Matrix and Tensor Factorization

Instructors: Anh-Huy Phan and Andrzej Cichocki
TAs: Salman Ahmadi Asl, Julia Gusak, Valetin Leplat
Igor Vorona, Konstantin Sobolev, Konstantin Sozykin
Dmitry Ermilov and Nikolay Kozyrskiy

Center for Artificial Intelligence Technology
Skolkowo Institute of Science and Technology (SKOLTECH), Moscow, Russia

January 31, 2022

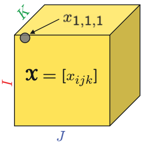
Skoltech

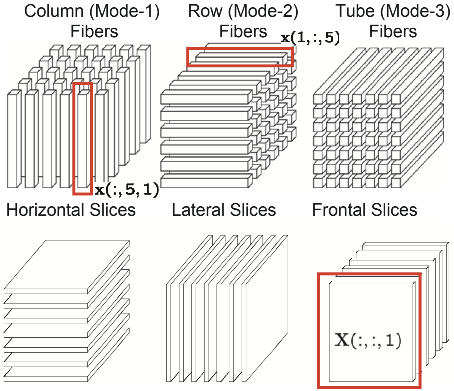
At a Glance

What is Tensor?

Tensor is multiway array often of high order

An $I \times J \times K$ tensor



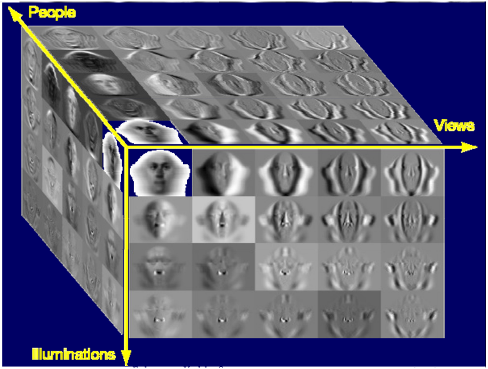


At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient

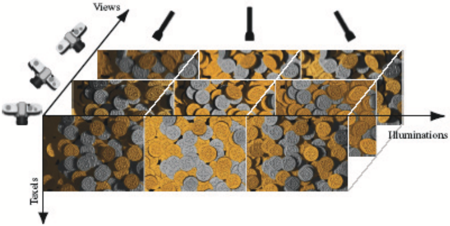


At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient

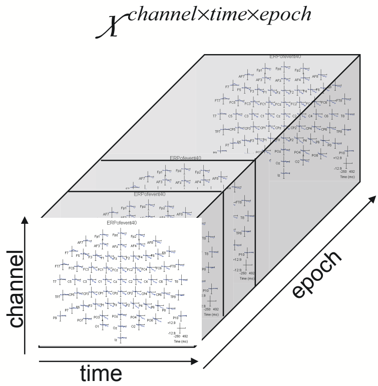


At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient



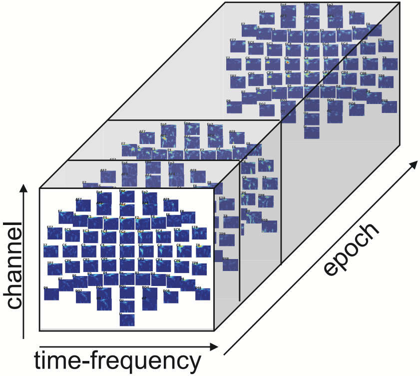
Skoltech
Skolkovo Institute of Science and Technology

At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient

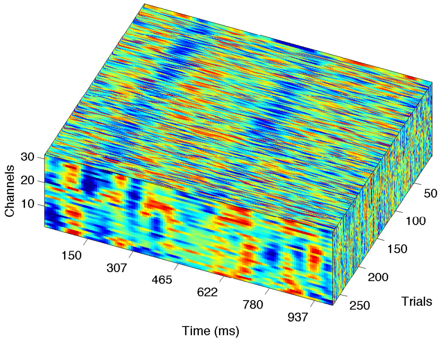


At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient



Skoltech
Skolkovo Institute of Science and Technology

At a Glance

What is Tensor?

Tensor is multiway array often of high order

- Faces: people identities $\times$ views $\times$ illumination $\times$ facial expression $\times$ pixels
- Textures: illumination $\times$ view $\times$ texels (position)
- EEG signals: channel $\times$ sample (time) $\times$ trial
- EEG spectra: channel $\times$ frequency bin $\times$ time frame $\times$ trial
- Alzheimer EEG signals: channel $\times$ sample (time) $\times$ patient
- Enron mail: employee $\times$ employee $\times$ month
- Fluorescence data: sample $\times$ emission $\times$ excitation
- Social network analysis: author $\times$ keyword $\times$ time
- Network traffic data in cyber security: time $\times$ senderIP $\times$ receiverIP $\times$ port

At a Glance

Why do we organize data as Tensor?

- Due to the mechanism of data collection
  - Multiple sensors, antennas
  - Multiple conditions such as subjects, tasks

At a Glance

Why do we organize data as Tensor?

- Due to the mechanism of data collection
  - Multiple sensors, antennas
  - Multiple conditions such as subjects, tasks
- Due to data transformations (wavelets), multilinear conversion (Toeplitz, Hankel)

At a Glance

Why do we organize data as Tensor?

- Due to the mechanism of data collection
  - Multiple sensors, antennas
  - Multiple conditions such as subjects, tasks
- Due to data transformations (wavelets), multilinear conversion (Toeplitz, Hankel)
- Model parameters e.g., in multivariate polynomial approximation, Volterra series representation

$$\begin{aligned} h(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) &= w_0 + \sum_k \boldsymbol{w}_k^T \boldsymbol{x}_k + \sum_{k,l} \boldsymbol{x}_k^T \mathbf{W}_{k,l} \boldsymbol{x}_l \\ &+ \sum_{k,l,m} \mathcal{W}_{k,l,m} \bar{\times}_1 \boldsymbol{x}_k \bar{\times}_2 \boldsymbol{x}_l \bar{\times}_3 \boldsymbol{x}_m \end{aligned}$$

At a Glance

Why do we need Tensor decompsition?

- Vectorization of a data tensor to exploit conventional approaches for vectors and matrices
  - Risk of losing correlation of data entries
  - Generate long vectors $\Longrightarrow$ high computational cost
  - ...
- Tensor decomposition is a natural tool for multiway data
  - hidden components, structures
  - low computational cost in comparison with vectorial approach for the same data
  - high dimensionality reduction

Cyber Security through Multidimensional Data Decompositions

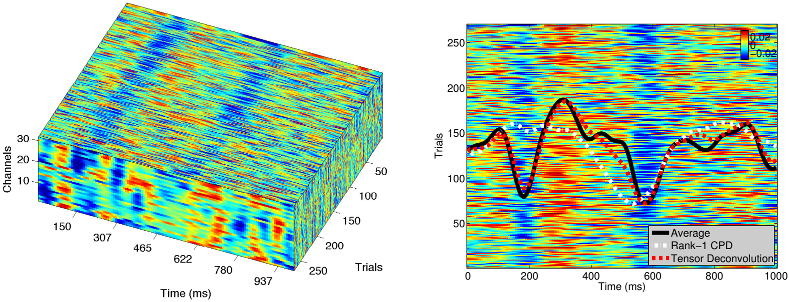
Network data can associate with attributes timestamp, source IP address, destination IP address, source port, destination port, protocol.

Reservoir Labs introduces ENSIGN

`https://www.reservoir.com/product/ensign-cyber/`

- ENSIGN uses a high-performance implementation of a collection of tensor decomposition algorithms to enable the unsupervised discovery of subtle undercurrents and deep, cross-dimensional correlations in multidimensional data.
- ENSIGN incorporates novel data formats to efficiently represent real-world structured, sparse data and optimizations that utilize the power of large High Performance Computing (HPC) systems, to solve large problems practically.

Multiway analysis for EEG signals I



- EEG signals were recorded when subject counted how often a target image was flashed on a screen.
- Target P300 responses were arranged in an array of size 32 channels $\times$ 128 sample points $\times$ 271 trials.

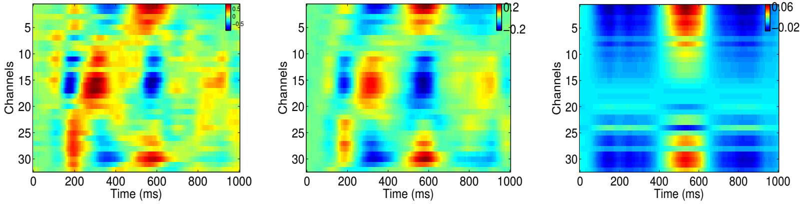
Multiway analysis for EEG signals II

Hidden common components across multiple aspects can be extracted using CPD.

Variations in EEG signals (response latencies, frequency variations, spatial distribution)

- Neural responses are usually evoked with different latencies even by the same stimuli,
- Electrodes might not be placed at identical locations.
- Variations are also observed from physiological effects such as habituations.
- Variations together with independent background activities and heavy noise in EEG signals make the extraction of common hidden components more difficult.

Multiway analysis for EEG signals III



(a) Grand average (b) Tensor deconvolution (c) Rank-1 CPD

Figure: Grand average and Common spatio-temporal patterns of P300 responses estimated by tensor decomposition and deconvolution.

Topics and Tentative Schedule I

- **Introduction and Motivations**

  Challenges in Large Scale Data Processing Linear and Multilinear Algebra and their links to Machine Learning (ML) and AI,

  Curse of Dimensionality and Generalized Separation of Variables for Multivariate Functions

  Advantages of Multiway Analysis via Tensor Factorizations

  Potential Applications of Tensors

Topics and Tentative Schedule II

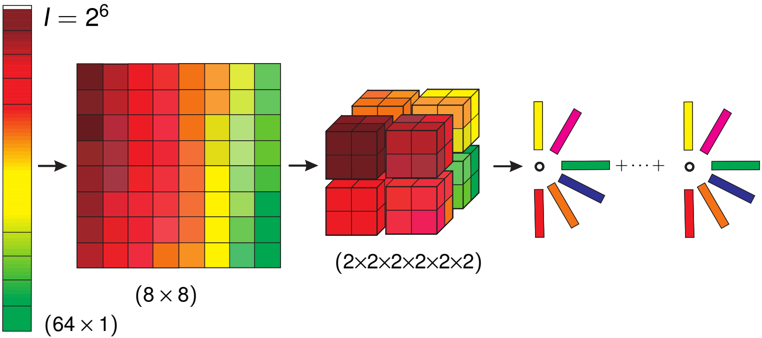


Figure: Construction of tensors. Tensorization of a vector or matrix into the so-called *quantized format*; in scientific computing this facilitates super-compression of large-scale vectors or matrices, (e.g., through a polyadic or a TT decomposition).

Topics and Tentative Schedule III

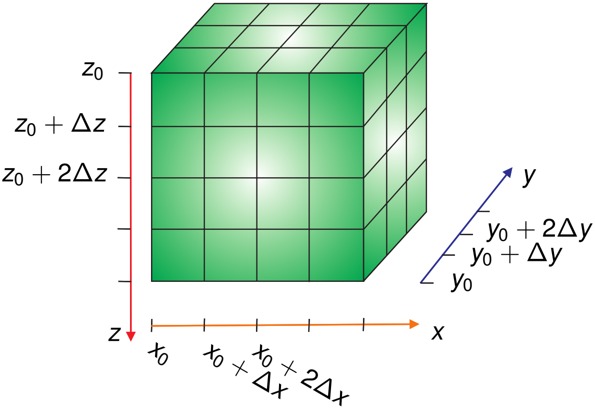


Figure: Construction of tensors. Tensor formed through the discretization of a trivariate function $f(x, y, z)$.

Topics and Tentative Schedule IV

- **Matrix/Tensor Operations and Tensor Network Diagrams**

  Matrix and Tensor Notations and Graphical Representations ,

  Basic algebraic and mathematical multilinear operations.,

  Properties of Multilinear Operations between tensors, matrices and vectors,

  Graphical Representation of Fundamental Tensor Decompositions CPD and Tucker Decomposition,

  Basic optimization algorithms for CPD

Topics and Tentative Schedule V

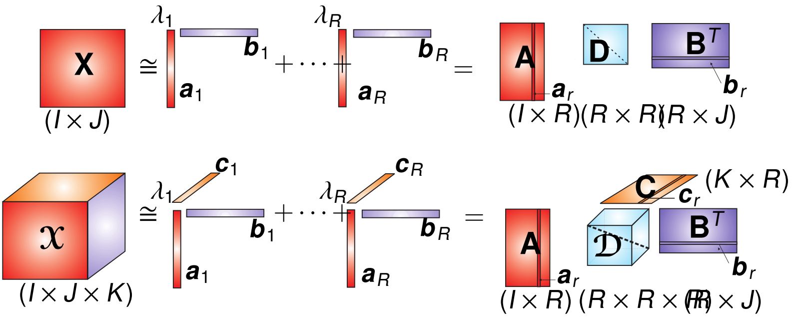


Figure: Analogy between dyadic (top) and polyadic (bottom) decompositions; the Tucker format has a diagonal core. The uniqueness of these decompositions is a prerequisite for blind source separation and latent variable analysis.

Topics and Tentative Schedule VI

- **Singular Value Decomposition (SVD) and Higher Order Singular Value Decomposition (HOSVD)**

  SVD and its Extension to tensors HOSVD, Higher Order Orthogonal Iterations (HOOI) Tucker decomposition with bounded approximation error

Topics and Tentative Schedule VII

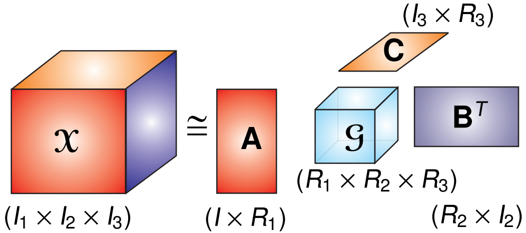


Figure: Tucker decomposition of a third-order tensor. The column spaces of **A**, **B**, **C** represent the signal subspaces for the three modes. The core tensor $\mathcal{G}$ is nondiagonal, accounting for possibly complex interactions among tensor components.

Topics and Tentative Schedule VIII

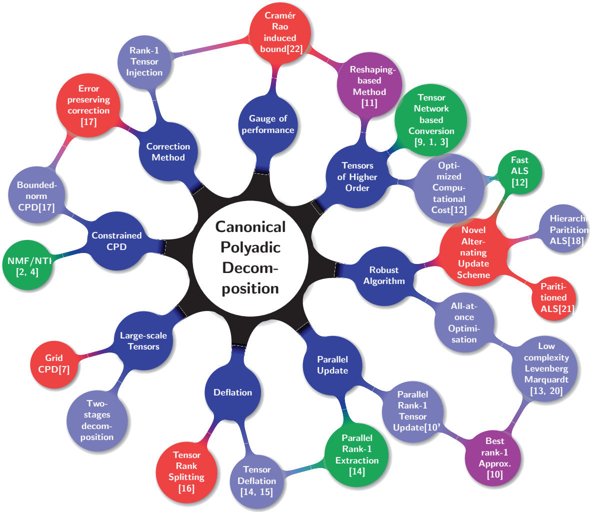
- **Nonnegative Matrix and Tensor Factorizations, Algorithms and Applications**

  Matrix Factorization Models with Nonnegativity and Sparsity Constraints Basic NMF Model, Symmetric NMF, Semi-Orthogonal NMF, NMF with Offset (Affine NMF), Multi-layer NMF, Projective and Convex NMF, Kernel NMF, Convolutive NMF

- **Advanced Algorithms for Tensor Factorizations**

  Algorithms for CP /PARAFAC (ALS, NTF, HALS) Algorithm for large scale Tucker Tensor Format SVD and Higher Order SVD (HOSVD) for Large-Scale Problems Two-way and Multiway Component Analysis (MWCA)

Topics and Tentative Schedule IX



Skoltech
Skolkovo Institute of Science and Technology

Topics and Tentative Schedule X

- **Tensor Train and Tensor Chain Decompositions** Graphical Interpretations and Algorithms Tensor Train Decomposition - Matrix Product State Matrix TT Decomposition - Matrix Product Operator Links between CPD, BTD formats and TT/TC formats Quantized Tensor Train (QTT) - Blessing of Dimensionality Basic Operations in TT Formats Algorithms for TT Decompositions (ALS, TT-SVD, SGD)

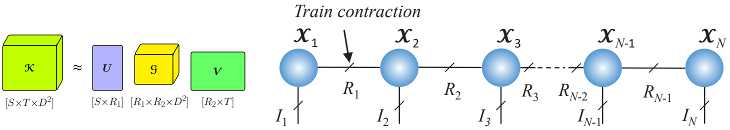


Figure: Graphical illustration of a TK2 tensor (left) and a TT-tensor (right) $\mathcal{X} = \mathcal{X}_1 \bullet \mathcal{X}_2 \bullet \cdots \bullet \mathcal{X}_N$. A node represents a 3-rd order core tensor $\mathcal{X}_n$.

Topics and Tentative Schedule XI

- **Tensorization and Structured Tensors**

  Advanced Reshaping or Folding Tensorization through a Toeplitz/Hankel Tensor Tensorization by Means of Löwner Matrix Tensorization based on Cumulants and Derivatives of GCFs Tensorization by Learning Local Structures Tensorization based on Divergences, Similarities or Information Exchange Tensor Structures in Multivariate Polynomial Regression Tensor Structures in Vector-variate Regression Tensor Structure in Volterra Models of Nonlinear Systems Tensor Representations of Sinusoid Signals and BSS

Topics and Tentative Schedule XII

- **Supervised Learning with Tensors**

  Linear regression and Tensor Regression Regularized Tensor Models Higher-Order Low-Rank Regression (HOLRR) Kernelized HOLRR Higher-Order Partial Least Squares (HOPLS) Kernel HOPLS Kernel Functions in Tensor Learning Tensor Variate Gaussian Processes (TVGP) Support Tensor Machines . Higher Rank Support Tensor Machines (HRSTM) Kernel Support Tensor Machines Tensor Fisher Discriminant Analysis (FDA) Linear Discriminant Analysis (LDA) and Higher Order Discriminant Analysis (HODA) Tensor Canonical Correlation Analysis

- **Randomized Algorithms for Matrix and Tensor Factorizations**
  Randomized SVD, Randomized PCA, Randomized CPD, Randomized HOSVD and Tucker and randomized TT algorithms

- **Matrix and Tensor Completions**

  Low-rank tensor approach, Singular Value Threshold (SVT) approach, Stochastic Gradient Descent (SGD) Approach

Topics and Tentative Schedule XIII

- **Constrained Tensor Decomposition/Tensor Networks**

  Low-rank tensor approach, Singular Value Threshold (SVT) approach, Stochastic Gradient Descent (SGD) Approach

- **Multiview, Fusion network**

  Canonical Correlation Analysis, Partial Least Regression, and their extension to multiway data

  Fusion network, multimodal analysis

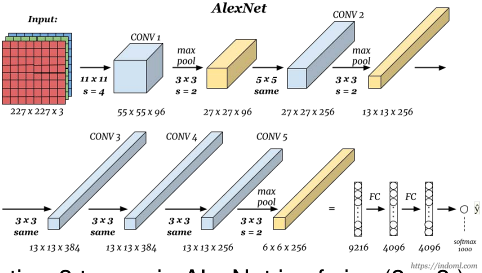
- **Recommender system**
- **Image compression, denoising, completion**
- **NLP: document representation**

Topics and Tentative Schedule XIV

- **Tensor methods in Deep Learning**

Model	Params	SIZE(MB)
SimpleNet	6.4M	24.4
SqueezeNet	1.25M	4.7
Inception v4*	42.71M	163
Inception v3*	23.83M	91
Incep-Resv2*	55.97M	214
ResNet-152	60.19M	230
ResNet-50	25.56M	97.70
AlexNet	60.97M	217.00
GoogleNet	7M	40
VGG16	138.36M	512.2

Topics and Tentative Schedule XV



- Convolution-3 tensor in AlexNet is of size $(3 \times 3\ ) \times 256 \times 384$.
- Low-rank approximation reduces the number of parameters in convolutional layers.
  Thereby accelerate the inference of the network (Jaderberg et. al, 2014)
  $\mathcal{Y} \approx \sum_{r=1}^{200} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r$
- However Lebedev et. al, 2015 observed that the fine-tuning of the whole network was *unstable*.

Skoltech
Skolkovo Institute of Science and Technology

Grades

- Homework 40%
- Midterm test 20%
- Project or Final exam (40%)
- Bonus : In-class assignment 5-10%

Early submission bonus

Early submission will be awarded a bonus for submission before the due date one or 2 days, respectively.
The bonus is awarded only to the correct and complete answers

Project I

Team project will be assessed according to the following criteria

- Scientific work (15 points)
  - Review of Literature
  - Complexity of the problem and its difficulty
  - Implementation
  - Comparison with existing methods for the similar problems
  - Teamwork
  - Creative
- Presentation (10 points)
  - Motivation of the project
  - Importance and contribution of the study methods or scientific papers
  - Application of the methods or problems investigated
  - How to solve the problem
  - Discussion

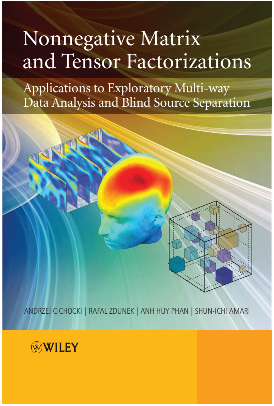
Project II

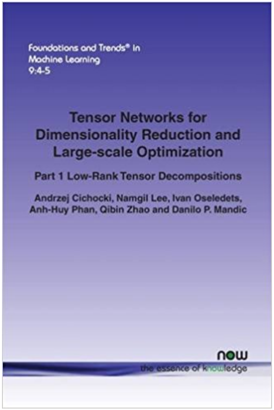
- Demonstration (10 points)
  - Show examples to demonstrate the method as presented in the original scientific paper
  - Quality of the implementation
- Audience assessment (5 points)
  - Appropriateness to problem investigated
  - Clarity, conciseness
  - Significance of the project and possible contribution
  - Review of Literature (thoroughness and comprehensiveness)
  - Quality of the examples

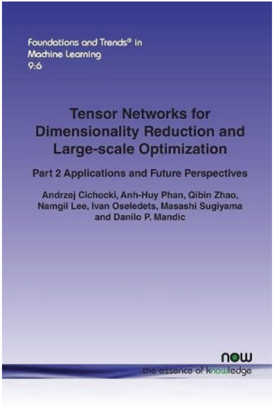
Project topics

- Algorithms for Tensor decompositions/ Tensor Networks
- Low-rank TNs for image denoising
- Low-rank TNs for image/video compression
- Low-rank tensor layer for compression of CNN, transformer
- Signal Separation, blind deconvolution
- Gradient free Optimization
- Tensor regression
- Multiview and Fusion network
-

Literature


Nonnegative Matrix and Tensor Factorizations
Applications to Exploratory Multi-way Data Analysis and Blind Source Separation
ANDRZEJ CICHOCKI | RAFAL ZDUNEK | ANH HUY PHAN | SHUN-ICHI AMARI
WILEY


Foundations and Trends® in Machine Learning 9:4-5
Tensor Networks for Dimensionality Reduction and Large-scale Optimization
Part 1 Low-Rank Tensor Decompositions
Andrzej Cichocki, Namgil Lee, Ivan Oseledets, Anh-Huy Phan, Qibin Zhao and Danilo P. Mandic
now
the essence of knowledge


Foundations and Trends® in Machine Learning 9:6
Tensor Networks for Dimensionality Reduction and Large-scale Optimization
Part 2 Applications and Future Perspectives
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
now
the essence of knowledge

Software

Matlab

- Tensor toolbox (Kolda and Bader)
- TensorBox (Phan, Tichavsky and Cichocki)
- TensorLab (L. de Lathauwer et al.)
- N-way toolbox (R. Bro)
- TT toolbox (Oseledets)
- Hierarchical Tucker toolbox

Python

- Tensorly
- Scikit-tensor
- tntorch
- HOTTBOX