ORDER-D ($D \geq 4$) T-SVD ALGEBRAIC FRAMEWORK

Wenjin Qin[1], *Hailin Wang*[2], *Jianjun Wang*[1,3,*]

[1]School of Mathematics and Statistics, Southwest University, China
[2] School of Mathematics and Statistics, Xi'an Jiaotong University, China
[3] Research Institute of Intelligent Finance and Digital Economics, Southwest University, China

1. BASIC ALGEBRAIC FRAMEWORK

The main notions and preliminaries for order-d tensor are listed in Table 1, some of which originate from [1–5].

Definition 1.1. ***(Order-d t-product)*** *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3 \times \cdots \times n_d}$, then their tensor-tensor product (t-product) is defined as follows:*

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \operatorname{bfold}(\operatorname{bcirc}(\mathcal{A}) \times \operatorname{bunfold}(\mathcal{B})). \tag{1}$$

The order-d t-product in (1) can be converted to the matrix-matrix multiplication in the transform domain. That is,

$$\mathcal{C} = \mathcal{A} *_L \mathcal{B} = L^{-1}(\mathcal{A}_L \triangle \mathcal{B}_L). \tag{2}$$

Definition 1.2. ***(Order-d conjugate transpose)*** *The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$ is the tensor $\mathcal{A}^* \in \mathbb{C}^{n_2 \times n_1 \times n_3 \times \cdots \times n_d}$ if ${\mathcal{A}^*}_L(:,:,i_3,\cdots,i_d) = (\mathcal{A}_L(:,:,i_3,\cdots,i_d))^*$, for $i_j \in \{1,\cdots,n_j\}, j \in \{3,\cdots,d\}$.*

Definition 1.3. ***(order-d identity tensor)*** *The order-d identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3 \times \cdots \times n_d}$ is the tensor such that $\mathcal{I}_L(:,:,i_3,\cdots,i_d) = \mathrm{I}_n$ for $i_3 \in \{1,\cdots,n_3\},\cdots,i_d \in \{1,\cdots,n_d\}$, where I_n denotes a $n \times n$ sized identity matrix.*

Definition 1.4. ***(Order-d orthogonal tensor)*** *An order-d tensor $\mathcal{Q} \in \mathbb{C}^{n \times n \times n_3 \times \cdots \times n_d}$ is orthogonal if it satisfies $\mathcal{Q}^* *_L \mathcal{Q} = \mathcal{Q} *_L \mathcal{Q}^* = \mathcal{I}$.*

Definition 1.5. ***(Order-d f-diagonal tensor)*** *An order-d tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times n_3 \times \cdots \times n_d}$ is called f-diagonal if $\mathcal{A}(:,:,i_3,\cdots,i_d)$ is a diagonal matrix for any $i_j \in \{1,\cdots,n_j\}, j \in \{3,\cdots,d\}$.*

Definition 1.6. ***(Order-d Gaussian random tensor)*** *$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is called a Gaussian random tensor, if all $\mathcal{A}_L(:,:,i_3,\cdots,i_d)$ satisfy the standard normal distribution for any $i_3 \in \{1,\cdots,n_3\},\cdots,i_d \in \{1,\cdots,n_d\}$.*

Definition 1.7. ***(Order-d t-QR)*** *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$, then it can be decomposed as*

$$\mathcal{A} = \mathcal{Q} *_L \mathcal{R}, \tag{3}$$

where $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_1 \times n_3 \times \cdots \times n_d}$ is an order-d orthogonal tensor, $\mathcal{R} \in \mathbb{R}^{n_2 \times n_2 \times n_3 \times \cdots \times n_d}$ is f-upper triangular tensor.

*✉ wjj@swu.edu.cn.

Definition 1.8. ***(Order-d t-SVD, Order-d tensor rank)*** *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$, then it can be factorized as*

$$\mathcal{A} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^*, \tag{4}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3 \times \cdots \times n_d}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3 \times \cdots \times n_d}$ are order-d orthogonal tensors, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$ is an order-d f-diagonal tensor (also named singular value tensor). Further tensor rank of $\mathcal{A}$ can be defined as

$$\operatorname{rank}_{tr}(\mathcal{A}) = \#\{i : \mathcal{S}(i,i,i_3,\cdots,i_d) \neq \mathbf{0}\},$$

for $i_3 \in \{1,\cdots,n_3\},\cdots,i_d \in \{1,\cdots,n_d\}$.

Definition 1.9. ***(Order-d tensor spectral norm)*** *The spectral norm of $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is defined as*

$$\|\mathcal{A}\| := \|\operatorname{bdiag}(\mathcal{A}_L)\|.$$

Definition 1.10. ***(Order-d TNN)*** *Let $\mathcal{A}_L$ has the t-SVD $\mathcal{A}_L = \mathcal{U}' *_L \mathcal{S}' *_L \mathcal{V}'^*$ for any $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, then the tensor nuclear norm of $\mathcal{A}$ is defined as*

$$\|\mathcal{A}\|_{\star,L} := \frac{1}{\rho}\|\operatorname{bdiag}(\mathcal{A}_L)\|_\star = \frac{1}{\rho}\sum_{i=1}^{m}\sum_{i_3=1}^{n_3}\cdots\sum_{i_d=1}^{n_d} \mathcal{S}'(i,i,i_3,\cdots,i_p),$$

where $\rho > 0$ is a constant, and $m = \min(n_1, n_2)$.

Definition 1.11. ***(Order-d WTSN)*** *Let $\mathcal{A}_L$ has the t-SVD $\mathcal{A}_L = \mathcal{U}' *_L \mathcal{S}' *_L \mathcal{V}'^*$ for any $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, then the weighted tensor Schatten-p norm (WTSN) of $\mathcal{A}$ is defined as*

$$\|\mathcal{A}\|_{\mathcal{W},S_p} := \Bigg(\frac{1}{\rho}\sum_{i=1}^{m}\sum_{i_3=1}^{n_3}\cdots\sum_{i_d=1}^{n_d}\mathcal{W}(i,i,i_3,\cdots,i_p) \left|\mathcal{S}'(i,i,i_3,\cdots,i_p)\right|^p\Bigg)^{\frac{1}{p}}, \tag{5}$$

where $\rho > 0$ is a constant, $\mathcal{W}$ denotes the order-d f-diagonal tensor (i.e., weight parameter), and $m = \min(n_1, n_2)$.

The WTSN defined in (5) can be equivalently reformulated as follows:

$$\|\mathcal{A}\|_{\mathcal{W},S_p}^p = \frac{1}{\rho}\sum_{i_3,\cdots,i_d}\operatorname{tr}\left(\mathrm{W}^{(i_3,\cdots,i_p)}\left|\mathrm{S}'^{(i_3,\cdots,i_p)}\right|^p\right). \tag{6}$$

Table 1: The main notions and preliminaries for order-d tensor.

Notations	Descriptions	Notations	Descriptions
$\mathcal{A} \in \mathbb{C}^{n_1\times\cdots\times n_d}$	order-d tensor	$\mathcal{A}^* \in \mathbb{C}^{n_1\times\cdots\times n_d}$	conjugate transpose
$\mathcal{A}_{i_1\cdots i_d}$ or $\mathcal{A}(i_1,\cdots,i_d)$	$(i_1,\cdots,i_d)$-th entry	$\mathrm{A}_{(k)} \in \mathbb{C}^{n_k\times\prod_{j\neq k} n_j}$	mode-k matricization of $\mathcal{A}$
$\Vert\mathcal{A}\Vert_\infty = \max_{i_1\cdots i_d} \vert\mathcal{A}_{i_1\cdots i_d}\vert$	tensor infinity norm	$\mathcal{A}(i_1,\cdots,i_{k-1},:,i_{k+1},\cdots,i_d)$	fiber along mode-k
$\Vert\mathcal{A}\Vert_q = \left(\sum_{i_1\cdots i_d}\vert\mathcal{A}_{i_1\cdots i_d}\vert^q\right)^{\frac{1}{q}}$	tensor ℓ_q-norm	$\mathcal{A}(:,:,i_3,\cdots,i_d)$ or $\mathrm{A}^{(i_3,\cdots,i_d)}$	slice along mode-1, mode-2
$\Vert\mathcal{A}\Vert_F = (\sum_{i_1\cdots i_d}\vert\mathcal{A}_{i_1\cdots i_d}\vert^2)^{1/2}$	tensor Frobenius norm	$\langle\mathcal{A},\mathcal{B}\rangle = \sum\langle \mathrm{A}^{(i_3,\cdots,i_d)}, \mathrm{B}^{(i_3,\cdots,i_d)}\rangle$	tensor inner product
$L(\cdot): \mathbb{C}^{n_1\times\cdots\times n_d} \to \mathbb{C}^{n_1\times\cdots\times n_d}$	invertible linear transforms	$\mathcal{A} *_L \mathcal{B}$	transforms L based t-product
$\bar{A} = \mathrm{fft}(\mathcal{A},[\,],i)$ for $i=3,\cdots,d$	Fast Fourier Transform	$\mathcal{A} = \mathrm{ifft}(\bar{A},[\,],j)$ for $j=d,\cdots,3$	inverse Fast Fourier Transform
$\mathcal{A}_i \in \mathbb{R}^{n_1\times\cdots\times n_{d-1}}$	order-$(d-1)$ tensor constructed by keeping the d-th index of $\mathcal{A}$ fixed at i, $\mathcal{A}_i := \mathcal{A}(:,\cdots,:,i)$.		
$\mathrm{A}^j \in \mathbb{R}^{n_1\times n_2}$	$\mathrm{A}^j = \mathcal{A}(:,:,i_3,\cdots,i_d),\ j=(i_d-1)n_3\cdots n_{d-1}+\cdots+(i_4-1)n_3+i_3,\ i_d\in\{1,\cdots,n_d\}$.		
$\mathcal{A}\times_n \mathrm{U}$	the mode-n product of tensor $\mathcal{A}$ with matrix U, $\mathcal{B} = \mathcal{A}\times_n\mathrm{U} \Leftrightarrow \mathrm{B}_{(n)} = \mathrm{U}\cdot\mathrm{A}_{(n)}$.		
$\mathcal{A}_L \triangleq L(\mathcal{A})$	$L(\mathcal{A}) = \mathcal{A}\times_3 \mathrm{U}_{n_3}\times_4 \mathrm{U}_{n_4}\cdots\times_d \mathrm{U}_{n_d}$, $\mathrm{U}_{n_i}\in\mathbb{C}^{n_i\times n_i}$ denotes an invertible transform matrix.		
$L^{-1}(\mathcal{A})$	$L^{-1}(\mathcal{A}) = \mathcal{A}\times_d \mathrm{U}_{n_d}^{-1}\times_{d-1}\mathrm{U}_{n_{d-1}}^{-1}\cdots\times_3\mathrm{U}_{n_3}^{-1}$, $L^{-1}(L(\mathcal{A}))=\mathcal{A}$.		
$\mathrm{circ}\,(\mathcal{A})$	$\mathrm{circ}\,(\mathcal{A}) = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_{n_d} & \mathcal{A}_{n_d-1} & \cdots & \mathcal{A}_2 \\ \mathcal{A}_2 & \mathcal{A}_1 & \mathcal{A}_{n_d} & \cdots & \mathcal{A}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{n_d} & \mathcal{A}_{n_d-1} & \mathcal{A}_{n_d-2} & \cdots & \mathcal{A}_1 \end{bmatrix} \in \mathbb{R}^{n_1n_d\times\cdots\times n_{d-2}n_d\times n_{d-1}}$.		
$\mathrm{bcirc}(\mathcal{A})$	a $(n_1n_3\cdots n_d\times n_2n_3\cdots n_d)$ block circulant matrix at the base level of the operator $\mathrm{circ}(\mathcal{A})$.		
$\mathrm{unfold}\,(\mathcal{A})$	$\mathrm{unfold}\,(\mathcal{A}) = [\mathcal{A}_1,\mathcal{A}_2,\cdots,\mathcal{A}_{n_d-1},\mathcal{A}_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_1n_d\times n_2\times\cdots\times n_{d-1}}$.		
$\mathrm{bunfold}(\mathcal{A})$	a $(n_1n_3\cdots n_d\times n_2)$ matrix formed by applying $\mathrm{unfold}(\cdot)$ repeatedly until a block matrix result.		
$\mathrm{bfold}(\mathcal{A})$	the operation takes $\mathrm{bunfold}\,(\mathcal{A})$ back to order-d tensor form, i.e., $\mathrm{bfold}(\mathrm{bunfold}\,(\mathcal{A}))=\mathcal{A}$.		
$\mathrm{bdiag}(\mathcal{A})$	$\mathrm{bdiag}(\mathcal{A}) = \mathrm{diag}(\mathrm{A}^1,\cdots,\mathrm{A}^j,\cdots,\mathrm{A}^J),\ J=n_3\cdots n_d,\ j\in\{1,\cdots,J\}$.		
$\mathcal{A}\triangle\mathcal{B}$	face-wise product of two order-d tensor, $\mathcal{C}=\mathcal{A}\triangle\mathcal{B}\Leftrightarrow\mathrm{bdiag}(\mathcal{C})=\mathrm{bdiag}(\mathcal{A})\cdot\mathrm{bdiag}(\mathcal{B})$.		

Remark 1.1. *Throughout the article, the constant ρ appeared in the key definition and theorem is the constant ρ obtained when corresponding matrices of the invertible linear transform L satisfy the equation:*

$$\begin{aligned}&(\mathrm{U}_{n_p}\otimes\mathrm{U}_{n_{p-1}}\otimes\cdots\otimes\mathrm{U}_{n_3})(\mathrm{U}_{n_p}^*\otimes\mathrm{U}_{n_{p-1}}^*\otimes\cdots\otimes\mathrm{U}_{n_3}^*)\\ &=(\mathrm{U}_{n_p}^*\otimes\mathrm{U}_{n_{p-1}}^*\otimes\cdots\otimes\mathrm{U}_{n_3}^*)(\mathrm{U}_{n_p}\otimes\mathrm{U}_{n_{p-1}}\otimes\cdots\otimes\mathrm{U}_{n_3})\\ &=\rho\mathrm{I}_{n_3n_4\cdots n_p},\end{aligned} \tag{7}$$

the constant ρ is related to the linear transforms. For instance, for a fifth-order tensor, if corresponding matrices of the invertible linear transform L satisfy: $\mathrm{U}_{n_3}\cdot\mathrm{U}_{n_3}^* = \mathrm{U}_{n_3}^*\cdot\mathrm{U}_{n_3} = n_3\mathrm{I}_{n_3}$, $\mathrm{U}_{n_4}\cdot\mathrm{U}_{n_4}^* = \mathrm{U}_{n_4}^*\cdot\mathrm{U}_{n_4} = n_4\mathrm{I}_{n_4}$ and $\mathrm{U}_{n_5}\cdot\mathrm{U}_{n_5}^* = \mathrm{U}_{n_5}^*\cdot\mathrm{U}_{n_5} = n_5\mathrm{I}_{n_5}$, then $\rho = n_3\cdot n_4\cdot n_5$.

Remark 1.2. *Let $\mathcal{X}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ with $\mathrm{rank}_{tr}(\mathcal{X}) = r$, the skinny t-SVD of tensor $\mathcal{X}$ is $\mathcal{X} = \mathcal{U}*_L\mathcal{S}*_L\mathcal{V}^*$, where $\mathcal{U}\in\mathbb{R}^{n_1\times r\times n_3\times\cdots\times n_d}$ and $\mathcal{V}\in\mathbb{R}^{n_2\times r\times n_3\times\cdots\times n_d}$ satisfy $\mathcal{U}^**_L\mathcal{U}=\mathcal{I}$ and $\mathcal{V}^**_L\mathcal{V}=\mathcal{I}$, and $\mathcal{S}\in\mathbb{R}^{r\times r\times n_3\times\cdots\times n_d}$ is a f-diagonal tensor.*

2. ORDER-D WTSN OPERATOR

Definition 2.1. ***(Order-d WTSN proximal operator)*** *For any $\mathcal{A}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ with t-SVD that $\mathcal{A}=\mathcal{U}*_L\mathcal{S}*_L\mathcal{V}^*$, and any $\tau>0$, the order-d WTSN proximal operator is defined as follows*

$$\mathcal{D}_{\mathcal{W},p,\tau}(\mathcal{A}) = \mathcal{U}*_L\mathcal{S}_{\mathcal{W},p,\tau}*_L\mathcal{V}^*, \tag{8}$$

where $\mathcal{S}_{\mathcal{W},p,\tau} = L^{-1}(GST(\mathcal{S}_L,\tau\mathcal{W},p,J))$, in which J is the number of GST iterations, p $(0<p<1)$ represents the adjustable parameters appeared in order-d WTSN, and $\mathcal{W}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ denotes an order-d f-diagonal tensor (i.e., weight parameter).

Algorithm 1 Order-d t-product

Input: $\mathcal{A}\in\mathbb{R}^{n_1\times n_2\times n_3\times\cdots\times n_d}$,$\mathcal{B}\in\mathbb{R}^{n_2\times l\times n_3\times\cdots\times n_d}$, and invertible linear transform L.

Output: $\mathcal{C}=\mathcal{A}*_L\mathcal{B}\in\mathbb{R}^{n_1\times l\times n_3\times\cdots\times n_d}$.

1. Compute the result of linear transform on $\mathcal{A}$ and $\mathcal{B}$
   $\mathcal{A}_L\leftarrow L(\mathcal{A})$, $\mathcal{B}_L\leftarrow L(\mathcal{B})$.
2. Compute each matrix slice of $\mathcal{C}_L$ by
   for $i_3\in\{1,\cdots,n_3\},\cdots,i_d\in\{1,\cdots,n_d\}$ **do**
   $\mathcal{C}_L(:,:,i_3,\cdots,i_d) = \mathcal{A}_L(:,:,i_3,\cdots,i_d)\cdot\mathcal{B}_L(:,:,i_3,\cdots,i_d)$.
   end for
3. Compute the result of inverse linear transform on $\mathcal{C}_L$
   $\mathcal{C}\leftarrow L^{-1}(\mathcal{C}_L)$.

Algorithm 2 Order-d t-SVD

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and invertible linear transform L.

Output: t-SVD components $\mathcal{U}$, $\mathcal{S}$ and $\mathcal{V}$ of $\mathcal{A}$.

1. Compute the result of linear transform on $\mathcal{A}$

$\mathcal{A}_L \leftarrow L(\mathcal{A})$.

2. Compute each slice of $\mathcal{U}_L$, $\mathcal{S}_L$ and $\mathcal{V}_L$ from $\mathcal{A}_L$ by

for $i_3 \in \{1, \cdots, n_3\}, \cdots, i_d \in \{1, \cdots, n_d\}$ **do**

$[\mathrm{U}, \mathrm{S}, \mathrm{V}] = \mathrm{svd}(\mathcal{A}_L(:, :, i_3, \cdots, i_d))$,

$\mathcal{U}_L(:, :, i_3, \cdots, i_d) = \mathrm{U}$, $\mathcal{S}_L(:, :, i_3, \cdots, i_d) = \mathrm{S}$, $\mathcal{V}_L(:, :, i_3, \cdots, i_d) = \mathrm{V}$.

end for

3. Compute the result of inverse linear transform on $\mathcal{U}_L$, $\mathcal{S}_L$ and $\mathcal{V}_L$

$\mathcal{U} \leftarrow L^{-1}(\mathcal{U}_L)$, $\mathcal{S} \leftarrow L^{-1}(\mathcal{S}_L)$ and $\mathcal{V} \leftarrow L^{-1}(\mathcal{V}_L)$.

Algorithm 3 PowerMethod $(\mathrm{Z}, \mathrm{R}, \eta)$ [6, 7].

Input: $\mathrm{Z} \in \mathbb{R}^{m \times n}$, $\mathrm{R} \in \mathbb{R}^{n \times k}$, and the number of power iterations η;
Initialize: $\mathrm{Y}_1 = \mathrm{ZR}$;
for $j = 1, 2, \ldots, \eta$ **do**
$\mathrm{Q}_j = qr(\mathrm{Y}_j)$; // qr(·) is QR factorization
$\mathrm{Y}_{j+1} = \mathrm{Z}(\mathrm{Z}^\top \mathrm{Q}_j)$;
end for
return Q_j.

Algorithm 4 GST $\big(s, w, p, J\big)$ [8, 9].

Input: s, w, p, J;
Output: $\Gamma_p^{GST}(s; w)$;
$\delta_p^{GST}(w) = [2w(1-p)]^{\frac{1}{2-p}} + wp[2w(1-p)]^{\frac{p-1}{2-p}}$;
if $|s| \le \delta_p^{GST}(w)$ **then**
$\Gamma_p^{GST}(s; w) = 0$;
else
$j = 0, s^{(k)} = |s|$;
for $j = 0, 1, \cdots, J$ **do**
$s^{(k+1)} = |s| - wp(s^{(j)})^{p-1}$;
$j = j + 1$;
end for
$\Gamma_p^{GST}(s; w) = \mathrm{sgn}(s)s^{(k)}$;
end if

Lemma 2.1. *[9] Let the SVD of* $\mathrm{Y} \in \mathbb{R}^{n_1 \times n_2}$ *be* $\mathrm{Y} = \mathrm{U}\Sigma\mathrm{V}^\top$ *with* $\Sigma = \mathrm{diag}\{\sigma_1, \ldots, \sigma_m\}$*, for any* $\tau \ge 0$ *and* $0 \le w_1 \le w_2 \le \cdots \le w_m$ *(*$m = \min\{n_1, n_2\}$*), a global optimal solution to the optimization problem*

$$\min_{\mathrm{X}} \tau \|\mathrm{X}\|_{w, S_p}^p + \frac{1}{2}\|\mathrm{X} - \mathrm{Y}\|_F^2$$

is given by

$$\mathcal{D}_{\mathbf{w},\tau,p}(\mathrm{Y}) = \mathrm{U} \cdot \mathrm{S}_{\mathbf{w},\tau,p}(\mathrm{X}) \cdot \mathrm{V}^{\mathrm{T}},$$

where $\mathrm{S}_{\mathbf{w},\tau,p}(\mathrm{X}) = \mathrm{diag}\{GST(\sigma_i(\mathrm{X}), \tau w_i, p, J),\ i = 1, \cdots, m\}$.

Algorithm 5 Order-d WTSN proximal operator

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $\tau > 0$, $0 < p < 1$, weight parameter: $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, the number of GST iterations: J and invertible linear transform L.

Output: $\mathcal{D}_{\mathcal{W},p,\tau}(\mathcal{A}) = \mathcal{U} *_L \mathcal{S}_{\mathcal{W},p,\tau} *_L \mathcal{V}^*$.

1. Compute the result of linear transform on $\mathcal{A}$

$\mathcal{A}_L \leftarrow L(\mathcal{A})$.

2. Compute each slice of $\mathcal{A}_L$ by

for $i_3 \in \{1, \cdots, n_3\}, \cdots, i_d \in \{1, \cdots, n_d\}$ **do**

$w = \mathrm{diag}\big(\mathcal{W}(:, :, i_3, \cdots, i_d)\big)$;

$[\mathrm{U}, \mathrm{S}, \mathrm{V}] = \mathrm{svd}\big(\mathcal{A}_L(:, :, i_3, \cdots, i_d)\big)$;

$diagSS = \mathrm{GST}\big(\mathrm{diag}(\mathrm{S}), \tau w, p, J\big)$;

$\mathcal{A}_L(:, :, i_3, \cdots, i_d) = \mathrm{U} \cdot \mathrm{diag}(diagSS) \cdot \mathrm{V}^*$;

end for

3. Compute the result of inverse linear transform on $\mathcal{A}_L$

$\mathcal{D}_{\mathcal{W},p,\tau}(\mathcal{A}) \leftarrow L^{-1}(\mathcal{A}_L)$.

Theorem 2.1. *Let* $m = \min(n_1, n_2)$*, for any* $\tau > 0$ *and* $\mathcal{Z} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$*, then the order-d WTSN proximal operator (8) obeys*

$$\mathcal{D}_{\mathcal{W},p,\tau}(\mathcal{Z}) = \arg\min_{\mathcal{X}} \tau \|\mathcal{X}\|_{\mathcal{W},S_p}^p + \frac{1}{2}\|\mathcal{X} - \mathcal{Z}\|_F^2, \quad (9)$$

if the weight parameter satisfies $0 \le \mathcal{W}(1, 1, i_3, \cdots, i_d) \le \cdots \le \mathcal{W}(m, m, i_3, \cdots, i_d)$*, for any* $i_3 \in \{1, \cdots, n_3\}, \cdots, i_d \in \{1, \cdots, n_d\}$.

Proof. Based on the definition of order-d WTSN, on the one hand, we have

$$\tau \|\mathcal{X}\|_{\mathcal{W},S_p}^p = \frac{1}{\rho} \sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \tau \|\mathcal{X}_L(:, :, i_3, \cdots, i_d)\|_{\mathrm{W}^{(i_3,\cdots,i_p)}, S_p}^p. \quad (10)$$

Utilizing the property (7), on the other hand, we have the following important equations:

$$\|\mathcal{A}\|_F = \frac{1}{\sqrt{\rho}} \|\,\mathrm{bdiag}(\mathcal{A}_L)\|_F, \quad (11)$$

$$\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{\rho} \langle \mathrm{bdiag}(\mathcal{A}_L), \mathrm{bdiag}(\mathcal{B}_L) \rangle, \quad (12)$$

thus

$$\begin{aligned}\frac{1}{2}\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{Z}}\|_F^2 &= \frac{1}{2\rho}\|\operatorname{bdiag}(\boldsymbol{\mathcal{X}}_L)-\operatorname{bdiag}(\boldsymbol{\mathcal{Z}}_L)\|_F^2\\ &= \frac{1}{2\rho}\sum_{i_3=1}^{n_3}\cdots\sum_{i_d=1}^{n_d}\|\boldsymbol{\mathcal{X}}_L(:,:,i_3,\cdots,i_d)\\ &\quad -\boldsymbol{\mathcal{Z}}_L(:,:,i_3,\cdots,i_d)\|_F^2. \end{aligned} \tag{13}$$

Then, the problem (9) is equivalent to

$$\begin{aligned}\arg\min_{\boldsymbol{\mathcal{X}}}\frac{1}{\rho}\sum_{i_3=1}^{n_3}\cdots\sum_{i_d=1}^{n_d}(\tau\|\boldsymbol{\mathcal{X}}_L(:,:,i_3,\cdots,i_d)\|^p_{\mathrm{W}^{(i_3,\cdots,i_p)},S_p}\\ +\frac{1}{2}\|\boldsymbol{\mathcal{X}}_L(:,:,i_3,\cdots,i_d)-\boldsymbol{\mathcal{Z}}_L(:,:,i_3,\cdots,i_d)\|_F^2).\end{aligned} \tag{14}$$

Since $0\leq\boldsymbol{\mathcal{W}}(1,1,i_3,\cdots,i_d)\leq\cdots\leq\boldsymbol{\mathcal{W}}(m,m,i_3,\cdots,i_d)$, the $(i_3,\cdots,i_d)$-th matrix slice of $L(\boldsymbol{\mathcal{D}}_{\boldsymbol{\mathcal{W}},p,\tau}(\boldsymbol{\mathcal{Z}}))$ solves the $(i_3,\cdots,i_d)$-th subproblem of (14) through Lemma 2.1. Hence, $\boldsymbol{\mathcal{D}}_{\boldsymbol{\mathcal{W}},p,\tau}(\boldsymbol{\mathcal{Z}})$ solves the problem (9). □

3. REFERENCES

[1] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[2] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Alg. Appl.*, vol. 435, no. 3, pp. 641–658, 2011.

[3] 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 J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.

[4] C. D. Martin, R. Shafer, and B. LaRue, "An order-p tensor factorization with applications in imaging," *SIAM J. Sci. Comput.*, vol. 35, no. 1, pp. A474–A490, 2013.

[5] E. Kernfeld, M. Kilmer, and S. Aeron, "Tensor–tensor products with invertible linear transforms," *Linear Alg. Appl.*, vol. 485, pp. 545–570, 2015.

[6] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011.

[7] Q. Yao and J. T. Kwok, "Accelerated and inexact soft-impute for large-scale matrix and tensor completion," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 9, pp. 1665–1679, 2018.

[8] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 217–224.

[9] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, 2016.