CSE 392/CS 395T/M 397C: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin

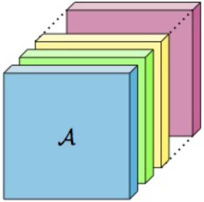
Spring 2025

Lecture 22: t-SVD, $\star_M$-product

Outline

1. t-SVD
2. $\star_M$-product

Recall: The t-product



$$\xrightarrow{\texttt{circ}(\mathcal{A})} \begin{pmatrix} A^{(1)} & A^{(n)} & \dots & A^{(3)} & A^{(2)} \\ A^{(2)} & A^{(1)} & A^{(n)} & \dots & A^{(3)} \\ A^{(3)} & A^{(2)} & A^{(1)} & \ddots & \dots \\ \vdots & \ddots & \ddots & \ddots & A^{(n)} \\ A^{(n)} & \dots & A^{(3)} & A^{(2)} & A^{(1)} \end{pmatrix}$$

The t-product is defined as:

$$\mathcal{A} * \mathcal{B} = \texttt{fold}(\texttt{circ}(\mathcal{A}) \cdot \texttt{unfold}(\mathcal{B})).$$

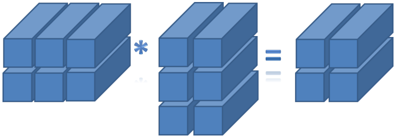
It is obvious that if $\mathcal{A}$ is $m \times p \times n$, need $\mathcal{B}$ to be $p \times k \times n$, and the result is $m \times k \times n$.

Kilmer, Martin, **Factorization Strategies for Third-Order Tensors**, LAA, 2011

T-product

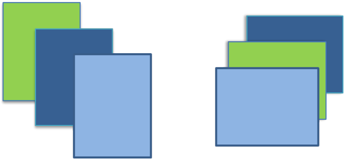
Block circulants block-diagonalized via 1D FFTs $\Rightarrow$ The t-product can be computed in-place using FFTs:

- $\widehat{\mathcal{A}} \leftarrow \texttt{fft}(\mathcal{A}, [\,], 3)$
- $\widehat{\mathcal{B}} \leftarrow \texttt{fft}(\mathcal{B}, [\,], 3)$
- $\widehat{\mathcal{C}}_{:,:,i} = \widehat{\mathcal{A}}_{:,:,i} \cdot \widehat{\mathcal{B}}_{:,:,i}, \; i = 1, \ldots, n$
- $\mathcal{C} = \texttt{ifft}(\widehat{\mathcal{C}}, [\,], 3)$



Transpose and Orthogonality

$\mathcal{A} \in \mathbb{R}^{\ell \times m \times n} \Rightarrow \mathcal{A}^\top \in \mathbb{R}^{m \times \ell \times n}$ is obtained by transposing each frontal slice & reversing order of transposed frontal slices 2 through n.



$\mathcal{U} \in \mathbb{R}^{m \times m \times n}$ is **orthogonal** if $\mathcal{U}^\top * \mathcal{U} = \mathcal{I} = \mathcal{U} * \mathcal{U}^\top$.

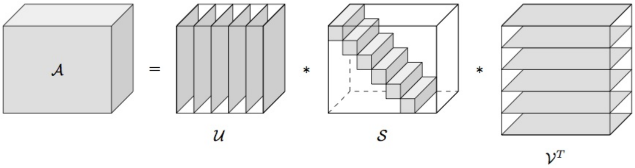
Can show **Frobenius norm invariance**: $\|\mathcal{U} * \mathcal{A}\|_F = \|\mathcal{A}\|_F$.
Exercise: show $(\mathcal{A} * \mathcal{B})^\top = \mathcal{B}^\top * \mathcal{A}^\top$

t-SVD

Theorem: For $\mathcal{A} \in \mathbb{R}^{m\times\ell\times n}$ there exists a full tensor-SVD

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

with $m \times m \times n$ orthogonal tensor $\mathcal{U}$, $\ell \times \ell \times n$ orthogonal tensor $\mathcal{V}$, and $m \times \ell \times n$ f-diagonal tensor $\mathcal{S}$ ordered such that the singular tubes $\mathbf{s}_i = \mathcal{S}_{i,i,:}$ have $\|\mathbf{s}_1\|_F^2 \geq \|\mathbf{s}_2\|_F^2 \geq \cdots$.



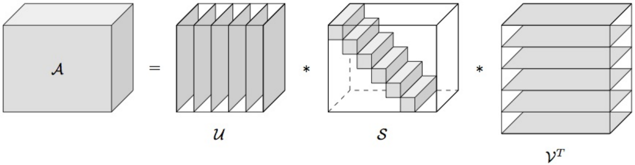
The **t-rank** is the number of non-zero tube-fibers in $\mathcal{S}$.

t-SVD

Theorem: For $\mathcal{A} \in \mathbb{R}^{m\times \ell \times n}$ there exists a full tensor-SVD

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

with $m \times m \times n$ orthogonal tensor $\mathcal{U}$, $\ell \times \ell \times n$ orthogonal tensor $\mathcal{V}$, and $m \times \ell \times n$ f-diagonal tensor $\mathcal{S}$ ordered such that the singular tubes $\mathbf{s}_i = \mathcal{S}_{i,i,:}$ have $\|\mathbf{s}_1\|_F^2 \geq \|\mathbf{s}_2\|_F^2 \geq \cdots$.



Exercise: Prove the claim that the $\|\mathbf{s}_i\|_F^2$ are non-increasing.

t-SVD Computation

The t-SVD can be computed efficiently (in parallel) by moving to the Fourier domain.

- Compute $\widehat{\mathcal{A}}$
- For $i = 1, \ldots, n$, find matrix SVD of each frontal slice: $\widehat{\mathcal{U}}_{:,:,i}\widehat{\mathcal{S}}_{:,:,i}\widehat{\mathcal{V}}^H_{:,:,i} = \widehat{\mathcal{A}}_{:,:,i}$
- To get $\mathcal{U}, \mathcal{S}, \mathcal{V}$, inverse FFT along tube fibers of $\widehat{\mathcal{U}}, \widehat{\mathcal{S}}, \widehat{\mathcal{V}}$.

t-SVD and Optimality in Truncation

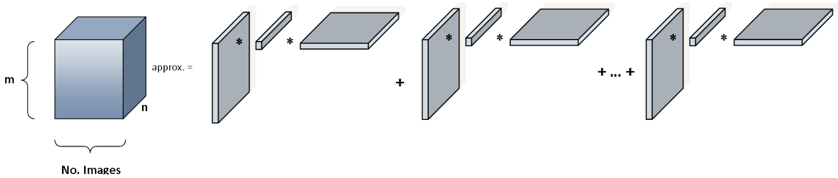
$\mathcal{A} \in \mathbb{R}^{m \times p \times n}$. For $k < \min(m, p)$, define

$$\mathcal{A}_k = \sum_{i=1}^{k} \mathcal{U}_{:,i,:} * \left(\mathcal{S}_{i,i,:} * \mathcal{V}_{:,i,:}^\top \right) = \mathcal{U}_k * (\mathcal{S}_k * \mathcal{V}_k^\top)$$

Then

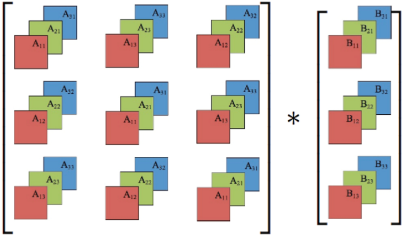
$$\mathcal{A}_k = \arg\min_{\widetilde{\mathcal{A}} \in \Omega} \|\mathcal{A} - \widetilde{\mathcal{A}}\|$$

where $\Omega = \{\mathcal{X} * \mathcal{Y} \,|\mathcal{X} \in \mathbb{R}^{m \times k \times n}, \mathcal{Y} \in \mathbb{R}^{k \times p \times n}\}$



The t-product, and the t-SVD, generalize to higher dimensions through recursion[1].

$$\begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_3 & \mathcal{A}_2 \\ \mathcal{A}_2 & \mathcal{A}_1 & \mathcal{A}_3 \\ \mathcal{A}_3 & \mathcal{A}_2 & \mathcal{A}_1 \end{bmatrix} * \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \\ \mathcal{B}_3 \end{bmatrix}$$



Treatment of change of pose or lighting information (as motion) $\rightarrow$ 4D.

[1] Martin, Shafer, LaRue, An Order-p Tensor Factorization with Applications in Imaging, SISC, 2013

Generalization?

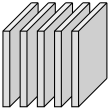
Is the DFT the only possibility for defining this matrix-mimetic type of framework, or are there other possibilities?

Generalization?

Is the DFT the only possibility for defining this matrix-mimetic type of framework, or are there other possibilities?

Now we will show: Whole family of options of tensor-tensor products for which this is possible! Offers the option of tailoring the product to the type of data or operator at hand!

Recall Mode-3 Multiplication



$m \times p \times n$ tensor$\mathcal{A}$

Let $\mathbf{M}$ be $r \times n$. To find $\mathcal{A} \times_3 \mathbf{M}$:

- Compute matrix-matrix product $\mathbf{M}\mathcal{A}_{(3)}$,
- Reshape the result to an $m \times p \times r$ tensor.

Equivalent to applying $\mathbf{M}$ along tube fibers.

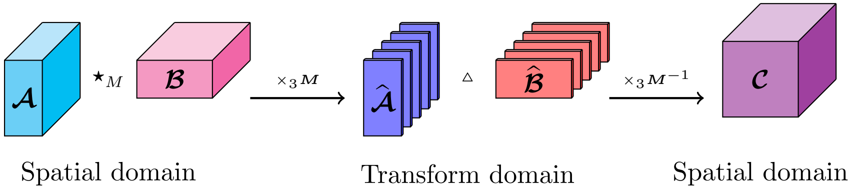
Star-M Product

Let $\mathbf{M}$ be any invertible, $n \times n$ matrix. Then

$$\widehat{\mathcal{A}} := \mathcal{A} \times_3 \mathbf{M} \text{ so that } \mathcal{A} = \widehat{\mathcal{A}} \times_3 \boldsymbol{M}^{-1}.$$

Definition

Given any invertible, $n \times n$ $\boldsymbol{M}$, $\mathcal{A} \in \mathbb{C}^{m \times p \times n}$ and $\mathcal{B} \in \mathbb{C}^{p \times \ell \times n}$, $\mathcal{C} = \mathcal{A} \star_M \mathcal{B}$ is defined via $\widehat{\mathcal{C}}_{:,:,i} = \widehat{\mathcal{A}}_{:,:,i} \widehat{\mathcal{B}}_{:,:,i}$.



Kernfeld, Kilmer, Aeron, LAA 2015

Special Case

If $\mathbf{M}$ is the (unnormalized) DFT matrix, we recover the t-product framework!

Other Properties

Definition (Conjugate Transpose)

Given $\mathcal{A} \in \mathbb{C}^{m \times p \times n}$ its $p \times m \times n$ **conjugate transpose** under $\star_M$ $\mathcal{A}^{\mathrm{H}}$ is defined such that $(\widehat{\mathcal{A}^{\mathrm{H}}})^{(i)} = (\widehat{\mathcal{A}}^{(i)})^{\mathrm{H}}, \qquad i = 1, \ldots, n.$

Definition (Unitary/Orthogonal Tensors)

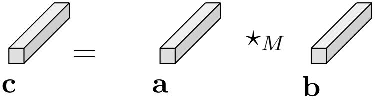
$\mathcal{Q} \in \mathbb{C}^{m \times m \times n}$ ($\mathcal{Q} \in \mathbb{R}^{m \times m \times n}$) is called $\star_M$-unitary ($\star_M$-orthgonal) if

$$\mathcal{Q}^{\mathrm{H}} \star_M \mathcal{Q} = \mathcal{I} = \mathcal{Q} \star_M \mathcal{Q}^{\mathrm{H}},$$

where H is replaced by transpose for real tensors. Note that $\mathcal{I}$ also defined under $\star_M$.

Kernfeld, Kilmer, Aeron, LAA 2015

Entry-wise $\mathbf{M}$-product

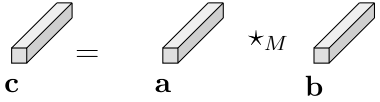


Tube fiber interpretation:

$$
\begin{aligned}
\mathbf{c} &= \texttt{fold}\left((\mathbf{M}^{-1}\mathrm{diag}(\hat{\mathbf{a}})\mathbf{M})\texttt{vec}(\mathbf{b})\right) \\
&= \texttt{fold}\left((\mathbf{M}^{-1}\mathrm{diag}(\hat{\mathbf{b}})\mathbf{M})\texttt{vec}(\mathbf{a})\right)
\end{aligned}
$$

Commutativity, and characterization using set of diagonal matrices diagonalized by $\mathbf{M}$ and its inverse.

Entry-wise $\mathbf{M}$-product



Tube fiber interpretation:

$$\begin{aligned}\mathbf{c} &= \texttt{fold}\left((\mathbf{M}^{-1}\mathrm{diag}(\hat{\mathbf{a}})\mathbf{M})\texttt{vec}(\mathbf{b})\right)\\ &= \texttt{fold}\left((\mathbf{M}^{-1}\mathrm{diag}(\hat{\mathbf{b}})\mathbf{M})\texttt{vec}(\mathbf{a})\right)\end{aligned}$$

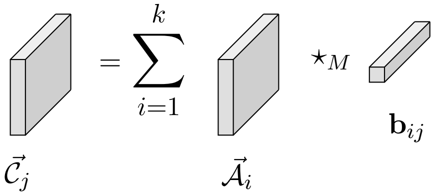
Commutativity, and characterization using set of diagonal matrices diagonalized by $\mathbf{M}$ and its inverse.

Special Case: $\boldsymbol{M}$ is DFT $\Rightarrow$ convolution, circulant matrices

Matrix-mimeticity

<u>Observation</u>: overloading scalar products with $\star_M$ in matrix-matrix algorithms gives product for larger dimensional tensors.

If $\mathcal{A}$ is $m \times k \times n$ and $\mathcal{B}$ is $k \times p \times n$, then $\mathcal{C}$ is $m \times p \times n$, and



$$j = 1, \ldots, p$$

Unitary Invariance

Theorem

If $\boldsymbol{M}$ *a non-zero multiple of a unitary/orthogonal matrix*[a]

$$\|\mathcal{Q} \star_M \mathcal{A}\|_F = \|\mathcal{A}\|_F$$

[a] Kilmer, Horesh, Avron, Newman (2021)

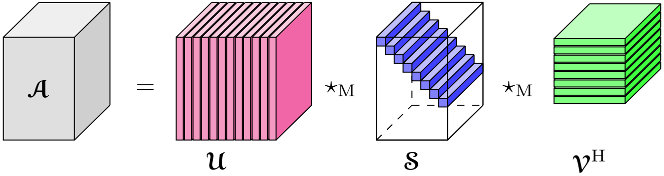
Tensor-tensor SVDs

Theorem (Kilmer, Horesh, Avron, Newman)

Let $\mathcal{A}$ be a $m \times p \times n$ tensor and $\boldsymbol{M}$ a non-zero multiple of a unitary/orthogonal matrix. The (full) $\star_M$ tensor SVD (t-SVDM) is

$$\mathcal{A} = \mathcal{U} \star_M \mathcal{S} \star_M \mathcal{V}^{\mathrm{H}} = \sum_{i=1}^{\min(m,p)} \mathcal{U}_{:,i,:} \star_M \mathcal{S}_{i,i,:} \star_M \mathcal{V}^{\mathrm{H}}_{:,i,:}$$

with $\mathcal{U}$, $\mathcal{V}$ $\star_M$-unitary, & $\|\mathcal{S}_{1,1,:}\|_F^2 \geq \|\mathcal{S}_{2,2,:}\|_F^2 \geq \dots$



Algorithm

$\widehat{\mathcal{A}} \leftarrow \mathcal{A} \times_M \boldsymbol{M}$
$i = 1, \ldots, n$
$[\widehat{\mathcal{U}}_{:,:,i}, \widehat{\mathcal{S}}_{:,:,i}, \widehat{\mathcal{V}}_{:,:,i}] = \texttt{svd}(\widehat{\mathcal{A}}_{:,:,i})$
$\mathcal{U} = \widehat{\mathcal{U}} \times_3 \boldsymbol{M}^{-1},\ \mathcal{S} = \widehat{\mathcal{S}} \times_3 \boldsymbol{M}^{-1},\ \mathcal{V} = \widehat{\mathcal{V}} \times_3 \boldsymbol{M}^{-1}.$

Perfectly (i.e. embarrassingly) parallelizable!

For face i, exist singular values $\hat{\sigma}_i^{(j)}$, $j = 1, .., \rho_i$

Eckart-Young

$\mathcal{A} \in \mathbb{R}^{m\times p\times n}$. For $k < \min(m,p)$, and $\boldsymbol{M}$ as previously, define

$$\mathcal{A}_k = \sum_{i=1}^{k} \mathcal{U}_{:,i,:} \star_M \left(\mathcal{S}_{i,i,:} \star_M \mathcal{V}_{:,i,:}^\top\right)$$

Then

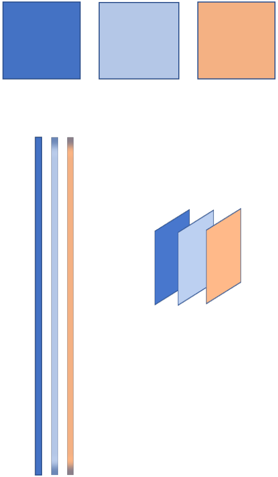
$$\mathcal{A}_k = \arg\min_{\widetilde{\mathcal{A}}\in\Omega} \|\mathcal{A} - \widetilde{\mathcal{A}}\|_F$$

where $\Omega = \{\mathcal{X} \star_M \mathcal{T} \mid \mathcal{X} \in \mathbb{R}^{m\times k\times n}, \mathcal{T} \in \mathbb{R}^{k\times p\times n}\}$

Error: $\|\mathcal{A} - \mathcal{A}_k\|_F^2 = \sum_{j>k} \|\mathcal{S}_{j,j}\|_F^2 = c\sum_{i=1}^{n}\sum_{j>k} \hat{\sigma}_j^{(i)}$, c depends on $\boldsymbol{M}$.

Data Comparison

In general, consider J pieces of 2D, $m \times n$ data. Storage as $mn \times J$ matrix $\boldsymbol{A}$ or $m \times J \times n$ tensor $\mathcal{A}$. Which is more compressible?



Theoretical Result

Theorem (Kilmer, Horesh, Avron, Newman (2021))

Suppose $\mathcal{A}_k$ is optimal k-term t-SVDM approximation to $\mathcal{A}$, and let $\boldsymbol{A}_k$ is optimal k-term matrix SVD approximation to $\boldsymbol{A}$. Then

$$\|\mathcal{A} - \mathcal{A}_k\|_F \leq \|\boldsymbol{A} - \boldsymbol{A}_k\|_F,$$

where **strict inequality is achievable**.

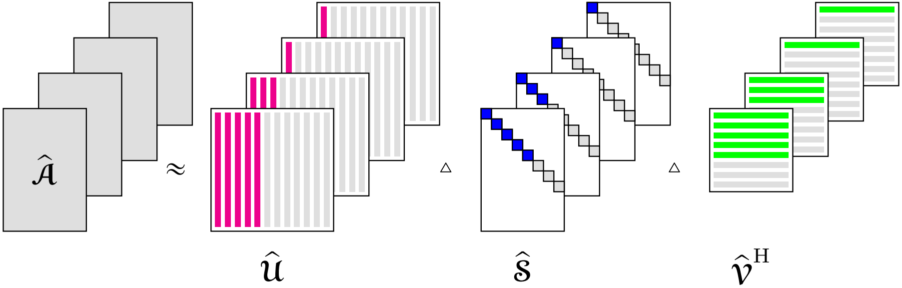
- Result works for any $\boldsymbol{M}$ that is multiple of unitary (orthogonal) matrix.
- Why? Takes advantage of latent structure in data.

t-SVDMII

Truncated t-SVDM ignores relative importance of faces.

Global approach: order $\hat{\sigma}_i^{(j)} := \widehat{\mathcal{S}}_{i,i,j}$, truncate on energy level.

Gives $\mathcal{A}_\rho$, with $\rho_i = \text{rank}(\widehat{\mathcal{A}}^{(i)})$



Comparison

Implicit rank = total number of non-zero $\hat{\sigma}_i^{(j)}$.

Theorem (Kilmer,Horesh,Avron,Newman, 2021)

Let $\mathcal{A}_k$ be the t-SVDM t-rank k approximation to $\mathcal{A}$, and suppose its implicit rank is r. Define $\mu = \|\mathcal{A}_k\|_F^2 / \|\mathcal{A}\|_F^2$. There exists $\gamma \leq \mu$ such that the t-SVDMII approximation, $\mathcal{A}_\rho$, obtained for this γ, has implicit rank less than or equal to the implicit rank of $\mathcal{A}_k$ and

$$\|\mathcal{A} - \mathcal{A}_\rho\|_F \leq \|\mathcal{A} - \mathcal{A}_k\|_F \leq \|\boldsymbol{A} - \boldsymbol{A}_k\|_F.$$

Summary

- Matrix Mimetic properties make $\star_M$ framework desirable - extensions of traditional matrix-based algorithms are possible
- Orientation dependent approach (not blackbox)
- Theoretical analysis comparing to matrix-based and other tensor based approaches is now possible, in third order.
- Algorithmic extensions to higher-order, but theory?
- Exercise Sequential t-SVD – what might this look like?
- Randomized methods more directly applicable.
- $\mathbf{M}$ learned/tailored to data

Matlab Demo