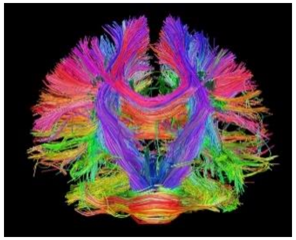



CS-570
Statistical Signal Processing

Lecture 12: Tensor models

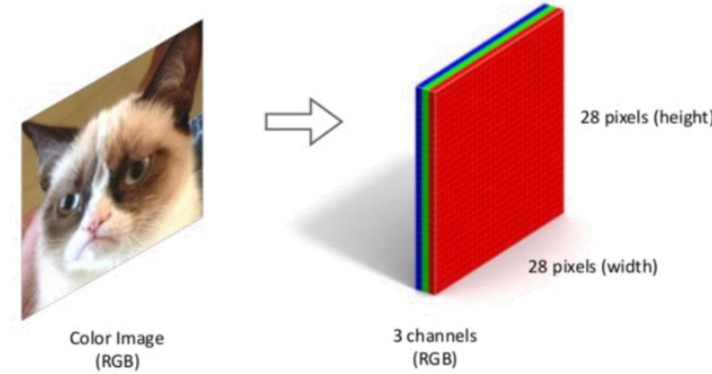
Spring Semester 2019

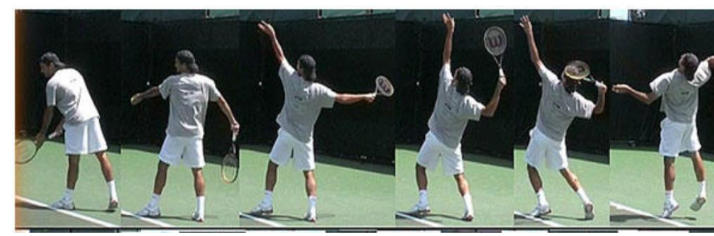
Grigorios Tsagkatakis

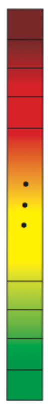
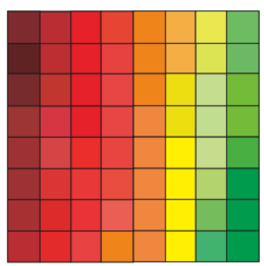
High-dimensional signal models

Encoding of multiple variables

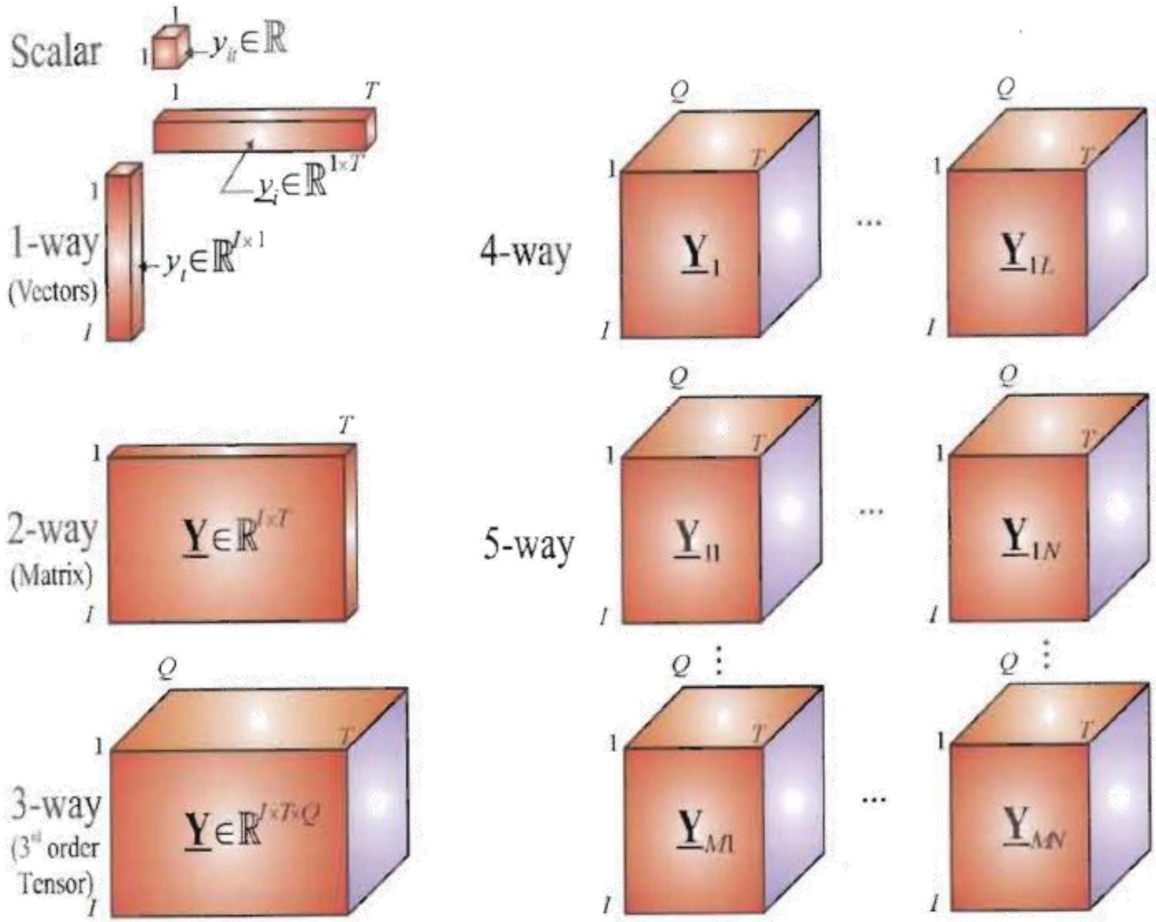
- Time, Space, Frequency, Modality





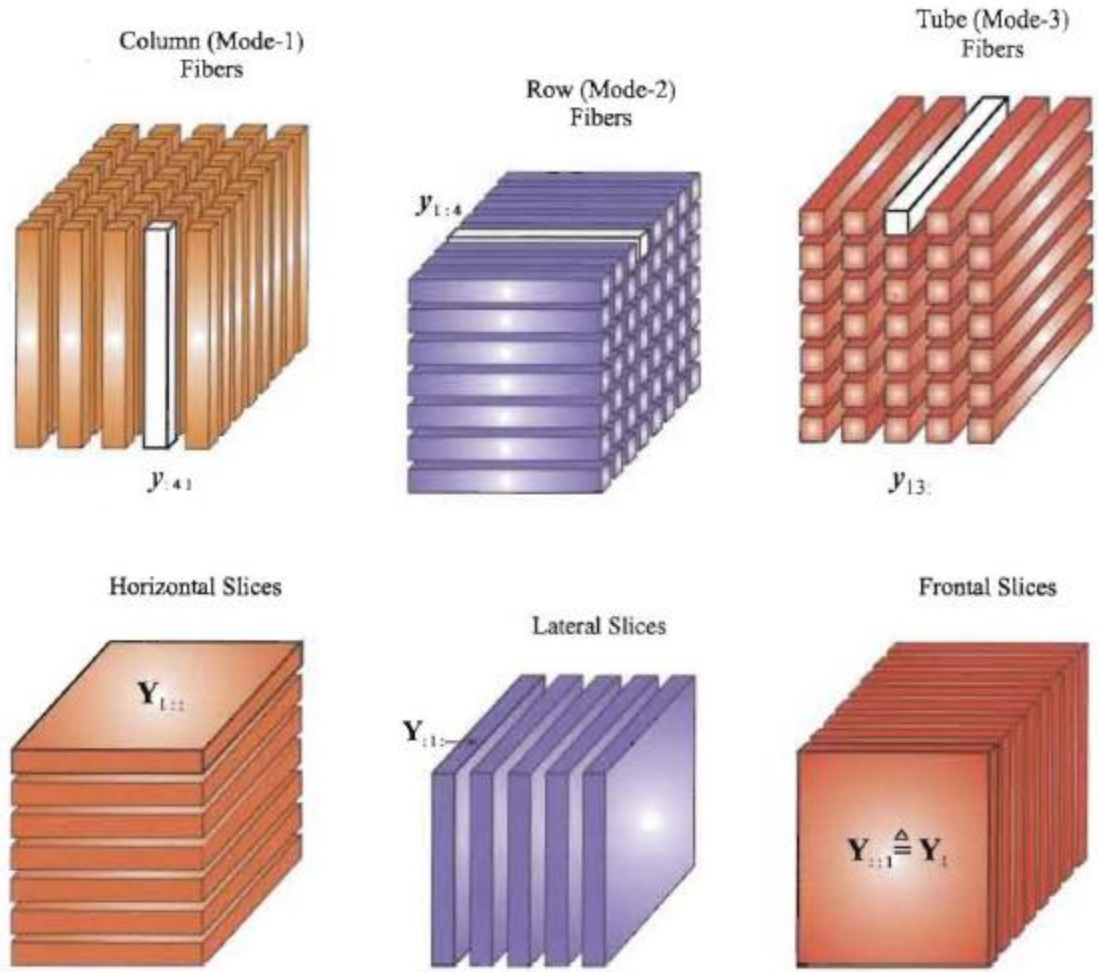
vector	matrix	tensor
		
$\mathbf{v} \in \mathbb{R}^{64}$	$X \in \mathbb{R}^{8x8}$	$\mathcal{X} \in \mathbb{R}^{4x4x4}$

Tensors



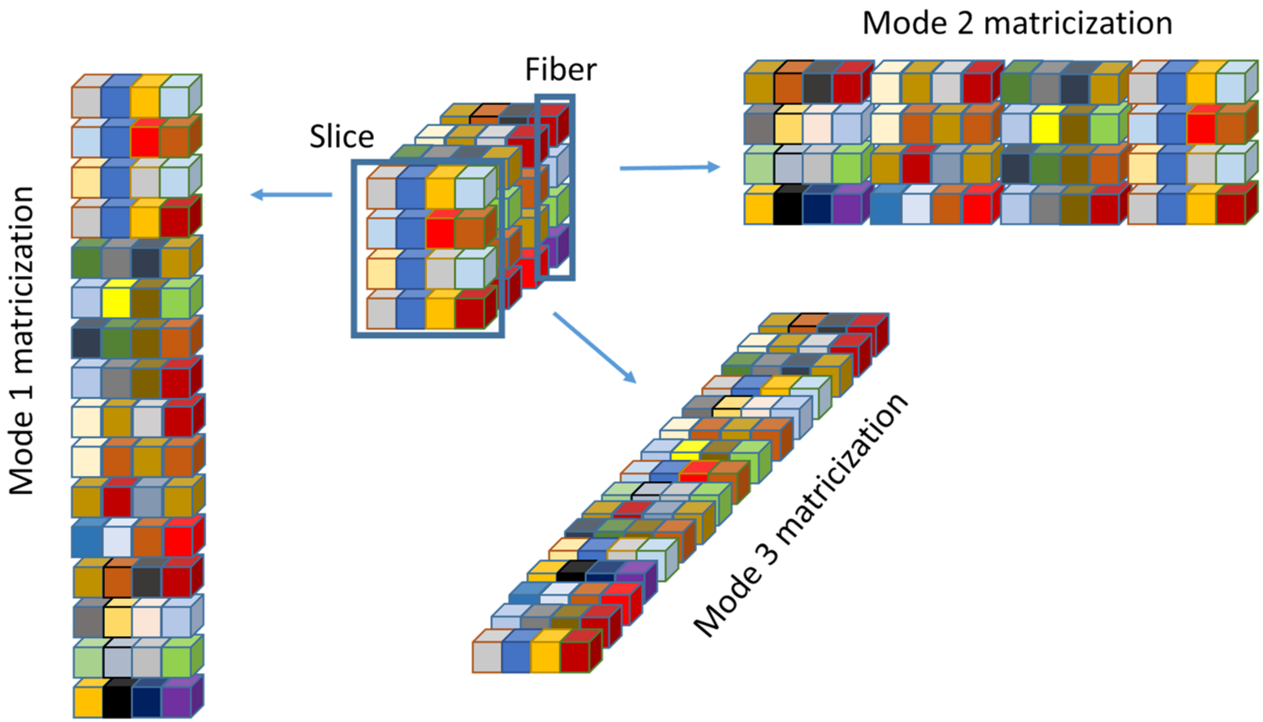
Includes materials from: Introduction to tensor, tensor factorization and its applications, by Mu Li, iPAL Group Meeting, Sept. 17, 2010

Fiber and slice



Tensor unfoldings: Matricization and vectorization

- Matricization: convert a tensor to a matrix



Tensor Mode-n Multiplication

$$\mathcal{X} \in \mathbb{R}^{I \times J \times K}, \ \mathbf{B} \in \mathbb{R}^{M \times J}, \ \mathbf{a} \in \mathbb{R}^{I}$$

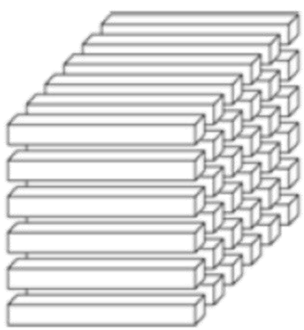
- Tensor x Matrix

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{B} \in \mathbb{R}^{I \times M \times K}$$

$$y_{imk} = \sum_j x_{ijk} \, b_{mj}$$

$$\mathbf{Y}_{(2)} = \mathbf{B}\mathbf{X}_{(2)}$$

Multiply each row (mode-2) fiber by **B**

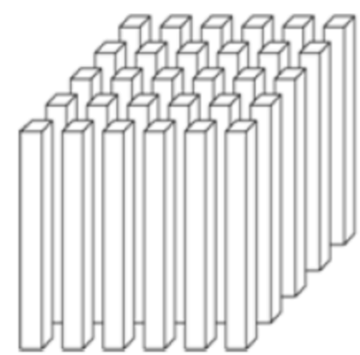


- Tensor x Vector

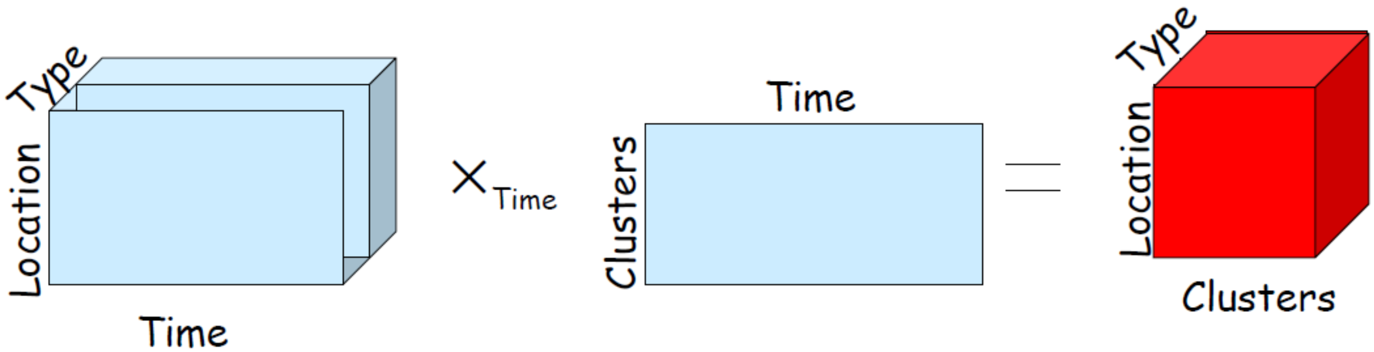
$$\mathbf{Y} = \mathcal{X} \,\bar{\times}_1\, \mathbf{a} \in \mathbb{R}^{J \times K}$$

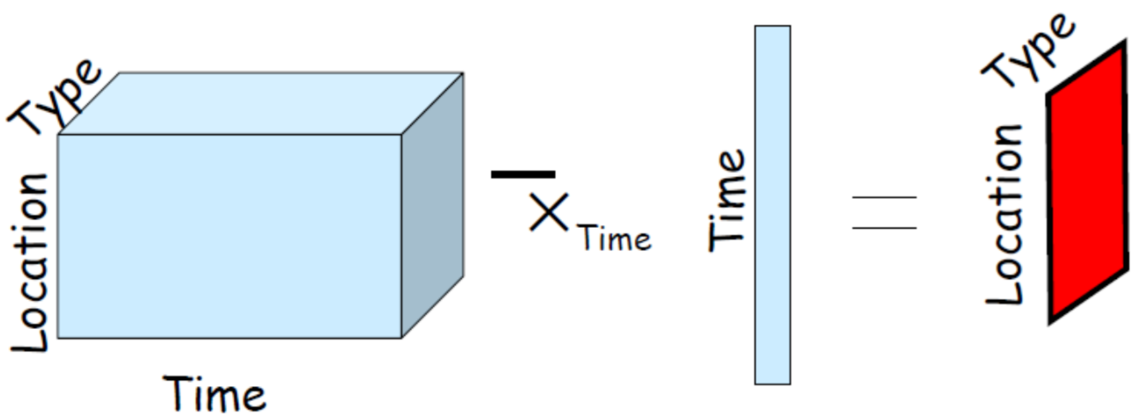
$$y_{jk} = \sum_i x_{ijk} \, a_i$$

Compute the dot product of a and each column (mode-1) fiber



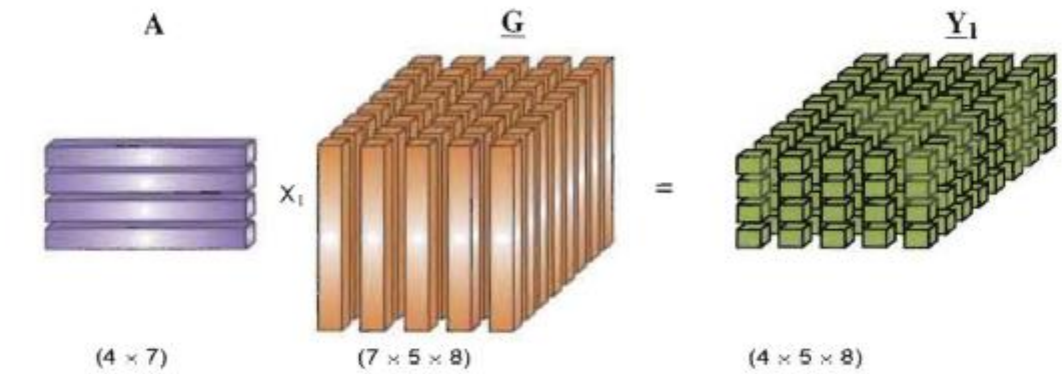
Examples

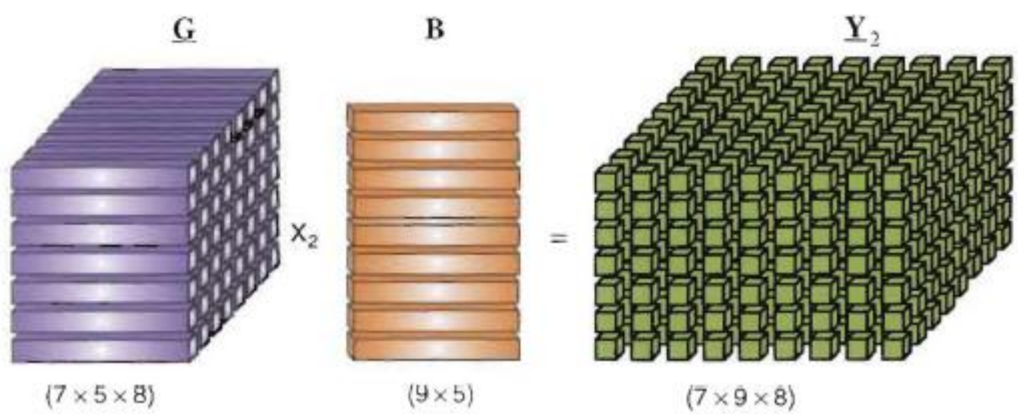

Type
Location
Time
×Time
Time
Clusters
=
Type
Location
Clusters


Type
Location
Time
×Time
Time
=
Type
Location

Tensor multiplication: the n-mode product: multiplied by a matrix

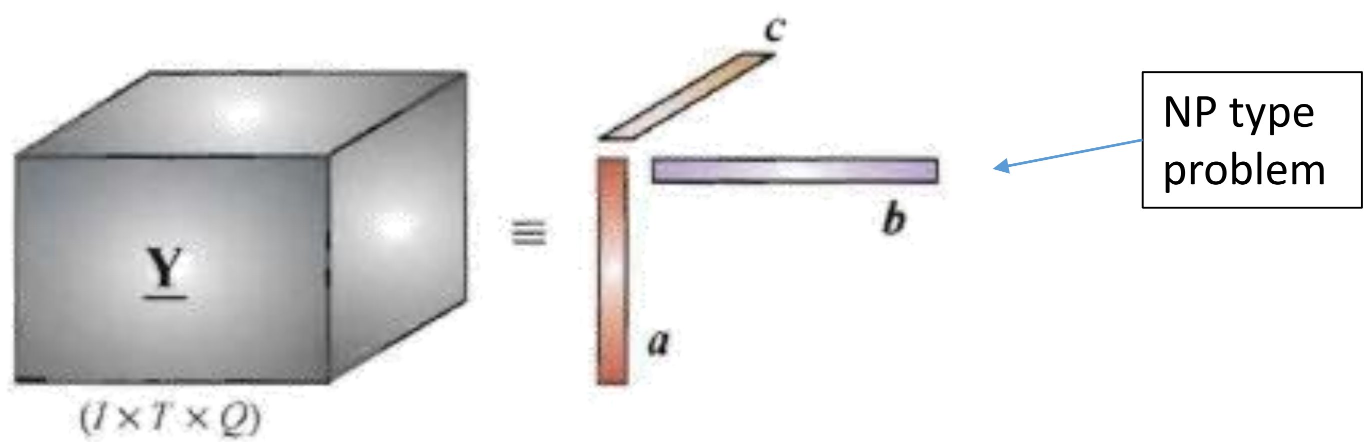
$$(\mathcal{X} \times_n \mathbf{U})_{i_1 \cdots i_{n-1} j\, i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N}\, u_{j i_n}.$$





Tensor models

- For two vectors $\mathbf{a}$ $(I \times 1)$ and $\mathbf{b}$ $(J \times 1)$, $\mathbf{a} \circ \mathbf{b}$ is an $I \times J$ rank-one matrix with (i, j)-th element $\mathbf{a}(i)\mathbf{b}(j)$; i.e., $\mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{b}^T$.
- For three vectors, $\mathbf{a}$ $(I \times 1)$, $\mathbf{b}$ $(J \times 1)$, $\mathbf{c}$ $(K \times 1)$, $\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ is an $I \times J \times K$ rank-one three-way array with (i, j, k)-th element $\mathbf{a}(i)\mathbf{b}(j)\mathbf{c}(k)$.
- The *rank of a three-way array* $\underline{\mathbf{X}}$ is the smallest number of outer products needed to synthesize $\underline{\mathbf{X}}$.
- Rank – 1 Tensor $\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$.



Kronecker and Khatri-Rao products

$\otimes$ stands for the Kronecker product:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{B}\mathbf{A}(1,1), \mathbf{B}\mathbf{A}(1,2), \cdots \\ \mathbf{B}\mathbf{A}(2,1), \mathbf{B}\mathbf{A}(2,2), \cdots \\ \vdots \end{bmatrix}$$

$\odot$ stands for the Khatri-Rao (column-wise Kronecker) product: given $\mathbf{A}$ $(I \times F)$ and $\mathbf{B}$ $(J \times F)$, $\mathbf{A} \odot \mathbf{B}$ is the $JI \times F$ matrix

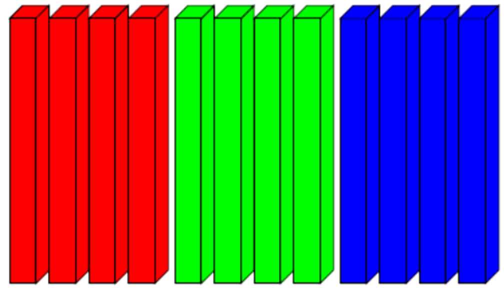
$$\mathbf{A} \odot \mathbf{B} = \left[\mathbf{A}(:,1) \otimes \mathbf{B}(:,1) \cdots \mathbf{A}(:,F) \otimes \mathbf{B}(:,F)\right]$$

$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$

If $\mathbf{D} = \text{diag}(\mathbf{d})$, then $\text{vec}(\mathbf{ADC}) = (\mathbf{C}^T \odot \mathbf{A})\mathbf{d}$

μ-mode matrix products

Consider 1-mode matricization $X^{(1)} \in \mathbb{R}^{n_1 \times (n_2 \cdots n_d)}$:



Seems to make sense to multiply an $m \times n_1$ matrix A from the left:

$$Y^{(1)} := A X^{(1)} \in \mathbb{R}^{m \times (n_2 \cdots n_d)}.$$

Can rearrange $Y^{(1)}$ back into an $m \times n_2 \times \cdots \times n_d$ tensor $\mathcal{Y}$.

This is called 1-mode matrix multiplication

$$\mathcal{Y} = A \circ_1 \mathcal{X} \qquad \Leftrightarrow \qquad Y^{(1)} = AX^{(1)}$$

More formally (and more ugly):

$$\mathcal{Y}_{i_1, i_2, \ldots, i_d} = \sum_{k=1}^{n_1} a_{i_1, k} \mathcal{X}_{k, i_2, \ldots, i_d}.$$

Rank decomposition for tensors

- Tensor:

$$\underline{\mathbf{X}} = \sum_{f=1}^{F} \mathbf{a}_f \circ \mathbf{b}_f \circ \mathbf{c}_f$$

- Scalar:

$$\underline{\mathbf{X}}(i,j,k) = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}, \quad \begin{array}{l} \forall i \in \{1,\cdots,I\} \\ \forall j \in \{1,\cdots,J\} \\ \forall k \in \{1,\cdots,K\} \end{array}$$

- Slabs:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T, \; k = 1,\cdots,K$$

- Matrix:

$$\mathbf{X}^{(KJ \times I)} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T$$

- Tall vector:

$$\mathbf{x}^{(KJI)} := vec\left(\mathbf{X}^{(KJ \times I)}\right) = (\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}))\,\mathbf{1}_{F \times 1} = (\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})\,\mathbf{1}_{F \times 1}$$

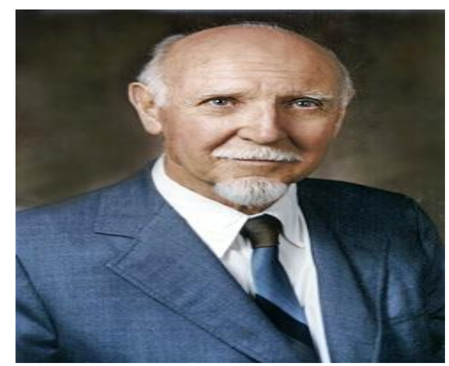
Tensor Products

The tensor product $\mathcal{A} \otimes \mathcal{B}$ between two tensors $\mathcal{A} \in \mathcal{S}_1 \otimes \mathcal{S}_2$ and $\mathcal{B} \in \mathcal{S}_3 \otimes \mathcal{S}_4$ is a tensor of $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3 \otimes \mathcal{S}_4$. The consequence is that the orders add up under tensor product.

Let $\mathcal{A}$ be represented by a three-way array $\mathcal{A} = [A_{ijk}]$ and $\mathcal{B}$ by a four-way array $\mathcal{B} = [B_{\ell mnp}]$; then tensor $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ is represented by the seven-way array of components $C_{ijk\ell mnp} = A_{ijk} B_{\ell mnp}$. With some abuse of notation, the tensor product is often applied to arrays of coordinates, so that notation $\boldsymbol{C} = \boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}$ may be encountered.

Tensor Decompositions-Historical Background

- **Founding fathers**:
  - ➢ Frank L. Hitchkock, in 1927 [1]
  - ➢ Raymond b. Katell, in 1944 [2]



- Regained interest due to:
  - ➢ Ledyard Tucker, in 1966

Tucker Decomposition

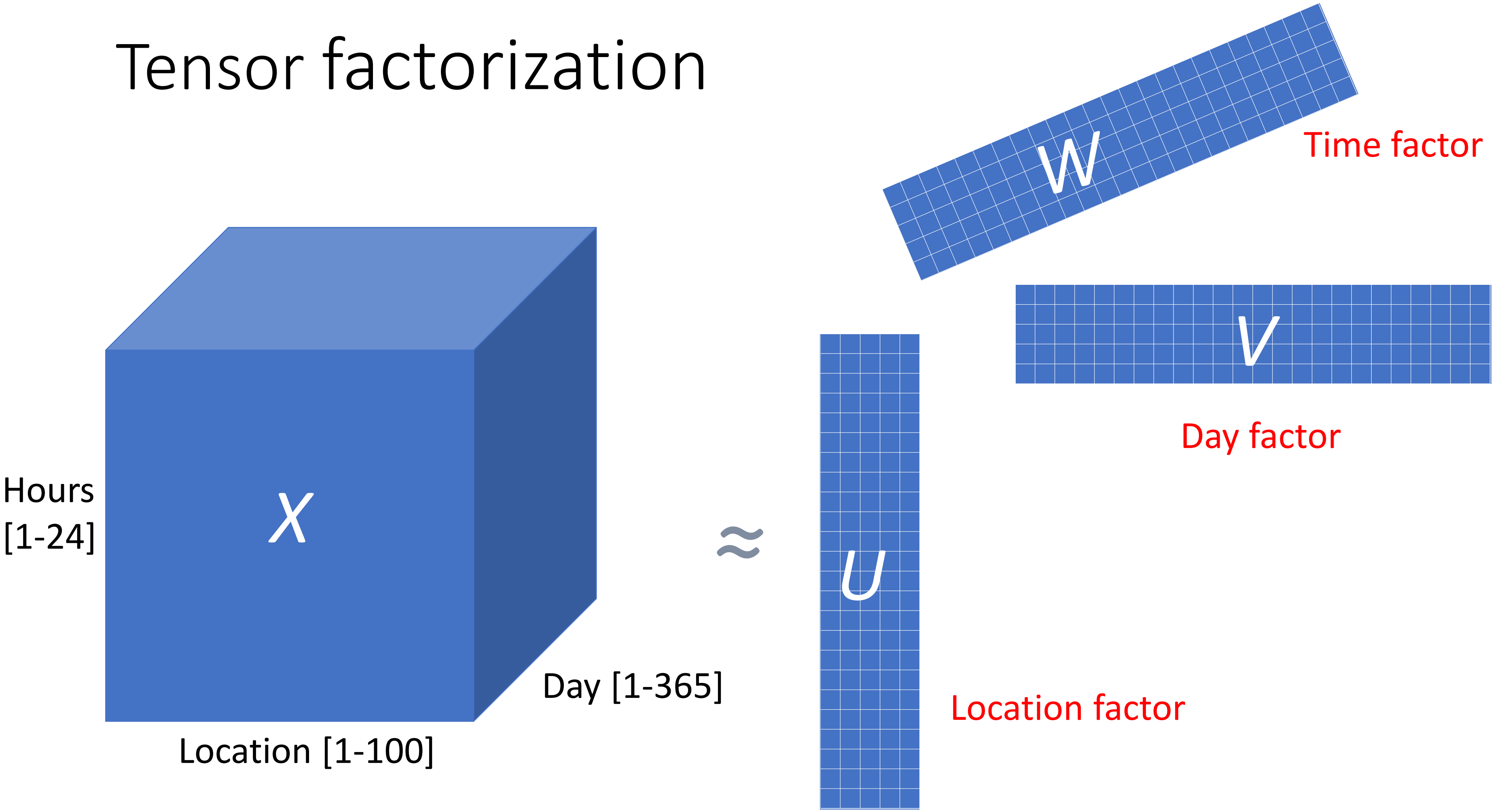
  - ➢ J. Douglas Caroll, in 1970
  - ➢ Richard A. Harshman, in 1970

PARAFAC/CANDECOMP

- First results in:
  - ➢ Psychometrics (Caroll, Harshman)
  - ➢ Chemometrics (Appelof, Davidson, R. Bro)

Slides by Michalis Giannopoulos

Tensor factorization



$$X \approx U \otimes V \otimes W$$

$$X_{i,j,k} \approx \sum_{r=1}^{\text{Rank}} U_{i,r} V_{j,r} W_{k,r}$$

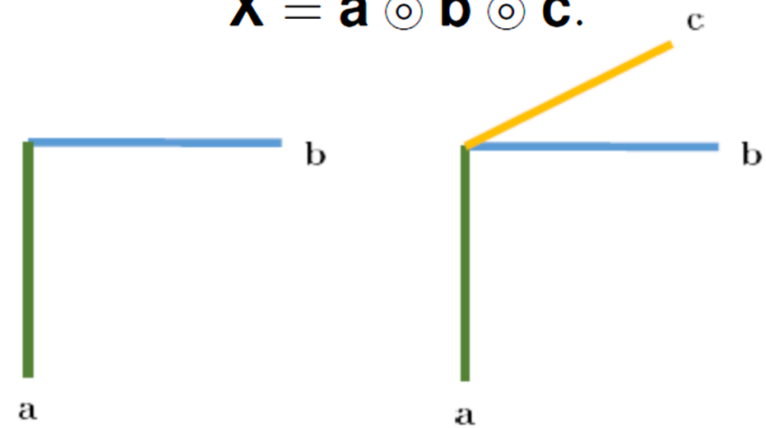
Rank-1 matrices and tensors

A **rank-1 matrix** $\mathbf{X}$ of size $I \times J$ is an outer product of two vectors: $\mathbf{X}(i,j) = \mathbf{a}(i)\mathbf{b}(j), \forall i \in \{1, \cdots, I\}, j \in \{1, \cdots, J\}$; i.e.,

$$\mathbf{X} = \mathbf{a} \circledcirc \mathbf{b}.$$

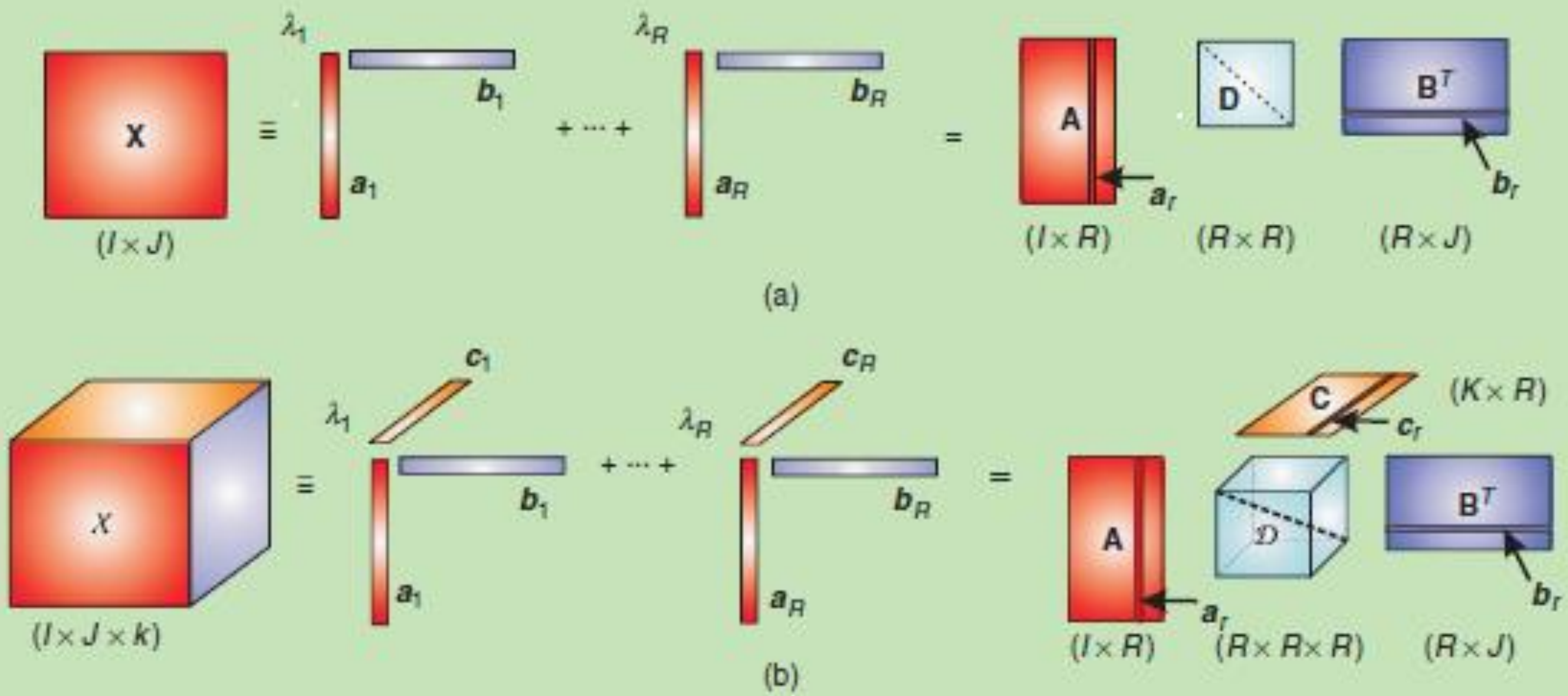
A **rank-1 third-order tensor** $\mathbf{X}$ of size $I \times J \times K$ is an outer product of three vectors: $\mathbf{X}(i,j,k) = \mathbf{a}(i)\mathbf{b}(j)\mathbf{c}(k)$; i.e.,

$$\mathbf{X} = \mathbf{a} \circledcirc \mathbf{b} \circledcirc \mathbf{c}.$$



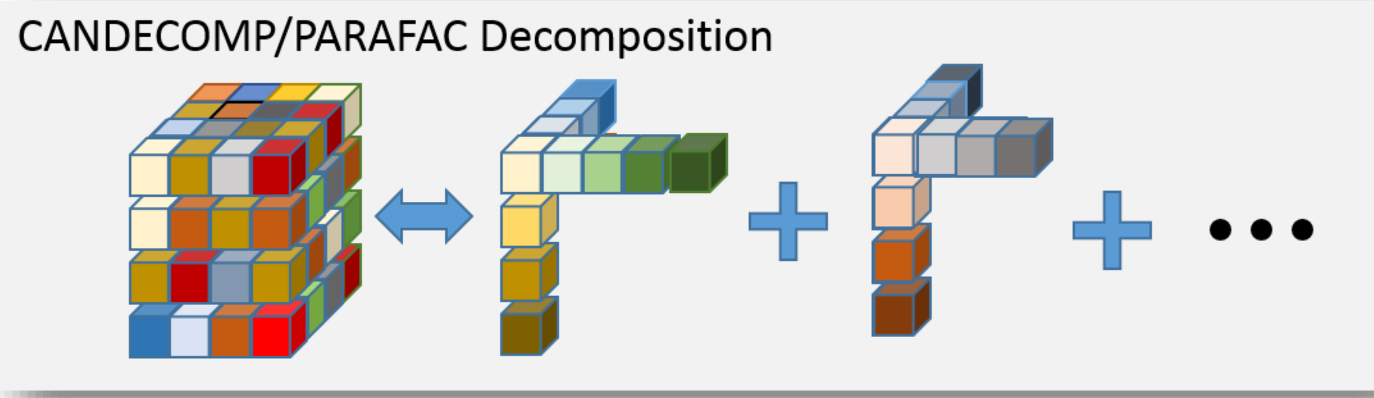
Extension of SVD

The analogy between dyadic and polyadic decompositions



CANDECOMP/PARAFAC

- Rank 1 Tensor models



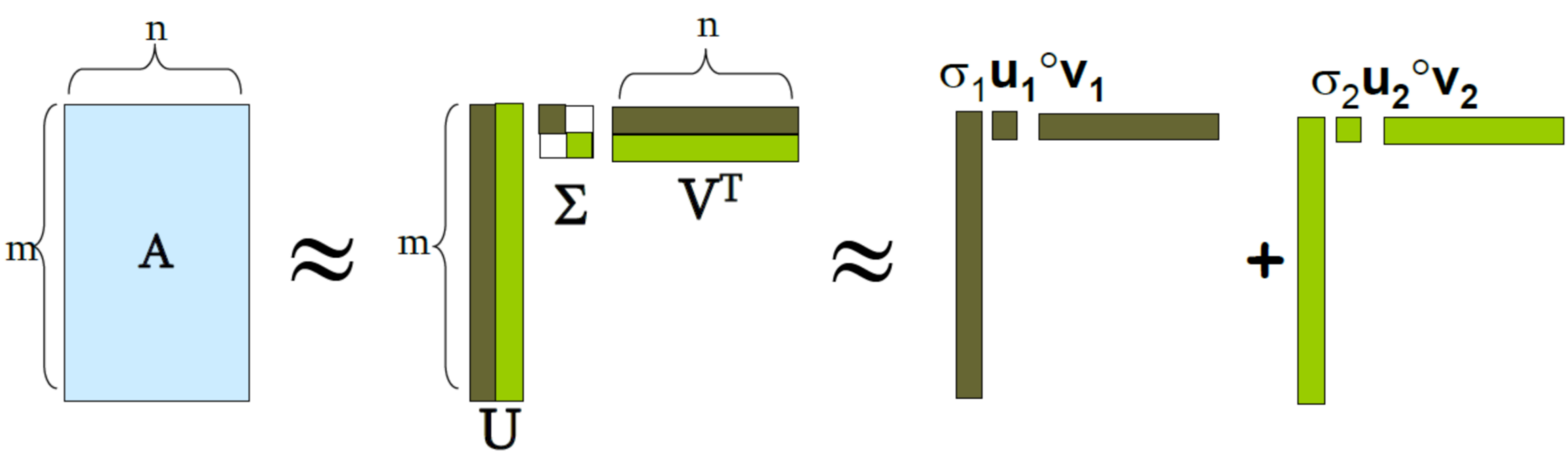
- CP factorization:

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \boldsymbol{b}_r^{(1)} \circ \boldsymbol{b}_r^{(2)} \circ \dots \circ \boldsymbol{b}_r^{(N)}.$$

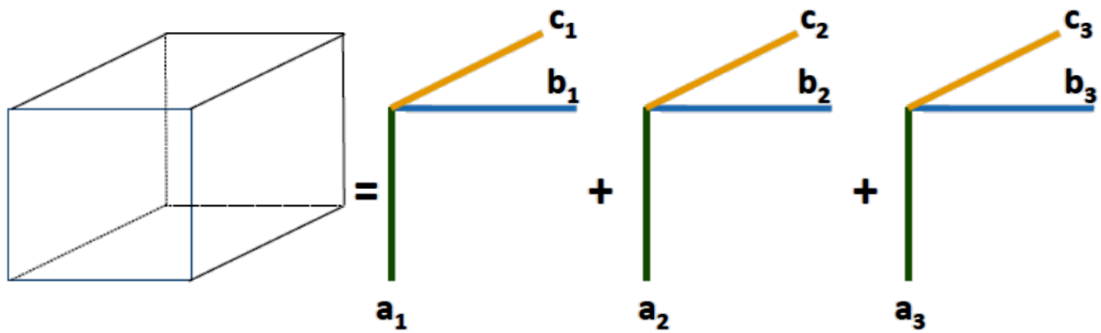
$$\mathcal{X} = \mathcal{D} \times_1 \mathbf{B}^{(1)} \times_2 \mathbf{B}^{(2)} \cdots \times_N \mathbf{B}^{(N)} = [\![\mathcal{D}; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)}]\!],$$

Reminder: SVD

$$\mathbf{A} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



Uniqueness

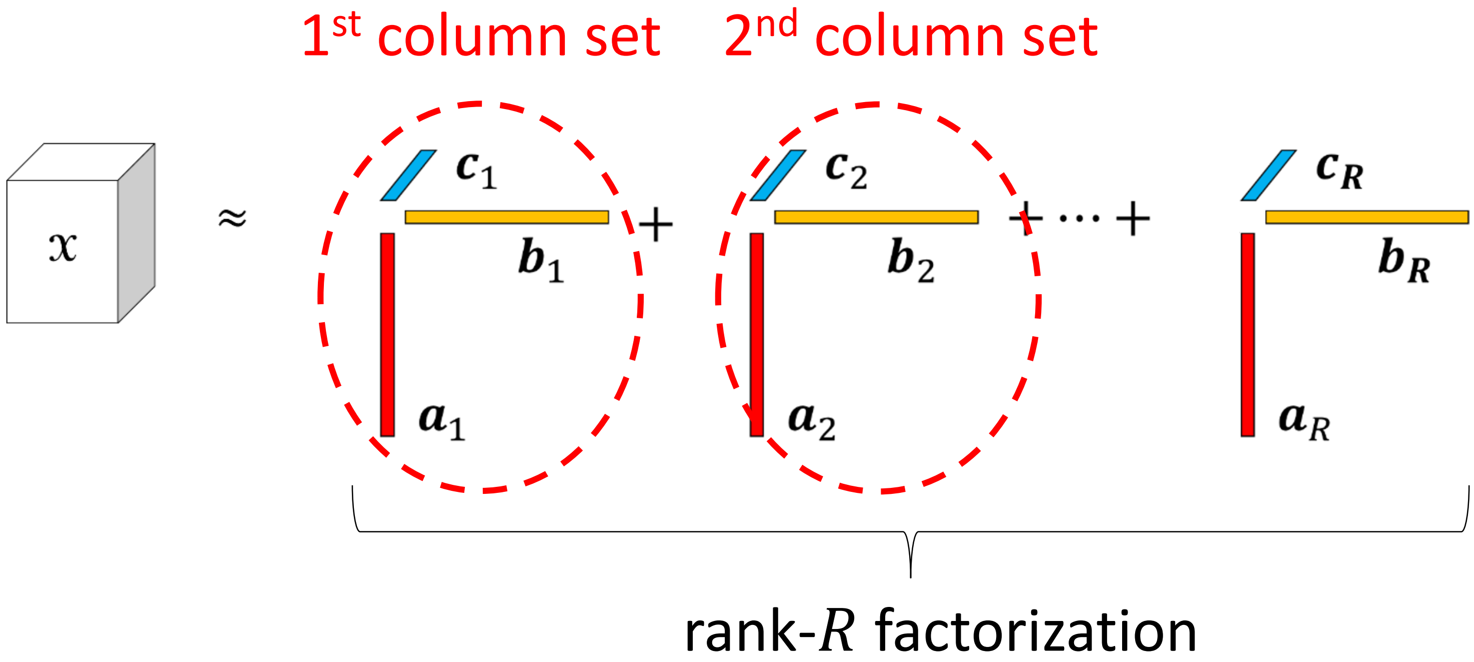


Given tensor **X** of rank F, its CPD is *essentially unique* iff the F rank-1 terms in its decomposition (the outer products or "chicken feet") are unique;

i.e., there is no other way to decompose **X** for the given number of terms.

Can of course permute "chicken feet" without changing their sum $\rightarrow$ permutation ambiguity.

Low rank Tensor Approximation



Low rank Tensor Approximation

Adopting a least squares criterion, the problem is

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} ||\mathbf{X} - \sum_{f=1}^{F} \mathbf{a}_f \circledcirc \mathbf{b}_f \circledcirc \mathbf{c}_f||_F^2,$$

Equivalently, we may consider

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} ||\mathbf{X}_1 - (\mathbf{C} \odot \mathbf{B})\mathbf{A}^T||_F^2.$$

Alternating optimization:

$$\mathbf{A} \leftarrow \arg\min_{\mathbf{A}} ||\mathbf{X}_1 - (\mathbf{C} \odot \mathbf{B})\mathbf{A}^T||_F^2,$$

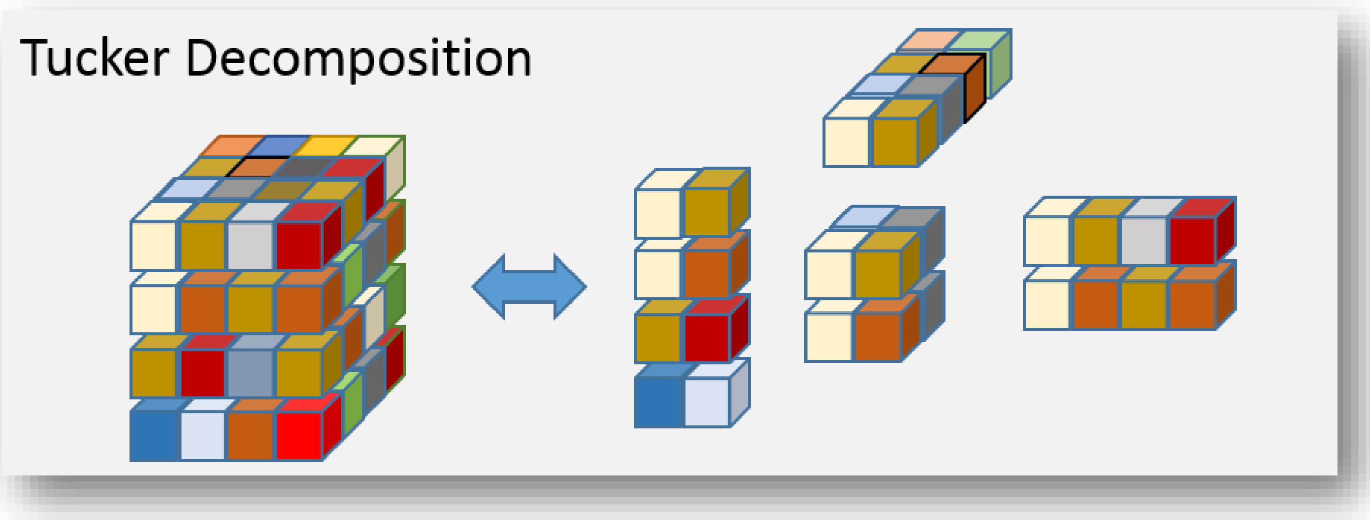
$$\mathbf{B} \leftarrow \arg\min_{\mathbf{B}} ||\mathbf{X}_2 - (\mathbf{C} \odot \mathbf{A})\mathbf{B}^T||_F^2,$$

$$\mathbf{C} \leftarrow \arg\min_{\mathbf{C}} ||\mathbf{X}_3 - (\mathbf{B} \odot \mathbf{A})\mathbf{C}^T||_F^2,$$

The above is widely known as **Alternating Least Squares (ALS)**.

TUCKER

- Tucker(3) factorization $\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} = [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]$

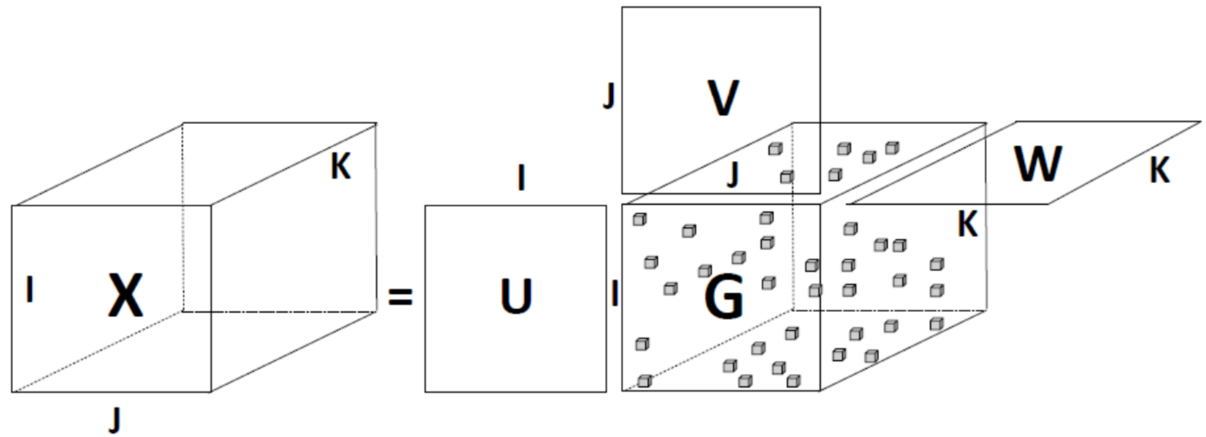


- The associated model-fitting problem is

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G}} ||\mathbf{X} - (\mathbf{B} \otimes \mathbf{A})\mathbf{G}\mathbf{C}^T||_F^2,$$

which is usually solved using an alternating least squares procedure.

Tucker and Multilinear SVD (MLSVD)



- Note that each column of **U** interacts with every column of **V** and every column of **W** in this decomposition.
- The strength of this interaction is encoded in the corresponding element of **G**.
- Different from CPD, which only allows interactions between corresponding columns of **A**, **B**, **C**, i.e., the only outer products that can appear in the CPD are of type $\mathbf{a}_f \circledcirc \mathbf{b}_f \circledcirc \mathbf{c}_f$.
- The *Tucker model* in (14) also allows "mixed" products of non-corresponding columns of **U**, **V**, **W**.

The n-Rank

- $R_n = rank_n(\mathscr{X})$[1], [2]: The dimension of the vector space which is spanned by the mode-n fibers of column rank of $\mathscr{X}$

- Rank-$(R_1, R_2, \cdots R_N)$ tensor → R_n: Column-rank of the mode-n unfolding $\boldsymbol{X}_{(n)}$

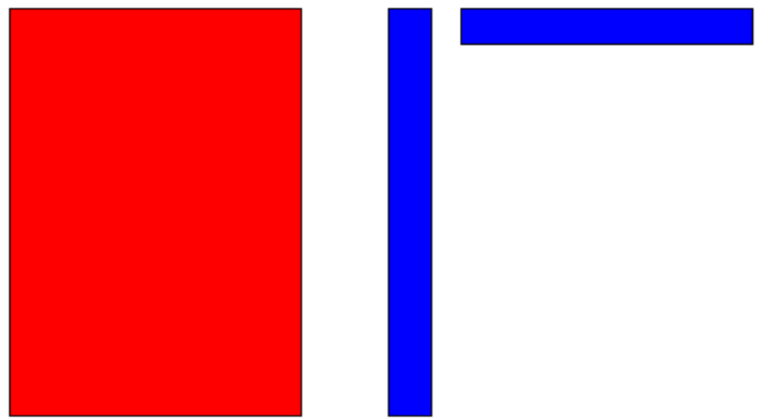
- **<u>Usefulness</u>**: Tensor approximation → Compression
  - For ≥ 1 dimensions: $R_n < rank_n(\mathscr{X})$ → **<u>Trimmed version of original tensor</u>**

- <u>Lack of Uniqueness</u>:
  - “Transform” the core tensor $\mathscr{G}$
  - Apply the inverse ”transform” to the factor matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$
  - **Sometimes desired**: Sketching arithmetic solutions for Tucker decomposition computation

Low rank approximation

Setting: Matrix $X \in \mathbb{R}^{n \times m}$, m and n too large to compute/store X explicitly.

Idea: Replace X by RS^T with $R \in \mathbb{R}^{n \times r}$, $S \in \mathbb{R}^{m \times r}$ and $r \ll m, n$.



	X	RS^T
Memory	nm	$nr + rm$
Cost	$\text{ops}(m, n)$	$\text{ops}(m, n) \times \frac{r}{\min\{m,n\}}$ (?)

Construction from SVD

SVD: Let matrix $X \in \mathbb{R}^{n\times m}$ and $k = \min\{m, n\}$. Then $\exists$ orthonormal matrices

$$U = [u_1,\ u_2,\ \ldots,\ u_k] \in \mathbb{R}^{n\times k}, \quad V = [v_1,\ v_2,\ \ldots,\ v_k] \in \mathbb{R}^{m\times k},$$

such that

$$X = U\Sigma V^T, \qquad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k).$$

Choose $r \leq k$ and partition

$$X = [U_1,\ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} [V_1,\ V_2]^T = \underbrace{U_1\,\Sigma_1}_{=:R}\ \underbrace{V_1^T}_{=:S^T} + U_2\Sigma_2 V_2^T.$$

Then $\|X - RS^T\|_2 = \|\Sigma_2\|_2 = \sigma_{r+1}$.

Good low rank approximation if singular values decay sufficiently fast.

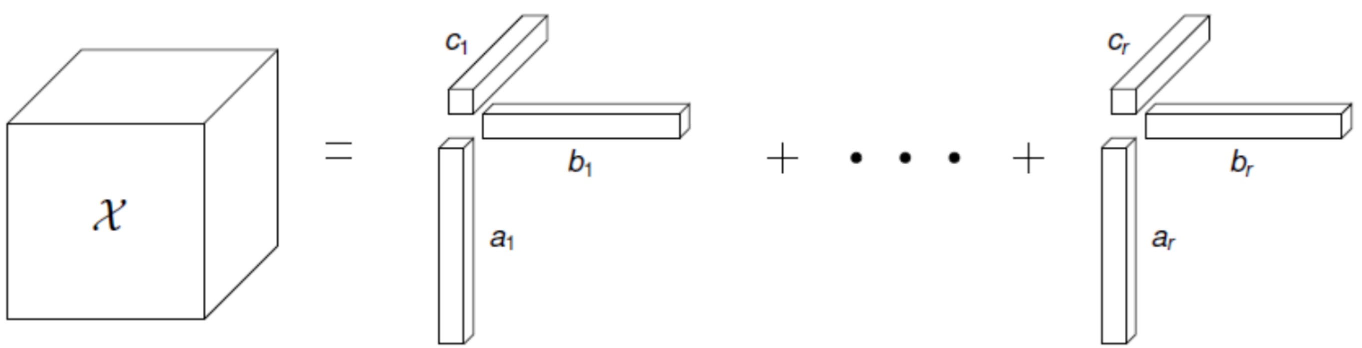
CP decomposition

Canonical Polyadic decomposition of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ defined via

$$\mathrm{vec}(\mathcal{X}) = c_1 \otimes b_1 \otimes a_1 + c_2 \otimes b_2 \otimes a_2 + \cdots + c_R \otimes b_R \otimes a_R$$

for vectors $a_j \in \mathbb{R}^{n_1}$, $b_j \in \mathbb{R}^{n_2}$, $c_j \in \mathbb{R}^{n_3}$.

Tensor rank of X = minimal possible R



CP decomposition

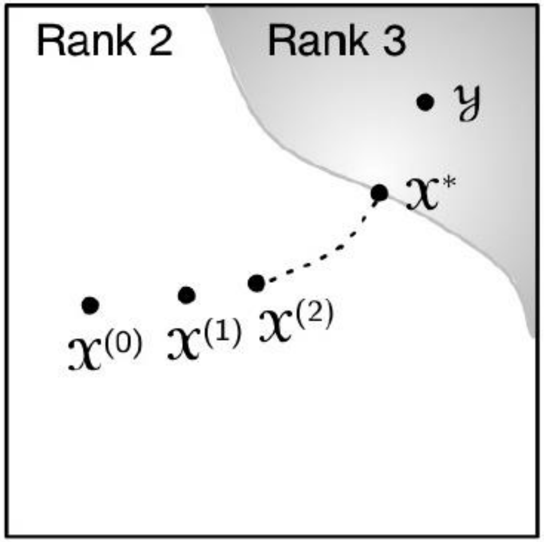
- For **matrices**:
  - rank r is upper semi-continuous $\rightsquigarrow$ closedness property: sequence of rank$=r$ matrices can only converge to rank$\leq r$ matrix.
  - best low-rank approximation possible by successive rank-1 approximations.
  - Robust black-box algorithms/software available (`svd`, Lanczos).

For tensors of order $d \geq 3$:

- tensor rank R is **not** upper semi-continuous $\rightsquigarrow$

  lack of closedness

- successive rank-1 approximations fail
- all algorithms based on optimization techniques (ALS, Gauss-Newton)



Tucker decomposition

- SVD: $\mathrm{vec}(X) = (V \otimes U) \cdot \mathrm{vec}(\Sigma).$

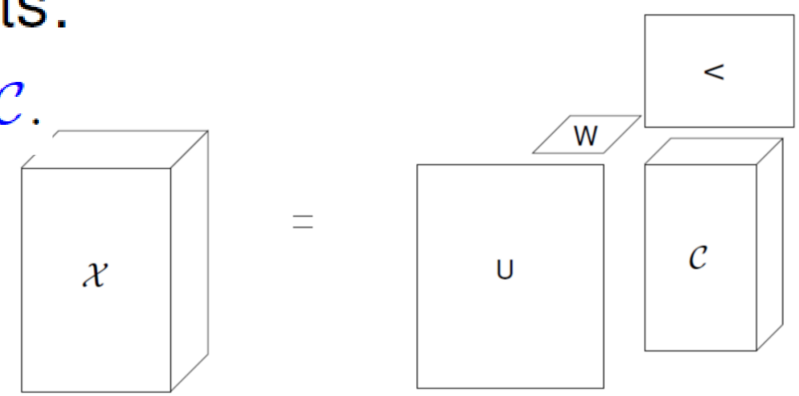
Tucker decomposition of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ defined via

$$\mathrm{vec}(\mathcal{X}) = (W \otimes V \otimes U) \cdot \mathrm{vec}(\mathcal{C})$$

with $U \in \mathbb{R}^{n_1 \times r_1}$, $V \in \mathbb{R}^{n_2 \times r_2}$, $W \in \mathbb{R}^{n_3 \times r_3}$,
and core tensor $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.

In terms of μ-mode matrix products:

$$\mathcal{X} = U \circ_1 V \circ_2 W \circ_3 \mathcal{C} =: (U, V, W) \circ \mathcal{C}.$$



Tucker decomposition

Consider all three matricizations:

$$\begin{aligned} X^{(1)} &= U \cdot C^{(1)} \cdot (W \otimes V)^T, \\ X^{(2)} &= V \cdot C^{(2)} \cdot (W \otimes U)^T, \\ X^{(3)} &= W \cdot C^{(3)} \cdot (V \otimes U)^T. \end{aligned}$$

These are low rank decompositions $\rightsquigarrow$

$$\text{rank}(X^{(1)}) \leq r_1, \quad \text{rank}(X^{(2)}) \leq r_2, \quad \text{rank}(X^{(3)}) \leq r_3.$$

Multilinear rank of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ defined by tuple

$$(r_1, r_2, r_3), \quad \text{with} \quad r_i = \text{rank}(X^{(i)}).$$

Higher-order SVD (HOSVD)

Goal: Approximate given tensor $\mathcal{X}$ by Tucker decomposition with prescribed multilinear rank (r_1, r_2, r_3).

1. Calculate SVD of matricizations:

$$X^{(\mu)} = \widetilde{U}_\mu \widetilde{\Sigma}_\mu \widetilde{V}_\mu^T \quad \text{for } \mu = 1, 2, 3.$$

2. Truncate basis matrices:

$$U_\mu := \widetilde{U}_\mu(:, 1 : r_\mu) \quad \text{for } \mu = 1, 2, 3.$$

3. Form core tensor:

$$\text{vec}(\mathcal{C}) := \left(U_3^T \otimes U_2^T \otimes U_1^T\right) \cdot \text{vec}(\mathcal{X}).$$

Truncated tensor produced by HOSVD [Lathauwer/De Moor/Vandewalle'2000]:

$$\text{vec}(\widetilde{\mathcal{X}}) := \left(U_3 \otimes U_2 \otimes U_1\right) \cdot \text{vec}(\mathcal{C}).$$

Higher-order SVD (HOSVD)

Tensor $\widetilde{\mathcal{X}}$ resulting from HOSVD satisfies quasi-optimality condition

$$\|\mathcal{X} - \widetilde{\mathcal{X}}\| \leq \sqrt{d}\|\mathcal{X} - \mathcal{X}_{\text{best}}\|,$$

where $\mathcal{X}_{\text{best}}$ is best approximation of $\mathcal{X}$ with multilinear ranks $(r_1, \ldots, r_d)$.

ALGORITHM 2: Higher-Order Singular Value Decomposition (HOSVD)

Input: $N-$mode tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and ranks $R_1, \ldots, R_N$.
Output: Tucker factors $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times R_1}, \cdots \mathbf{U}_N \in \mathbb{R}^{I_N \times R_N}$ and core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times \cdots \times R_N}$

1: **for** $n = 1 \cdots N$ **do**
2: $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] \leftarrow \text{SVD}(\mathbf{X}_{(n)})$
3: $\mathbf{U}_n \leftarrow \mathbf{U}(:, 1 : R_n)$, i.e., set $\mathbf{U}_n$ equal to the R_n left singular vectors of $\mathbf{X}_{(n)}$
4: **end for**
5: $\mathcal{G} \leftarrow \mathcal{X} \times_N \mathbf{U}_N^T \times_{N-1} \mathbf{U}_{N-1}^T \cdots \times_1 \mathbf{U}_1^T$

Take home messages

- CP

+ Exploratory model
+ Unique (under mild conditions)
+ Easy to interpret

- Hard to determine appropriate rank
- Global minimum

➢ Extract latent factors for interpretation
➢ Exploratory clustering

Take home messages

- TUCKER

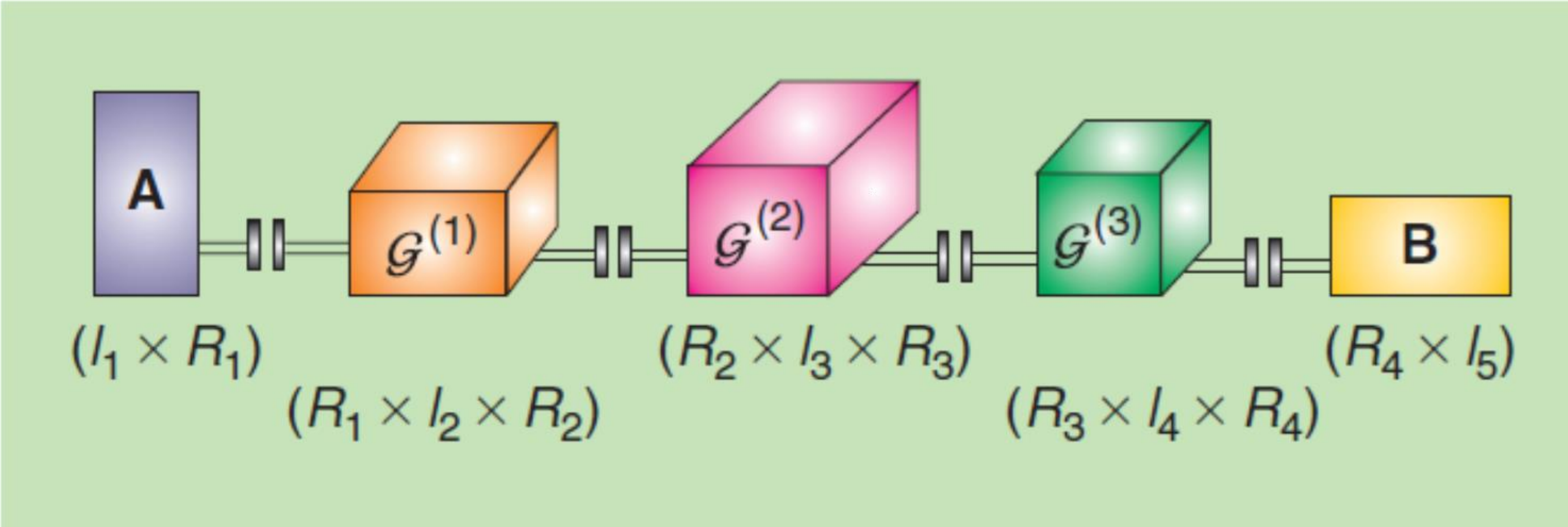
+ Non-trilinear interactions

+ Optimal tensor compression

- Non-unique
- Hard to interpret

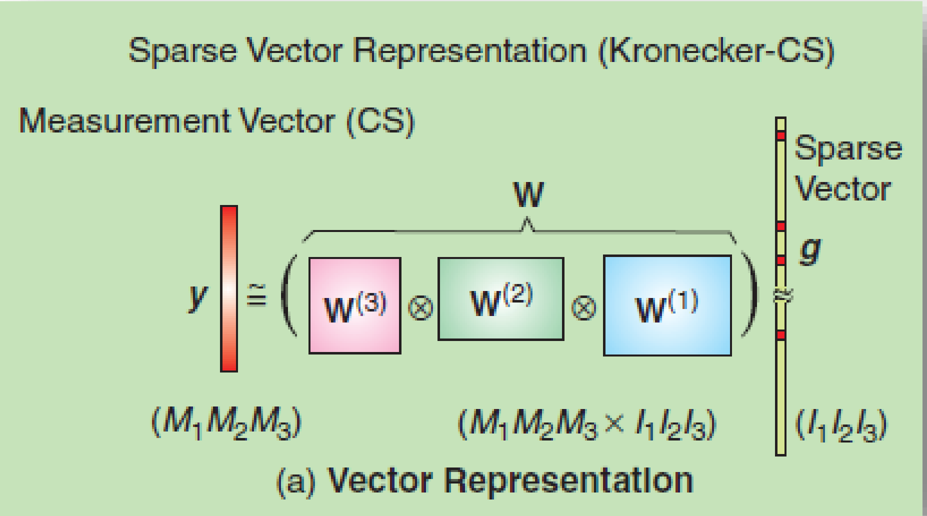
➢ Tensor compression

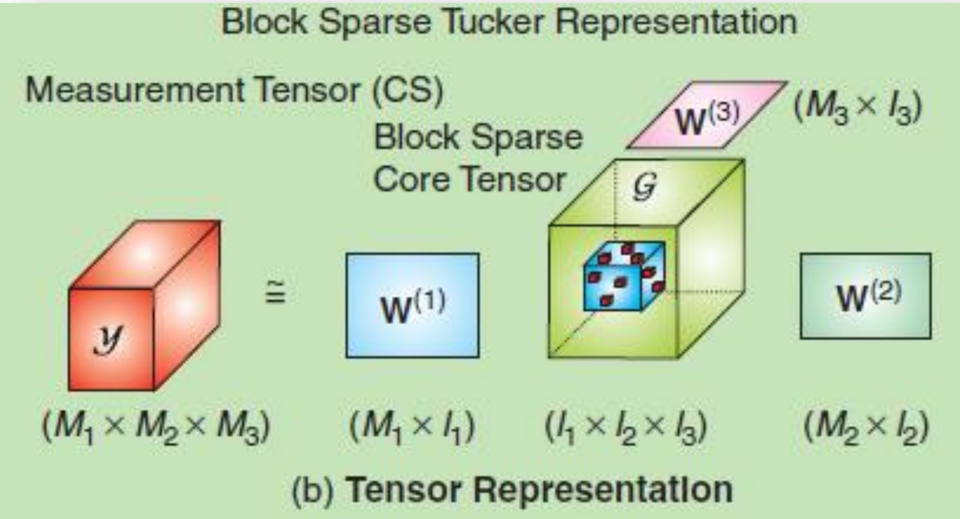
Tensor Trains



[FIG9] The TT decomposition of a fifth-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_5}$, consisting of two matrix carriages and three third-order tensor carriages. The five carriages are connected through tensor contractions, which can be expressed in a scalar form as $x_{i_1,i_2,i_3,i_4,i_5} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_5=1}^{R_5} a_{i_1,r_1} g^{(1)}_{r_1,i_2,r_2} g^{(2)}_{r_2,i_3,r_3} g^{(3)}_{r_3,i_4,r_5} b_{r_4,i_5}$.

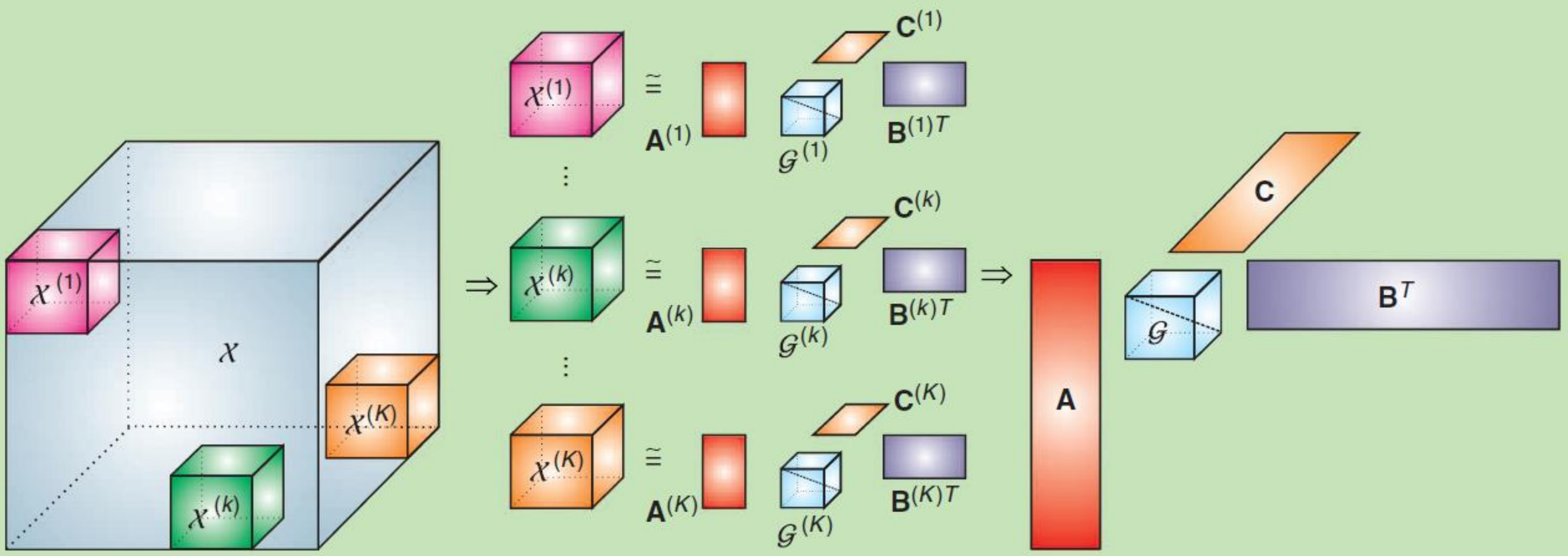
Higher-order CS with Tensors





$$\mathcal{Y} \cong \mathcal{G} \times_1 \mathbf{W}^{(1)} \times_2 \mathbf{W}^{(2)} \cdots \times_N \mathbf{W}^{(N)}, \quad \text{with } \|\mathcal{G}\|_0 \leq K,$$

Parallel decomposition



Sparse Non-Negative Tensor Factorization

Sparse encoding

$$\min_{U,V,W} \|X - U \otimes V \otimes W\|_F + \lambda_U \|U\|_1 + \lambda_V \|V\|_1 + \lambda_W \|W\|_1$$

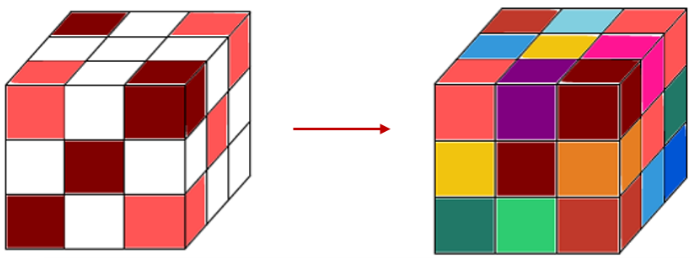
Non-negativity:

$$U \geq 0$$
$$V \geq 0$$
$$W \geq 0$$

More interpretable results

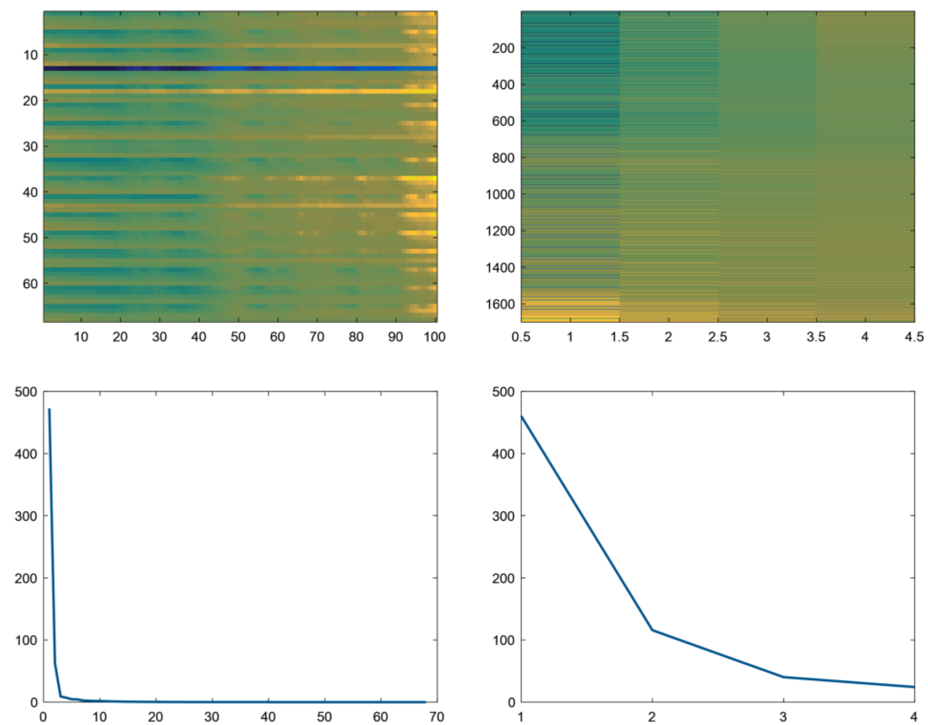
Tensor Completion

- Low rank Tensor/Matrices



$$\begin{aligned} \underset{\mathcal{X}}{\text{minimize}} \quad & \text{rank}_*(\mathcal{X}) \\ \text{subject to} \quad & \mathcal{X}_\Omega = \mathcal{T}_\Omega, \end{aligned}$$

$$\begin{aligned} \underset{\mathcal{X}}{\text{minimize}} \quad & \text{rank}_*(\mathcal{X}) \\ \text{subject to} \quad & \mathcal{X}_\Omega = \mathcal{T}_\Omega + \mathcal{E}_\Omega, \end{aligned}$$



Low-rank Tensor completion

- CP based approach

$$\underset{\mathcal{X}}{\text{minimize}} \quad \mathcal{D}(\mathcal{T}_\Omega, \mathcal{X}_\Omega)$$
$$\text{subject to} \quad \text{rank}_{\text{CP}}(\mathcal{X}) = r,$$

$$\mathcal{D}(\mathcal{T}_\Omega, \mathcal{X}_\Omega) = \|\mathcal{X}_\Omega - \mathcal{T}_\Omega\|_F$$

- N-rank approach

$$\underset{\mathcal{X}}{\text{minimize}} \quad f(\text{rank}_{\text{tc}}(\mathcal{X}))$$
$$\text{subject to} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega,$$

$$\text{rank}_{\text{tc}}(\mathcal{X}) = (\text{rank}(\mathbf{X}_{(1)}), \ldots, \text{rank}(\mathbf{X}_{(N)})).$$

$$f(\text{rank}_{\text{tc}}(\mathcal{X})) = \sum_{i=1}^{N} \text{rank}(\mathbf{X}_{(i)}).$$

Decomposition based approaches

PARAFAC/TUCKER

- Assume $$\mathcal{X} = \mathcal{P}_{\Omega}(\mathcal{T}) + \mathcal{P}_{\Omega^c}(\hat{\mathcal{X}}) = \mathcal{W} * \mathcal{T} + (1 - \mathcal{W}) * \hat{\mathcal{X}},$$

$$\mathcal{W}(i_1, i_2, \dots, i_N) = \begin{cases} 1 & \text{if } \mathcal{T}(i_1, i_2, \dots, i_N) \text{ is observed.} \\ 0 & \text{if } \mathcal{T}(i_1, i_2, \dots, i_N) \text{ is unobserved.} \end{cases}$$

Tensor Completion via Parallel Matrix Factorization

- Generalization of MC problem:

$$\begin{aligned} &\underset{\mathscr{X}}{\text{minimize}} \quad \|\mathscr{X}\|_* \\ &\text{subject to} \quad \mathscr{A}(\mathscr{X}_{i_1 i_2 i_3}) = \mathscr{A}(\mathscr{T}_{i_1 i_2 i_3}), \quad \forall (i_1 i_2 i_3) \in \Omega \end{aligned}$$

- Sampling operator: $\mathscr{A}(\mathscr{T}) = \begin{cases} \tau_{i_1 i_2 i_3}, & \text{if } (i_1 i_2 i_3) \in \Omega \\ 0, & \text{otherwise} \end{cases}$

- Tensor Nuclear Norm Definition [1]: $\|\mathscr{X}\|_* = \sum_{i=1}^n \alpha_i \|\mathbf{X}_{(i)}\|_*$, where $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$

- **<u>Problem reformulation</u>**:

$$\begin{aligned} &\underset{\mathscr{X}}{\text{minimize}} \quad \sum_{i=1}^{n} \alpha_i \|\mathbf{X}_{(i)}\|_* \\ &\text{subject to} \quad \mathscr{A}(\mathscr{X}_{i_1 i_2 i_3}) = \mathscr{A}(\mathscr{T}_{i_1 i_2 i_3}), \quad \forall (i_1 i_2 i_3) \in \Omega \end{aligned}$$

Tensor Completion via Parallel Matrix Factorization

1.2. **Problem formulation.** We aim at recovering an (approximately) low-rank tensor $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ from partial observations $\boldsymbol{\mathcal{B}} = \mathcal{P}_\Omega(\boldsymbol{\mathcal{M}})$, where Ω is the index set of observed entries, and $\mathcal{P}_\Omega$ keeps the entries in Ω and zeros out others. We apply low-rank matrix factorization to each mode unfolding of $\boldsymbol{\mathcal{M}}$ by finding matrices $\mathbf{X}_n \in \mathbb{R}^{I_n \times r_n}$, $\mathbf{Y}_n \in \mathbb{R}^{r_n \times \Pi_{j \neq n} I_j}$ such that $\mathbf{M}_{(n)} \approx \mathbf{X}_n \mathbf{Y}_n$ for $n = 1, \ldots, N$, where r_n is the estimated rank, either fixed or adaptively updated. Introducing one common variable $\boldsymbol{\mathcal{Z}}$ to relate these matrix factorizations, we solve the following model to recover $\boldsymbol{\mathcal{M}}$

$$\min_{\mathbf{X}, \mathbf{Y}, \boldsymbol{\mathcal{Z}}} \sum_{n=1}^{N} \frac{\alpha_n}{2} \|\mathbf{X}_n \mathbf{Y}_n - \mathbf{Z}_{(n)}\|_F^2, \text{ subject to } \mathcal{P}_\Omega(\boldsymbol{\mathcal{Z}}) = \boldsymbol{\mathcal{B}}, \quad (2)$$

where $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$. In the model, α_n, $n = 1, \ldots, N$, are weights and satisfy $\sum_n \alpha_n = 1$. The constraint $\mathcal{P}_\Omega(\boldsymbol{\mathcal{Z}}) = \boldsymbol{\mathcal{B}}$ enforces consis-

TC via Parallel Matrix Factorization

- Similar to the matrix case

$$\min_{\mathcal{Z}} \sum_{n=1}^{N} \alpha_n \|\mathbf{Z}_{(n)}\|_*, \text{ subject to } \mathcal{P}_\Omega(\mathcal{Z}) = \mathcal{B},$$

where $\alpha_n \geq 0, n = 1, \ldots, N$ are preselected weights satisfying $\sum_n \alpha_n = 1$.

- Tensor nuclear norm

$$\|\boldsymbol{X}\|_* = \max_{\|\boldsymbol{W}\|=1} \langle \boldsymbol{W}, \boldsymbol{X} \rangle$$

- Nuclear norm minimization

$$\min_{\boldsymbol{X} \in \mathbb{R}^{d_1 \times d_2 \times d_3}} \|\boldsymbol{X}\|_* \qquad \text{subject to } \mathcal{P}_\Omega \boldsymbol{X} = \mathcal{P}_\Omega \boldsymbol{T},$$

where $\mathcal{P}_\Omega : \mathbb{R}^{d_1 \times d_2 \times d_3} \mapsto \mathbb{R}^{d_1 \times d_2 \times d_3}$ such that

$$(\mathcal{P}_\Omega \boldsymbol{X})(i,j,k) = \begin{cases} \boldsymbol{X}(i,j,k) & \text{if } (i,j,k) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

Tensor Signal Analysis for WSN Data

Experimental data collected from a WSN operating at a pilot desalination plant, located at La Tordera, Spain [1]

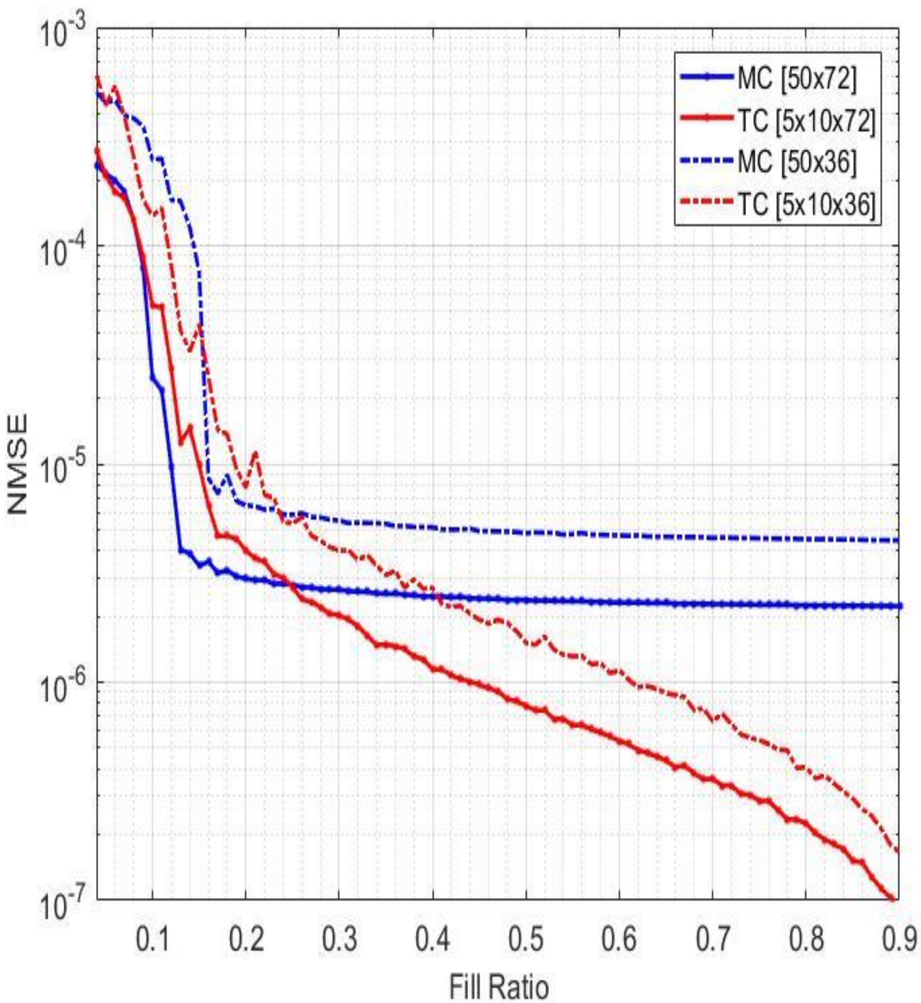
- Water impedance measurements (Ohms)
  - 5 sensors
  - 10 different channels/sensor
  - 3 day period → Sampling every 1 and 2 hours

Matrices: 50×72, 50×36

Tensors: $5 \times 10 \times 72$, $5 \times 10 \times 36$

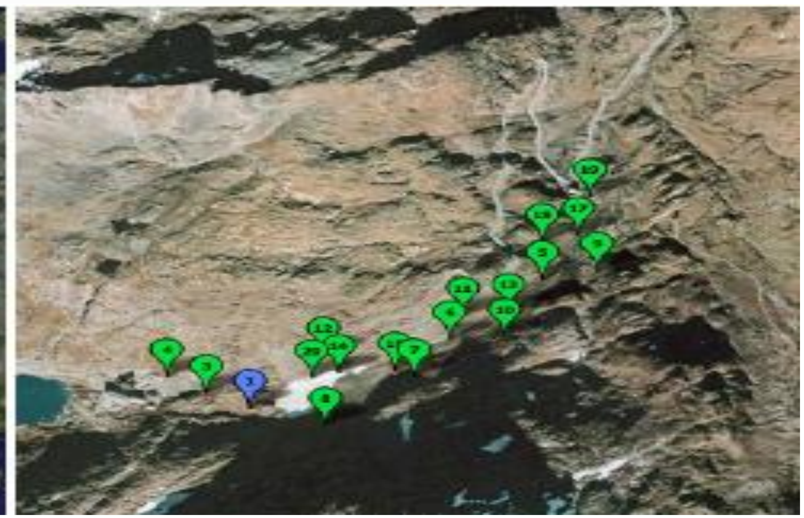
Effects of Data Structuring

- Higher fill-ratio
  - Better reconstruction quality quantified
- <u>Regardless matrix/tensor size</u>
  - TC outperforms MC from low fill-ratio regimes
- <u>NMSE convergence</u>
  - MC reaches a plateau
  - TC decreases (nearly) monotonically



WSN Outdoors Dataset

Experimental data collected from a WSN operating at a Grand-St-Bernard pass between Switzerland and Italy



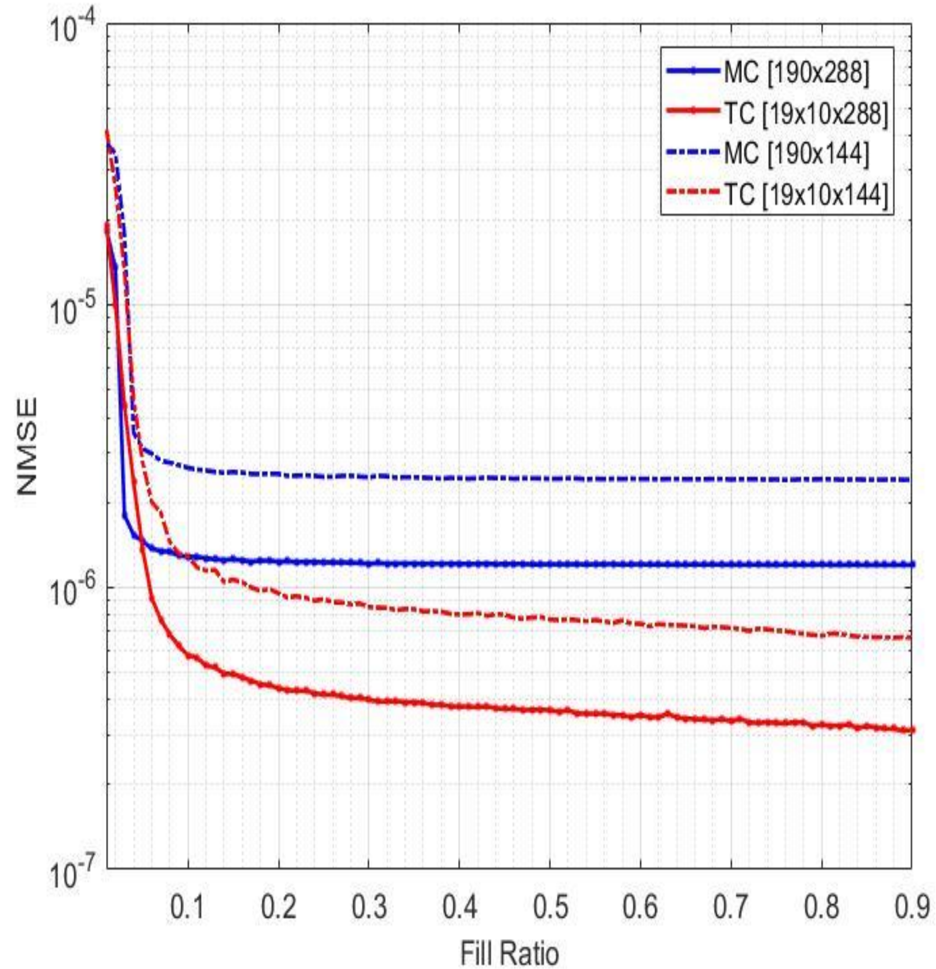
- Temperature measurements
  - 19 sensors
  - 10 day period
  - Sampling every 5 and 10 minutes

Matrices: 190×288, 190×144

Tensors: $19 \times 10 \times 288$, $19 \times 10 \times 144$

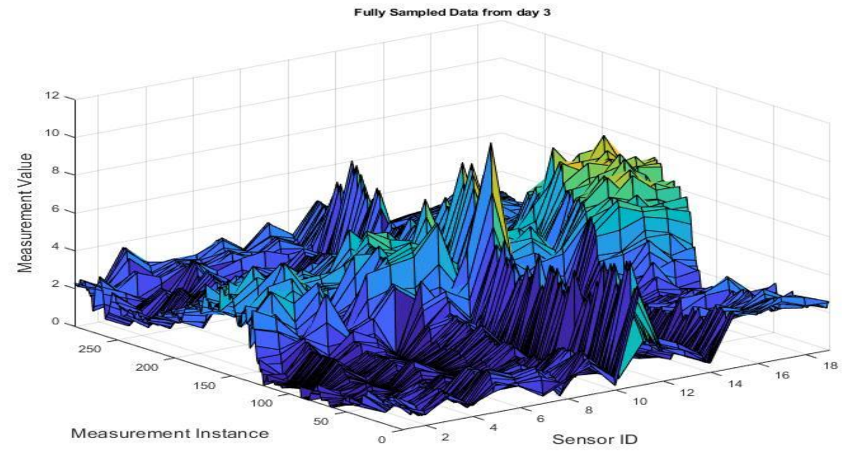
Effects of Data Structuring

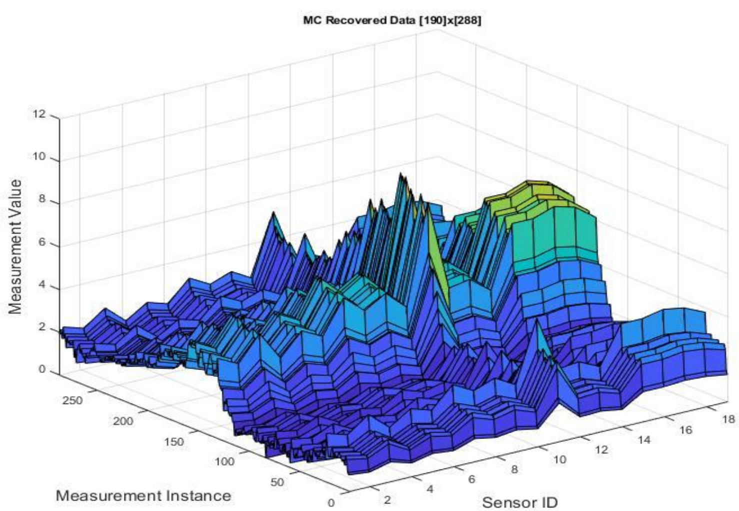
- Higher fill-ratio
  - ➢ Better reconstruction quality quantified
- <u>Larger Dataset</u>
  - ➢ TC outperforms MC <u>from lower</u> fill-ratio regimes
- <u>NMSE convergence</u>
  - ➢ MC reaches a plateau
  - ➢ TC keeps decreasing

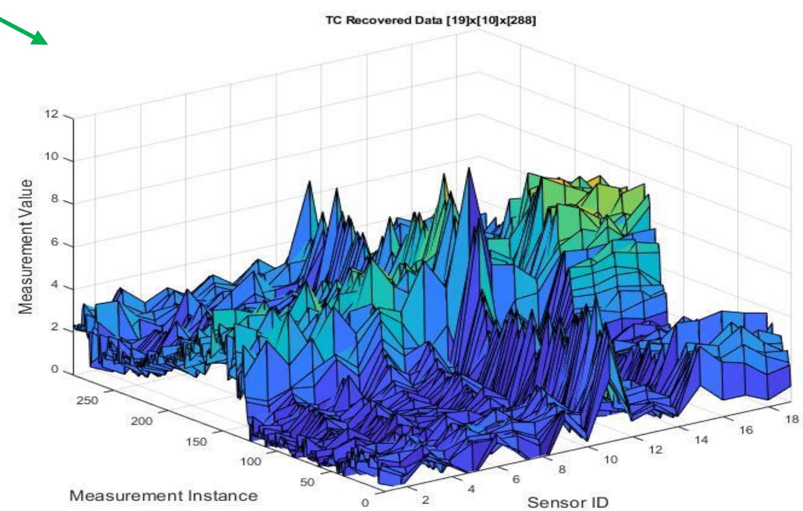


Effects of Fill-Ratio

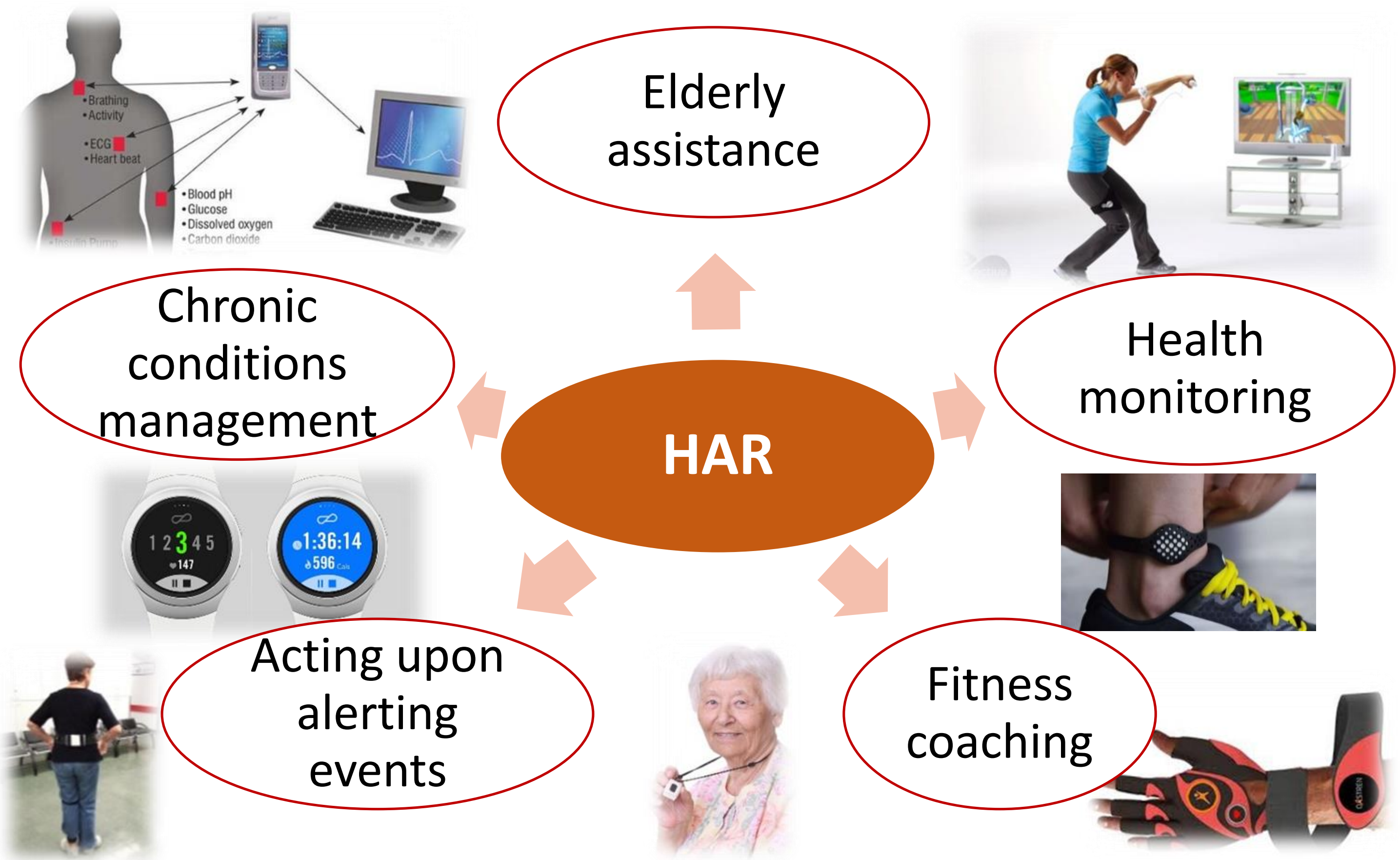
- Data sampled from single day
  - GT data
  - MC reconstructed data
  - TC reconstructed data
- $f = 0.2$



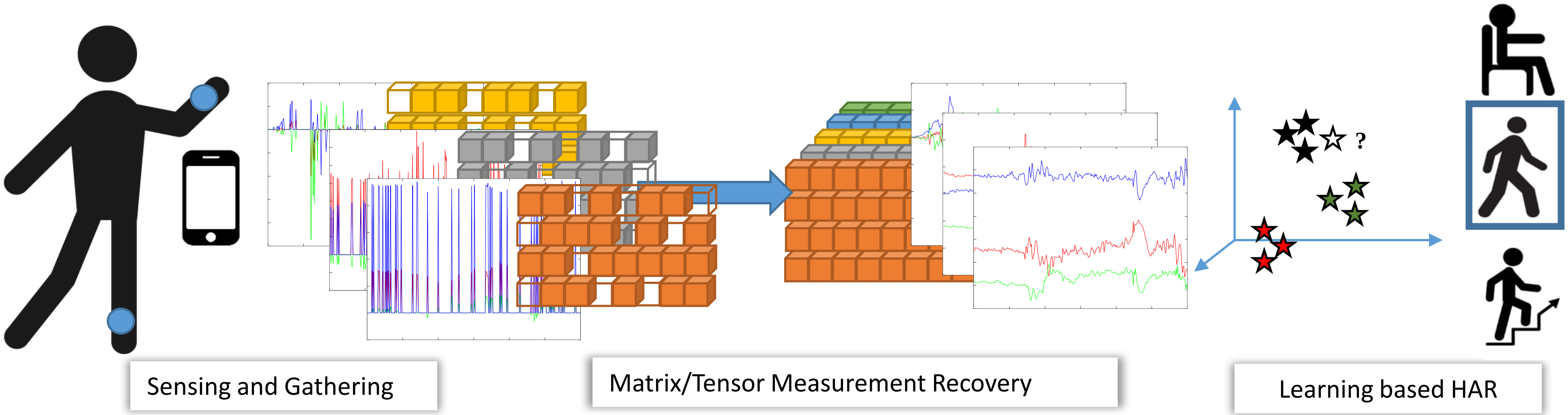


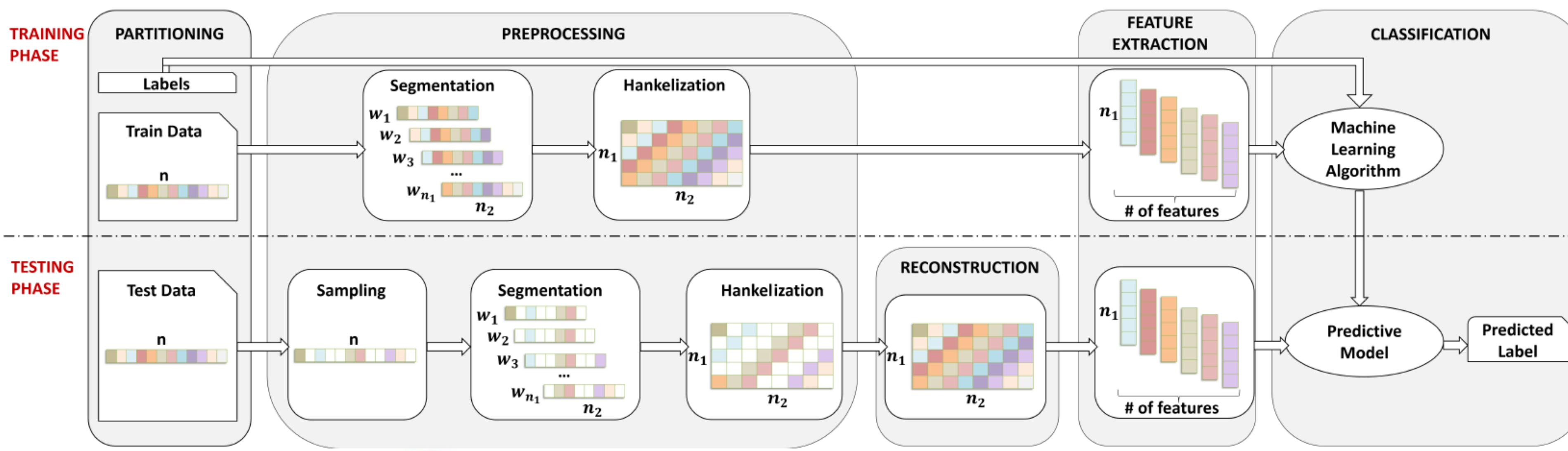


WSNs for Human Activity Recognition



Problem formulation

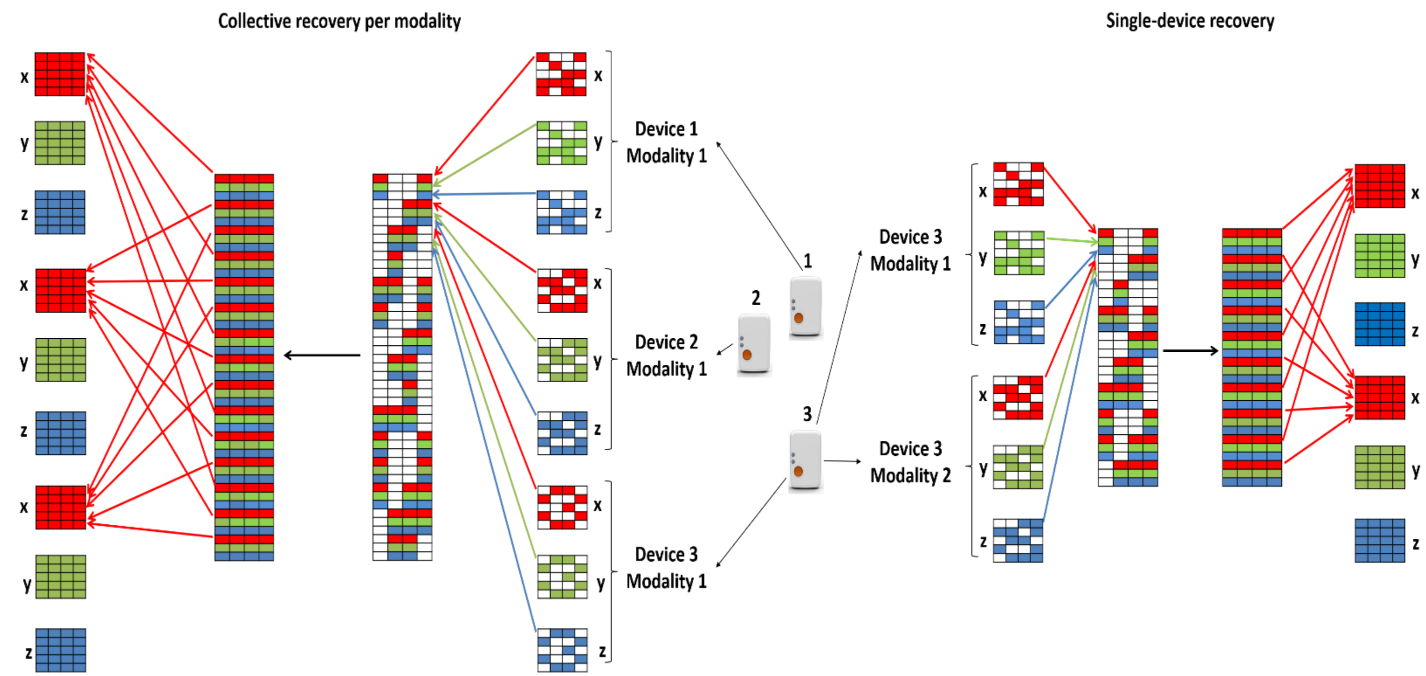




Single-device vs collective recovery: matrices

Scenario 2
Collective per modality

Scenario 1
Single-device

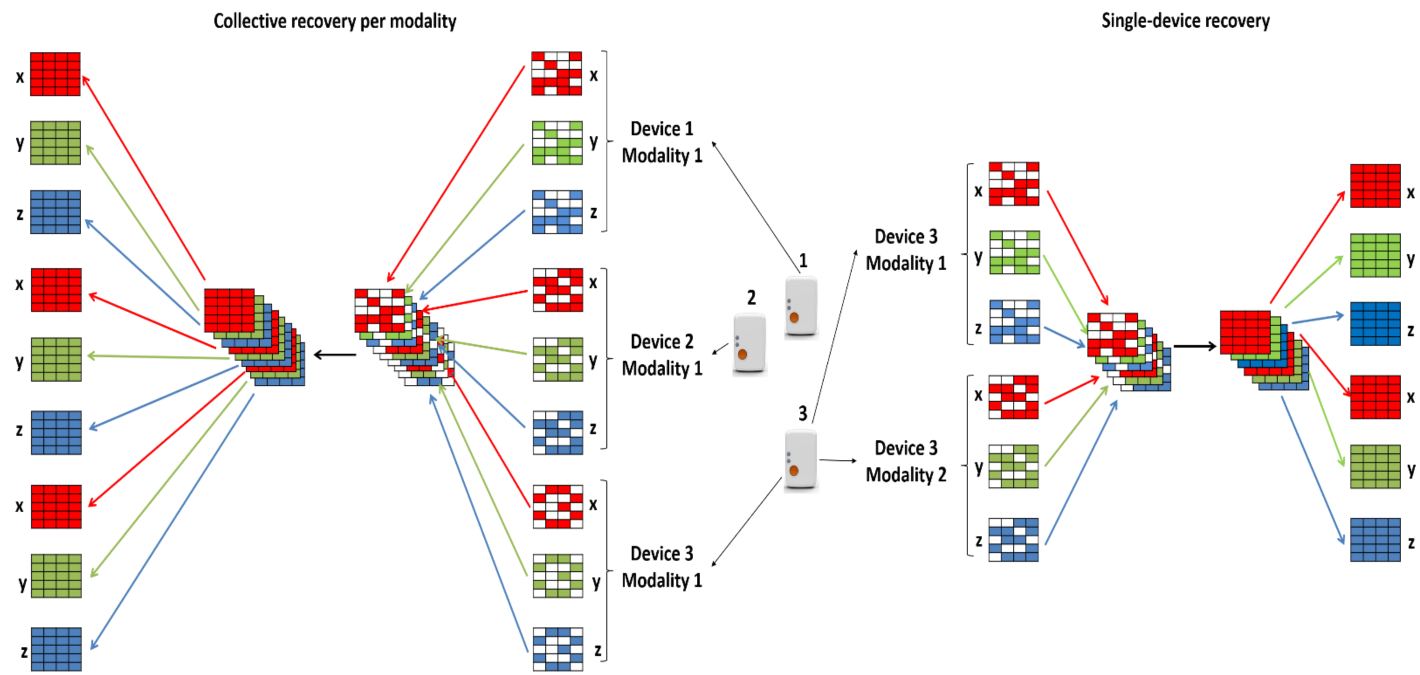


Scenario 3: Overall collective recovery structured similarly

Single-device vs collective recovery: tensors

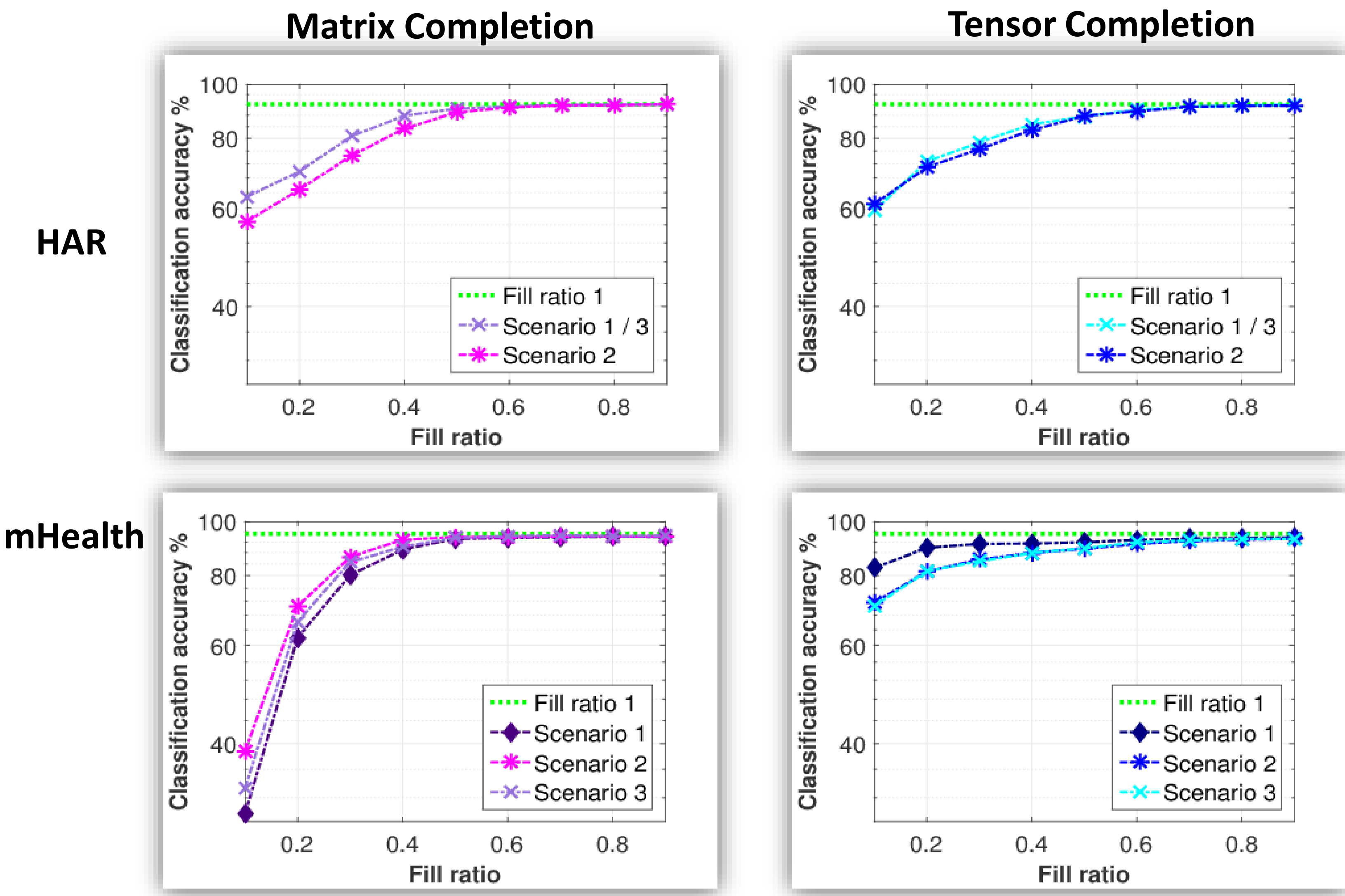
Scenario 2
Collective per modality

Scenario 1
Single-device



Scenario 3: Overall collective recovery structured similarly

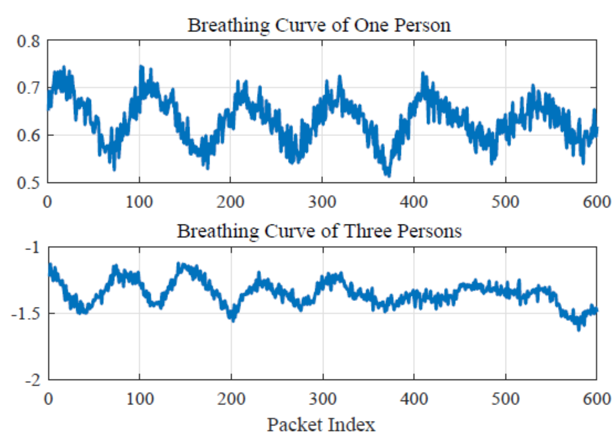
Some results



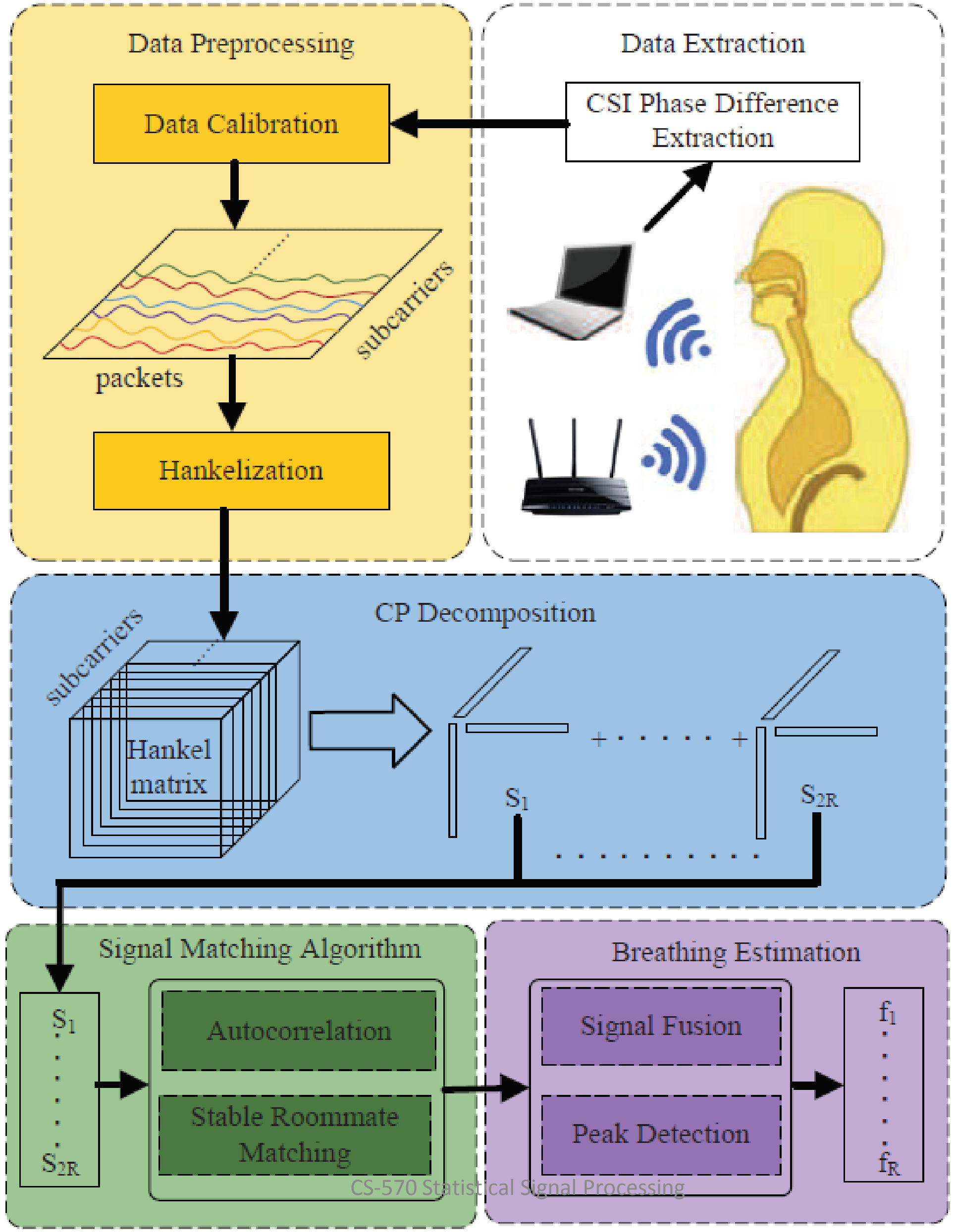
TensorBeat

Tensor Decomposition for Monitoring Multi-Person Breathing Beats with Commodity WiFi

- channel state information (CSI) phase difference
- CP decomposition of a two dimensional Hankel matrix with phase difference data from back-to-back received packets extracted from each subcarrier at each antenna
- leveraging the phase differences from the 60 subcarriers, i.e., that between antennas 1 and 2, and between antennas 2 and 3, we can construct the third dimension of the CSI tensor data.



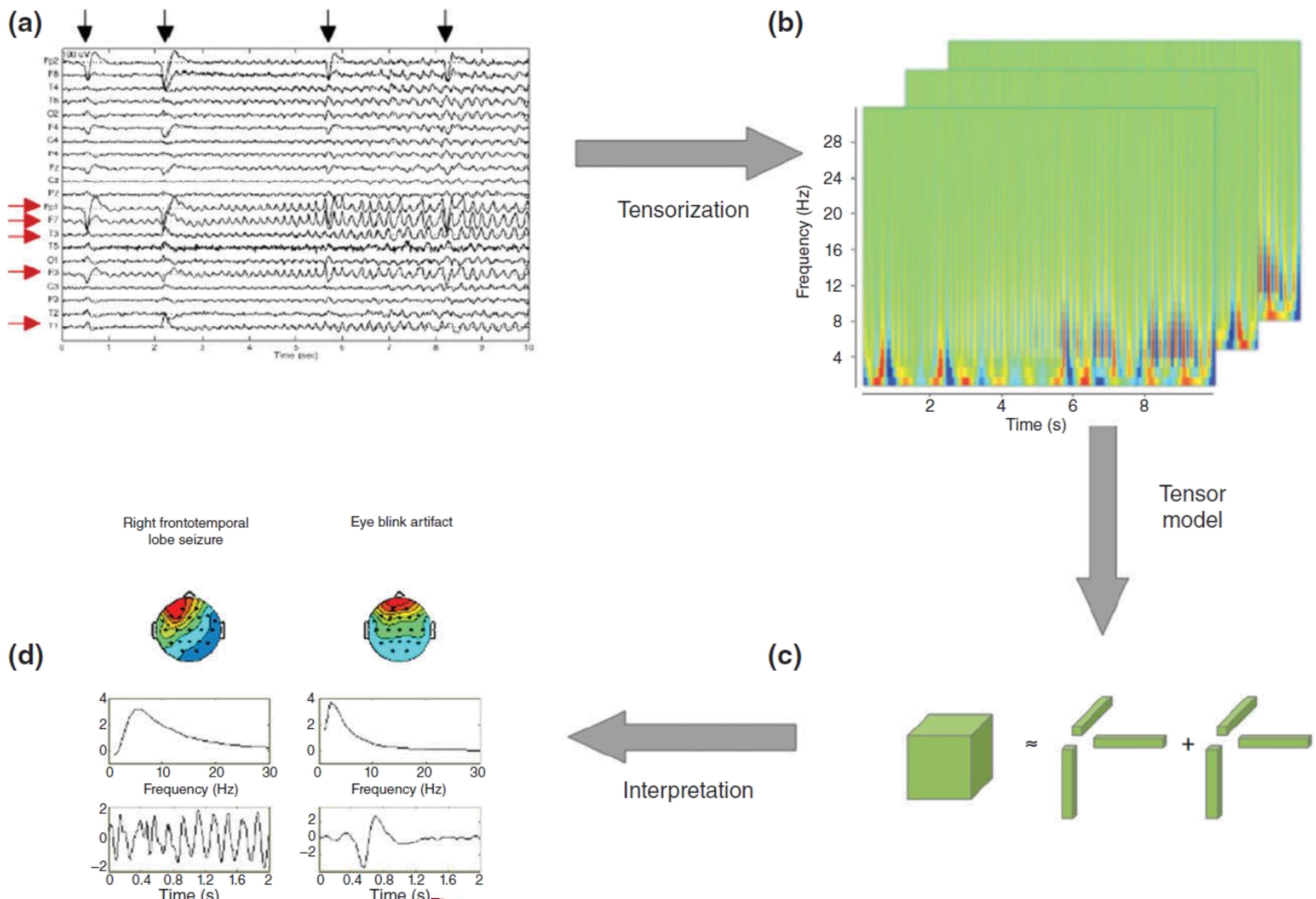
TensorBeat



Data Preprocessing
Data Calibration
packets
subcarriers
Hankelization
Data Extraction
CSI Phase Difference Extraction
CP Decomposition
subcarriers
Hankel matrix
S1
S2R
Signal Matching Algorithm
S1
S2R
Autocorrelation
Stable Roommate Matching
Breathing Estimation
Signal Fusion
Peak Detection
f1
fR

Analysis of EEG and f-MRI

- Tensor decompositions and data fusion in epileptic EEG and functional magnetic resonance imaging data



channels × time samples × measurements × patients

Impact of structuring