CSE 392: Matrix and Tensor Algorithms for Data Spring 2024

Lecture 21 — 2024.04.03

Instructor: Shashanka Ubaru *Scribe: Tong Wang*

1 Algebraic Semantics

1.1 Algebraic Context - Semantics

Definitions. The following concepts are introduced in this lecture:

- A **binary composition** * on a set $\mathcal{S}$ is a function $\mathcal{S} \times \mathcal{S} \ni (x, y) \mapsto x * y \in \mathcal{S}$ (*e.g.*, multiplication).
- **Semigroup**: A set $\mathcal{S}$ with a binary composition * that is associative.
- **Monoid**: A semigroup which has a unit element.
- **Group**: A monoid where each element has an inverse.
- **Abelian group**: A group whose binary composition is commutative.

Figure 1 introduces the relationship between the aforementioned concepts.

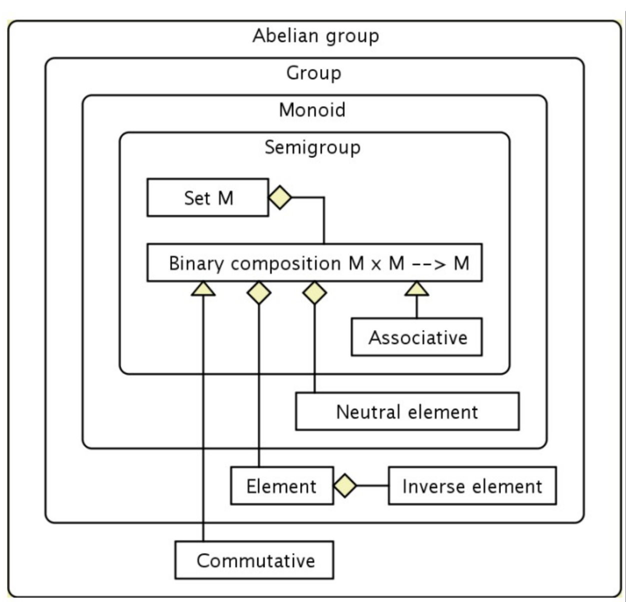

Figure 1: The relationship between the introduced arithmetic concepts.

1.2 Algebraic Context - Ring and Field

Definitions. Based on the definitions introduced in Section 1.1, we can further define the following concepts:

- **Ring**: A set $\mathcal{R}$ with two binary compositions, called addition (denoted by $+$) and multiplication such that $(\mathcal{R}, +)$ is an Abelian group, $(\mathcal{R}, \cdot)$ is a semigroup, and multiplication is distributive over addition.
- **Ring with unit**: A ring $\mathcal{R}$ where $(\mathcal{R}, \cdot)$ is a monoid.
- **Commutative ring**: A ring with commutative multiplication.
- **Field**: A commutative ring with unit where each non-zero element has a multiplicative inverse.

Figure 2 has provided an introduction between the concepts of rings and fields with the basic concepts introduced in Section 1.1.

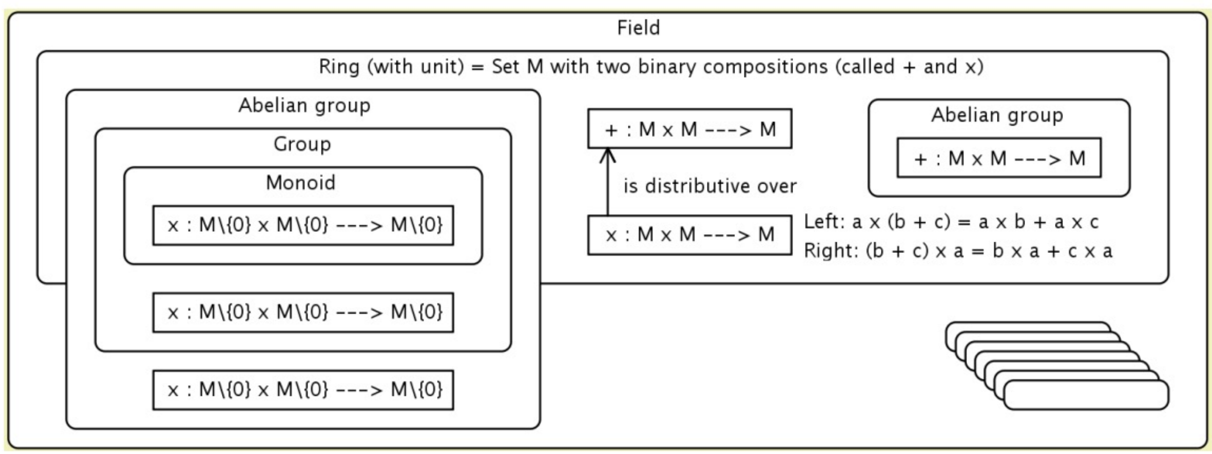

Figure 2: The relationship between the basic arithmetic concepts, rings and fields.

1.3 Algebraic Context - Module and Vector Space

Definitions:

- **Module**: An Abelian group $(\mathcal{M}, \oplus)$ is called a module over the ring $(\mathcal{R}, +, \cdot)$, or an $\mathcal{R}$ module, if there is a function $\mathcal{R} \times \mathcal{M} \ni (r, m) \mapsto rm \in \mathcal{M}$, such that:
  1. 0m $=$ 0 where 0 is the additive unit of $\mathcal{R}$
  2. 1m $=$ m if $\mathcal{R}$ has a multiplicative unit 1
  3. $(r + r')m = (rm) \oplus (r'm)$
  4. $r(m \oplus m') = (rm) \oplus (rm')$
  5. $(r \cdot r')m = r(r'm)$
- **Vector space**: A module over a field is called a vector space.

1.4 Algebraic Context - Algebra

Definitions.

Algebra: A structure comprising of addition, multiplication, and scalar multiplication (may also include additional assumptions as associativity, commutativity, etc.).

2 Tube-fiber product

2.1 Step Back to the Matrix SVD

Traditional workhorse for dimensional reduction or feature extraction is to use matrix singular value decomposition (SVD):

- Using PCA to find directions of most variability; projections in 'dominant' directions allows for dimension reduction/relative comparison
- Compression (reduce near redundancies) via truncated SVD expansion is optimal (according to Eckart-Young)

For $\mathbf{A} \in \mathcal{R}^{m \times n}$ and $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top = \sum_{i=1}^{r} \sigma_i(\mathbf{u}^{(i)} \circ \mathbf{v}^{(i)})$,

$$\mathbf{B} = \sum_{i=1}^{p} \sigma_i(\mathbf{u}^{(i)} \circ \mathbf{v}^{(i)}) \quad \text{solves the optimization problem of}$$

$$\min \|\mathbf{A} - \mathbf{B}\|_F \quad \text{s.t. B has rank p} \leq \text{r.}$$

Implicit storage: only $O(p(n+m))$ numbers stored in the SVD results compared to the full version, which needs mn.

The question is, how to make an extension to higher dimensions?

2.2 An Ideal Tensor Algebra

An ideal algorithm should have the four characteristic shown in Figure 3. For example, Tucker Decomposition is efficient to compute but not optimal. Many algorithms are not matrix mimetic and have to define more product and other things. The critical problem is that there is no one-to-one connection between matrix decomposition and tensor matrix decomposition.

2.3 Introduction of Mimeticity

Mimeticity is used to establish the connection between matrix operation and tensor operation. Essentially, we want to establish the following equivalence:

Matrices are linear operators $\leftrightarrow$ Tensors are t-linear operators

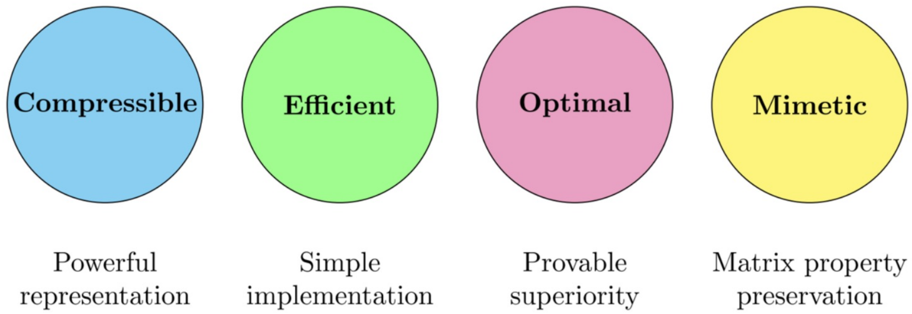

Figure 3: Ideal Properties of Tensor Algorithms

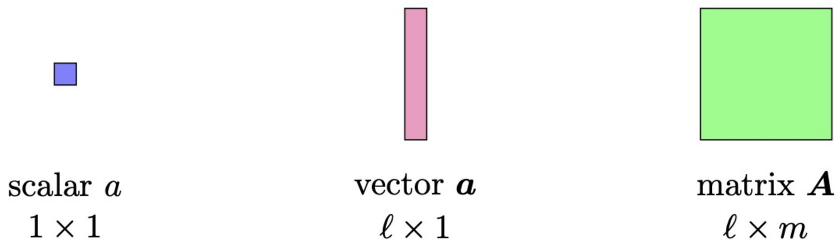

Figure 4: Basic units when handling matrix operations.

Orientation Dependence. CP and Tucker decomposition are orientation independent. For the remainder, let us first fix the orientation and then think of a tensor as **a matrix of tube fibers or matrix of lateral slices** (refer to Figure 5 and Figure 6). Using this fixed orientation, we can first consider a simple case, *i.e.*, the product between two tube fibers.

Circulant Matrices and Convolution. Let $\boldsymbol{v} \in \mathbb{R}^n$. Then the $n \times n$ circulant matrix generated by $\boldsymbol{v}$ is

$$\mathbf{C} = \mathrm{circ}(\boldsymbol{v}) = \begin{bmatrix} v_1 & v_n & v_{n-1} & \cdots & v_2 \\ v_2 & v_1 & v_n & \cdots & v_3 \\ \vdots & \ddots & \ddots & \ddots & v_4 \\ v_n & v_{n-1} & v_{n-2} & \cdots & v_1 \end{bmatrix}.$$

This matrix is well known to be diagonalized by the (unitary) DFT matrix:

$$\mathbf{C} = \mathbf{F}^H \Lambda \mathbf{F},$$

where the eigenvalues can be computed from the fast-Fourier transform(FFT) of $\boldsymbol{v}$.

Tensor Product for Tube Fibers. We can further define the discrete convolution between $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$ as follows:

$$\mathbf{a} * \mathbf{b} := \mathrm{circ}(\mathbf{a})\mathbf{b} = \mathbf{F}^H \mathrm{diag}(\hat{\mathbf{a}})\mathbf{F}\mathbf{b} = \frac{1}{\sqrt{n}}\mathbf{F}^H(\mathrm{diag}(\hat{\mathbf{a}})\hat{\mathbf{b}}) = \frac{1}{\sqrt{n}}\mathbf{F}^H(\hat{\mathbf{a}} \odot \hat{\mathbf{b}}) = \mathtt{ifft}(\hat{\boldsymbol{c}}),$$

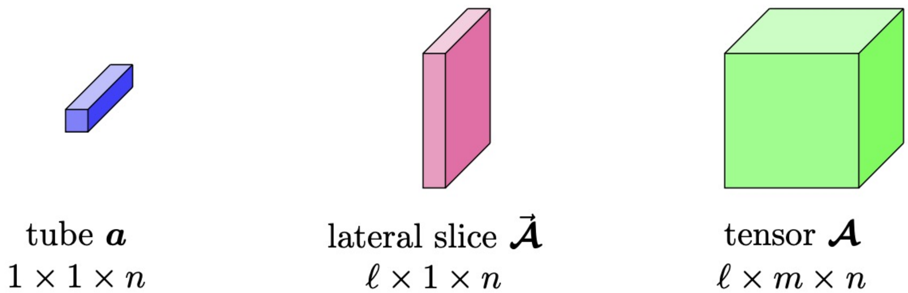

Figure 5: Basic units when handling tensor operations.

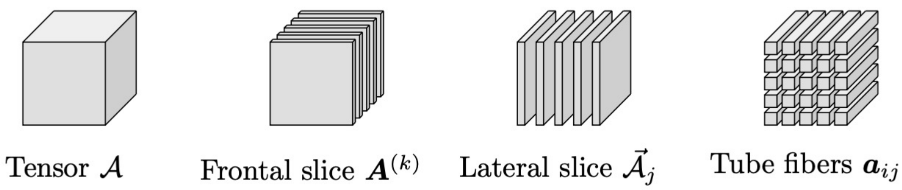

Figure 6: More basic units when handling tensor operations.

where $\hat{\mathbf{c}} = \hat{\mathbf{a}} \odot \hat{\mathbf{b}}$. It can be easily seen that discrete convolution between two vectors is commutative. If we establish an 1-1 correspondence between $1 \times 1 \times n$ tube fiber (element in $\mathbb{K}_n$) and a vector in $\mathbb{C}^n$, we can define the product of 2, length-n tube fibers as the discrete convolution. Since this is a commutative ring, there is an identity element $\boldsymbol{e} = \{1, 0, 0, 0, \dots\}$.

Tube-fiber Product. To compute $\mathbf{c} = \mathbf{a} * \mathbf{b}$, it can be done by calculating n independent scalar multiplications in the Fourier domain as shown in Figure 7. In the figure, $\triangle$ denotes face-wise scalar multiplication, analogous to the $\odot$ notation between two vectors. Consequently, $\mathbf{c}$ is obtained by applying the inverse transform to $\hat{\boldsymbol{c}}$.

The algorithmic cost of the operation can be broken down to three parts: (1) the cost to compute $\hat{a}, \hat{b}$, which requires $O(n \log n)$ FLOPs; (2) the cost for the n scalar products, which requires $O(n)$ FLOPs; and the cost of the inverse transform on $\hat{\mathbf{c}}$, which requires $O(n \log n)$ FLOPs.

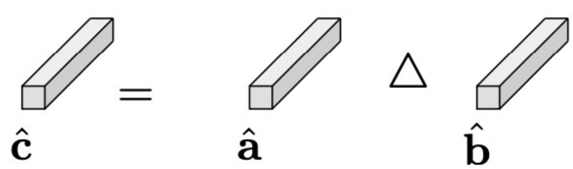

Figure 7: Tube-fiber Product.

3 The t-product

Next, we can define an entirely different tensor decomposition based on **circulant algebra**. In this factorization, a tensor in $\mathbb{R}^{n_1 \times n_2 \times n_3}$ is viewed as a $n_1 \times n_2$ matrix of "tubes" also known as **elements** of the **ring** $\mathbb{K}_{n_3}$ where addition is defined as vector addition and multiplication as **circular convolution**.

This "matrix-of-tubes" formalism leads to definitions of a **new multiplication for tensors** ("tubal multiplication"), a **new rank** for tensors ("tubal rank"), and a new notion of a **SVD for tensors** ("tubal SVD").

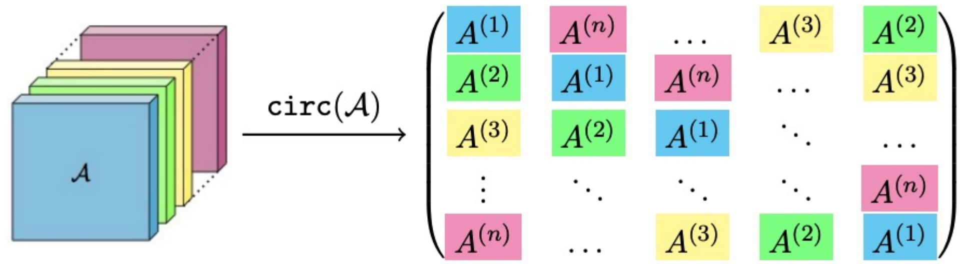

Figure 8: `circ` operation on tensors.

Firstly, the t-product can be defined as follows:

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

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

Block Circulants. A block circulant can be block-diagonalized by a (normalized) DFT in the 2^{nd} dimension:

$$(\mathbf{F} \otimes \mathbf{I})\texttt{circ}(\mathcal{A})(\mathbf{F}^* \otimes \mathbf{I}) = \begin{bmatrix} \hat{A}_1 & 0 & \dots & 0 \\ 0 & \hat{A}_2 & 0 & \dots \\ 0 & \dots & \ddots & 0 \\ 0 & \dots & 0 & \hat{A}_n \end{bmatrix},$$

where $\otimes$ is a Kronecker product of matrices. If $\mathbf{F}$ is a $n \times n$ matrix, and $\mathbf{I}$ is a $m \times m$ matrix, then $(\mathbf{F} \otimes \mathbf{I})$ is a $mn \times mn$ block matrix of n block rows and columns, where each block is $m \times m$ and the ij^{th} block is $f_{i,j} I$.

In practice, one never explicitly implement it this way because an FFT **along tube fibers** of $\mathcal{A}$ yields a tensor, $\hat{\mathcal{A}}$, whose frontal slices are the $\hat{A}_i$.

3.1 General Case for the t-product

The formula of t-product is defined as follows:

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

Since the term in the middle can be computed using the previous observation (compare to the case tube-fiber special case), we know that the t-product can be computed in the Fourier domain with

n independent matrix-matrix products $\hat{A}_i \cdot \hat{B}_i, i = 1, \ldots, n$, and then moving back to the spatial domain with an inverse transform of the result.

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

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

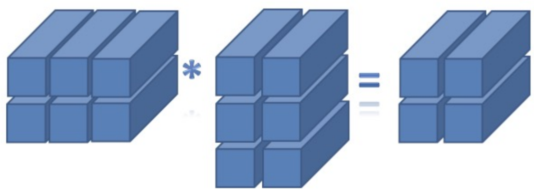

Figure 9: Block circulants block-diagonalized via 1D FFTs

3.2 Additional Properties

Now that the t-product is defined, we can define additional linear algebraic properties. For example:

- The $l \times l \times n$ identity tensor $\mathcal{I}$ is the tensor whose frontal slice is the $l \times l$ identity matrix, and whose other frontal slices are all zeros.
- An $l \times l \times n$ tensor $\mathcal{A}$ has an inverse $\mathcal{B}$ provided that $\mathcal{A} * \mathcal{B} = \mathcal{I}$, and $\mathcal{B} * \mathcal{A} = \mathcal{I}$.
- $\mathcal{A}^T \in \mathbb{R}^{m \times l \times n}$ is obtained by transposing each frontal slice and reversing order of transposed frontal slices 2 through n of $\mathcal{A}$.
- $\mathcal{U} \in \mathbb{R}^{m \times m \times n}$ is orthogonal if $\mathcal{U}^T * \mathcal{U} = \mathcal{I} = \mathcal{U} * \mathcal{U}^T$.