} = \Big\{ & \mathbf{A}_{\mathcal{F}} : \|\mathbf{A}_{\mathcal{F}}\|_F = \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_2 \leq n_3, \\ & \sum_{i=1}^{n_3} rank(\mathbf{A}_{\mathcal{F}}^{(i)}) = rank(\mathbf{A}_{\mathcal{F}}) = \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_0 \leq k \Big\}. \end{aligned} \tag{5}$$

In particular, we extract the part of the singular values from above (note that $\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}/n_3$ has the same cardinality as $\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}$) and denote the singular vector by

$$\mathcal{C}_k^{(sv)} = \Big\{ \boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}} : \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}/n_3\|_2 \leq 1, \|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}/n_3\|_0 \leq k \Big\}, \tag{6}$$

which amounts to

$$\mathcal{C}_k^{(2)} = \Big\{ \mathbf{v} \in \mathbb{R}^D : \|\mathbf{v}\|_2 \leq 1, \|\mathbf{v}\|_0 \leq k \Big\}. \tag{7}$$

With the above equivalence relationships, the Fenchel conjugate of $\mathcal{X}_{\mathcal{C}_k^{(Fro)}}(\mathcal{A})$ can be computed by

$$\begin{aligned} \mathcal{X}^*_{\mathcal{C}_k^{(Fro)}}(\mathcal{B}) &= \sup_{\mathcal{A}} \langle \mathcal{A}, \mathcal{B} \rangle - \mathcal{X}_{\mathcal{C}_k^{(Fro)}}(\mathcal{A}) \\ &= \sup_{\mathbf{A}_{\mathcal{F}}} \frac{1}{n_3} \langle \mathbf{A}_{\mathcal{F}}, \mathbf{B}_{\mathcal{F}} \rangle - \mathcal{X}_{\mathcal{C}_k^{(Fro)}}(\mathcal{A}) \\ &\overset{(i)}{=} \sup_{\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}} \frac{1}{n_3} \langle \boldsymbol{\sigma}_{\mathcal{A}}, \boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}} \rangle - \mathcal{X}_{\mathcal{C}_k^{(Fro)}}(\mathcal{A}) \\ &= \sup_{\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}} \langle \boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}/n_3, \boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}} \rangle - \mathcal{X}_{\mathcal{C}_k^{(sv)}}(\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}) \\ &= \|(\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}})_{[1:k]}\|_2 = \|\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}}\|^*_{vp,k} = \|\mathbf{B}_{\mathcal{F}}\|^*_{msp,k}, \end{aligned} \tag{8}$$

Jian Lou is with the Department of Computer Science, Emory University, Georgia, GA30322, USA, and with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. Yiu-ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China (e-mail: jian.lou@emory.edu, ymc@comp.hkbu.edu.hk). Yiu-ming Cheung is the Corresponding Author.

where the equality (i) is by Von Neumann's trace inequality with $\mathbf{A}_{\mathcal{F}}$ sharing the same unitary matrices $\mathbf{U}_{\mathcal{F}}$ and $\mathbf{V}_{\mathcal{F}}$ from SVD with $\mathbf{B}_{\mathcal{F}}$. Then, the Fenchel bi-conjugate is computed by the Fenchel conjugate of $\mathcal{X}^*_{\mathcal{C}_k^{(Fro)}}(\boldsymbol{\mathcal{B}})$ as follows,

$$\begin{aligned}\mathcal{X}^{**}_{\mathcal{C}_k^{(Fro)}}(\boldsymbol{\mathcal{A}}) &= \sup_{\boldsymbol{\mathcal{A}}}\langle\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{A}}\rangle - \mathcal{X}^*_{\mathcal{C}_k^{(Fro)}}(\boldsymbol{\mathcal{B}})\\ &= \sup_{\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}}}\frac{1}{n_3}\langle\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}},\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\rangle - \|\boldsymbol{\sigma}_{\mathbf{B}_{\mathcal{F}}}\|_2^* \qquad (9)\\ &= \mathcal{X}_{\|\frac{1}{n_3}\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_{vp,k}\le 1}(\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}),\end{aligned}$$

where the last equality is because the Fenchel conjugate of a norm (i.e. dual norm of the k-support norm) is the indicator function of the unit ball of its dual norm (i.e. the k-support norm). Also, the constraint $\|\frac{1}{n_3}\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_2 \le 1$ (again by applying the property of the k-support norm to vector $\frac{1}{n_3}\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}$) gives $\|\frac{1}{n_3}\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_2 = \frac{\sqrt{n_3}}{n_3}\|\boldsymbol{\mathcal{A}}\|_F \le 1$, which is the $\sqrt{n_3}$-scaled tensor Frobenius norm ball of $\|\boldsymbol{\mathcal{A}}\|_F \le \sqrt{n_3}$. As a result, the TSP-$k$ norm takes the form as

$$\|\boldsymbol{\mathcal{A}}\|_{tsp,k} = \frac{1}{n_3}\|\boldsymbol{\sigma}_{\mathbf{A}_{\mathcal{F}}}\|_{vp,k} = \frac{1}{n_3}\|\mathbf{A}_{\mathcal{F}}\|_{msp,k}. \qquad (10)$$

□

A3. Proof of Proposition 11

Proof. When $k=1$, we have $l=0$, and subsequently we have $\|\boldsymbol{\mathcal{A}}\|_{tsp,1} = \frac{1}{n_3}\|\boldsymbol{\sigma}_{\boldsymbol{\mathcal{A}}_{\mathcal{F}}}\|_1 = \frac{1}{n_3}\sum_{i=1}^{n_3}\|\boldsymbol{\sigma}^{(i)}_{\boldsymbol{\mathcal{A}}_{\mathcal{F}}}\|_1 = \|\boldsymbol{\mathcal{A}}\|_{t*,avg}$; When $k=D$, the dual norm becomes

$$\|\boldsymbol{\mathcal{A}}\|^*_{tsp,k} = \|(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{A}}_{\mathcal{F}}})_{[1:D]}\|_2 = \|\boldsymbol{\sigma}_{\boldsymbol{\mathcal{A}}_{\mathcal{F}}}\|_2 = \|\mathbf{A}_{\mathcal{F}}\|_F = \|\mathbf{A}_{\mathcal{F}}\|^*_{msp,k}, \qquad (11)$$

which indicates the primal norm $\|\boldsymbol{\mathcal{A}}\|_{tsp,k} = \frac{1}{n_3}\|\mathbf{A}_{\mathcal{F}}\|_{msp,k} = \frac{1}{n_3}\|\boldsymbol{\sigma}\|_2 = \frac{1}{n_3}\|\mathbf{A}_{\mathcal{F}}\|_F = \frac{1}{\sqrt{n_3}}\|\boldsymbol{\mathcal{A}}\|_F$. □

Appendix B
Proofs for Section IV

B1. Proof of Proposition 16

Proof. In order to compute $\boldsymbol{\mathcal{L}}^{\#}$:

$$\boldsymbol{\mathcal{L}}^{\#} = \arg\min_{\boldsymbol{\mathcal{L}}}\frac{\beta}{2}(\|\boldsymbol{\mathcal{L}}\|^*_{tsp,k})^2 + \frac{1}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{T}}\|_F^2, \qquad (12)$$

we can first convert the problem to Fourier domain via FFT and then recover the result via IFFT. The equivalent problem after FFT is

$$\boldsymbol{\mathcal{L}}^{\#}_{\mathcal{F}} = \arg\min_{\boldsymbol{\mathcal{L}}_{\mathcal{F}}}\frac{\beta}{2}(\|\boldsymbol{\mathcal{L}}_{\mathcal{F}}\|^*_{msp,k})^2 + \frac{1}{2n_3}\|\boldsymbol{\mathcal{L}}_{\mathcal{F}} - \boldsymbol{\mathcal{T}}_{\mathcal{F}}\|_F^2, \quad (13)$$

With $\boldsymbol{\mathcal{L}}^{\#}_{\mathcal{F}}$ sharing the same unitary matrices $\boldsymbol{\mathcal{U}}_{\mathcal{F}}$ and $\boldsymbol{\mathcal{V}}_{\mathcal{F}}$ with $\boldsymbol{\mathcal{T}}_{\mathcal{F}}$, it suffices to compute the proximal operator of the vector dual k-support norm:

$$\boldsymbol{\sigma}_{\boldsymbol{\mathcal{L}}^{\#}_{\mathcal{F}}} = \underset{\boldsymbol{\sigma}_{\boldsymbol{\mathcal{L}}_{\mathcal{F}}}}{\operatorname{argmax}}\frac{\beta n_3}{2}(\|\boldsymbol{\sigma}_{\boldsymbol{\mathcal{L}}_{\mathcal{F}}}\|^*_{vp,k})^2 + \frac{1}{2}\|\boldsymbol{\sigma}_{\boldsymbol{\mathcal{L}}_{\mathcal{F}}} - \boldsymbol{\sigma}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}}\|_2^2. \quad (14)$$

To obtain $\boldsymbol{\sigma}_{\boldsymbol{\mathcal{L}}^{\#}_{\mathcal{F}}}$, we follow the derivation of the proximal operator the $\frac{1}{2}(\|\cdot\|^*_{vp,k})^2$ in [1] (i.e. the proximal operator of the dual k-support norm in the vector case). Let $[\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}}, \mathtt{idx}] = \mathtt{sort}(\boldsymbol{\sigma}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}}, \text{‘descend’})$. Substituting the form of $\frac{1}{2}(\|\cdot\|^*_{vp,k})^2$ in to Eq.(14), it can then be written as

$$\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{L}}^{\#}_{\mathcal{F}}} = \arg\min_{\mathbf{v}_i \ge \mathbf{v}_{i+1}} h(\mathbf{v}) = \mathbf{v}^\top\mathbf{G}^{\frac{1}{\beta n_3}}\mathbf{v} - \frac{2}{\beta n_3}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})^\top\mathbf{v}, \qquad (15)$$

where $\mathbf{G}^{\frac{1}{\beta n_3}}_{ij} = \begin{cases}1+\frac{1}{\beta n_3}, & i=j\le k\\ \frac{1}{\beta n_3}, & i=j>k\\ 0, & i\ne j\end{cases}$. Apparently, for large enough $(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i$, $\mathbf{v}_i = \frac{\frac{1}{\beta n_3}}{1+\frac{1}{\beta n_3}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i$ and for small enough $(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i$, $\mathbf{v}_i = (\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i$, where the "large/small enough" means we need not worry $\frac{\frac{1}{\beta n_3}}{1+\frac{1}{\beta n_3}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i < (\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j$, for $i<j$. However, for intermediate $i<j$ around k, there is possibility that $\frac{\frac{1}{\beta n_3}}{1+\frac{1}{\beta n_3}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_i < (\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j$, which breaks the non-increasing constraint of "$\mathbf{v}_i \ge \mathbf{v}_j$, for $i<j$". To avoid this, we find an interval $[k^{low}, k^{upp}] = \mathbb{I}^*$ around k, and $\mathbf{v}_i$ for $i\in\mathbb{I}^*$ will take the same value: $\mathbf{v}_i = \frac{\frac{1}{\beta n_3}\sum_{j\in\mathbb{I}^*}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j}{\sum_{j\in\mathbb{I}^*}(\mathbf{G}^{\frac{1}{\beta n_3}})_{jj}} = \frac{\sum_{j=k^{low}}^{k^{upp}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j}{(1+\frac{1}{\beta n_3})(k-k^{low}+1)+\frac{1}{\beta n_3}(k^{upp}-k)}$. To find the interval $[k^{low}, k^{upp}] = \mathbb{I}^*$, we follow [1] to repeatedly use two binary searches from the intervals of $[1,k]$ and $[k,D]$ for k^{low} and k^{upp}, respectively. In particular, we search for the largest k^{low} that $(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_{k^{low}} < \frac{(1+\frac{1}{\beta n_3})\sum_{j=k^{low}}^{k^{upp}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j}{(1+\frac{1}{\beta n_3})(k-k^{low}+1)+\frac{1}{\beta n_3}(k^{upp}-k)}$ and search for the smallest k^{upp} that $(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_{k^{upp}} > \frac{\sum_{j=k^{low}}^{k^{upp}}(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j}{(1+\frac{1}{\beta n_3})(k-k^{low}+1)+\frac{1}{\beta n_3}(k^{upp}-k)}$, which are the conditions in eq.(22). □

B2. Proof of Proposition 21

Proof. We proceed the derivation with the FFT transformed block diagonal matrix, which gives the equivalent formulation of eq.(25) in the paper in Fourier domain as

$$\mathbf{A}^{\#}_{\mathcal{F}} = \underset{\frac{1}{n_3}\|\mathbf{A}_{\mathcal{F}}\|_{msp,k}\le 1}{\operatorname{argmax}}\frac{1}{n_3}\langle\mathbf{T}_{\mathcal{F}},\mathbf{A}_{\mathcal{F}}\rangle. \qquad (16)$$

The maximum is obtained when $\mathbf{A}^{\#}_{\mathcal{F}}/n_3$ shares the same $\mathbf{U}_{\mathcal{F}}, \mathbf{V}_{\mathcal{F}}$ of $\mathbf{T}_{\mathcal{F}}$, which further converts computation to

$$\boldsymbol{\sigma}^{\downarrow}_{\mathbf{A}^{\#}_{\mathcal{F}}/n_3} = \underset{\|\mathbf{v}\|_{vp,k}\le 1}{\operatorname{argmax}}\langle\mathbf{v},\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}}\rangle. \qquad (17)$$

The above eq.(17) is the polar operator of the vector k-support norms which ensures $\mathbf{v}$ the following closed-form computation

$$(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{A}}^{\#}_{\mathcal{F}}}/n_3)_j = \mathbf{v}_j = \begin{cases}\frac{(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_j}{\|(\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{T}}_{\mathcal{F}}})_{[1:k]}\|_2}, & j\in[1:k],\\ 0, & j\in[k:D].\end{cases} \qquad (18)$$

Hence, after reshuffling the elements of $\boldsymbol{\sigma}^{\downarrow}_{\boldsymbol{\mathcal{A}}^{\#}_{\mathcal{F}}}$ back to their position in the original frontal slices according to `idx` kept during the `sort` operation, the polar map is proved to take the form as in eq.(26) and (27) in the paper. □

APPENDIX C
PROOFS FOR SECTION V

C1. Proof of Proposition 24

Proof. Beginning with the Lagrangian dual reformulation, the following sequence of equivalence relationship holds,

$$\begin{aligned}
&\max_{\mathfrak{J}} \min_{\mathcal{L},||\mathcal{E}||_s\le\tau} \Big[\frac{1}{2}||\mathcal{L}||_{tsp,k}^2 + \langle \mathfrak{J}, \mathfrak{M}(\mathcal{L}) + \mathcal{E} - \mathfrak{M}(\mathcal{X})\rangle\Big] \\
&\Leftrightarrow \max_{\mathfrak{J}} \Big[\Big(\min_{\mathcal{L}} \frac{1}{2}||\mathcal{L}||_{tsp,k}^2 + \langle \mathfrak{J}, \mathfrak{M}(\mathcal{L})\rangle\Big) \\
&\qquad + \Big(\min_{||\mathcal{E}||_s\le\tau} \langle \mathfrak{J}, \mathcal{E}\rangle\Big) - \Big(\langle \mathfrak{J}, \mathfrak{M}(\mathcal{X})\rangle\Big)\Big] \\
&\Leftrightarrow \max_{\mathfrak{J}} \Big[\min_{\mathcal{L}} -\Big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\Big) \\
&\qquad + \min_{||\mathcal{E}||_s\le\tau} -\Big(\langle -\mathfrak{J}, \mathcal{E}\rangle\Big) - \Big(\langle \mathfrak{J}, \mathfrak{M}(\mathcal{X})\rangle\Big)\Big] \\
&\Leftrightarrow \max_{\mathfrak{J}} \Big[-\underbrace{\max_{\mathcal{L}} \Big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\Big)}_{(i)} \\
&\qquad - \underbrace{\max_{||\mathcal{E}||_s\le\tau} \Big(\langle -\mathfrak{J}, \mathcal{E}\rangle\Big)}_{(ii)} - \Big(\langle \mathfrak{J}, \mathfrak{M}(\mathcal{X})\rangle\Big)\Big]
\end{aligned} \tag{19}$$

$$\begin{aligned}
&\Leftrightarrow \max_{\mathfrak{J}} -\Big[\frac{1}{2}(||-\mathfrak{M}^\top(\mathfrak{J})||_{tsp,k}^*)^2 + \langle \mathfrak{J}, \mathfrak{M}(\mathcal{X})\rangle + \tau||-\mathfrak{J}||_s^*\Big] \\
&\Leftrightarrow -\min_{\mathfrak{J}} \Big[\underbrace{\frac{1}{2}(||-\mathfrak{M}^\top(\mathfrak{J})||_{tsp,k}^*)^2 + \langle \mathfrak{J}, \mathfrak{M}(\mathcal{X})\rangle}_{\mathfrak{f}(\mathfrak{J})} + \underbrace{\tau||-\mathfrak{J}||_s^*}_{\mathfrak{h}(\mathfrak{J})}\Big].
\end{aligned} \tag{20}$$

In above, (i) is by the definition of Fenchel conjugate of $\frac{1}{2}||\cdot||_{tsp,k}^2$, i.e.

$$\max_{\mathcal{L}} \big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\big) = \frac{1}{2}(||-\mathfrak{M}^\top(\mathfrak{J})||_{tsp,k}^*)^2; \tag{21}$$

and (ii) is by the definition the dual norm of $||\cdot||_s$, i.e. $(ii) = \tau||-\mathfrak{J}||_s^*$. As a result, we can equivalently solve the dual objective: $\min_{\mathfrak{J}} \mathfrak{D}(\mathfrak{J}) := \min_{\mathfrak{J}} \mathfrak{f}(\mathfrak{J}) + \mathfrak{h}(\mathfrak{J})$. ☐

C2. Proof of Proposition 25

Proof. By taking the (sub)gradient of $\mathfrak{f}(\Gamma)$ in eq.(12), we have

$$\mathfrak{g}(\mathfrak{J}) = \partial\Big(\frac{1}{2}(||-\mathfrak{M}^\top(\mathfrak{J})||_{tsp,k}^*)^2\Big) + \mathfrak{M}(\mathcal{X}). \tag{22}$$

By the relation of eq.(21), it gives

$$\begin{aligned}
&\partial\Big(\frac{1}{2}(||-\mathfrak{M}^\top(\mathfrak{J})||_{tsp,k}^*)^2\Big) \\
&=\partial\Big(\max_{\mathcal{L}} \big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\big)\Big) \\
&=\partial\big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}^\#\rangle - \frac{1}{2}||\mathcal{L}^\#||_{tsp,k}^2\big) \\
&= -\mathfrak{M}(\mathcal{L}^\#),
\end{aligned} \tag{23}$$

where the last equality is because the (sub)gradient is taken with respect to $\mathfrak{J}$. The optimum $\mathcal{L}^\#$ in the second equality should satisfy,

$$\mathcal{L}^\# = \operatorname*{argmax}_{\mathcal{L}} \big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\big), \tag{24}$$

which has closed-form solution for every norm function. We provide the details for our TSP-k norm in the following for completeness. In order to compute $\mathcal{L}^\#$, we have

$$\begin{aligned}
&\big(\langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{L}\rangle - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\big) \\
&\le \big(\|\mathcal{L}\|_{tsp,k} \cdot \|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^* - \frac{1}{2}||\mathcal{L}||_{tsp,k}^2\big) \\
&= -\frac{1}{2}\big(\|\mathcal{L}\|_{tsp,k} - \|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*\big)_2^2 + \frac{1}{2}(\|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*)^2 \\
&\le \frac{1}{2}(\|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*)^2,
\end{aligned} \tag{25}$$

where both inequalities are obtained at $\mathcal{L}$ satisfying $\|\mathcal{L}\|_{tsp,k} = \|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*$. Hence, the "scale" of $\mathcal{L}^\#$ under the TSP-k norm is $\|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*$. Fixing this scale, we need to decide the "direction" of $\mathcal{L}^\#$, i.e., a tensor $\mathcal{A}^\#$ with unit TSP-k norm and the maximization becomes,

$$\operatorname*{argmax}_{\|\mathcal{A}\|_{tsp,k}\le 1} \langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{A}\|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*\rangle + \frac{1}{2}(\|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^*)^2, \tag{26}$$

which is the polar operator of TSP-k norm at $\mathfrak{T} = -\mathfrak{M}^\top(\mathfrak{J})$:

$$\mathcal{A}^\# = \operatorname*{argmax}_{\|\mathcal{A}\|_{tsp,k}\le 1} \langle -\mathfrak{M}^\top(\mathfrak{J}), \mathcal{A}\rangle. \tag{27}$$

As a result, we have proved that

$$\mathcal{L}^\# = \|-\mathfrak{M}^\top(\mathfrak{J})\|_{tsp,k}^* \cdot \mathcal{A}^\#. \tag{28}$$

☐

C3. Proof of Corollary 26

Proof. To prove the Corollary, it suffices to show the singular values of $\mathcal{L}^\#$. With $\mathfrak{T}_{\mathcal{F}} = -\mathfrak{M}^\top(\mathfrak{J})$, note that $\|\mathfrak{T}_{\mathcal{F}}\|_{tsp,k}^* = \|\boldsymbol{\sigma}_{\mathfrak{T}_{\mathcal{F}}}\|_{vp,k}^* = \|(\boldsymbol{\sigma}_{\mathfrak{T}_{\mathcal{F}}}^\downarrow)_{[1:k]}\|_2$. By eq.(28) and (18) and the linearity of FFT, the singular values $\boldsymbol{\sigma}_{\mathcal{L}_{\mathcal{F}}^\#}$ are

$$(\boldsymbol{\sigma}_{\mathcal{L}_{\mathcal{F}}^\#}(\mathtt{idx}))_{\mathtt{j}} = \begin{cases} n_3(\boldsymbol{\sigma}_{\mathfrak{T}_{\mathcal{F}}}^\downarrow)_j, & j \in [1:k], \\ 0, & j \in [k:D]. \end{cases} \tag{29}$$

The remaining is followed by Proposition 21. ☐

APPENDIX D
DETAILED DEFINITIONS IN TABLE II

Definition D.1. (Tensor Conjugate Transpose [2]) The conjugate transpose of a tensor $\mathcal{A}$ of size $n_1 \times n_2 \times n_3$ is the $n_2 \times n_1 \times n_3$ tensor $\mathcal{A}^\top$ obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3.

Definition D.2. (Identity Tensor [2]) A tensor $\mathfrak{J} \in \mathbb{R}^{n\times n\times n_3}$ is called identity tensor if its first frontal slice $\mathfrak{J}^{(1)}$ is the $n \times n$ identity matrix and all its other frontal slices, i.e. $\mathfrak{J}^{(i)}$ for $i = 2, ..., n_3$, are zero matrices.

Definition D.3. (Orthogonal Tensor [2]) A tensor $\mathfrak{Q} \in \mathbb{R}^{n\times n\times n_3}$ is called orthogonal if the following condition holds,

$$\mathfrak{Q}^\top * \mathfrak{Q} = \mathfrak{Q} * \mathfrak{Q}^\top = \mathfrak{J}, \tag{30}$$

where $\mathfrak{J} \in \mathbb{R}^{n\times n\times n_3}$ is an identity tensor as in Definition D.2 and $*$ is the t-Product.

Definition D.4. (f-Diagonal Tensor [2]) For a tensor $\mathcal{A}$, if all its frontal slices $\mathcal{A}^{(i)}$, $i = 1, ..., n_3$ are diagonal matrices, then it is defined to be an f-diagonal tensor.

Definition D.5. (Tensor Spectral Norm [3]) For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor spectral norm $\|\mathcal{A}\|_2$ is defined to be the spectral norm of $\mathbf{A}_{\mathcal{F}}$, i.e. $\|\mathcal{A}\|_2 := \|\mathbf{A}_{\mathcal{F}}\|_2$.

Definition D.6. (Tensor Frobenius Norm [3]) For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor Frobenius norm is denoted by $\|\mathcal{A}\|_F$, i.e. $\|\mathcal{A}\|_F := \langle \mathcal{A}, \mathcal{A} \rangle^{\frac{1}{2}} = \frac{1}{\sqrt{n_3}} \|\mathcal{A}_{\mathcal{F}}\|_F = \frac{1}{\sqrt{n_3}} \|\mathbf{A}_{\mathcal{F}}\|_F = \sqrt{\sum_i \sum_j \sum_k \mathcal{A}_{ijk}^2}$.

Appendix E
Additional Materials for Section V

E1. Detailed derivation of preconditioned ADMM

To solve eq.(29) in the paper via ADMM-type algorithm, the augmented Lagrangian is given by

$$\begin{aligned} \mathfrak{D}_\rho(\mathcal{L}, \mathcal{E}, \mathcal{J}) = \frac{1}{2} \|\mathcal{L}\|_{tsp,k}^2 + \lambda \|\mathcal{E}\|_1 \\ + \langle \mathcal{J}, \mathfrak{M}(\mathcal{X} - \mathcal{L}) - \mathcal{E} \rangle + \frac{\rho}{2} \|\mathfrak{M}(\mathcal{X} - \mathcal{L}) - \mathcal{E}\|_F^2. \end{aligned} \tag{31}$$

We then carry out alternative update to the variables $\mathcal{L}$ and $\mathcal{E}$ at every iteration.

1) Update of $\mathcal{L}_t$:

$$\begin{aligned} \mathcal{L}_t = \underset{\mathcal{L}}{\operatorname{argmin}} \frac{1}{2} \|\mathcal{L}\|_{tsp,k}^2 \\ + \langle \mathcal{J}_{t-1}, \mathfrak{M}(\mathcal{X} - \mathcal{L}) - \mathcal{E}_{t-1} \rangle + \frac{\rho}{2} \|\mathfrak{M}(\mathcal{X} - \mathcal{L}) - \mathcal{E}_{t-1}\|_F^2. \end{aligned} \tag{32}$$

To separate the linear map $\mathfrak{M}$ apart from $\mathcal{L}$, the preconditioned ADMM approximates $\frac{1}{2} \|\mathfrak{M}(\mathcal{X} - \mathcal{L}) - \mathcal{E}_{t-1}\|_F^2$ with second order Taylor expansion around $\mathcal{L}_{t-1}$, as

$$\begin{aligned} \frac{1}{2} \|\mathfrak{M}(\mathcal{X} - \mathcal{L}_{t-1}) - \mathcal{E}_{t-1}\|_F^2 - \\ \langle \mathfrak{M}^\top (\mathfrak{M}(\mathcal{X} - \mathcal{L}_{t-1}) - \mathcal{E}_{t-1}), \mathcal{L} - \mathcal{L}_{t-1} \rangle + \frac{\eta}{2} \|\mathcal{L} - \mathcal{L}_{t-1}\|_F^2. \end{aligned} \tag{33}$$

Incorporating eq.(33) into eq.(32) gives

$$\begin{aligned} \mathcal{L}_t = \underset{\mathcal{L}}{\operatorname{argmin}} \frac{1}{2\rho\eta} \|\mathcal{L}\|_{tsp,k}^2 + \\ \frac{1}{2} \|\mathcal{L} - (\mathcal{L}_{t-1} + \frac{1}{\eta} \mathfrak{M}^\top (\mathfrak{M}(\mathcal{X} - \mathcal{L}_{t-1}) - \mathcal{E}_{t-1}) + \frac{\mathcal{J}_{t-1}}{\rho\eta})\|_F^2 \\ = \mathtt{Prox}_{\frac{1}{2\rho\eta} \|\cdot\|_{tsp,k}^2} \Big(\mathcal{L}_{t-1} + \frac{1}{\eta} \mathfrak{M}^\top (\mathfrak{M}(\mathcal{X} - \mathcal{L}_{t-1}) - \mathcal{E}_{t-1}) \\ + \frac{\mathcal{J}_{t-1}}{\rho\eta} \Big). \end{aligned} \tag{34}$$

2) Update of $\mathcal{E}_t$:

$$\begin{aligned} \mathcal{E}_t = \underset{\mathcal{E}}{\operatorname{argmin}} \lambda \|\mathcal{E}\|_1 \\ + \langle \mathcal{J}_{t-1}, \mathfrak{M}(\mathcal{X} - \mathcal{L}_t) - \mathcal{E} \rangle + \frac{\rho}{2} \|\mathfrak{M}(\mathcal{X} - \mathcal{L}_t) - \mathcal{E}\|_F^2 \\ = \underset{\mathcal{E}}{\operatorname{argmin}} \frac{\lambda}{\rho} \|\mathcal{E}\|_1 + \frac{1}{2} \|\mathcal{E} - (\mathfrak{M}(\mathcal{X} - \mathcal{L}_t) + \frac{1}{\rho} \mathcal{J}_{t-1})\|_F \\ = \mathtt{Prox}_{\frac{\lambda}{\rho} \|\cdot\|_1} \Big(\mathfrak{M}(\mathcal{X} - \mathcal{L}_t) + \frac{1}{\rho} \mathcal{J}_{t-1} \Big). \end{aligned} \tag{35}$$

which then follows by the element-wise soft-thresholding operation.

E2. Computational Complexity Analysis for Algorithms 2&3

Suppose the input tensor size is $n_1 \times n_2 \times n_3$.

Algorithm 2:

- Step 1: Computing `fft` takes $O(n_1 n_2 n_3 \log(n_3))$;
- Step 2-4: Compute n_3 full SVD of $n_1 \times n_2$ matrices takes $O(n_1 n_2 n_3 \min\{n_1, n_2\})$;
- Step 5: Sorting a vector of length $\min\{n_1, n_2\} n_3$ takes $O(\min\{n_1, n_2\} n_3 \log(\min\{n_1, n_2\} n_3))$;
- Step 6-7: Repeating binary search at most k and $\min\{n_1, n_2\} n_3 - k$ times for ranges of $[1, k]$ and $[k, \min\{n_1, n_2\} n_3]$ each takes $O(k \log(k))$ and $(\min\{n_1, n_2\} n_3 - k) O(\min\{n_1, n_2\} n_3 - k)$, correspondingly;
- Step 8: Element-wise vector arithmetic takes $O(\min\{n_1, n_2\} n_3)$;
- Step 9: Rearranging vector elements back according to `idx` kept during `sort` operation takes $O(\min\{n_1, n_2\} n_3)$;
- Step 10-12: n_3 matrix multiplications take $O(n_1 n_2 n_3 \min\{n_1, n_2\})$;
- Step 13: `ifft` takes $O(n_1 n_2 n_3 \log(n_3))$;

Algorithm 3:

- Step 1: Computing `fft` takes $O(n_1 n_2 n_3 \log(n_3))$;
- Step 2-4: Compute n_3 partial SVD of $n_1 \times n_2$ matrices takes $O(k n_1 n_2 n_3)$;
- Step 5: Sorting a vector of length $k n_3$ takes $O k n_3 \log(k n_3))$;
- Step 6: Element-wise vector arithmetic takes $O(k)$;
- Step 7: Rearranging vector elements back according to `idx` kept during `sort` operation takes $O(k)$;
- Step 8-10: n_3 matrix multiplications with $O(k n_1 n_2 n_3)$;
- Step 13: `ifft` takes $O(n_1 n_2 n_3 \log(n_3))$;

E3. Line-search Subroutine for Algorithm 5

Ideally, the algorithm performs better when adapting to the local smoothness of the dual loss function. To investigate such possibility, we study the structure of the (sub)gradient set. As the key ingredient of the (sub)gradient is the differentiation of the dual tensor spectral k-support norm, we describe its structure in the following proposition.

In particular, let $\mathcal{Y} = -\mathfrak{M}^\top(\mathcal{J})$. We describe the (sub)gradient set of $\|\mathcal{Y}\|_{tsp,k}^*$ in the following.

Proposition E.1. *Denote a particular tensor SVD of $\mathcal{Y}$ by $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ and the associated Fourier transformed diagonal matrix by $\mathbf{Y}_{\mathcal{F}} = \mathbf{U}_{\mathcal{F}} \mathtt{diag}(\boldsymbol{\sigma}) \mathbf{V}_{\mathcal{F}}^\top$. Denote the sorted non-increasing singular vectors by $\boldsymbol{\sigma}^\downarrow$. Assume the sorted singular values satisfy*

$$\boldsymbol{\sigma}_1^\downarrow \geq ... > \underbrace{\boldsymbol{\sigma}_{k-a+1}^\downarrow = ... = \boldsymbol{\sigma}_k^\downarrow = ... = \boldsymbol{\sigma}_{k+b}^\downarrow}_{(i)} > ... \geq \boldsymbol{\sigma}_{n_2 \cdot n_3}^\downarrow. \tag{36}$$

Let $\mathbb{K}_1$ denote the corresponding index set in $\boldsymbol{\sigma}$ of entries in $\boldsymbol{\sigma}_{[1:k-a]}^\downarrow$, and $\mathbb{K}_2$ for the indices of entries in $\boldsymbol{\sigma}_{[k-a+1:k+b]}^\downarrow$, *and $\mathbb{K}$ denotes the indices of the leading k singular values. The (sub)gradient set of the dual spectral k-support norm at $\mathbf{Y}_{\mathcal{F}}$ is denoted by $\mathfrak{G}_{\mathcal{F}}$, of which each i-th frontal slice in the Fourier domain $\mathfrak{G}_{\mathcal{F}}^{(i)} =$*

$$\begin{aligned}\frac{1}{\|\mathbf{Y}_{\mathcal{F}}\|_{msp,k}^{*}}\Big\{ &\sum_{k_1\in\mathbb{K}_1}(\mathbf{U}_{\mathcal{F}})_{k_1}^{i}\boldsymbol{\sigma}_{k_1}^{(i)}((\overline{\mathbf{V}})_{k_1}^{(i)})^{\top} \\ &+\sum_{k_2\in\mathbb{K}_2}(\mathbf{U}_{\mathcal{F}})_{k_2}^{i}\mathbf{T}((\mathbf{V}_{\mathcal{F}})_{k_2}^{(i)})^{\top}\Big\},\end{aligned} \tag{37}$$

where $\mathbf{T}$ satisfies $\mathbf{T}=\mathbf{T}^{\top}$, $\|\mathbf{T}\|_2 \le \boldsymbol{\sigma}_k^{\downarrow}$ and $\|\mathbf{T}\|_* = a\boldsymbol{\sigma}_k^{\downarrow}$. In particular, $\|\mathfrak{Y}\|_{tsp,k}^{*}$ is differentiable at $\mathfrak{Y}$, if $\boldsymbol{\sigma}_k^{\downarrow} > \boldsymbol{\sigma}_{k+1}^{\downarrow}$ or $\boldsymbol{\sigma}_k^{\downarrow}=0$, during when $\mathfrak{G}_{\mathcal{F}}$ denotes the unique gradient, which satisfies,

$$\mathfrak{G}_{\mathcal{F}}^{(i)} = \frac{1}{\|\mathbf{Y}_{\mathcal{F}}\|_{msp,k}^{*}}\sum_{k\in\mathbb{K}}(\mathbf{U}_{\mathcal{F}})_{k}^{i}\boldsymbol{\sigma}_{k}^{(i)}((\mathbf{V}_{\mathcal{F}})_{k}^{(i)})^{\top}. \tag{38}$$

The singular value subsequence of (i) in eq.(36) is the longest subsequence which equals to each other containing the k-the largest singular value. Under the condition $\boldsymbol{\sigma}_k^{\downarrow} > \boldsymbol{\sigma}_{k+1}^{\downarrow}$ or $\boldsymbol{\sigma}_k^{\downarrow}=0$, the leading k singular values are well-separated with the remaining smaller singular values starting from $k+1$-th entry. Also, eq.(38) is always a particular choice in eq.(37), since eq.(38) corresponds to $\mathbf{T}=\boldsymbol{\sigma}_k^{\downarrow}\mathtt{diag}(\mathbf{1}_{[1:a]})$ (obviously $\|\mathbf{T}\|_2 \le \boldsymbol{\sigma}_k^{\downarrow}$ and $\|\mathbf{T}\|_* = a\boldsymbol{\sigma}_k^{\downarrow}$ are satisfied). In particular, when the singular value are "well-separated", eq.(38) becomes the unique element of the subgradient set of eq.(37), i.e. the dual tensor spectral k-support norm is differentiable at $\mathfrak{Y}$ with "well-separated" singular values under t-SVD. Finally, by the computation of the polar operator and the dual (sub)gradient, we actually always choose eq.(38) in the dual objective optimization.

Despite the smoothness variation, we always choose eq.(38) during the dual (sub)gradient computation. Hence, it would be desirable the our optimization procedure would adaptive to the smoothness to the dual loss function. The $\nu \ge 1$ is called the Hölder smoothness order, and the associated parameter H_ν is defined as

$$H_\nu = H_\nu(\mathfrak{f}) := \sup_{\mathbf{x}_1\neq\mathbf{x}_2}\Big\{\frac{\|\nabla\mathfrak{f}(\mathbf{x}_1)-\nabla\mathfrak{f}(\mathbf{x}_2)\|_2}{\|\mathbf{x}_1-\mathbf{x}_2\|_2^{\nu}}\Big\}, \tag{39}$$

where $\nabla\mathfrak{f}(\mathbf{x})$ denotes a (sub)gradient of $\mathfrak{f}$ at $\mathbf{x}$.

In particular, we propose to rely on the concept of the Hölder smoothness $\nu\in[0,1]$ and the line-search strategy that adapts to this smoothness variation. By Hölder smoothness, the smooth case corresponds to $\nu=1$ in the Hölder smoothness, while when the (sub)gradient set is not unique, it corresponds to $\nu=0$ in the Hölder smoothness. H_t is determined a backtracking line search, which searches the $\mathfrak{J}_{t+1}$ satisfying,

$$\begin{aligned}&\mathfrak{f}(\mathfrak{J}_{t+1})+\mathfrak{h}(\mathfrak{J}_{t+1}) \\ &\le \mathfrak{f}(\breve{\mathfrak{J}}_t)+\langle\mathfrak{g}(\breve{\mathfrak{J}}_t),\mathfrak{J}_{t+1}-\breve{\mathfrak{J}}_t\rangle+\frac{H_t}{2}\|\mathfrak{J}_{t+1}-\breve{\mathfrak{J}}_t\|_F^2+\mathfrak{h}(\mathfrak{J}_{t+1})\end{aligned} \tag{40}$$

Algorithm 1 presents the backtracking line search for determining H_t and updating $\mathfrak{J}_{t+1}$. The dominating per-iteration complexity, with the remaining being simple tensor inner product and element-wise operation, lines in Line 3 and 4: 1) proximal mapping with respect to the dual regularizer and 2) the computation of $\mathfrak{J}_{\mathrm{t,i+1}}$. The former computational cost mainly comes from the projection onto the unit $\|\cdot\|_1$ ball which has $O(n_1n_2n_3)$ complexity algorithm. The latter mainly needs to compute the dual tensor spectral k-support norm on $\mathfrak{J}_{t,i+1}$, which relies on the leading k singular values of each $\hat{\mathfrak{J}}_{t,i+1}$ and can be obtained in $O(kn_1n_2n_3)$ computation by *partial* SVD. Note that such line search is only possible after our dual reformulation, because the primal norm computation during such backtracking would require *full* SVD whichs cost super-linear complexity $O(n_1(n_2)^2n_3)$. Also, by [4], the line search requires roughly two rounds on average. In sum, the line search step takes the complexity $O(kn_1n_2n_3)$.

Algorithm 1 line-search subroutine: $(\mathfrak{J}_{t+1},H_t) = \mathtt{line\_search}(\breve{\mathfrak{J}}_t,\mathfrak{g}(\breve{\mathfrak{J}}_t),H_{t-1},\epsilon,\theta_t)$

Input: $\breve{\mathfrak{J}}_t,\mathfrak{g}(\breve{\mathfrak{J}}_t),H_{t-1},\epsilon,\theta_t$

1: $H_{t,0}=H_{t-1}/2$;
2: **for** $i=0,1,...,i_{max}$ **do**
3:   $\mathfrak{J}_{t,i+1}=\mathtt{Prox}_{H_{t,i}^{-1}\mathfrak{h}(\mathfrak{J})}\big(\breve{\mathfrak{J}}_t-H_{t,i}^{-1}\mathfrak{g}(\breve{\mathfrak{J}}_t)\big)$;
4:   **if** $\mathfrak{f}(\mathfrak{J}_{t,i+1})\le\mathfrak{f}(\breve{\mathfrak{J}}_t)+\langle\mathfrak{g}(\breve{\mathfrak{J}}_t),\mathfrak{J}_{t,i+1}-\breve{\mathfrak{J}}_t\rangle+\frac{H_{t,i}}{2}\|\mathfrak{J}_{t,i+1}-\breve{\mathfrak{J}}_t\|_F^2+\frac{\epsilon}{2\theta_t}$ **then**
5:     **break**;
6:   **else**
7:     $H_{t,i+1}=2H_{t,i}$;
8:   **end if**
9: **end for**
10: **Return:** $\mathfrak{J}_{t,i},H_{t,i}$

References

[1] A. Eriksson, T. T. Pham, T.-J. Chin, and I. Reid, "The k-support norm and convex envelopes of cardinality and rank," in *CVPR*. IEEE, 2015, pp. 3349–3357.

[2] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 148–172, 2013.

[3] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2017.

[4] A. Yurtsever, Q. T. Dinh, and V. Cevher, "A universal primal-dual convex optimization framework," in *NIPS*, 2015, pp. 3150–3158.