Relevant textbook materials: Chapter 2.1-2.5., 2.7.

Concepts

- Vector, examples, operations.
- Matrix, examples, operations, decompositions
- Cauchy inequalities and min-max principle
- Tensor, examples and operations.

Vector

Denote two vectors with n dimensions by $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ and $\mathbf{y} = (y_1, y_2, ..., y_n)^T$. Special vectors include:

- Basis vectors of n-dimensional Euclidean space (or Cartesian coordinate system) can be defined as: $\mathbf{e}_1 = (1, 0, ..., 0)^T$, $\mathbf{e}_2 = (0, 1, 0, ..., 0)^T$, ..., $\mathbf{e}_n = (0, ..., 0, 1)^T$.
- Any vector can be represented by basis vectors $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$.
- Vector of zeros: $\mathbf{0} = (0, 0, ..., 0)^T$. For an n-dimensional vector (or n-vector) of zeros, we often denote it by $\mathbf{0}_n = (0, 0, ..., 0)^T$.
- Vector of ones: $\mathbf{1} = (1, 1, ..., 1)^T$. For an n-dimensional vector (or n-vector) of ones, we often denote it by $\mathbf{1}_n = (1, 1, ..., 1)^T$.
- Unit vector. For any vector $\tilde{\mathbf{x}}$, we call $\tilde{\mathbf{x}}$ a unit vector, if the "length" of $\mathbf{x}$, denoted by $||\tilde{\mathbf{x}}|| = 1$. The "length" is often defined by the Euclidean distance. For any vector $\mathbf{x}$ (not a vector of zeros), $\mathbf{x}/||\mathbf{x}||$ is a unit vector.

Basic vector operations and properties include

- Vector summation. $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$.
- Scalar product. Assume $c \in \mathbb{R}$. $c\mathbf{x} = (cx_1, cx_2, ..., cx_n)^T$.
- Dot (inner) product. $\mathbf{x}^T\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + ... + x_ny_n = \sum_{i=1}^n x_iy_i$. In `R`, the dot product is `t(x)%*%y` or you may write `sum(x*y)`.
- Cross product. $\mathbf{x} \times \mathbf{y} = ||\mathbf{x}||||\mathbf{y}||\sin(\theta)\mathbf{n}$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$ and $\mathbf{n}$ is the unit vector perpendicular to the plane containing $\mathbf{x}$ and $\mathbf{y}$ in the direction of right hand rule. Suppose $\mathbf{x} = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ and $\mathbf{y} = (y_1, y_2, y_3)^T \in \mathbb{R}^3$. Then

$$\mathbf{z} = \mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)^T,$$

  See Wikipedia for more properties of the cross product.

- Outer (tensor) product. $\mathbf{x}\mathbf{y}^T = \begin{pmatrix} x_1y_1 & x_1y_2 & ... & x_1y_n \\ x_2y_1 & x_2y_2 & ... & x_ny_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & ... & x_ny_n \end{pmatrix}$ We denote outer product by $\mathbf{x} \otimes_{outer} \mathbf{y}$.

- Length of a vector (L_2 norm or Euclidean distance). $||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} = (\sum_{i=1}^n x_i^2)^{1/2}$.

- Angle between two vectors. Denote the angle between $\mathbf{x}$ and $\mathbf{y}$ by θ. $\cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}$. Thus $\theta = \arccos\left(\frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}\right)$. When $\mathbf{x}^T\mathbf{y} = 0$, it means the two vectors are orthogonal or perpendicular.

- Vector projection. Projecting vector $\mathbf{x}$ onto $\mathbf{y}$ will give a vector: $||\mathbf{x}||\cos(\theta)\frac{\mathbf{y}}{||\mathbf{y}||}$, where $\cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}$.

- Linear dependence. For two vector $\mathbf{x}$ and $\mathbf{y}$ with same dimensions, if these exists nonzero constant c_1 and c_2, such that

$$c_1\mathbf{x} + c_2\mathbf{y} = \mathbf{0},$$

then we call $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent. For a set of vectors with the same dimension $\mathbf{x}_1$, $\mathbf{x}_2$,...,$\mathbf{x}_n$, if there exists constants c_1, c_2,...,c_n, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_n\mathbf{x}_n = \mathbf{0},$$

we call this set of vectors are linearly dependent.

Matrix

Consider two matrices $\mathbf{A}_{[n_1 \times n_2]}$ and $\mathbf{B}_{[n_1 \times n_2]}$, where the (i, j)th entry of $\mathbf{C}$ was defined as a_{ij} and b_{ij}, respectively.

Special matrices

- Square matrix. An $n_1 \times n_2$ matrix is a square matrix if $n_1 = n_2$.

- Symmetric matrix. An $n \times n$ matrix is call symmetric, if $a_{ij} = a_{ji}$.

- Diagonal matrix. An $n \times n$ matrix $\mathbf{D}$ is said to be a diagonal matrix if at least one diagonal value is not zero, while off-diagonal entries are all zeros.

- Identity matrix. An $n \times n$ identity matrix $\mathbf{I}_n$ with ones on the diagonal entries and zeros else.

- Tridiagonal matrix. A tridiagonal matrix $\mathbf{T}$ is a band matrix with nonzero elements on the diagonal entries, the first diagonal below/above this, and zero elsewhere:

$$\mathbf{T} = \begin{pmatrix} a_1 & b_1 & 0 & 0 & \dots & 0 \\ c_1 & a_2 & b_2 & 0 & \dots & 0 \\ 0 & c_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & 0 & \dots & 0 & c_{n-1} & a_n \end{pmatrix}.$$

  Tridiagonal matrix can simplify computation for Markov models.

- Lower/upper triangular matrix. A matrix is called a lower/upper triangular matrix, if the entries above/below the diagonal entries are zero.

- Positive semi-definite matrix. An $n \times n$ matrix $\mathbf{M}$ is positive semi-definite if and only if $\mathbf{x}^T\mathbf{M}\mathbf{x} \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. A $\mathbf{M}$ is positive definite matrix if and only if $\mathbf{x}^T\mathbf{M}\mathbf{x} > 0$ for any $\mathbf{x} \in \mathbb{R}^n$. Every positive definite matrix is invertible and its inverse is also positive definite.

- Rank one matrix. Consider an $n_1 \times n_2$ matrix $\mathbf{A}$. If $\mathbf{A} = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ is an n_1 vector and $\mathbf{v}$ is an n_2 vector.

Basic matrix operations

- Matrix summation. Let $\mathbf{C} = \mathbf{A} + \mathbf{B}$. Each entry of $\mathbf{C}_{[n_1 \times n_2]}$ is c_{ij}.

- Scalar product. Let $c \in \mathbb{R}$, the (i, j)th entry of $c\mathbf{A}$ is ca_{ij}.

- Matrix product. Let $\mathbf{C} = \mathbf{A}\mathbf{B}^T$, where the (i, j)th entry of $\mathbf{C}_{[n_1 \times n_1]}$ was $c_{ij} = \sum_{t=1}^{n_2} a_{it} b_{jt}$, respectively. In R, the matrix product is `A%*%t(B)`. What is the number of operations of the matrix product $\mathbf{A}\mathbf{B}^T$?

- Hadamard (element-wise) product. Let $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$, where $\circ$ denotes the Hadamard product and the (i, j)th entry of $\mathbf{C}_{[n_1 \times n_2]}$ is $c_{ij} = a_{ij} b_{ij}$. In R, the element-wise product is `A * B`.

- Kronecker product. Let $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$, where $\otimes$ denotes the Kronecker product. Then $\mathbf{C}$ is an $n_1 n_1 \times n_2 n_2$ matrix defined as

$$\begin{aligned} \mathbf{C} &= \mathbf{A} \otimes \mathbf{B} \\ &= \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n_2}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n_2}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_1 1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{n_1 n_2}\mathbf{B} \end{pmatrix}. \end{aligned}$$

We may also write the Kronecker product by $\otimes_{Kron}$ to differentiate the Kronecker product and the outer product. What is the difference between the Kronecker product and the outer product? In R, the Kronecker product is $\texttt{kronecker(A, B, FUN = "*")}$.

- Matrix inversion. A square matrix $\mathbf{A}$ is called invertible (or nonsingular or nondegenerate), if there exists a matrix $\mathbf{A}^{-1}$, such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$. Inverting a matrix can be implemented by the Gaussian-elimination algorithm. The computational complexity of inversion is $O(n^3)$ in general[1].

- Woodbury matrix identity. Let $\mathbf{A}_{[n\times n]}$, $\mathbf{U}_{[n\times k]}$, $\mathbf{C}_{[k\times k]}$ and $\mathbf{V}_{[k\times n]}$ be 4 matrices. Further assume $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{A}+\mathbf{UCV}$ are invertible. Then

$$(\mathbf{A}+\mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}.$$

- Trace. The trace of a $n\times n$ matrix $\mathbf{A}$ is defined as $\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.

- Determinant. The determinant of a $n\times n$ square matrix $\mathbf{A}$ is often denoted as $|\mathbf{A}|$ or $\det(\mathbf{A})$. If $n=1$, $|\mathbf{A}| = \mathbf{A}$ and If $n>1$, $\mathbf{A} = \sum_{j=1}^n a_{1j}|\mathbf{A}_{1j}|(-1)^{1+j}$, where $\mathbf{A}_{1j}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the first row and jth column of $\mathbf{A}$. Also $|\mathbf{A}| = \sum_{j=1}^n a_{ij}|\mathbf{A}_{ij}|(-1)^{i+j}$. Furthermore, $|\mathbf{A}| = \lambda_1\lambda_2...\lambda_n$ where λ_i is the ith eigenvalue of the matrix. $|\mathbf{A}| = L_{11}^2 L_{22}^2 ... L_{nn}^2$ where L_{ii} is the ith diagonal entry in the Cholesky decomposition of $\mathbf{A}$ (if $\mathbf{A}$ is positive definite, see discussion below). Computing the determinant of an $n\times n$ matrix costs $O(n^3)$ in general.

- Vectorization. Vectorizing a matrix $\mathbf{A}_{[n_1\times n_2]}$ gives an $n_1 n_2$-vector:

$$\mathrm{Vec}(\mathbf{A}) = [a_{11}, a_{21}, ..., a_{n_1 1}, a_{12}, a_{22}..., a_{n_1 2}, ..., a_{1n_2}, a_{2n_2}..., a_{n_1 n_2}]^T.$$

Matrix decomposition

- Eigendecomposition (spectral decomposition). For an $n\times n$ square matrix $\mathbf{A}$, the eigendecomposition follows

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n]$ is the matrix of the eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues with the ith diagonal term defined as λ_i. We have $\mathbf{u}_i^T\mathbf{u}_i = 1$ and $\mathbf{u}_i^T\mathbf{u}_j = 0$ if $i \neq j$. The eigenpairs of a matrix satisfy the condition: $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$.

[1] We often avoid directly computing the inversion of a matrix. There are other ways to compute terms such as $\mathbf{A}^{-1}\mathbf{x}$ and $|\mathbf{A}|$. If $\mathbf{A}$ is sparse, conjugate gradient and Krylov subspace method may be efficient to approximate $\mathbf{A}^{-1}\mathbf{x}$.

- Singular value decomposition. For an $n_1 \times n_2$ matrix $\mathbf{A}$, suppose $n_1 \le n_2$. The singular value decomposition (SVD) of $\mathbf{A} = \mathbf{UDV}$, where $\mathbf{U}$ is an $n_1 \times n_1$ orthogonal matrix, $\mathbf{D}$ is $n_1 \times n_2$ rectangular diagonal matrix with at most n_1 nonzero entries, and $\mathbf{V}$ is an $n_2 \times n_2$ orthogonal matrix. The SVD is connected to eigendecomposition through

$$\mathbf{AA}^T = \mathbf{UDV}(\mathbf{UDV})^T = \mathbf{UDD}^T\mathbf{U}^T,$$

  where the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{AA}^T$ and the diagonal entries of $\mathbf{DD}^T$ give the eigenvalues. In R, singular value decomposition can be computed by `svd`(.).

- Cholesky decomposition. For a symmetric positive definite matrix $\mathbf{A}$, the Cholesky decomposition means $\mathbf{A} = \mathbf{LL}^T$, where $\mathbf{L}$ is a lower triangular matrix with real and positive diagonal entries. Cholesky decomposition is unique.

- Lower–upper (LU) decomposition. For an $n \times n$ square matrix $\mathbf{A}$, LU decomposition means $\mathbf{A} = \mathbf{LU}$, where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{U}$ is an upper triangular matrix.

Matrix norm, rank and conditional number

- Matrix norm. For an $n_1 \times n_2$ real-valued matrix $\mathbf{A}$. The Frobenius norm (also called L_2 norm or Euclidean norm) of $\mathbf{A}$ is

$$||\mathbf{A}||_F = \sqrt{\sum_{j=1}^{n_2}\sum_{i=1}^{n_1} a_{ij}^2}.$$

- Rank of a matrix. The rank of a matrix is the dimension of the vector space generated by its column vectors (i.e. maximum number of linear independent columns).

- For a positive definite matrix, one often uses conditional number $\kappa = \lambda_{max}/\lambda_{min}$. If κ is very large (e.g. $> 10^{10}$), it means that the matrix is ill-conditioned or near-singular. In R, `kappa(.)` is the function to evaluate the condition number.

Cauchy-Swartz inequality and min-max principle

Denote two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$.

Theorem 1 (Cauchy-Schwarz Inequality)**.** .

$$|\mathbf{x}_1^T\mathbf{x}_2| \le ||\mathbf{x}_1||||\mathbf{x}_2||,$$

with equality if and only $\mathbf{x}_1 = c\mathbf{x}_2$ *(or* $\mathbf{x}_1 = c\mathbf{x}_2$*) for some constant* c.

See the proof from textbook or Wikipedia.

Remark 1. *Denote the inner product* $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^T \mathbf{x}_2$. *Cauchy-Schwarz Inequality means that the inner product of two vectors is always smaller than the product of the length of each vector, because the inner product measured the product of the length of the projected vector* $\mathbf{x}_1$ *onto* $\mathbf{x}_2$ *and the length of* $\mathbf{x}_1$. *Given any inner product, the quotient*

$$cos(\theta) = \frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{||\mathbf{x}_1||||\mathbf{x}_2||},$$

is often used to define the "angle" between vector $\mathbf{x}_1$ *and* $\mathbf{x}_2$.

Remark 2. *Define* X_1 *and* X_2 *two random variables. Suppose* $\mathbb{E}[X_1^2]$ *and* $\mathbb{E}[X_2^2]$ *exists. Cauchy-Schwarz Inequality means*

$$|\mathbb{E}(X_1 X_2)| \leq \sqrt{\mathbb{E}[X_1^2]\mathbb{E}[X_2^2]}$$

Corollary 1 (Extended Cauchy-Schwarz Inequality)**.** *Let* $\mathbf{A}$ *be a positive definite matrix. Then*

$$\mathbf{x}_1^T \mathbf{x}_2 \leq (\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1)(\mathbf{x}_2^T \mathbf{A}^{-1} \mathbf{x}_2),$$

with equality if and only if $\mathbf{x}_1 = c\mathbf{A}^{-1}\mathbf{x}_2$ *(or* $\mathbf{x}_2 = c\mathbf{A}\mathbf{x}_1$*) for some constant* c. *for some constant* c.

Corollary 2 (Inner product Maximization)**.** *Let* $\mathbf{A}$ *be a positive definite matrix. Then for any nonzero* $\mathbf{x} \in \mathbb{R}^p$ *and a positive definite matrix* $\mathbf{A}$

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{(\mathbf{x}^T \mathbf{v})}{\mathbf{x}^T \mathbf{A} \mathbf{x}} = \mathbf{v}^T \mathbf{A}^{-1} \mathbf{v}.$$

Theorem 2 (Courant–Fischer–Weyl min-max principle)**.** . *Let* $\mathbf{A}_{[p \times p]}$ *be a positive definite matrix with eigenvalues* $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ *associated with eigenvectors* $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p$. *Then*

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_1 \quad (attained\ when\ \mathbf{x} = \mathbf{u}_1)$$

$$\min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_p \quad (attained\ when\ \mathbf{x} = \mathbf{u}_p),$$

$$Moreover, \max_{\mathbf{x} \perp \mathbf{u}_1, \ldots \mathbf{u}_k} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_{k+1} (attained\ when\ \mathbf{x} = \mathbf{u}_{k+1}, k = 1, 2..., p-1).$$

See the proof from textbook or Wikipedia.

Remark 3. *The quotient*

$$R(\mathbf{A}, \mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

is called Rayleigh quotient.

Tensor

The following notes are mostly from Kolda and Bader (2009).

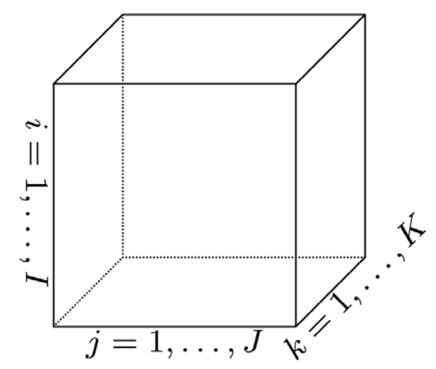

Figure 1: A three-way tensor from Kolda and Bader (2009).

An Nth-order tensor may be denoted as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$. The order (or way or mode) of a tensor $\mathcal{X}$ is the number of dimensions. E.g. 3rd-order (or three-way) tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is plotted in Figure 1. For a 3rd-order $\mathcal{X}$, the entry may be defined as x_{ijk} and for an Nth-order tensor, the entry may be defined as $x_{i_1 i_2 \ldots i_N}$. The fibers of a tensor is analogue to rows or columns in a matrix. E.g. a third-order tensor $\mathcal{X}^{I \times J \times K}$ has column, row, and tube fibers, denoted as $\mathbf{x}_{\cdot jk}$, $\mathbf{x}_{i \cdot k}$ and $\mathbf{x}_{ij\cdot}$ (shown in lower panels in Figure 2). The slices are two-dimensional sections of a tensor, defined by fixing all but two indices. E.g. a third-order tensor $\mathcal{X}^{I \times J \times K}$ has slices $\mathbf{X}_{i\cdot\cdot}$, $\mathbf{X}_{\cdot j\cdot}$ and $\mathbf{X}_{\cdot\cdot k}$ (graphed in lower panels in Figure 2). Sometimes the slide $\mathbf{X}_{\cdot\cdot k}$ of third-order tensor $\mathcal{X}$ may be simply denoted as $\mathbf{X}_k$.

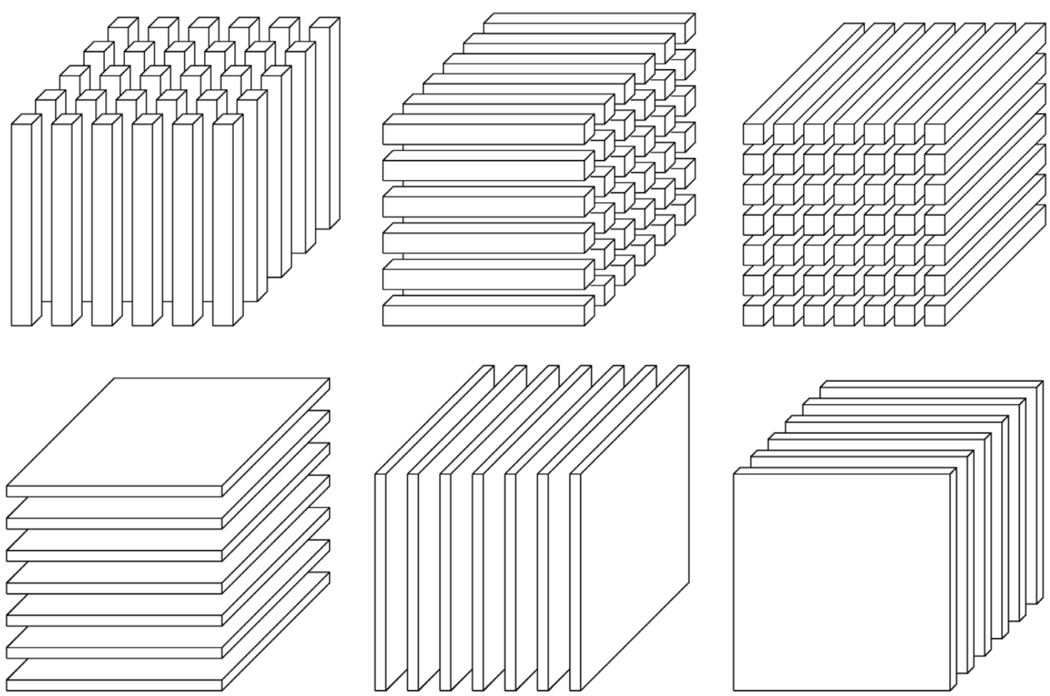

Figure 2: Upper panels from left to right: Mode-1 (column) fibers ($\mathbf{x}_{\cdot jk}$), Mode-2 (row) fibers ($\mathbf{x}_{i \cdot k}$) and Mode-3 (tube) fibers: ($\mathbf{x}_{ij\cdot}$). Lower panels from left to right: Horizontal slices ($\mathbf{X}_{i\cdot\cdot}$), Lateral slices ($\mathbf{X}_{\cdot j\cdot}$) and Frontal slices ($\mathbf{X}_{\cdot\cdot k}$). Figures are from Kolda and Bader (2009).

Special tensors

- All zeros tensor or all one tensors if all entries are zeros or ones.
- Diagonal tensor. A Nth order tensor $\mathcal{X}$ is diagonal if $x_{i_1 i_2 \dots i_n} \neq 0$ only if $i_1 = i_2 = \dots = i_N$.
- Rank one tensor. A Nth order tensor $\mathcal{X}$ has rank one if it can be written as the outer product of N vectors:

$$\mathcal{X} = \mathbf{u}_1 \otimes_{outer} \mathbf{u}_2 \otimes_{outer} \dots \otimes_{outer} \mathbf{u}_N.$$

- Cubical tensor. A tensor is called cubical if every mode is the same side, i.e. $\mathcal{X} \in \mathbb{R}^{I \times I \times \dots \times I}$.
- Symmetric and supersymmetric tensor. A cubical tensor is supersymmetric if its elements remain constant under any permutation of the indices. E.g. for a three-way tensor $\mathcal{X} \in \mathbb{R}^{I \times I \times I}$ is supersymmetric if

$$x_{ijk} = x_{ikj} = x_{jik} = x_{jki} = x_{kij} = x_{kji},$$

for any $i, j, k = 1, 2, \dots, I$. The tensor can be symmetric in two or more modes. E.g. a three-way tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is symmetric in modes one and two if all its frontal slices are symmetric, i.e. $\mathbf{X}_k = \mathbf{X}_k^T$ for $k = 1, 2, \dots, K$.

Summation and scalar product on tensor follows naturally.

Other tensor operations

- Matricization: transformation a tensor into a matrix. Matricization (unfolding or flattening) is to reorder the entries of an N-way tensor into a matrix (analogy to vectorizing a matrix).

Example 1. *Suppose the frontal slices of $\mathcal{X} \in \mathbb{R}^{3 \times 4 \times 2}$ be:*

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{pmatrix}$$

Then the three mode-n unfoldings are

$$\mathbf{X}_{(1)} = \begin{pmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{pmatrix}$$

$$\mathbf{X}_{(2)} = \begin{pmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{pmatrix}$$

$$\mathbf{X}_{(3)} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \end{pmatrix}$$

- Tensor product: the n-mode (matrix) product. For a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the n-mode product with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n \mathbf{U}$ and is of size $I_1 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N$. Each element of the $\mathcal{X} \times_n \mathbf{U}$ follows:
$$(\mathcal{X} \times_n \mathbf{U})_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} x_{i_1} x_{i_2} \ldots x_{i_N} u_{j i_n}$$
The n-mode product is equivalent to have each mode-n fiber multiplied by the matrix U:
$$\mathcal{Y} = \mathcal{X}_n \mathbf{U} \equiv \mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}.$$
Example 2. *Assume $\mathcal{X}$ is defined in Example 1 and let* $\mathbf{U} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$. *The 1-mode product $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U} \in \mathbb{R}^{2\times 4 \times 2}$ is*
$$\mathbf{Y}_{(1)} = \begin{pmatrix} 22 & 49 & 76 & 103 & 130 & 157 & 184 & 211 \\ 128 & 64 & 100 & 136 & 172 & 208 & 244 & 280 \end{pmatrix}$$
or equivalently the front slides of $\mathcal{Y}$ can be written as
$$\mathbf{Y}_1 = \begin{pmatrix} 22 & 49 & 76 & 103 \\ 128 & 64 & 100 & 136 \end{pmatrix}, \quad \mathbf{Y}_2 = \begin{pmatrix} 130 & 157 & 184 & 211 \\ 172 & 208 & 244 & 280 \end{pmatrix}.$$
Other properties of tensor product include
  - The distinct mode product is exchangeable. E.g.
  $$\mathcal{X} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n \mathbf{B} \times_m \mathbf{A}, \quad \text{if } m \neq n.$$
  - The same mode product can be written as
  $$\mathcal{X} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n (\mathbf{B}\mathbf{A}).$$

See Kolda and Bader (2009) for tensor decompositions.

Reading materials

- Textbook Chapter 2.1-2.5., 2.7.
- For derivatives of vector and matrix, please see Matrix Cookbook.
- Please read Kolda and Bader (2009) for more about tensor decomposition.

References

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.