# Greedy Orthogonal Pivoting for Non-negative Matrix Factorization

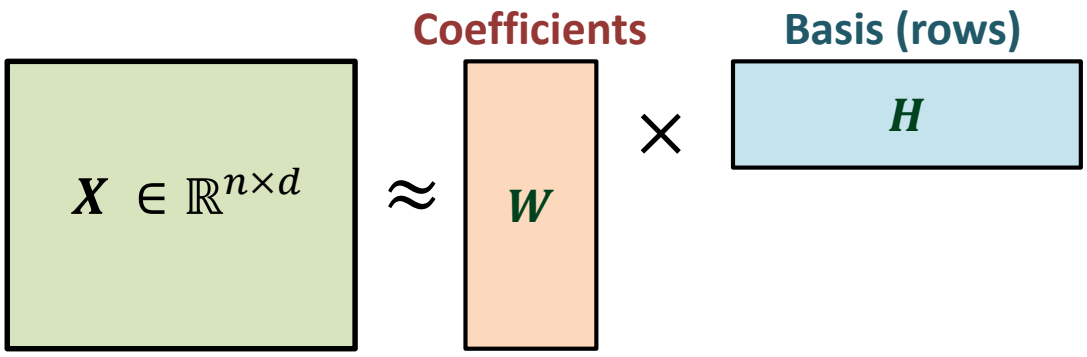
**Kai Zhang, Jun Liu, Jie Zhang, Jun Wang**

**Infinia ML Inc., Fudan University, East China Normal University**

# Non-negative Matrix Factorization

- Represent data with non-negative basis [Lee & Seung, 2000][Ding et al. 2006]

$$\min_{W,H \geq 0} \quad \|X - WH\|^2$$




- Applications
  - Signal separation, Image classification, Gene expression analysis, Clustering…

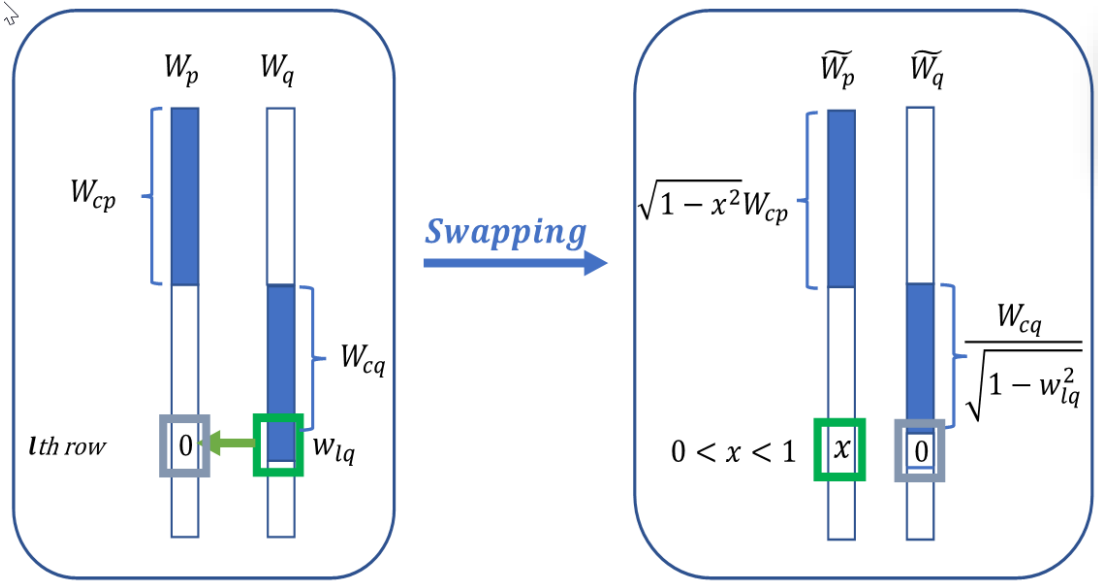
# Orthogonal NMF

$$\min_{W,H \geq 0} \quad \|X - WH\|^2 \qquad s.t. \;\; W'W = I$$

- Motivation
  - NMF optimization is ill-posed
  - Task Preferences (cluster indicator matrix)
- Existing Methods
  - Multiplicative updates [Ding et. al. 2006]
  - Soft orthogonality constraints [Shiga et al. 2014, Lin 2007]
  - Clustering-based formulation [Pompili et al. 2014]
- Challenges
  - Zero-locking problem
  - Level of orthogonality hard to control

# Greedy Orthogonal Pivoting Algorithm

- A Group-coordinate-descent with adaptive updating variables and closed-form iterations



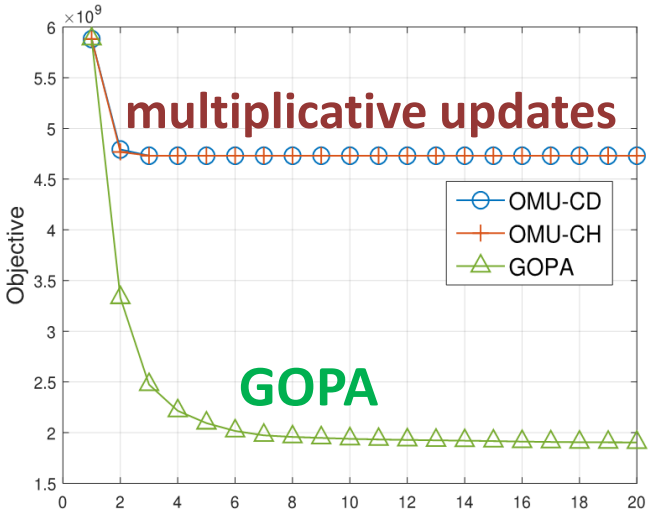

$$\max_{0<x<1} \sqrt{1-x^2}(\mathbf{W}'_{cp}\mathbf{R}_{p[\mathcal{I}_p]}) + x\mathbf{R}_{p[l]}.$$

$$x^* = \frac{\mathbf{R}_{p[l]}}{\sqrt{\mathbf{R}^2_{p[l]} + (\mathbf{W}'_{cp}\mathbf{R}_{p[\mathcal{I}]})^2}}.$$

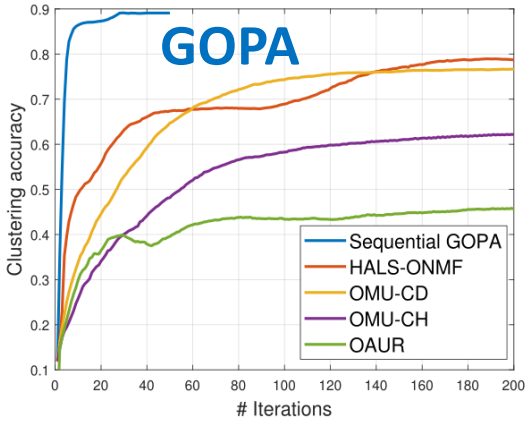
- Exact orthogonality, easy to implement, faster convergence (batch-mode and randomized version)

# Empirical Observations

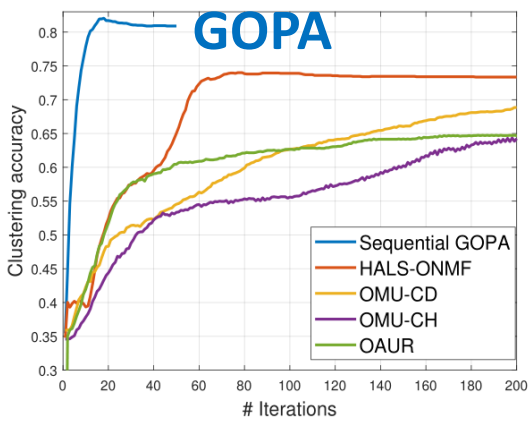
- Avoid zero-locking
  - when starting from a feasible (sparse) solution, GOPA avoids pre-mature convergence



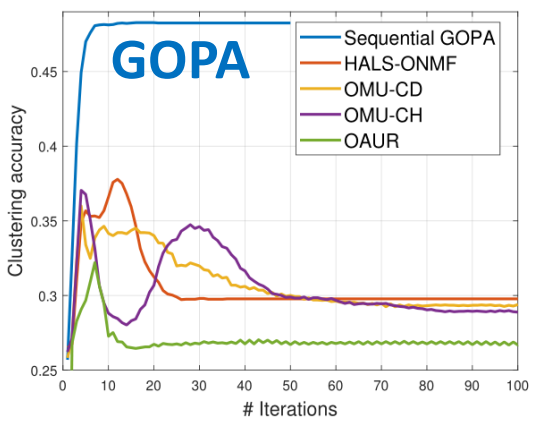

- Faster Convergence




(a) UCI data




(b) DNA data




(c) Newsgroup4 Data

# Future Work

- Adaptive control of sparsity (or orthogonality)
- New way of decomposition into sub-problems
- Probabilistic error guarantee

# Thank You !