CSE 392: Matrix and Tensor Algorithms for Data Spring 2024

Lecture 22 — 2024.04.08

Instructor: Shashanka Ubaru *Scribe: Tong Wang*

1 t-SVD

Theorem: For any $\mathcal{A} \in \mathbb{R}^{m \times l \times n}$, there exists a full tensor-SVD such that

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

with $\mathcal{U}$ being an $m \times m \times n$ orthogonal tensor, $\mathcal{V}$ being an $l \times l \times n$ orthogonal tensor, and $\mathcal{S}$ an $m \times l \times n$ f-diagonal tensor ordered such that the singular tubes $s_i = S_{i,i,:}$ have $\|s_1\|_F^2 \geq \|s_2\|_F^2 \geq \cdots$. The **t-rank** is the number of non-zero tube-fibers in $\mathcal{S}$.

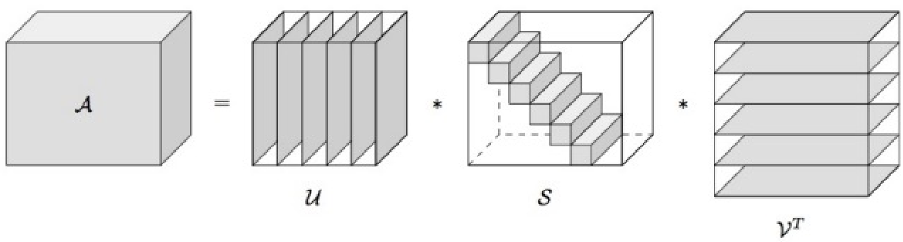

Figure 1: Tensor SVD formulation.

1.1 t-SVD computation

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

- Compute $\hat{\mathcal{A}}$ using FFT;
- For $i = 1, \ldots, n$, find the matrix SVD of each frontal slice: $\hat{\mathcal{U}}_{:,:,i} \hat{\mathcal{S}}_{:,:,i} \hat{\mathcal{V}}_{:,:,i}^H = \hat{\mathcal{A}}_{:,:,i}$;
- To get $\mathcal{U}$, $\mathcal{S}$ and $\mathcal{V}$, just apply the inverse FFT along tube fibers of $\hat{\mathcal{U}}$, $\hat{\mathcal{S}}$ and $\hat{\mathcal{V}}$.

1.2 t-SVD and optimality in truncation

Optimality. Let $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$ be a tensor. For $k < \min(m, p)$, we can define

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

then

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

where $\Omega = \{\mathcal{X} * \mathcal{Y} | \mathcal{X} \in \mathbb{R}^{m \times k \times n}, \mathcal{Y} \in \mathbb{R}^{k \times p \times n}\}$. This states the optimality of t-SVD.

1.3 Generalization to higher dimensions

The t-product, and the t-SVD, can generalize to higher dimensions through recursion.

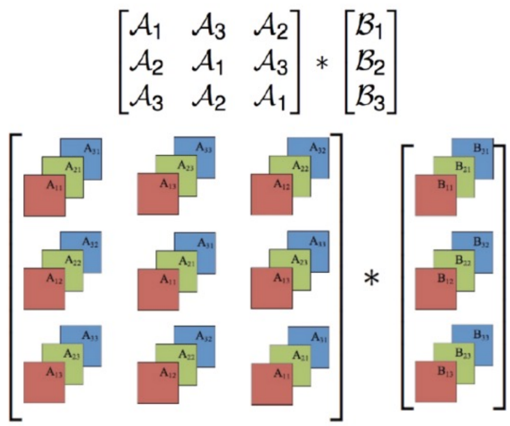

Figure 2: Generalization to higher dimensions through recursion.

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

2 $\star_M$ Product

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

2.1 Recall Mode-3 Multiplication

Let $\mathbf{M}$ be a $r \times n$ matrix. To find $\mathcal{A}_{\times 3}\mathbf{M}$, we can do the following:

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

The above pipeline is equivalent to applying $\mathbf{M}$ along tube fibers.

2.2 Definition of $\star_M$ Product

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

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

Definition: Given any invertible, $\mathbf{M} \in \mathcal{R}^{n\times n}$, $\mathcal{A} \in \mathbb{C}^{m\times p\times n}$ and $\mathcal{B} \in \mathbb{C}^{p\times l\times n}$, $\mathcal{C} = \mathcal{A} \star_M \mathcal{B}$ is defined via $\hat{\mathcal{C}}_{:,:,i} = \hat{\mathcal{A}}_{:,:,i}\hat{\mathcal{B}}_{:,:,i}$.

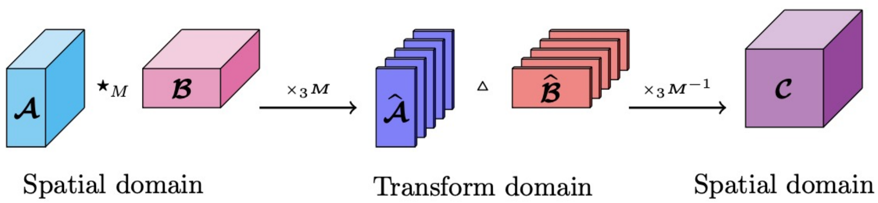

Figure 3: Pipeline of $\star_M$ product.

2.3 Special Case

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

2.4 Other Properties

Definition (Conjugate Transpose): Given $\mathcal{A} \in \mathbb{C}^{m\times p\times n}$, its $m \times p \times n$ conjugate transpose under $\star_M$ $\mathcal{A}^H$ is defined such that $(\hat{\mathcal{A}^H})^{(i)} = (\hat{\mathcal{A}^{(i)}})^H$, $i = 1, \ldots, n$.

Definition (Unitary/Orthogonal Tensors): $\mathcal{Q} \in \mathbb{C}^{m\times m\times n}$(or $\mathcal{Q} \in \mathbb{R}^{m\times m\times n}$) is called $\star_M$-unitary (or $\star_M$-orthogonal) if

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

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

2.5 Entry-wise M-product

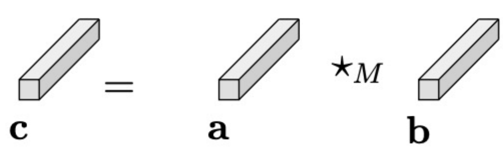

Figure 4: Entry-wise M-product

Tube fiber interpretation:

$$c = \texttt{fold}((\mathbf{M}^{-1}\text{diag}(\hat{\boldsymbol{a}})\mathbf{M})vec(\boldsymbol{b})) = \texttt{fold}((\mathbf{M}^{-1}\text{diag}(\hat{\boldsymbol{b}})\mathbf{M})vec(\boldsymbol{a}))$$

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

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

2.6 Matrix-mimeticity

Observation: 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

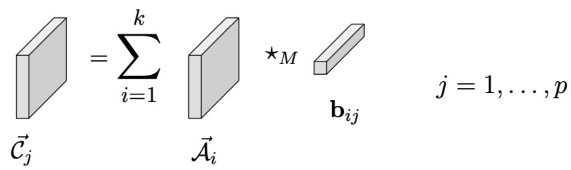

Figure 5: Product operation on $\mathcal{C}$.

2.7 Unitary Invariance

Theorem: If $\mathbf{M}$ a non-zero multiple of a unitary/orthogonal matrix, then

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

2.8 Tensor-tensor SVDs

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

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

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

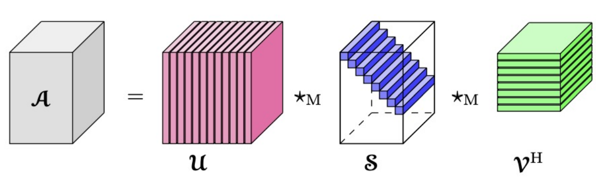

Figure 6: Tensor-Tensor SVD

Practical algorithm:

- $\hat{\mathcal{A}} \leftarrow \mathcal{A} \times_M \mathbf{M}$
- $\left[\hat{\mathcal{U}}_{:,:,i}, \hat{\mathcal{S}}_{:,:,i}, \hat{\mathcal{V}}_{:,:,i,}\right] = \text{SVD}(\hat{\mathcal{A}}_{:,:,i,})$, for $i = 1, \dots, n$
- $\mathcal{U} = \hat{\mathcal{U}} \times_3 \mathbf{M}^{-1}, \mathcal{S} = \hat{\mathcal{S}} \times_3 \mathbf{M}^{-1}, \mathcal{V} = \hat{\mathcal{V}} \times_3 \mathbf{M}^{-1}$

Eckart-Young Theorem. Given $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$, for $k < min(m, p)$ and $\mathbf{M}$ as previously, we define

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

Then,

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

where $\Omega = \{\mathcal{X} \star_M \mathcal{T} \mid \mathcal{X} \in \mathbb{R}^{m \times k \times n}, \mathcal{T} \in \mathbb{R}^{k \times p \times n}\}$. The error is calculated as:

$$\|\mathcal{A} - \mathcal{A}_k\|_F = \sum_{j>k} \|\mathcal{S}_{j,j}\|_F^2 = c \sum_{i=1}^{n} \sum_{j>k} \hat{\sigma_i}^{(j)},$$

where c depends on $\mathbf{M}$.

2.9 Theoretical Result

Theorem (Kilmer, Horesh, Avron, Newman (2021)) Suppose $\mathcal{A}_k$ is optimal k-term t-SVDM approximation to $\mathcal{A}$, and let $\mathcal{A}_k$ is optimal k-term matrix SVD approximation to $\mathbf{A}$. Then

$$\|\mathcal{A} - \mathcal{A}_k\|_F \leq \|\mathbf{A} - \mathbf{A}_k\|_F$$

where **strict inequality is achievable**. This result works for any $\mathbf{M}$ that is multiple of unitary (orthogonal) matrix.

2.10 t-SVDMII

Truncated t-SVDM ignores relative importance of faces. Global approach: order $\hat{\sigma_i}^{(j)} := \hat{\mathcal{S}_{i,i,j}}$, truncate on energy level.

Given $\mathcal{A}_\rho$, with $\rho_i = rank(\hat{\mathcal{A}}^{(i)})$

2.11 Comparison

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

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

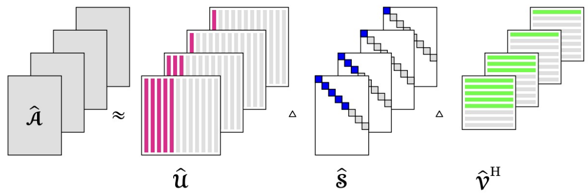

Figure 7: Demonstration of t-SVDMII.

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

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