

CS-541
Wireless Sensor Networks

Lecture 4: Data models and data acquisition

Spring Semester 2017-2018

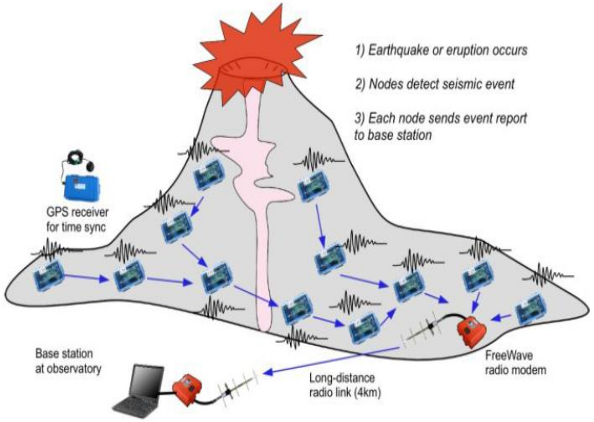
Prof Panagiotis Tsakalides, Dr Athanasia Panousopoulou, Dr Gregory Tsagkatakis

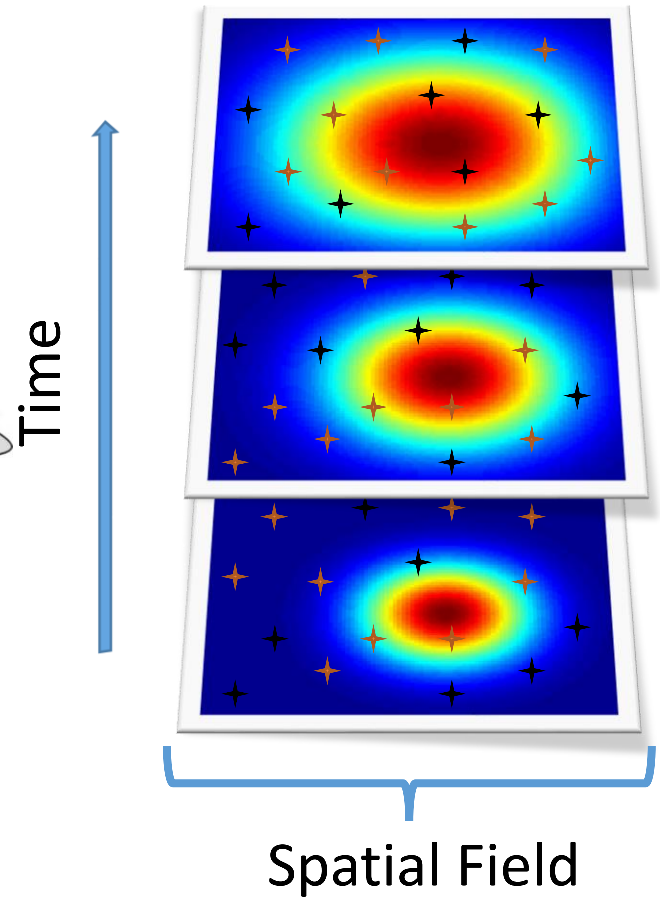
Disclaimer

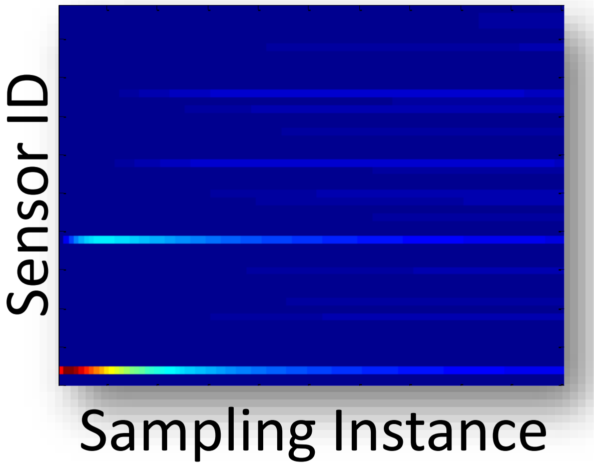
Material adapted from:

- Tensor Decomposition for Signal Processing and Machine Learning, by N.D. Sidiropoulos, L. De Lathauwer, X. Fu, E.E. Papalexakis, ICASSP 2017 Tutorial
- M. Giannopoulos presentation

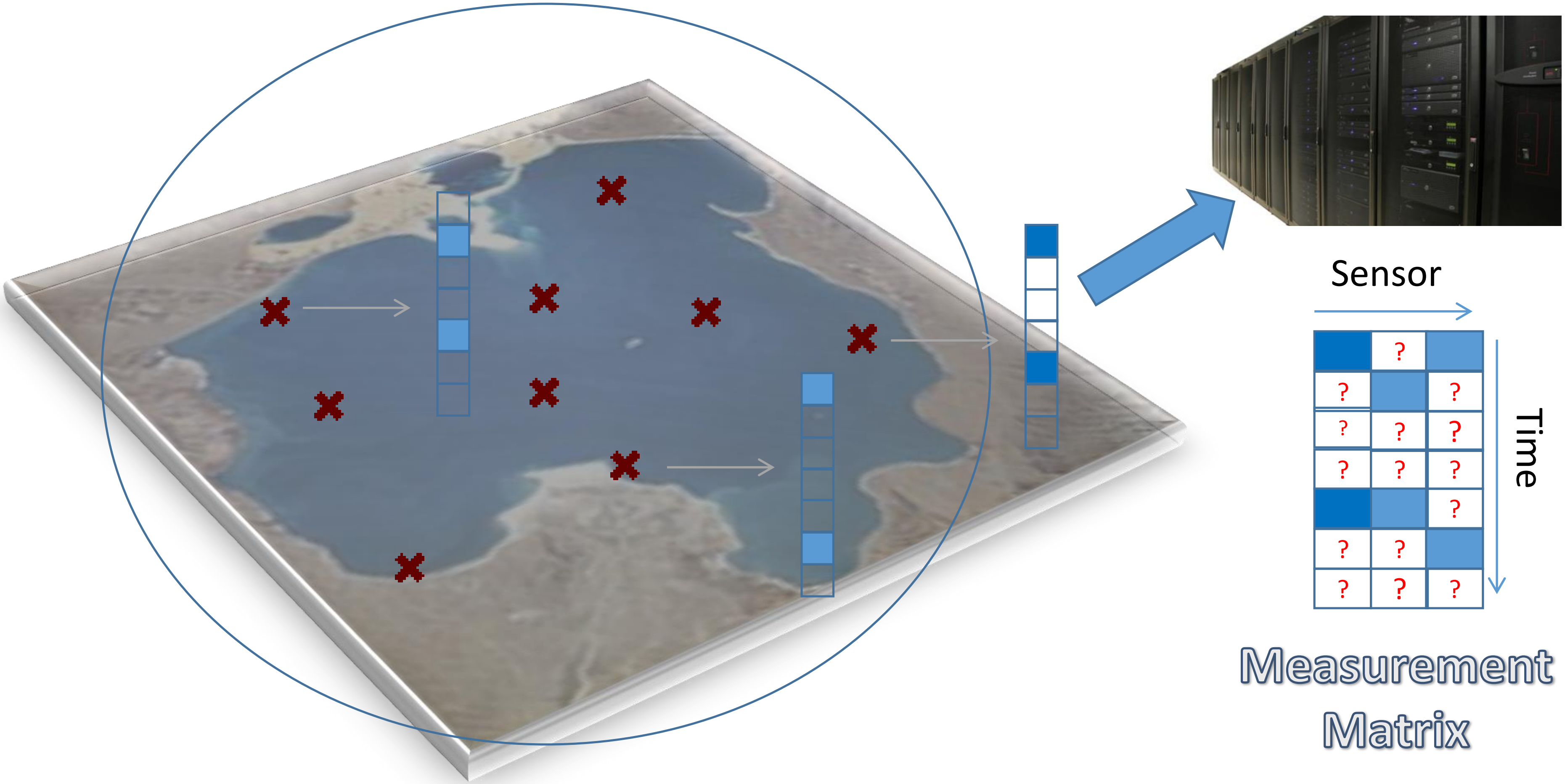
Multivariate based WSN models







Sampling a WSN



Incomplete Matrices

Problem “first” appeared in Netflix challenge

- Given user-movie rating
- Guess missing entries

	John	Anne	Scot	Mark	Alice
Chicago	2	5	?	?	?
Matrix	5	?	5	?	?
Star wars	?	?	5	?	1
Inception	?	3	?	2	?
Alien	4	1	?	?	?
Pulp Fiction	?	?	4	?	2

Matrix Rank

The **rank** of a matrix M is the size of the largest collection of linearly independent *columns of M* (the **column rank**) or the size of the largest collection of linearly independent *rows of M* (the **row rank**)

- Row Echelon Form

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} R_2 \rightarrow 2r_1 + r_2 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 3 & 5 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 3 & 5 & 0 \end{bmatrix} R_3 \rightarrow -3r_1 + r_3 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{bmatrix} R_3 \rightarrow r_2 + r_3 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} R_1 \rightarrow -2r_2 + r_1 \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank=2

A matrix is in **row echelon form** if

(i) all nonzero rows are above any rows of all zeroes

(ii) The leading coefficient of a nonzero row is always strictly to the right of the leading coefficient of the row above it

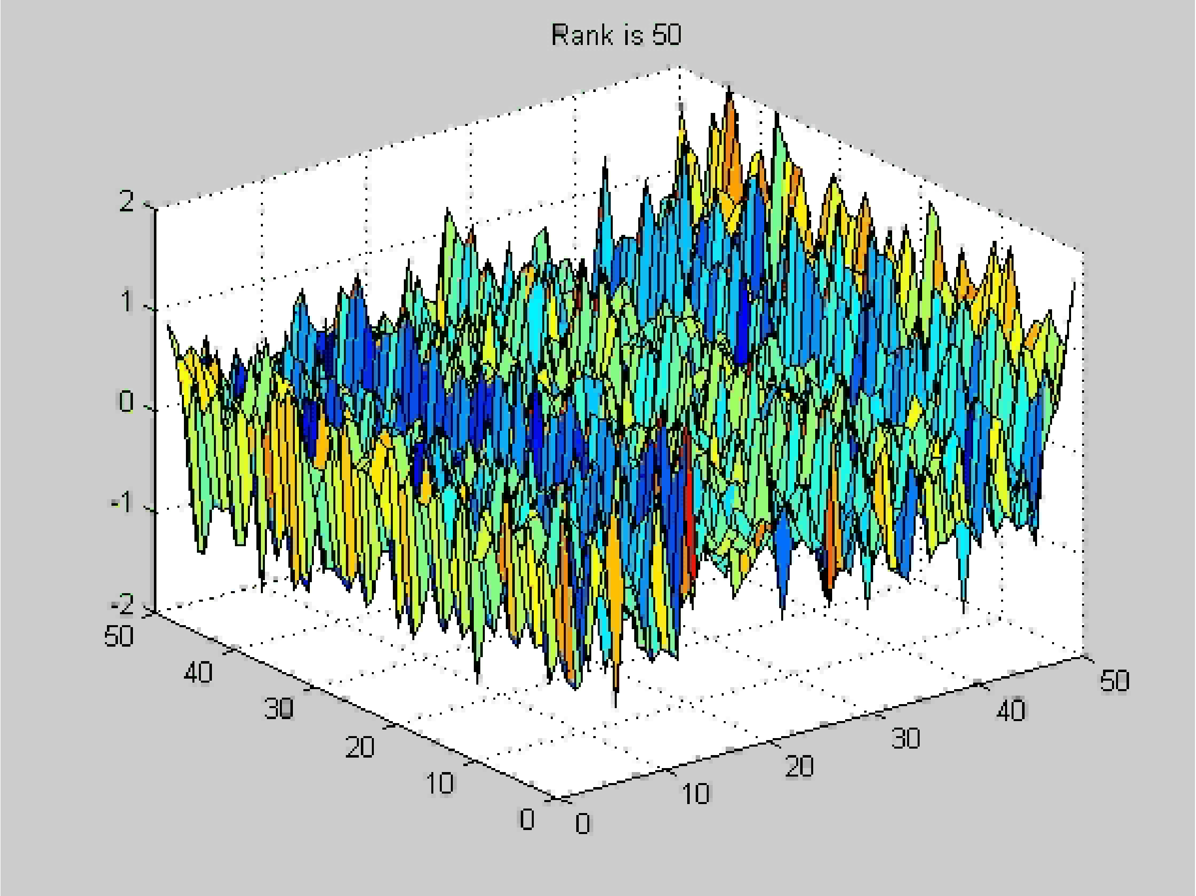
Matrix Rank

- The rank of an $m \times n$ matrix is a nonnegative integer and cannot be greater than either m or n. That is, rank(M) ≤ min(m, n).
- A matrix that has a rank as large as possible is said to have **full rank**; otherwise, the matrix is **rank deficient**.

$$\mathrm{rank}(AB) \leq \min(\mathrm{rank}\ A, \mathrm{rank}\ B).$$

$$\mathrm{rank}(A^TA) = \mathrm{rank}(AA^T) = \mathrm{rank}(A) = \mathrm{rank}(A^T)$$

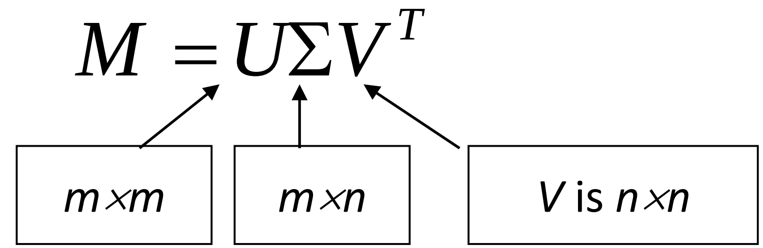
Matrix Rank



Singular Value Decomposition (SVD)

Given any $m \times n$ matrix **M**, algorithm to find matrices **U**, **Σ**, and **V** such that $\mathbf{M} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}}$

- **U**: left singular vectors (orthonormal)
- **Σ**: diagonal containing singular values
- **V**: right singular vectors (orthonormal)



$$\begin{pmatrix} \\ \\ M \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \\ \mathbf{U} \\ \\ \\ \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix} \begin{pmatrix} & & \\ & \mathbf{V} & \\ & & \end{pmatrix}^{\mathrm{T}}$$

Singular Value Decomposition (SVD)

Properties

- The s_i are called the singular values of **M**
- If **M** is singular, some of the s_i will be 0
- In general *rank*(**M**) = number of nonzero s_i
- SVD is mostly unique (up to permutation of SV)

Low rank approximations

- Denoising
- Dimensionality reduction

Low rank approximation

Matrix norms

- Frobenius norm can be computed from SVD $\|M\|_F = \sum_i \sum_j m_{ij}^2$
- Changes to a matrix ↔ changes to singular values $\|M\|_F = \sum_i s_i^2$

Low rank approximation

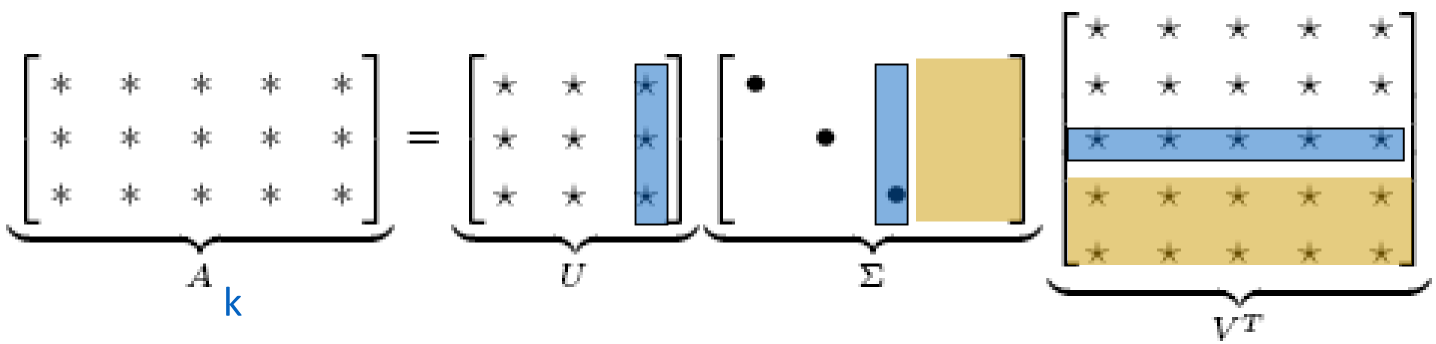
Approximation problem: Find ***M_k*** of rank ***k*** such that

$$M_k = \min_{X:rank(X)=k} \|M - X\|_F$$

Singular Value Decomposition (SVD)

- Solution via SVD $M_k = U \operatorname{diag}(\sigma_1, \ldots, \sigma_k, \underbrace{0, \ldots, 0}) V^T$

set smallest r-k
singular values to zero



$$M_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

← *column notation: sum of rank 1 matrices*

Approximation error

- How good (bad) is this approximation?
- It's the best possible, measured by the Frobenius norm of the error:

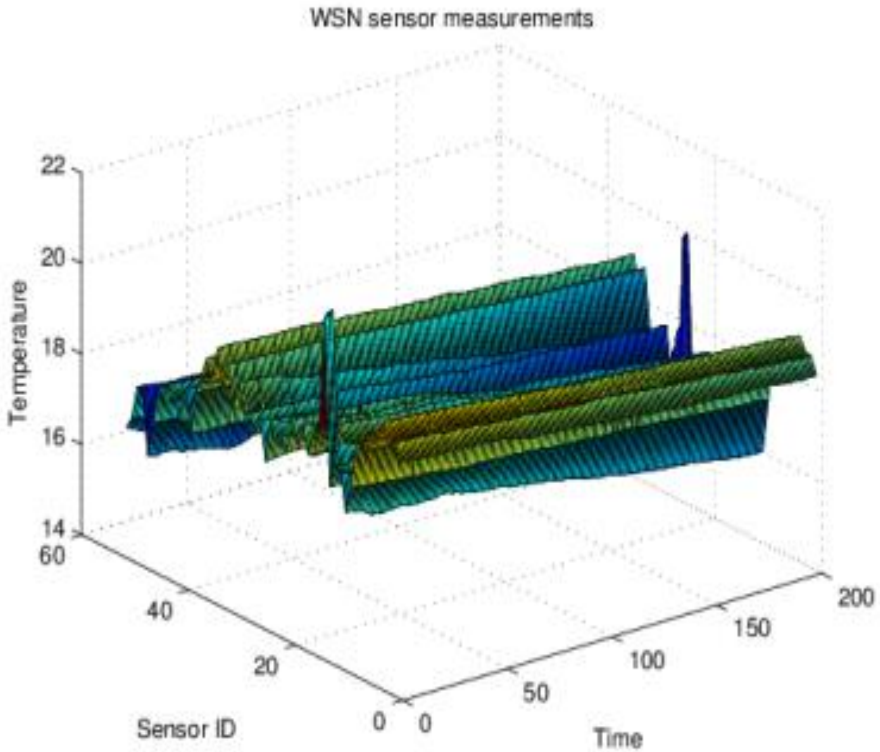
$$\min_{X:rank(X)=k} \|M - X\|_F = \|M - M_k\|_F = \sigma_{k+1}$$

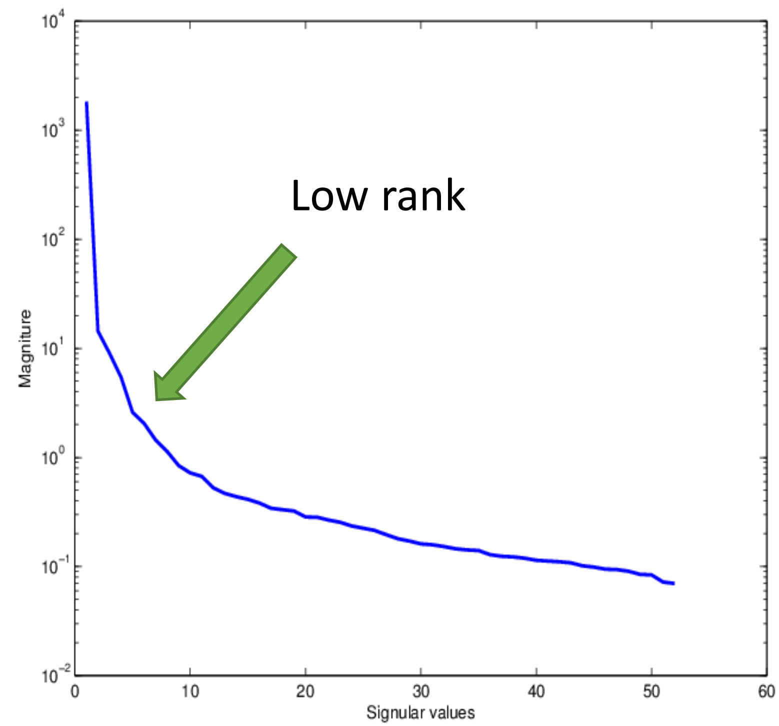
where the σ_i are ordered such that $\sigma_i \geq \sigma_{i+1}$.

Suggests why Frobenius error drops as *k* increased.

Data model

- WSN data modeling
  - Spatio-temporal correlations <-> Low rank measurement matrix





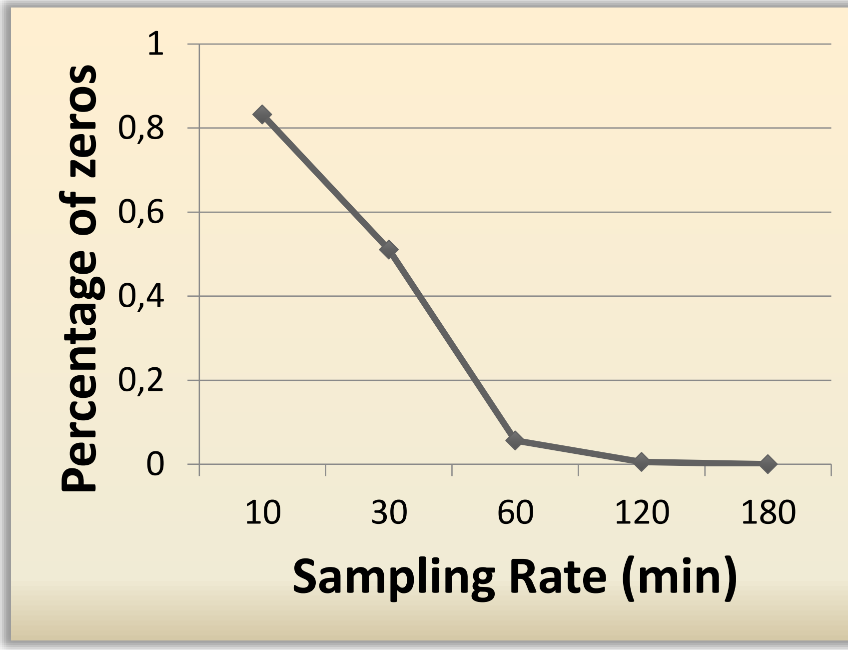
The case of missing values

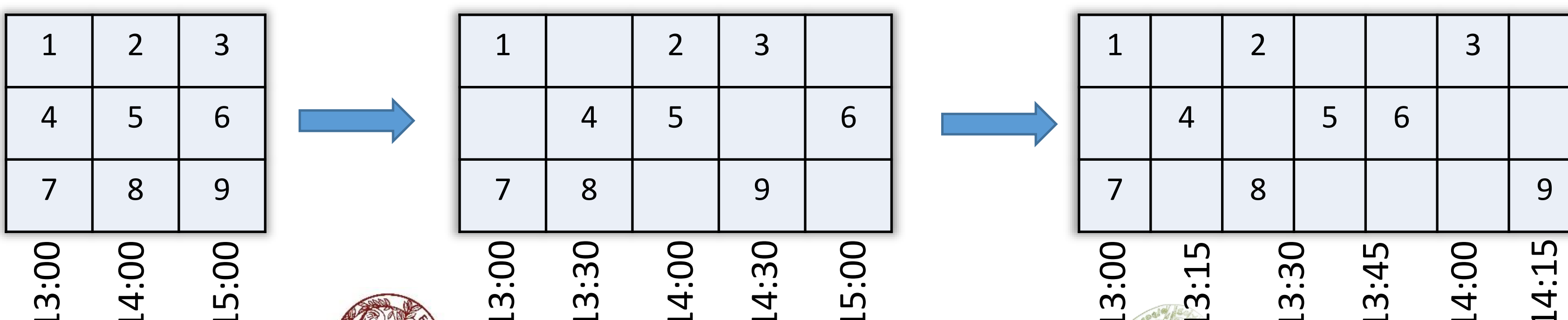
Power consumption

Packet losses

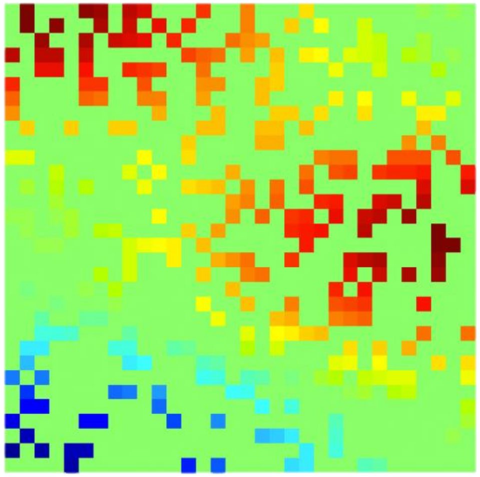
Temporal sampling of WSN

- Sampling rate
- De-synchronization
- Temporal resolution

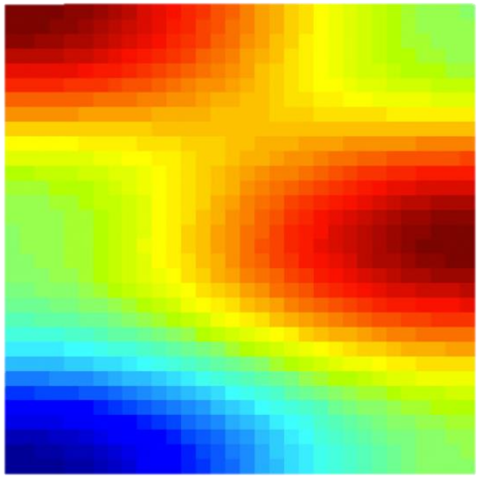




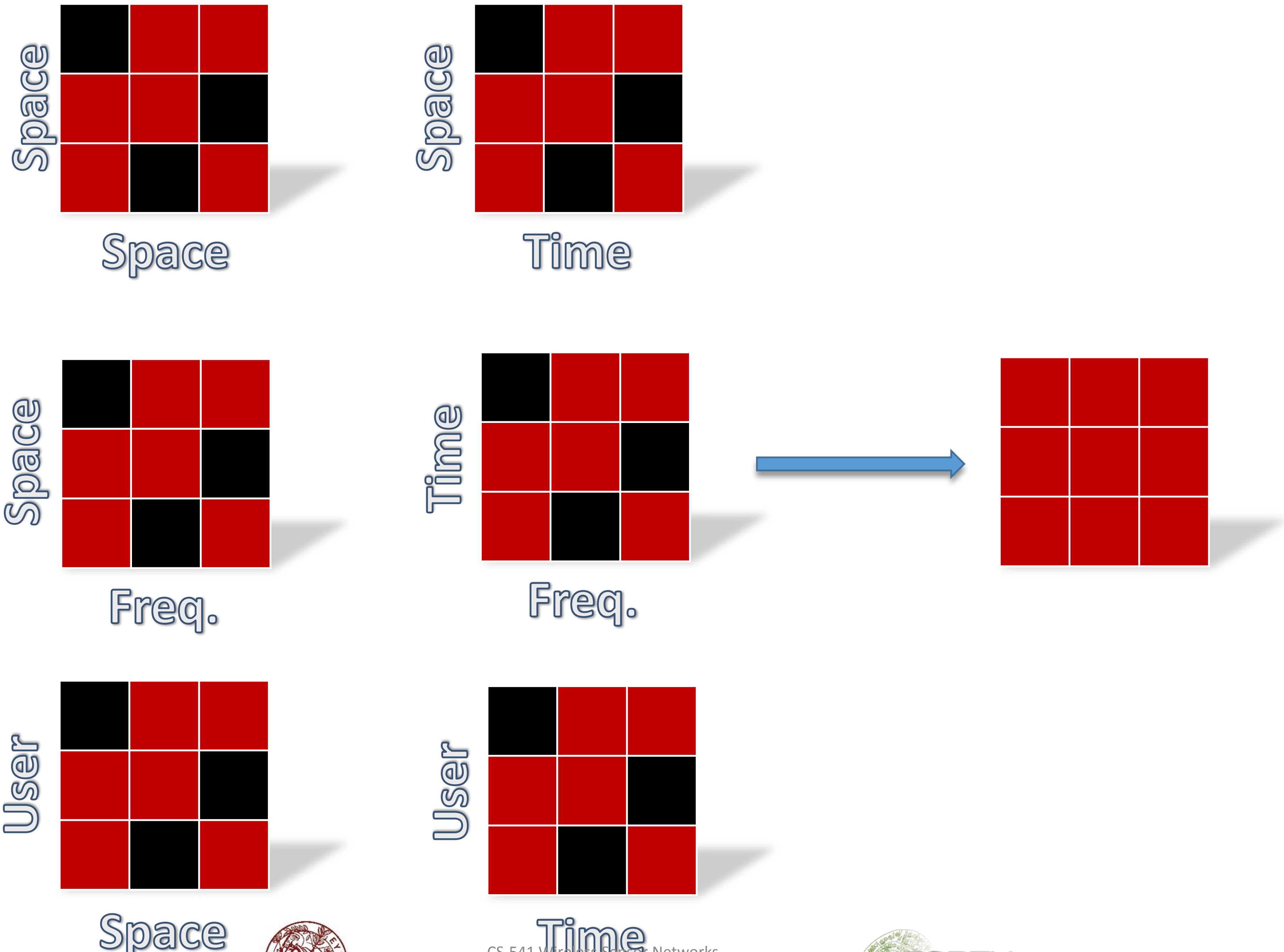
Matrix completion



low rank matrix with
missing entries



low rank
matrix


Space
Space
Space
Time
Space
Freq.
Time
Freq.
User
Space
User
Time

Matrix Completion (MC)

Let $\mathbf{M} = [\mathbf{M_0}, ..., \mathbf{M_1}] \in \mathbf{R^{n \times s}}$ be a measurement matrix consisting of *s* measurements from *n* different sources.

Recovery of **M** is possible from *k<<ns* random entries if matrix **M** is *low rank* and $k \geq Cn^{6/5} r log(n)$

To recover the unknown matrix, solve:

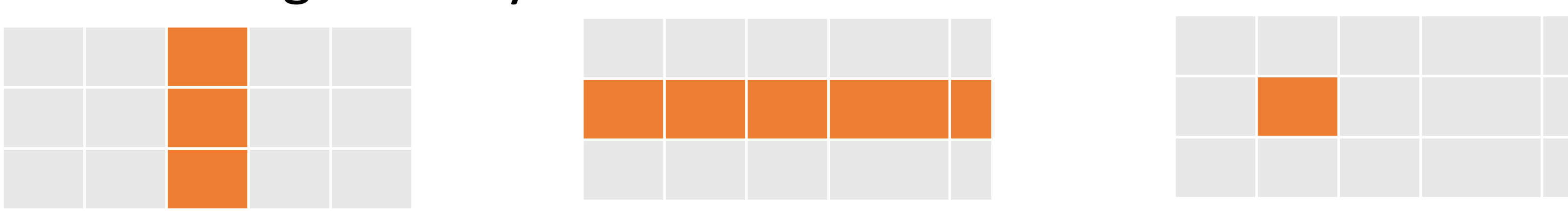
$$\min\{ \ rank(\mathbf{X}) \ : \ \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M})\}$$

Rank constraint makes problem NP-hard....

Sampling operator

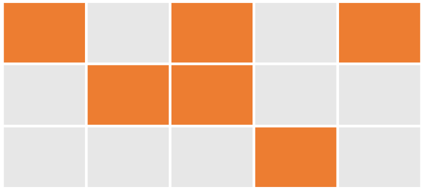
Sampling operator $\mathcal{A}_{ij}(\mathbf{M}) = \begin{cases} M_{ij}, & \text{if } ij \in S \\ 0, & \text{otherwise} \end{cases}$

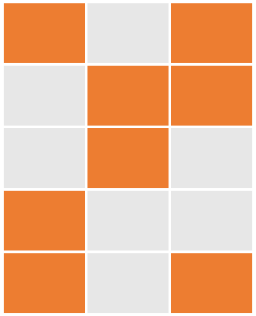
- Not all low-rank matrices can be recovered from partial measurements!
  - … a matrix containing zeroes everywhere except the top-right corner.
  - This matrix is low rank, but it ***cannot*** be recovered from knowledge of only a fraction of its entries!



Matrix Coherence

The coherence of subspace ***U*** of $\Re^n$ and having dimension *r* with respect to the canonical basis $\{\mathbf{e}_i\}$ is defined as: $\mu(U) = \frac{n}{r} \max_{1 \le i \le n} \|Ue_i\|^2$





$\mu(U) = O(1)$

– sampled from the uniform distribution with $r > \log n$

Formal definition of key assumptions

- Consider an underlying matrix **M** of size n_1 by n_2. Let the SVD of **M** be given as follows:

$$M = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

- We make the following assumptions about **M**:

(A0) $\mu_1 \sqrt{r/(n_1 n_2)}, \mu_1 > 0$

(A1) The maximum entry in the n_1 by n_2 matrix $\sum_{k=1}^{r} u_k v_k^T$ is upper bounded by

$$\exists \mu_0 \text{ such that } \max(\mu(U), \mu(V)) \leq \mu_0$$

What do these assumptions mean

(A0) means that the singular vectors of the matrix are sufficiently **incoherent** with the canonical basis.

(A1) means that the singular vectors of the matrix are **not spiky**

- canonical basis vectors are spiky signals – the spike has magnitude 1 and the rest of the signal is 0;
- a vector of n elements with all values equal to 1/square-root(n) is not spiky.

What is the trace-norm of a matrix?

- The nuclear / trace norm of a matrix is the **sum of its singular values.**

$$\|\mathbf{M}\|_* = \sum_{i=1}^{k} \sigma_i$$

- It is a **softened version of the rank** of a matrix
- Similar to the $L_0 \longrightarrow L_1$-norm of a vector
- Minimization of the trace-norm is a **convex optimization problem** and can be solved efficiently.
- This is similar to the L_1-norm optimization (in compressive sensing) being efficiently solvable.

Matrix Completion (MC)

Relaxation

$$\min\{ \ \|\mathbf{M}\|_* \ : \ \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M})\}$$

Performance

$$\left\|M - M^*\right\|_F^2 \le 4\sqrt{\frac{(2+p)\min(n_1, n_2)}{p}}\delta + 2\delta,$$

$$\text{where } p = \text{fraction of known entries} = \frac{m}{n_1 n_2} = \frac{|\Omega|}{n_1 n_2}$$

Noisy case

$$\min\{ \ \|\mathbf{M}\|_* \ : \ \|\mathcal{A}(\mathbf{X}) - \mathcal{A}(\mathbf{M})\|_F^2 \le \epsilon\}$$

Recovery guarantees

Theorem 1.3 *Let* $\boldsymbol{M}$ *be an* $n_1 \times n_2$ *matrix of rank* r *obeying* **A0** *and* **A1** *and put* $n = \max(n_1, n_2)$. *Suppose we observe* m *entries of* $\boldsymbol{M}$ *with locations sampled uniformly at random. Then there exist constants* C, c *such that if*

$$m \geq C \max(\mu_1^2, \mu_0^{1/2}\mu_1, \mu_0 n^{1/4}) \, nr(\beta \log n) \tag{1.9}$$

for some $\beta > 2$, *then the minimizer to the problem* (1.5) *is unique and equal to* $\boldsymbol{M}$ *with probability at least* $1 - cn^{-\beta}$. *For* $r \leq \mu_0^{-1} n^{1/5}$ *this estimate can be improved to*

$$m \geq C \, \mu_0 \, n^{6/5} r(\beta \log n) \tag{1.10}$$

with the same probability of success.

the trace-norm minimizer

Matrix Completion solvers

- Matrix Completion via ALM
  - Objective

$$\begin{aligned} &\text{minimize}_{\mathbf{X}} \quad \|\mathbf{X}\|_* \\ &\text{subject to} \;\; \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M}) \end{aligned}$$

  - Reformulation

$$\begin{aligned} &\text{minimize}_{\mathbf{X},\mathbf{E}} \quad \|\mathbf{X}\|_* \\ &\text{subject to} \;\; \mathbf{X} + \mathbf{E} = \mathbf{M} \\ &\qquad\qquad\quad \mathcal{A}(\mathbf{E}) = 0 \end{aligned}$$

Matrix Completion solvers

- Let y=A(M)

$$\underset{X}{\text{minimize}} \, \|\mathcal{A}(X) - y\|_2^2 + \lambda \|X\|_*.$$

- Iterative Hard Thresholding

$$Y_{k+1} = X_k - \gamma_k \mathcal{A}^*(\mathcal{A}(X_k) - y))$$

$$X_{k+1} = \text{ProjectRank}_R(Y_{k+1}).$$

← **SVD**

CS and MC

	Sparse recovery	***Rank minimization***
Unknown	Vector *x*	Matrix *A*
Observations	*y = Ax*	*y = L[A]* (linear map)
Combinatorial objective	$\#\{\mathbf{x}_i \neq 0\} = \|\mathbf{x}\|_0$	$\text{rank}(A) = \#\{\sigma_i(A) \neq 0\} = \|\sigma(A)\|_0$
Convex relaxation	$\|\mathbf{x}\|_1 = \sum_i \lvert\mathbf{x}_i\rvert$	$\|A\|_* = \sum_i \sigma_i(A)$
Algorithmic tools	Linear programming	Semidefinite programming

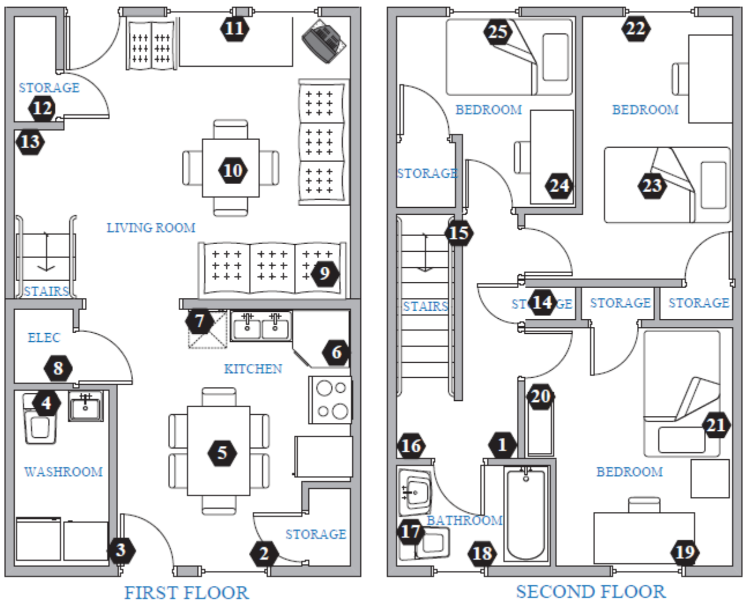
Yi Ma et al, "Matrix Extensions to Sparse Recovery", CVPR2009

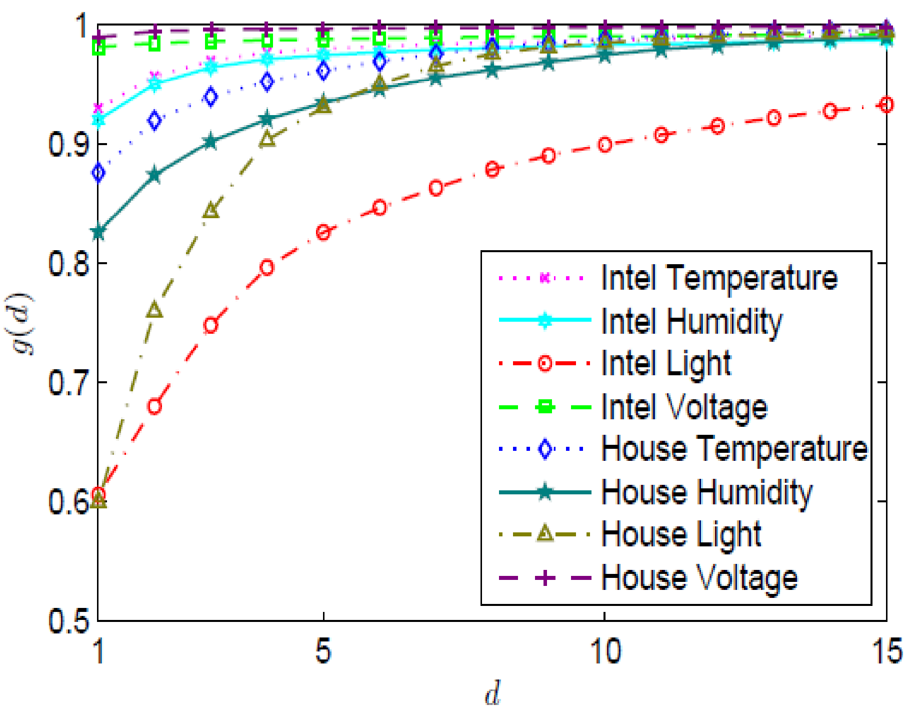
Applications of MC

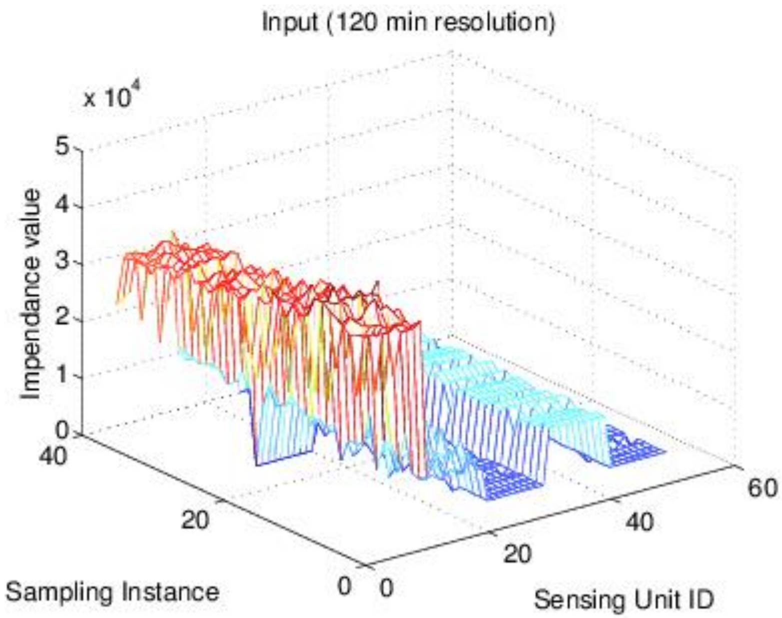
- Recommendation systems
  - Matrix (user, preference/quality/intention)
- Sensor localization
  - Matrix (location, physical quantity)
- Data recovery in Wireless Sensor Networks
  - Matrix (sensor, time)

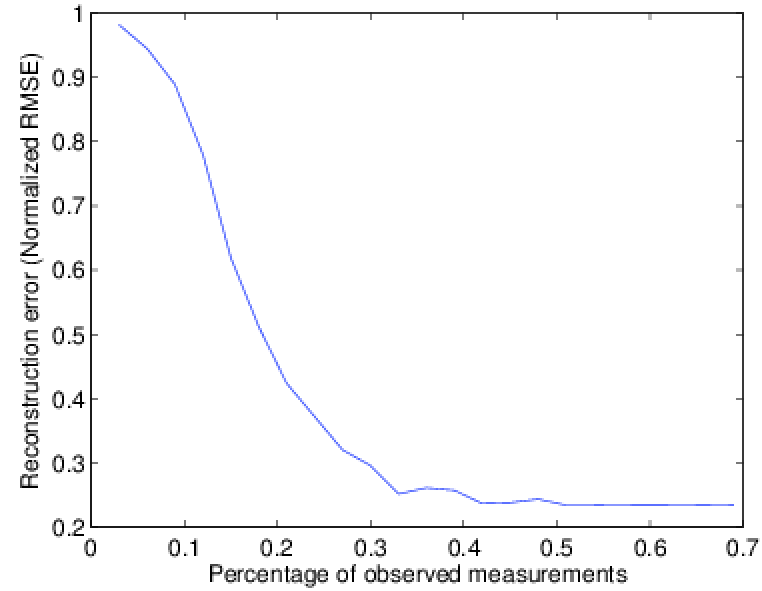
Data Gathering

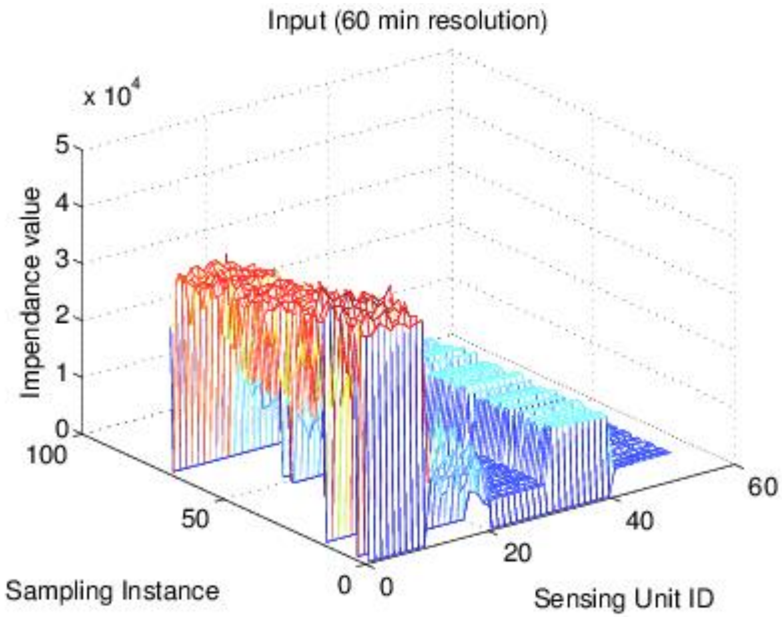
- STCDG: An Efficient Data Gathering Algorithm Based on Matrix Completion for Wireless Sensor Networks

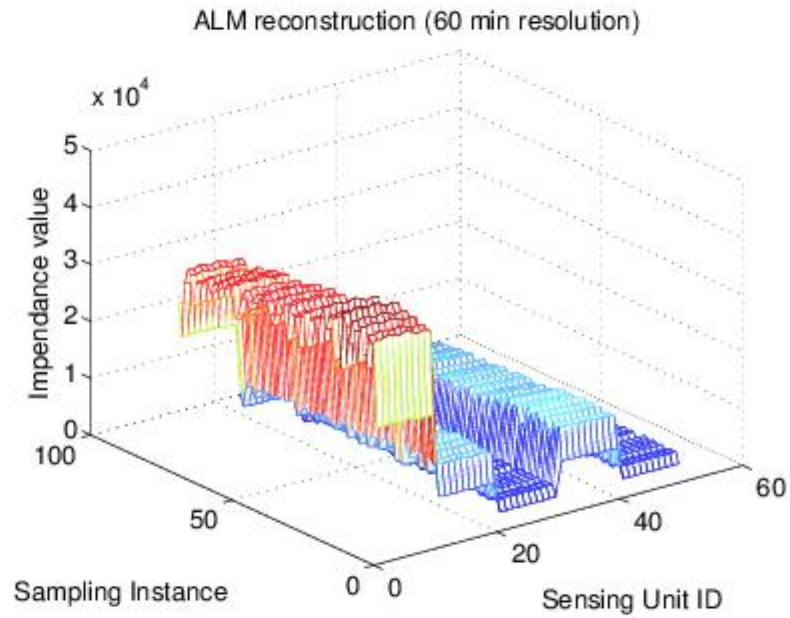




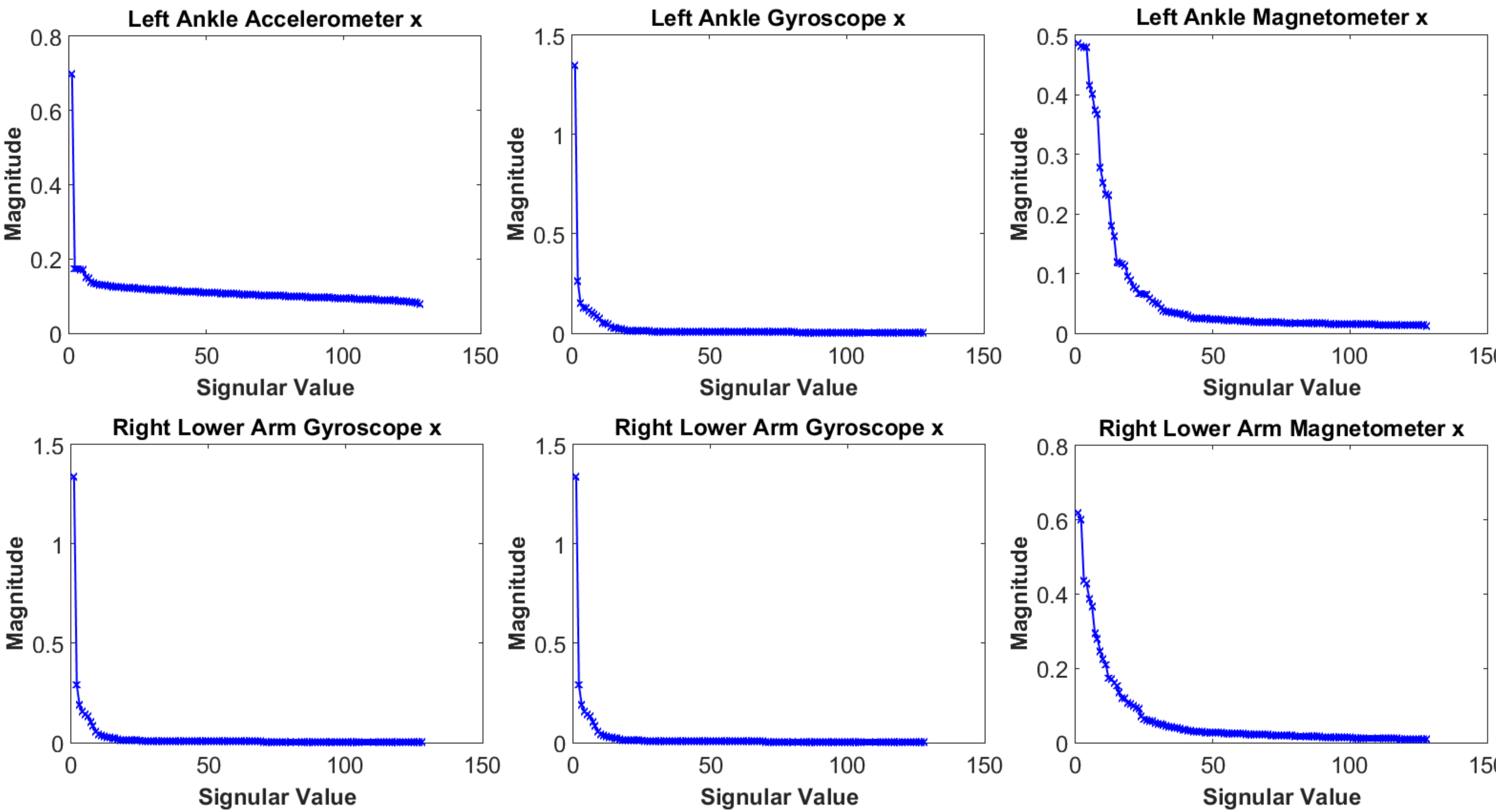

Input (120 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID


Reconstruction error (Normalized RMSE)
Percentage of observed measurements


Input (60 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID


ALM reconstruction (60 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID

Body sensor network



RTT estimation

Decentralized Matrix Factorization by Stochastic Gradient Descent (DMFSGD),

Estimation of end-to-end network distances

- Network nodes exchange messages with each other
- Each node collects and processes local measurements

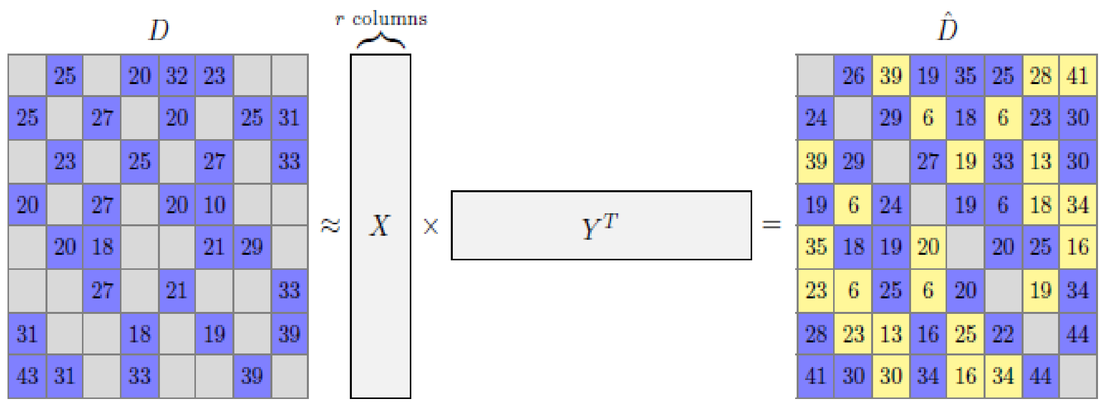


Fig. 2. Network distance prediction by matrix factorization. Note that the diagonal entries of D and $\hat{D}$ are empty.

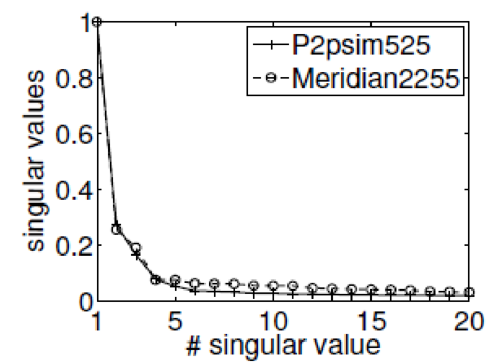
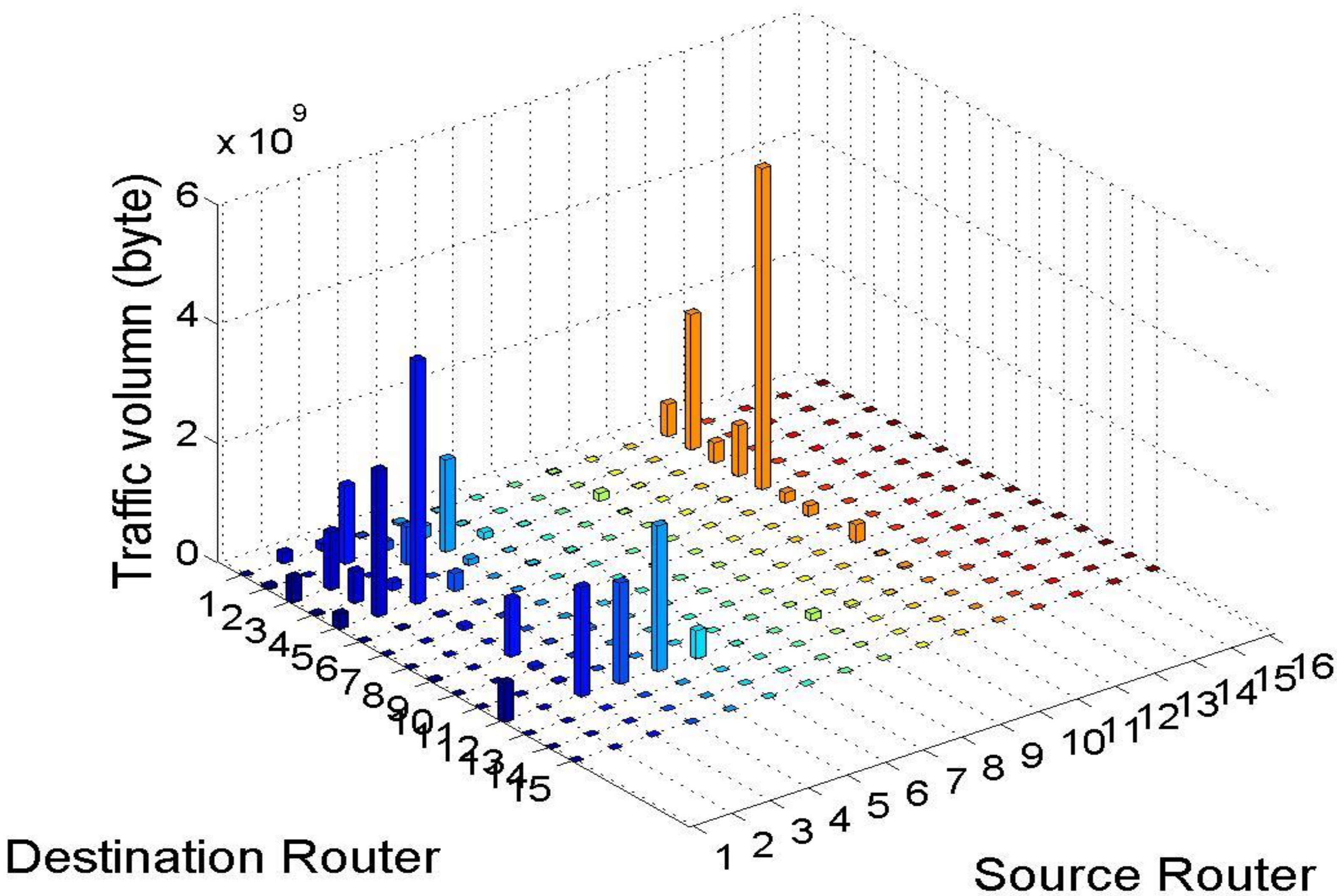
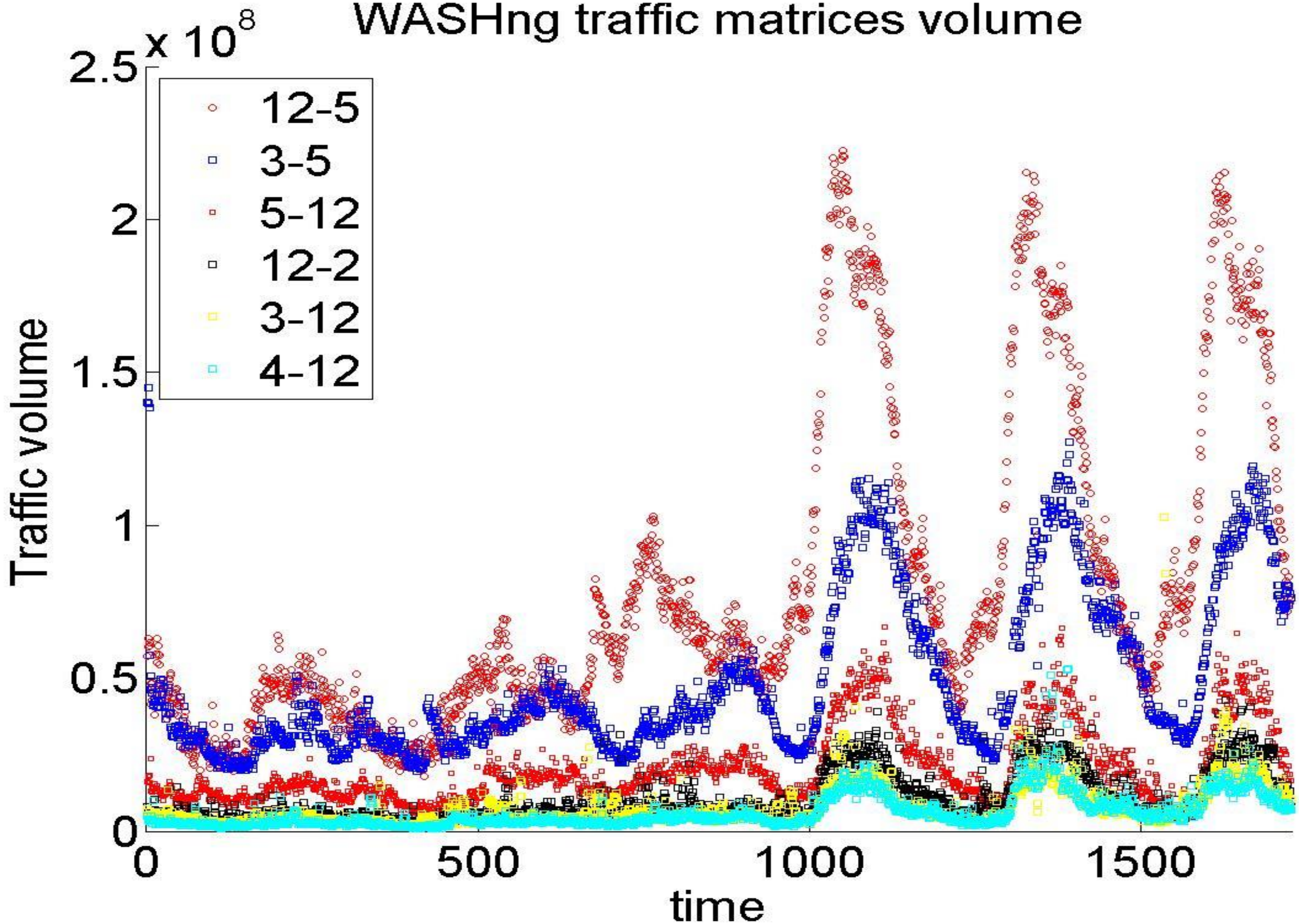


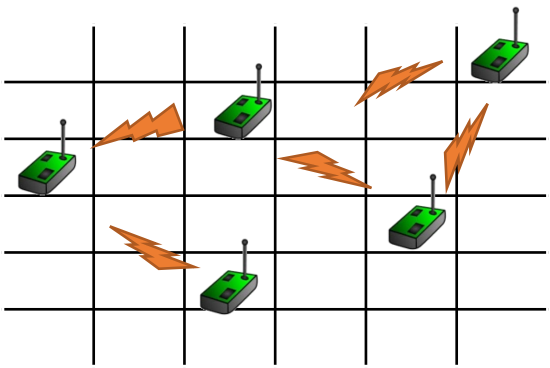
Fig. 3. The singular values of a RTT matrix of 2255×2255, extracted from the Meridian dataset [30] and called "Meridian2255", and of a RTT matrix of 525×525, extracted from the P2psim dataset [30] and called "P2psim525". The singular values are normalized so that the largest singular values of both matrices are equal to 1.

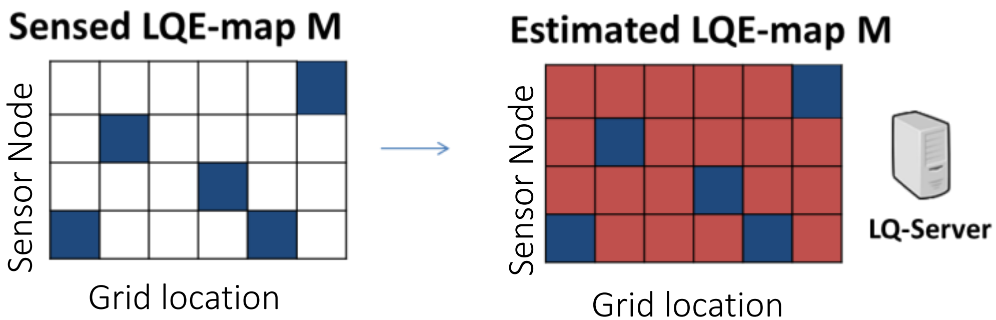
Traffic Matrix of router WASHng

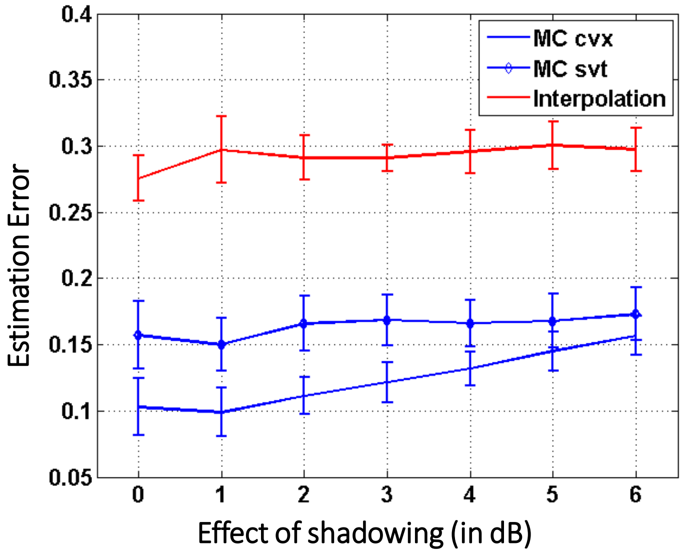


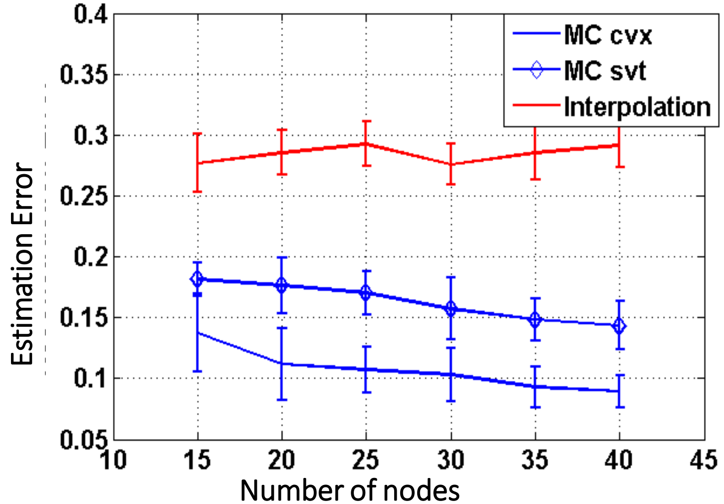

WASHng traffic matrices volume
x 10^8
2.5
2
1.5
1
0.5
0
12-5
3-5
5-12
12-2
3-12
4-12
Traffic volume
0
500
1000
1500
time

LQM Estimation



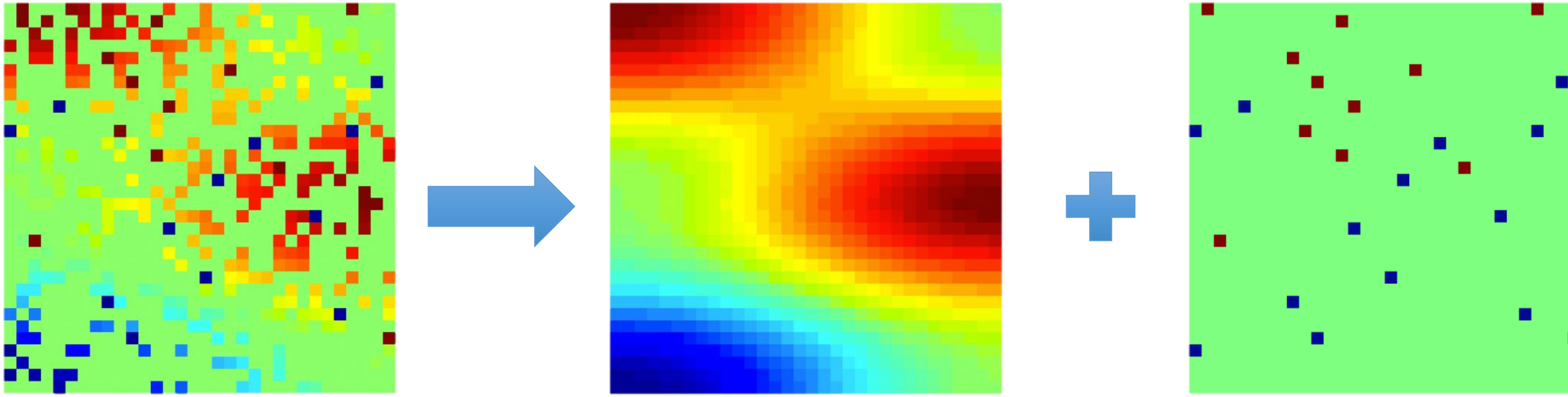






- Dataset
  - Testbed @ FORTH (144m^2, 1x1m grid)
  - RSSI values (channel quality)
  - 13 IEEE802.11b/g channels

Robust PCA



missing + **corrupted** entries

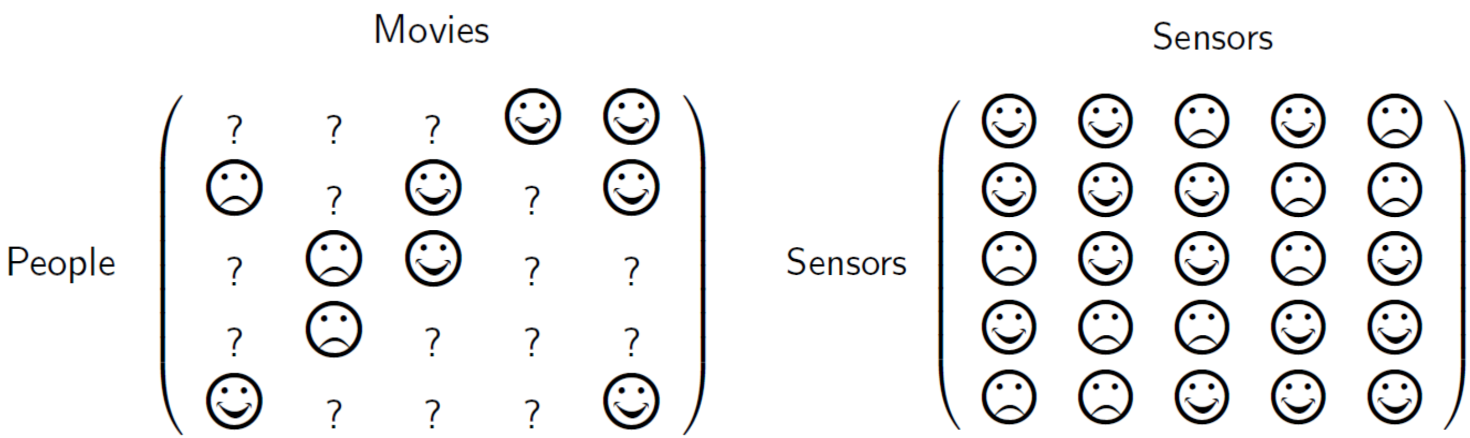
low rank matrix

sparse corruptions

$$\text{minimize}_{\mathbf{X},\mathbf{E}} \quad \|\mathbf{X}\|_* + \|\mathbf{E}\|_1$$

$$\text{subject to} \quad \mathcal{A}(\mathbf{X}+\mathbf{E}) = \mathcal{A}(\mathbf{M})$$

1-Bit MC


Movies
People
Sensors
Sensors

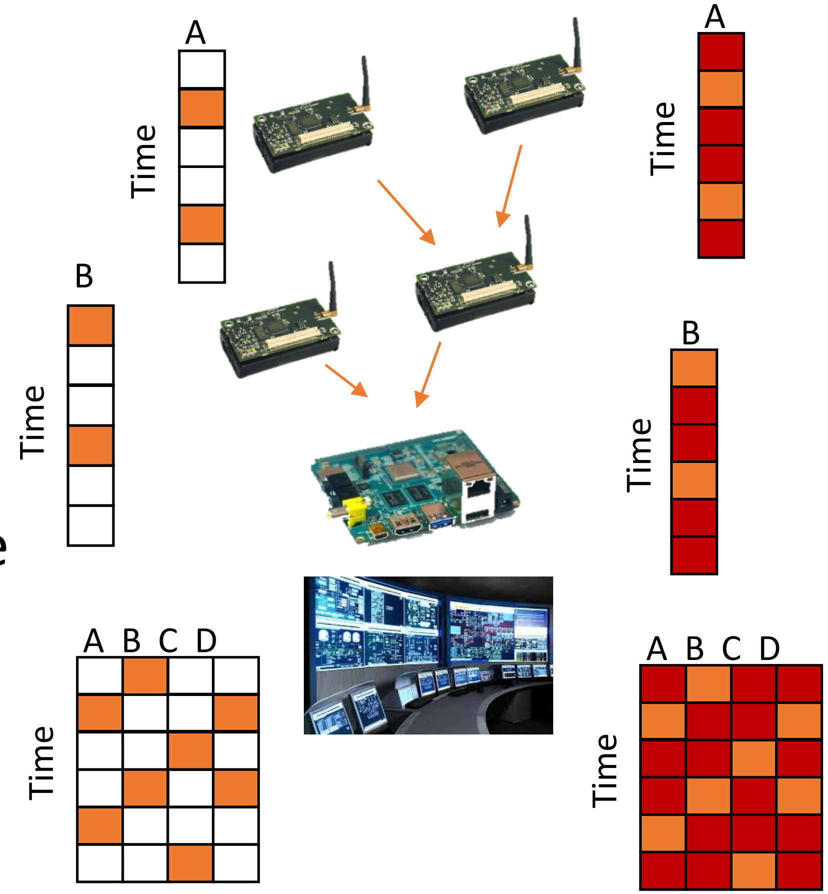

M
$\mathbb{P}\left(Y_{i,j}=☺\right)=f(M_{i,j})$
"Preference" Matrix

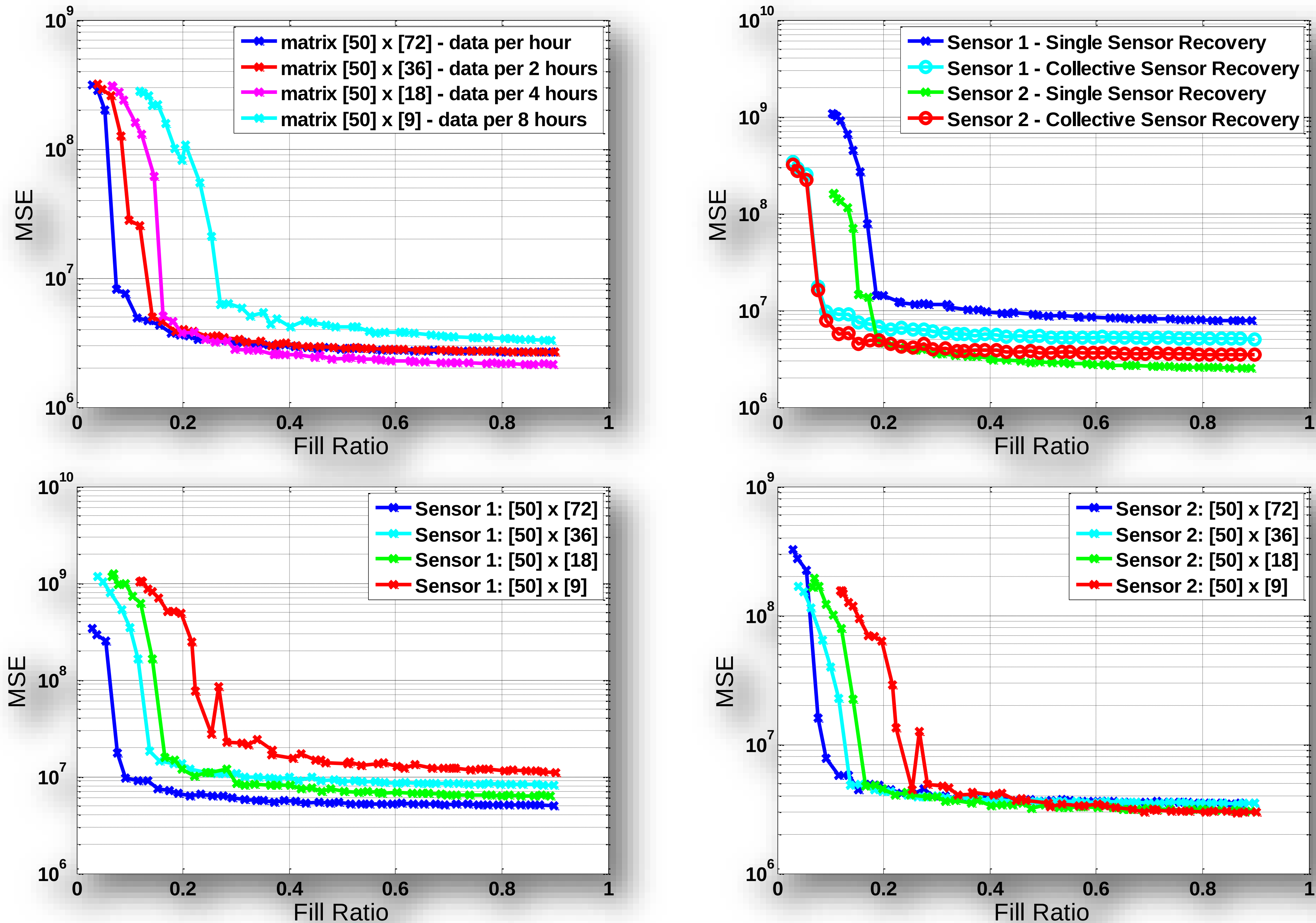
Distributed vs. Centralized Storage

- Centralized
  - ➢ Access to resources
  - ➢ Controlled environment
- Decentralized
  - ➢ Increased network lifetime
  - ➢ Autonomy

Performance comparison

- Per sensor vs. collective
- Temporal resolution

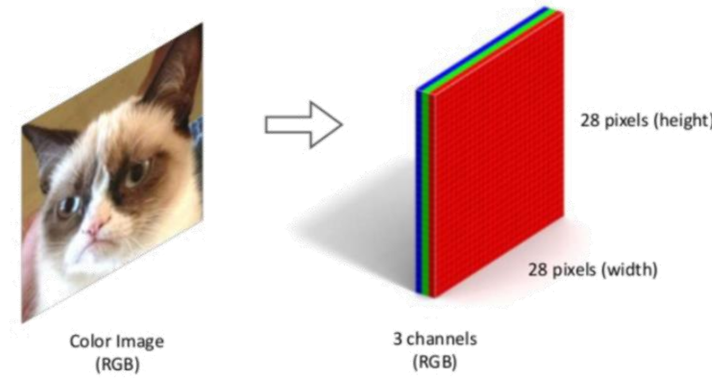



MSE
Fill Ratio
matrix [50] x [72] - data per hour
matrix [50] x [36] - data per 2 hours
matrix [50] x [18] - data per 4 hours
matrix [50] x [9] - data per 8 hours
Sensor 1 - Single Sensor Recovery
Sensor 1 - Collective Sensor Recovery
Sensor 2 - Single Sensor Recovery
Sensor 2 - Collective Sensor Recovery
Sensor 1: [50] x [72]
Sensor 1: [50] x [36]
Sensor 1: [50] x [18]
Sensor 1: [50] x [9]
Sensor 2: [50] x [72]
Sensor 2: [50] x [36]
Sensor 2: [50] x [18]
Sensor 2: [50] x [9]

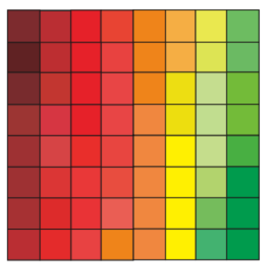
High-dimensional signal models

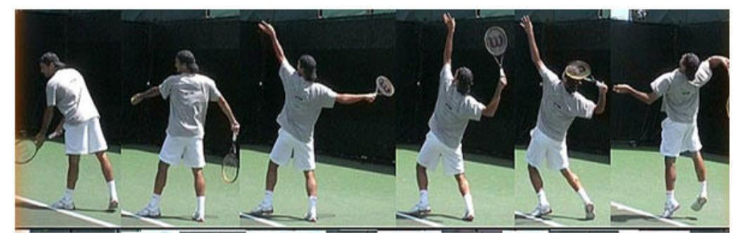
Encoding of multiple variables

- Time, Space, Frequency, Modality



vector | matrix | tensor

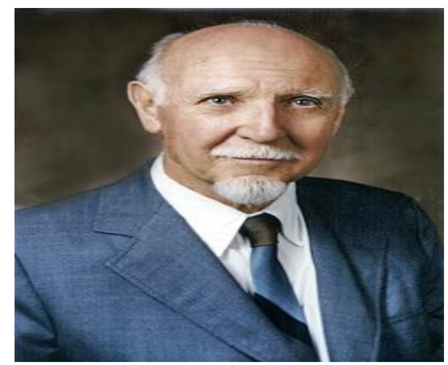




$\mathbf{v} \in \mathbb{R}^{64}$ $\quad X \in \mathbb{R}^{8x8}$ $\quad \mathcal{X} \in \mathbb{R}^{4x4x4}$

Tensor Decompositions-Historical Background

- **<u>Founding fathers</u>**:
  - 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

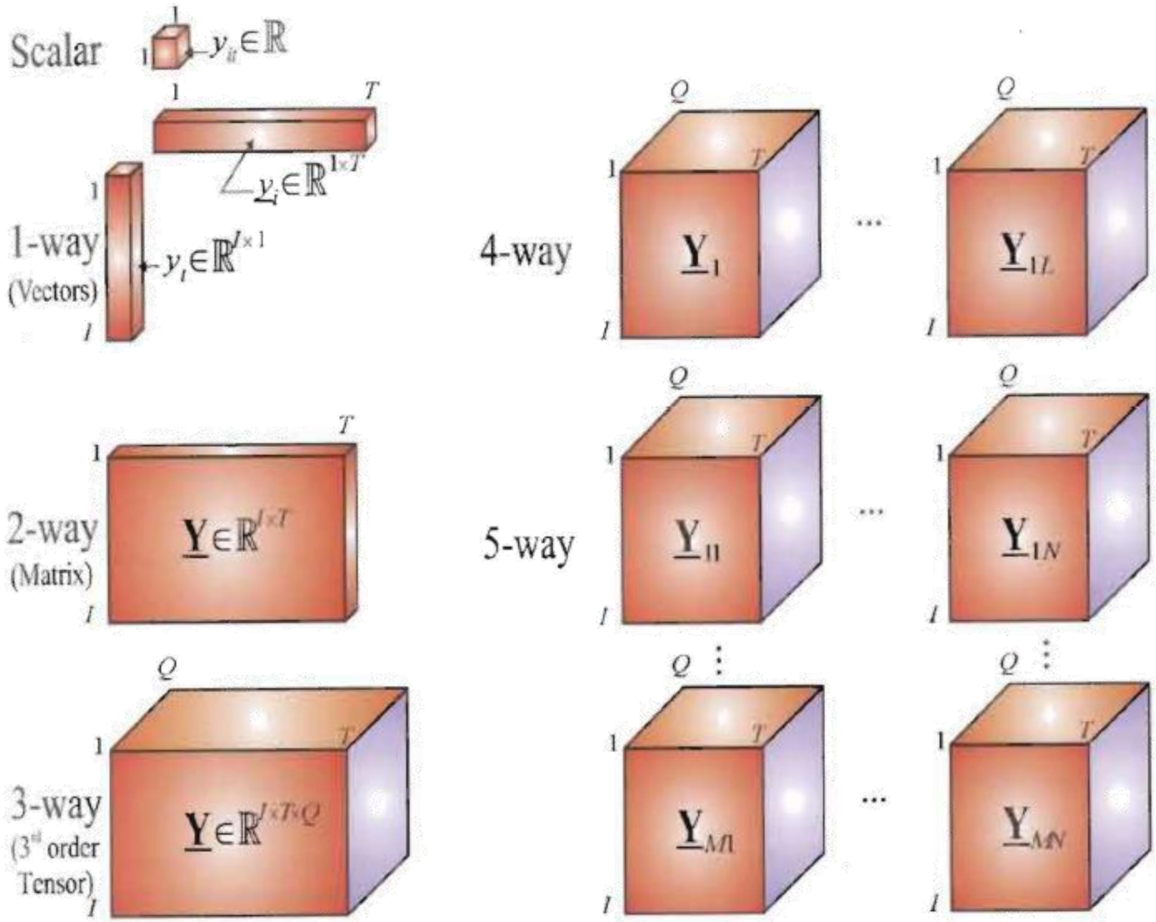
PARAFAC/CANDECOMP

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

Slides by Michalis Giannopoulos

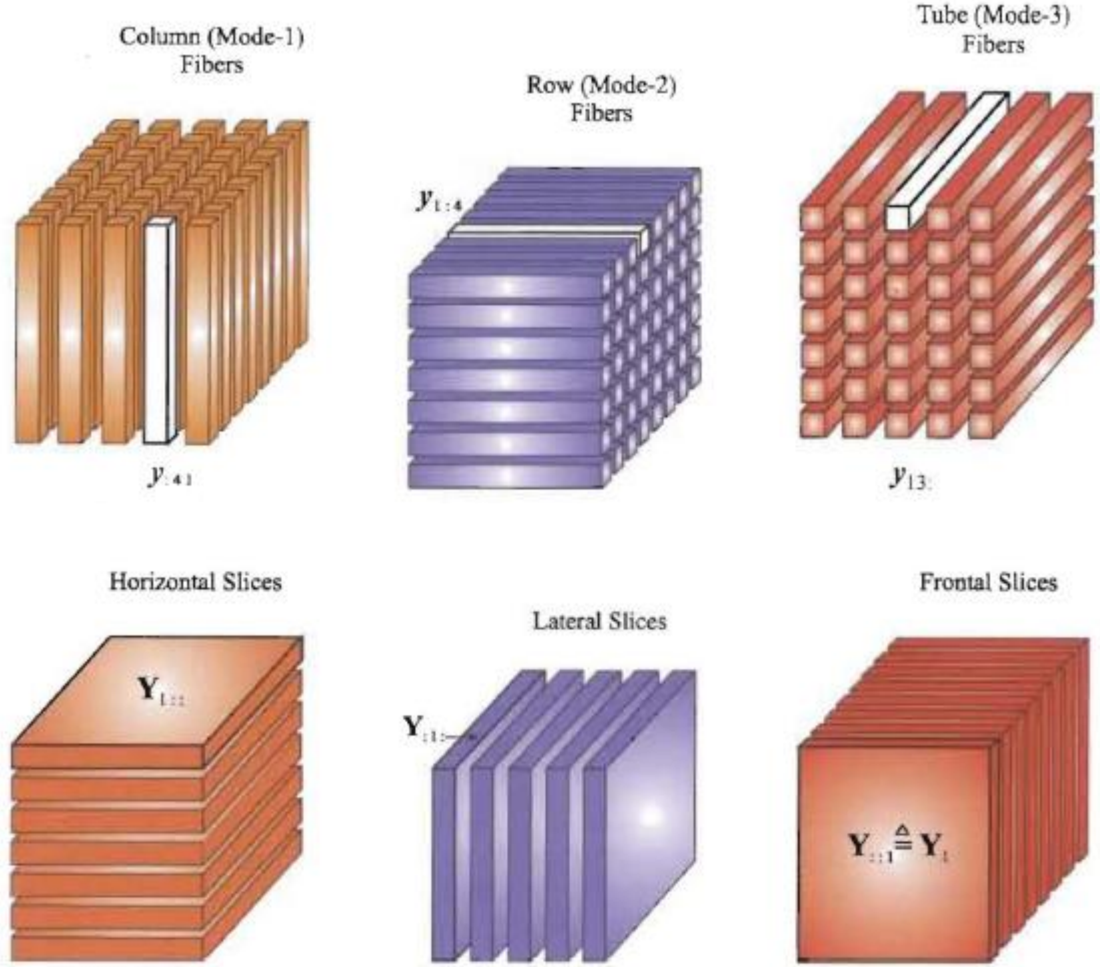
[1] F. L. Hitchcock, “The expression of a tensor or a polyadic as a sum of products”, Studies in Applied Mathematics, 6(1-4):164-189, 1927.
[2] R. B. Cattell, ”Parallel proportional profiles and other principles for determining the choice of factors by rotation”, Psychometrika, 9(4):267-283, 1944.

Tensors



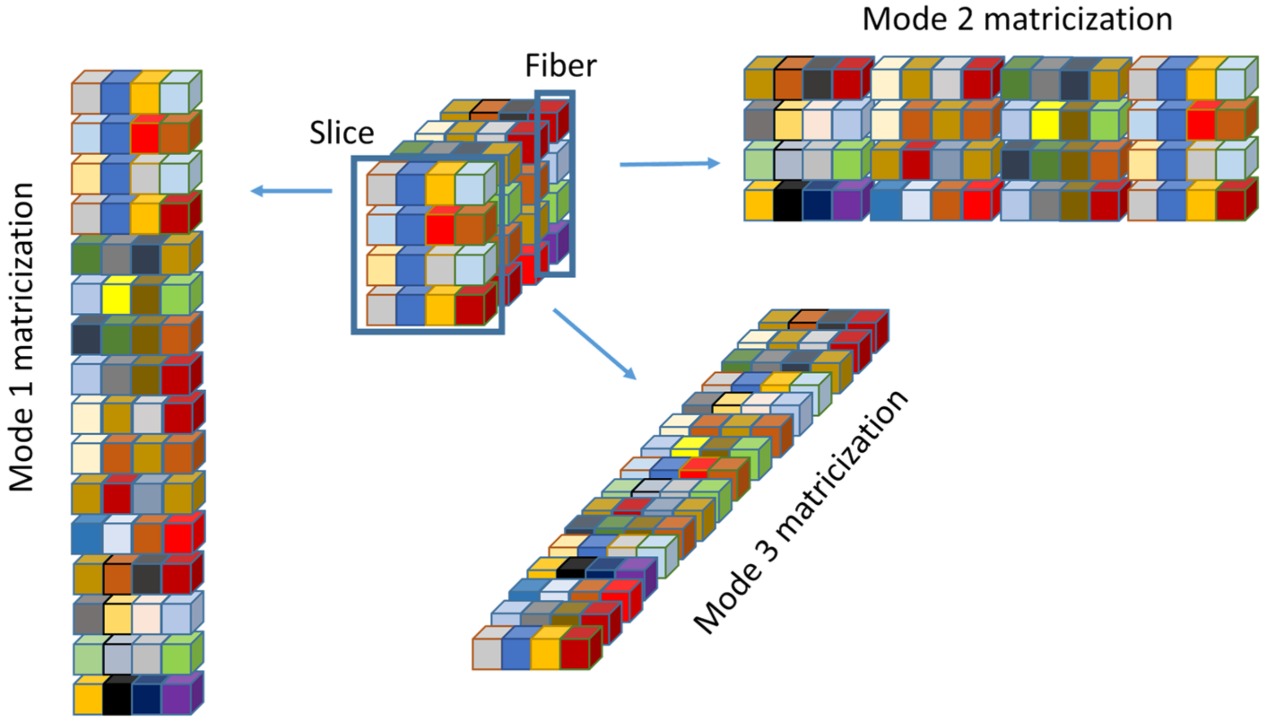
Includes materials from: Introduction to tensor, tensor factorization and its applications, by Mu Li, iPAL Group Meeting, Sept. 17, 2010

Fiber and slice



Tensor unfoldings: Matricization and vectorization

- Matricization: convert a tensor to a matrix



Tensor Mode-n Multiplication

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

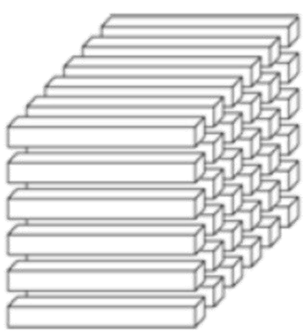
- Tensor x Matrix

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

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

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

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

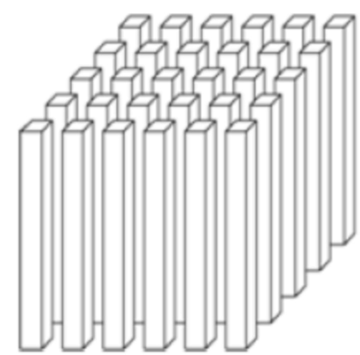


- Tensor x Vector

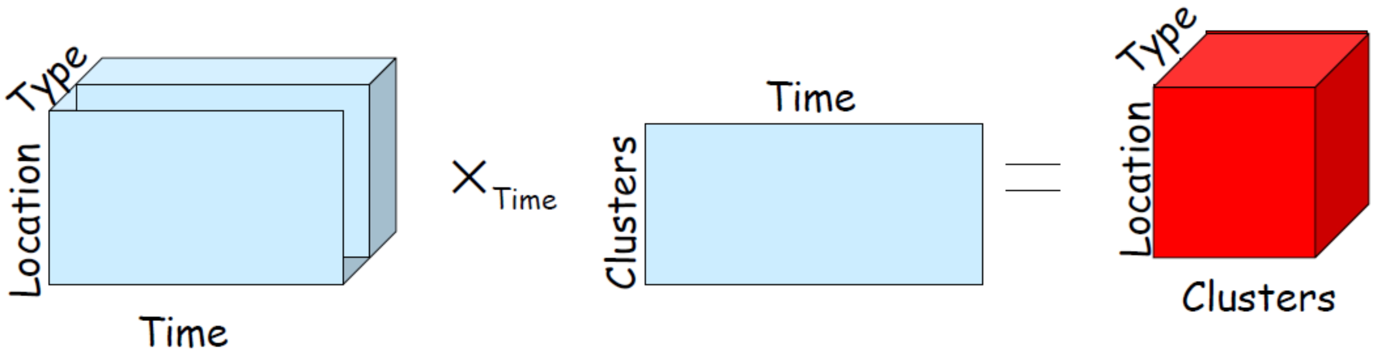
$$\mathbf{Y} = \mathcal{X} \,\bar{\times}_1\, \mathbf{a} \in \mathbb{R}^{J \times K}$$

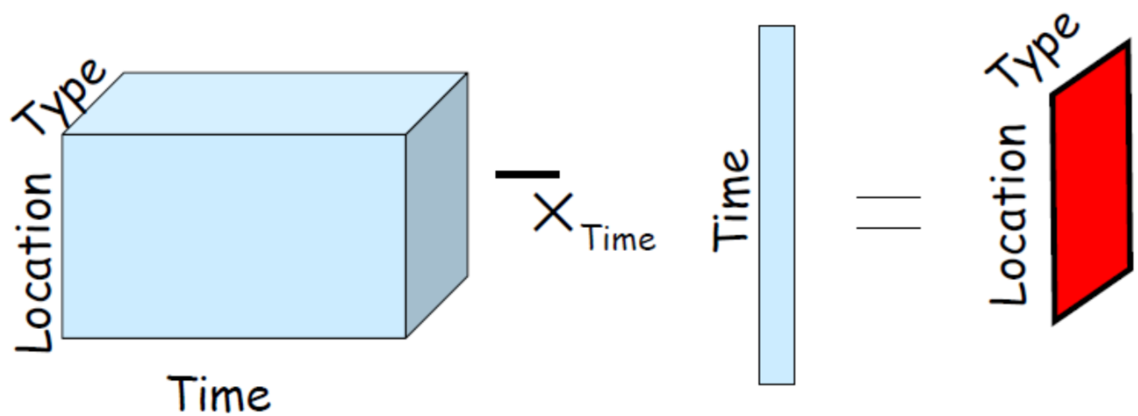
$$y_{jk} = \sum_i x_{ijk} \, a_i$$

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



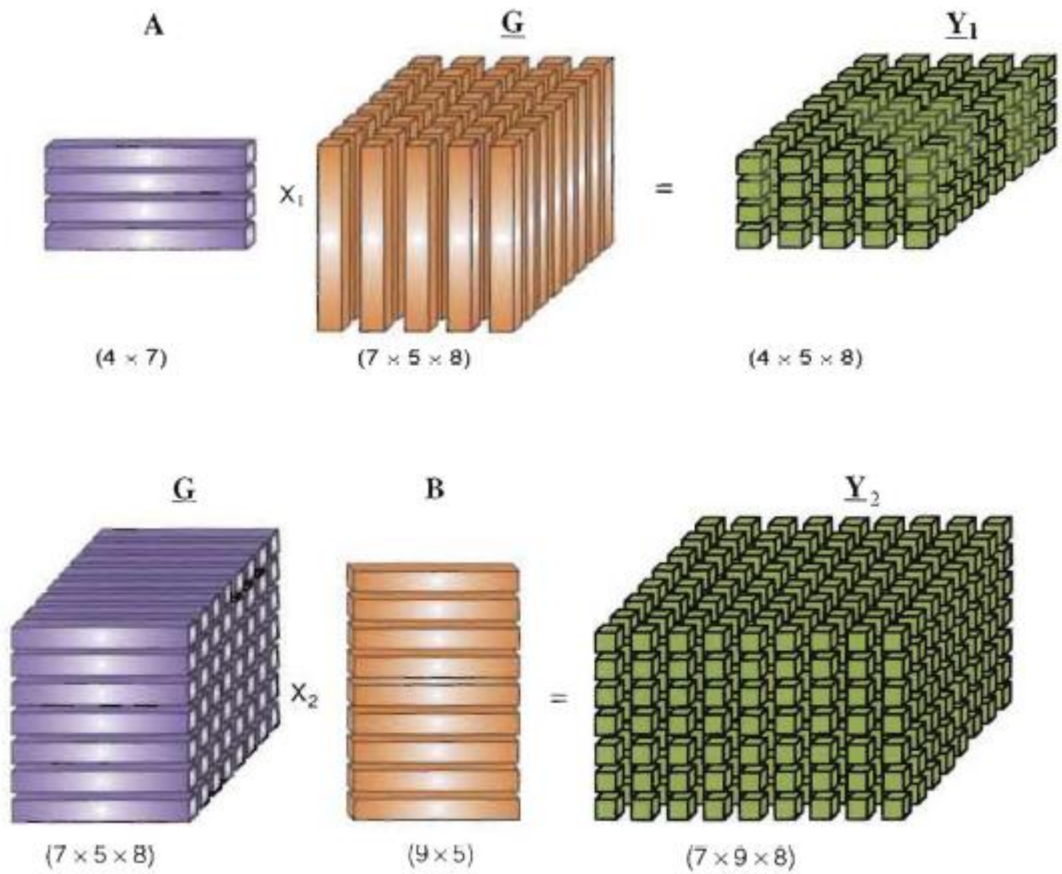
Examples


Type
Location
Time
×Time
Time
Clusters
Type
Location
Clusters


Type
Location
Time
×Time
Time
Type
Location

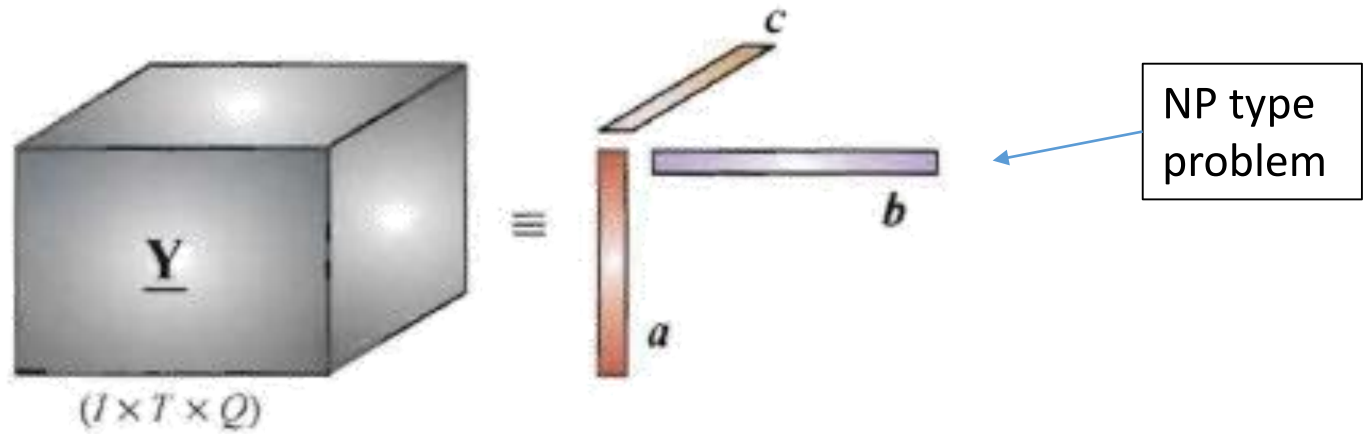
Tensor multiplication: the n-mode product: multiplied by a matrix

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



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

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

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

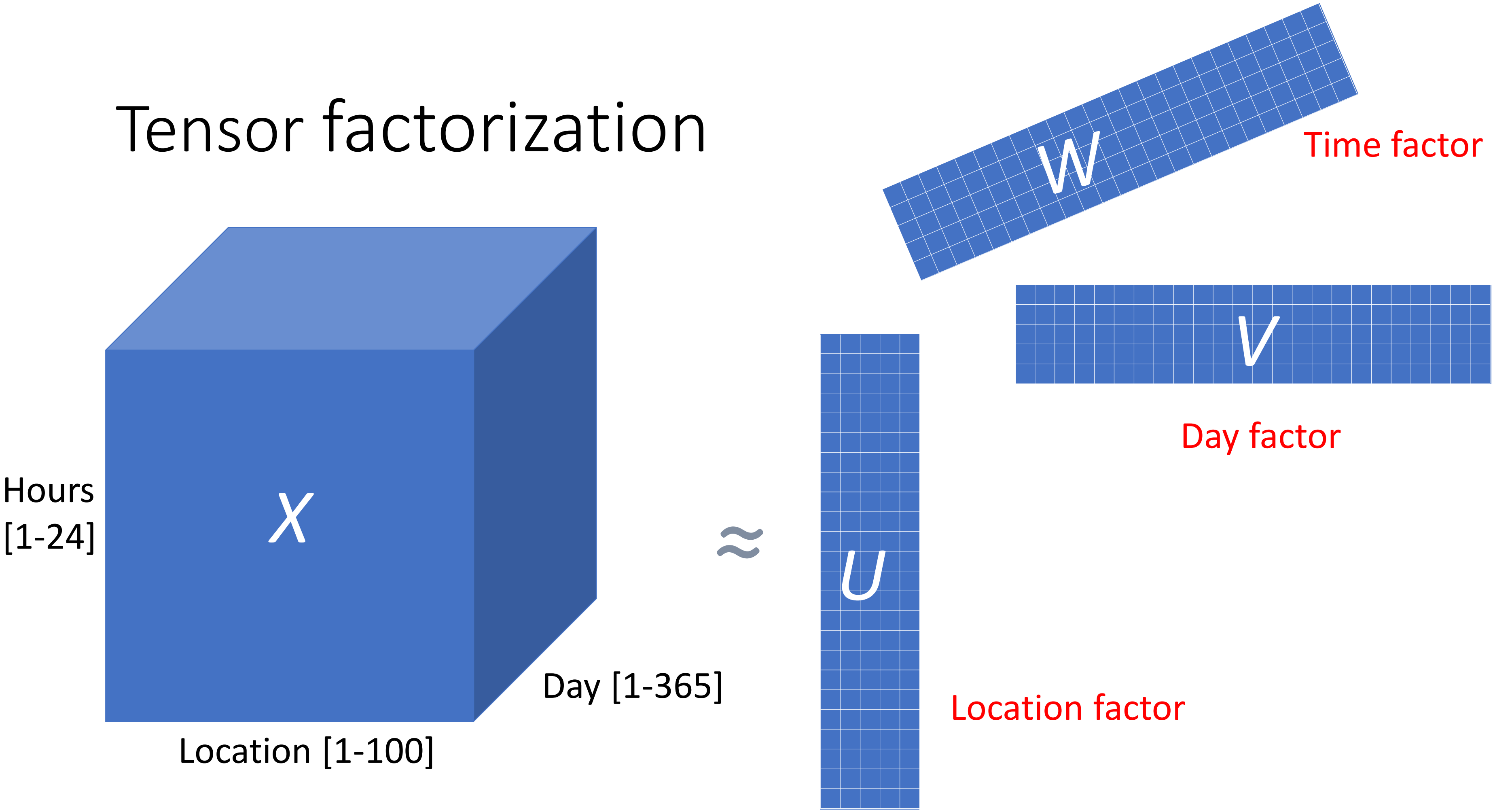
If $\mathbf{D} = \text{diag}(\mathbf{d})$, then $\text{vec}(\mathbf{ADC}) = (\mathbf{C}^T \odot \mathbf{A})\mathbf{d}$

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 $C = \mathcal{A} \otimes \mathcal{B}$ 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 factorization



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

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

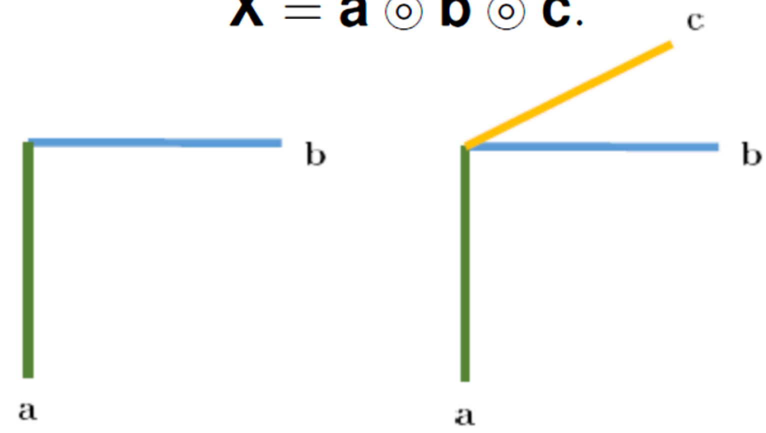
Tensor Rank

A **rank-1 matrix 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 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.,

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



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:

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

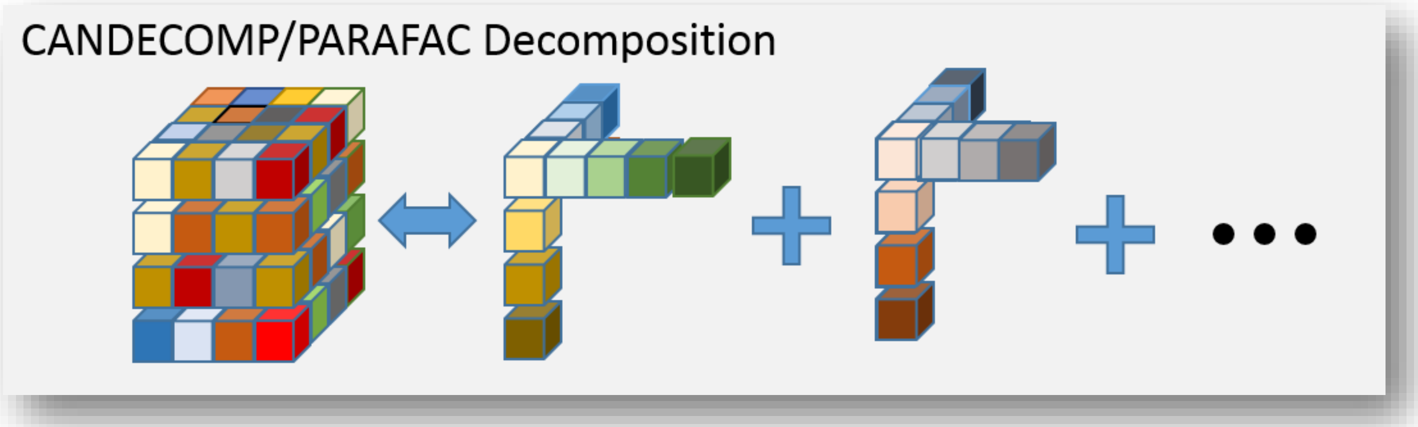
$$\mathbf{B} \leftarrow \arg\min_{\mathbf{B}} ||\mathbf{X}_2 - (\mathbf{C} \odot \mathbf{A})\mathbf{B}^T||_F^2,$$

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

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

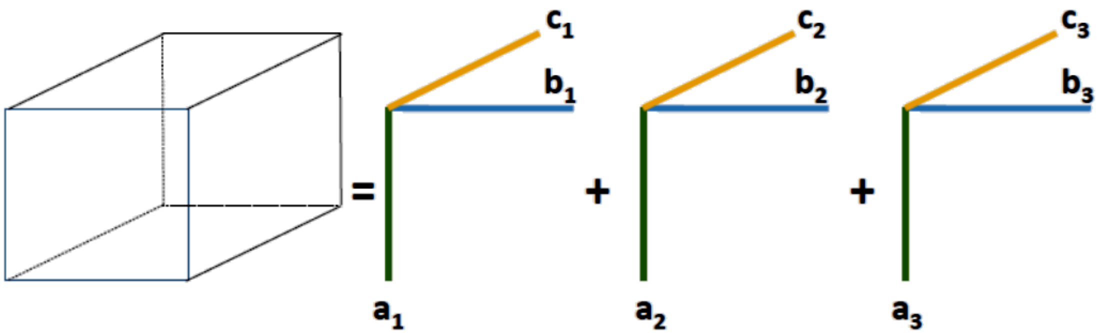
CANDECOMP/PARAFAC

- Rank 1 Tensor models



- CP factorization: $\mathfrak{X} \approx [\![\lambda \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$
- CP of tensor is unique under some general conditions

Uniqueness



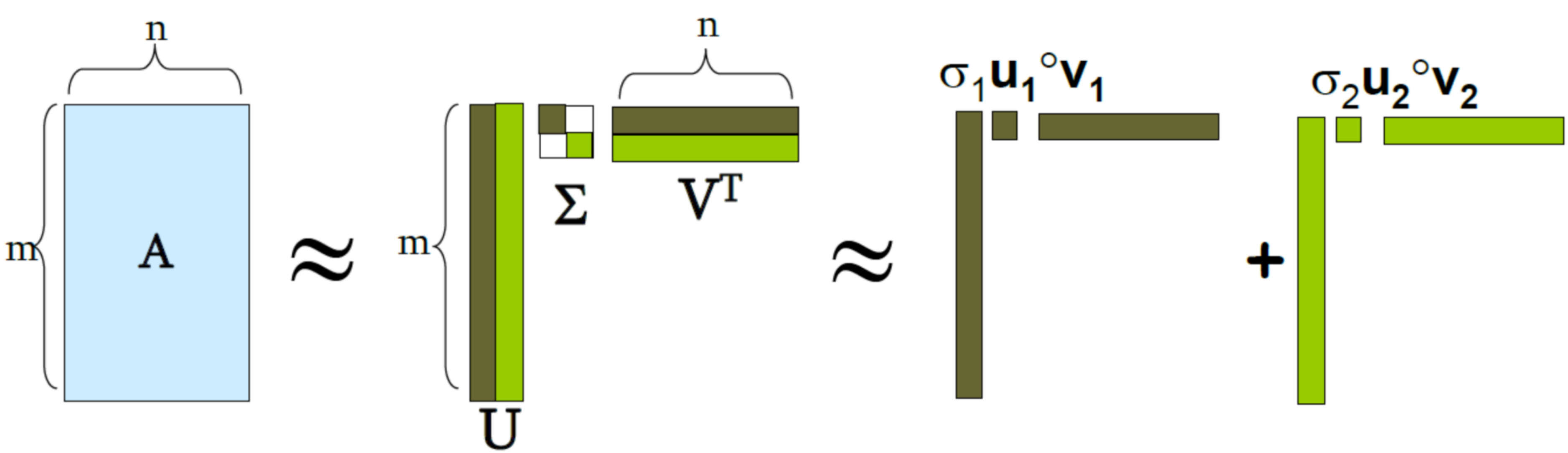
Given tensor **X** of rank *F*, its CPD is *essentially unique* iff the *F* rank-1 terms in its decomposition (the outer products or "chicken feet") are unique;

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.

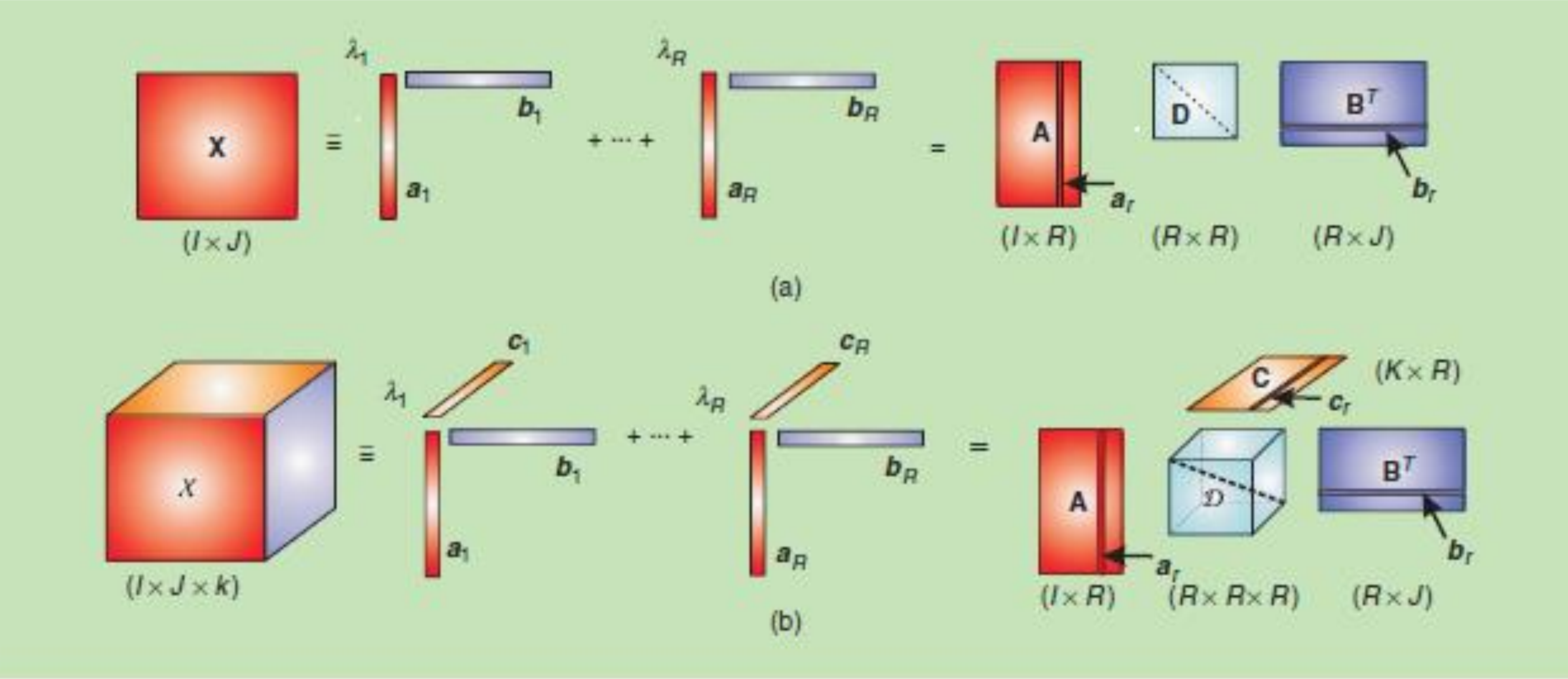
Reminder: SVD

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



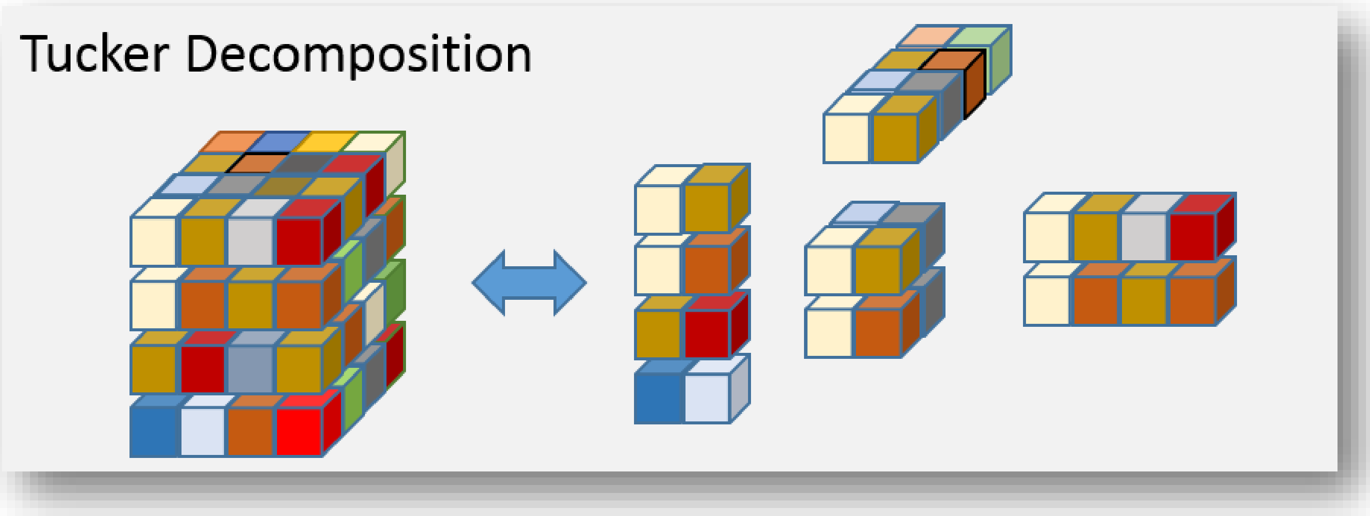
Relationship to SVD

- **The analogy between (a) dyadic decompositions and (b) polyadic decompositions**



TUCKER

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

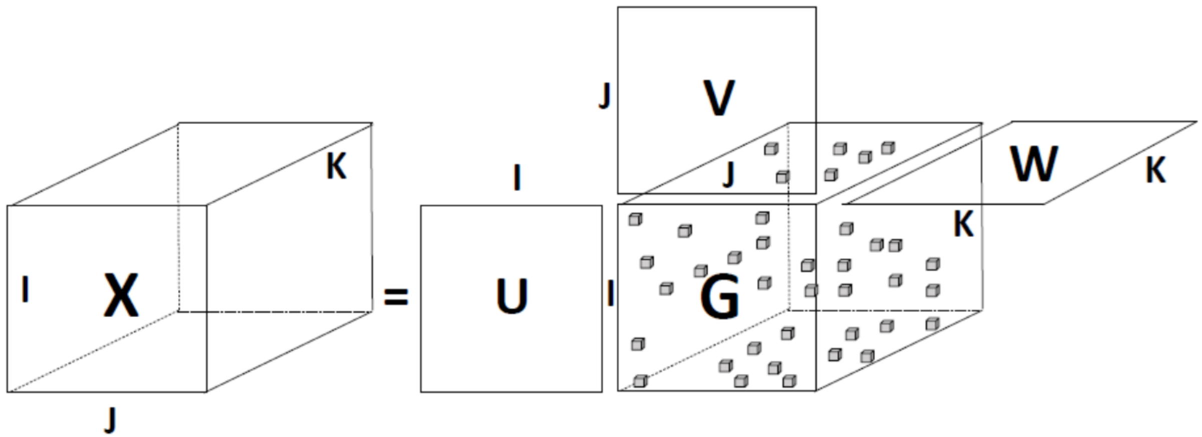


- The associated model-fitting problem is

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

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>**

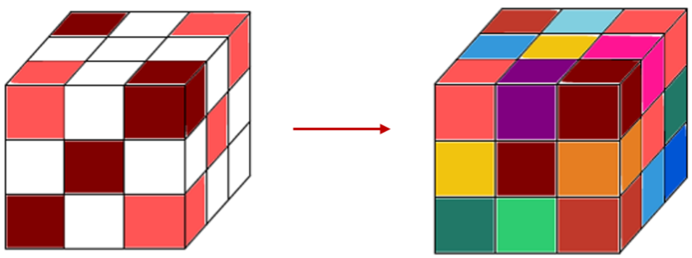
- <u>Lack of Uniqueness</u>:
  - “Transform” the core tensor $\mathscr{G}$
  - Apply the inverse ”transform” to the factor matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$
  - **Sometimes desired**: Sketching arithmetic solutions for Tucker decomposition computation

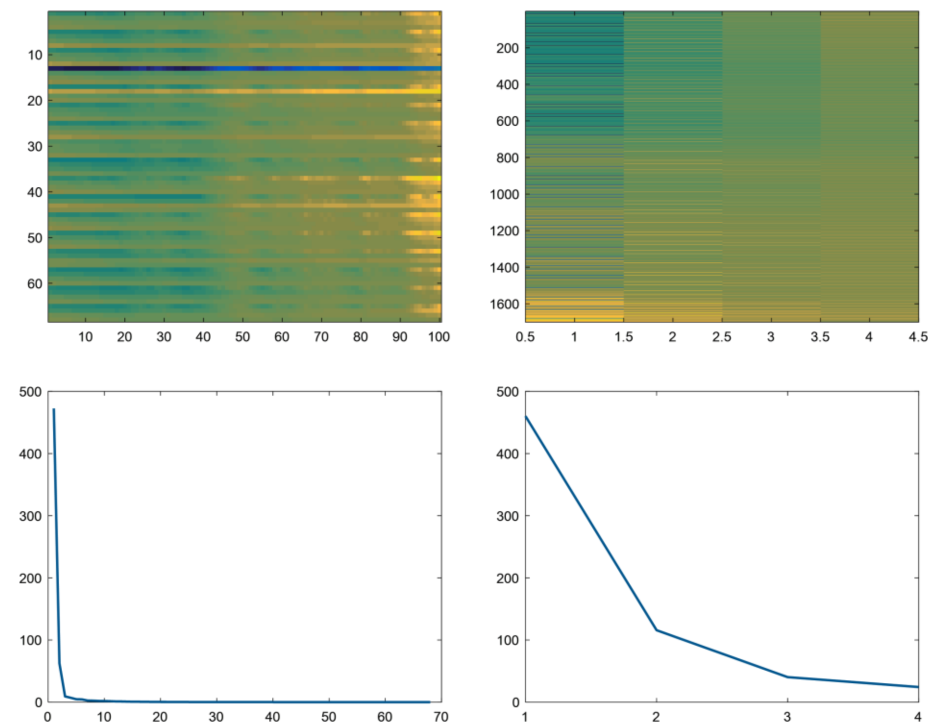
[1] R Coppi and S Bolasco. “Rank, decomposition, and uniqueness for 3-way and n-way arrays”, 1989.
[2] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. “A multilinear singular value decomposition”. SIAM journal on Matrix Analysis and Applications, 21(4):1253–1278, 2000.

FORTH Institute of Computer Science

Tensor Completion

- Low rank Tensor/Matrices





Extension: Tensor Completion

- 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]: $\|\mathscr{X}\|_* = \sum_{i=1}^{n} \alpha_i \|\mathbf{X}_{(i)}\|_*$

  $\alpha_i \geq 0$

  $\sum_{i=1}^{n} \alpha_i = 1$

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

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

[1] Ji Liu, P. Musialski, P. Wonka, and [illegible]sor completion for estimating missing values in visual data", IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1):208-220, 2013.

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 $\boldsymbol{\mathcal{M}}$

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

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

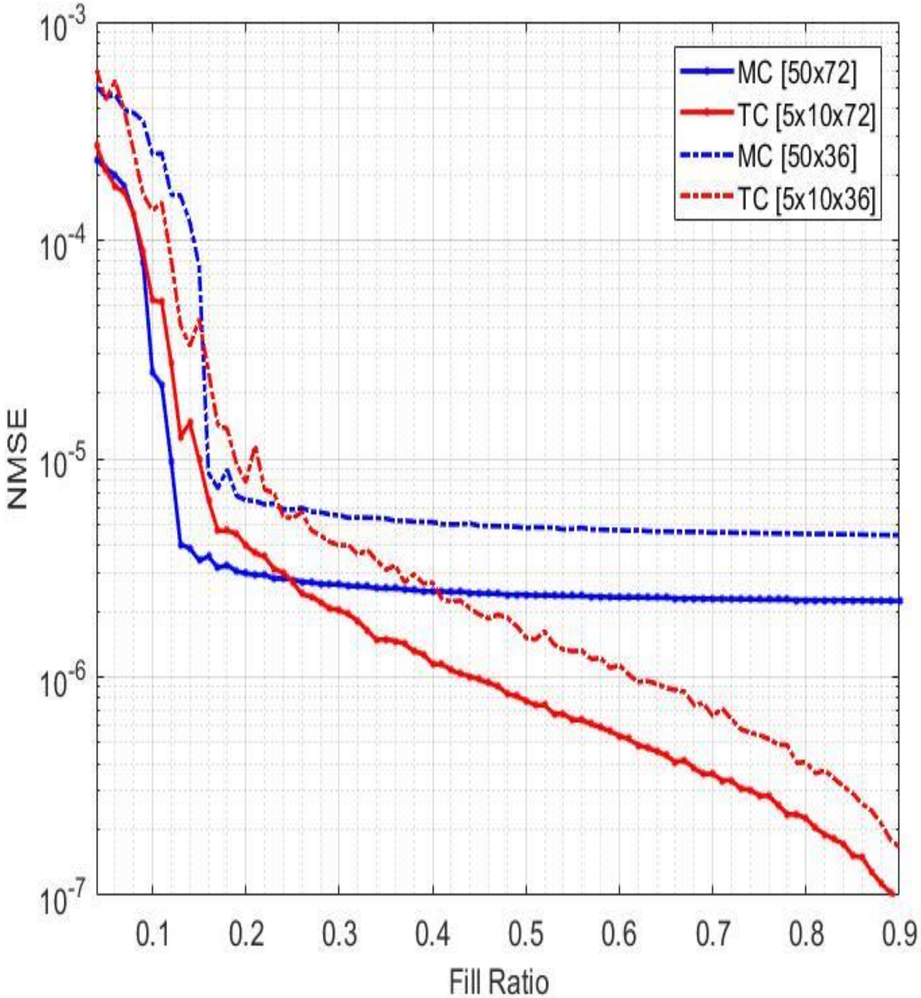
Matrices: 50×72, 50×36

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

FORTH
Institute of Computer Science

Effects of Data Structuring

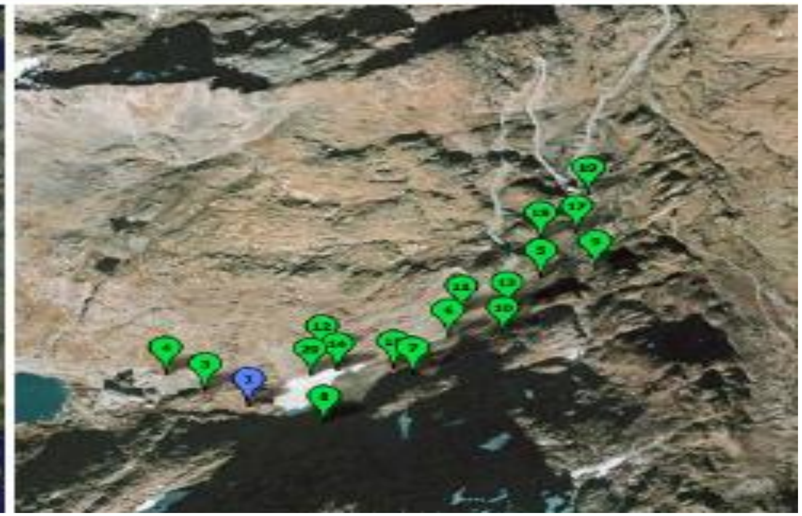
- Higher fill-ratio
  - ➢ Better reconstruction quality quantified
- <u>Regardless matrix/tensor size</u>
  - ➢ TC outperforms MC from low fill-ratio regimes
- <u>NMSE convergence</u>
  - ➢ MC reaches a plateau
  - ➢ TC decreases (nearly) monotonically



FORTH
Institute of Computer Science

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

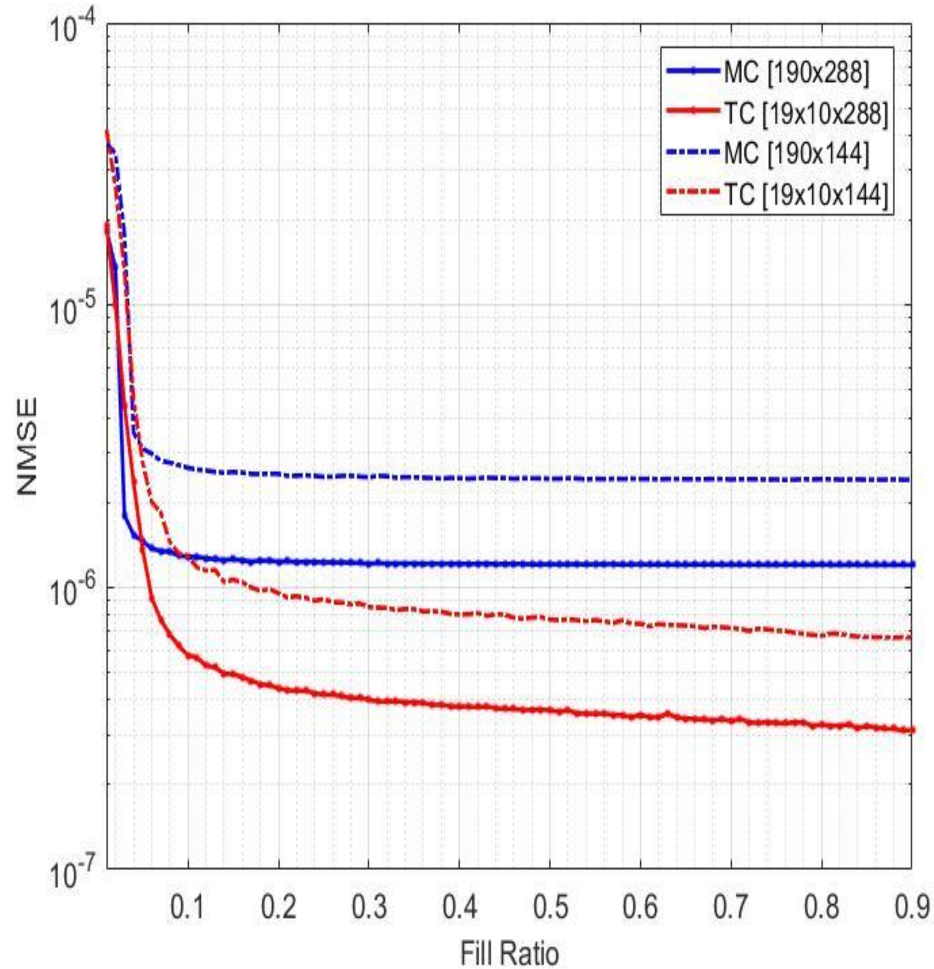
Matrices: 190×288, 190×144

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

[1] Guillermo Barrenetxea, Francois Ingelrest, Gunnar Schaefer, Martin Vetterli, Olivier Couach, and Marc Parlange, "Sensorscope: Out-of-the-box environmental monitoring", In Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on, pages 332-343, IEEE, 2008.

Effects of Data Structuring

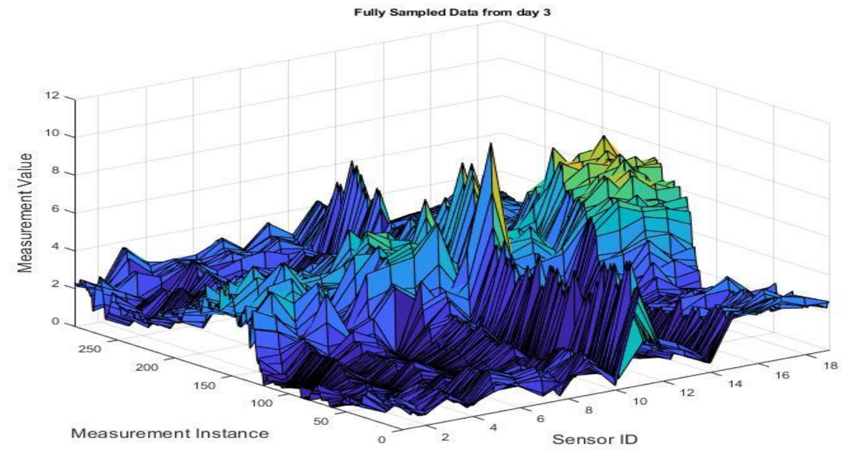
- Higher fill-ratio
  - ➢ Better reconstruction quality quantified
- <u>Larger Dataset</u>
  - ➢ TC outperforms MC <u>from lower</u> fill-ratio regimes
- <u>NMSE convergence</u>
  - ➢ MC reaches a plateau
  - ➢ TC keeps decreasing

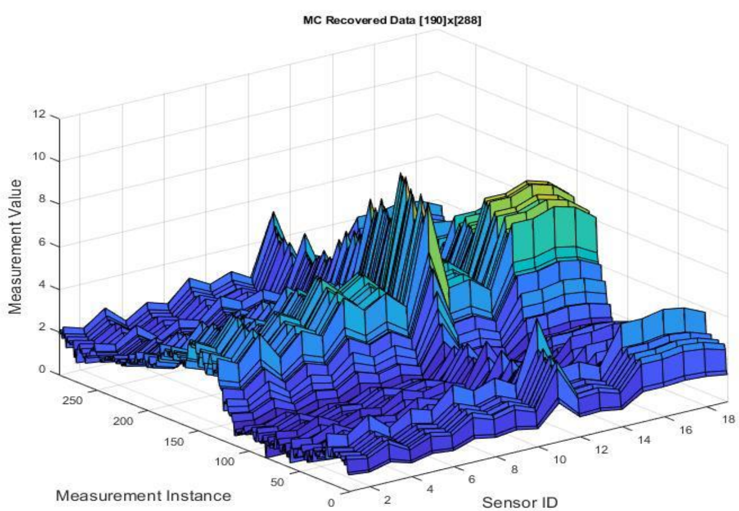


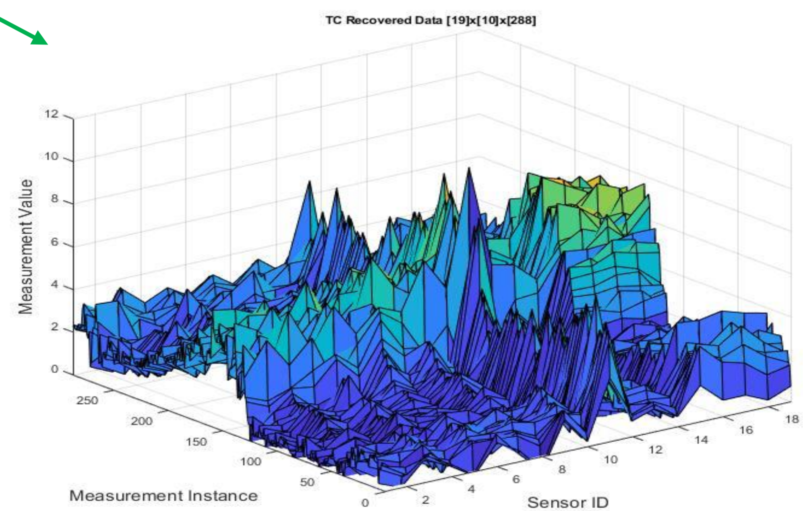
FORTH Institute of Computer Science

Effects of Fill-Ratio

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



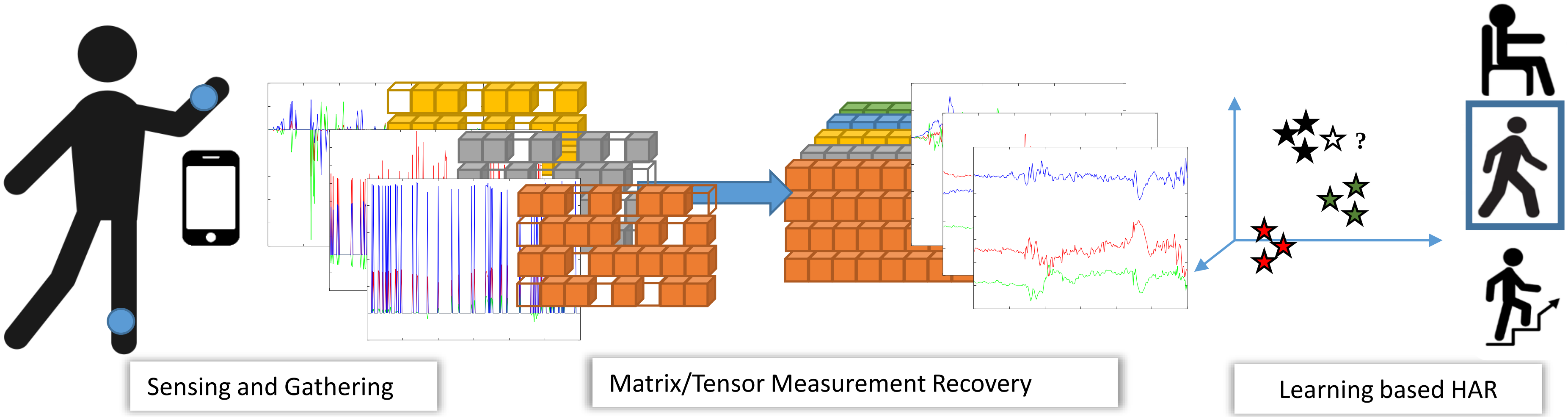


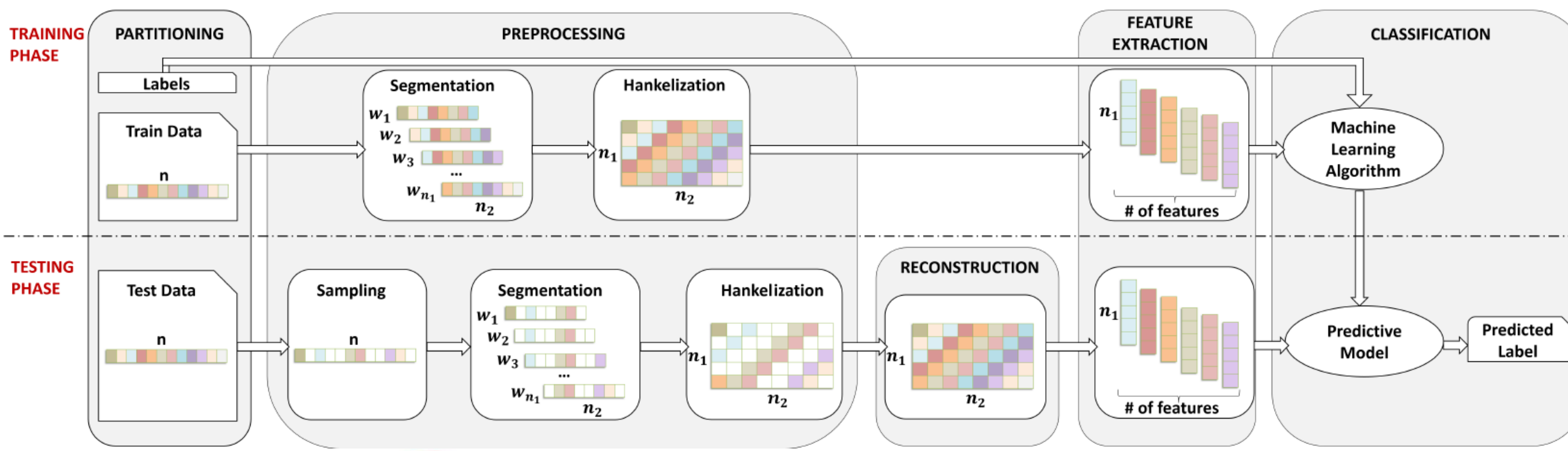


FORTH Institute of Computer Science

WSNs for Human Activity Recognition

Problem formulation



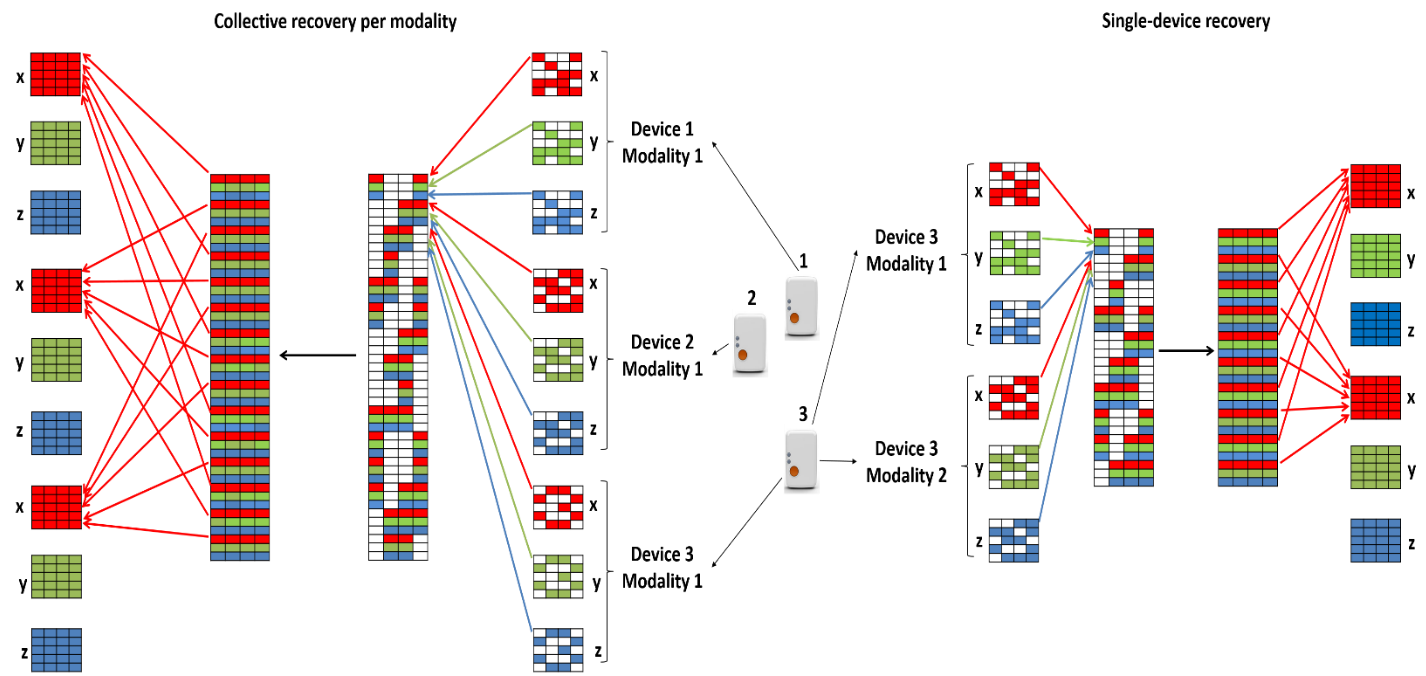


FORTH
Institute of Computer Science

Single-device vs collective recovery: matrices

Scenario 2
Collective per modality

Scenario 1
Single-device

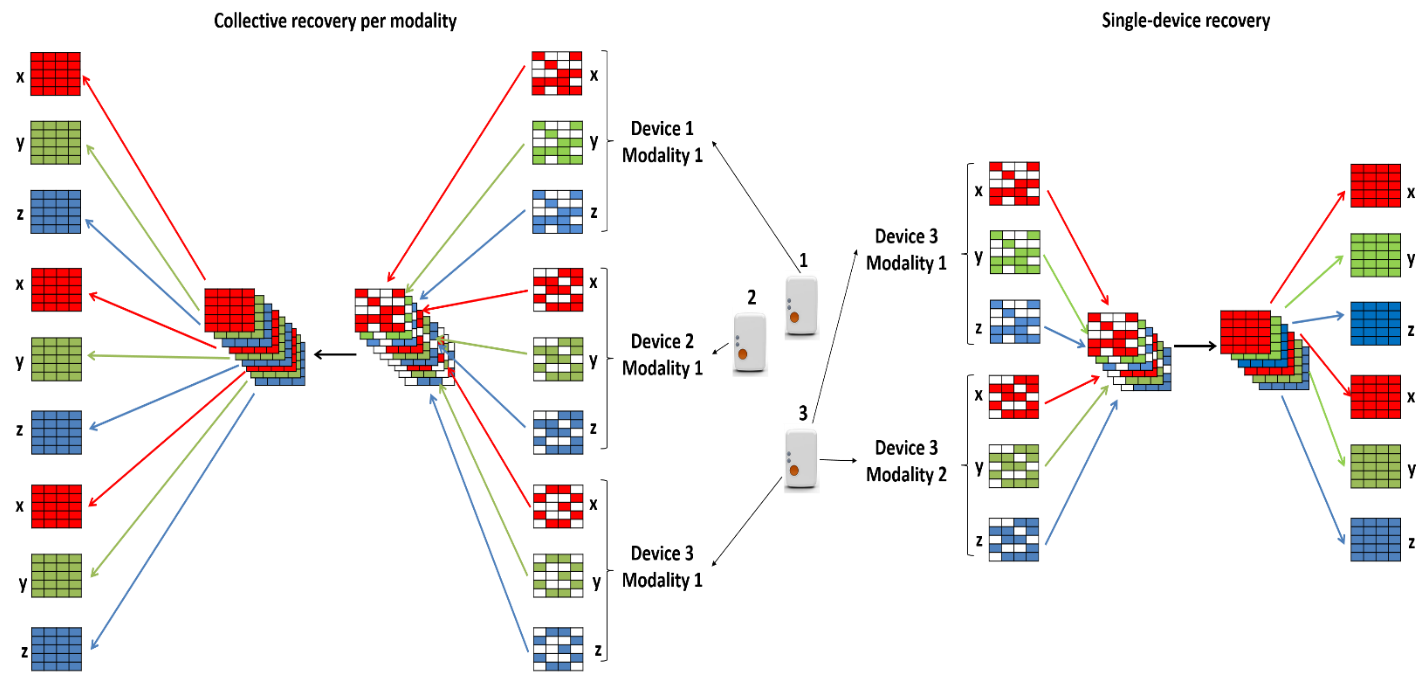


Scenario 3: Overall collective recovery structured similarly

Single-device vs collective recovery: tensors

Scenario 2
Collective per modality

Scenario 1
Single-device

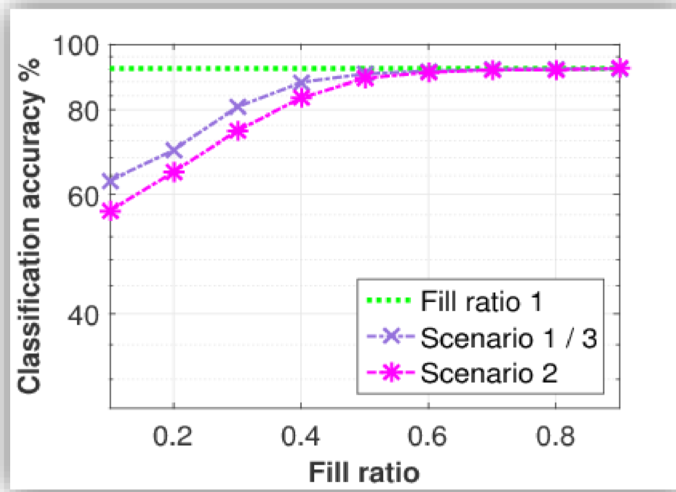


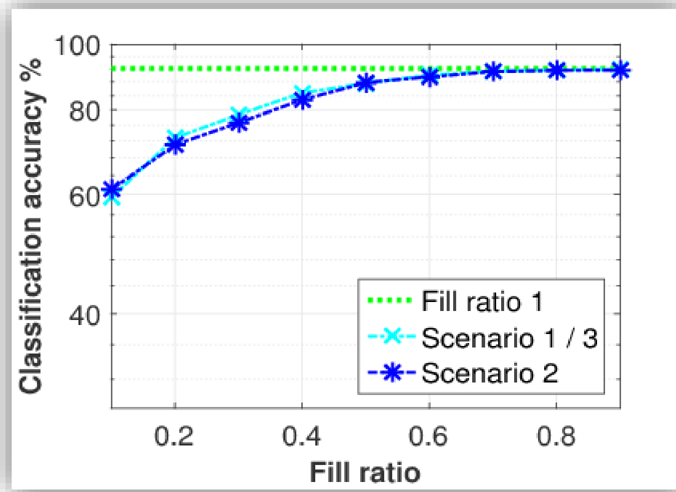
Scenario 3: Overall collective recovery structured similarly

Some results

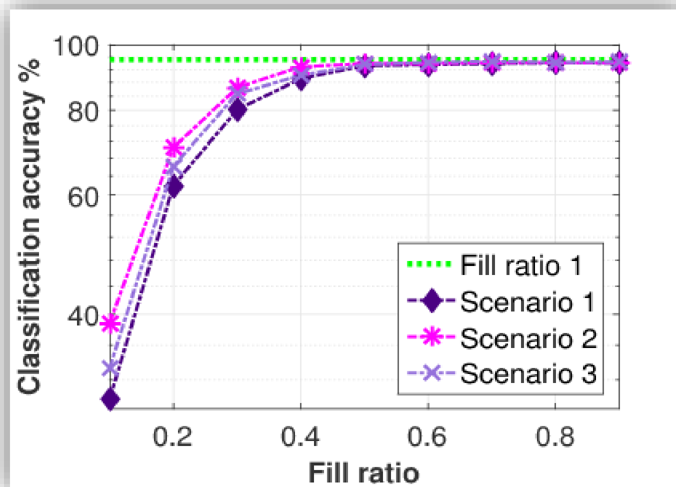
Matrix Completion | **Tensor Completion**

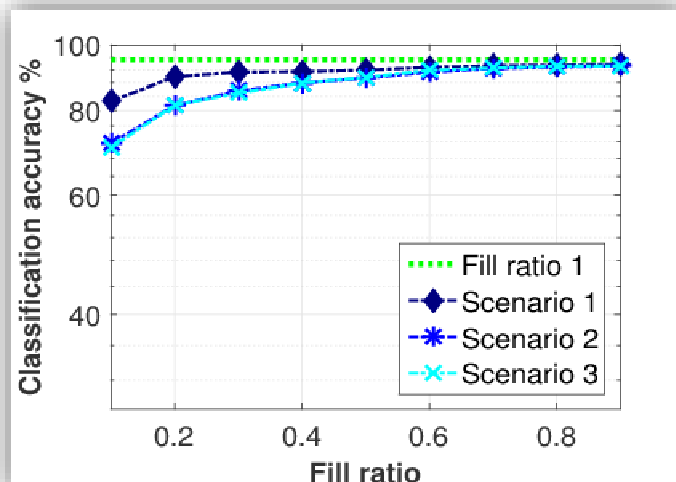
HAR





mHealth





Reading Material

- Davenport, Mark A., and Justin Romberg. "An overview of low-rank matrix recovery from incomplete observations." IEEE Journal of Selected Topics in Signal Processing 10.4 (2016): 608-622.
- Cichocki, Andrzej, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. "Tensor decompositions for signal processing applications: From two-way to multiway component analysis." *IEEE Signal Processing Magazine* 32, no. 2 (2015): 145-163.
- Savvaki, Sofia, Grigorios Tsagkatakis, Athanasia Panousopoulou, and Panagiotis Tsakalides. "Matrix and Tensor Completion on a Human Activity Recognition Framework." *IEEE journal of biomedical and health informatics* 21, no. 6 (2017): 1554-1561.