A t-SVD-based Nuclear Norm with Imaging Applications

Oguz Semerci\textsuperscript{1} \quad Ning Hao\textsuperscript{2} \quad Misha E. Kilmer\textsuperscript{2} \quad Eric Miller\textsuperscript{2} \\
Shuchin Aeron\textsuperscript{2} \quad Gregory Ely\textsuperscript{2} \quad Zemin Zhang\textsuperscript{2}

\textsuperscript{1}Schlumberger-Doll Research \\
\textsuperscript{2}Tufts University

Kilmer and Hao’s work supported by NSF-DMS 0914957
\( \mathcal{A}_{(i,j,k)} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k \)
\( \mathcal{A}_{(i,j,k)} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k \)
$\mathcal{A}_{(i,j,k)} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k$
\( \mathcal{A}_{(i,j,k)} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k \)
Motivation

The application drives the choice of factorization (e.g. CP, Tucker) and the constraints. Today we are concerned with imaging applications, orientation dependent.

Talk builds on:

- Closed multiplication operation between two tensors, factorizations reminiscent of matrix factorizations [K., Martin, Perrone, 2008; K., Martin 2010; Martin et al, 2012].
- View of Third order tensors as operators on matrices, [K., Braman, Hoover, Hao, 2013]
Toward Defining Tensor-Tensor Multiplication

For $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$, let $A_i = \mathcal{A}_{:, :, i}$. 

$\text{unfold} \rightarrow \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_n \end{bmatrix} \in \mathbb{R}^{mn \times p}$
Toward Defining Tensor-Tensor Multiplication

For $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$, let $A_i = \mathcal{A}_{:, :, i}$.
The block circulant matrix generated by $\text{unfold}(\mathcal{A})$ is

$$
circ(\mathcal{A}) = 
\begin{bmatrix}
A_1 & A_n & \cdots & A_3 & A_2 \\
A_2 & A_1 & A_n & \cdots & \cdots \\
A_3 & A_2 & \ddots & \ddots & \ddots \\
\vdots & \vdots & \ddots & \ddots & \ddots \\
A_n & \cdots & \cdots & A_2 & A_1 
\end{bmatrix}
$$
A block circulant can be block-diagonalized by a (normalized) DFT in the 2nd dimension:

\[(F \otimes I) \text{circ}(A)(F^* \otimes I) = \begin{bmatrix}
\hat{A}_1 & 0 & \cdots & 0 \\
0 & \hat{A}_2 & 0 & \cdots \\
0 & \cdots & \ddots & 0 \\
0 & \cdots & 0 & \hat{A}_n
\end{bmatrix}\]

Conveniently, an FFT along tube fibers of \(A\) gives \(\hat{A}\).
[K., Martin, Perrone ‘08]: For $\mathbf{A} \in \mathbb{R}^{m \times p \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times q \times n}$, define the t-product

$$\mathbf{A} \ast \mathbf{B} \equiv \text{fold}\left(\text{circ}\left(\mathbf{A}\right) \cdot \text{unfold}\left(\mathbf{B}\right)\right).$$

Result is $m \times q \times n$. 
[K., Martin, Perrone ‘08]: For $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$ and $\mathcal{B} \in \mathbb{R}^{p \times q \times n}$, define the t-product

$$\mathcal{A} \ast \mathcal{B} \equiv \text{fold} \left( \text{circ} \left( \mathcal{A} \right) \cdot \text{unfold} \left( \mathcal{B} \right) \right).$$

Result is $m \times q \times n$.

Example: $\mathcal{A} \in \mathbb{R}^{m \times p \times 3}$ and $\mathcal{B} \in \mathbb{R}^{p \times q \times 3}$,

$$\mathcal{A} \ast \mathcal{B} = \text{fold} \left( \begin{bmatrix} A_1 & A_3 & A_2 \\ A_2 & A_1 & A_3 \\ A_3 & A_2 & A_1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \right).$$
[K., Martin, Perrone ‘08]: For $A \in \mathbb{R}^{m \times p \times n}$ and $B \in \mathbb{R}^{p \times q \times n}$, define the t-product

$$A \ast B \equiv \text{fold} \left( \text{circ} (A) \cdot \text{unfold} (B) \right).$$

Result is $m \times q \times n$.

Example: $A \in \mathbb{R}^{m \times p \times 3}$ and $B \in \mathbb{R}^{p \times q \times 3}$,

$$A \ast B = \text{fold} \left( \begin{bmatrix} A_1 & A_3 & A_2 \\ A_2 & A_1 & A_3 \\ A_3 & A_2 & A_1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \right).$$

This tensor-tensor multiplication generalizes to higher-order tensors through recursion - see Martin et al, 2012.
A $1 \times 1 \times n$ tensor is called a tubal scalar. The t-product between tubal scalars is commutative $\Rightarrow$ the t-product resembles matrix-matrix product with scalar mult replaced by t-product mult among tubal scalars.

The $\ell \times \ell \times n$ identity tensor $I$ is the tensor whose frontal slice is the $\ell \times \ell$ identity matrix, and whose other frontal slices are all zeros.
Transpose

**Definition**

If $\mathcal{A}$ is $\ell \times m \times n$, then $\mathcal{A}^T$ is the $m \times \ell \times n$ tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through $n$.

**Example**

If $\mathcal{A} \in \mathbb{R}^{\ell \times m \times 4}$

\[
\mathcal{A}^T = \text{fold} \left( \begin{bmatrix} A_1^T \\ A_4^T \\ A_3^T \\ A_2^T \end{bmatrix} \right)
\]
Orthogonality

**Definition**

$\mathbf{U} \in \mathbb{R}^{m \times m \times n}$ is orthogonal if $\mathbf{U}^T \ast \mathbf{U} = \mathbf{I} = \mathbf{U} \ast \mathbf{U}^T$.

Can show Frobenius norm invariance: $\| \mathbf{U} \ast \mathbf{A} \|_F = \| \mathbf{A} \|_F$. 
The t-SVD

**Theorem (K. and Martin, 2011)**

Let $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$. Then $\mathcal{A}$ can be factored as

$$\mathcal{A} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^T$$

where $\mathcal{U}, \mathcal{V}$ are orthogonal $\ell \times \ell \times n$ and $m \times m \times n$, and $\mathcal{S}$ is an $\ell \times m \times n$ f-diagonal tensor. Also, $\mathcal{B} = \mathcal{U}_{1:k,1:k,:} \ast \mathcal{S}_{1:k,1:k,:} \ast \mathcal{V}_{1:k,1:k,:}^T$ satisfies

$$\mathcal{B} = \arg \min_M \| \mathcal{A} - \mathcal{B} \|_F, \quad M = \{ \mathcal{B} = \mathcal{X} \ast \mathcal{Y}, \mathcal{X} \in \mathbb{R}^{\ell \times k \times n}, \mathcal{Y} \in \mathbb{R}^{k \times m \times n} \}.$$
Let $A$ be $2 \times 2 \times 2$.

\[
(F \otimes I) \text{circ} (A) (F^* \otimes I) = \begin{bmatrix}
\hat{A}_1 & 0 \\
0 & \hat{A}_2
\end{bmatrix}
\]

\[
\begin{bmatrix}
\hat{A}_1 & 0 \\
0 & \hat{A}_2
\end{bmatrix} = \begin{bmatrix}
\hat{U}_1 & 0 \\
0 & \hat{U}_2
\end{bmatrix} \begin{bmatrix}
\hat{\sigma}_1^{(1)} & 0 \\
0 & \hat{\sigma}_2^{(1)}
\end{bmatrix} \begin{bmatrix}
\hat{\sigma}_1^{(2)} & 0 \\
0 & \hat{\sigma}_2^{(2)}
\end{bmatrix} \begin{bmatrix}
\hat{V}_1^* & 0 \\
0 & \hat{V}_2^*
\end{bmatrix}
\]

The $U$, $S$, $V^T$ are formed by putting the hat matrices as frontal slices, ifft along tubes.

e.g. $s_1 = \begin{bmatrix}
\hat{\sigma}_1^{(1)} \\
\hat{\sigma}_1^{(2)}
\end{bmatrix}$ oriented into the screen.
Definition (K., Braman, Hoover, Hao, 2013)

Let $\mathcal{A} \in \mathbb{R}^\ell \times m \times n$. The multi-rank of $\mathcal{A}$ is a length $n$ vector consisting of the ranks of all the $\mathcal{A}^{(i)}$, which must be symmetric about the “middle”.

Example

$\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 4}$, multi-rank possible: $[i, j, k, j]^T$, $1 \leq i, j, k \leq 2$.

Example

$\mathcal{A} \in \mathbb{R}^{5 \times 4 \times 3}$, multi-rank possible: $[i, j, j]^T$, $1 \leq i \leq 4, 1 \leq j \leq 4$. 

Misha E. Kilmer (Tufts University)  Tensor Nuclear Norm  June 2013 13 / 28
If $A$ is an $\ell \times m$, $\ell \geq m$ matrix with singular values $\sigma_i$, the nuclear norm $\|A\|_\diamondsuit = \sum_{i=1}^{m} \sigma_i$.

However, in the t-SVD, we have singular tubes (the entries of which need not be positive), which sum to a singular tube!

The entries in the $j$th singular tube are the inverse Fourier coefficients of the length-$n$ vector of the $j$th singular values of $\hat{A}_{:,i,i}, i = 1..n$.

**Definition**

For $A \in \mathbb{R}^{\ell \times m \times n}$, our tensor nuclear norm is

$$\|A\|_\diamondsuit = \sum_{i=1}^{\min(\ell,m)} \|\sqrt{n}F s_i\|_1 = \sum_{i=1}^{\min(\ell,m)} \sum_{j=1}^{n} \hat{S}_{i,i,j}.$$  (Same as the matrix nuclear norm of $\text{circ}(A)$.)
Theorem (Semerci, Hao, Kilmer, Miller)

The tensor nuclear norm is a valid norm.

Since the t-SVD extends to higher-order tensors [Martin et al, 2012], the norm does, as well.
Yes, the t-SVD is orientation dependent, as is the norm. There are applications where this is particularly useful!

- Collection of “structurally similar” $m \times n$ images
- Video frames (3D, 4D= color); completing missing data
$k$ energy bins.

$\mu(\mathbf{r}, E_k) \rightarrow X_k \in \mathbb{R}^{N_1 \times N_2}$

$x_k = \text{vec}(X_k)$

$(\phi, t)$ space into $N_m$ source/det pairs

Then $\mathbf{A} \in \mathbb{R}^{N_m \times N_p}$ where $[\mathbf{A}]_{ij}$ represents the length of that segment of ray $i$ passing through pixel $j$. 
The log-likelihood function to be optimized, assuming Poisson noise

\[ L_k(x_k) = \| D_k^{-1/2}(Ax_k - m_k) \|_2^2 \]

where \( D_k \) is diagonal, \( m_k \) is log of scaled projection data.
Regularized Problem

Let $X_{s.t.}X_{:,k} = X_k$; recall $x_k = \text{vec}(X_k)$.

$$\min_{X} \left( \sum_{k=1}^{N_3} L_k(x_k) + \alpha_k R(x_k) \right) + \gamma \|Z\|_*, \text{ sbj to } Z = X$$

where $R()$ denotes possible additional regularization.

Optimized via an Alternating Direction Method of Multipliers (ADMM) (see Boyd et al 2011, Bertsekas 1999)

Let $L_\eta(X, Z, Y)$ denote the augmented Lagrangian. The updates are

$$X^{n+1} := \arg\min_X L_\eta(X, Z^n, Y^n),$$

$$Z^{n+1} := \arg\min_Z L_\eta(X^{n+1}, Z, Y^n)$$

$$Y^{n+1} := Y^n + \eta(X^{n+1} - Z^{n+1})$$
\[ \mathcal{Z}^{n+1} := \arg\min_{\mathcal{Z}} L_\eta (\mathcal{X}^{n+1}, \mathcal{Z}, \mathcal{Y}^n) \]

Compute \( (\frac{1}{\eta} \mathcal{Y}^n + \mathcal{X}^{n+1}) = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T \). Recall \( \hat{\mathbf{S}} \) contains \( \hat{\Sigma}_k \) for frontal slices of transformed tensor.

\[ \mathcal{Z}^{n+1} := \mathbf{U} \ast \rho(\mathbf{S}) \ast \mathbf{V}^T, \]  where \( \rho(\mathbf{S}) \) takes the non-zeros in \( \hat{\mathbf{S}} \) and replaces them with the difference between them and rho, if that difference is greater than 0, and 0 otherwise. \( \rho \) depends on the parameter in the problem. 

Shown that \( \mathcal{Z}^{n+1} := \mathbf{U} \ast (\mathbf{S} \ast \mathbf{D}) \ast \mathbf{V}^T \)
Numerical Results

FBP, TNN, TV, TNN+TV

Figure: Reconstruction results for 25 keV.
Numerical Results

Figure: Reconstruction results for 85 keV.

Misha E. Kilmer (Tufts University)
Tensor Nuclear Norm
June 2013
Tensor Completion

Given unknown tensor $\mathcal{M}$ of size $n_1 \times n_2 \times n_3$, given a subset of entries $\{\mathcal{M}_{ijk} : (i, j, k) \in \Omega\}$ where $\Omega$ is an indicator tensor of size $n_1 \times n_2 \times n_3$. Recover the entire $\mathcal{M}$:

$$\min \|X\| \mathcal{\Psi} \quad \text{subject to } P_\Omega(X) = P_\Omega(\mathcal{M})$$

The $(i, j, k)_{th}$ component of $P_\Omega(X)$ is equal to $\mathcal{M}_{ijk}$ if $(i, j, k) \in \Omega$ and zero otherwise.

Similar to the previous problem, this can be solved by ADMM, with 3 update steps, one which decouples, one that is a shrinkage/thresholding step.
Numerical Results


with thanks to Dr. Amit Agrawal
Numerical Results

![Graph 1](image1)

![Graph 2](image2)
Numerical Results

Basketball video data of size $144 \times 256 \times 3 \times 80$. 

![Image 1](image1.png) ![Image 2](image2.png) ![Image 3](image3.png)
Introduced the notion of a tensor nuclear norm around the concept of the t-SVD for tensors
Discussed in 3rd order case, but generalizes to higher order
The tensor nuclear norm is useful in imaging applications where we can exploit certain features that are orientation dependent
Efficiency in implementations (parallelism); exploit complex conjugacy in the Fourier domain
A different t-SVD and associated factorizations and norm based on fast trig-transform [Kernfeld, et al, 2013]