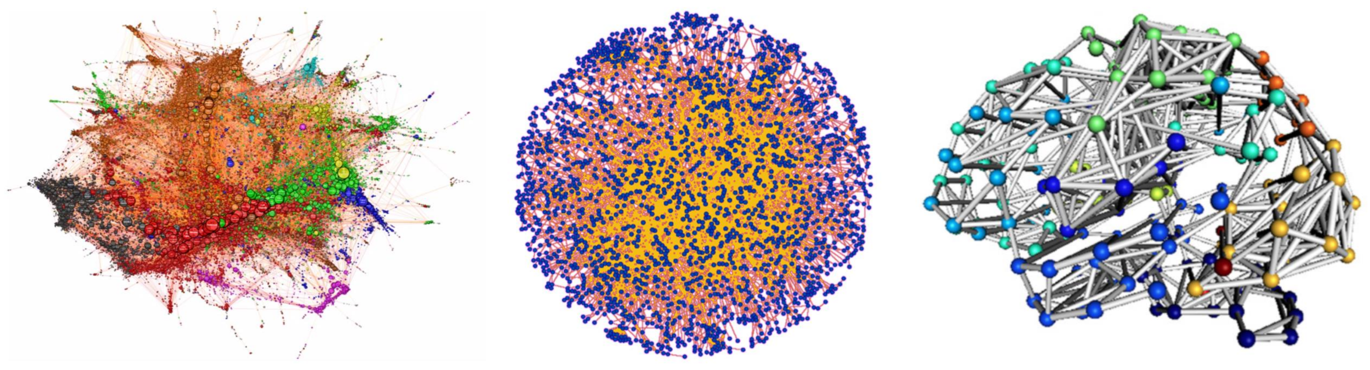
Constrained Tensor Network Nonconvex Optimization with Graph Structured Data via Randomized Iterations

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 multidimensional data with high dimensionality structures;
- Tensor decompositions aim to represent a higher-order (or multi-dimensional) data as a multilinear product of several latent factors;
- Based on low rank approximation which avoids the “curse of dimensionality”;
- Constrained tensor decomposition is a powerful tool for the extraction of parts-based and physically meaningful latent components while preserving multilinear structure;
- From the viewpoint of optimization, the objective function is a nonconvex, nonsmooth with a fidelity term and a regularization term;

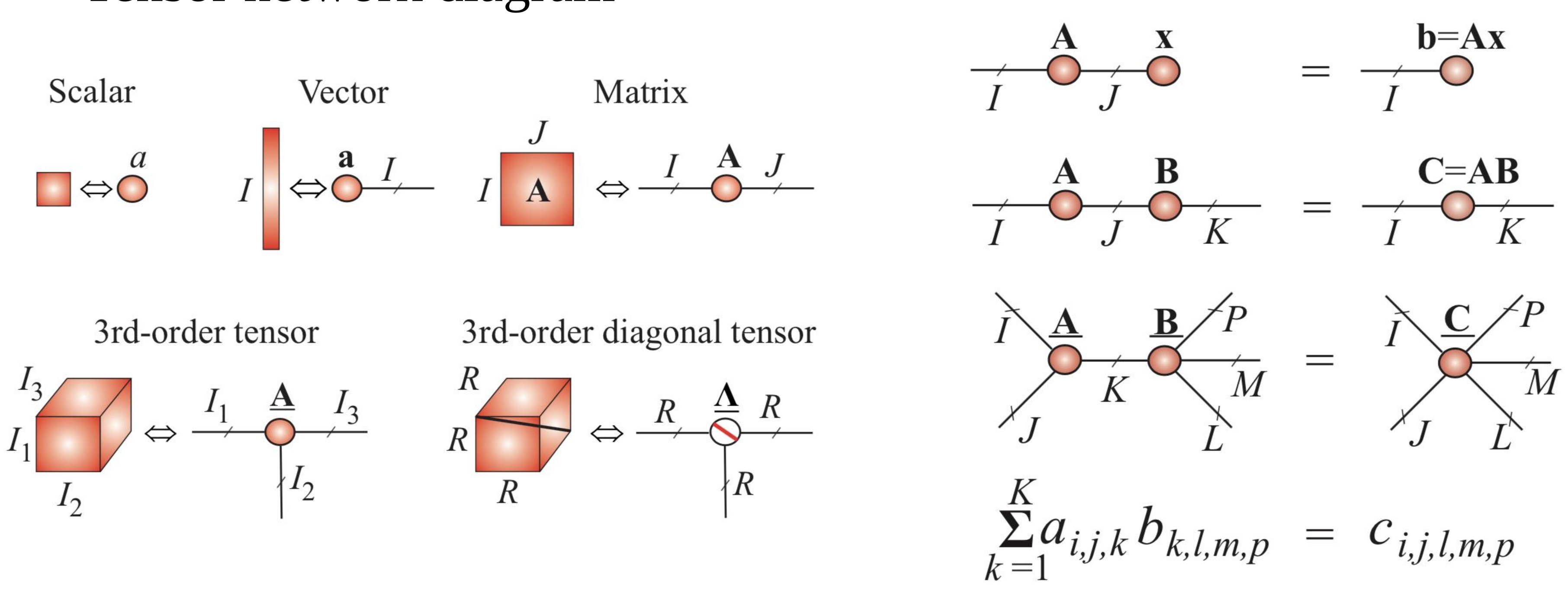


Internet, Biological Networks, Brain Graphs

- Robust and accurate numerical methods are required since the problem is ill-posed, and the numerical solutions are sensitive to the perturbation of input data.

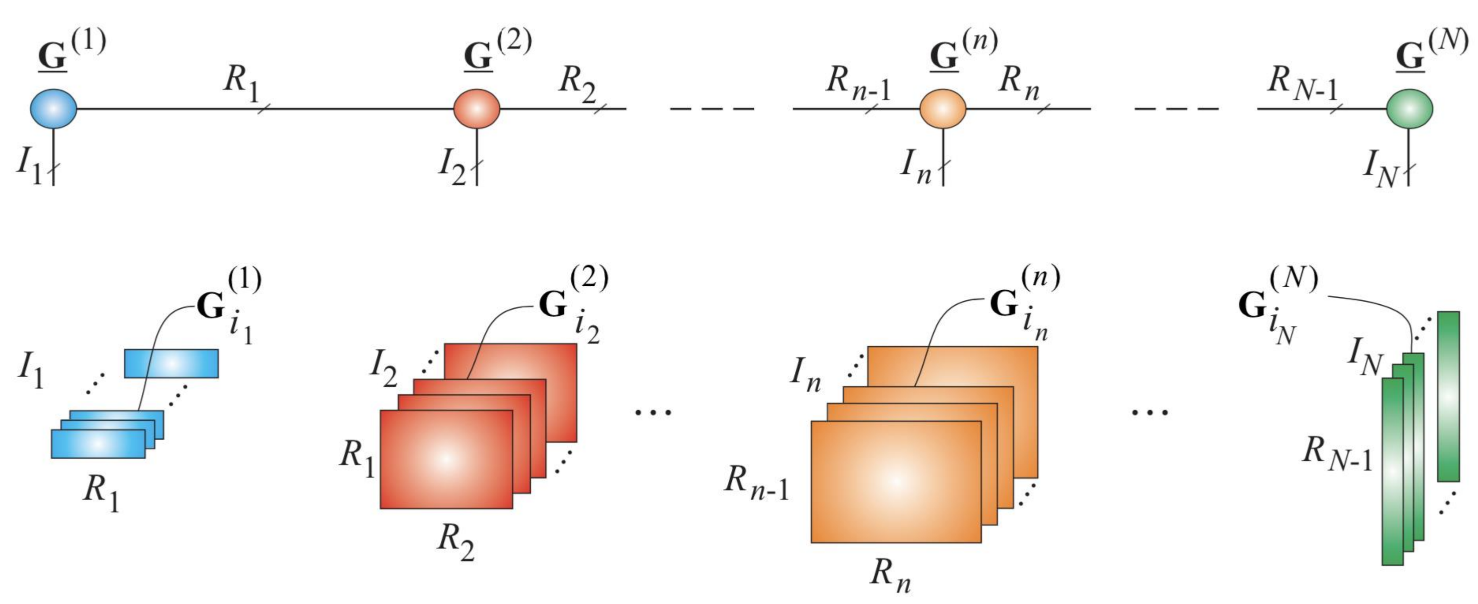
Tensor Networks

- 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;
- Tensor network diagram

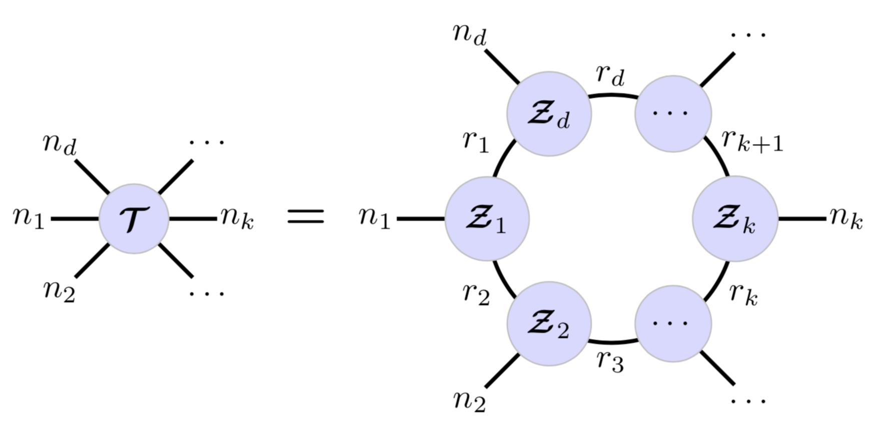


Tensor Train / Ring and Matrix Product State

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



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



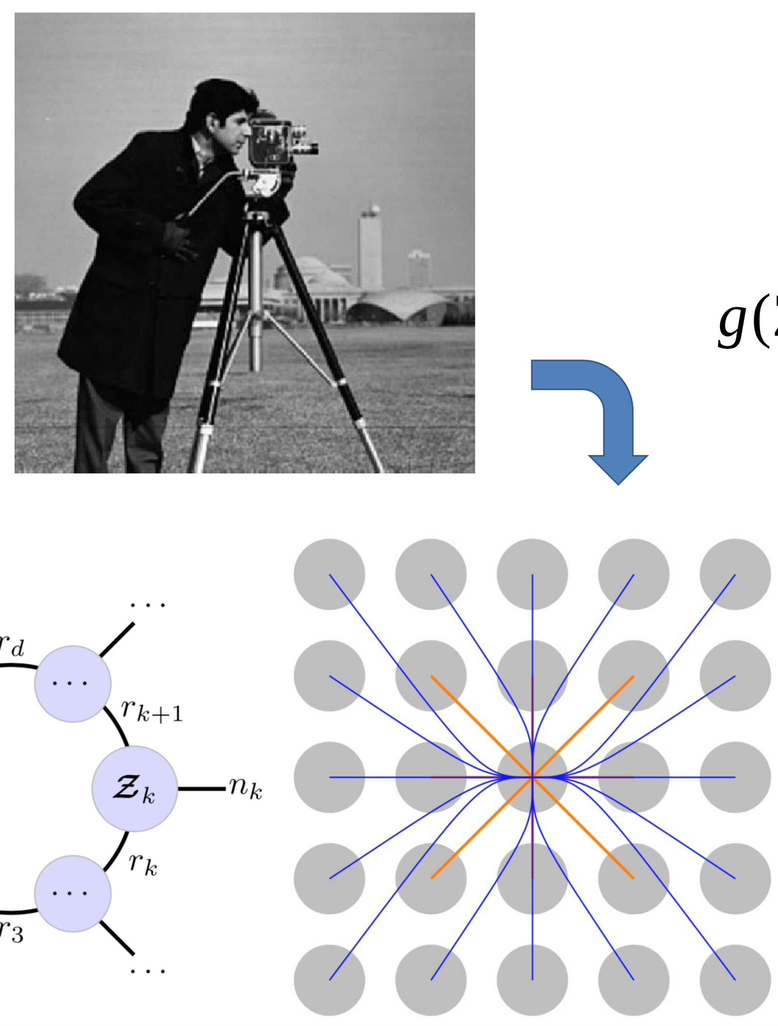
Tensor Networks Nonconvex Optimization

- 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:

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



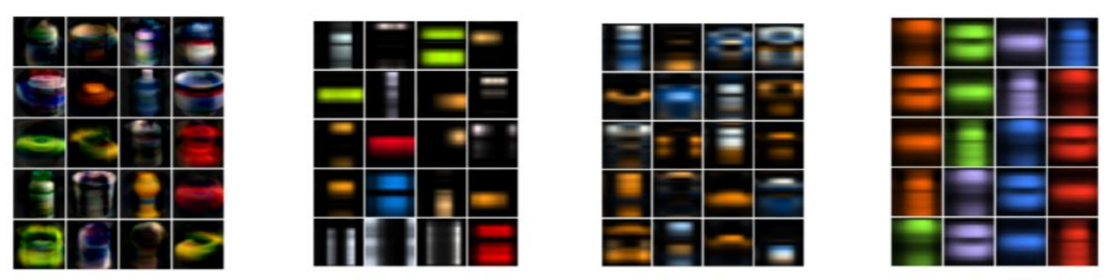
- Task 1: graph topology, geometrical information of data can be obtained by modeling a neighbor graph;

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

- Task 2: nonnegative tensor network optimization. Other related models: nonnegative matrix factorization (NMF), NCP, NTD, NTT, etc.
- Task 3: rank-selection randomized greedy block coordinate descent iterative algorithm. Solve the subproblem via randomized iterations.

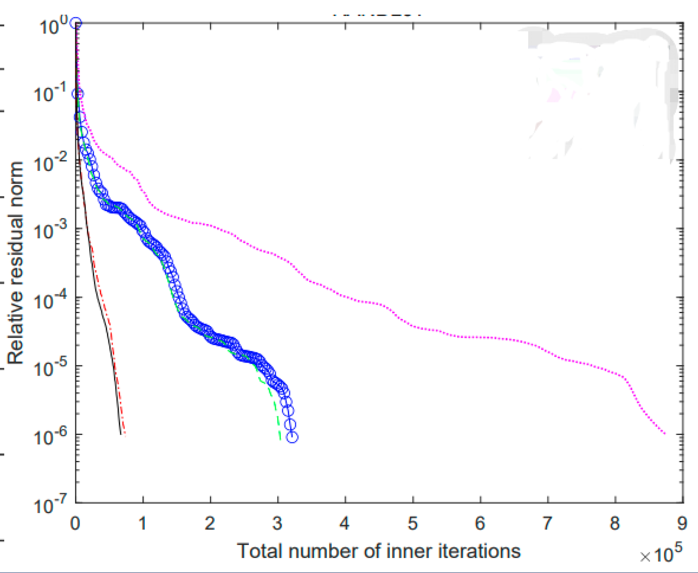
Numerical Experiments

- Randomly generated tensor with Gaussian distributions;
- COIL-100 (first 20 objects);
- Georgia Tech Face.



(f) GNMF (g) LRNTF (h) GNTD (i) GNTR

Databases	Metric	Methods					
		NMF	NCP	NTD	NTT	NTR	GNMF
AT&T ORL	AC	67.6±2.7	66.1±2.4	66.8±2.6	69.6±1.9	71.5±2.8	70.9±2.7
	NMI	82.6±1.9	81.7±1.4	82.1±1.6	83.7±1.2	85.6±1.4	85.2±1.4
Yale	AC	51.9±2.7	49.8±3.7	52.5±5.5	52.1±2.8	47.7±3.0	51.8±3.7
	NMI	55.5±1.9	52.1±2.8	55.3±3.5	55.4±2.3	51.3±2.6	54.1±2.5
Umist	AC	41.2±2.1	41.8±2.9	40.1±2.6	38.4±1.6	43.5±3.2	43.7±2.2
	NMI	59.5±1.7	61.5±2.9	57.4±2.1	57.3±1.4	62.3±2.0	63.4±1.7
COIL-100	AC	70.0±3.2	69.4±2.5	69.5±3.0	70.9±3.8	74.0±3.5	72.3±2.9
	NMI	80.8±1.9	79.5±1.5	79.5±1.7	80.7±2.2	84.2±1.2	83.2±1.0
GT	AC	45.3±1.5	44.9±1.9	39.5±2.0	41.1±2.8	50.3±3.8	46.3±1.8
	NMI	62.8±0.9	62.6±1.4	58.4±1.3	59.3±2.2	67.7±3.0	63.4±1.1



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