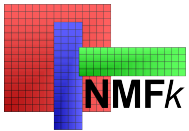
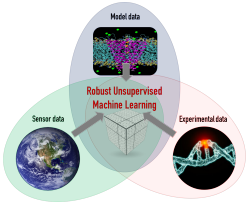
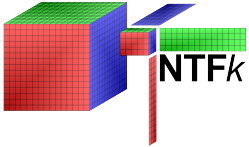
Unsupervised Machine Learning

Velimir V. Vesselinov (monty) (vvv@lanl.gov)
Earth and Environmental Sciences Division, Los Alamos National Laboratory, NM, USA
http://tensors.lanl.gov







Why unsupervised Machine Learning (ML)?

- **Supervised** ML: requires “labeling” (prior categorization (knowledge) about the processed data)

  Example: Recognizes images of cats and dogs after extensive training; but cannot recognize horses if not trained

  Cannot discover something that we do not know already

- **Unsupervised** ML: extracts hidden features (signals) in the processed data without any prior information (**exploratory analysis**; **data-driven science**)

  Example: Identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.)

Why unsupervised Machine Learning (ML)?

- **Supervised** ML: requires “labeling” (prior categorization (knowledge) about the processed data)

  Example: Recognizes images of cats and dogs after extensive training; but cannot recognize horses if not trained

  Cannot discover something that we do not know already

- **Unsupervised** ML: extracts hidden features (signals) in the processed data without any prior information (**exploratory analysis**; **data-driven science**)

  Example: Identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.)

Why unsupervised Machine Learning (ML)?

- **Supervised** ML: requires “labeling” (prior categorization (knowledge) about the processed data)

  Example: Recognizes images of cats and dogs after extensive training; but cannot recognize horses if not trained

  Cannot discover something that we do not know already

- **Unsupervised** ML: extracts hidden features (signals) in the processed data without any prior information (**exploratory analysis**; **data-driven science**)

  Example: Identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.)

Why unsupervised Machine Learning (ML)?

- **Supervised** ML: requires “labeling” (prior categorization (knowledge) about the processed data)

  Example: Recognizes images of cats and dogs after extensive training; but cannot recognize horses if not trained

  Cannot discover something that we do not know already

- **Unsupervised** ML: extracts hidden features (signals) in the processed data without any prior information (**exploratory analysis**; **data-driven science**)

  Example: Identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.)

Why unsupervised Machine Learning (ML)?

- **Supervised** ML: requires “labeling” (prior categorization (knowledge) about the processed data)

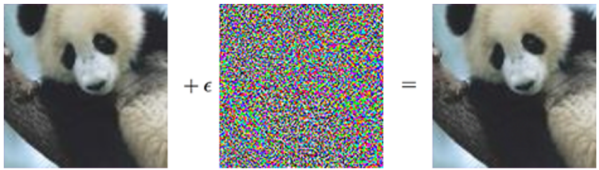
  Example: Recognizes images of cats and dogs after extensive training; but cannot recognize horses if not trained

  Cannot discover something that we do not know already

- **Unsupervised** ML: extracts hidden features (signals) in the processed data without any prior information (**exploratory analysis**; **data-driven science**)

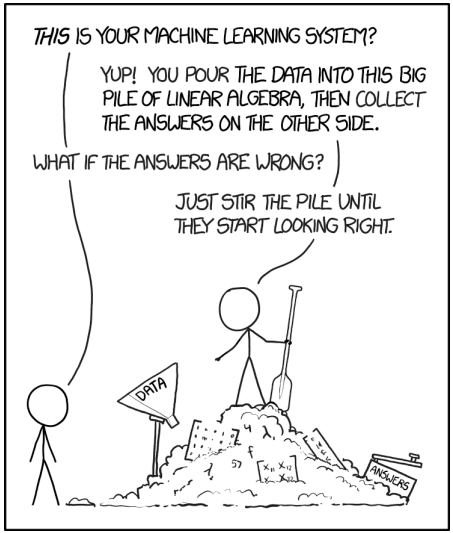
  Example: Identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.)

- **Supervised** ML
  - introduces subjectivity (through the labeling process)
  - does not provide insights why horses are different from dogs / cats
  - cannot make predictions (that we do not know already)
  - requires huge training (labeled) datasets
  - we do not know why it works
  - is impacted by “adversarial examples”



$\Rightarrow$ major limitations of the **supervised** ML methods for **science** applications

Supervised Machine Learning (xkcd)



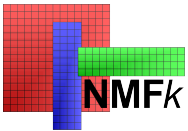
Unsupervised Machine Learning Applications

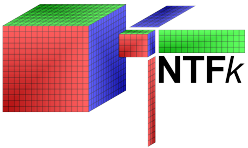
- **Data Analytics**: Identify signals (features) in datasets
  - Feature extraction (**FE**)
  - Blind source separation (**BSS**)
  - Detection of disruptions / anomalies
  - Image recognition
  - Discover unknown dependencies and phenomena
  - Develop reduced-order/surrogate models
  - Guide development of physics models representing the data
  - Make predictions
  - Optimize data acquisition
  - “Label” datasets for supervised ML analyses

- **Model Diagnostics**: Identify processes (features) in model outputs
  - Separate processes (inseparable during modeling)
  - Model reduction
  - Identify dependencies between model inputs and outputs
  - Discover unknown dependencies and phenomena
  - Make predictions

Nonnegative Matrix/Tensor Factorization

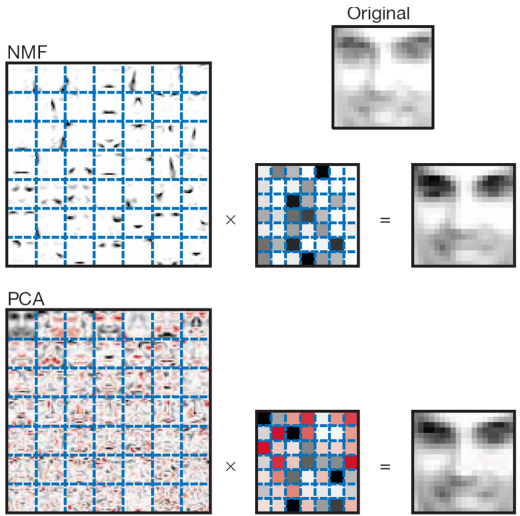
- Novel LANL-patented, open-source, unsupervised Machine Learning (ML) methods and computational techniques
- Based in matrix/tensor factorization coupled with custom k-means clustering and nonnegativity/sparsity constraints:
  - NMFk: Nonnegative **Matrix** Factorization
  - NTFk: Nonnegative **Tensor** Factorization
  - `https://github.com/TensorDecompositions`
- Capable to efficiently process large datasets (GB/TB's) utilizing GPU's & TPU's (julia, TensorFlow, PyTorch, MXNet)





Why nonnegativity?

- NMF vs PCA (Lee & Seung, 1999)
- NMF: Nonnegative Matrix Factorization
- PCA: Principal Component Analysis

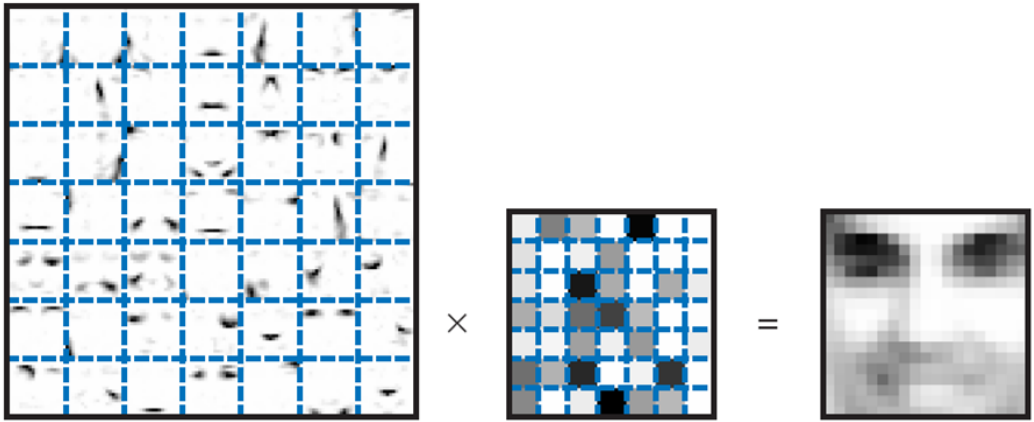


Nonnegativity constraints provide meaningful and interpretable results (+sparsity)

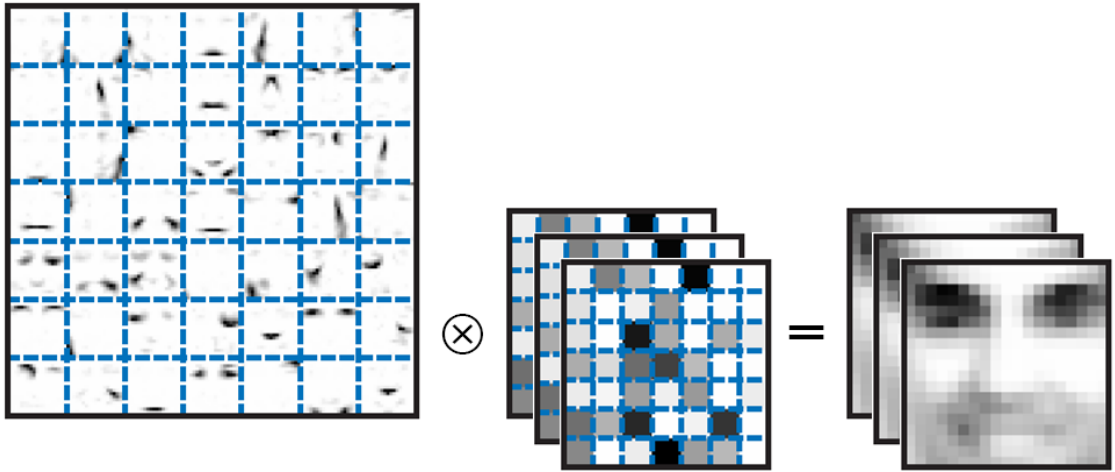
Why tensors?

- **Tensors** (multi-dimensional/multi-modal/multi-way datasets) are everywhere:
  - observational data are typically a 5-D tensor (x, y, z, t, attributes)
  - model outputs are typically a 5-D tensor (x, y, z, t, attributes)
  - data dependency to N parameters will form a $(N + 5)$-D tensor

NMF: Nonnegative Matrix Factorization



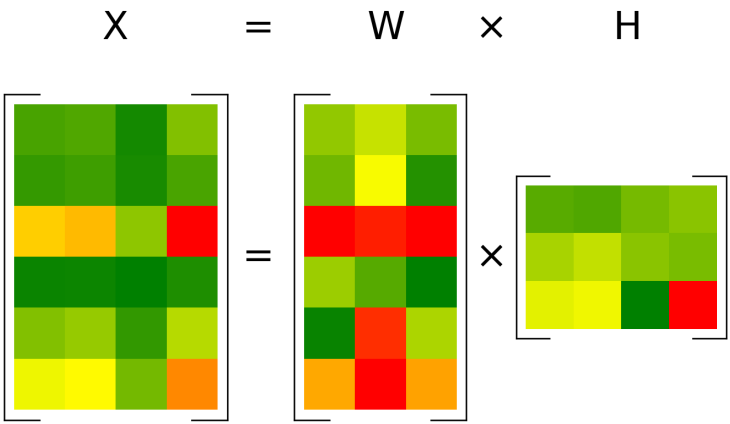
NTF: Nonnegative Tensor Factorization



Why not SVD (PCA)?

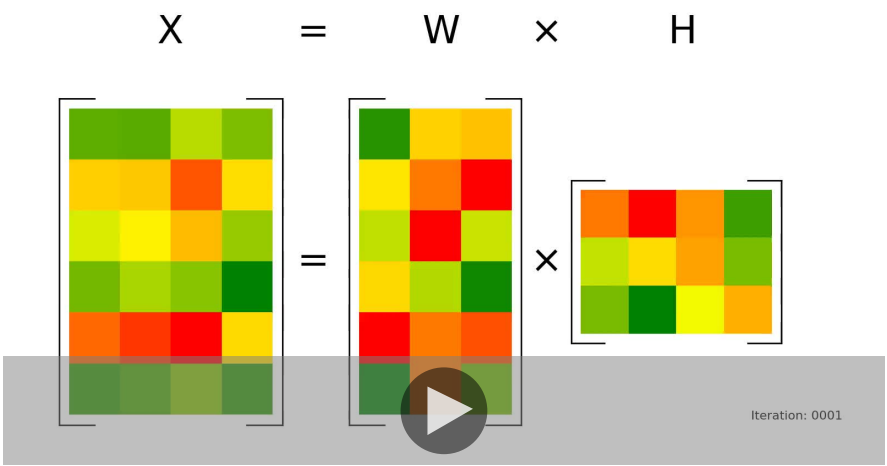
- **SVD** cannot be applied on multi-dimensional (tensor) datasets
- **HOSVD** cannot tell us the number of features (signals)

NMFk example: Truth

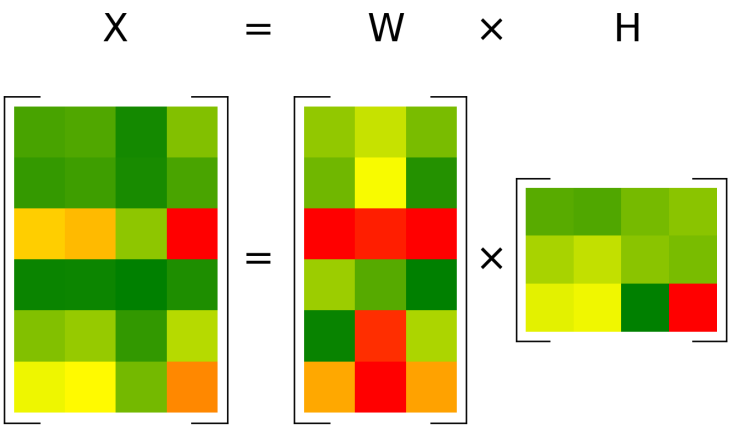


24 knowns $(6 \times 4) \rightarrow 30$ unknowns $(6 \times 3) + (3 \times 4)$

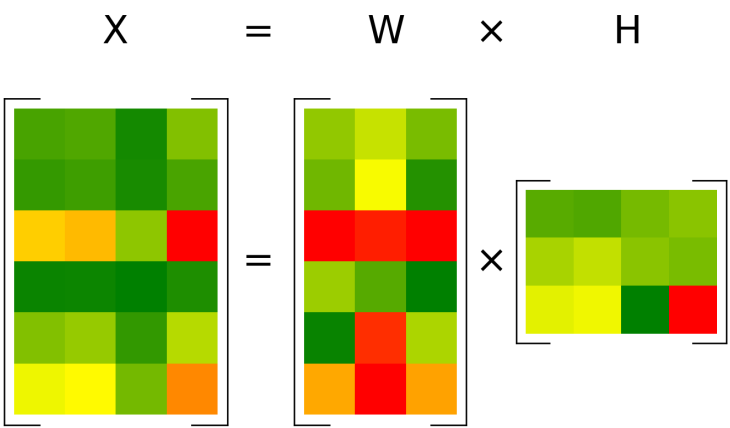
NMFk example: Factorization process



NMFk example: Estimate



NMFk example: Truth



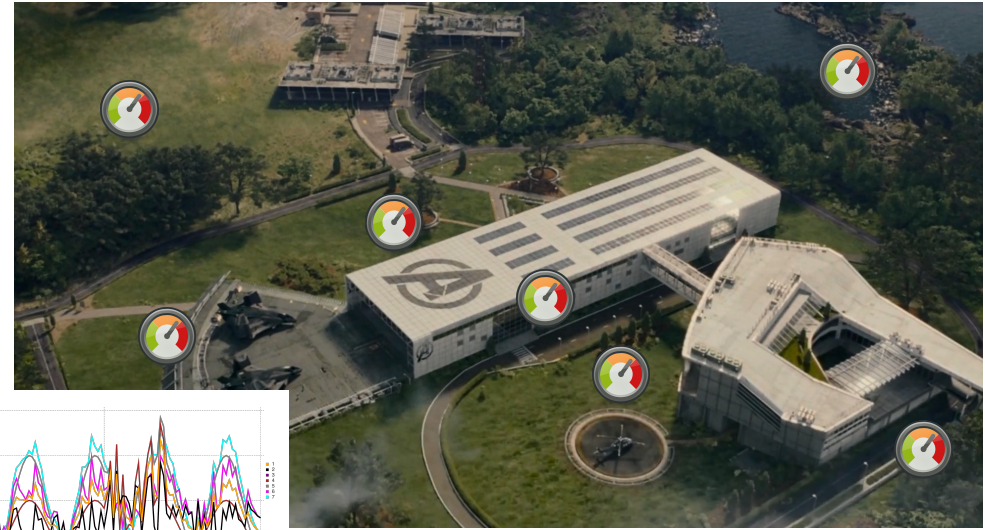
NMFk: Extraction of features / anomalies / disruptions

NMFk: Extraction of features / anomalies / disruptions

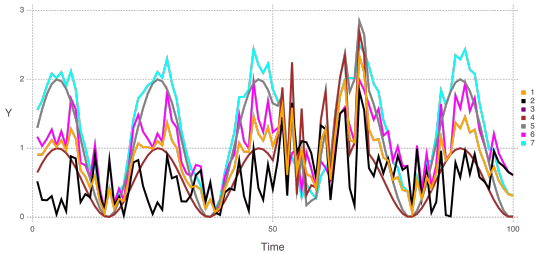
NMFk: Extraction of features / anomalies / disruptions

NMFk: Extraction of features / anomalies / disruptions

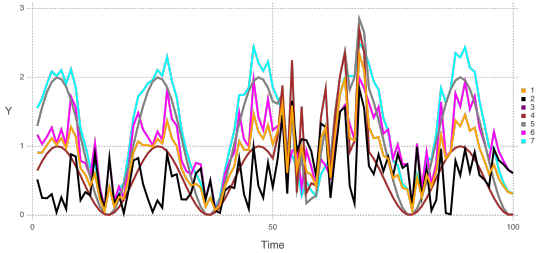
NMFk: Extraction of features / anomalies / disruptions

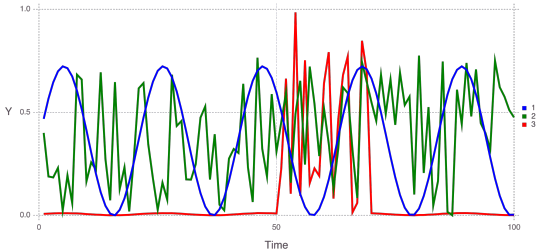


NMFk: Extraction of features / anomalies / disruptions

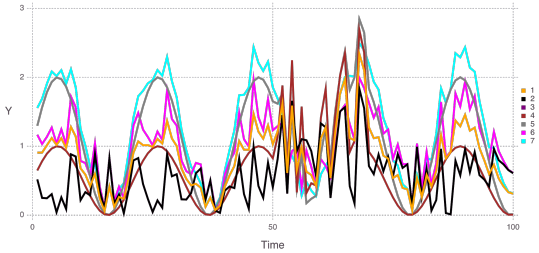


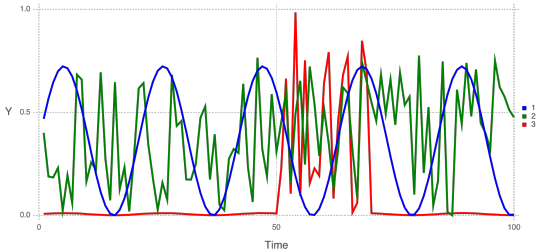
NMFk: Extraction of features / anomalies / disruptions

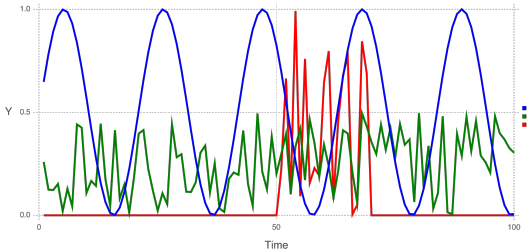




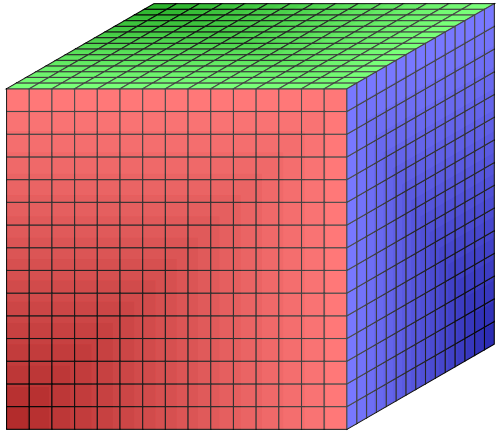
NMFk: Extraction of features / anomalies / disruptions





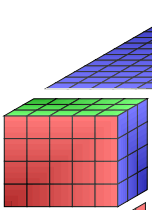


Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(3,4,5)

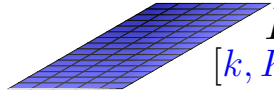


X
$[K, M, N]$

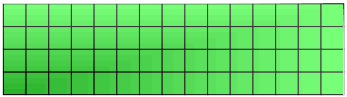
$\approx$



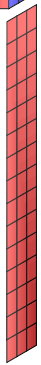
G
$[k, m, n]$



H
$[k, K]$



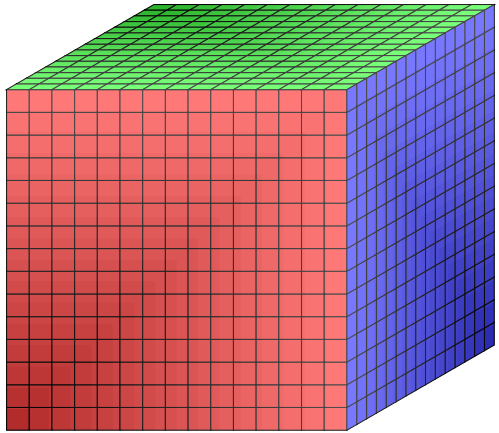
W
$[n, N]$



V
$[m, M]$

$$X \approx G \otimes H \otimes W \otimes V$$

Canonical Polyadic Tensor Decomposition (3D case): CPD-3 Rank-1



$$X \approx \lambda(H \otimes W \otimes V)$$

X
$[K, M, N]$

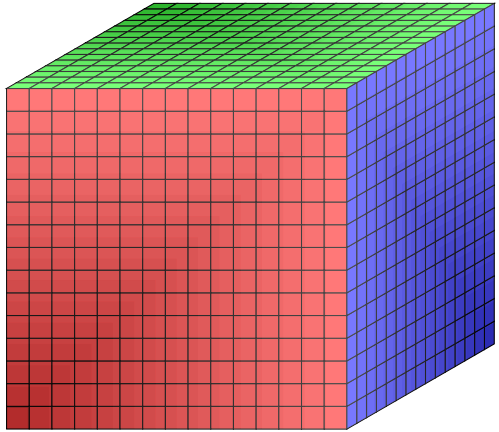
H
$[1, K]$

λ

W
$[1, N]$

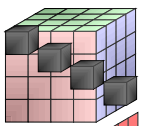
V
$[1, M]$

Canonical Polyadic Tensor Decomposition (3D case): CPD-3 Rank-4

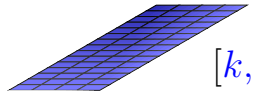


X
$[K, M, N]$

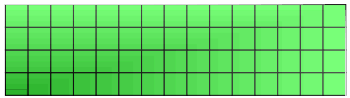
$\approx$



D
$[k, k, k]$



H
$[k, K]$

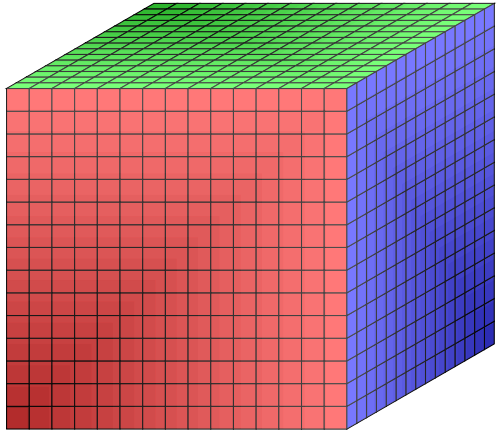


W
$[k, N]$

V
$[k, M]$

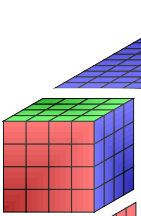
$$X \approx D \otimes H \otimes W \otimes V$$

Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(4,4,4)



X
$[K, M, N]$

$\approx$



G
$[k, k, k]$

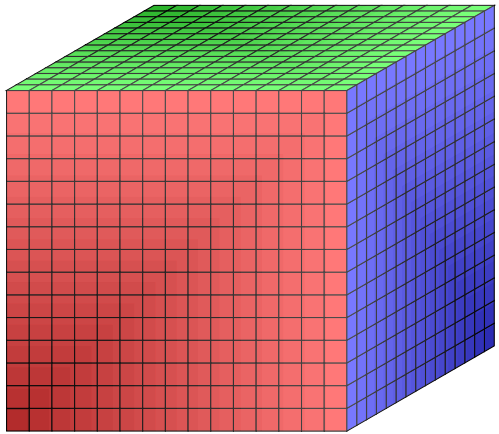
H
$[k, K]$

W
$[k, N]$

V
$[k, M]$

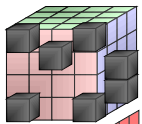
$$X \approx G \otimes H \otimes W \otimes V$$

Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(4,4,4)

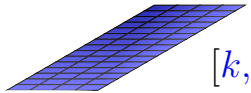


X
$[K, M, N]$

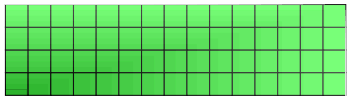
$\approx$



D
$[k, k, k]$



H
$[k, K]$



W
$[k, N]$

V
$[k, M]$

$$X \approx D \otimes H \otimes W \otimes V$$

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

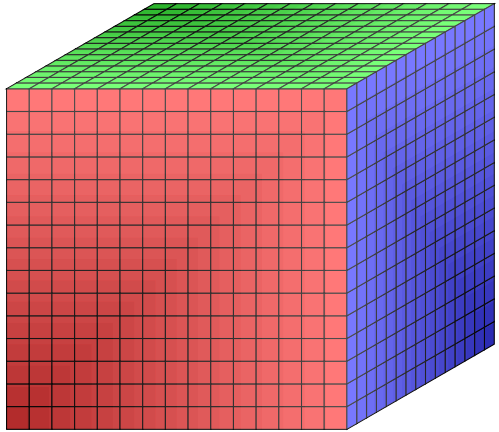
Tensor Rank and Tensor Decomposition

- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tensor Rank and Tensor Decomposition

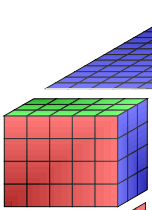
- **Rank-1 Tensor**: a single tensor product of vectors
- **Tensor Rank**: smallest number r of rank-1 tensors under Canonical Polyadic Decomposition
- **Tensor Multi-Rank**: smallest dimensions of core tensor G under Tucker Decomposition (it always exists)
- **Tensor Rank** is always equal to the rank of Tucker tensor core: $rank(X) = rank(G)$
- **Nonnegative Tensor Rank**: smallest number r of rank-1 tensors under Nonnegative Canonical Polyadic Decomposition
- **Nonnegative Tensor Multi-Rank**: smallest dimension of core tensor G under Nonnegative Tucker Decomposition (existence cannot be guaranteed)
- **Nonnegative Tensor Rank** may not be equal to the rank of Nonnegative Tucker tensor core: $rank_+(X) \leq rank_+(G)$
- **Nonnegative Tensor Rank**: smallest number r of nonzero entries in tensor core under Nonnegative Tucker Decomposition

Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(3,4,5)

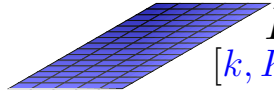


X
$[K, M, N]$

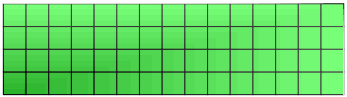
$\approx$



G
$[k, m, n]$



H
$[k, K]$

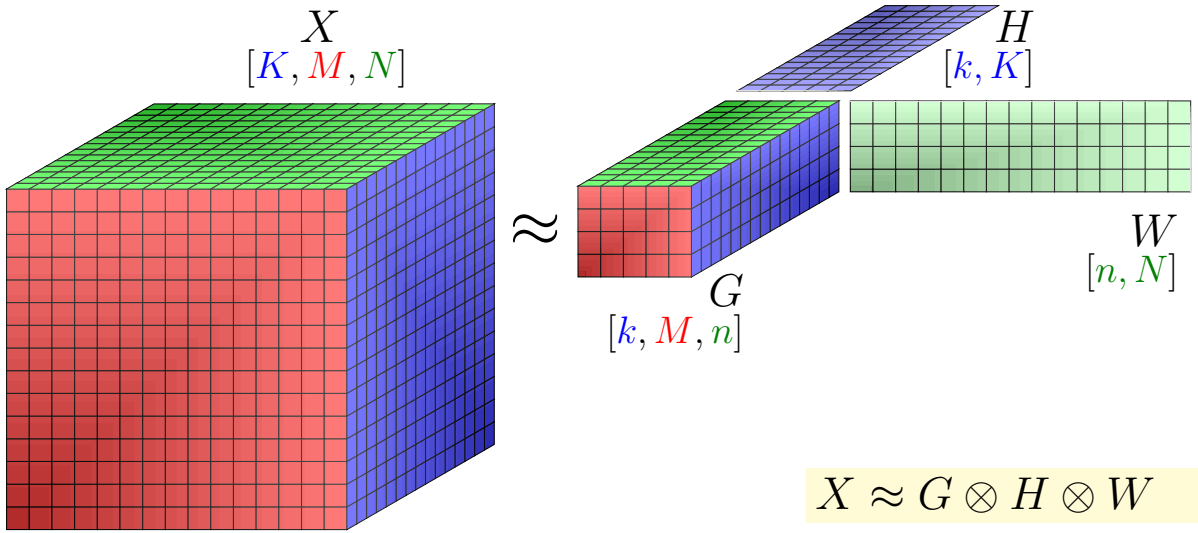


W
$[n, N]$

V
$[m, M]$

$$X \approx G \otimes H \otimes W \otimes V$$

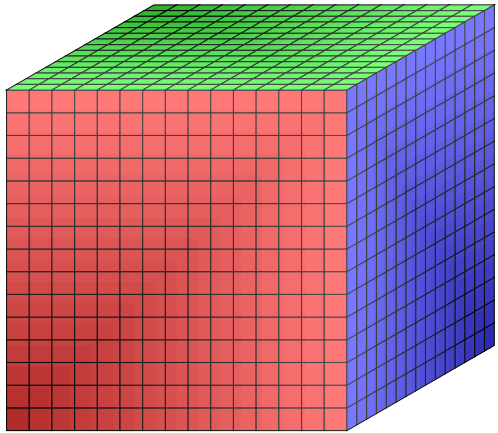
Tucker Tensor Decomposition: Tucker-2 (three possible alternatives)



$$X \approx G \otimes H \otimes W$$

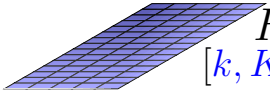
Tucker Tensor Decomposition: Tucker-1 (three possible alternatives)

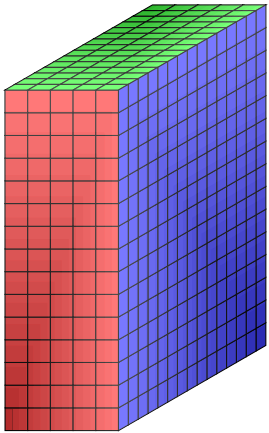
X
$[K, M, N]$



$\approx$

H
$[k, K]$

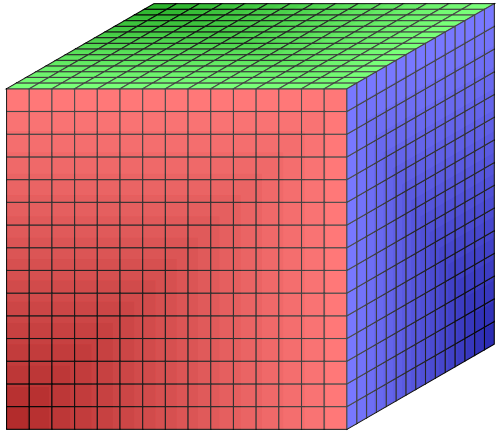




G
$[k, M, N]$

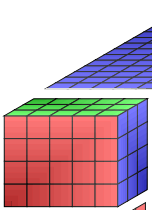
$$X \approx G \otimes H$$

Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(3,4,5)

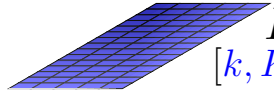


X
$[K, M, N]$

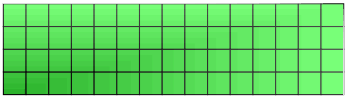
$\approx$



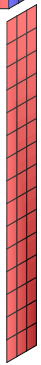
G
$[k, m, n]$



H
$[k, K]$



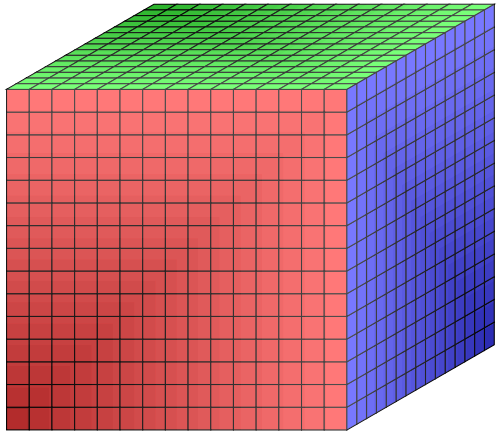
W
$[n, N]$



V
$[m, M]$

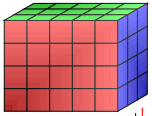
$$X \approx G \otimes H \otimes W \otimes V$$

Tucker Tensor Decomposition: Feature extraction

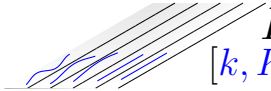


X
$[K, M, N]$

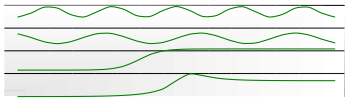
$\approx$



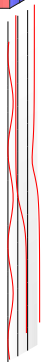
G
$[k, m, n]$



H
$[k, K]$



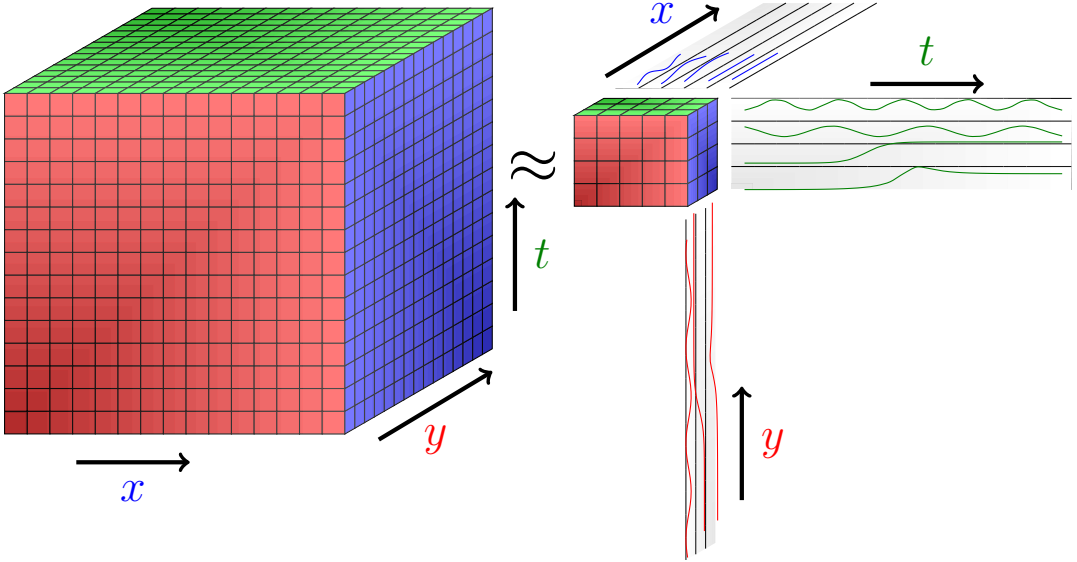
W
$[n, N]$



V
$[m, M]$

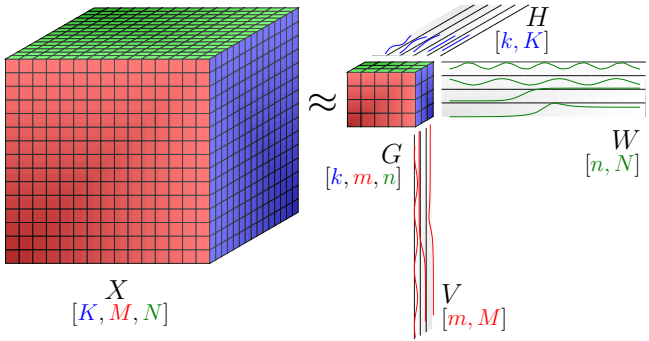
$$X \approx G \otimes H \otimes W \otimes V$$

Tucker Tensor Decomposition: Feature extraction

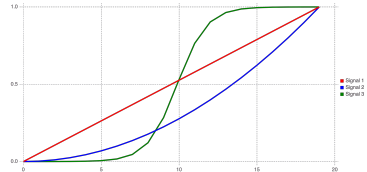


Tensor Decomposition

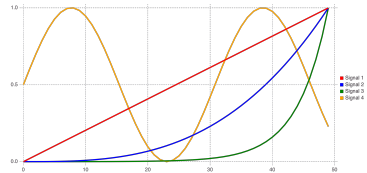
- Tucker/CPD decomposition is achieved through minimization
- Nonnegativity and sparsity constraints applied
- Optimal number of features [k, m, n] is estimated through k-means clustering of a series minimization solutions with random initial guesses



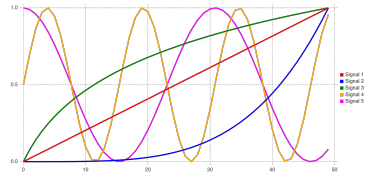
Tucker Tensor Decomposition: Example



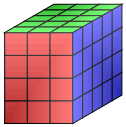
$$V = \begin{bmatrix} t \\ t^2 \\ tanh(t-10)+1 \end{bmatrix}$$



$$W = \begin{bmatrix} x \\ x^3 \\ e^x \\ sin(x)+1 \end{bmatrix}$$



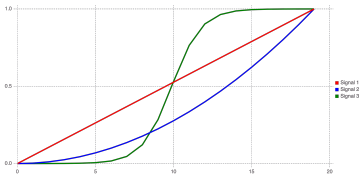
$$H = \begin{bmatrix} y \\ y^4 \\ ln(y) \\ sin(2y)+1 \\ cos(y)+1 \end{bmatrix}$$

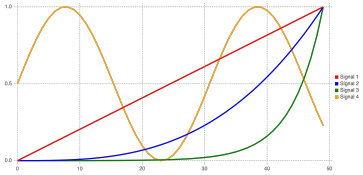


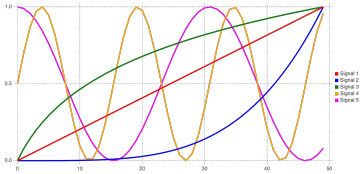
(12 elements)

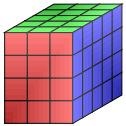
$$\begin{aligned} X &= G \otimes H \otimes W \otimes V \\ &= xyt + xy^4t + xtln(y) + \\ &xt(sin(2y)+1) + x^3yt + \\ &xt(cos(y)+1) + yte^x + \\ &yt(sin(x)+1) + \\ &xyt^2 + xy(1+tanh(t-10)) + \\ &t^2e^x(sin(2y)+1) + \\ &(1+tanh(t-10))(sin(x)+1) \\ &(cos(y)+1) \end{aligned}$$

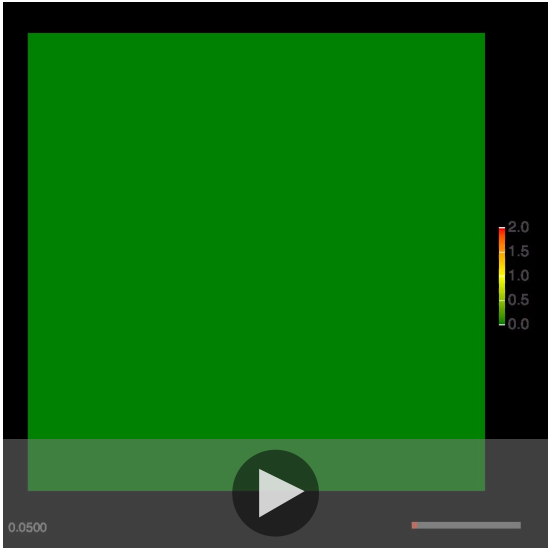
Tucker Tensor Decomposition: Example



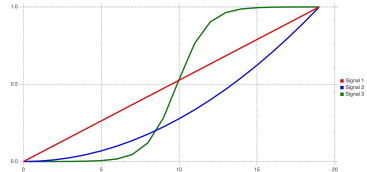


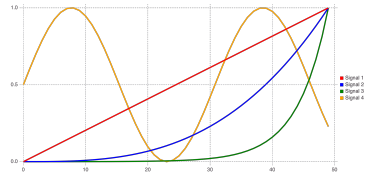


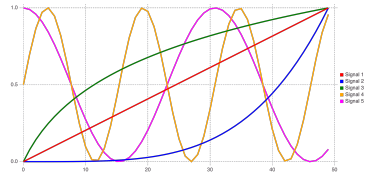




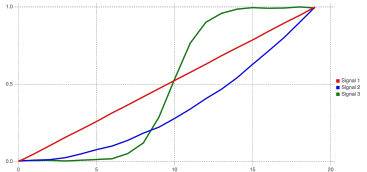
Tucker Tensor Decomposition: Example

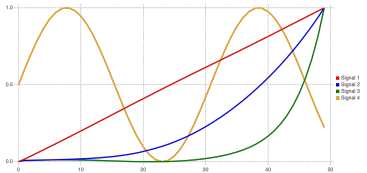


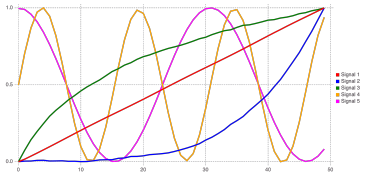




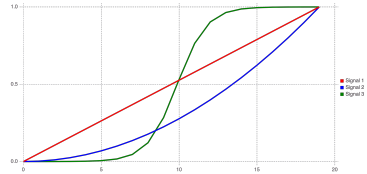
← **Truth** vs **Predictions** →



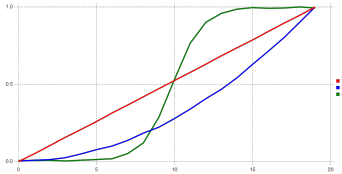




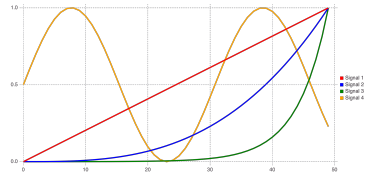
Tucker Tensor Decomposition: Example

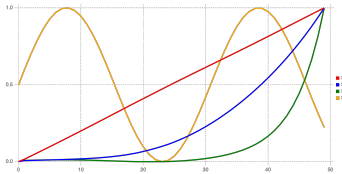


← Truth vs Predictions →

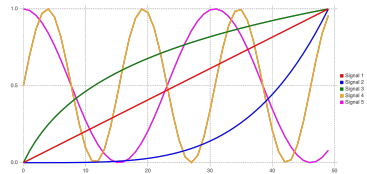


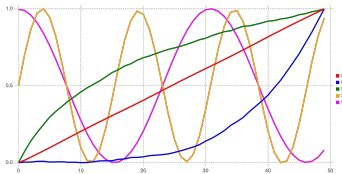
$$V = \begin{bmatrix} t \\ t^2 \\ tanh(t-10)+1 \end{bmatrix}$$





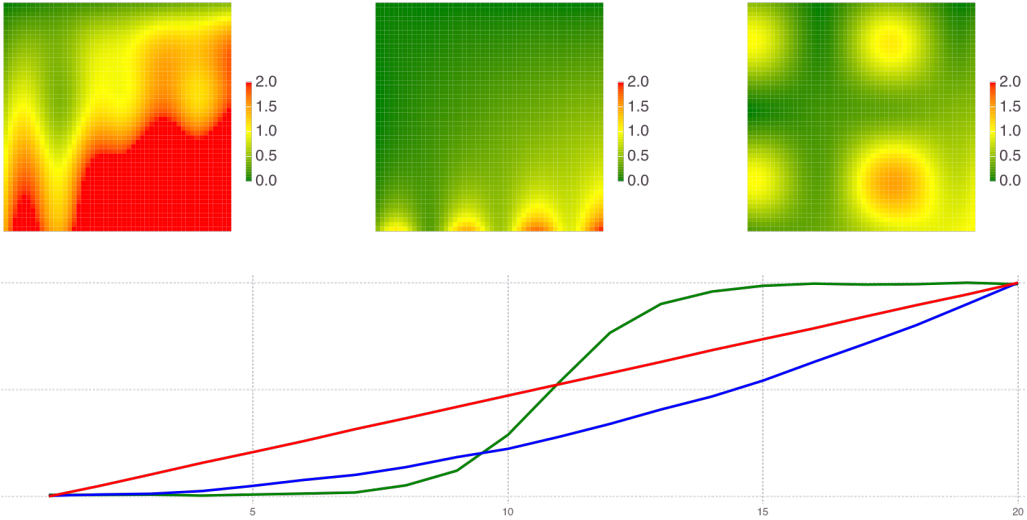
$$W = \begin{bmatrix} x \\ x^3 \\ e^x \\ sin(x)+1 \end{bmatrix}$$



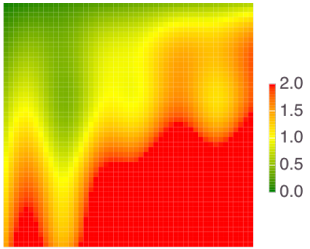


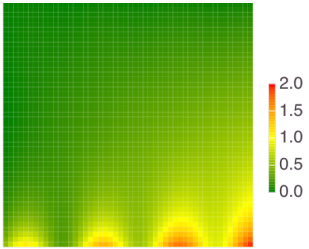
$$H = \begin{bmatrix} y \\ y^4 \\ ln(y) \\ sin(2y)+1 \\ cos(y)+1 \end{bmatrix}$$

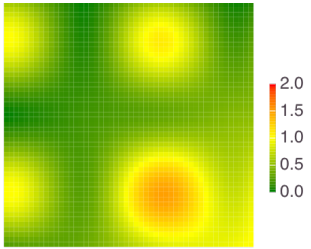
Tucker Tensor Decomposition: Example



Tucker Tensor Decomposition: Example







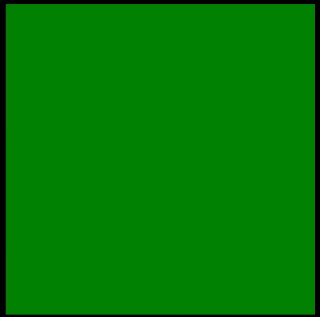
$$M = G \otimes H \otimes W$$

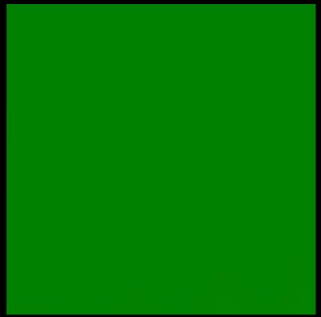
$$M_1 = xy + xy^4 + xlog(y+1) + x(sin(2y)+1) + ye^x + x(cos(y)+1) + x^3y + y(sin(x)+1)$$

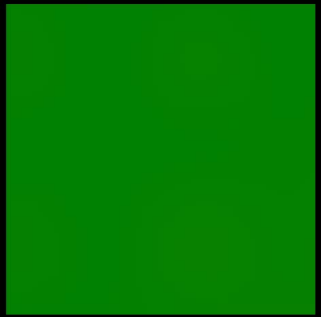
$$M_2 = xy + e^x(cos(y)+1)$$

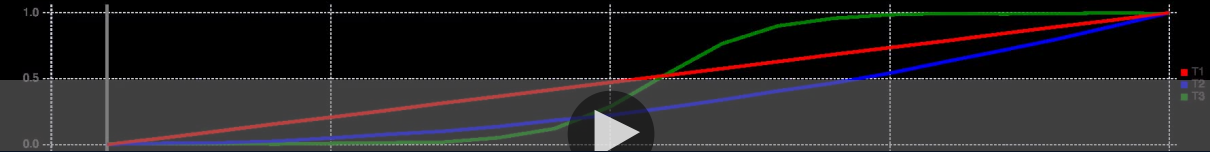
$$M_3 = xy + (sin(x)+1)(cos(y)+1)$$

Tucker Tensor Decomposition: Example



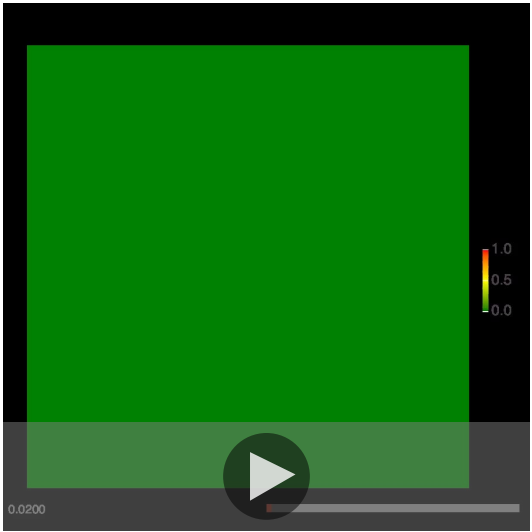




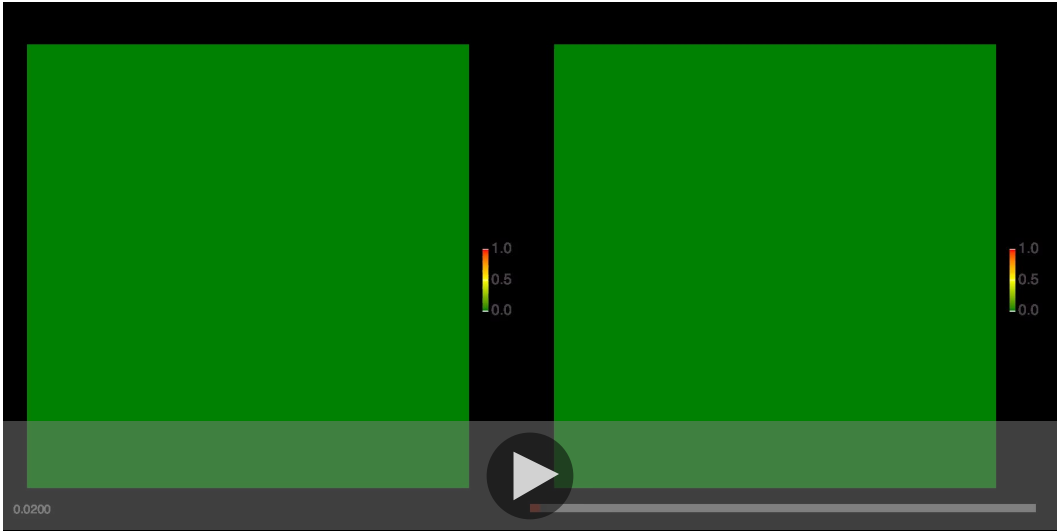


Tucker Tensor Decomposition: Example II

- $(50 \times 50 \times 50)$ tensor
- 50 columns in x
- 50 rows in y
- 50 time frames
- ‘ones‘ swimming in a sea of ‘zeros‘

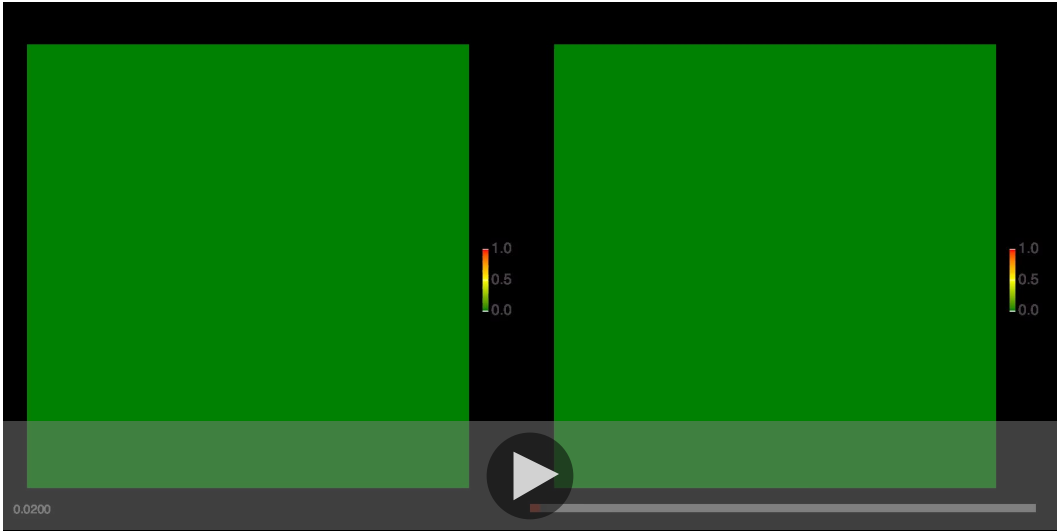


Tucker Tensor Decomposition: Example II



Factorizing all **3** dimensions $(50 \times 50 \times 50) \rightarrow (6 \times 44 \times 48)$

Tucker Tensor Decomposition: Example II



6 groups of swimmers (x); 44 lanes occupied (y); 48 time frames (first/last empty)

NMFk / NTFk Challenges

- **Identifying the number of unknown features**:
  - resolved using custom k-means clustering and sparsity constraints on the core tensor
  - number of features identified based on the reconstruction quality (e.g., Frobenius norm) and cluster Silhouettes
- **Solving a non-unique optimization problem**:
  - addressed through multistarts, regularization and nonnegativity constraints
  - applying diverse optimization techniques (Multiplicative/Alternating Least Squares algorithms, NLopt, Ipopt, Gurobi, MOSEK, GLPK, Clp, Cbc, ...)
- **Processing Big Data**:
  - GPU's / TPU's / Distributed computing
  - Account for data sparsity and structure
  - Nonnegative Tensor Trains
- **Dealing with Noisy Data**:
  - Random noise impacts accuracy but it is accountable
  - Systematic noise is identified as separate signals

4GB Tensor ($1000 \times 1000 \times 1000$)

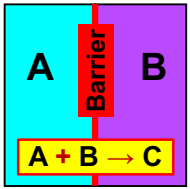
Framework	Execution time (seconds)
MATLAB	2634
NumPy	881
MXNet	644
PyTorch	121
TensorFlow	119
Julia	109

- **Field Data**:
  - Groundwater contamination
  - US Climate data
  - Geothermal data
  - Seismic data
- **Lab Data**:
  - X-ray Spectroscopy
  - UV Fluorescence Spectroscopy
  - Microbial population analyses
- **Operational Data**:
  - LANSCE: Los Alamos Neutron Accelerator
  - Oil/gas production
- **Model Outputs**:
  - Reactive mixing $A + B \rightarrow C$
  - Phase separation of co-polymers
  - Molecular Dynamics of proteins
  - EU Climate modeling

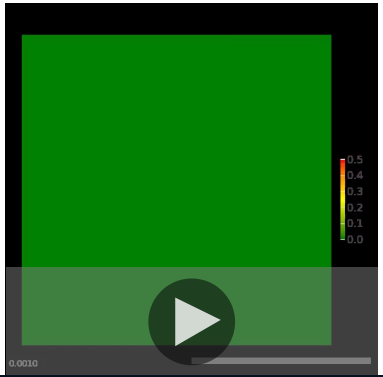
NMFk / NTFk Publications

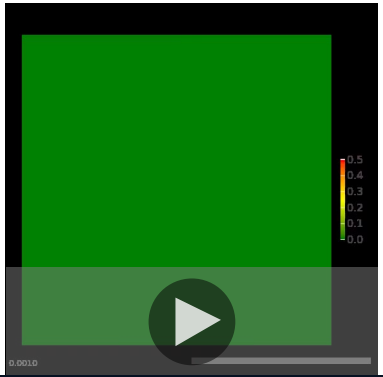
- Vesselinov, Munuduru, Karra, O'Maley, Alexandrov, Unsupervised Machine Learning Based on Non-Negative Tensor Factorization for Analyzing Reactive-Mixing, **Journal of Computational Physics**, Special issue: Machine Learning, 2019.
- Stanev, Vesselinov, Kusne, Antoszewski, Takeuchi, Alexandrov, Unsupervised Phase Mapping of X-ray Diffraction Data by Nonnegative Matrix Factorization Integrated with Custom Clustering, **Nature Computational Materials**, 2018.
- Vesselinov, O'Malley, Alexandrov, Nonnegative Tensor Factorization for Contaminant Source Identification, **Journal of Contaminant Hydrology**, 2018.
- O'Malley, Vesselinov, Alexandrov, Alexandrov, Nonnegative/binary matrix factorization with a D-Wave quantum annealer, **PLOS ONE**, 2018.
- Vesselinov, O'Malley, Alexandrov, Contaminant source identification using semi-supervised machine learning, **Journal of Contaminant Hydrology**, 2017.
- Alexandrov, Vesselinov, Blind source separation for groundwater level analysis based on nonnegative matrix factorization, **WRR**, 2014.

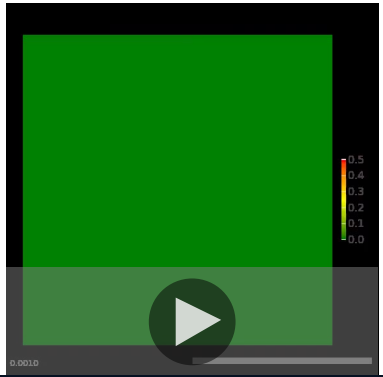
Reactive mixing



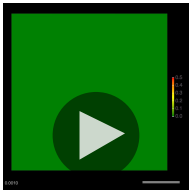
- > 2000 simulations of C concentrations in time/space with varying model inputs representing reactive mixing (5 input model parameters)
- **NTF**k identifies physics processes impacting C concentrations and their relationship to model inputs



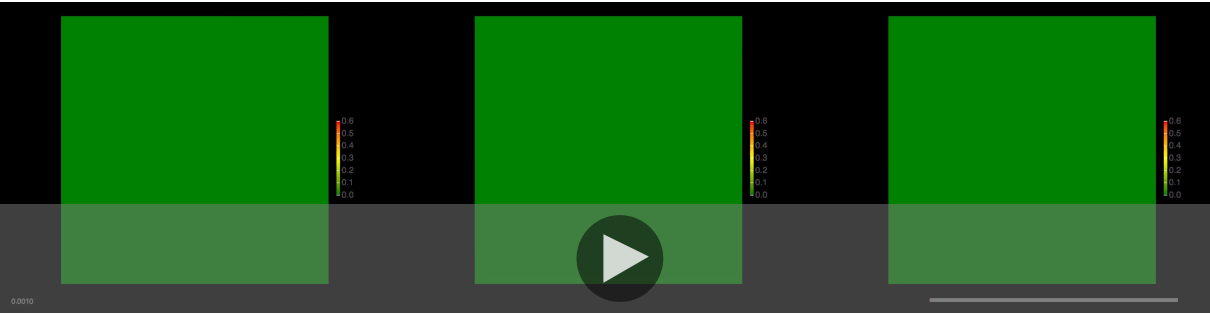




Reactive mixing: NTFk results



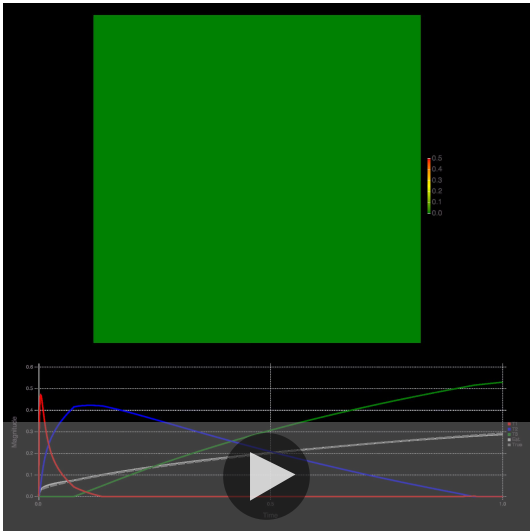
- **NTFk** extracts the dominant time/space features (**processes** / **vortices**) and compresses the model outputs
- Compression: > 200GB $\rightarrow \sim 70$MB (ratio ~ 3000
  Here, $(1000 \times 81 \times 81) \rightarrow (3 \times 12 \times 13)$ $(time \times rows \times columns)$



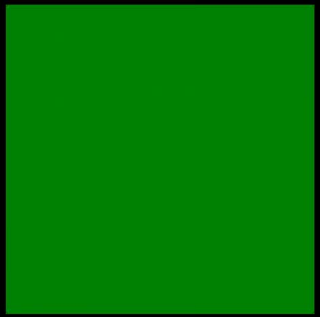
Advection **Dispersion** **Diffusion**

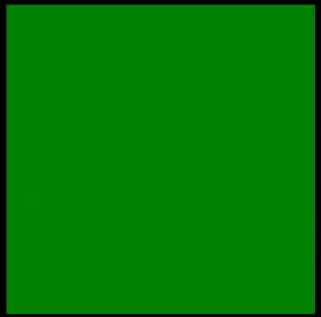
Reactive mixing: NTFk results

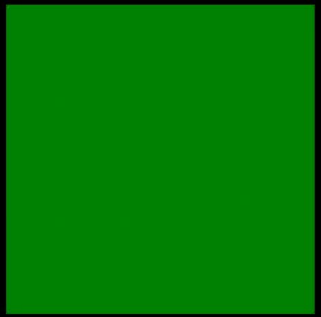
- T1: Advection
- T2: Dispersion
- T3: Diffusion

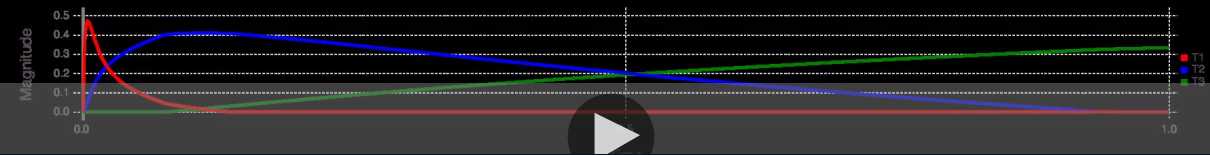


Reactive mixing: NTFk results



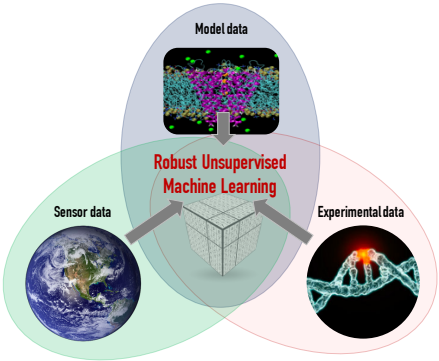






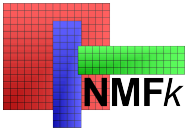
Summary

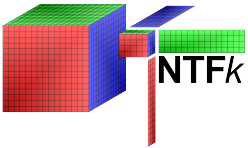
- Developed **novel** unsupervised ML methods and computational tools based on Nonnegative Factorization (Matrices/Tensors)
- Our ML methods have been used to solve various real-world problems (brought breakthrough discoveries related to human cancer research)



- Codes:

NMFk MADS NTFk





- Examples:

http://madsjulia.github.io/Mads.jl/Examples/blind_source_separation
http://tensors.lanl.gov
http://tensordecompositions.github.io
https://github.com/TensorDecompositions
https://hub.docker.com/u/montyvesselinov