Lecture 1: introduction to Tensors and Applications

Neriman Tokcan

Masterclass, University of Trento
November 25-29 2024

TensorDec Laboratory
algebra, geometry and applications of tensor decompositions

UNIVERSITÀ DI TRENTO
Dipartimento di Matematica

25 - 29 November 2024 | Polo Ferrari - Povo 1

Masterclass

Tensor Decompositions and Applications in Multi-Omics Data Analysis

Neriman Tokcan
(U. Massachusetts Boston)

Omics technologies, including genomics, transcriptomics, proteomics, and metabolomics, have revolutionized biological research by enabling comprehensive, high-throughput analysis of molecular components within cells and organisms. The resulting high-dimensional datasets pose significant analytical challenges, particularly in integrating diverse data types and uncovering complex biological relationships. Tensor-based approaches have emerged as powerful tools for analyzing these high-dimensional omics data, offering advantages over traditional matrix-based methods in capturing complex, multi-way relationships.

SCHEDULE

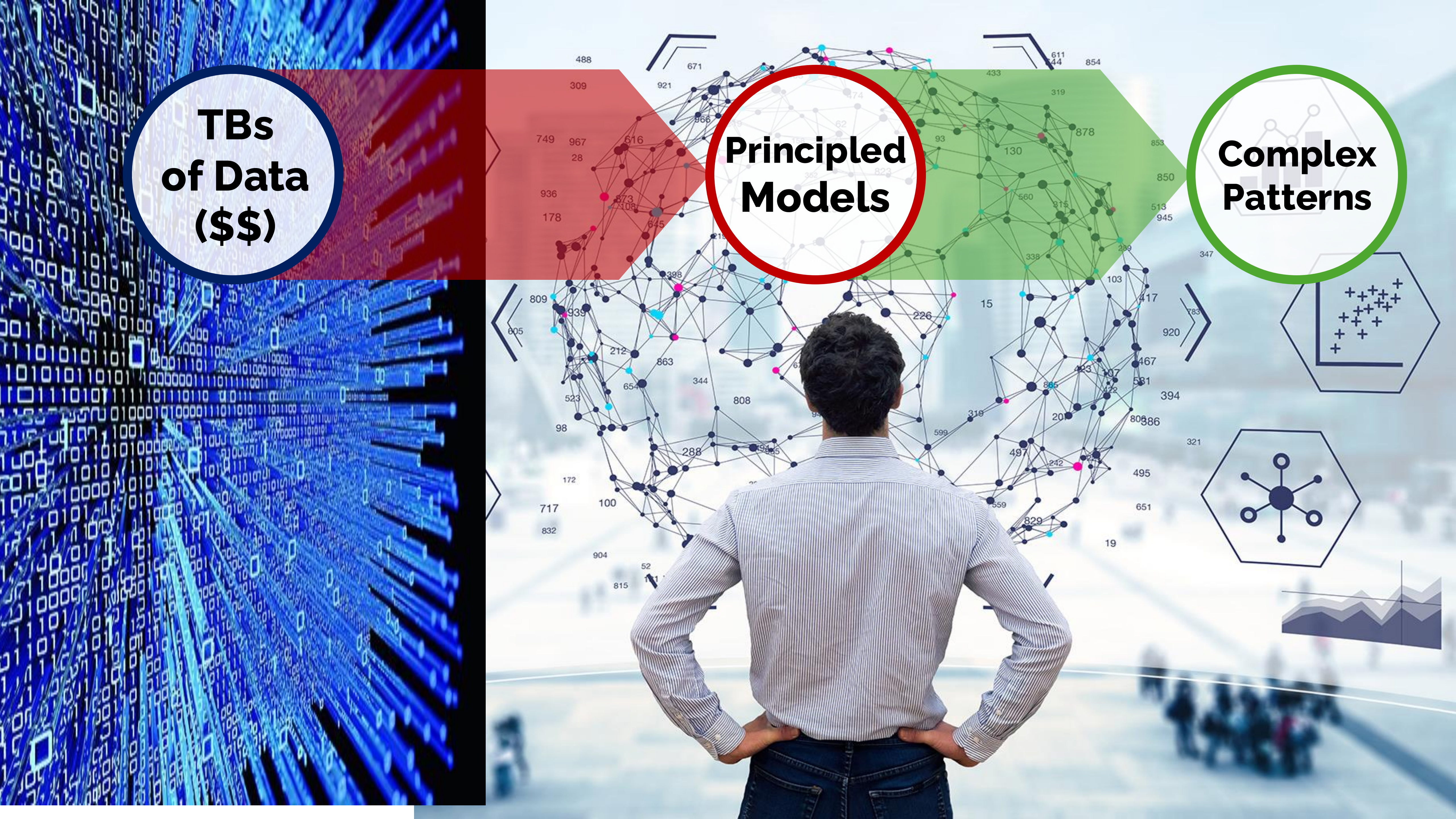
Monday 25	11:30 - 13:30	Room A108
Tuesday 26	15:30 - 17:30	Room A102
Wednesday 27	12:30 - 14:30	Room A203
Thursday 28	10:30 - 12:30	Room A209
Friday 29	11:30 - 13:30	Room A209

The lecture will also be available in streaming via ZOOM.
The link will be shared to the registered participants.

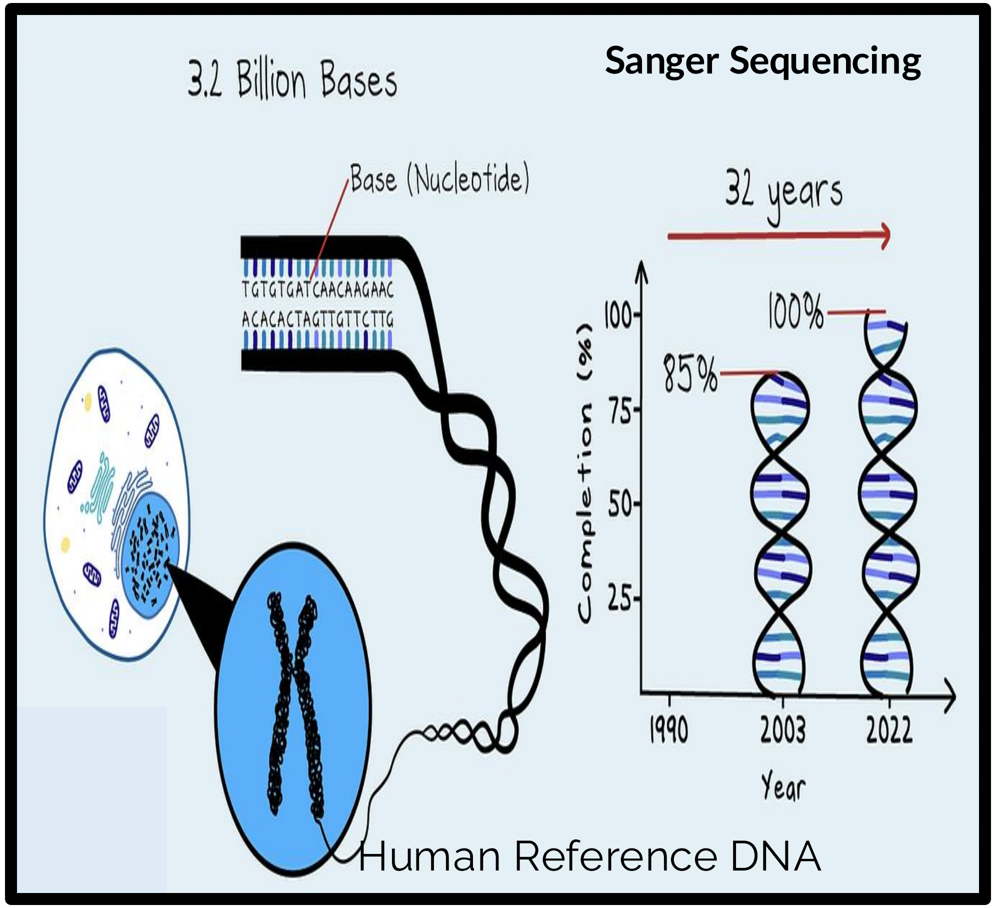
Scan the QR code for info and registration

0

The era of **Biotech**

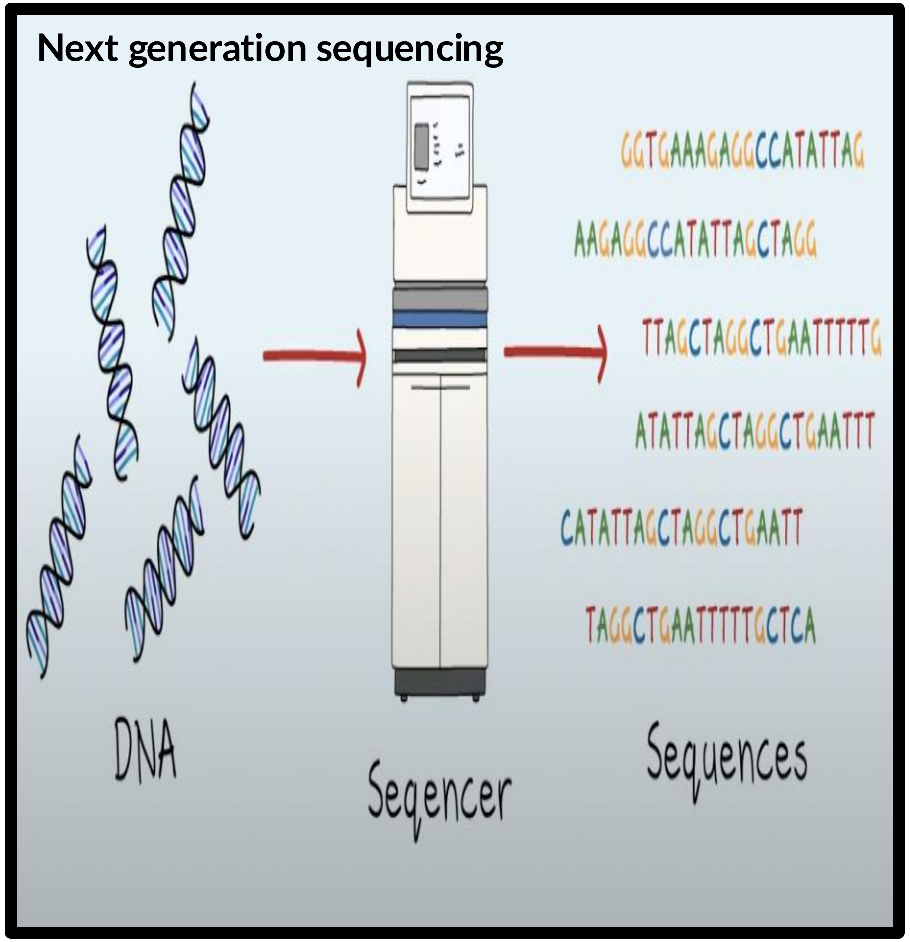

TBs of Data ($$)
Principled Models
Complex Patterns

The Human Genome Project



Time: 32 years

Cost: $3 billion



Time: 1 day

Cost: $1,000- $5,000

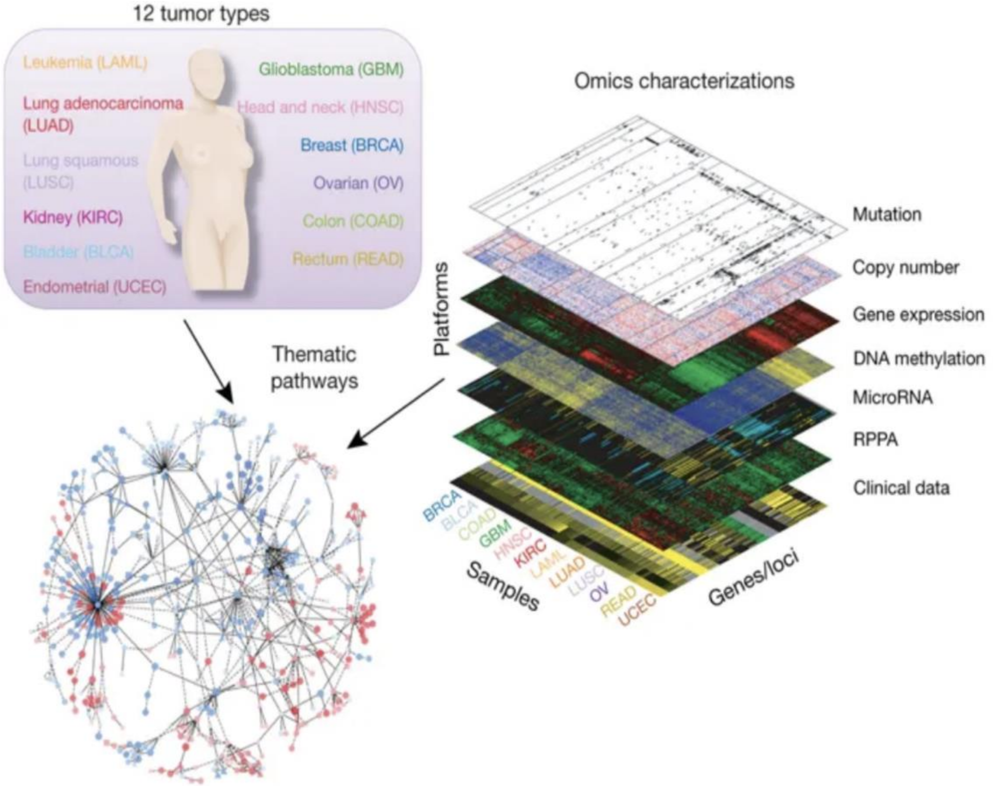
www.clevaLab.com

The omics era

Bulk genomics

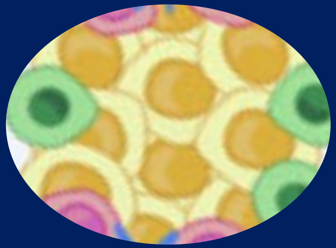
Single-cell genomics

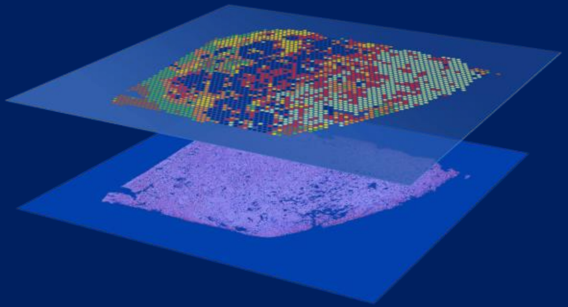
Spatial transcriptomics

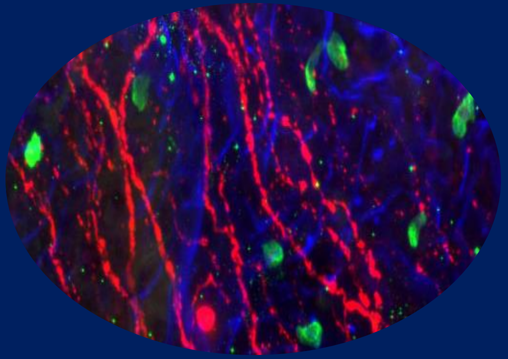


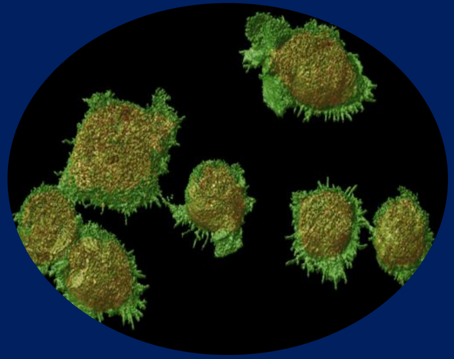
Multi-omics data

$$donors \times features \times omics\ platforms$$

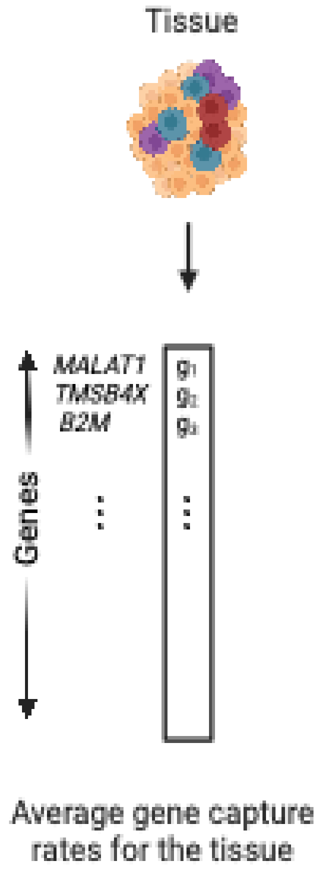


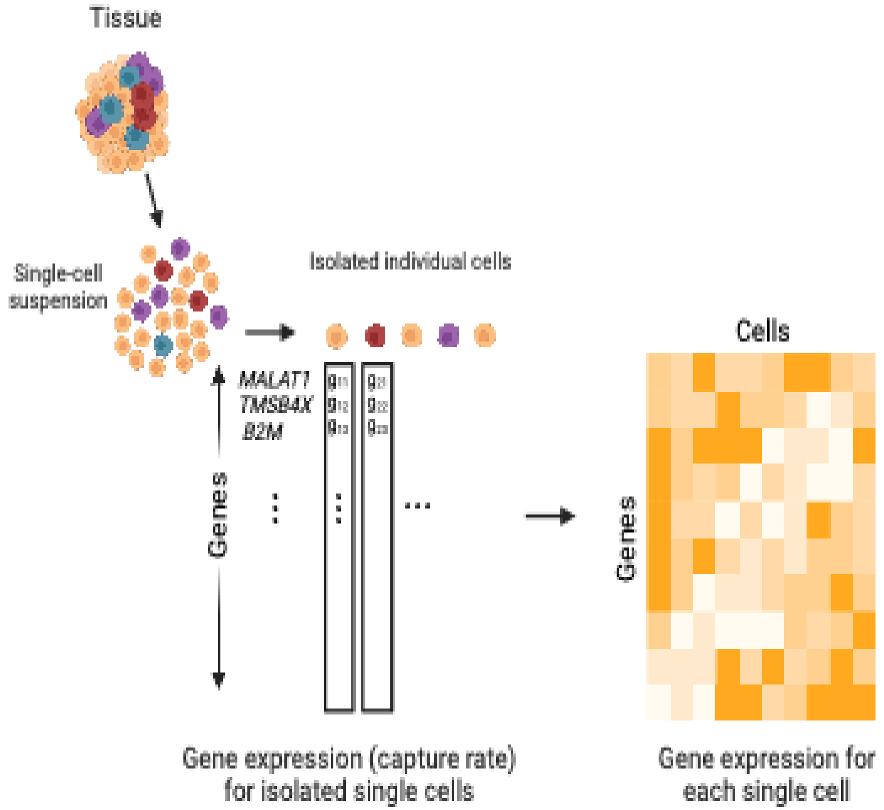


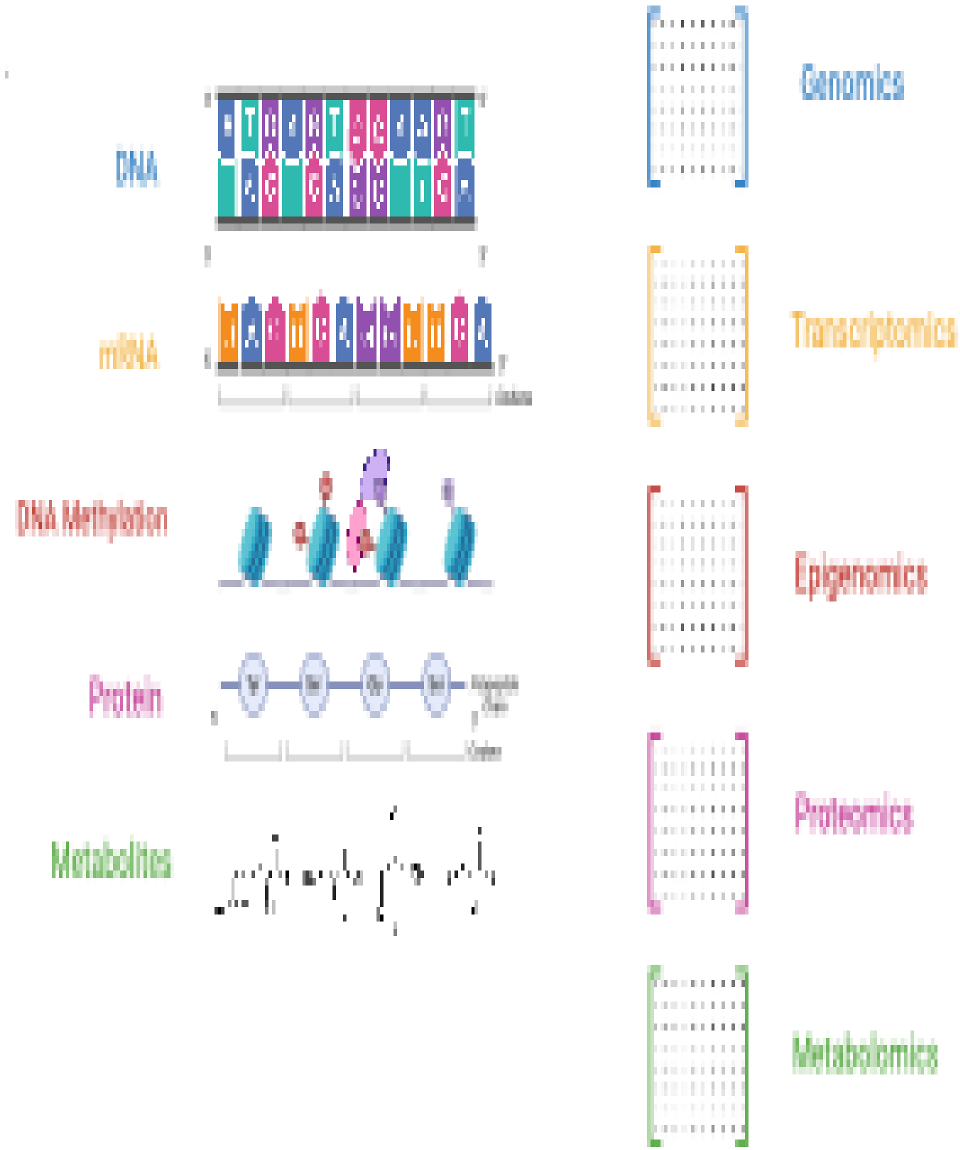




The omics era





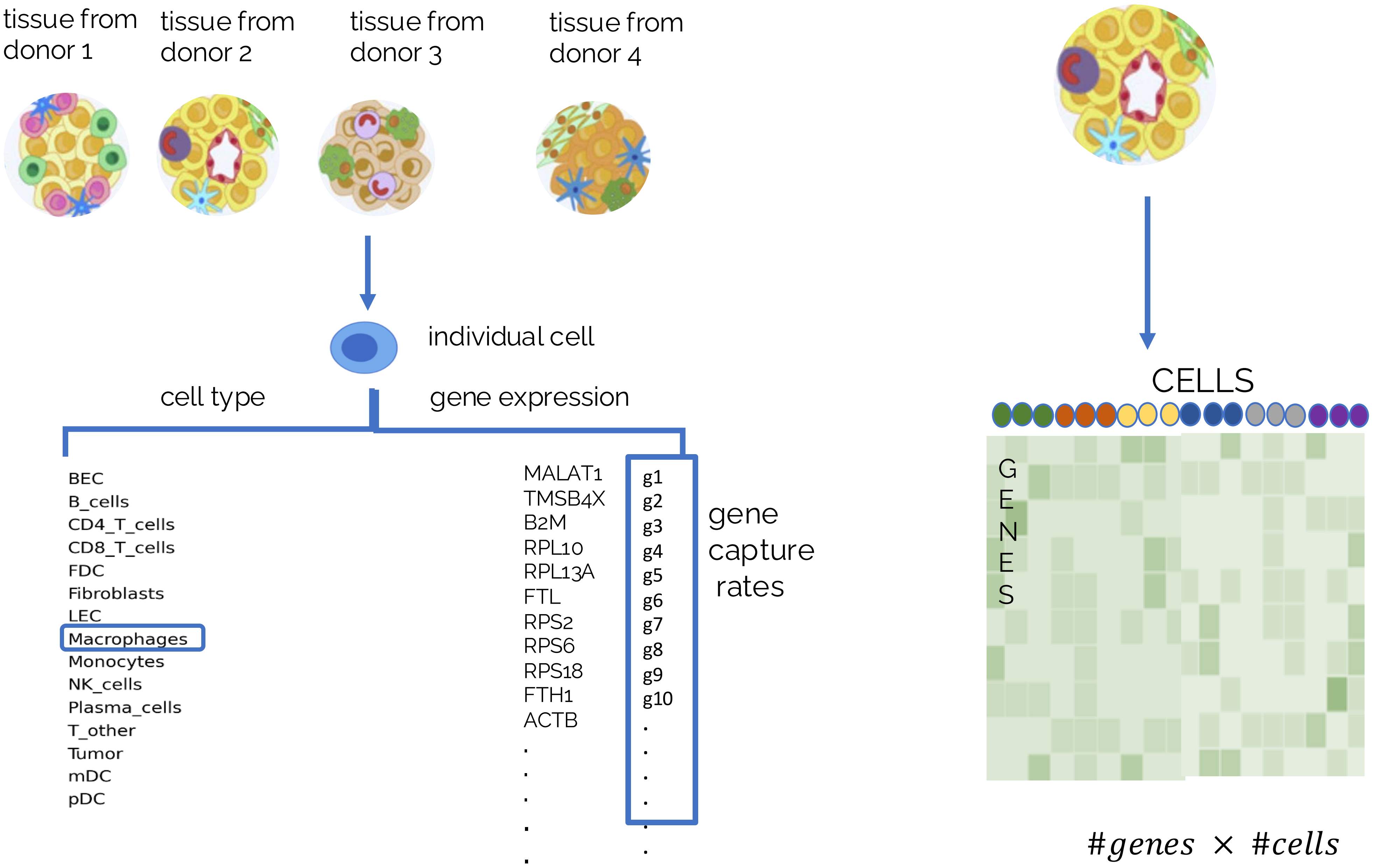


bulk-sequencing

single-cell

multi-omics

Multi-modal genomics data



How to represent represent and study this triple interaction between genes, cell types and donors

1

What is a tensor?

What is a tensor?

A tensor is a generalization of matrices to higher dimensions. We will explore this generalization from two perspectives:

Multilinear algebra

Matrices correspond to linear maps, whereas tensors correspond to multilinear maps.
- We will introduce multilinear maps, tensor product spaces

Data structure

Tensors are multi-dimensional arrays, we will discuss representation of tensor data.

Tensors are multilinear maps

multilinear map

Let $U_1, U_2 \dots, U_d$ be vector spaces. A function $f: U_1 \times U_2 \times \cdots \times U_d \to \mathbb{C}$ is called **multilinear** if it is linear in each variable.

For any $u_2 \in U_2, \dots, u_d \in U_d$ $f(:, u_2, \dots, u_d) \to \mathbb{C}$ is linear. Same for all $1 \leq i \leq d$.

$$f(u_1, \dots, a\, u_i + b\, u_i', \dots, u_d) = a\, f(\mathrm{u}_1, \dots, \mathrm{u}_\mathrm{i}, \dots, u_d) + b f(\mathrm{u}_1, \dots, u_i', \dots, u_d)$$

The space of all multilinear maps is denoted by $U_1^* \otimes U_2^* \otimes \cdots \otimes U_d^*$

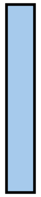
Elements $\mathcal{T} \in U_1^* \otimes U_2^* \otimes \cdots \otimes U_d^*$ are called **tensors.**

Tensor notations

Let $\mathcal{X}$ be tensor in $\mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$, then $(i_1, i_2, \ldots, i_d)$-th entry of $\mathcal{X}$ is denoted by $\boldsymbol{\mathcal{X}_{i_1 i_2 \ldots i_d}, 1 \leq i_1 \leq n_1, \ldots, 1 \leq i_d \leq n_d}$.

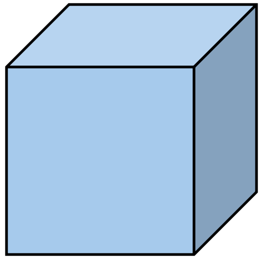
Order of a tensor

The **order** of a tensor is the number of dimensions, also known as **ways** or **modes.**
$\mathcal{X}$ in $\mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$ is an order-d (d-way) tensor, it has d modes.

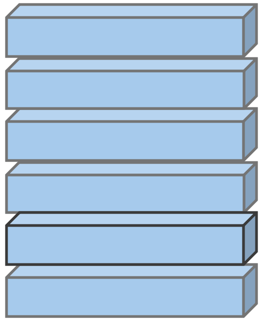


Vectors are order-1 tensors

matrices are order-2 tensors



order-3 tensor



order-4 tensor

Tensors are multi-dimensional arrays

a.k.a. multi-linear maps

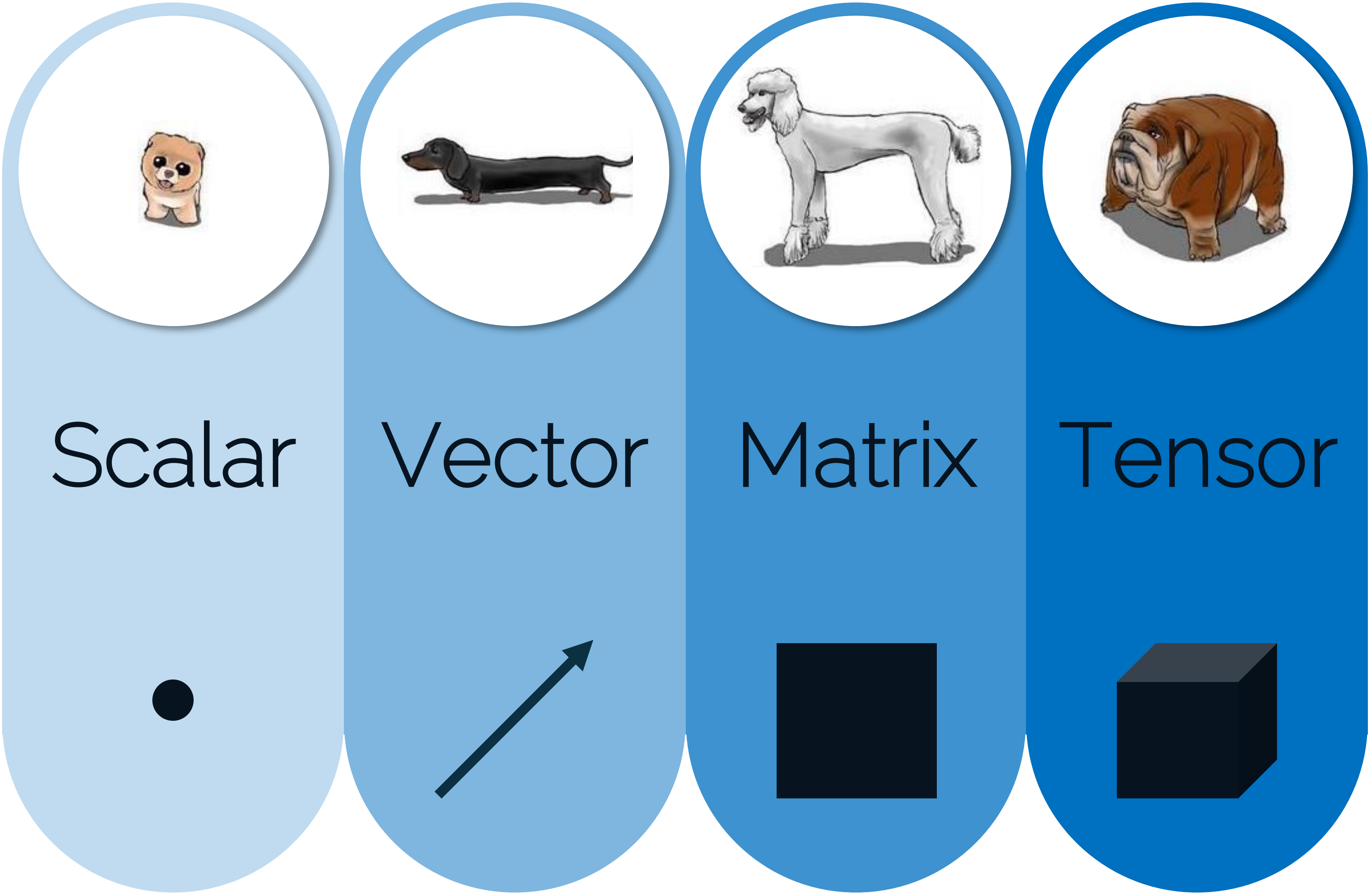
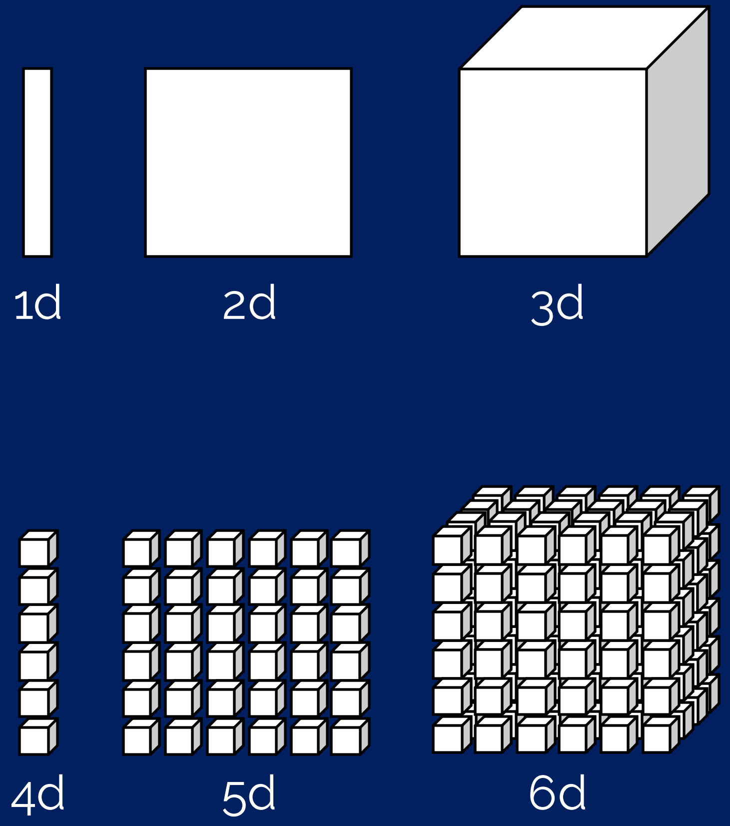
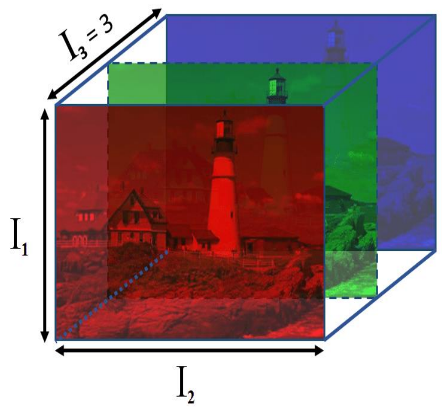


Figure credit: Anima Anandkumar

they can be considered as ***generalizations*** *of matrices to higher* ***dimensions***

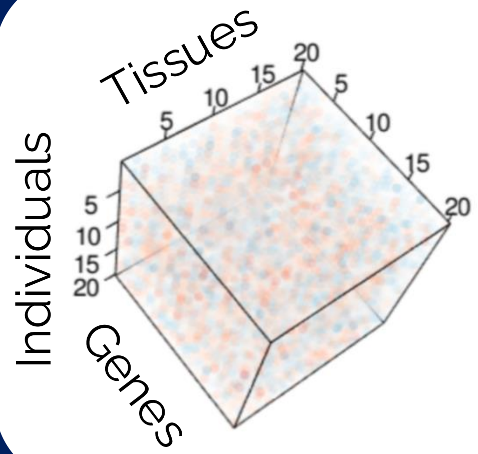


Tensors: a compact way to represent multi-modal data

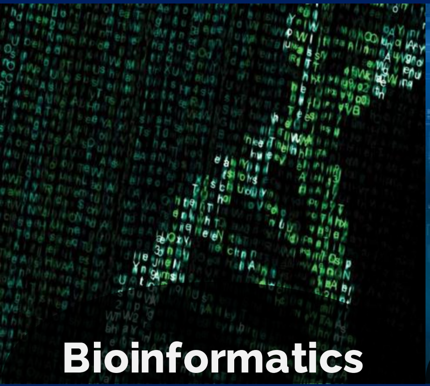


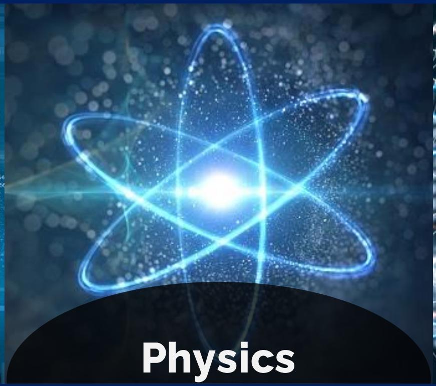
Color Images

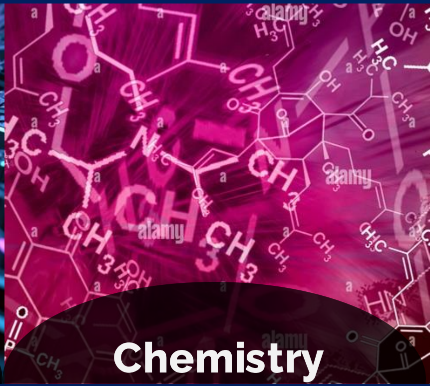
Multi-sample multi-tissue genomics data



Perturbational Datasets

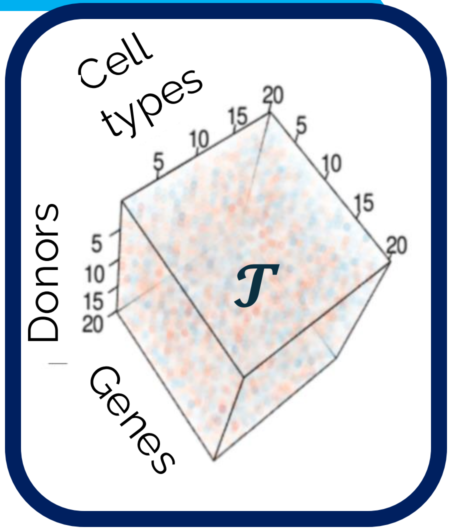






Tensor data

Multi-modal genomics data



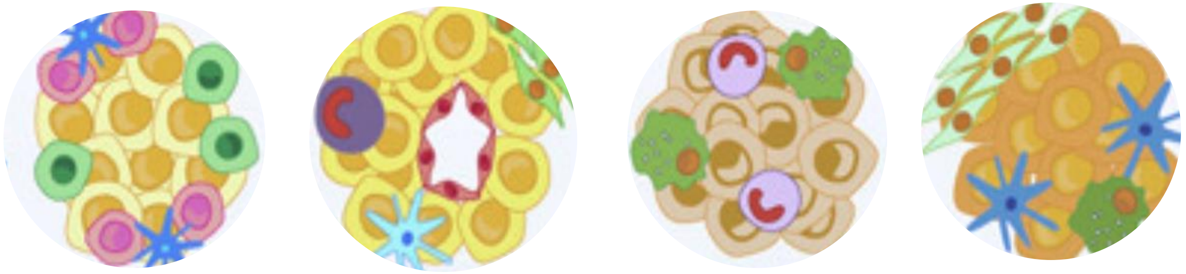
(Genes × Cell types × Donors)

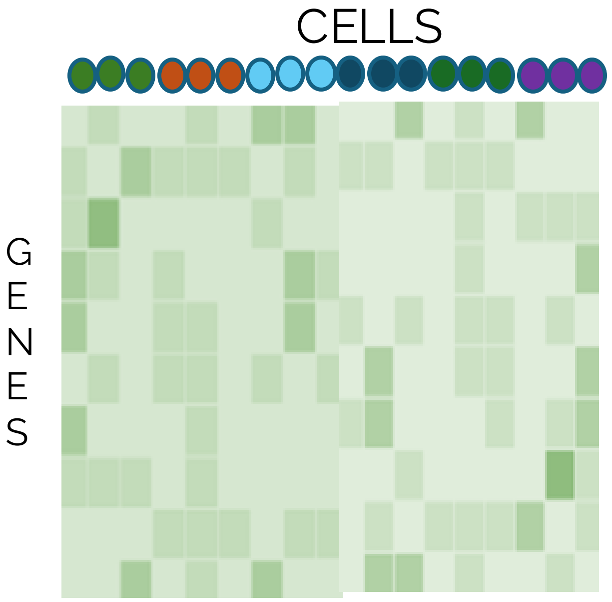
tissue from donor 1

tissue from donor 2

tissue from donor 3

tissue from donor 4





$\#genes \times \#cells$

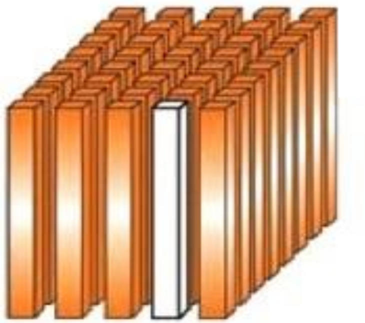
$\mathcal{T}_{123}$ = average expression of gene #1 at cell type #2 in the donor #3

Tensor fibers

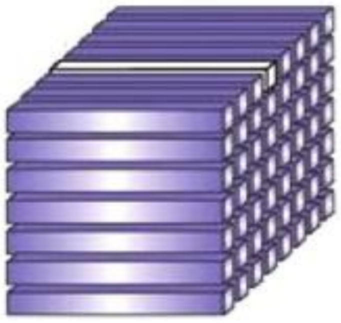
Fibers are generalizations of matrix rows and columns. A fiber is defined by fixing all but one index of a tensor.

Let $\mathcal{X}$ be a 3-way tensor of size $7 \times 5 \times 8$, then we can form fibers for each modality such as

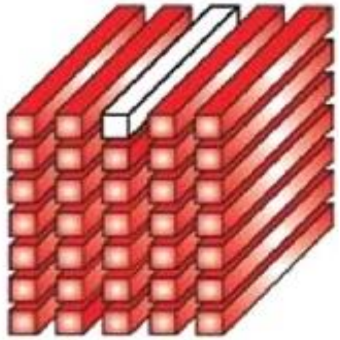
$$\mathcal{X}_{:12} \in \mathbb{C}^7, \mathcal{X}_{1:3} \in \mathbb{C}^5, \mathcal{X}_{34:} \in \mathbb{C}^8$$



mode-1
column fibers

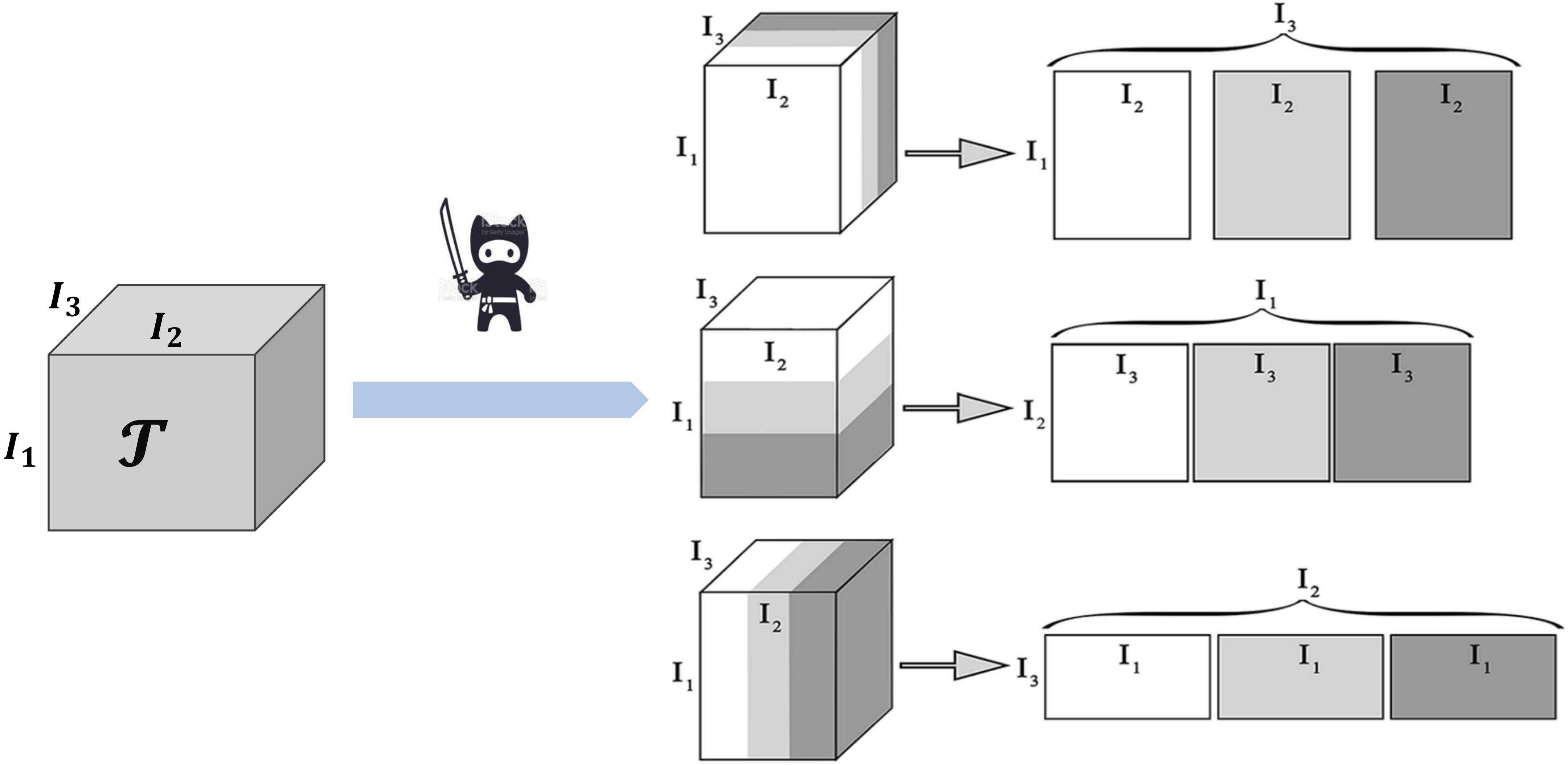


mode-2
row fibers



mode-3
tube fibers

Unfolding

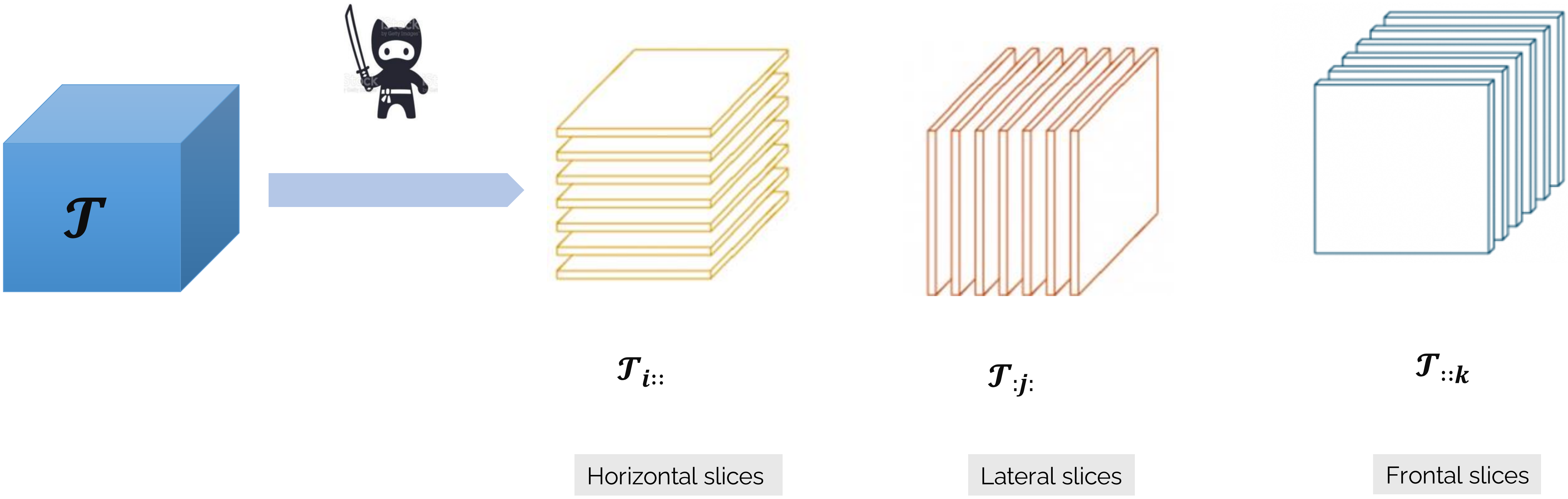


$\mathcal{T}_{(1)}$
$\mathcal{T}^{1|23}$

$\mathcal{T}_{(2)}$
$\mathcal{T}^{2|13}$

$\mathcal{T}_{(3)}$
$\mathcal{T}^{3|12}$

Slices


$\mathcal{T}$
$\mathcal{T}_{i::}$
$\mathcal{T}_{:j:}$
$\mathcal{T}_{::k}$
Horizontal slices
Lateral slices
Frontal slices

Slices

$\mathcal{T} \in \mathbb{R}^{3 \times 4 \times 2}$

$$\mathcal{T}_{::1} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \qquad \mathcal{T}_{::2} = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$

Fibers

$$\mathcal{T}_{:21} = [4\ 5\ 6] \qquad \mathcal{T}_{2:2} = [14\ 17\ 20\ 23] \qquad \mathcal{T}_{34:} = [12\ 24]$$

Slices

$$\mathcal{T}_{1::} = \begin{bmatrix} 1 & 13 \\ 4 & 16 \\ 7 & 19 \\ 10 & 20 \end{bmatrix} \qquad \mathcal{T}_{:2:} = \begin{bmatrix} 4 & 16 \\ 5 & 17 \\ 6 & 18 \end{bmatrix}$$

Unfolding

$\mathcal{T} \in \mathbb{R}^{3 \times 4 \times 2}$ $\quad \mathcal{T}_{::1} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$, $\quad \mathcal{T}_{::2} = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$

$$\mathcal{T}_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{bmatrix} \in \mathbb{R}^{3 \times 8}.$$

$$\mathcal{T}_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 19 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{bmatrix} \in \mathbb{R}^{4 \times 6}$$

$$\mathcal{T}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \end{bmatrix} \in \mathbb{R}^{2 \times 12}.$$

Outer product, Kronecker product

Rank 1 matrix

The vector **outer product** of $a \in \mathbb{R}^I, b \in R^J$ is $a \otimes b = ab^T \in \mathbb{R}^{I \times J}$

$$a \otimes b = \begin{bmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_J \\ a_2b_1 & a_2b_2 & \dots & a_2b_J \\ \vdots & \vdots & \vdots & \vdots \\ a_Ib_1 & a_Ib_2 & \dots & a_Ib_J \end{bmatrix} \in R^{I \times J}$$

The **Kronecker product** of $A \in \mathbb{R}^{I \times J}$, $\mathrm{B} \in \mathbb{R}^{K \times L}$

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1J}B \\ a_{21}B & a_{22}B & \dots & a_{2J}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{I1}B & a_{I2}B & \dots & a_{IJ}B \end{bmatrix} \in \mathbb{R}^{IK \times JL}$$

$$A \otimes B = [a_1 \otimes b_1 \quad a_1 \otimes b_2 \quad \dots \quad a_J \otimes b_1 \quad \dots \quad a_J \otimes b_L]$$

Kronecker product: examples (matrix direct product)

special case of the tensor product space

$$A \in \mathbb{R}^{I \times J},\ B \in \mathbb{R}^{K \times L} \Rightarrow \quad A : \mathbb{R}^{J} \to \mathbb{R}^{I} \ \ \& \ \ B : \mathbb{R}^{L} \to \mathbb{R}^{K}$$

Kronecker product constructs a bilinear map: $A \otimes B : \mathbb{R}^{J \times L} \to \mathbb{R}^{I \times K}$

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix}$$

Kronecker product results in a block matrix

$$A \otimes B = \begin{bmatrix} 1 \cdot B & 2 \cdot B \\ 3 \cdot B & 4 \cdot B \end{bmatrix}$$

$$A \otimes B = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix}$$

Properties of Kronecker product

1 $$A \otimes (B + C) = A \otimes B + A \otimes C$$
$$(A + B) \otimes C = A \otimes C + B \otimes C$$
distributive

2 $$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$
associative

3 $$c(A \otimes B) = (cA) \otimes B = A \otimes (cB)$$
scalar multiplication

4 $$(A \otimes B) \otimes (X \otimes Y) = (AX) \otimes (BY)$$
for compatible matrices
mixed product

5 $$(A \otimes B)^\top = A^\top \otimes B^\top$$
transpose

Properties of Kronecker product

$A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}$

λ_i's are eigenvalues of A, μ_j's are eigenvalues of B, then eigenvalues of $A \otimes B = \{\lambda_i \mu_j : \forall\, i, j\}$

$A\, v_i = \lambda_i v_i$ and $B u_j = \mu_j u_j \implies (A \otimes B)(v_i \otimes u_j) = (Av_i) \otimes (Bu_j)$

$(A \otimes B)(v_i \otimes u_j) = (\lambda_i\, v_i) \otimes (\mu_j\, u_j) \implies (A \otimes B)(v_i \otimes u_j) = \boldsymbol{\lambda_i\, \mu_j}(v_i \otimes u_j)$

$$\det(A \otimes B) = \prod_{i=1}^{m}\prod_{j=1}^{n}(\lambda_i \mu_j)$$

$$\prod_{i=1}^{m}\prod_{j=1}^{n}(\lambda_i \mu_j) = \left(\prod_{i=1}^{m}\lambda_i^n\right)\cdot\left(\prod_{j=1}^{n}\mu_j^m\right) \implies \det(A \otimes B) = (\det A)^n \cdot (\det B)^m.$$

Properties of Kronecker product

- $tr(A \otimes B) = tr(A)tr(B)$

- $\text{rank}(A \otimes B) = \text{rank}(A) \cdot \text{rank}(B).$

column space of A is spanned by $r_A = Rank(A)$ linearly independent columns $\{a_1, \ldots, a_{R_A}\}$

column space of B is spanned by $r_B = Rank(B)$ linearly independent columns $\{b_1, \ldots, b_{R_B}\}$

column space of $A \otimes B$ is spanned by $\{a_1 \otimes b_1, \ldots, a_{R_A} \otimes b_{R_B}\}$

- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if A and B are invertible.

- $(A \otimes B)^{\dagger} = A^{\dagger} \otimes B^{\dagger}$

- $(A \otimes B)vec(V) = vec(BVA^{\top})$

Khatri-Rao product, Hadamard product

The **Khatri-Rao product** of $A \in \mathbb{R}^{I \times K}$, $B \in \mathbb{R}^{J \times K}$ is the matching-columnwise Kronecker product

$$A \odot B = [a_1 \otimes b_1 \quad a_2 \otimes b_2 \quad \dots \quad a_K \otimes b_K] \in \mathbb{R}^{IJ \times K}$$

The **Hadamard product** of $A \in \mathbb{R}^{I \times J}$, $B \in \mathbb{R}^{I \times J}$ is the elementwise matrix product

$$A * B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2J}b_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \dots & a_{IJ}b_{IJ} \end{bmatrix} \in \mathbb{R}^{I \times J}$$

Properties of Khatri-Rao product

Matching columnwise Kronecker product– so previous properties listed for Kronecker product hold such as associativity, distributivity

- $(A \odot B) \odot C = A \odot (B \odot C)$

for other properties see Kronecker product

- $(A + B) \odot C = A \odot C + B \odot C$

- $(A \odot B)^{\top} (A \odot B) = A^{\top} A * B^{\top} B$

$$((A \odot B)^{\top} (A \odot B))_{ij} = \langle (a_i \otimes b_i), (a_j \otimes b_j) \rangle$$

$$\langle a_i \otimes b_i, a_j \otimes b_j \rangle = \langle a_i, a_j \rangle \cdot \langle b_i, b_j \rangle$$

$$((A \odot B)^{\top} (A \odot B))_{ij} = (A^{\top} A)_{ij} \cdot (B^{\top} B)_{ij}.$$

Properties of Khatri-Rao product

$$(A \odot B)^{\dagger} = ((A^{\top}A) * (B^{\top}B))^{\dagger}(A \odot B)^{\top}$$

first, we define the pseudoinverse

$A \in \mathbb{R}^{n\times m}$, pseudoinverse of A is defined as a matrix $A^{\dagger} \in \mathbb{R}^{n\times m}$ such that

1. $AA^{\dagger}$ maps all column vectors of A to themselves $AA^{\dagger}A = A$
2. $A^{\dagger}$ acts as a weak inverse $A^{\dagger}AA^{\dagger} = A^{\dagger}$
3. $A^{\dagger}A$ is Hermitian $(A^{\dagger}A)^{*} = A^{\dagger}A$
4. $AA^{\dagger}$ is Hermitian $(AA^{\dagger})^{*} = AA^{\dagger}$

It should satisfy these 4 properties
for every matrix there is one and only
One pseudo-inverse

We can also conclude $A^{\dagger} = (A^{*}A)^{\dagger}A^{*}$

Properties of Khatri-Rao product

$$(A \odot B)^{\dagger} = ((A^{\top}A) * (B^{\top}B))^{\dagger}(A \odot B)^{\top}$$ for real matrices

$$(A \odot B)^{\dagger} = ((A \odot B)^{\top}(A \odot B))^{\dagger}(A \odot B)^{\top}$$

we use the property $A^{\dagger} = (A^{*}A)^{\dagger}A^{*}$

$$= (A^{\top}A * B^{\top}B)^{\dagger}(A \odot B)^{\top}$$

we use the property $(A \odot B)^{\top}(A \odot B) = A^{\top}A * B^{\top}B$

Mode-n product

$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and $A \in \mathbb{R}^{J \times I_n}$, then the **mode-n** product can be given as

$$(\mathcal{X}_{\times n} A)_{i_1 i_2 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1 i_2 \ldots i_n \ldots i_N} A_{j i_n} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N}$$

every mode-n fiber is multiplied by the matrix A

mode-n product is related to a change of basis in the case when a tensor defines a multilinear map

Properties:

- $\mathcal{X}_{\times n} A_{\times m} B = \mathcal{X}_{\times m} B_{\times n} A \quad (m \neq n)$ order independence across modes
- $\mathcal{X}_{\times n} A_{\times n} B = \mathcal{X}_{\times n} (BA)$ if the modes are the same

Mode-n product

We can also express it in terms of unfolded tensors: $\mathcal{X} \times_n A = \mathcal{Y} \iff A\mathcal{X}_{(n)} = \mathcal{Y}_{(n)}$

Example: $\mathcal{T}_{::1} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ $\mathcal{T}_{::2} = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$ $\Rightarrow \mathcal{T}$ is $2 \times 3 \times$ $\mathbf{U} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$

we can multiply $\mathcal{T}$ and U along 2th mode

$\mathcal{T} \times_2 U = \mathcal{Y}$ of size $2 \times 4 \times 2$

$$\mathcal{T}_{(2)} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \Rightarrow U\mathcal{T}_{(2)} = \begin{bmatrix} 10 & 28 & 46 & 64 \\ 11 & 31 & 51 & 71 \\ 3 & 9 & 15 & 21 \\ 5 & 14 & 23 & 32 \end{bmatrix} = \mathcal{Y}_{(2)}$$

$\Rightarrow$ we need to reshape it

Notice that we expanded the second dimension

Common usage is reducing the dimension, i.e., compressing the tensor

Tensor inner product

For $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the inner product of tensors $\mathcal{X}, \mathcal{Y}$:

$$< \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} >= \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \boldsymbol{\mathcal{X}}_{i_1 i_2 \ldots i_d} \boldsymbol{\mathcal{Y}}_{i_1 i_2 \ldots i_d}$$

For $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, then the Frobenius norm of tensor $\mathcal{X}$ is given as

$$\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \boldsymbol{\mathcal{X}}^2_{i_1 i_2 \ldots i_d}}$$

Distance (or similarity) between tensors?

- $\boldsymbol{d(\mathcal{T}, \mathcal{T}') = \|\mathcal{T} - \mathcal{T}'\|_F}$ (not a probabilistic approach, assumes normal noise)

- **Kullback-Leibner divergence**

  $$D_{KL}(\mathcal{T} \| \mathcal{T}) = \sum_{z \in Z} \mathcal{T}(z) log \frac{\mathcal{T}(z)}{\mathcal{T}'(z)}$$

  (probabilistic approach) –distance metric for tensors is an active research area, we will discuss different distance/similarity matrices

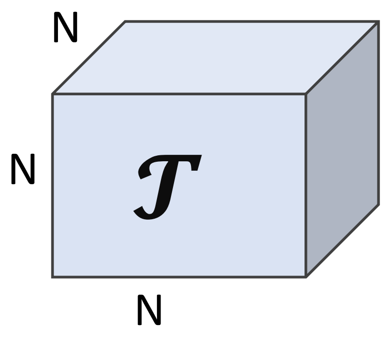
Symmetric tensors

Let **V** be a vector space of dimension N and $\mathcal{T} \in \underbrace{V \otimes V \otimes \cdots \otimes V}_{d \text{ times}} = V^{\otimes d}$

$\mathcal{T}$ is a d-way tensor of size $\underbrace{N \times N \times \cdots \times N}_{d \text{ times}}$, then $\mathcal{T}$ is symmetric

$$\mathcal{T}_{i_1 i_2 \dots i_d} = \mathcal{T}_{\sigma(i_1)\sigma(i_2)\dots\sigma(i_d)}$$

for all permutation $\sigma \in S_n$



$\mathcal{T}_{123} = \mathcal{T}_{132} = \mathcal{T}_{312} =$
$\mathcal{T}_{321} = \mathcal{T}_{231} = \mathcal{T}_{213}$

Assume $N \geq 3$

$S^d(V)$: set of the space of all symmetric tensors of order *d* defined on *V*

$S^d(\mathbb{C}^n)$: set of all symmetric tensors of order d represents the space of symmetric tensors over $\mathbb{C}^n$

$$S(\mathbb{C}^n) = \bigoplus_d S^d(\mathbb{C}^n)$$

$S(\mathbb{C}^n)$ space of symmetric tensors

Symmetric tensors :

homogenous polynomials

$\mathcal{T} \in S^d(\mathbb{C}^n) \Leftrightarrow f(\mathcal{T}) \in \mathbb{C}_d[x_1, x_2, x_n]$ (polynomial of degree d with n variables)

$$f_T(x_1, x_2, \dots, x_n) = \sum \mathcal{T}_{i_1 i_2 \dots i_d} x_{i_1} x_{i_2} \dots x_{i_n}$$

$$i_1 i_2 \dots i_d = \underbrace{1 \dots 1}_{j_1 \text{ times}} \underbrace{2 \dots 2}_{j_2 \text{ times}} \dots \underbrace{n \dots n}_{j_n \text{ times}}$$

$$j_1 + j_2 + \dots + j_n = d$$

$$\binom{d}{j_1, j_2, \dots, j_n} = \frac{d!}{j_1! j_2! \cdots j_n!}$$

$$f_T(x_1, \dots, x_n) = \sum_{j_1 + j_2 + \dots + j_n = d} \binom{d}{j_1 j_2 \dots j_n} \mathcal{T}_{\underbrace{1 \dots 1}_{j_1 \text{ times}} 2 \dots 2 \dots n \dots n} x_1^{j_1} x_2^{j_2} \dots x_d^{j_n}$$

Symmetric tensors :
matrix case

Let M be symmetric matrix, $M \in S^2(C^n)$ - what is the corresponding homogeneous polynomial?

$$f_M(x_1, x_2, \ldots, x_n) = \sum_{1 \leq i,j \leq n} M_{i,j} x_i x_j = \sum_{1 \leq i \leq n} M_{i,i} x_i^2 + \sum_{1 \leq i \leq j \leq n} 2 M_{i,j} x_i x_j$$

$$f_M(x_1, x_2, \ldots, x_n) = x^T M x, \text{ where } x = [x_1, x_2, \ldots, x_n]$$

homogeneous quadratic polynomial with n variables

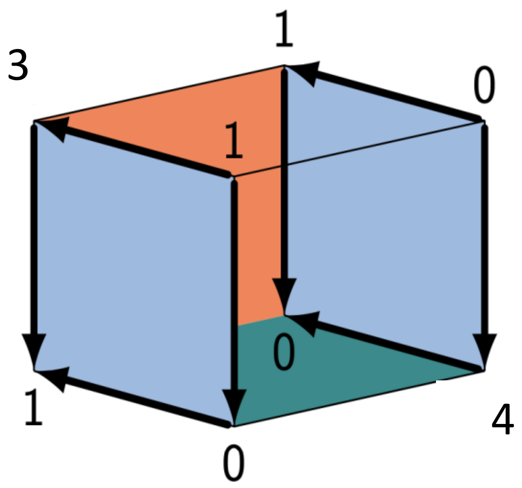
Example:

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \implies f_M(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= a\, x_1^2 + 2\, a\, b\, x_1 x_2 + c\, x_2^2$$

note that $x^T M\, x$ appears in quadratic programming

Symmetric tensors :



$$\mathcal{T} = \left[\begin{matrix} 3 & 1 \\ 1 & 0 \end{matrix}\middle|\begin{matrix} 1 & 0 \\ 0 & 4 \end{matrix}\right]$$

$\mathcal{T}$ is a $2 \times 2 \times 2$ symmetric tensor

$\mathcal{T}_{111}=3$	x_1^3
$\mathcal{T}_{112}=\mathcal{T}_{121}=\mathcal{T}_{211}=1$	$3x_1^2\,x_2$
$\mathcal{T}_{122}=\mathcal{T}_{212}=\mathcal{T}_{221}=0$	$3x_1\,x_2^2$
$\mathcal{T}_{222}=4$	x_2^3

$$f_T(x_1, x_2) = 3x_1^3 + 1 * 3\,x_1^2 x_2 + 0 * 3x_1\,x_2^2 + 4 * x_2^3$$

$$f_T(x_1, x_2) = 3x_1^3 + 3x_1^2 x_2 + 4\,x_2^3$$

Rank-1 tensor

rank 1 matrix

The vector outer product of $u \in \mathbb{R}^{n_1}, v \in R^{n_2}$ is $u \otimes v = uv^T \in \mathbb{R}^{n_1 \times n_2}$

A d-way rank 1 tensor $\mathcal{T}$ of size $n_1 \times n_2 \times \cdots \times n_d$ is written as outer product of d vectors

$$\mathcal{T} = u^1 \otimes u^2 \otimes \cdots \otimes u^d ,$$

rank 1 tensor

pure (simple) tensor

where $u^i \in \mathbb{R}^{n_i}, 1 \leq i \leq d.$

$$\mathcal{T}_{i_1 i_2 \dots i_d} = u^1_{i_1}\, u^2_{i_2} \dots u^d_{i_d}$$

represents a **separable signal** which can be expressed as the combination of independent factors from each mode.

Rank-1 tensor

example

$$u^1 = [1, 2, 3] \in \mathbb{R}^3$$

$$u^2 = [4, 5, 6, 7] \in \mathbb{R}^4$$

$$u^3 = [8, 9] \in \mathbb{R}^2$$

$\boldsymbol{\mathcal{T}} = \boldsymbol{u^1} \otimes \boldsymbol{u^2} \otimes \boldsymbol{u^3}$, $\quad \boldsymbol{\mathcal{T}_{ijk}} = \boldsymbol{u_i^1} \otimes \boldsymbol{u_j^2} \otimes \boldsymbol{u_k^3}$, $3 \times 4 \times 2$ tensor

$$\mathcal{T}_{::1} = 8\, u^1 \otimes u^2 = 8\,[1\ 2\ 3][4\ 5\ 6\ 7]^{\mathsf{T}} = 8 \begin{bmatrix} 4 & 5 & 6 & 7 \\ 8 & 10 & 12 & 14 \\ 12 & 15 & 18 & 21 \end{bmatrix} = \begin{bmatrix} 32 & 40 & 48 & 56 \\ 64 & 80 & 96 & 112 \\ 96 & 120 & 144 & 168 \end{bmatrix}$$

$$\mathcal{T}_{::2} = 9 u^1 \otimes u^2 = 9\,[1\ 2\ 3][4\ 5\ 6\ 7]^{\mathsf{T}} = 9 \begin{bmatrix} 4 & 5 & 6 & 7 \\ 8 & 10 & 12 & 14 \\ 12 & 15 & 18 & 21 \end{bmatrix} = \begin{bmatrix} 36 & 45 & 54 & 63 \\ 72 & 90 & 108 & 126 \\ 108 & 135 & 162 & 189 \end{bmatrix}$$

Hidden variable models

independent random variables: rank 1 tensors

Given independent random variables $X_1, X_2, \dots X_d$ with $X_i \in \{x_1, x_2, \dots, x_{n_i}\}$, their joint distribution can be written as product of their marginal distributions:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_d = x_d)$$

The joint distribution can be represented with d –way tensor $\boldsymbol{\mathcal{T}}$ such that

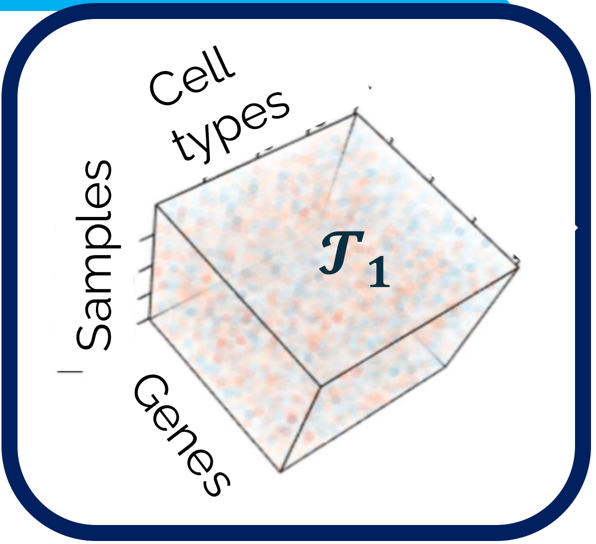
$$\boldsymbol{\mathcal{T}}_{\boldsymbol{i_1, i_2, \dots, i_d}} = P\left(X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_d = x_{i_d}\right)$$

$$\boldsymbol{\mathcal{T}} = P(X_1) \otimes P(X_2) \otimes \cdots \otimes P(X_d)$$

rank-1 tensor naturally represents a system where each dimension corresponds to an independent random variable, and the tensor entries represent the product of probabilities (or related measures) associated with each independent variable.

Rank-1 tensor : unit of an expression pattern

Multi-modal Genomics Data



(**Genes** × **Cell types** × **Samples**)

Assume that $\mathcal{T}_1$ is a rank 1 tensor of size $N_g \times N_c \times N_s$ where

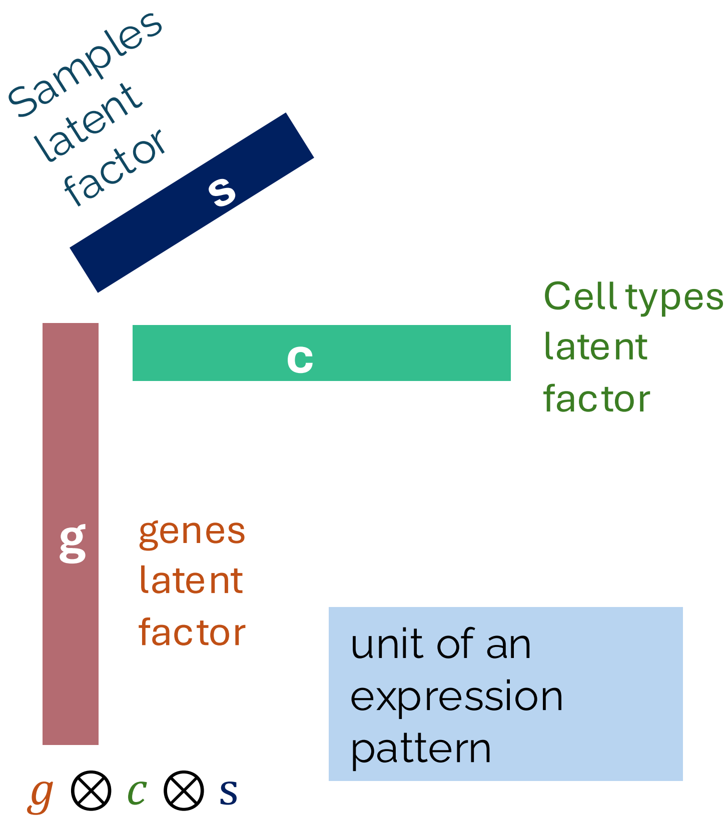
N_g = number of genes (20.000)
N_c = number of cell types (10)
N_s = number of samples (40)

$\mathcal{T}_1 = \boldsymbol{g} \otimes \boldsymbol{c} \otimes \boldsymbol{s}$

g : genes latent factor

c : cell types latent factor

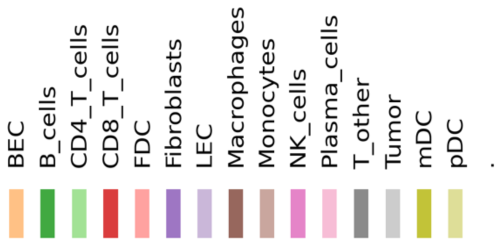
s : samples latent factor

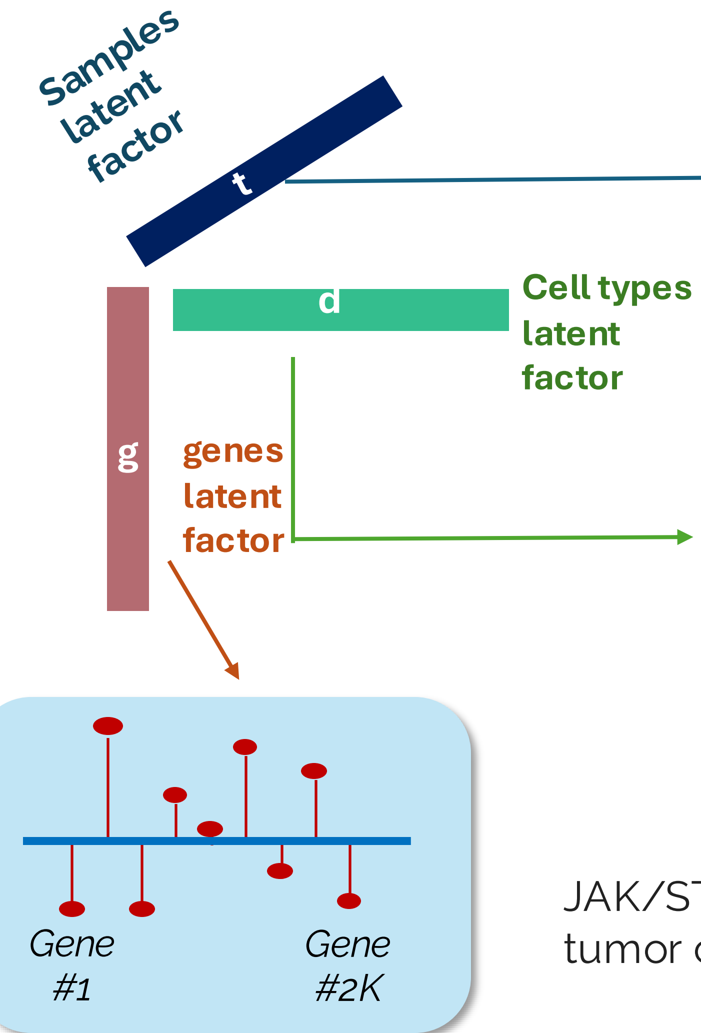


Samples/ patient groups

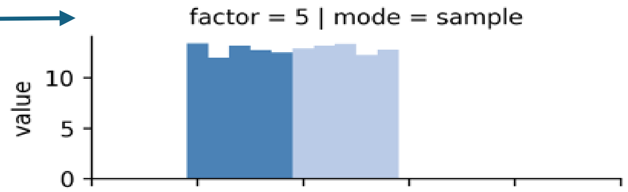
HL: Hodgkin lymphoma
EBV+: Epstein-Barr virus positive
EBV-: Epstein-Barr virus negative
RLN: Reactive Lymph node

cell types



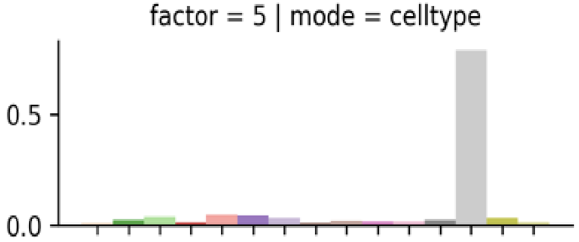


Samples latent factor



Donor type influence
Donor classification

Cell types latent



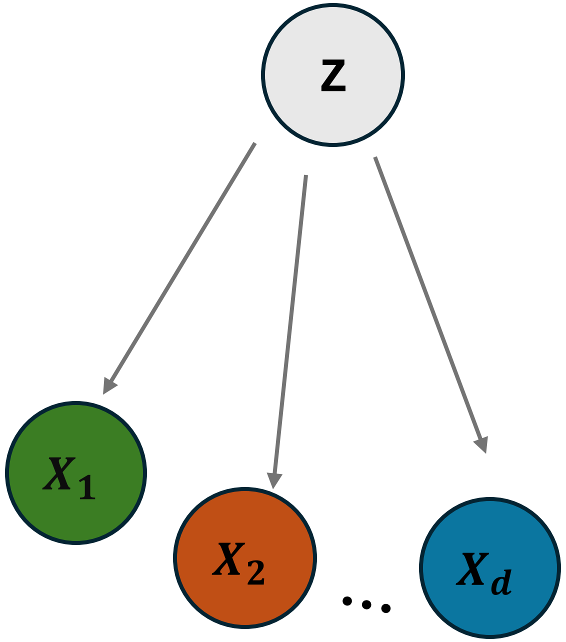
Clusters of cell types

Clusters of Gene Expression

JAK/STAT pathway promotes
tumor cell proliferation and survival of tumor cells

Conditionally independent variables

Suppose that the random variables $X_1, X_2, \ldots X_d$ are conditionally independent, given $Z = j$.



The **conditional distribution**

$$P(X_1, X_2, \ldots, X_d \mid Z = j) = P(X_1 \mid Z = j) P(X_2 \mid Z = j) \ldots P(X_d \mid Z = j)$$

The **total distribution** is obtained by summing over Z . Suppose $Z \in \{1, \ldots, r\}$, then the total distribution has the structure as a sum of r rank-1 components:

$$P(X_1, X_2, \ldots, X_d) = \sum_{j=1}^{R} P(Z = j) P(X_1 \mid Z = j) P(X_2 \mid Z = j) \ldots P(X_d \mid Z = j)$$

hidden variable distribution

conditional distribution

$$\mathcal{T}_{i_1, i_2, \ldots, i_d} = P\left(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_d = x_{i_d}\right) \Longrightarrow \mathcal{T} = \sum_{j=1}^{R} P(Z = j) P(X_1 \mid Z = j) \otimes P(X_2 \mid Z = j) \otimes \cdots \otimes P(X_d \mid Z = j)$$

Rank R tensor

Matrix factorization

$X \in \mathbb{R}^{m \times n}$ $X \approx AB^\top$ where $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{n \times r}$

Singular Value Decomposition (SVD)

data compression
extracts meaningful pattern

Decomposition into orthogonal matrices and singular values

$X \approx U\Lambda V^\top$ where U and V are orthogonal matrices (as in principal component analysis (PCA))

Non-negative matrix factorization (NMF)

$X \approx WH, \quad W, H \geq 0$ non-negativity constraints for interpretability applications in topic modeling

QR factorization

factorization into orthogonal matrix Q and upper-triangular matrix R

$X = QR$ often used to solve least squares problems

LU Decomposition

factorization into lower triangular matrix L and upper triangular matrix U

$X = LU$ often used to solve linear systems

Eckart-Young theorem

For a given matrix $A \in \mathbb{R}^{m\times n}$ of rank R with singular value decomposition (SVD) $A = U\Sigma V^\top$, the best rank-k approximation A_k, in terms of the Frobenius norm or spectral norm, is given by:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top,$$

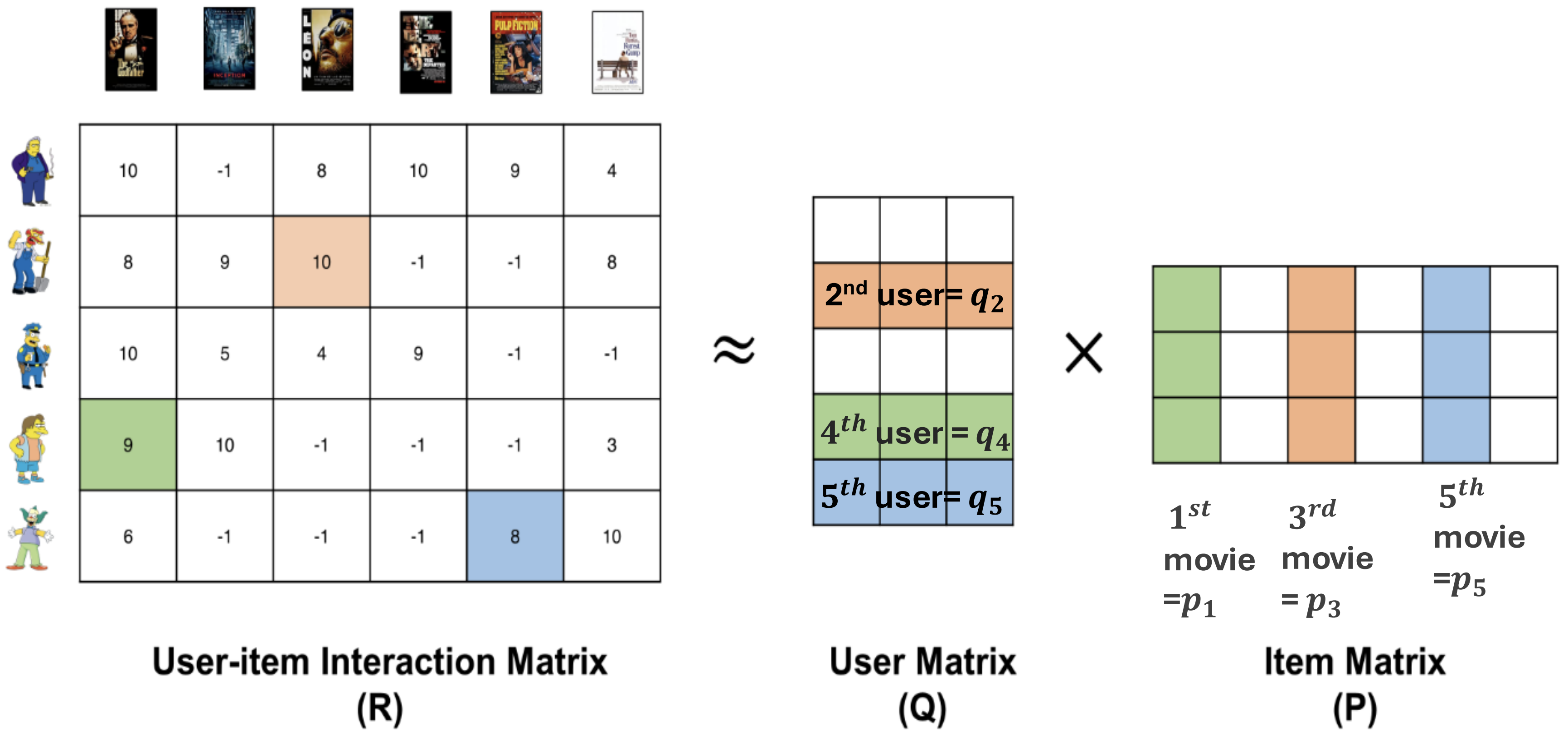
where σ_i are the singular values of A (in descending order), and u_i and v_i are the corresponding left and right singular vectors. This theorem ensures A_k minimizes the approximation error:

$$\|A - A_k\|_F \quad \text{or} \quad \|A - A_k\|_2,$$

where $\|\cdot\|_F$ is the Frobenius norm and $\|\cdot\|_2$ is the spectral norm.

Matrix factorization

collaborative filtering



$$R_{ij} \approx \mathbf{q_i}\, \mathbf{p_j}$$

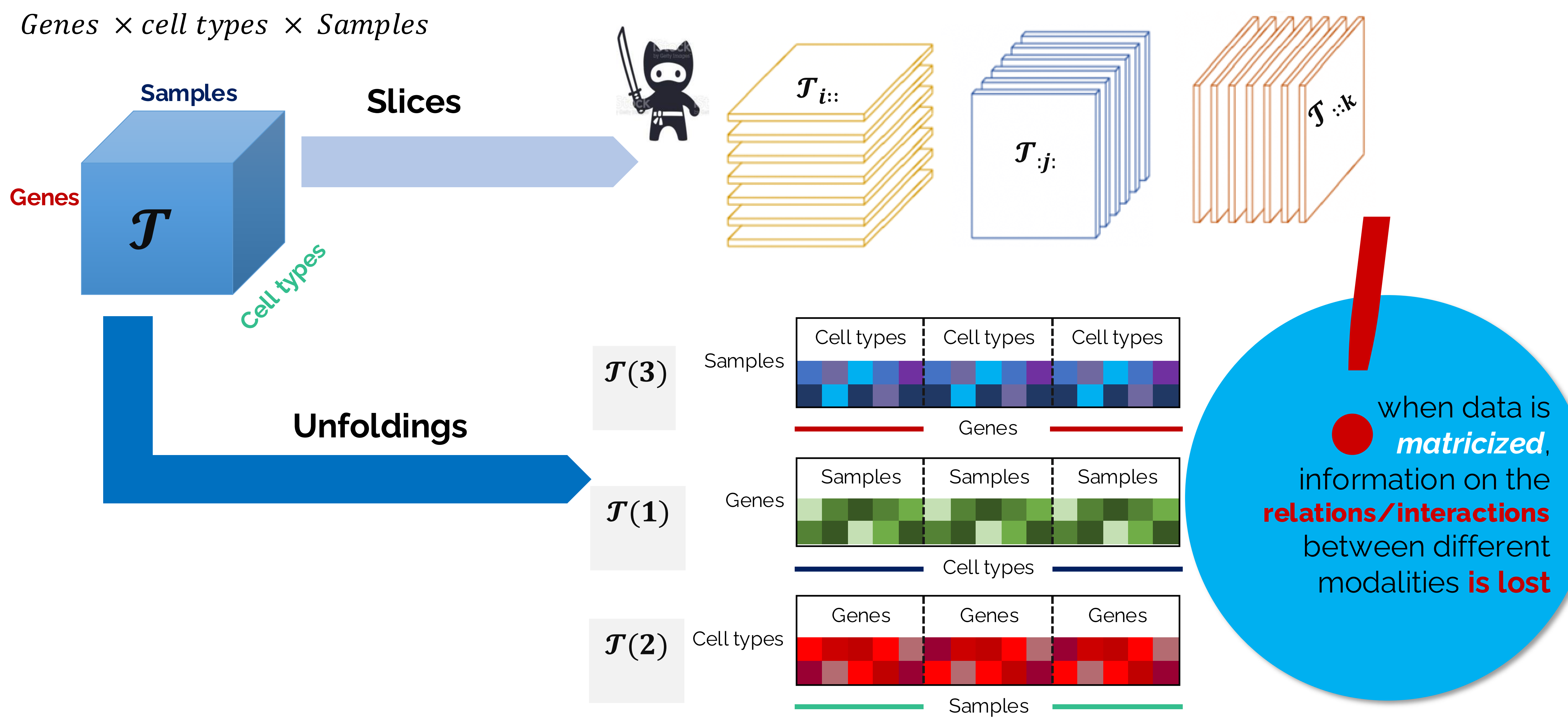
$$R_{23}\left(\text{estimated rate of } 2^{nd} \text{ user for } 3^{rd} \text{ movie}\right) \approx q_2\, p_3$$

Latent features from the factorization capture correlations in previous user-item interactions, enabling user and item matrices to approximate these patterns and predict unknown ratings

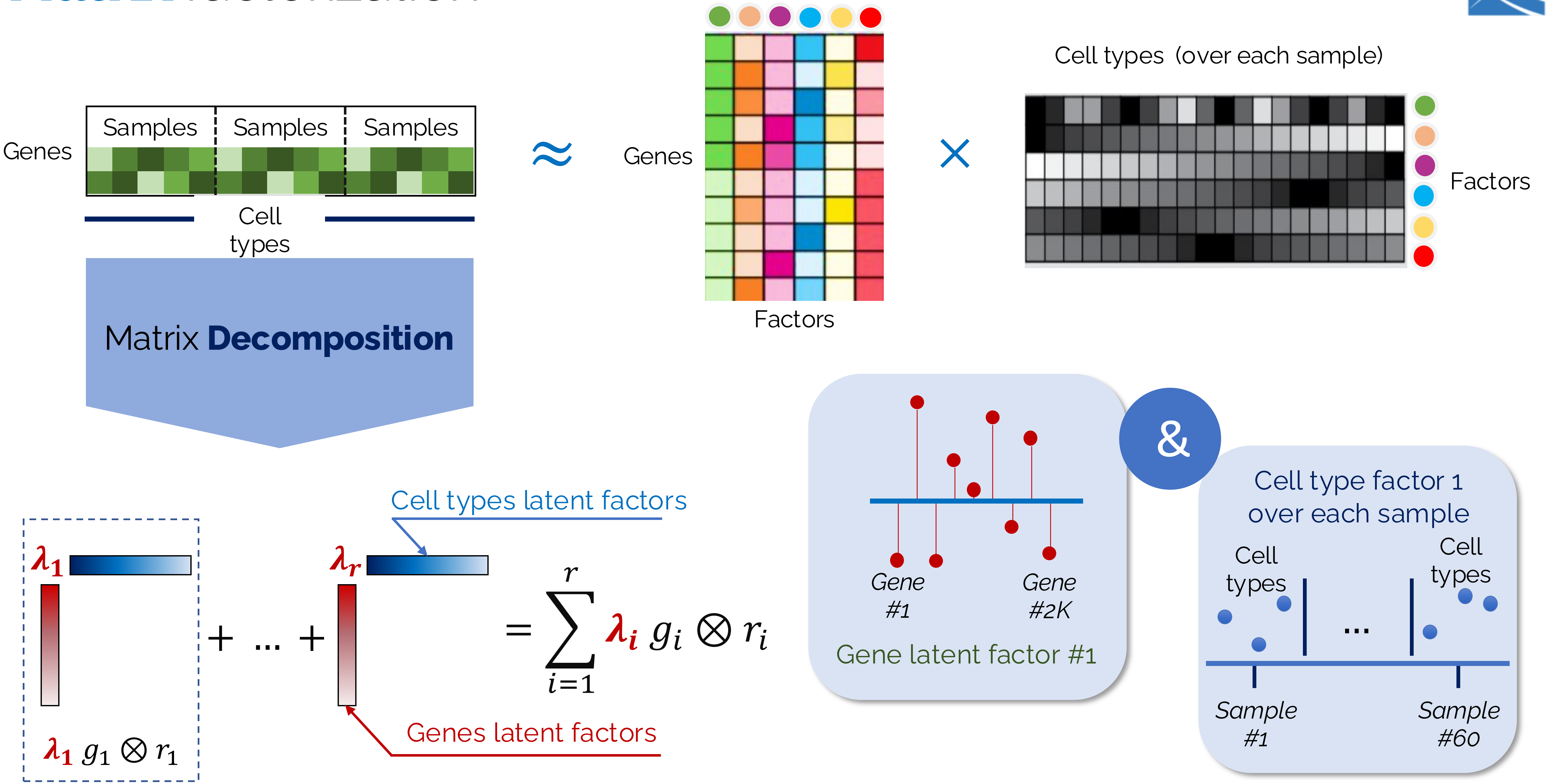
*This example and figures are taken from Codinfox Buomsoo Kim

Low rank factorization

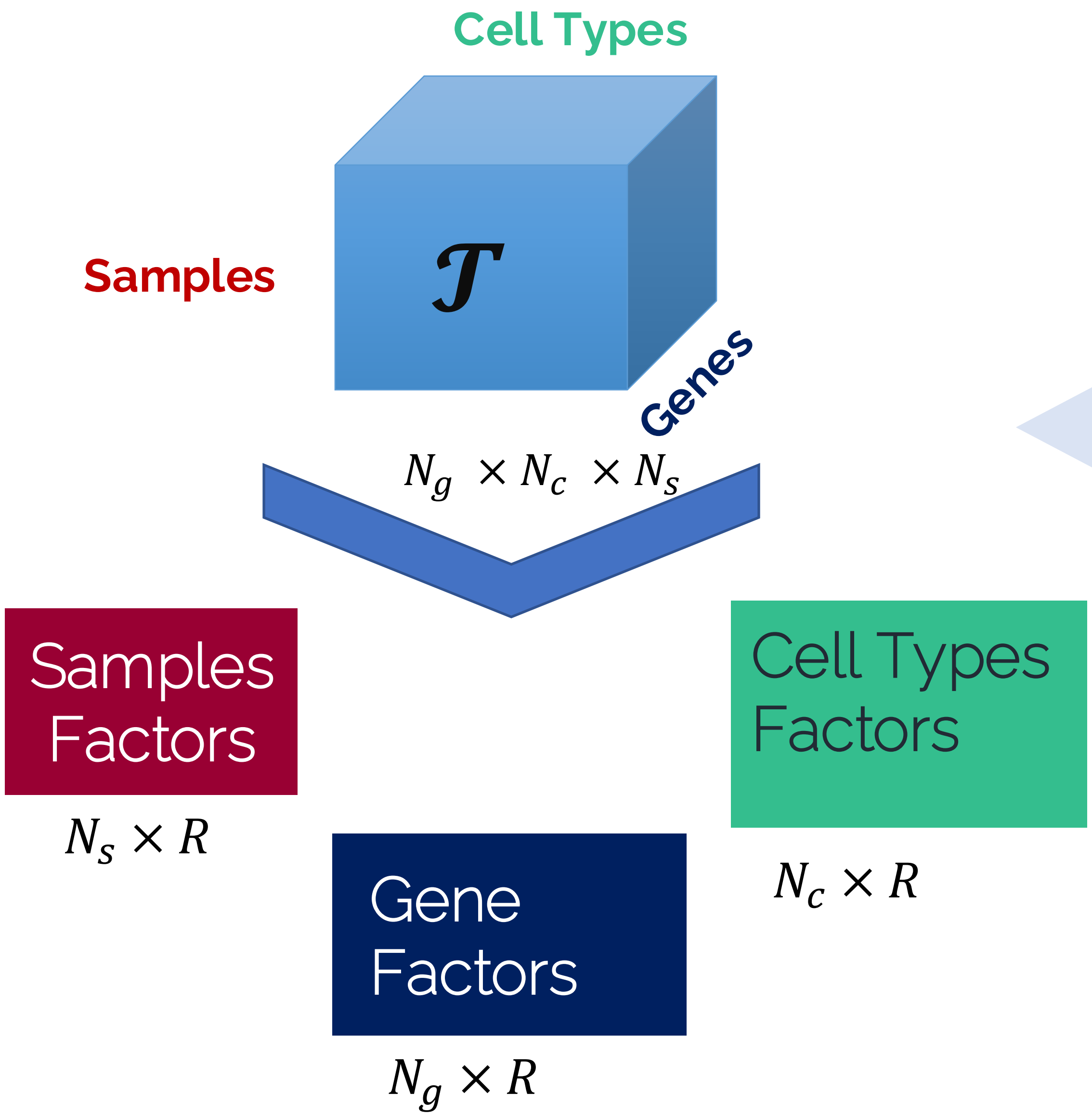
matrix case

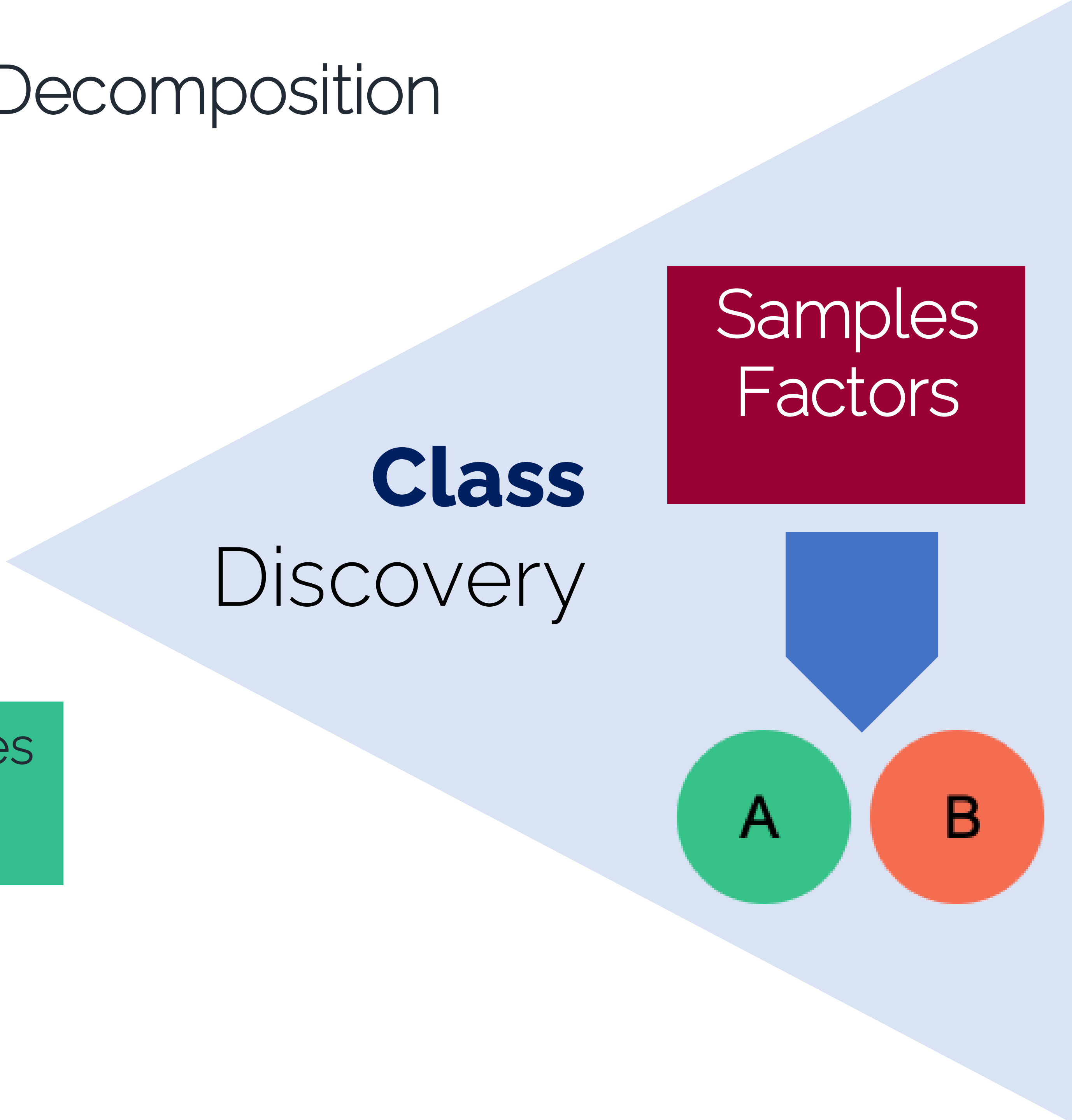


Matrix factorization

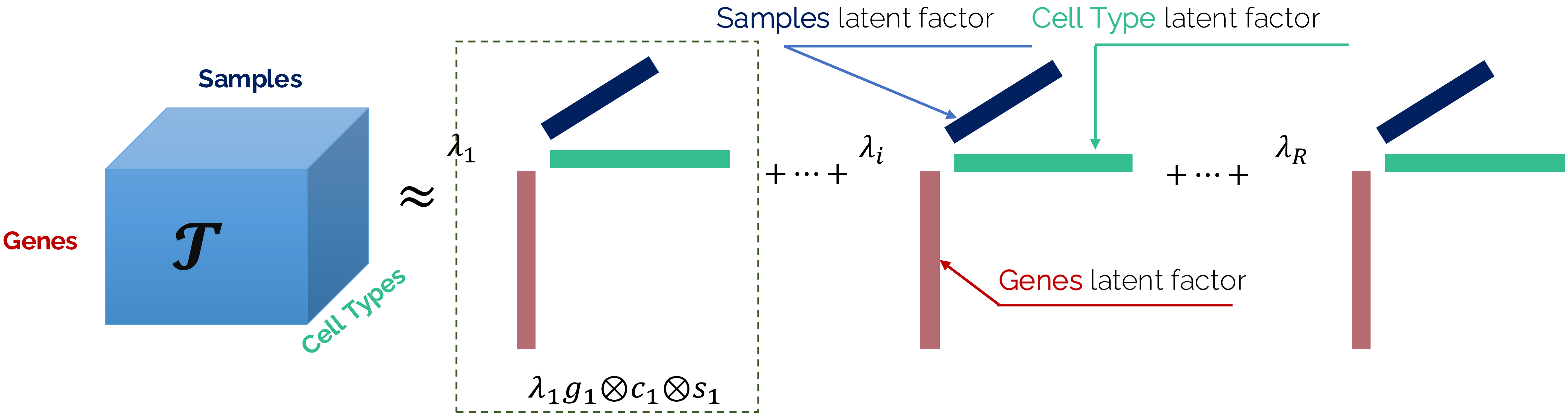


Candecomp/Parafac (CP) Decomposition





CP Decomposition



$\mathcal{T} = \mathcal{T}' + \varepsilon$ where $\mathcal{T}' = \sum_{r=1}^{R} \lambda_r \, g_r \otimes c_r \otimes s_r$

$\mathcal{T}' = [G, C, S]$ *where* $S = [s_1 \; s_2 \ldots s_R], C = [c_1 \; c_2 \ldots c_R], G = [g_1 \; g_2 \ldots g_R]$

we can concisely represent this factorization $\mathcal{T}' = [\boldsymbol{\Lambda}; \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{S}]$

noise assumption?
rank selection?
constraints?
uniqueness?

Candecomp/Parafac (CP) Decomposition

some of the different names

Hitchcock, 1927	Polyadic form of a tensor
Harshman, 1970	Parallel factor analysis (PARAFAC)
Carroll and Chang, 1970	Canonical Decomposition (CANDECOMP, CAND)
Möcks, 1988	Topographic components model
Kiers, 2000	CP(Candecomp/PARAFAC)

Other names?
- CPD
- rank approximation

source: Tamara G. Kolda† Brett W. Bader, Tensor Decompositions and Applications

CP Decomposition

For an N-way tensor $\mathcal{T}$ of size $I_1 \times I_2 \times \cdots \times I_N$, consider CP rank R approximation. We want to solve the following problem:

$$\min_{\mathcal{T}'} \|\mathcal{T} - \mathcal{T}'\| = \sqrt{\sum_{\mathrm{i,j,k}} \varepsilon_{\mathrm{ijk}}^2} \text{ where } \mathcal{T}' = [\lambda;\ A^{(1)}, A^{(2)}, \ldots, A^{(N)}], A^{(k)} \in \mathbb{R}^{I_k \times R}, 1 \leq k \leq R, \lambda = [\lambda_1, \lambda_2, \ldots, \lambda_R].$$

It is not a convex problem, but it can be given as N convex problems. For these, we need to consider the matricized version of the approximation

$$\mathcal{T}^{(k)} \approx A^{(k)} \Lambda \left(A^{(N)} \odot \cdots \odot A^{(k+1)} \odot A^{(k-1)} \odot \cdots \odot A^{(1)} \right)^{\top}, 1 \leq k \leq N, \Lambda = diag(\lambda)$$

In the next slide– we will check it for order-3 tensor

CP-decomposition– traditional approaches

A common method for CP decomposition and other tensor-related optimization problems is **alternating least squares**. We want to solve the following problem:

$$\min_{\mathcal{T}'} \|\mathcal{T} - \mathcal{T}'\| = \sqrt{\sum_{i,j,k} \varepsilon_{ijk}^2} \quad \text{where} \quad \mathcal{T}' = [\Lambda; A, B, C].$$

Fit (explained variance)=

$$1 - \frac{\|\mathcal{T} - \mathcal{T}'\|}{\|\mathcal{T}\|}$$

It is not a convex problem, but it can be given as 3 convex problems.

$$\min_{S} \left\|\mathcal{T}^{(1)} - A(C \odot B)^T\right\|$$
$$\min_{C} \left\|\mathcal{T}^{(2)} - B(C \odot A)^T\right\|$$
$$\min_{G} \left\|\mathcal{T}^{(3)} - C(B \odot A)^T\right\|$$

where $\mathcal{T}^{(i)}$ is the mode-1 matricization of the tensor $\mathcal{T}$, $\odot$ denotes the ``Khatri-Rao" product – matching column-wise Kronecker product $A \odot B = [a_1 \otimes b_1 \;\; a_2 \otimes b_2 \ldots a_R \otimes b_R]$

Other loss functions?

Kullback-Leibner divergence

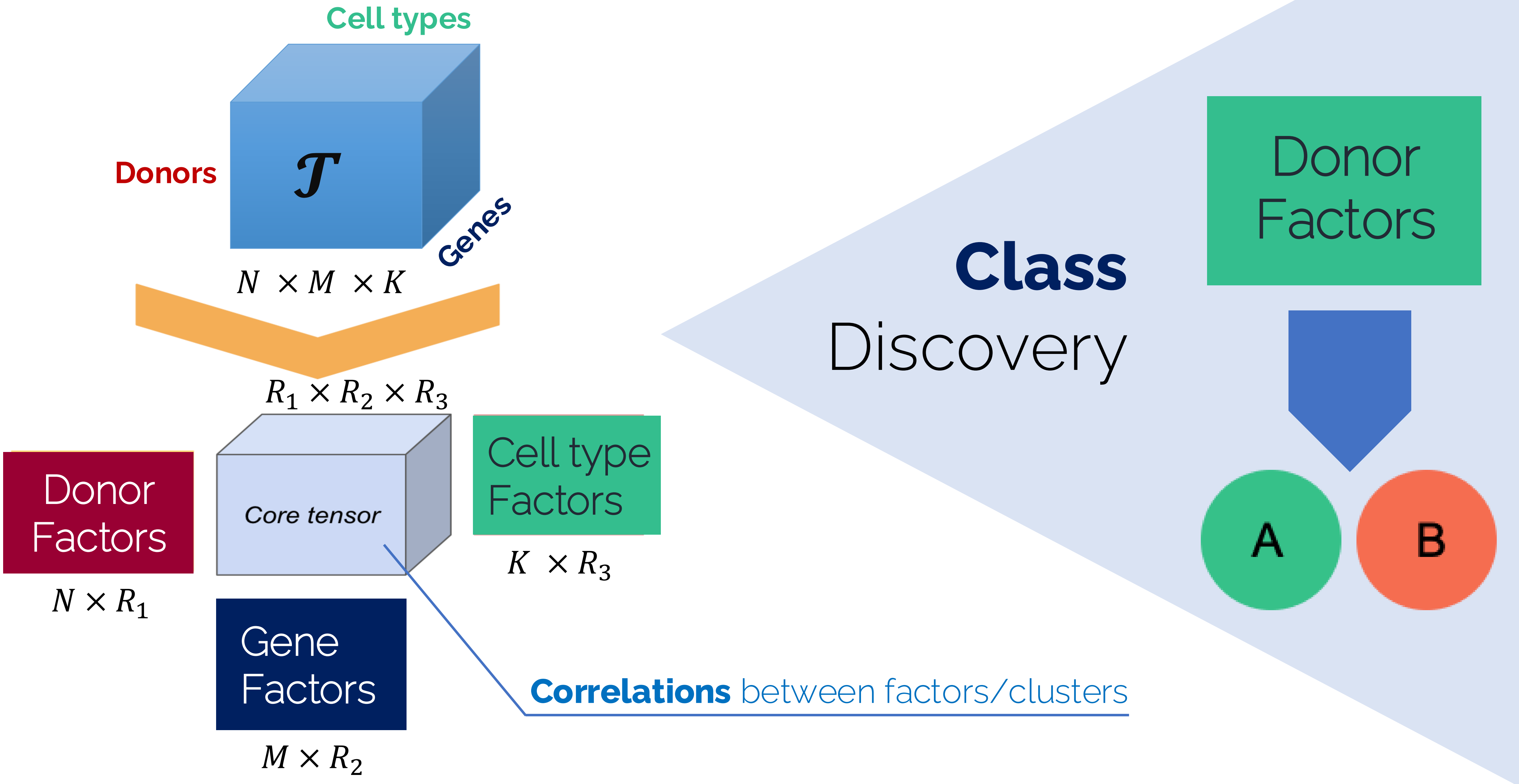
$$D_{KL}(\mathcal{T}||\mathcal{T}') = \sum_{z \in Z} \mathcal{T}(z) log \frac{\mathcal{T}(z)}{\mathcal{T}'(z)}$$

?

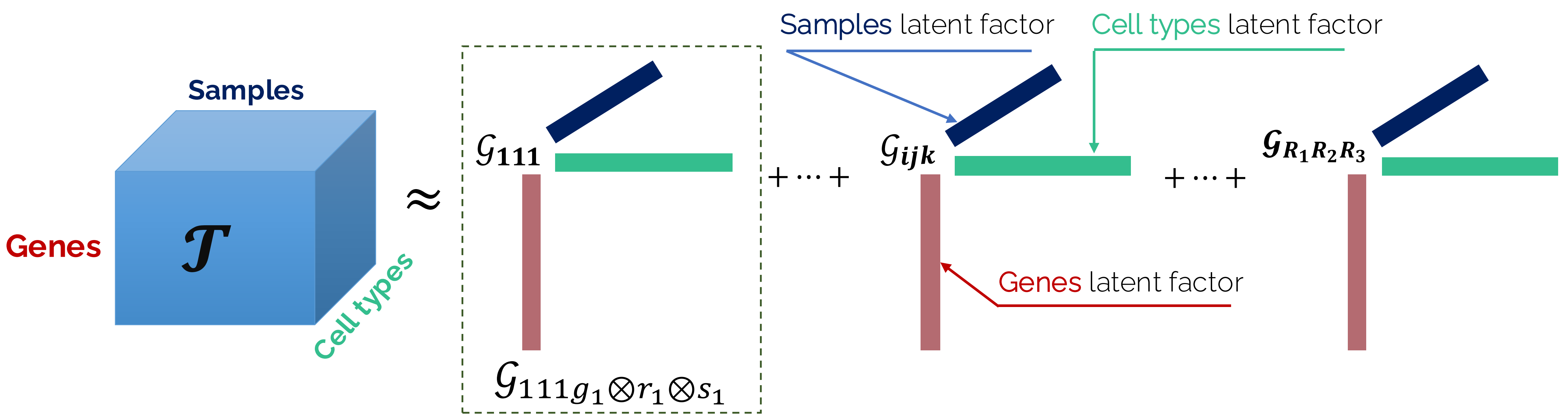
Choice of the loss function or distance metric?

Stability of the factorization?

Tucker Decomposition



Tucker Decomposition



Tensorized Data $\approx \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} \sum_{k=1}^{R_3} \mathcal{G}_{ijk} \; g_i \otimes c_j \otimes s_k = [\![\mathcal{G};$ **Genes Factors**, ***Cell types Factors***, ***Samples Factors*** $]\!]$

! A, B, C has orthogonal columns.

Tucker Decomposition

Let $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$, then Tucker decomposition can be given as

$$\mathcal{T} \approx \sum_{i_1=1}^{R_1} \sum_{i_2=1}^{R_2} \cdots \sum_{i_N=1}^{R_N} \mathcal{G}_{i_1 i_2 \ldots i_N} a_{i_1}^{(1)} \otimes a_{i_2}^{(2)} \otimes \ldots \otimes a_{i_N}^{(N)}$$

where $A^{(k)} = [a_1^{(k)} a_2^{(k)} \ldots a_{R_k}^{(k)}] \in \mathbb{R}^{I_k \times R_k}$, and the core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \ldots \times R_N}$.

We can concisely represent the approximation as $\mathcal{T} \approx \mathcal{G}_{\times 1} A^{(1)}_{\times 2} A^{(2)}_{\times 3} \ldots_{\times N} A^{(N)}$.

We can give the matricized version as the following

$$\mathcal{T}_{(k)} \approx A^{(k)} \mathcal{G}_{(k)} (A^{(N)} \otimes A^{(N-1)} \otimes \ldots \otimes A^{(k+1)} \otimes A^{(k-1)} \otimes \ldots \otimes A^{(1)})^\top.$$

Tucker Decomposition - storage complexity

$$\text{Storage for the core tensor} = R_1 \cdot R_2 \cdot \dots \cdot R_N.$$

$$\text{Storage for factor matrices} = \sum_{n=1}^{N} (I_n \cdot R_n).$$

If the ranks are uniform, i.e., $R_1 = R_2 = \dots = R_N = R$, then the storage complexity simplifies to:

$$\text{Total Storage Complexity} = R^N + \sum_{n=1}^{N} (I_n \cdot R).$$

- **Core tensor** storage scales **exponentially** with the tensor order N, as it involves R^N,
- **Factor matrix** storage scales **linearly** with N, as it involves $I_n R$ for each mode n.

Tucker Decomposition

some of the different names

Tucker, 1966	Three-mode factor analysis (3MFA/Tucker3)
Kroonenberg, De Leeuw, 1980	Three-mode PCA (3MPCA)
Kapteyn et al., 1986	N-mode PCA
De Lathauwer et al., 2000	Higher-order SVD (HOSVD)
Vasilescu and Terzopoulos, 2002	N-mode SVD

Other names?
MLSVD

source: Tamara G. Kolda† Brett W. Bader, Tensor Decompositions and Applications

Tensor Rank

The rank of a tensor $\mathcal{T}$, denoted **rank($\mathcal{T}$),** is defined as the smallest number of rank-one tensors needed to express $\mathcal{T}$ as their sum. In other words, this is the smallest number of components in an **exact** CP decomposition.

The concept of matrix rank and tensor rank are different.

1 ***Rank over different fields***

Consider N-way tensor $\mathcal{T} \in F^{I_1 \times I_2 \times \cdots \times I_N}$ for $F \subseteq \mathbb{C}$.

$\mathcal{T} = \sum_{r=1}^{R} \lambda_r\, g_r \otimes c_r \otimes s_r$, where $\lambda_r, g_r, s_r \in F$, the smallest such R is called the rank of $\mathcal{T}$ over F, $rank_F(\mathcal{T})$.
If $F = \mathbb{R}$, it is called **real rank** and if $F = \mathbb{C}$, it is called **complex rank**.

The rank of a matrix remains consistent across different fields, but this property does not extend to higher-order tensors.

Example: $\mathcal{T}_{::1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mathcal{T}_{::2} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ $\mathcal{T}$ is a $2 \times 2 \times 2$ tensor with **rank 3 over real numbers,** $\mathcal{T} = [A, B, C]$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

whereas it has **rank 2 over** $\mathbb{C}$ has the following factor matrices instead, $\mathcal{T} = [D, E, F]$

$$D = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}, \quad E = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}.$$

Tensor Rank

2 There is no specific algorithm to determine the rank of a specific tensor; the problem is NP-hard.

3 ***Typical and maximal ranks***

The **maximum rank** is the highest achievable rank, whereas the **typical rank** is any rank that occurs with positive probability, meaning it appears on a set with non-zero Lebesgue measure.

- For matrices of size $n \times m$, maximum rank=typical rank=$\min(n, m)$. For higher-order tensors, these two ranks can be different

maximum rank: $\mathcal{X} \in \mathbb{R}^{I \times J \times K} \implies \operatorname{rank}(\mathcal{X}) \leq \min\{IJ, IK, JK\}.$

for tensors of order $d \geq 3$,

- maximal rank and typical rank can be different
- there might be more than one typical rank over $\mathbb{R}$
- there is always one typical rank over $\mathbb{C}$, which is called **generic rank.**

Uniqueness

Let X be $n \times m$ matrix with the factorization

$$X = U V^{\top}, \mathrm{U} \in \mathbb{R}^{n \times r} \text{ and } \mathrm{V} \in \mathbb{R}^{m \times r}$$

Assume that $Q \in \mathbb{R}^{r \times r}$ is an orthogonal matrix, then

$$X = (UQ)\,(VQ)^{\top}$$

Thus, the presence of orthogonal transformations demonstrates that matrix factorizations are not unique.

- SVD is unique provided all the singular values are distinct
- For other factorizations, such as Non-Negative Matrix Factorization (NMF), strict conditions must be imposed to ensure uniqueness.

CP decomposition is unique under weaker conditions. Uniqueness means the factor matrices are uniquely determined, up to scaling and permutation.

uniqueness up to permutation

$\mathcal{T} = \sum_{i=1}^{R} a_r \otimes b_r \otimes c_r = [A, B, C]$= $[PA, PB, PC]$ for any permutation matrix P of size $R \times R$.

uniqueness up to scaling

$\mathcal{T} = \sum_{i=1}^{R} (\alpha_r a_r) \otimes (\beta_r b_r) \otimes (\gamma_r c_r) = [A, B, C]$ where $\alpha_r \beta_r \gamma_r = 1, 1 \leq r \leq R$.

CP uniqueness

Kruskal's result-sufficient condition

The k-rank of a matrix A, denoted k_A, is defined as the maximum value k such that any k columns are linearly independent.
The rank R CP decomposition of a 3 –way tensor $\mathcal{T} = [A, B, C]$ is unique if $k_A + k_B + \mathrm{k_C} \geq 2\mathrm{R} + 2$.

Sidiropoulos and Bro extended Kruskal's result to N-way tensors.

The rank R CP decomposition of a N–way tensor $\mathcal{T} = [A^{(1)}, A^{(2)}, \ldots, A^{(N)}]$ is unique if $\sum_{k=1}^{N} k_{A^{(k)}} \geq 2\mathrm{R} + (\mathrm{N} - 1)$.

necessary condition for uniqueness of rank R decomposition of a N-way tensor

$$\min_{1 \leq n \leq N} \; rank(A^{(1)} \odot \cdots \odot A^{(n-1)} \odot A^{(n+1)} \odot \cdots \odot A^{(N)}) = R$$

$$rank(A \odot B) \leq rank(A \otimes B) \leq rank(A) \cdot rank(B)$$

simplifies the necessary condition

$$\Rightarrow \min\left(\prod_{\substack{m=1 \\ m \neq n}}^{N} rank\left(A^{(m)}\right)\right) \geq R.$$

Border rank

border rank($\mathcal{X}$) = $\min\{r : for\ every\ \epsilon > 0\ there\ exist\ a\ tensor\ \bar{\boldsymbol{\mathcal{X}}}\ \ such\ that\ \|\ \boldsymbol{\mathcal{X}} - \bar{\boldsymbol{\mathcal{X}}}| < 0\ \}$

border rank($\mathcal{X}$)≤ rank($\mathcal{X}$)

- for matrices border rank = rank.
- it doesn't hold for tensors.

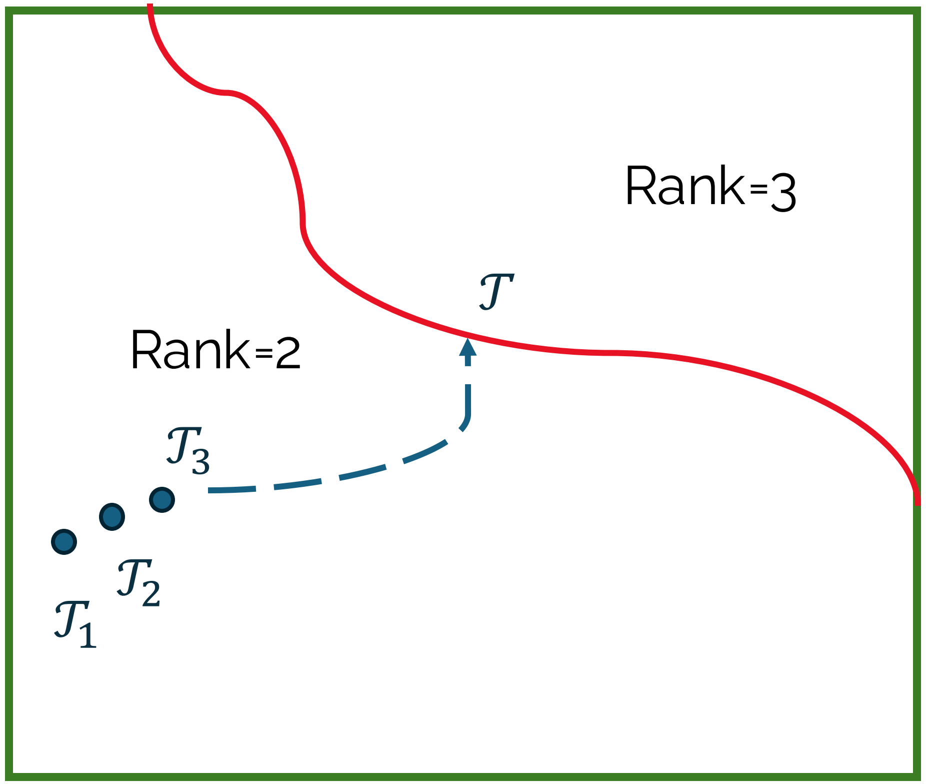
Set of tensors of rank at most r is not closed for $r \geq 2$.

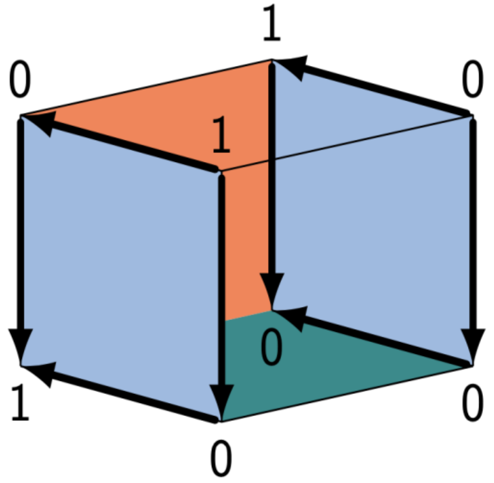
Strassen's algorithm for matrix multiplication relies on tensors with lower border rank to reduce computational complexity.

Matrix multiplication tensor: $\mathcal{T} : \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times p} \to \mathbb{R}^{m\times p}$ given by $\mathcal{T}(A, B) = C.$

$4 \times 4 \times 4$ multiplication tensor represents matrix multiplication of 2×2 matrices. It was shown that the rank and border rank of the tensor are both equal to 7

Border rank



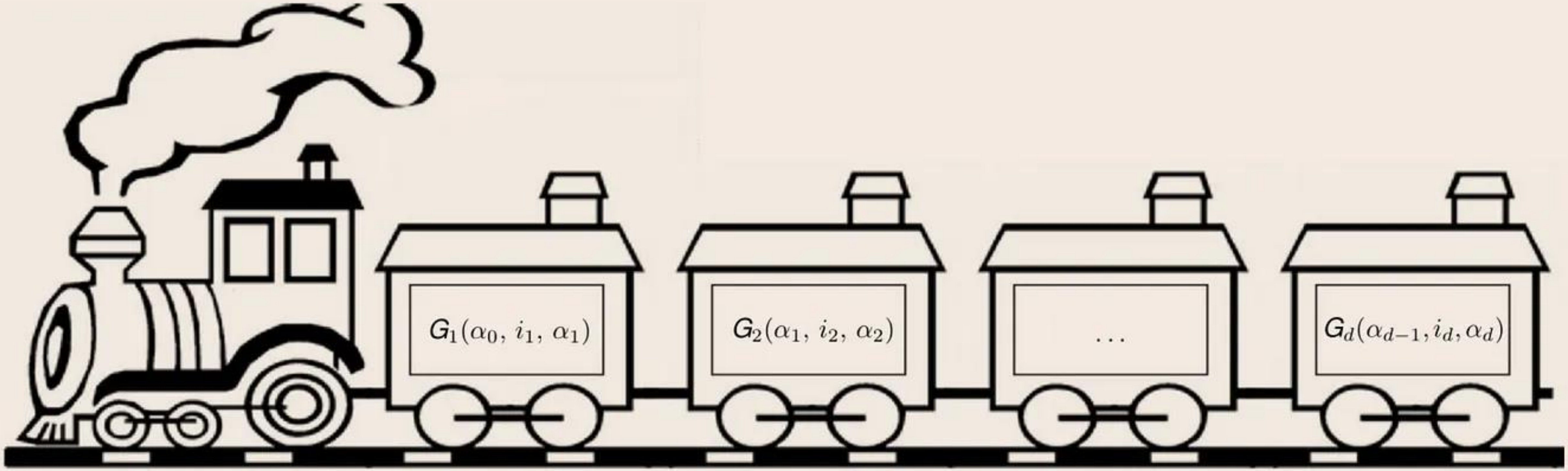


$$\mathcal{T} = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$$

$$\mathcal{T} = e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2$$

$\mathcal{T}$ has rank 3, but it can be approximated by rank 2 tensors

$$\lim_{n\to\infty} n \left(e_1 + \frac{1}{n} e_2 \right) \otimes \left(e_1 + \frac{1}{n} e_2 \right) \otimes \left(e_1 + \frac{1}{n} e_2 \right) - n e_1 \otimes e_1 \otimes e_1 = \mathcal{T}$$



Tensor Train Decomposition

Collage AIRI. Credit: iclcollective.com

Tensor Train Decomposition

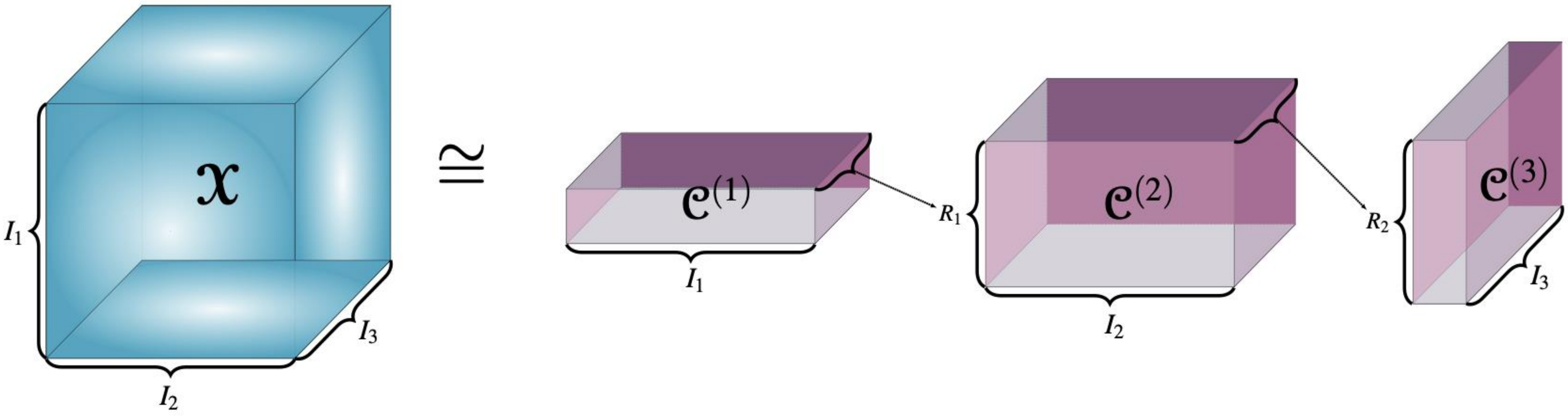


figure source: *Kour, Kirandeep, et al. "A weighted subspace exponential kernel for support tensor machines."*

Tensor Train Decomposition

Tensor Train factorizes a d-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ into a sequence of 3-way tensors:

$$\mathcal{X}_{i_1, i_2, \ldots, i_d} \approx \sum_{R_0, R_1, \ldots, R_d} \mathcal{G}^{(1)}_{R_0, i_1, R_1} \mathcal{G}^{(2)}_{R_1, i_2, R_2} \cdots \mathcal{G}^{(d)}_{R_{d-1}, i_d, R_d},$$

$(R_0 = R_d = 1)$

where each core $\mathcal{G}^{(k)} \in \mathbb{R}^{R_{k-1} \times I_k \times R_k}$ for $1 \leq k \leq d$.

The tuple of minimal integers $(R_0, R_1, \ldots, R_d)$ for which equality holds is the **TT rank** of the tensor.

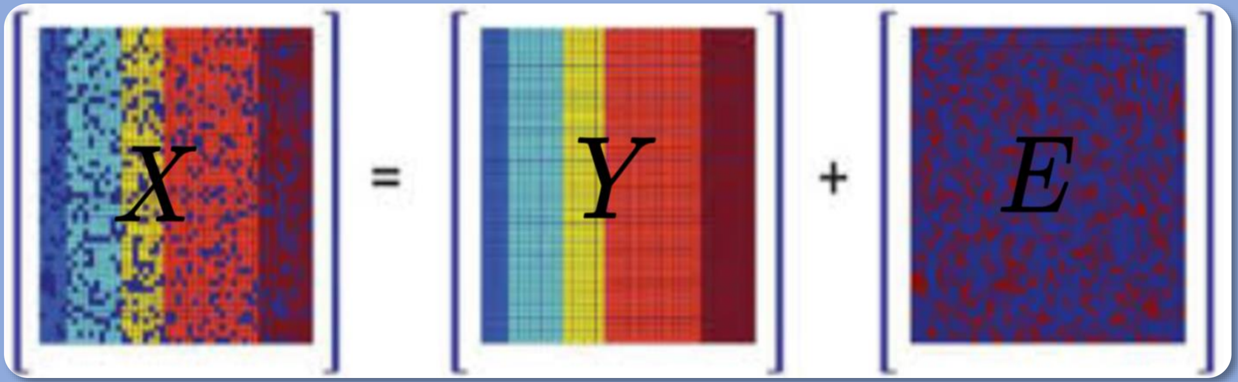
Storage complexity

storage requirement grows **linearly** with the number of modes, making it significantly more efficient than Tucker

$\sum_{i=1}^{d} R_{i-1} I_i R_i$ if we assume $R_i = R$ and $I_i = I$ $\implies dR^2 I$

Nuclear Norm for **Matrices**

Matrix **Recovery**



$$\min_{Y} rank(Y)$$

s.t. $X_{\Omega} = Y_{\Omega}$

$$\min_{Y} ||Y||_{*}$$

s.t. $X_{\Omega} = Y_{\Omega}$

Nuclear Norm ▶ **Convex surrogate** of rank

Missing **Data** Completion

Nuclear Norm (Trace norm) for **Matrices**

The **nuclear norm** (also called the **trace norm**) of A is defined as:

$$||A||_* = \sum_{i=1}^{r} \sigma_i,$$

where σ_i are the singular values of A.

Properties

1. **Convexity**: The nuclear norm is a convex function, making it useful in optimization problems.
2. **Dual Norm**: The nuclear norm is the dual of the spectral norm (the largest singular value).
3. **Low-Rank Promotion**: Minimizing the nuclear norm encourages solutions with lower rank, as the nuclear norm serves as a convex relaxation of the rank function.

Main applications: collaborative filtering, low rank approximation, compressed sensing

Spectral Norm for Matrices – dual of the nuclear norm

The **spectral norm** is the largest singular value of the matrix A, i.e.,

$$\|A\|_2 = \sigma_{\max}(A),$$

where $\sigma_{\max}(A)$ is the largest singular value of A.

For the nuclear norm: The nuclear norm is the dual of the spectral norm because the following holds for all matrices A and B:

$$\langle A, B \rangle = \mathrm{Tr}(A^T B),$$

where $\langle A, B \rangle$ is the Frobenius inner product, and:

$$\|A\|_* = \sup_B \langle A, B \rangle \quad \text{subject to} \quad \|B\|_2 \leq 1.$$

In other words, the nuclear norm of a matrix is the maximum of the Frobenius inner product over all matrices B whose spectral norm is less than or equal to 1.

For the spectral norm: Conversely, the spectral norm is the dual of the nuclear norm because:

$$\|A\|_2 = \sup_B \langle A, B \rangle \quad \text{subject to} \quad \|B\|_* \leq 1.$$

Spectral Norm for Matrices – dual of the nuclear norm

The **spectral norm** measures the maximum stretching factor of the matrix, which corresponds to the largest singular value. Applications include measuring sensitivity of linear systems, low rank approximation, data compression

Properties

1. **Sub-multiplicativity**:

   The spectral norm satisfies $\|AB\|_2 \leq \|A\|_2\|B\|_2$, meaning the norm of the product is at most the product of the norms.

2. **Dual Norm**:

   The spectral norm is the dual of the nuclear norm, which is the sum of the singular values.

3. **Operator Norm**:

   The spectral norm represents the largest stretching factor of a matrix, given by $\max \|Ax\|_2/\|x\|_2$.

4. **Computational Complexity**:

   The spectral norm, the largest singular value, is computed using SVD or methods like the power method.

Nuclear Norm for **Tensors**

Generalizes matrix nuclear norm to higher dimensions.

$\mathcal{X}$ is a 3-way tensor.
Then the ***nuclear norm*** of $\mathcal{X}$ is given by:

$$\|\mathcal{X}\|_* = min\{\sum_{i=1}^{r} |\lambda_i| : \mathcal{X} = \sum_{i=1}^{r} \lambda_i \, u_i \otimes v_i \otimes w_i, \|u_i \otimes v_i \otimes w_i\| = 1\}$$

nuclear decomposition

promotes sparsity in the tensor's decomposition, encouraging simpler, low-rank representations of multi-dimensional data.

Spectral Norm for **Tensors**

Dual to tensor nuclear norm

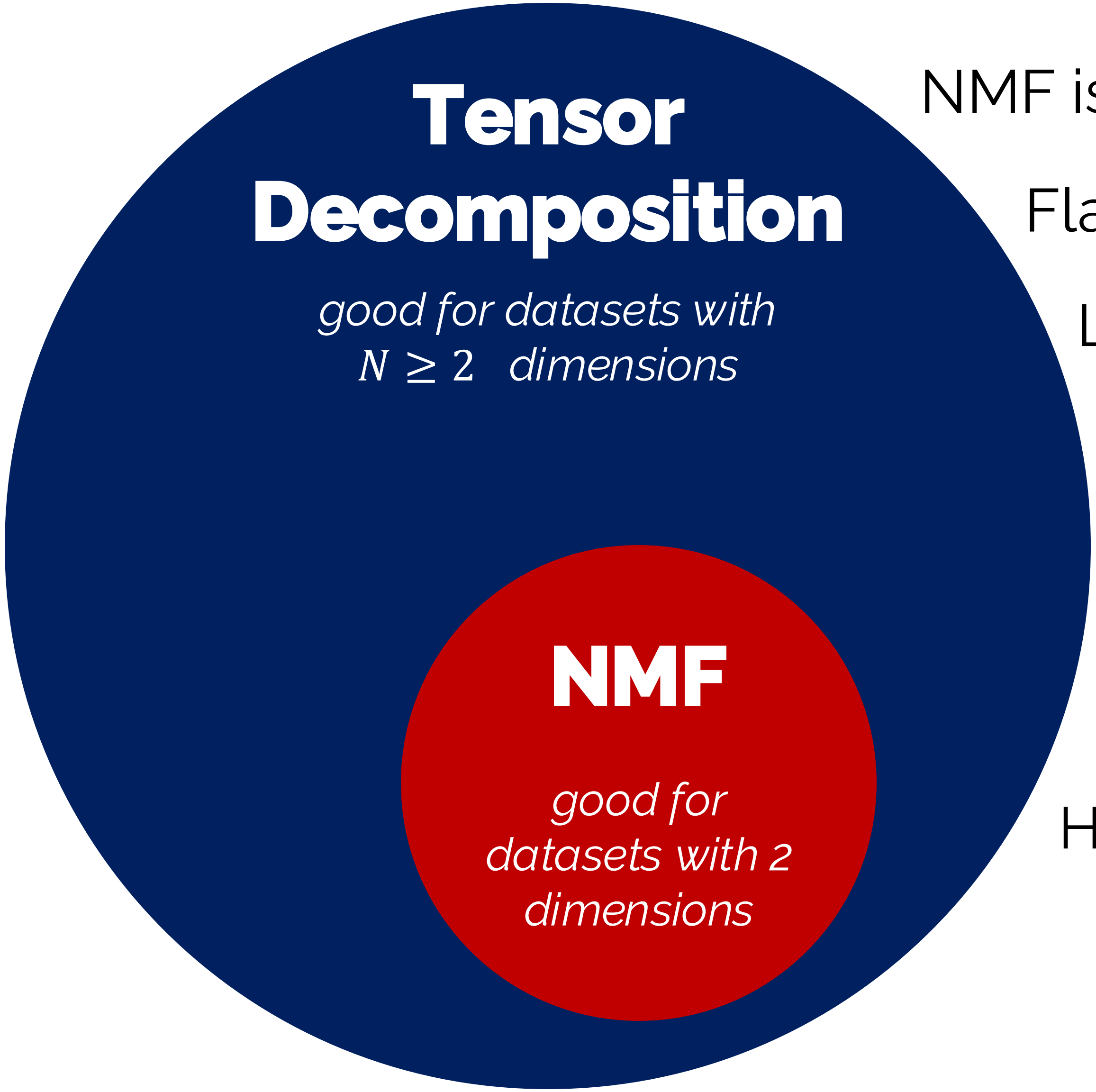
$\mathcal{X}$ is a 3-way tensor.
Then the ***spectral norm*** of $\mathcal{X}$ is given by:

$$\|\mathcal{X}\|_\sigma = sup\{\langle \mathcal{X}, u_i \otimes v_i \otimes w_i\rangle, \|u_i \otimes v_i \otimes w_i\| = 1\}$$

Spectral Norm ▶ Rank 1 **approximations**

If $\mathcal{Y}$ is a best rank-1 approximation of the tensor $\mathcal{X}$, then $\|\mathcal{X} - \mathcal{Y}\|_F = \sqrt{\|\mathcal{X}\|_F - \|\mathcal{X}\|_\sigma}$

Comparison #1: NMF, PCA



NMF is designed for 2-dimensional data

Flattens the data if dimensions > 2

Loss of interactions btw. different modalities

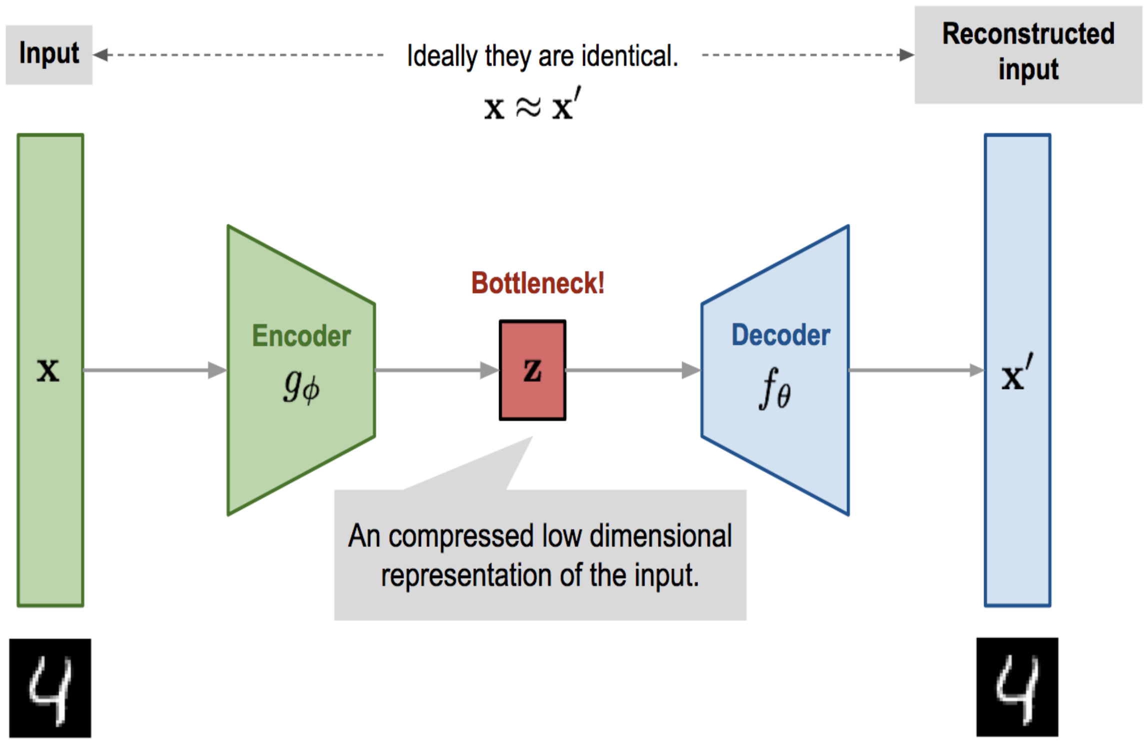
NMF is not unique

Assumes normally-distributed data

Rank selection is non-trivial

Higher-dimensions = challenging interpretation

Comparison #2: Deep learning approaches



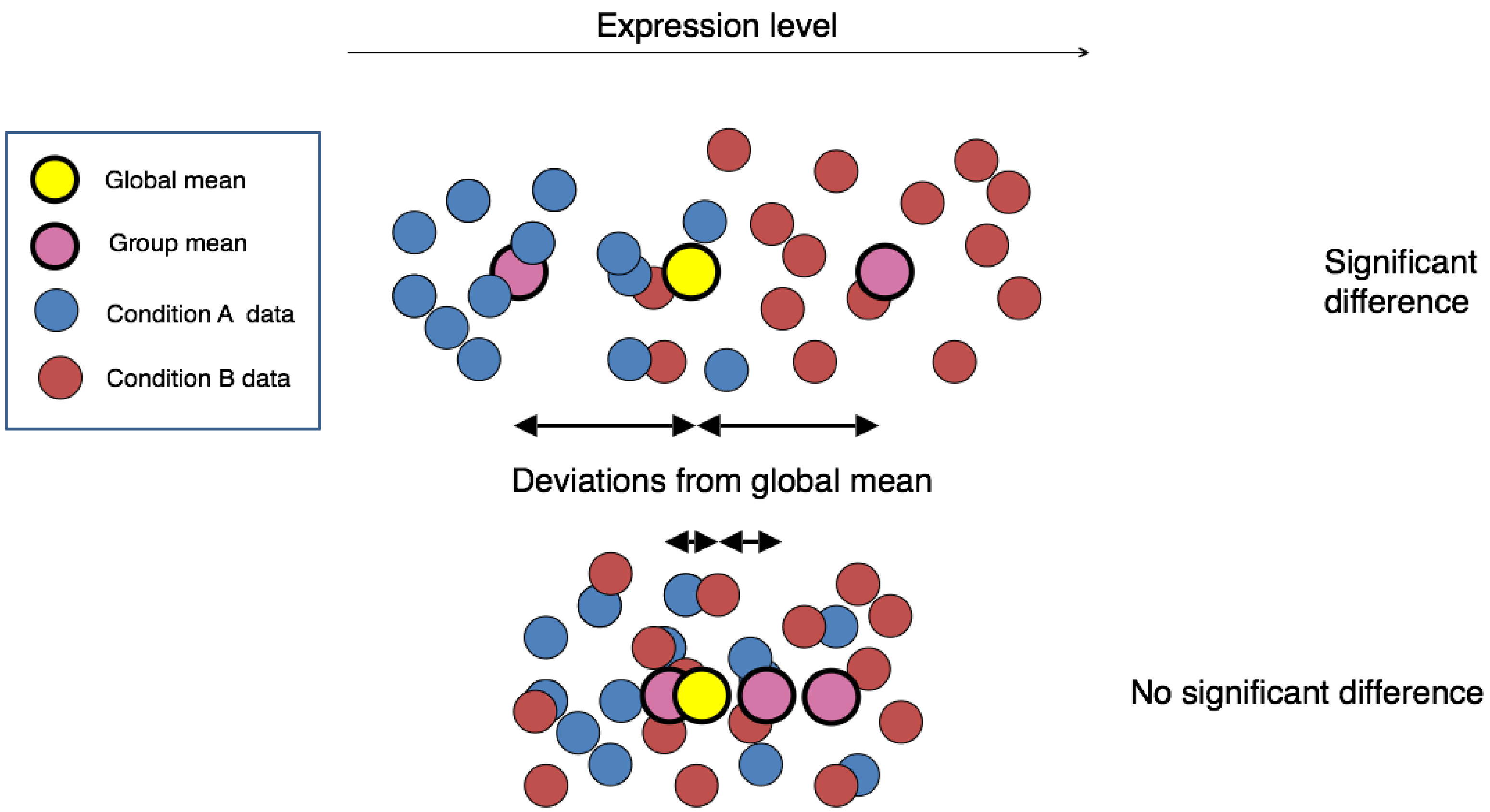
Latent space interpretation is **challenging**

1000's of training data

Train NN

Inference of patterns on a new dataset is as good as training data set

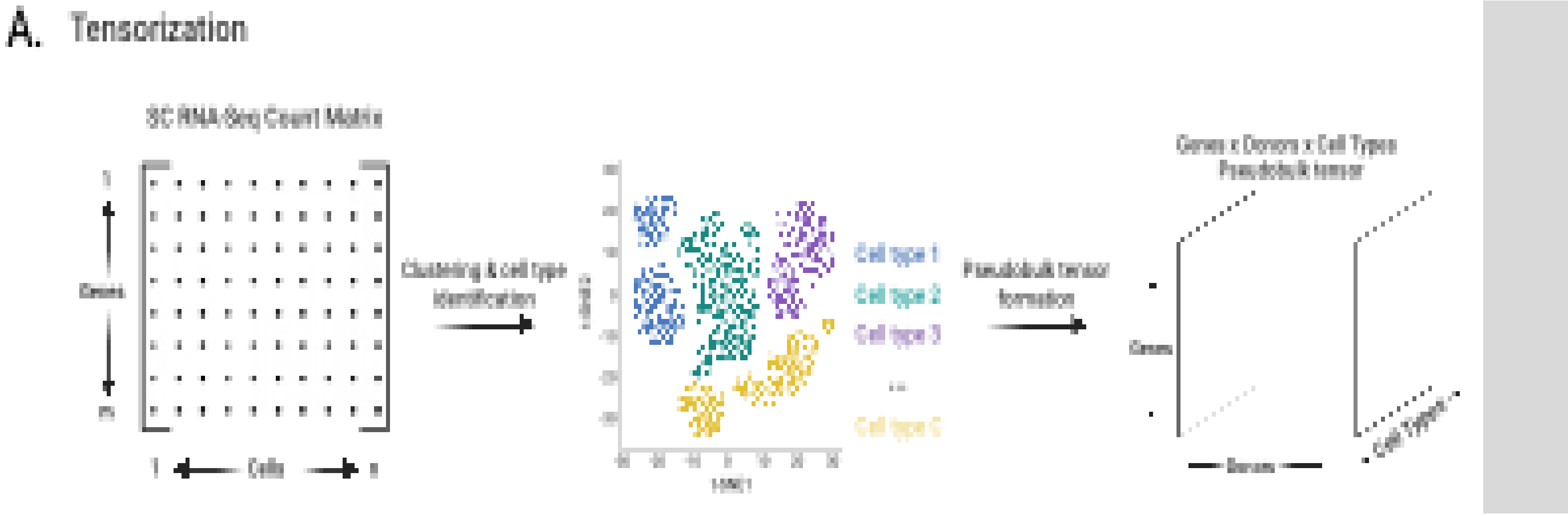
Comparison #3: Supervised DE methods



Need to give **predefined conditions**

Does not consider sample **heterogeneity**

Multi-sample Gene expression data



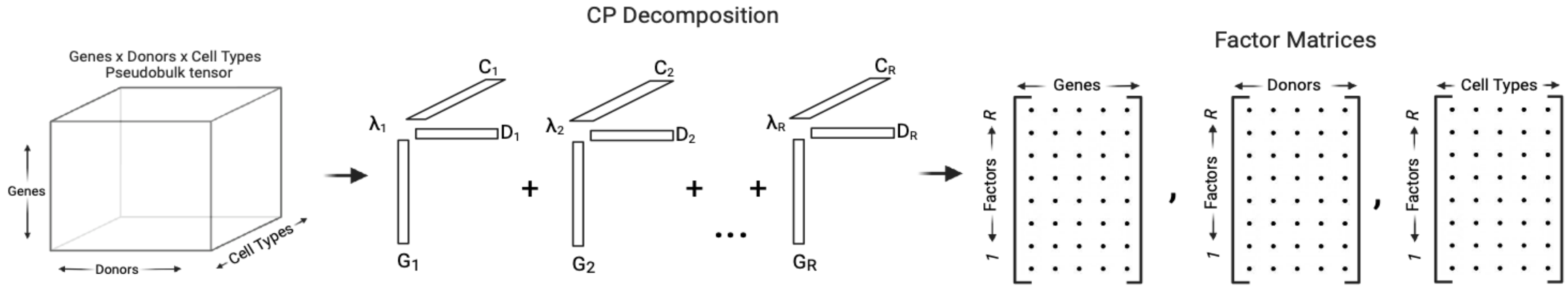
Genotype-Tissue Expression (GTEx) Portal
The TwinsUK cohort
The Illumina Body Map – 16 different human tissues
The Cancer Genome Atlas (TCGA)

Traditional approaches often assume that gene expression patterns remain consistent across different contexts or that samples are **independent** and **homogeneous**.

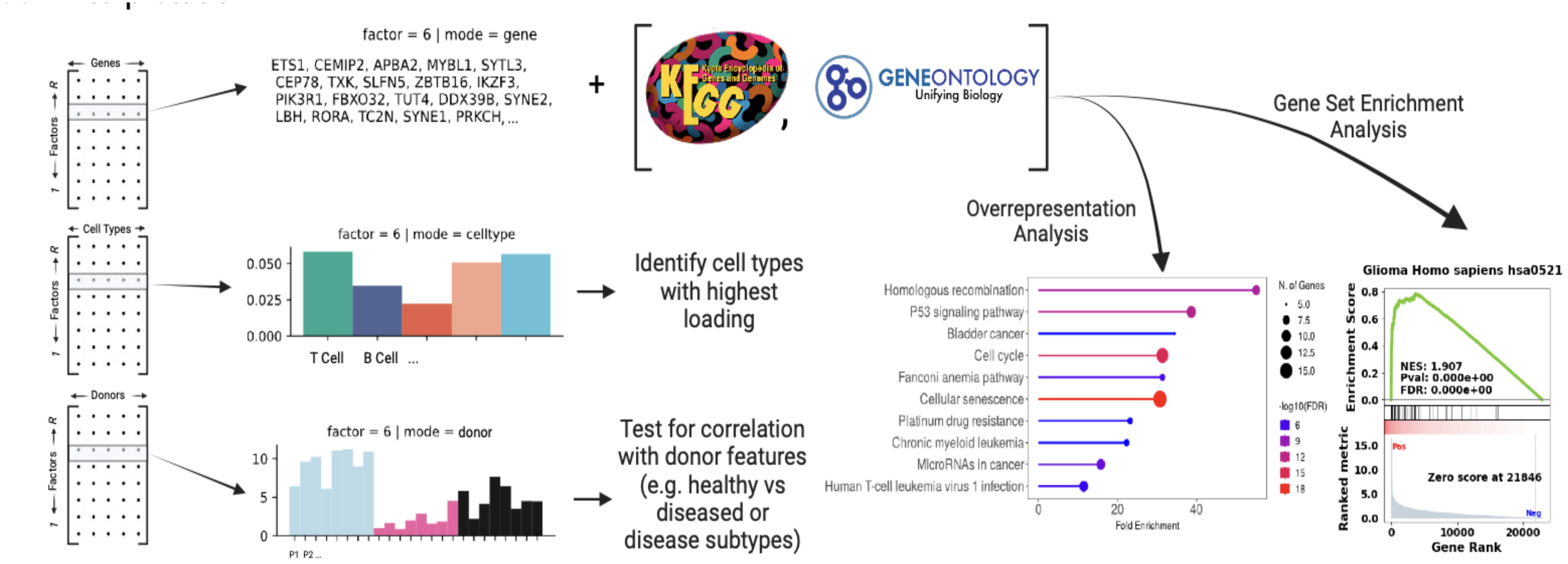
Structuring the high-dimensional genomics data as matrices poses **several challenges:**

- It may hinder the discernment of **cell-type specific**, tissue-specific, or individual-specific patterns.
- Inferring gene modules independently for each context might overlook shared characteristics among cell types or tissues and impede the **identification of differentially expressed genes**
- Neglecting **individual heterogeneity**, including biological factors like race, gender, and age, can compromise the accuracy of estimating correlations between genes and tissues.

Factorization



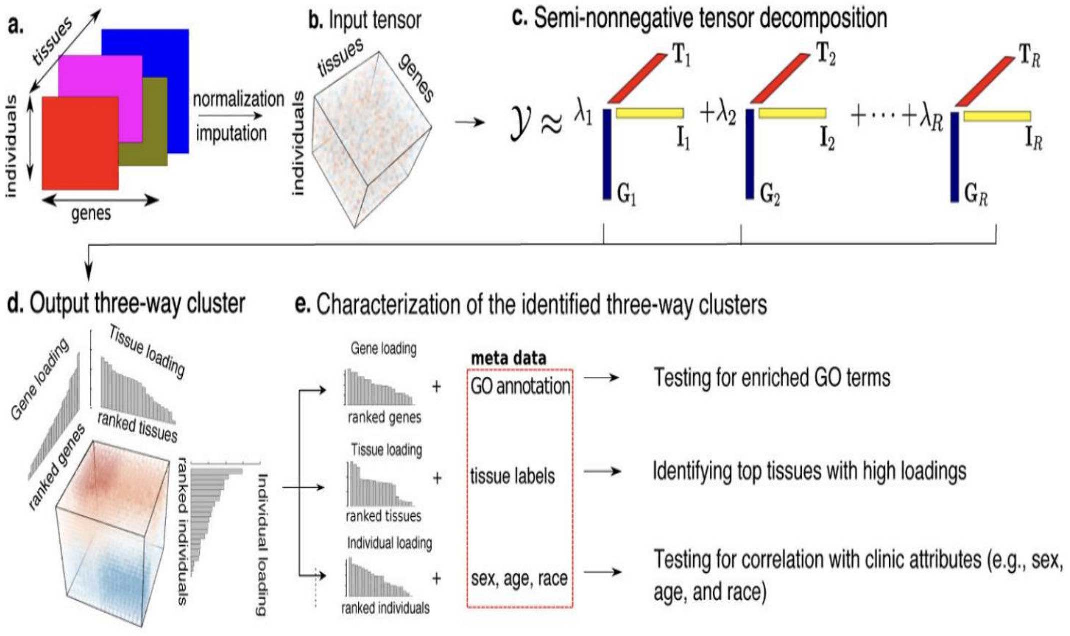
Interpretation

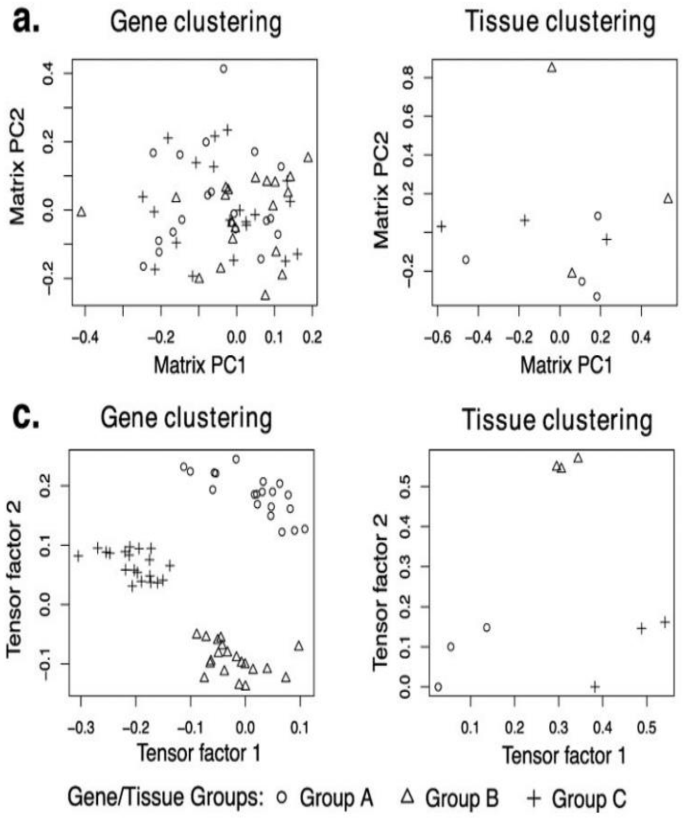


Case study: MultiCluster

data: GTEx v6 gene expression data, consisting of RNA-seq samples collected from 544 individuals across 53 human tissues, including 13 brain subregions, adipose tissue, heart, artery, skin, and more

$$18,481 \times 544 \times 53$$





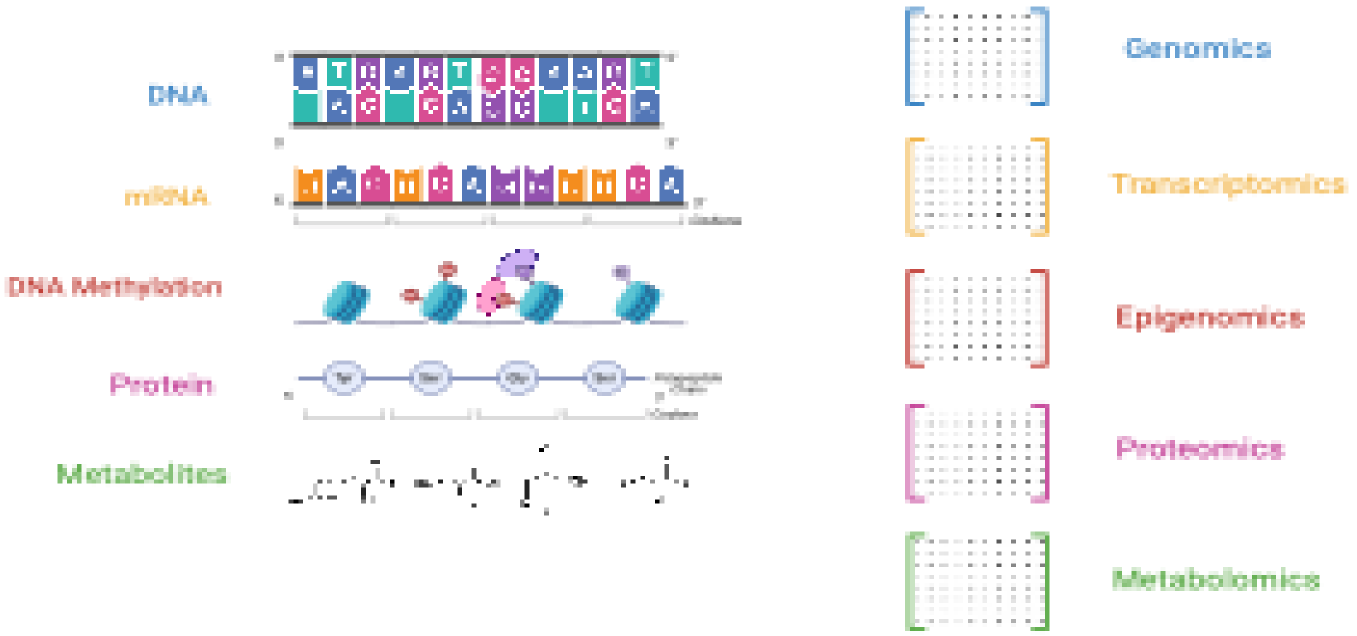
Clustering

Classification

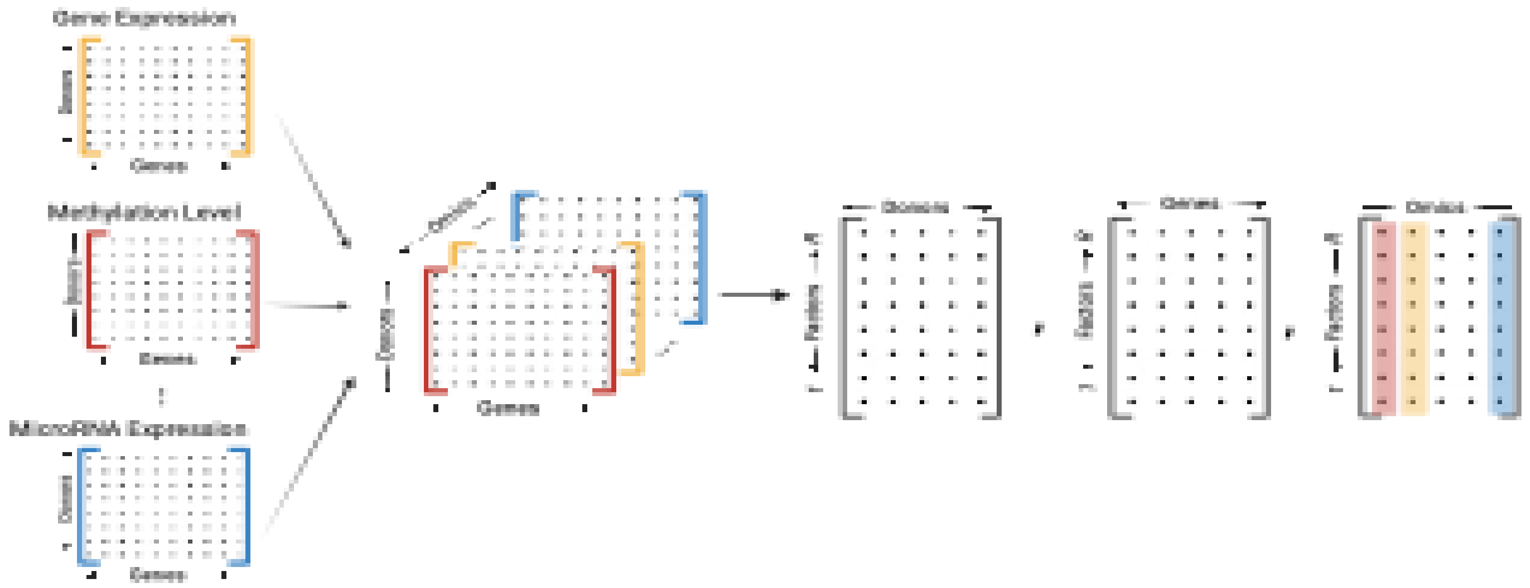
Co-variate effects

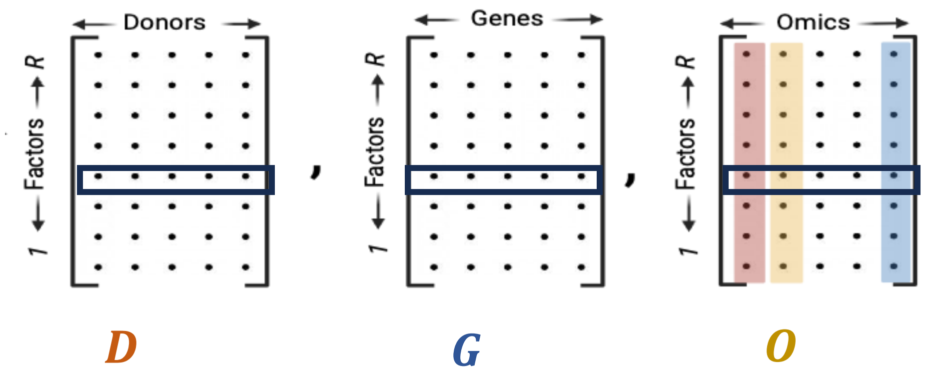
Wang, Miaoyan, Jonathan Fischer, and Yun S. Song. "Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition." *The annals of applied statistics* 13.2 (2019): 1103.

Multi-omics data



- TCGA, GTEx, ENCODE
- Human Functional Genomics Project (HFGP)





$D^T = [D^{(1)}\ D^{(2)} \ldots D^{(R)}]$ donors latent factor matrix

$G^T = [G^{(1)}\ G^{(2)} \ldots G^{(R)}]$ gene latent factor matrix

$O^T = [O^{(1)}\ O^{(2)} \ldots O^{(R)}]$ omics latent factor matrix

$$\mathcal{X}_{\text{omics}} \approx \sum_{r=1}^{R} \lambda_r G^{(r)} \otimes D^{(r)} \otimes O^{(r)},$$

r-th component expression pattern

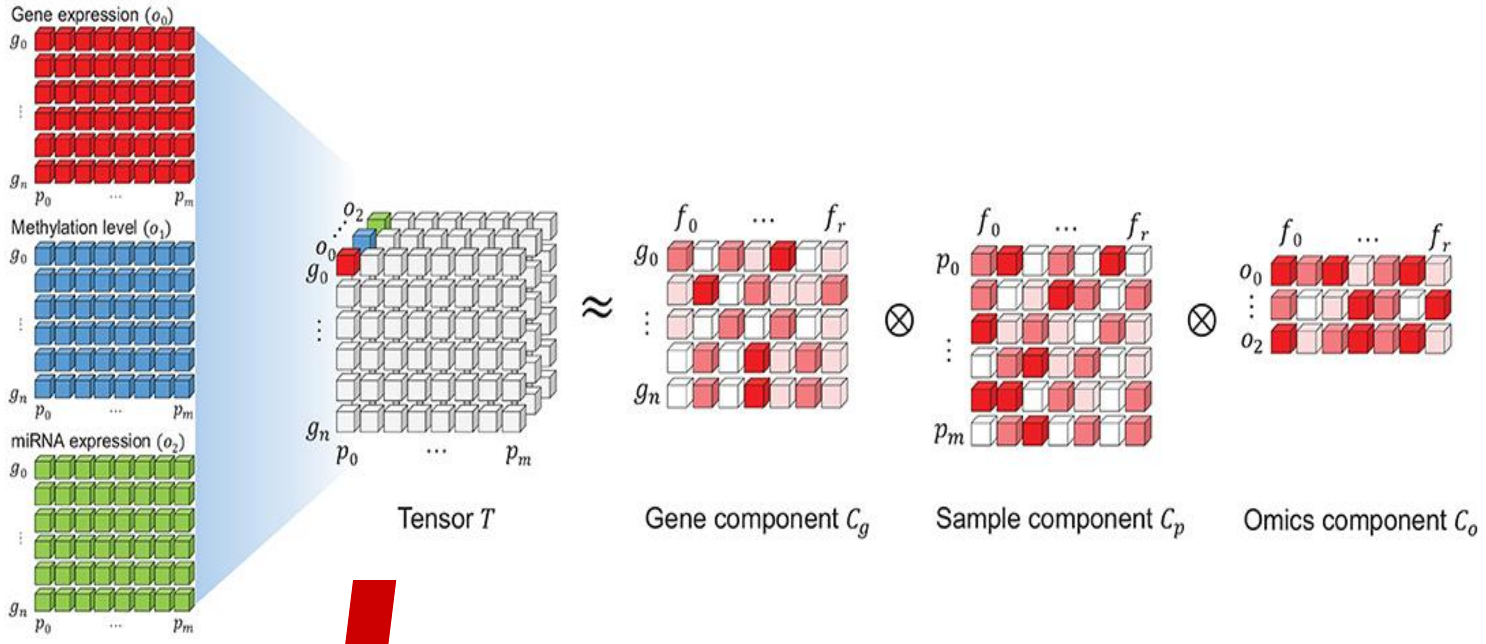
weights

$$(G^{(r)} \otimes D^{(r)} \otimes O^{(r)})_{i,j,k} = G_i^{(r)} D_j^{(r)} O_k^{(r)}$$

Captures the interaction between i-th gene, j-th donor and k-th omics platform

Case study: Monti

Applied to three case studies of 597 breast cancer, 314 colon cancer, and 305 stomach cancer cohorts.
Goal: subtype classification such as for breast cancer Luminal A, Luminal B, Her2, and Basal.



MONTI was able to detect subtype specific (or clinical attribute specific)gene sets that showed to be strongly regulated by certain omics, from which correlation between omics types could be inferred.

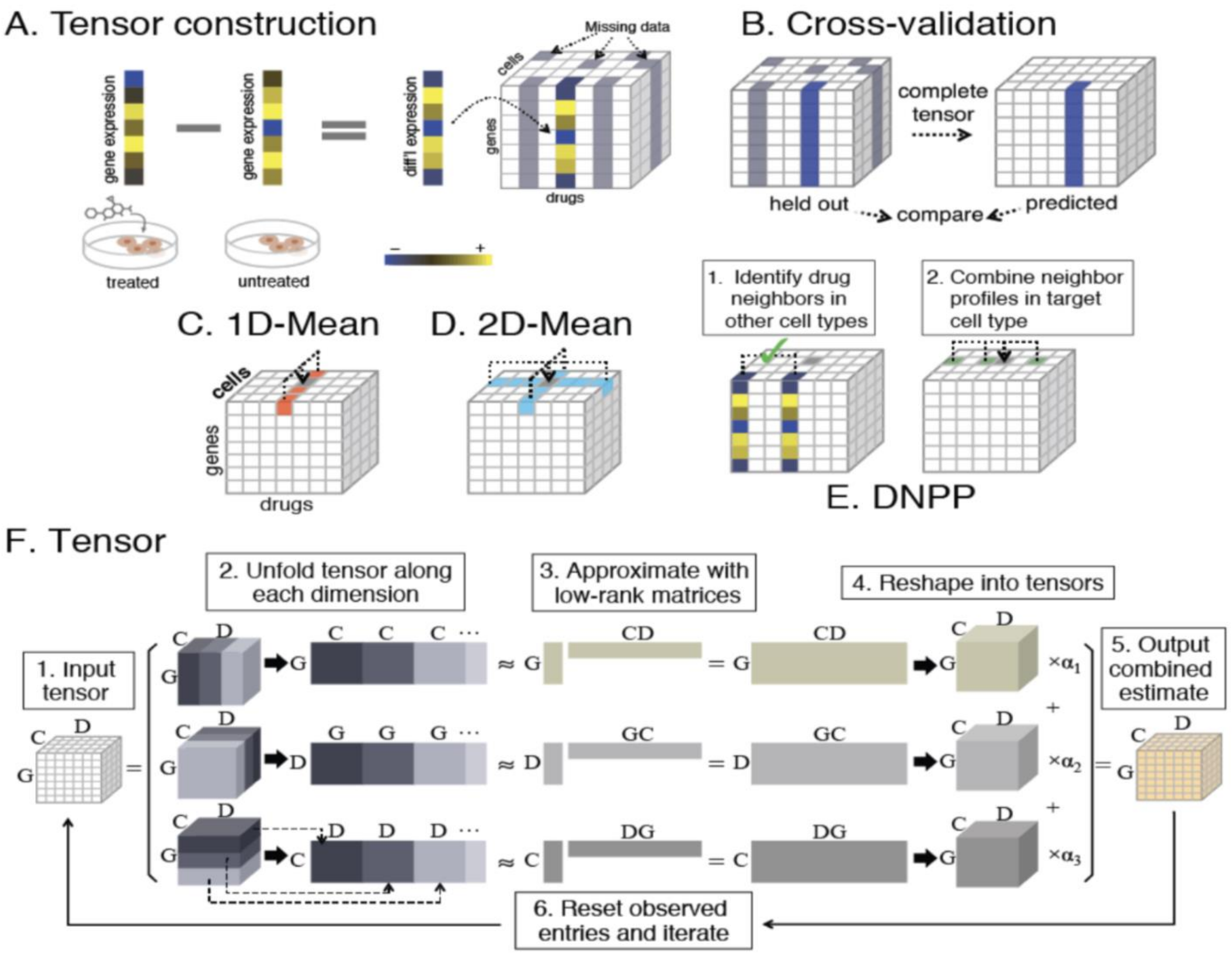
integrating multi-omics data in a gene centric manner improves detecting cancer subtype specific features and other clinical features

Jung, Inuk, et al. "MONTI: a multi-omics non-negative tensor decomposition framework for gene-level integrative analysis." *Frontiers in genetics* 12 (2021): 682841.

Data Imputation

Challenge

capture **multilinear** relationships between **drugs**, **targets**, and **diseases**



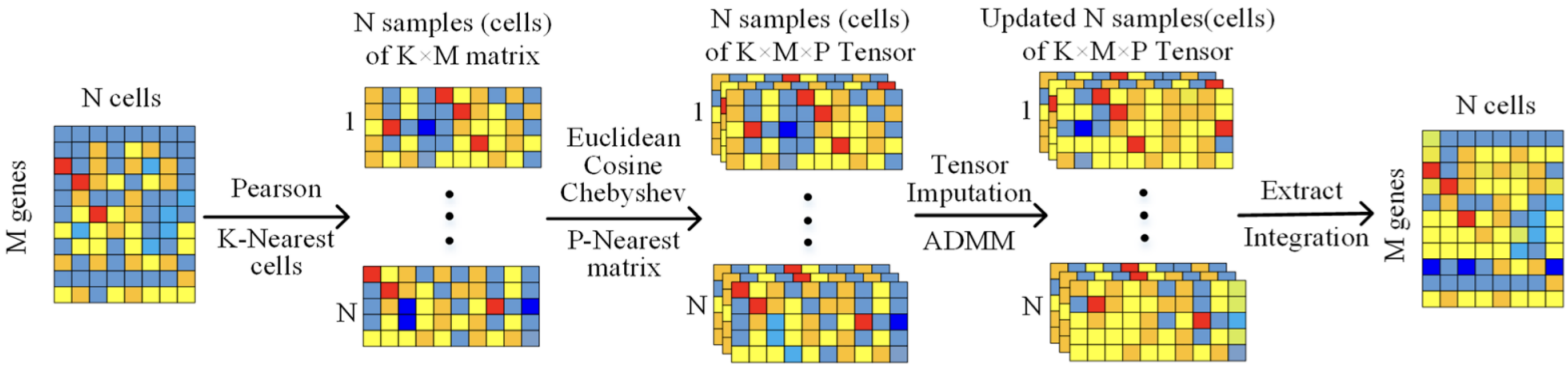
Hodos et.al, Cell-specific prediction and application of drug-induced gene expression profiles, 2018

Case study: ScLRTC

An important challenge in analyzing genomics data is the high prevalence of zero values, largely due to the **“drop-out”** effect.

Data sets: published scRNA-seq datasets, including Usoskin, Pollen, Yan, Zeisel, Mouse and PBMC

Compared methods: DrImpute, SAVER, scImpute, MAGIC, CMF-Impute and PBLR



Pan, Xiutao, et al. "ScLRTC: imputation for single-cell RNA-seq data via low-rank tensor completion." *BMC genomics* 22 (2021): 1-19.

Communication is the key!

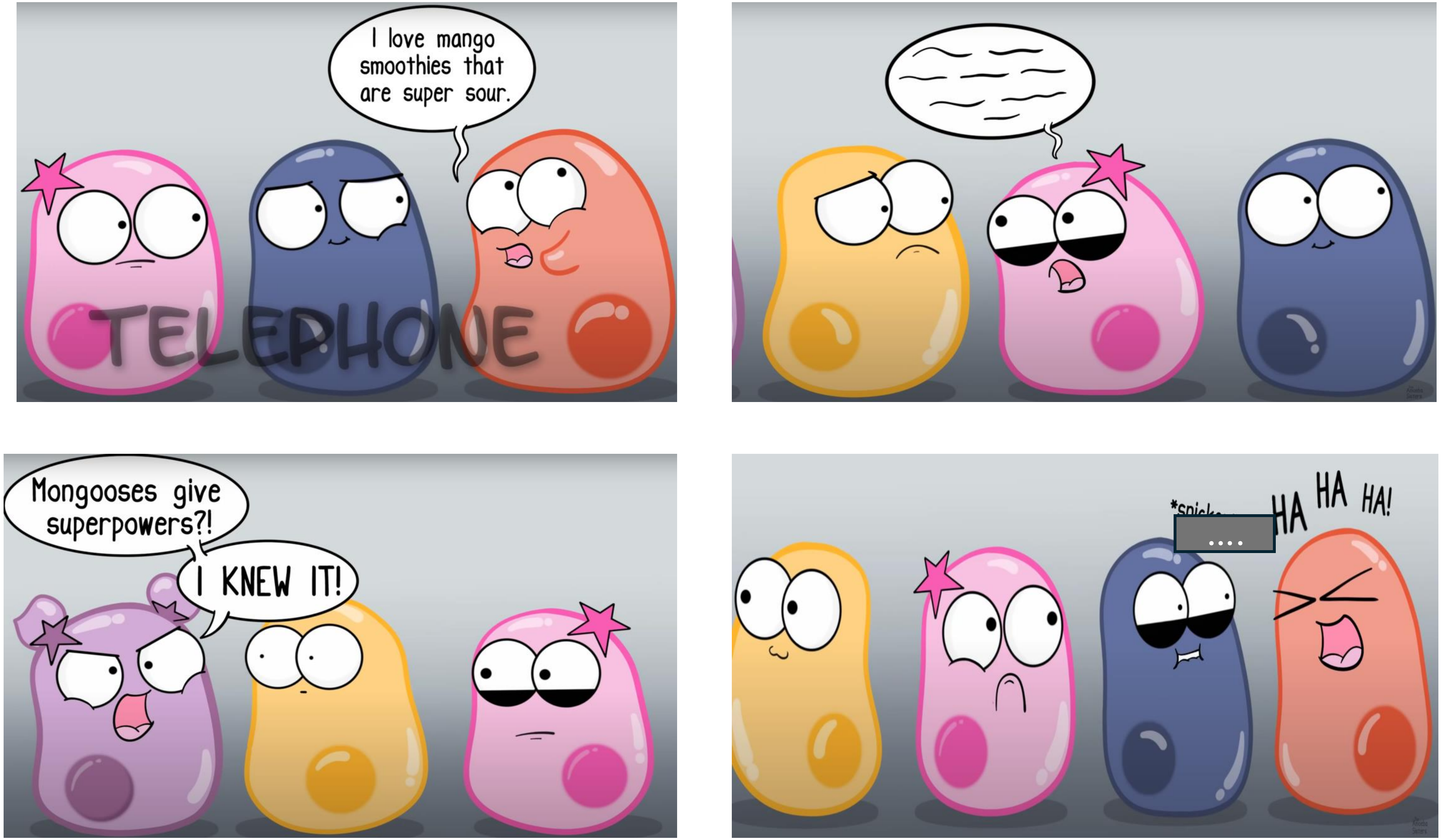
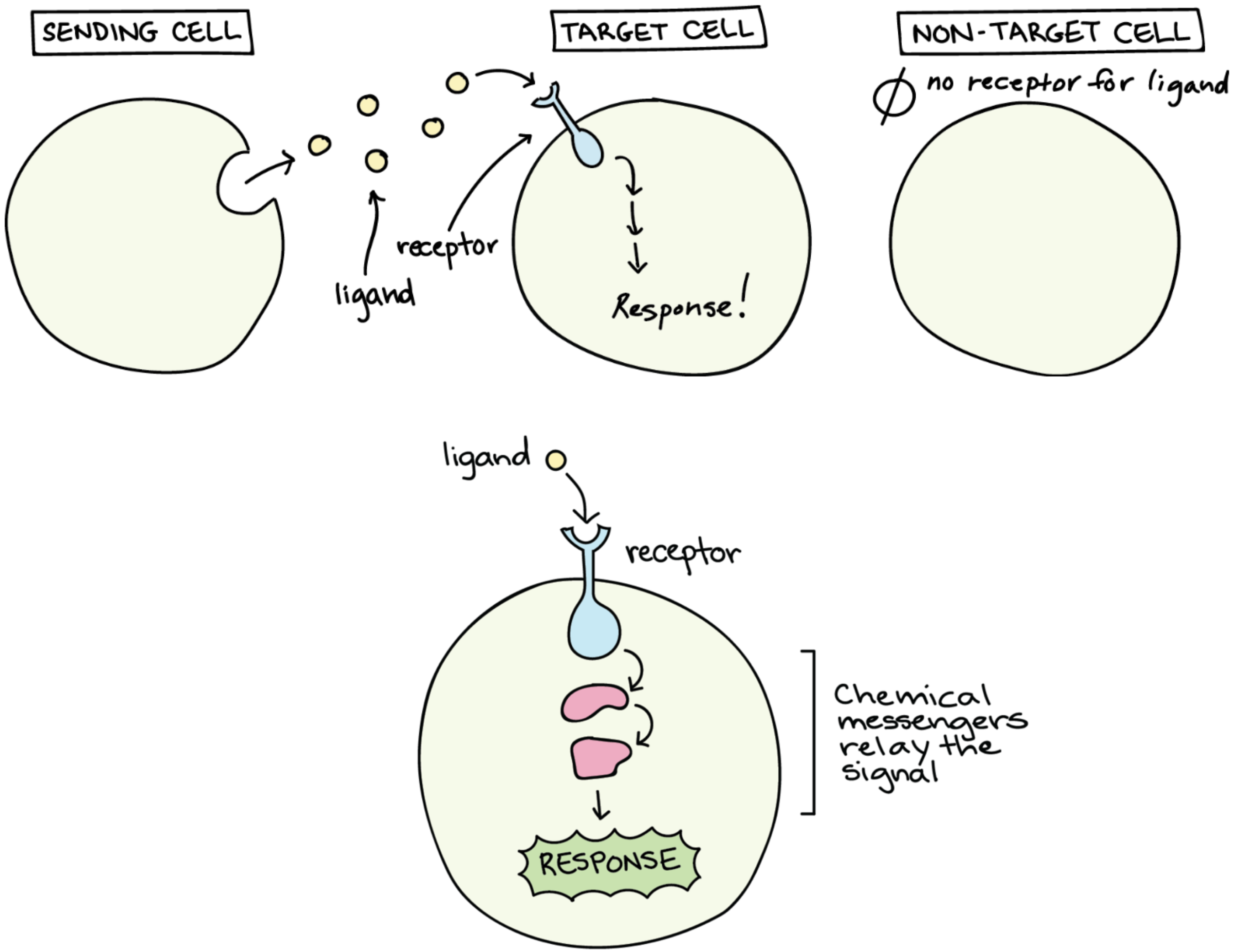


Figure credit: https://www.amoebasisters.com/

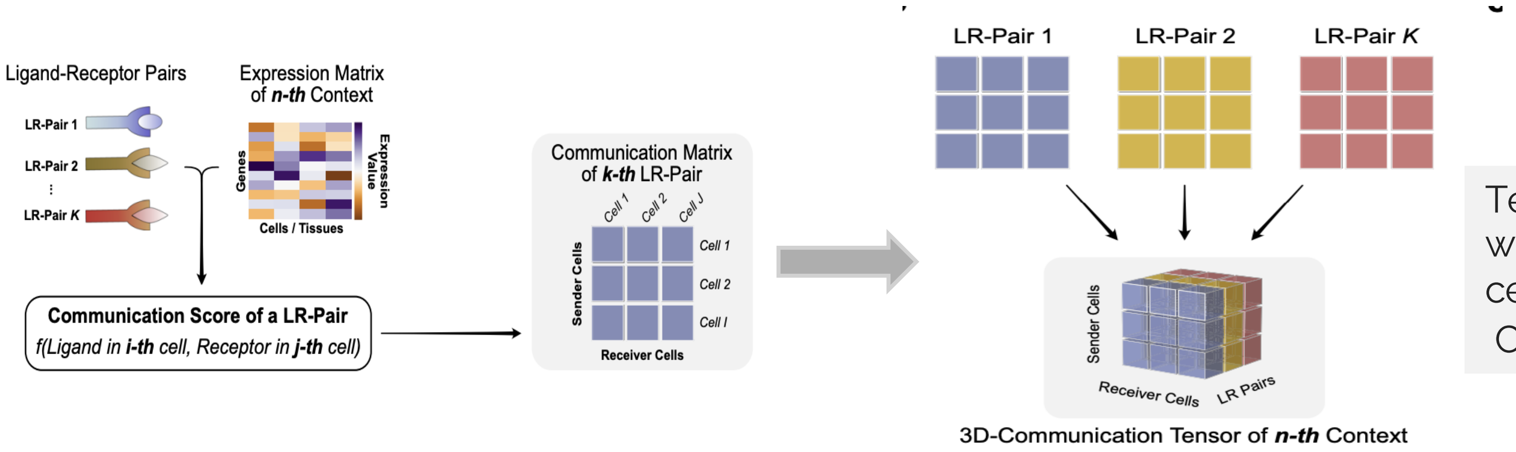
1- Ligand-receptor (LR) interactions



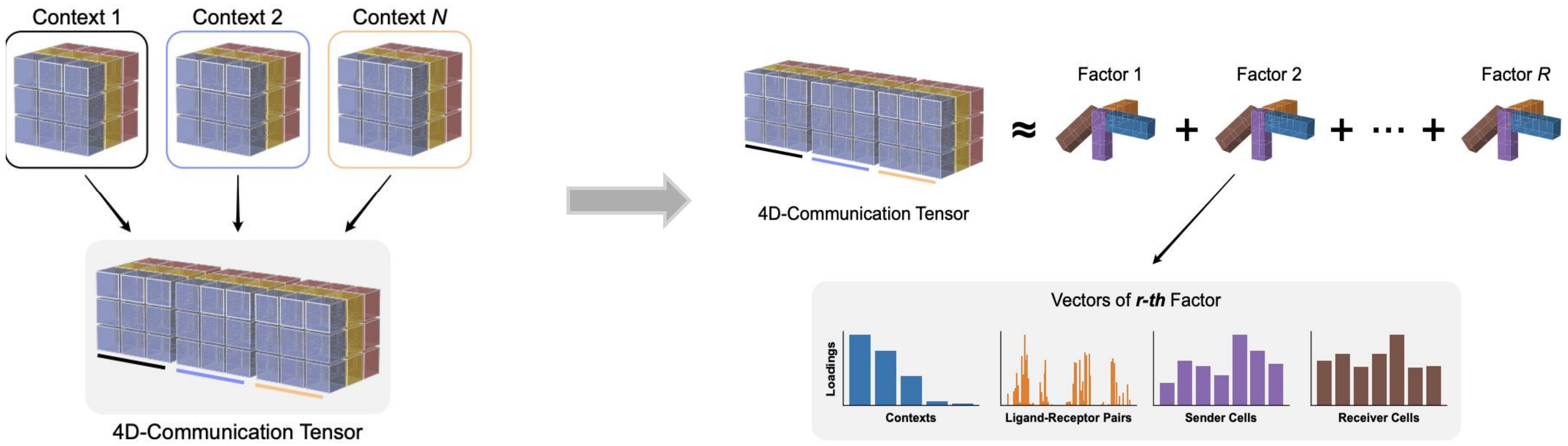
- sender cell types
- receiver cell types
- LR-pairs
- Multiple contexts

Figure credit: https://www.khanacademy.org/

Case study: Tensor-cell2cell



Tensor-cell2cell can identify multiple modules associated with distinct communication processes (e.g., participating cell–cell and ligand-receptor pairs) linked to severities of Coronavirus Disease 2019 and to Autism Spectrum Disorder.



?

- Choice of the rank
- Stability of the factorization
- Probabilistic model
- Using spatial map

multiple biological contexts or conditions
(e.g., time points, study subjects, and body sites)

Armingol, Erick, et al. "Context-aware deconvolution of cell–cell communication with Tensor-cell2cell." *Nature communications* 13.1 (2022): 3665.