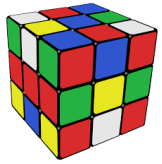
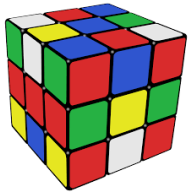
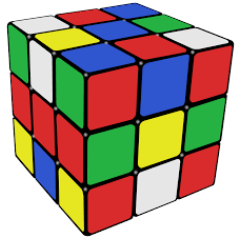
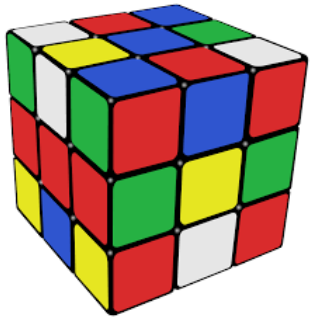
Tensors for Multi-Dimensional Data Analysis (with a brief survey on the applications)

Neriman Tokcan

What is a tensor?

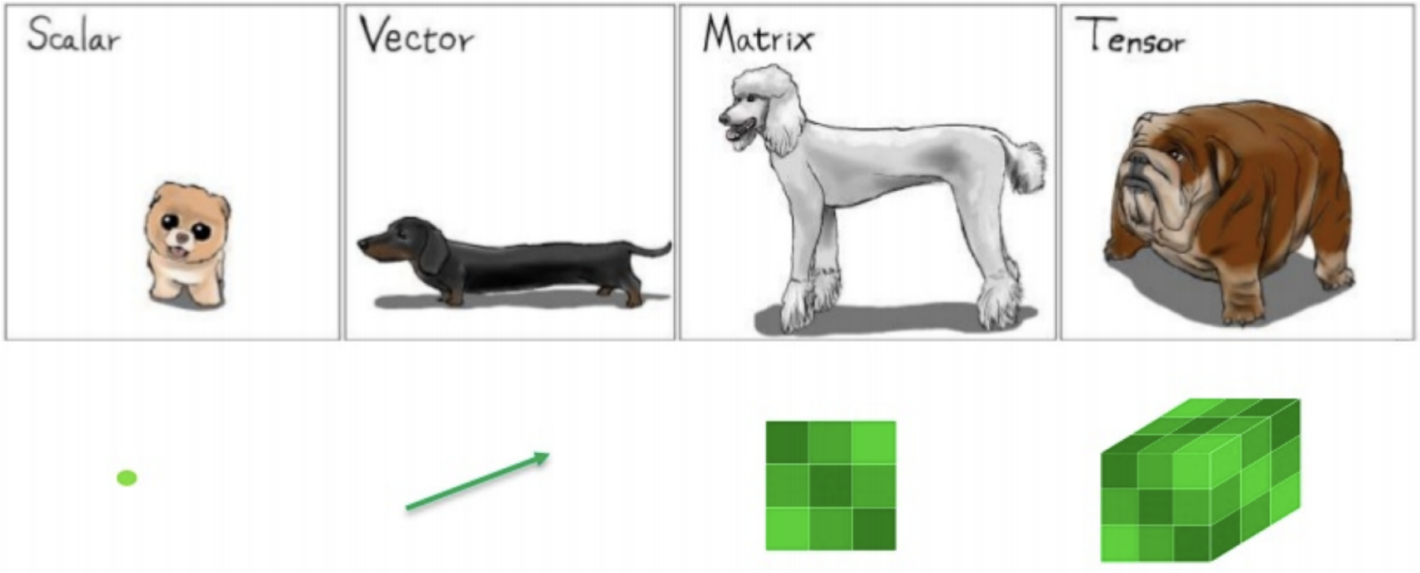
A tensor is a multi-way extension of a matrix:

- A multi-dimensional array
- A multi-linear map

We all know the following tensors:

- Scalars
- Vectors
- Matrices

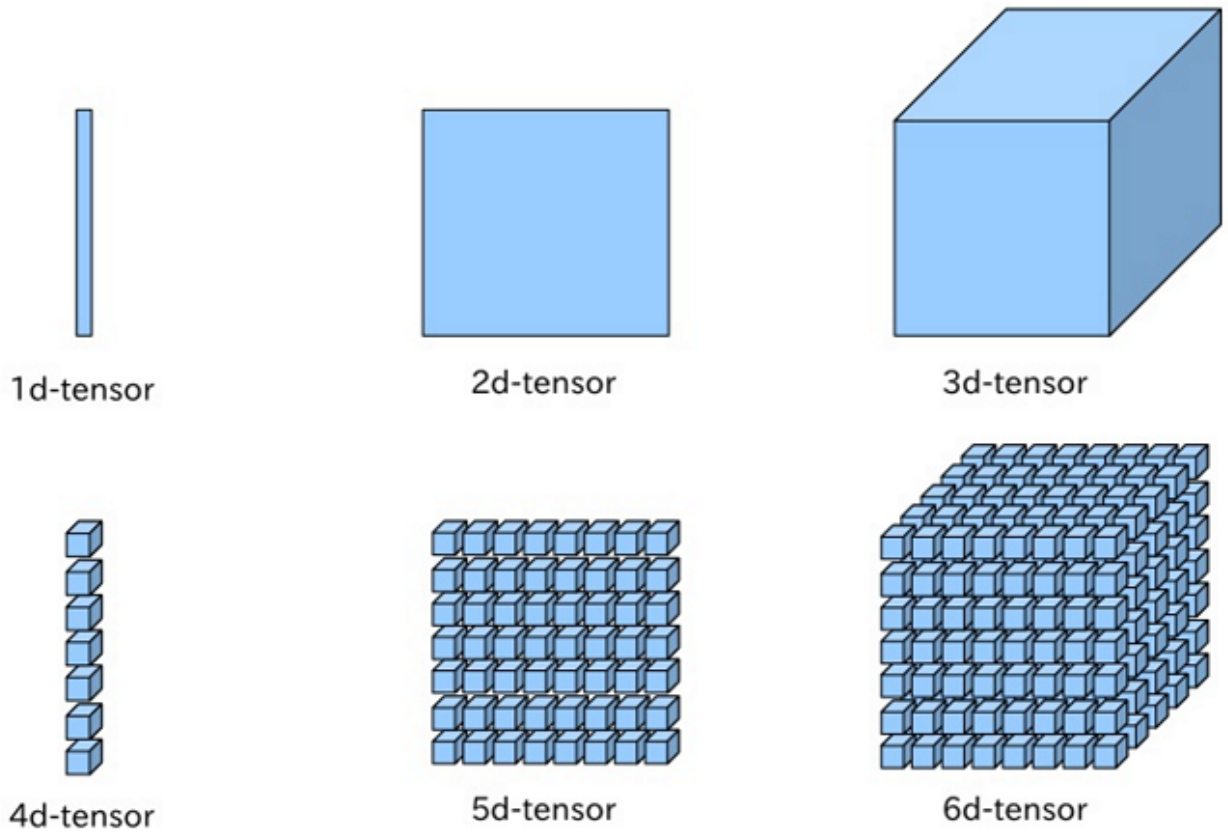
Multi-way extension?



1

[1] Figure: Anima Anandkumar

What is a tensor?



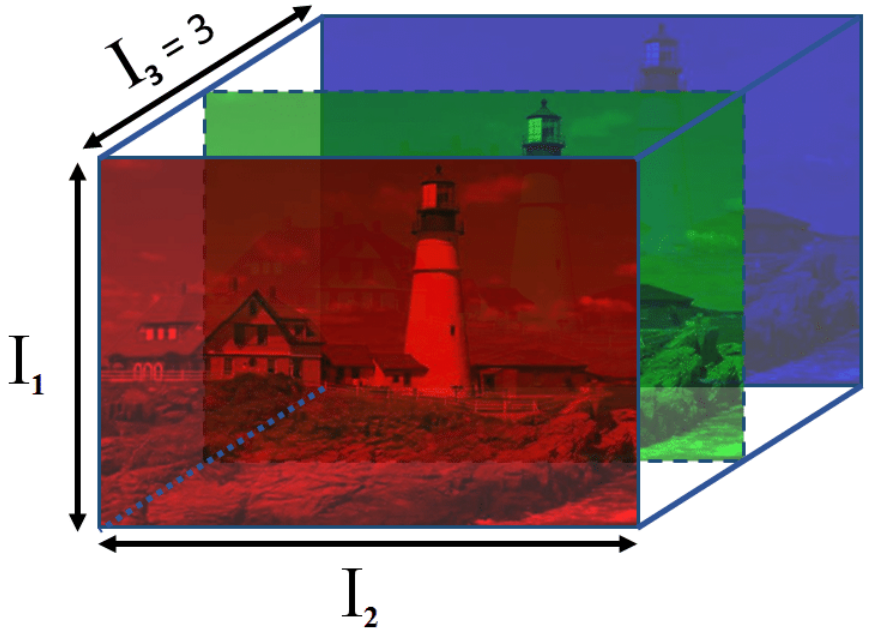
2

[2] https://www.slideshare.net/yokotatsuya/principal-component-analysis-for-tensor-analysis-and-eeg-classification

Notation and Preliminaries 1

- The *order* of a tensor is the number of dimensions, also known as ways, modes. A tensor $\mathcal{X}$ is an N-way (N-dimensional) array.
- *Fibers* are higher-order analogue of matrix rows and columns. A third order tensor has three modes: columns, rows and tubes.
- *Slices* are generated by fixing one of the indices. For a third order tensor, slices are two-dimensional sections.
- Unfolding (flattening, matricization) of a tensor is the process of reordering the elements of a tensor into a matrix. Foir instance a $3 \times 4 \times 5$ tensor can arranged as a 3×20 matrix, or a 4×15 matrix, and so on.

Examples



Order 3 tensor

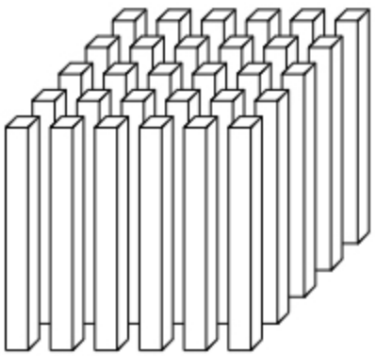
color video is 4th-order tensor

Order 4 tensor

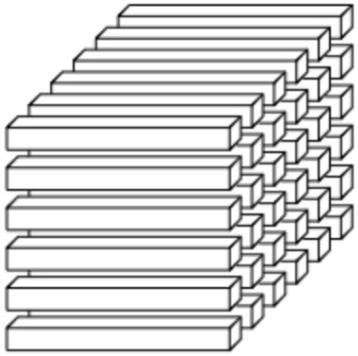
3

[3] M.A. Qureshi et al., *Quantifying Blur in Color images using Higher Order Singular Values*, https://www.slideshare.net/BertonEarnshaw/a-brief-survey-of-tensors

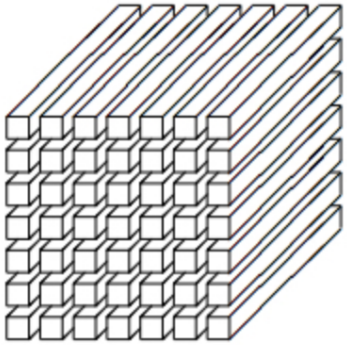
Fibers, Slices of tensors



(a) Mode-1 (column) fibers: $\mathbf{x}_{:jk}$

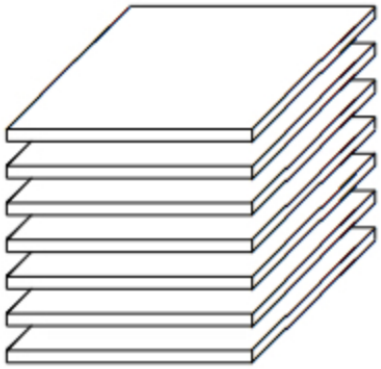


(b) Mode-2 (row) fibers: $\mathbf{x}_{i:k}$

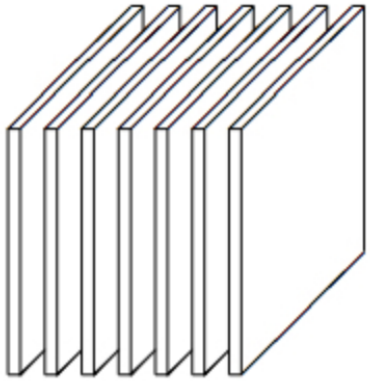


(c) Mode-3 (tube) fibers: $\mathbf{x}_{ij:}$

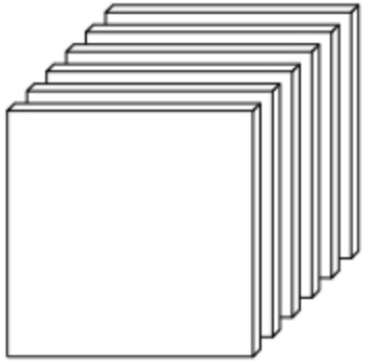
4



(a) Horizontal slices: $\mathbf{X}_{i::}$



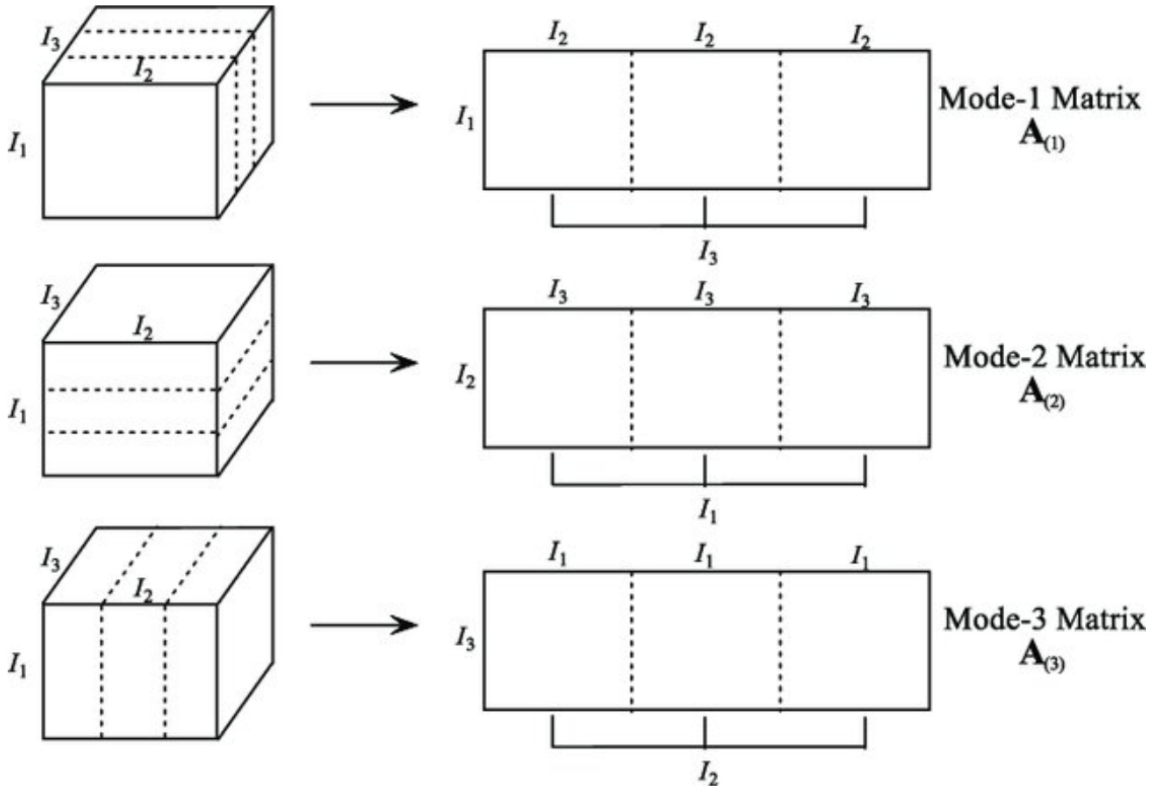
(b) Lateral slices: $\mathbf{X}_{:j:}$



(c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

[4] T. G. Kolda, B. W. Bader, *Tensor Decompositions and its Applications*, 2009

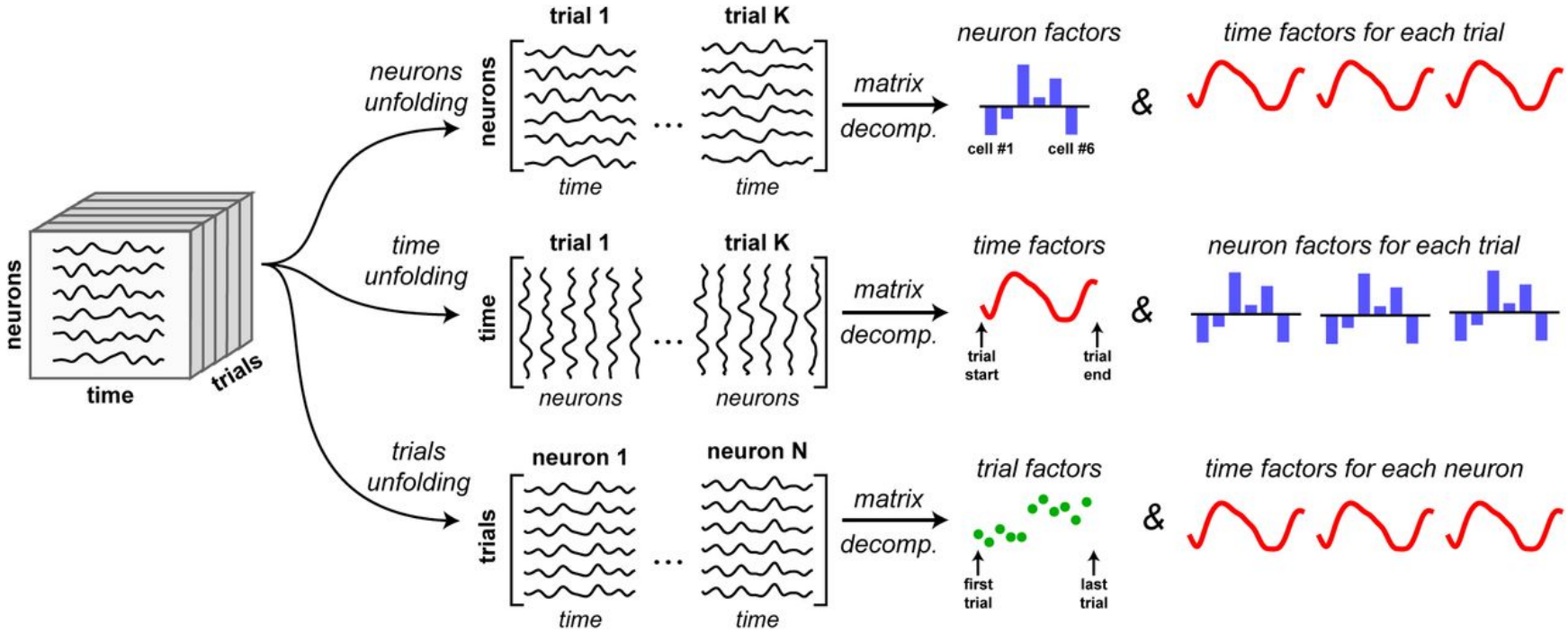
Unfolding of tensors



5

[5] Qiao et.al, *Generalized N-Dimensional Principal Component Analysis (GND-PCA) Based Statistical Appearance Modeling of Facial Images with Multiple Modes*, 2009.

Unfolding example



6

[6] Williams et al., *Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor components analysis*, 2017.

Tensors for Biomedical data

 Rapid growth in quantity and variety of biomedical data exceeds the capacity of matrix based data representations

 One of the highest challenges in biomedical data processing is the analysis of multi-modal data

 Tensors provide often natural and compact representation of such massive data

 Increasing number of multi-platform genome data of a single person, such as a cancer patient, are being generated.These data describe different biological aspects of a person and need to be integratively analyzed.

Why tensors and tensors decompositions?

- To analyze big data (As starting point express the tensor as sum of meaningful parts)
- For dimension reduction
- To exploit the structure of the data
- To reduce the computational complexity
- To deal with missing data (tensor completion)
- To deal with noisy data

Preliminaries - volume 2

Let u and v be vectors of dimensions n and m respectively, then their *outer product* is a *rank 1 matrix* of size $n \times m$:

$$u \otimes v = uv^T$$

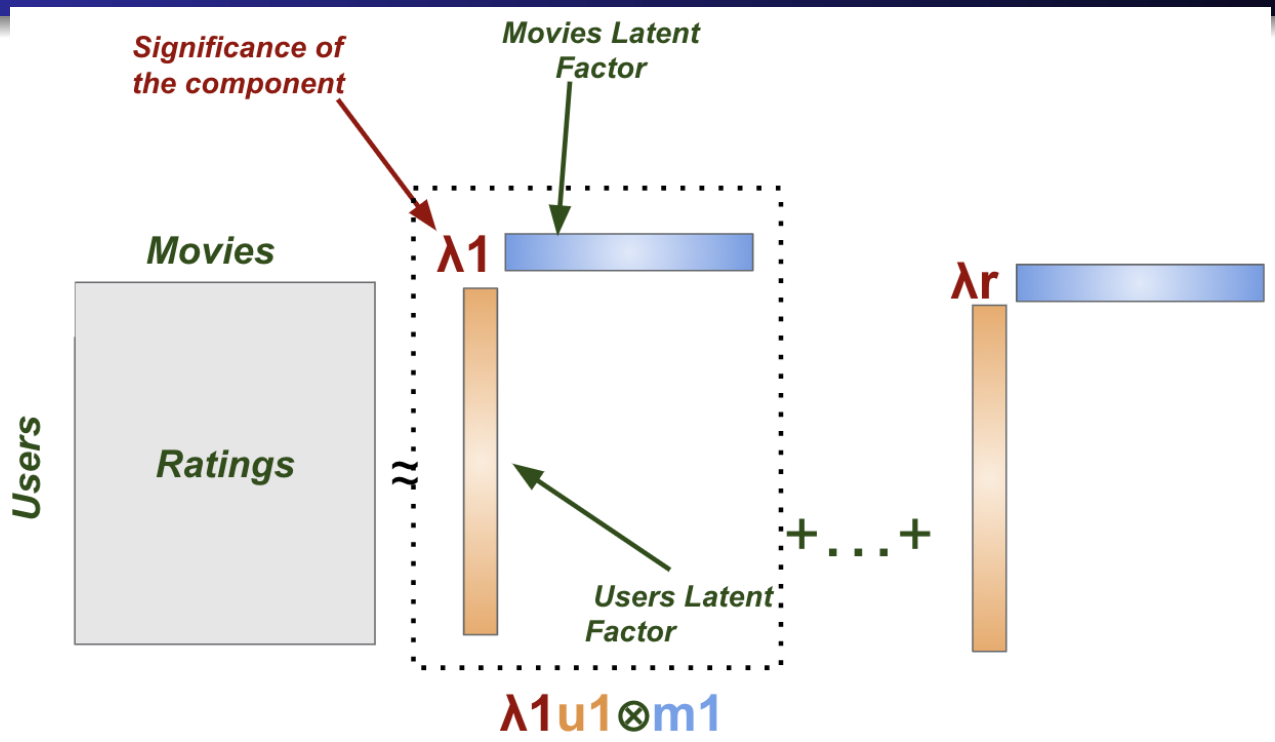
The outer product of vectors $u^{(i)}$ of size $I_i, 1 \leq i \leq N$ is a rank 1 (simple) tensor of size $I_1 \times I_2 \ldots \times I_N$:

$$(u^{(1)} \otimes u^{(2)} \ldots \otimes u^{(N)})_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

The *Frobenius norm* of an $I_1 \times I_2 \times I_3$ tensor $\mathcal{X}$ can be given as:

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i=1}^{I_1} \sum_{j=1}^{I_2} \sum_{k=1}^{I_3} \mathcal{X}_{ijk}^2}$$

Matrix factorization



$$Ratings \approx \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_r \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix} \begin{bmatrix} | & | & & | \\ m_1 & m_2 & \dots & m_r \\ | & | & & | \end{bmatrix}^T$$

$$Ratings \approx \sum_{i=1}^{r} \lambda_i u_i \otimes m_i$$

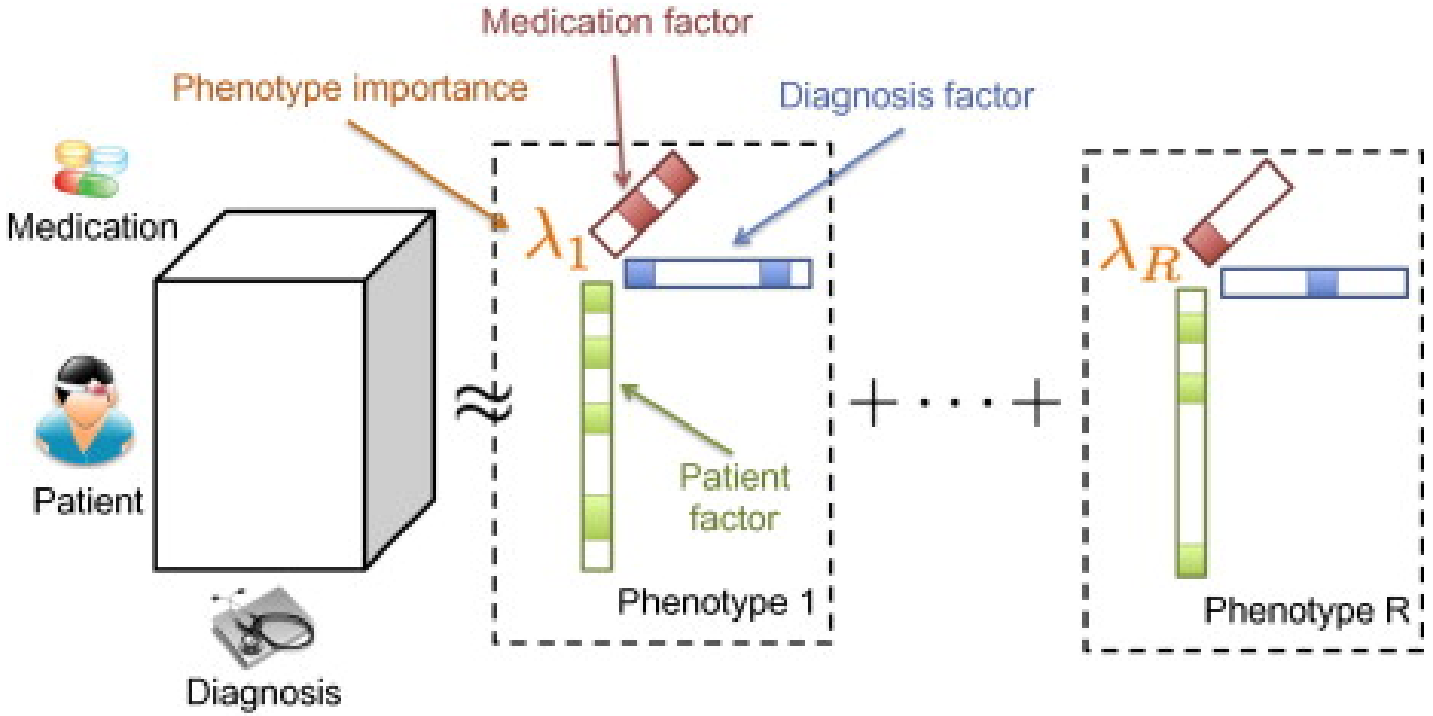
CP decomposition

Let $\mathcal{X}$ be a tensor of size $I_1 \times I_2 \times I_3$. The *rank* of $\mathcal{X}$ is the smallest r such that

$$\mathcal{X} = \sum_{i=1}^{r} \lambda_i a_i \otimes b_i \otimes c_i. \tag{1}$$

The decomposition given in (1) is known as CANDECOMP /PARAFAC (CP) decomposition.
The *factor matrices* refer to the combination of the vectors from the rank-one components, i.e., $A = [a_1 a_2 \dots a_r]$, $B = [b_1 b_2 \dots b_r]$ and $C = [c_1 c_2 \dots c_r]$.
The decomposition given in (1) can be concisely written as $\mathcal{X} \approx [[\Lambda; A, B, C]]$.

Summarizing data-phenotype generation

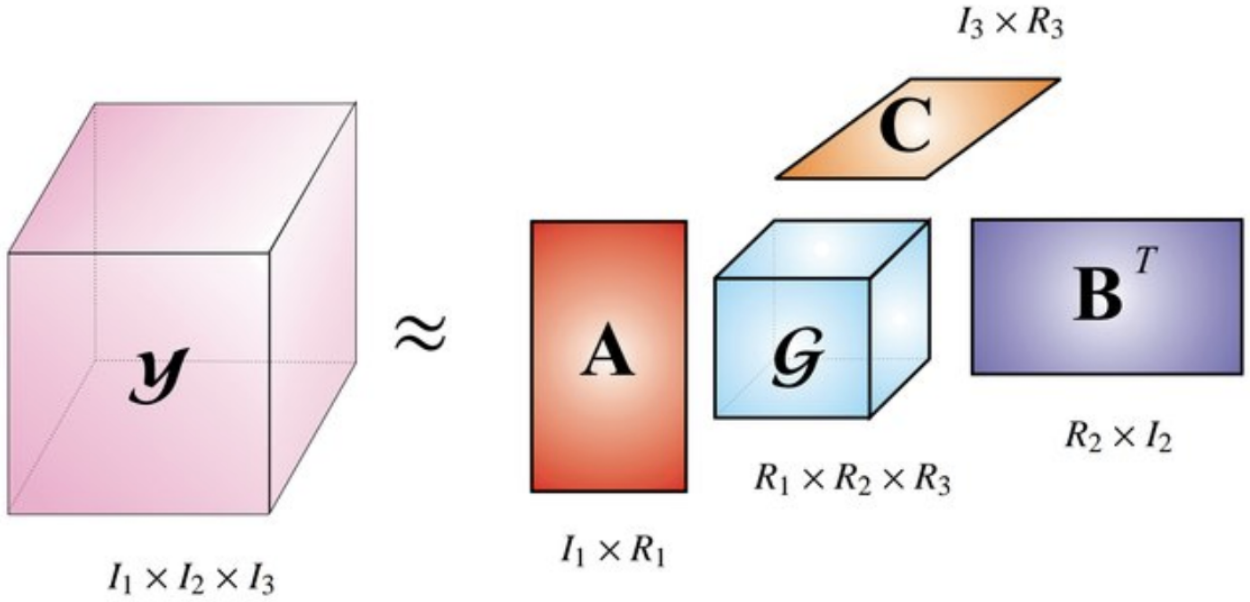


7

[7] Henderson et al., *Limestone: High-throughput candidate phenotype generation via tensor factorization, 2014*

Tucker decomposition

The *Tucker decomposition* is a form of higher-order PCA. It decomposes a tensor into a set of matrices and a small core tensor.



$$\mathcal{Y} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C$$

$$\mathcal{Y} \approx \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} \sum_{k=1}^{R_3} G_{ijk} a_i \otimes b_i \otimes c_i.$$

Alternating Least Squares for tensor decompositions

A common method for CP decomposition and other tensor-related optimization problems is Alternating Least Squares. We want to solve the following problem:

$min_{\tilde{\mathcal{X}}} \left\| \mathcal{X} - \tilde{\mathcal{X}} \right\|$ where $\mathcal{X} \approx \tilde{\mathcal{X}} = [[A, B, C]]$.

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

$$min_A \left\| \mathcal{X}^{(1)} - A(B \odot C)^T \right\|,$$
$$min_B \left\| \mathcal{X}^{(2)} - B(C \odot A)^T \right\|,$$
$$min_C \left\| \mathcal{X}^{(3)} - C(B \odot A)^T \right\|.$$

CP decomposition- Limitations

- CP decompositions are not always numerically stable
- Convergence is very slow
- Algorithm may not converge to a global minimum
- It is heavily dependent on the starting guess

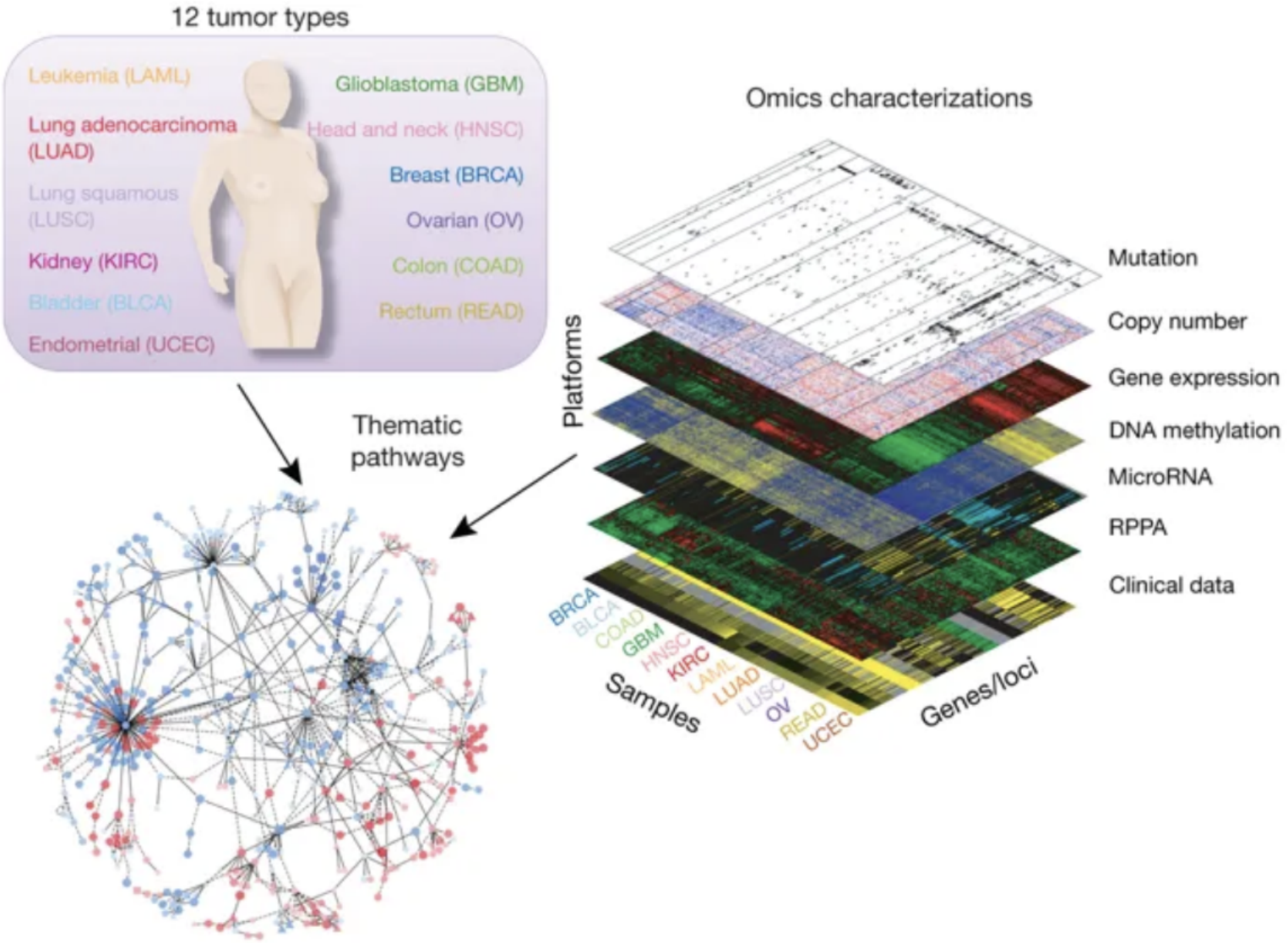
The Cancer Genome Atlas (TCGA)

TCGA is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies.

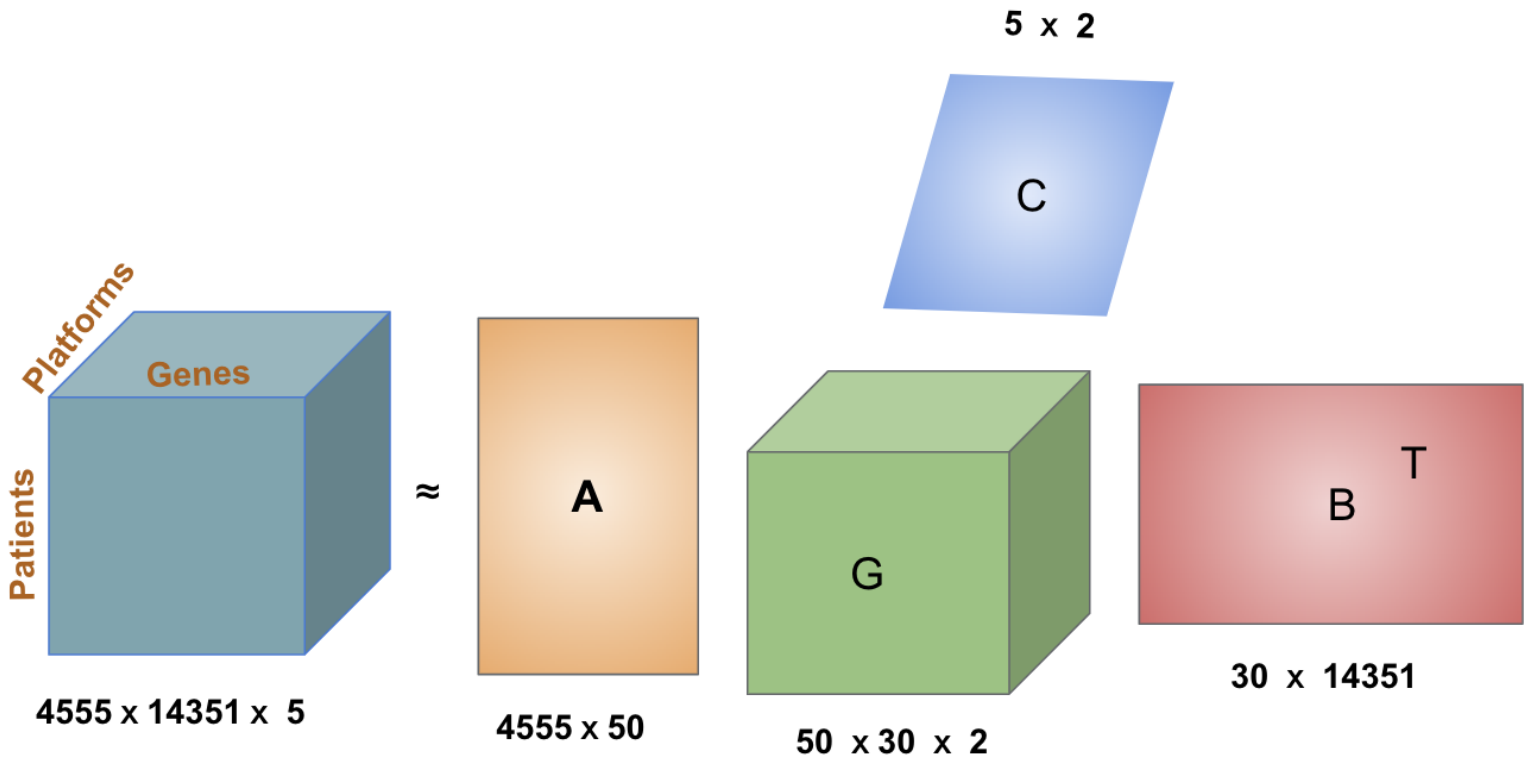
Outcomes and Impacts:

- Deepened our understanding of cancer through molecular characterizations
- Established a rich genomics data resource for the broad research community
- Helped advance health and science technologies
- Changed the way cancer patients are treated in the clinic

Pan-Cancer Project

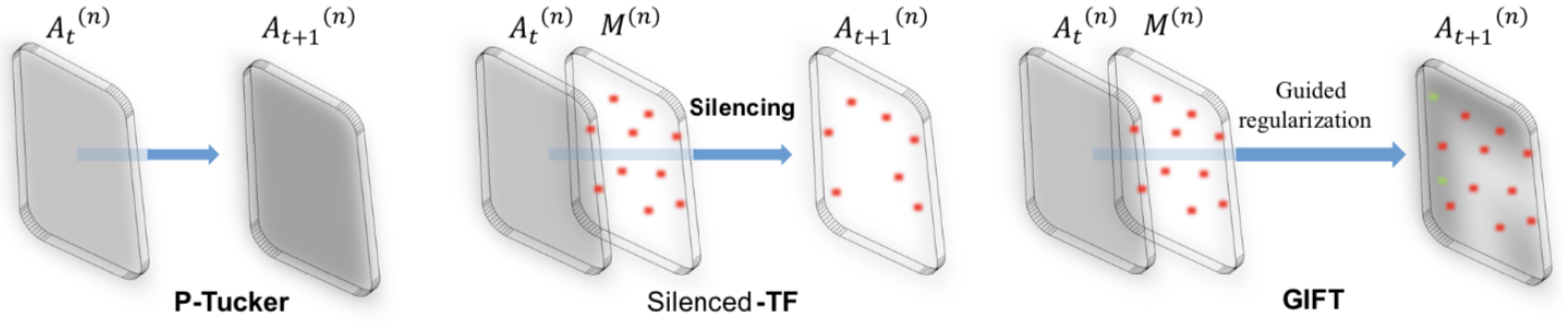


PANCAN12 Tensor – Tucker decomposition



Problems

- Data are embedded in a high-dimensional space with a low-dimensional flow pattern
- Genomics data are usually contaminated by noise
- Missing data or sparse tensors. Some effective methods: Partially Observable Tucker, Silenced Tucker, GIFT.

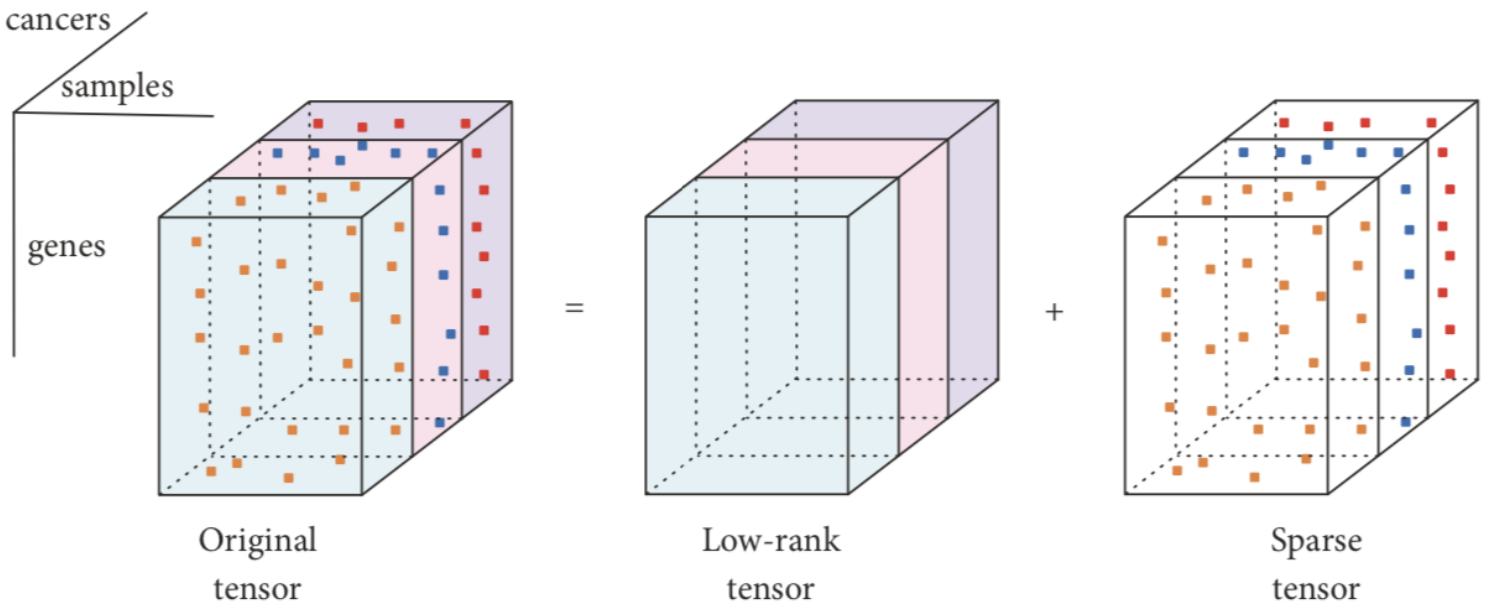


Oh S. et al., *GIFT: Guided and Interpretable Factorization for Tensors - An Application to Large-Scale Multi-platform Cancer Analysis*, 2018

Differentially Expressed Genes

- Gene expression data – low rank tensors: Gene expression data are close to some low-dimensional subspaces. It is natural to approximate nondifferentially expressed gene data with a low rank tensor.
- Differentially expressed genes – sparse tensors: Although the human body contains tens of thousands of genes, only a few are in fact related to biological processes. Therefore, the differentially expressed genes are treated as sparsely disturbed signals (sparse tensor) in the original data.

Low rank decomposition for differentially expressed genes



Y. Hu et al., *Differentially Expressed Genes Extracted by the Tensor Robust Principal Component Analysis (TRCPA) Method*, 2019.

First Matrix Case– Robust PCA

Original matrix | Low-rank matrix | Sparse matrix

https://kojinoshiba.com/robust-pca/

PCA – low rank matrix recovery

There are several ways to mathematically formulate PCA.
For a given data matrix X of size $m \times n$, (Trimmed) PCA can be formulated as follows:

$$\begin{aligned} & min_{Y,E} && ||E||_F, \\ & \text{subject to} && rank(Y) \leq r \text{ and } X = Y + E. \end{aligned}$$

- $r \leq min(m, n)$,
- $||.||_F$ is the Frobenius norm.

Note: PCA is severely affected by large-amplitude noise; not robust.

Robust PCA - RPCA

[Wright 2009]

$$\begin{aligned} min_{Y,E} \quad & rank(Y) + \lambda ||E||_0, \\ \text{s.t.} \quad & X = Y + E. \end{aligned}$$

- $||.||_0$ is ℓ_0 norm, number of non-zero elements in E.
- λ is a Lagrange multiplier.
- It is a matrix recovery problem
- $rank(Y)$ and $||.||_0$ are not continuous, not convex; very hard to solve.

Robust PCA– Convex version

[Candés2011] reformulated the problem:

$$\begin{aligned} min_{Y,E} \quad & ||Y||_* + \lambda ||E||_1, \\ \text{s.t.} \quad & X = Y + E. \end{aligned}$$

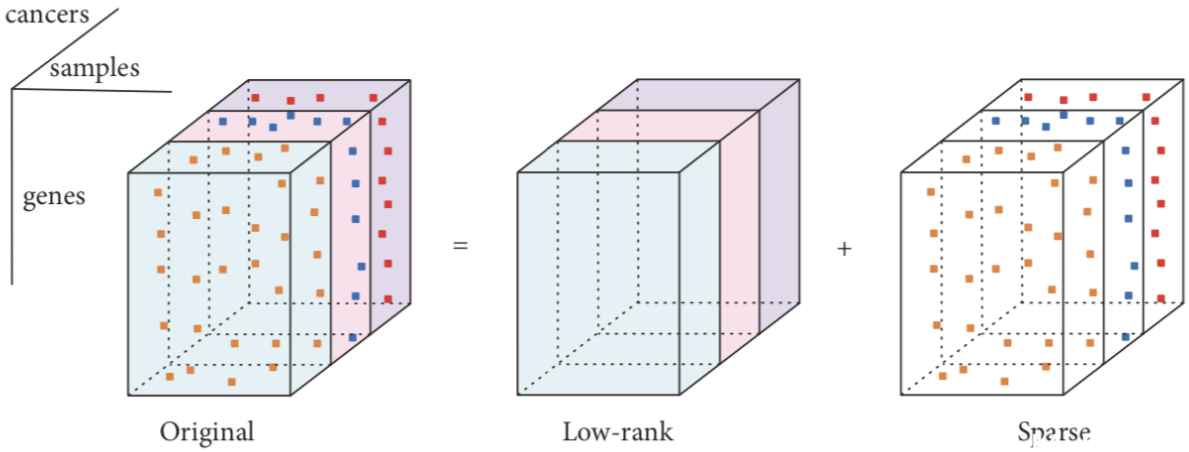
- $||Y||_*$ is the nuclear norm, sum of singular values of Y; convex surrogate for $rank(Y)$.
- $||E||_1$ is ℓ_1 norm, sum of absolute values of entries of E; surrogate for $||.||_0$.

Tensor Robust Principal Component Analysis for order 3

For a given tensor data $\mathcal{X}$, we want to decompose it $\mathcal{X} = \mathcal{Y} + \mathcal{E}$ into a low rank tensor and a sparse tensor. The objective function can be given as:

$$\begin{aligned} min_{\mathcal{Y},E} \quad & ||\mathcal{Y}||_* + \lambda||\mathcal{E}||_1, \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{Y} + \mathcal{E}. \end{aligned}$$

$$||\mathcal{Y}||_* = min\{\sum_{i=1}^{r} |\lambda_i| : \mathcal{Y} = \sum_{i=1}^{r} \lambda_i a_i \otimes b_i \otimes c_i, r \in \mathbb{R}\} \qquad (2)$$



Identification of Differentially Expressed Genes

Assume that $\mathcal{X}$ is a tensor of size $n \times m \times 3$, then $\mathcal{E}$ has 3 frontal slices: $\mathcal{E}_1 = \mathcal{E}(:,:,1), \mathcal{E}_2 = \mathcal{E}(:,:,2), \mathcal{E}_3 = \mathcal{E}(:,:,3)$.

Steps for each slice:

$$
\begin{aligned}
f_j &= \sum_{i=1}^{n} |\mathcal{E}_1(i,j)|, \ 1 \leq j \leq m, \\
\hat{\mathcal{E}}_1 &= (f_1, f_2, \ldots, f_m), \text{ arrange vectors in descending order,} \\
\bar{\mathcal{E}}_1 &= (\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_m).
\end{aligned}
$$

Filter out the top 500 maximum values and extract the corresponding genes. They utilized GO:TermFinder, an important for analysis of genomic data in which Gene Ontology information and rich Gene Ontology terms can be accessed.

Y. Hu et al., *Differentially Expressed Genes Extracted by the Tensor Robust Principal Component Analysis (TRCPA) Method*, 2019.

Data and Tests

The TCGA project included the 33 most common cancers and more than 11,000 tumor samples for sequencing.
COAD_HNSC_ESCA_GE: 20502 genes * 192 samples * 3 cancer types
(colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), esophageal carcinoma (ESCA))

- After getting feature vectors, GO:TermFinder is used,
- Performance computation of different methods were evaluated using P-values and hit counts.
- The experimental method corresponding to the smaller P-value indicates that the effect of differentially expressed genes is better
- Maximum value of p value is set to 0.01.

Results

ID	TRPCA		RPCA		LLRR		PCA		BRTF		Count in genome
	P-value	Hit Count	P-value	Hit Count	P-value	Hit Count	P-value	Hit Count	P-value	Hit Count	
GO:0006614	*6.49E-74*	*61*	5.91E-56	51	5.31E-56	51	4.35E-72	60	5.60E-63	55	94
GO:0006613	*5.10E-71*	*61*	8.65E-54	51	7.76E-54	51	2.84E-69	60	1.53E-60	55	101
GO:0045047	*1.24E-70*	*61*	8.65E-54	51	7.76E-54	51	2.84E-69	60	3.25E-60	55	102
GO:0072599	*3.78E-69*	*61*	2.24E-52	51	2.01E-52	51	1.87E-67	60	5.89E-59	55	106
GO:0070972	*4.88E-68*	*64*	1.03E-50	53	9.22E-51	53	*4.25E-68*	64	6.89E-57	57	125
GO:0000184	*6.19E-66*	*62*	1.20E-51	53	1.07E-51	53	*5.41E-66*	62	6.35E-58	57	121
GO:0000956	*5.56E-57*	*68*	1.27E-41	56	6.89E-43	57	*4.81E-57*	68	1.19E-46	60	202
GO:0006612	*2.63E-54*	*65*	1.81E-41	55	9.57E-43	56	4.92E-53	64	8.00E-48	60	194
GO:0019083	*2.35E-53*	*62*	9.12E-43	54	8.16E-43	54	4.73E-52	61	4.85E-48	58	176
GO:0006401	*2.34E-50*	*68*	1.85E-36	56	1.32E-37	57	*2.04E-50*	68	5.19E-41	60	247

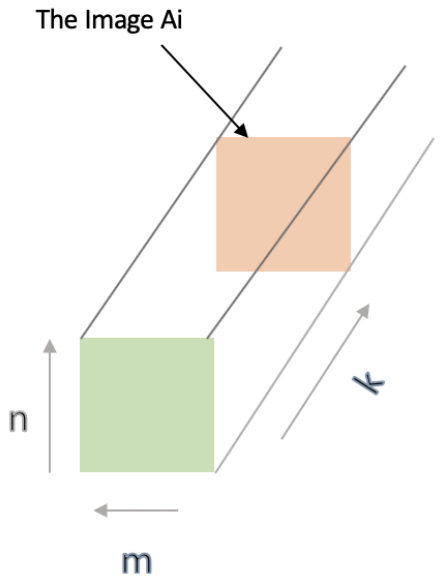
- The P-value indicates the enrichment degree of the gene.
- P-value of GO:0006614 was 6.49E-74, which is much smaller than the P-values of other methods.
- There were 94 genes in the GO:0006614 terminology, and RPCA, LLRR, PCA, and BRTF could detect 51, 51, 60, and 55 genes, respectively. However, 61 genes were identified using the TRPCA method.

Y. Hu et al., *Differentially Expressed Genes Extracted by the Tensor Robust Principal Component Analysis (TRCPA) Method*, 2019.

Question: What factorization principle would support a decomposition of training images of a class of objects into a basis of local parts?

Matrix based representation: Images are vectorized

Tensor based representation: 2D representation of images are preserved



Non-negative Matrix Factorization: Let V be a vector whose columns are the vectorized training images.

$$V \approx WH, W \geq 0, H \geq 0.$$

The columns of W form the new basis vectors and due to non-negativity constraint both the basis vectors and mixing coefficients tends to come out sparse.

Non-negative Tensor Factorization

Let $A_t,\ t = 1, \ldots, k$ be images of dimension $n \times m$. We stack them to get a 3rd order tensor $\mathcal{X}$ of size $n \times m \times k$. We want to factorize $\mathcal{X}$ and consider the following least-squares problem:

$$min_{u_i, v_i, w_i} \quad ||\mathcal{X} \approx \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i||_F$$
$$\text{subject to :} \quad u_i, v_i, w_i \geq 0.$$

We get three factor matrices:
$U = [u_1, u_2, \ldots, u_r],\ V = [v_1, v_2, \ldots, v_r],\ W = [w_1, w_2, \ldots, w_r]$.
How to capture relationship between the 2D images and the above factorization?

- $vec(A_t)$ is a linear combination of rank1-matrices $u_i \otimes v_i = vec(u_i v_i^T)$ with coefficients are taken from the t-th row of w_t.
- $A_t = U \lambda_t V^T$ where $\lambda_t = diag(w_{1t}, w_{2t}, \ldots w_{rt})$.

Sparse Image Coding—NTF versus NMF

	NMF	NTF
1	Vectorizes images	It represents the image collection as a 3-way array
2	Decomposition is not unique	Decomposition is unique under mild conditions (even without non-negativity)
3	Sparse decomposition, not separable	Sparse and separable decomposition

1. Vectorizing an image will lead to information loss as the local image structure (spatial redundancy) would be lost
2. It is not clear whether the NMF process will yield the underlying generative parts, even when there is a perfect fit. Problem: Invariant parts which in fact create "ghost" in the factors and contaminate the sparsity of basis vectors.
3. Efficiency- Factors of NTF are sparse and separable, they are significantly more compressed than NMF factors.

Experiment 1–Swimmer image set

Data: Swimmer image set of 256 images of dimensions 32×32. Each image contains a "torso" in the center and four "limbs" that can be in 4 different positions.

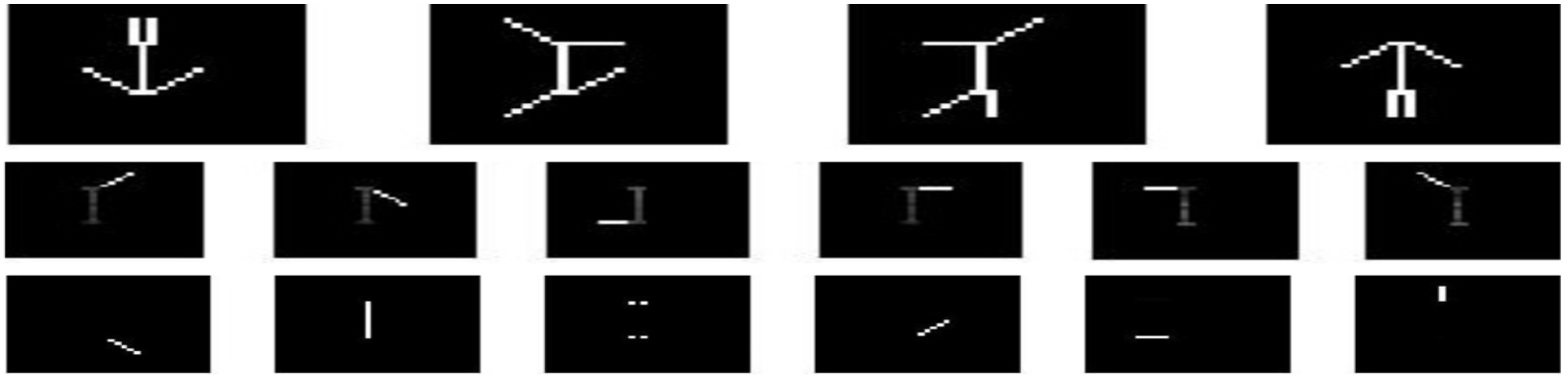


Figure: Factors generated by NMF (middle row), NTF (bottom row). The NMF factors contains "ghost" of invariant parts(the torso) which contaminate the sparse decomposition.

NMF and NTF both find 17 factors correctly resolves the local parts. NMF fails on the torso. NTF contains a unique factorization.

Hazan et al., *Sparse Image Coding Using a 3D Non-Negative factorization*

NTF resolving local parts a single image

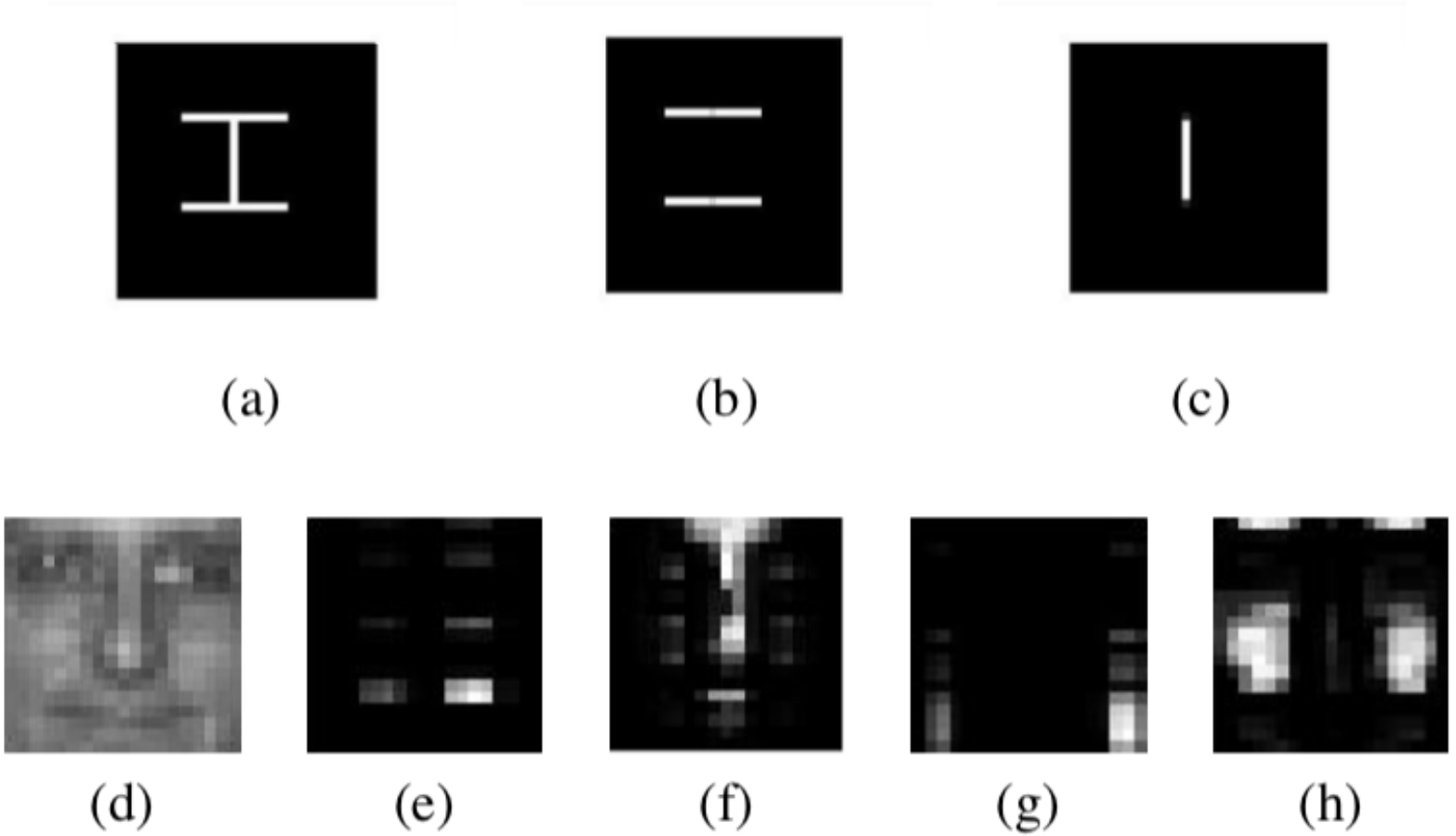


Figure 3: *Running NTF on a single image copied 20 times to form a 3D cube. Upper row: (a) the original image, (b),(c) the two recovered factors. Lower row: (d) the original image, (e)-(h) the recovered factors in 4 groups.*

Hazan et al., *Sparse Image Coding Using a 3D Non-Negative factorization*

Experiment 2 –MIT CBCL face dataset

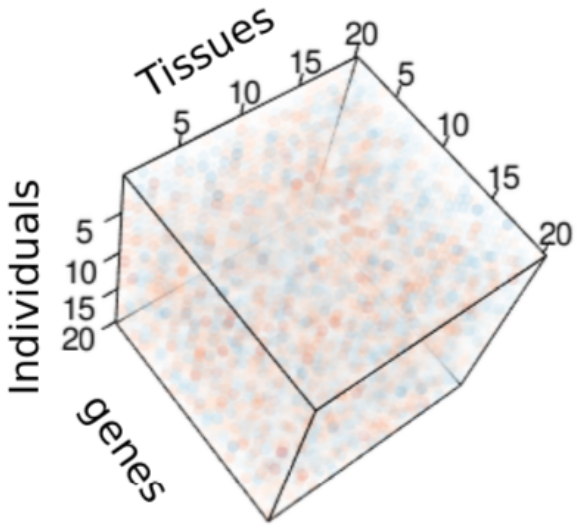
Data: MIT CBCL database, set of 2429, 19×19 face images.
Experiment:Using the filter responses of NMF, NTF, PCA as measurement for an SVM classifier, with linear, polynomial of degree five and RBF kernels, trained over the face dataset. 50 NTF factors are computed to reconstruct the original images. NTF framework has higher compression rate and each NMF is comparable to 19 NTF factors.

	linear	poly $d = 5$	RBF
NTF (50)	91.9%	95.3%	95.9%
NMF (50)	91.6%	94%	95%
NMF (20)	87.5%	90.1%	89%
NMF (6)	83.2%	84.3%	86%
PCA	90.8%	94%	91.7%

Figure: The NTF outperformed the NMF even 50 NMF factors were used (20-fold higher space than NTF)

Three-Way Clustering of Multi-Tissue Multi-Individual Gene Expression Data

A typical multi-tissue experiment collects gene expression profiles from different individuals in a number of different tissues, and variation in expression levels often results from complex interactions among genes, individuals, tissues.



Methods: Clustering has proven useful to reveal latent structure in high-dimensional expression data. Traditional methods: K-means, PCA, t-SNE,etc.

Wang et al., *Three-Way Clustering of Multi-Issue Multi-Individual Gene Expression Data Using Semi-nonnegative tensor decomposition*. 2019

Problems with the traditional methods

- These methods assume that gene expression patterns persist across one of the different contexts, or that samples are i.i.d or homogenous.
- Arranging the given data as matrices brings some problems:
  1. precluding potential insights into tissue $\times$ individual specificity,
  2. inferring gene modules separately for each tissue ignores commonalities among tissues and may hinder the discovery of differentially-expressed (DE) genes,
  3. ignoring individual heterogeneity (biological attributes such as race, gender, and age) impedes the accurate estimation of gene-and/or tissue-wise correlations.

Tensor based approaches have been proposed to handle heterogeneity in each mode and learns the clustering patterns across different modes of the data in an unsupervised manner analogous to PCA and SVD.

Dataset & Semi-nonnegative tensor decomposition

Genotype-Tissue Expression (GTEx) RNA-seq data, which consist of RNA-seq samples collected from 544 individuals across 53 human tissues. The GTEx data set contains categorical clinical variables such as sex, race, and age.
The expression tensor $\mathcal{Y} \in \mathbb{R}^{n_G \times n_I \times n_T}$ is modeled as a perturbed rank-r tensor,

$$\mathcal{Y} = \sum_{i=1}^{r} \lambda_r G_r \otimes I_r \otimes T_r + \mathcal{E},$$

where $\lambda_r \in \mathcal{R}_+$; $G_r, I_r,$ and T_r are norm-1 vectors; and
$\mathcal{E} = [[E_{i,j,k}]]$ is a noise tensor with each entry $E_{i,j,k}$ i.i.d. $N(0, \sigma_e^2)$.

$G_r \rightarrow$ eigen-genes

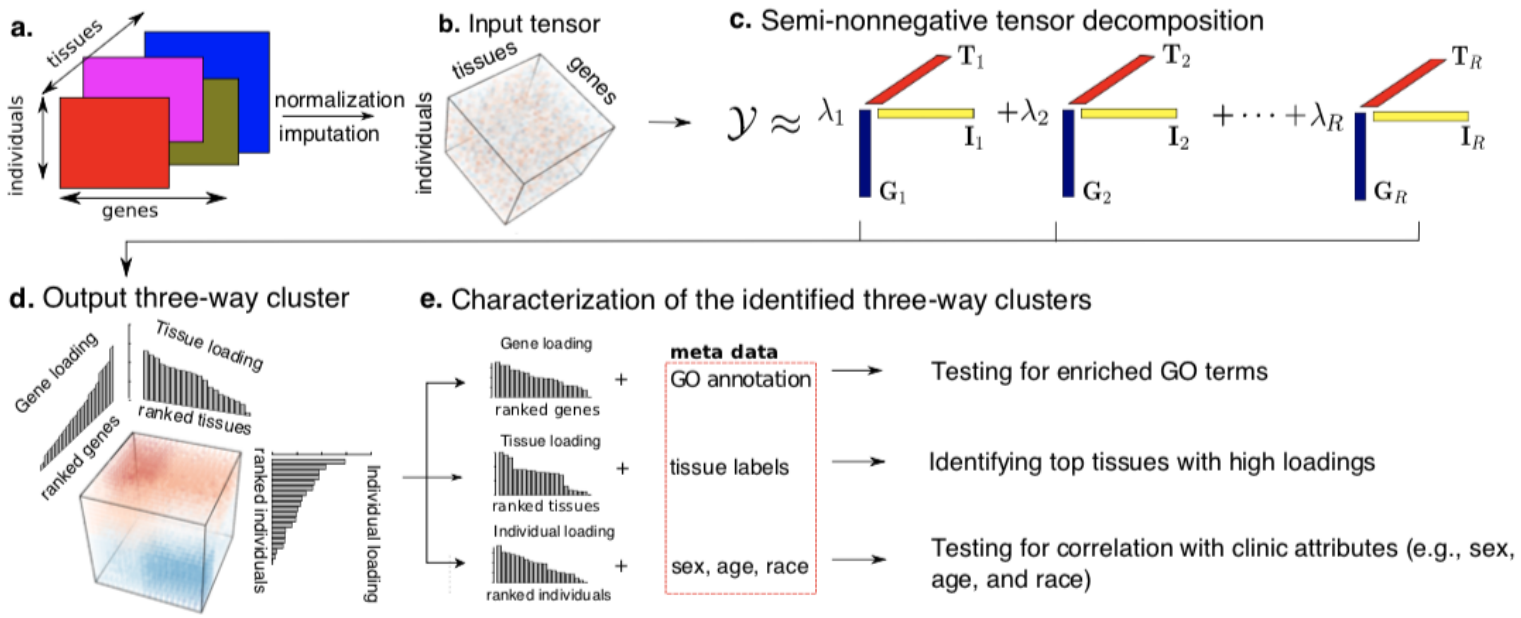
$I_r \rightarrow$ eigen-individuals

$T_r \rightarrow$eigen-tissues

$G_r \otimes T_r \otimes I_r \rightarrow$ basic unit of an expression pattern

where $(G_r \otimes T_r \otimes I_r)_{i,j,k} = G_{r,i} T_{r,j} I_{r,k}$.

Clustering data using Semi-nonnegative Tensor Decomposition



[9]

[9]Wang et al., *Three-Way Clustering of Multi-Issue Multi-Individual Gene Expression Data Using Semi-nonnegative tensor decomposition*, 2019

Experiment-1

Expression tensor consisting of 60 genes, 20 individuals and 10 tissues. The 20 individuals were partitioned into two groups (young vs. elderly), each of size 10. The genes and tissues were each partitioned into three groups (denoted by A, B, C).

TABLE 1
Mean expression value of the illustrative tensor.

	Tissue Group A		Tissue Group B		Tissue Group C	
Gene \ Individual	Young	Elderly	Young	Elderly	Young	Elderly
Gene Group A	1	−1	−1	1	0	0
Gene Group B	0	0	0.5	−0.5	0.1	−0.1
Gene Group C	0	0	0	0	0.5	−0.5

[10]

[10] Wang et al., *Three-Way Clustering of Multi-Issue Multi-Individual Gene Expression Data Using Semi-nonnegative tensor decomposition*. 2019

PCA and fixed-effect data analysis

- PCA: They averaged the expression over individuals and apply matrix PCA.

  Neither the mode-specific grouping nor the three-way interaction can be recovered.

- Fixed-effect meta analysis They tested the age effects in each tissue separately and combined the test statistics into a pooled estimate using z-score method

  It suffers from low power for detecting DE genes. The meta-analysis poor performance is due to the tissue- specificity of DE genes: genes in Gene Group A have opposite age effects in two of the tissue groups, so the signals partially cancel out; moreover, genes in Gene Groups B and C have age effects in only subsets of tissues, potentially diluting observed DE patterns.

Results

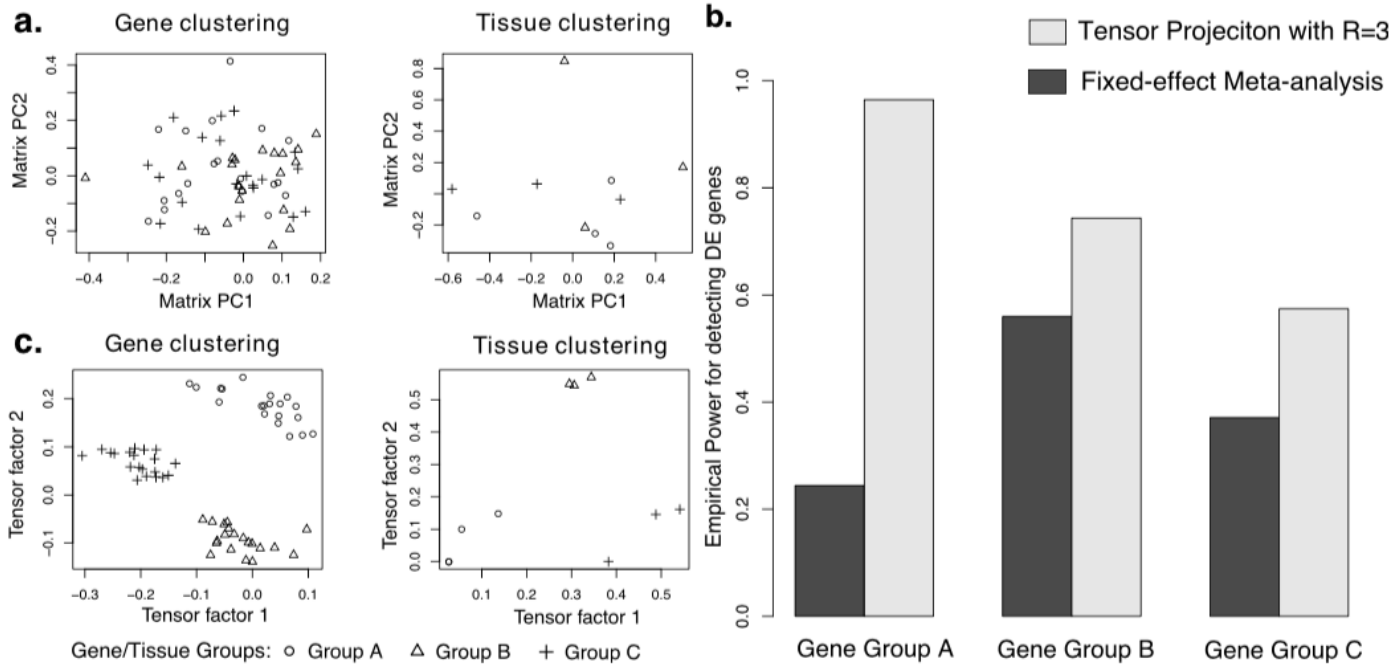


Fig 3 Performance comparison for the illustrative example. (a) First two gene/tissue factors in the matrix PCA. (b) Power comparison for detecting age effects in three gene groups. (c) First two gene/tissue factors in the tensor decomposition.

11

[11] Wang et al., *Three-Way Clustering of Multi-Issue Multi-Individual Gene Expression Data Using Semi-nonnegative tensor decomposition*. 2019