Nonnegative Matrix Factorization and Nonnegative Tensor Network with Graph-based Regularization

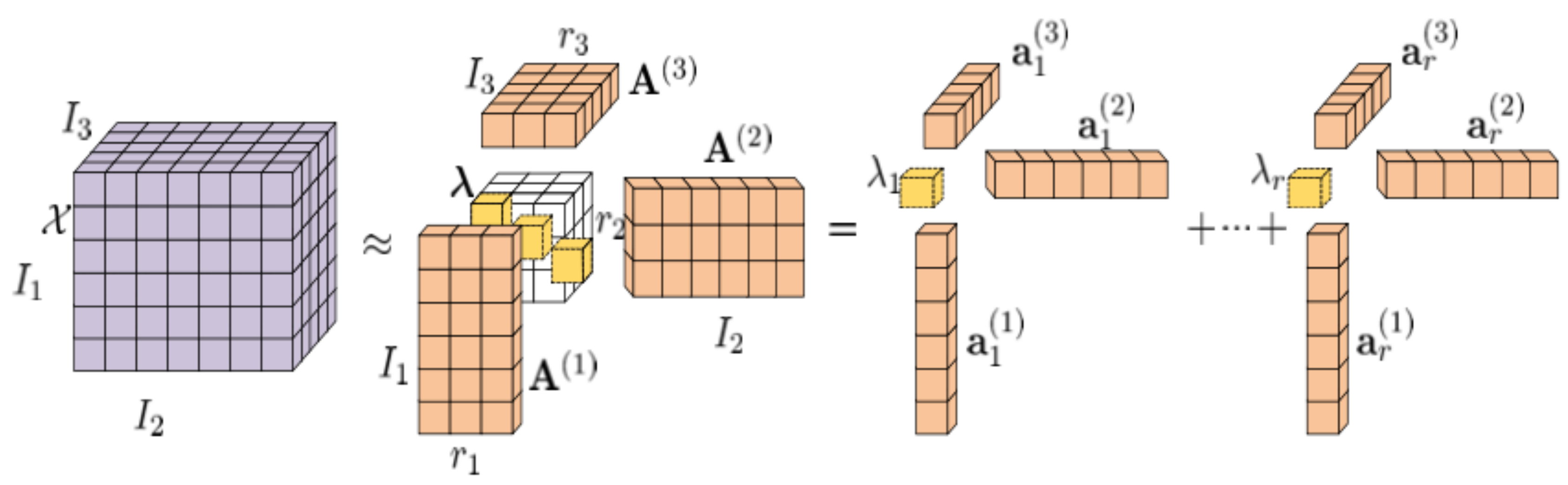
Ning Zheng (鄭 寧)
Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics

Goals and Challenges

- Real application problems in signal processing and machine learning generate multi-dimensional data with physically meaning;
- Low rank matrix/tensor approximation of multi-dimensional arrays aim to represent a higher-order data as a multilinear product of several latent factors (bless of dimensionality);
- Nonnegative constrained matrix/tensor decomposition is a powerful tool for extracting physically meaningful latent components while preserving multilinear structure;
- Linear algebra properties and statistical manifold properties [Ghalamkari, Sugiyama, 2023].

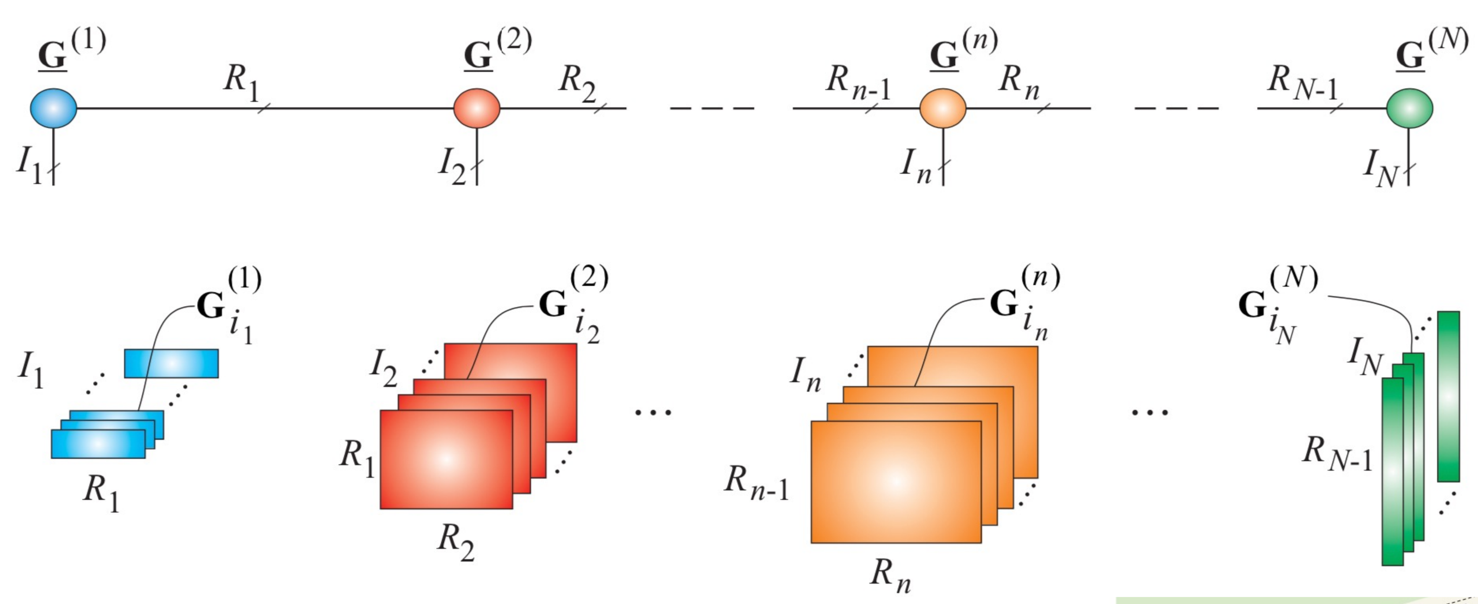
Tensor Networks

- Representation ability: a powerful tool to describe strongly entangled quantum many-body systems in physics;
- Dimensional / Model reduction: decompose a high-order tensor into a collection of low-order tensors connected according to a network pattern;
- Graphical representation of tensor network diagram

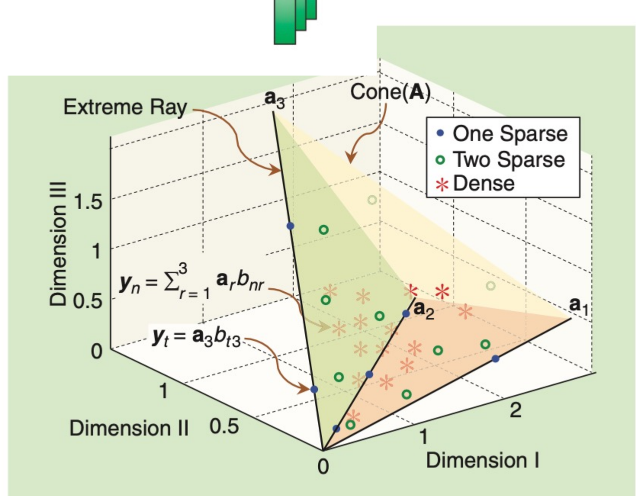


Tensor Train / Ring and Matrix Product State

- Tensor train (TT) decomposition [Oseledets, SIAM, 2011]:



- Tensor chain / ring (TR) decomposition: [Zhao, 2018]
- Sum of TT with shared core tensors
- Hierarchical Tucker decomposition



Graph Tensor Networks

- Given a d-th order tensor, compute the cores with given TR-ranks

$$\min \left\| T - R(Z_1, Z_2, \ldots, Z_d) \right\|_F + \mu g(Z_k)$$
$$s.t. \quad Z_1, Z_2, \ldots, Z_d \in M$$

- Fidelity term denotes the low rank approximation of given tensor;
- Regularization term denotes the prior knowledge for the output core tensors, which are sparse, nonnegative/box constrained, orthogonal (robust PCA), graph structures, etc.
- Alternating least squares method (block coordinate descent method, or block Gauss-Seidel method):
  - ■Alternatively update one core tensor and fix all the other cores tensors
  - ■Solve the subproblem: mode-k unfolding matrix representation:

$$\min \left\| \mathbf{T}_{[k]} - \mathbf{Z}_{k(2)} (\mathbf{Z}_{[2]}^{\neq k})^T) \right\|_F + \mu g(\mathbf{Z}_{k(2)})$$
$$s.t. \quad \mathbf{Z}_{k(2)} \in M$$

- Task 1: graph topology, geometrical information of data can be obtained by modeling a neighbor graph;

$$g(\mathbf{Z}) = J_{\text{graph}} = \sum_{i,j} (z_i - z_j)^2 W_{ij} = \text{Tr}(\mathbf{Z}^T (D - W)\mathbf{Z})$$

- Task 2: nonnegative tensor network optimization. Other related models: nonnegative matrix factorization (NMF), NCP, NTD, NTT, etc.

Nonnegative Matrix/Tensor Factorization

$$\min_{W \in \mathbb{R}_+^{m \times r}, H \in \mathbb{R}_+^{r \times m}} D_{\text{KL}}(V|WH),$$

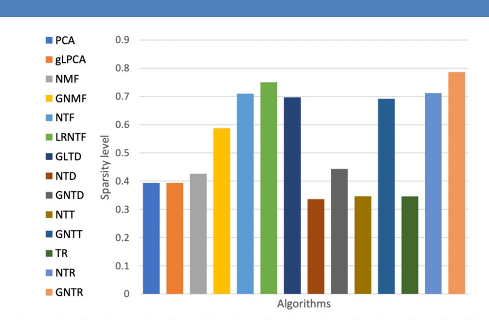
- Nonconvex optimization with infinite permuted local minima;
- NP-hard problem and solved by alternating nonnegative constrained least squares framework;
- Multiplicative update, hierarchical ALS, projected gradient, etc.;
- Frobenius norm for image/text problems, Kullback–Leibler divergence, alpha/beta divergence for audio data.

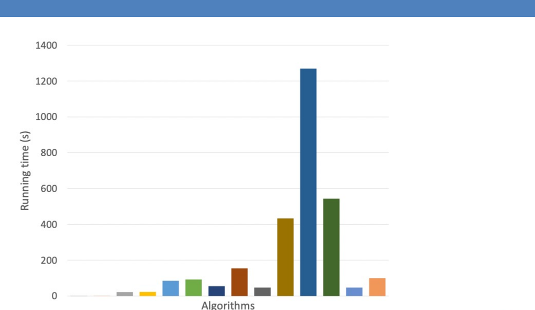
$$\min_{\mathcal{G}^{(n)}} \left\| \mathcal{X} - \text{NTR}\left(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \cdots, \mathcal{G}^{(N)}\right) \right\|_F^2$$
$$\text{s.t. } \mathcal{G}^{(n)} \geq 0, n = 1, 2, \cdots, N,$$

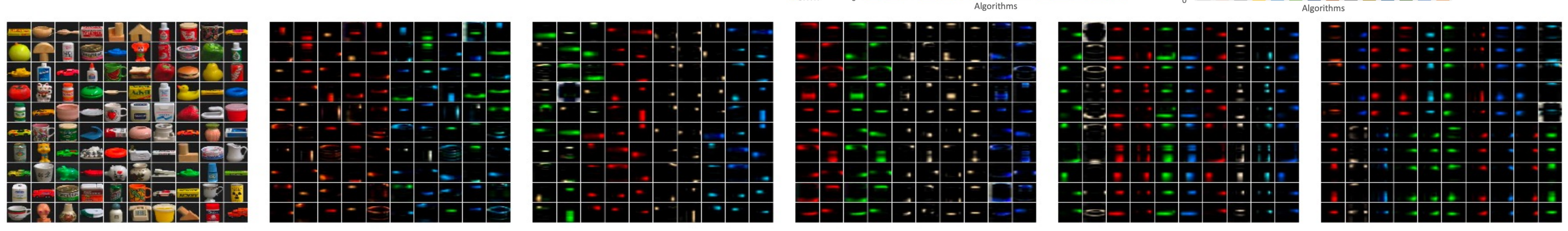
$$D_{\text{KL}}(V|WH) := \sum_{i=1}^{m} \sum_{j=1}^{n} \left((WH)_{ij} - V_{ij} \log(WH)_{ij} \right) + \sum_{i=1}^{m} \sum_{j=1}^{n} \left(V_{ij} \log V_{ij} - V_{ij} \right).$$

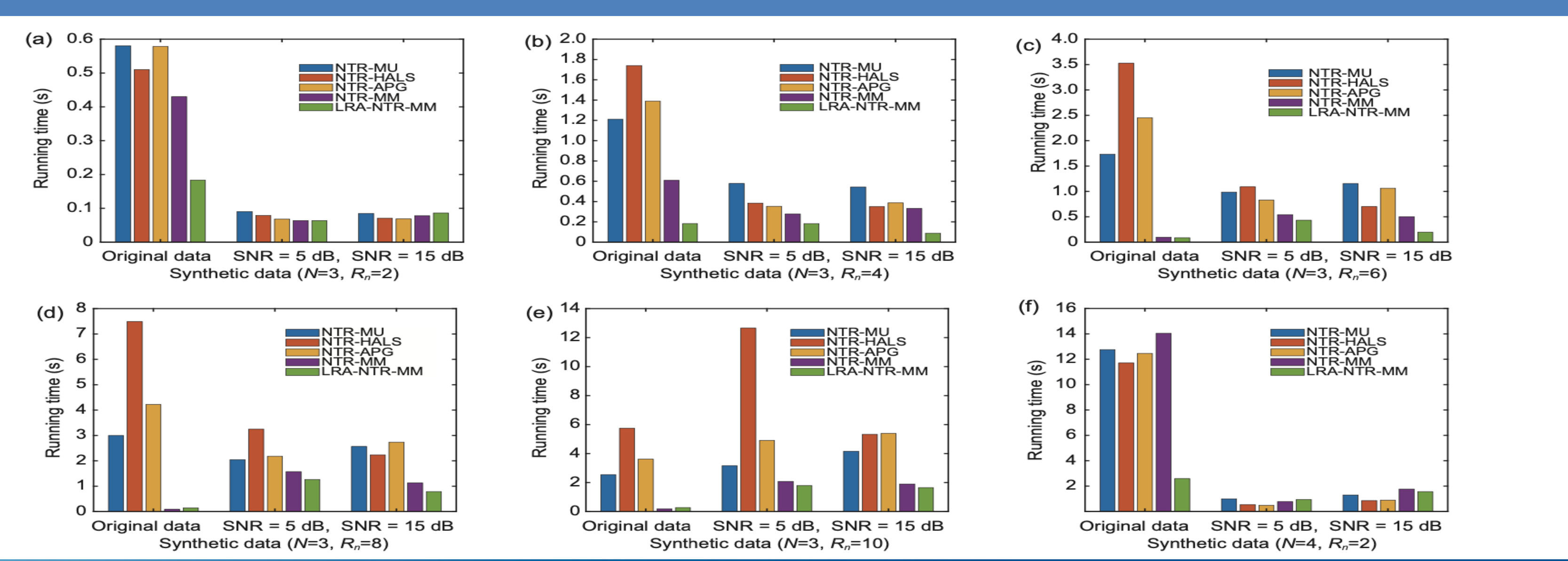
Numerical Experiments

- ORL Database: face images;
- COIL-100 Database;
- Faces PART 1 Database.









大学共同利用機関法人 情報・システム研究機構
統計数理研究所
連絡先：鄭 寧
Email：nzheng@ism.ac.jp
The Institute of Statistical Mathematics