Distributed-Memory Parallel Symmetric Nonnegative Matrix Factorization

Srinivas Eswar[1], Koby Hayashi[1], Grey Ballard[2], Ramakrishnan Kannan[3], Richard Vuduc[1], Haesun Park[1]
[1] Georgia Institute of Technology, [2] Wake Forest University, [3] Oak Ridge National Laboratory

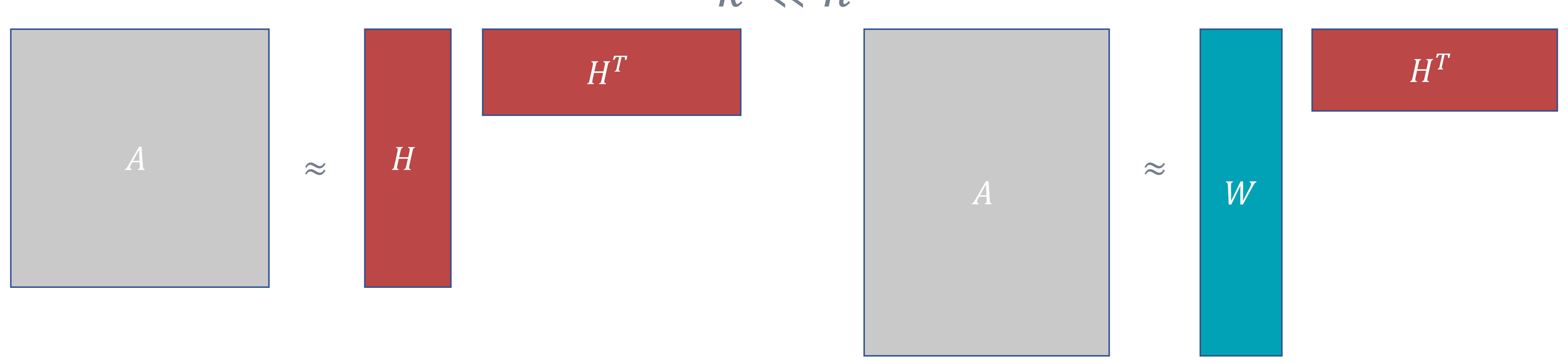
What is SymNMF?

1. Many **data mining** tasks can be cast as low-rank constrained matrix approximation.
2. Standard tasks which can be handled via **matrix factorization**.
   1. **Graph clustering** in undirected graphs.
   2. **Image segmentation** given pixel-pixel similarities.
   3. **Topic modelling** given word co-occurrences among documents.
3. Adding **nonnegativity** constraints often makes solutions **interpretable** and **parts-based**.

$$\min_{H \geq 0} \|A - HH^T\|_F$$

where,

$$A \in \mathbb{R}_+^{n \times n} \quad A = A^T$$
$$H \in \mathbb{R}_+^{n \times k}$$
$$k \ll n$$

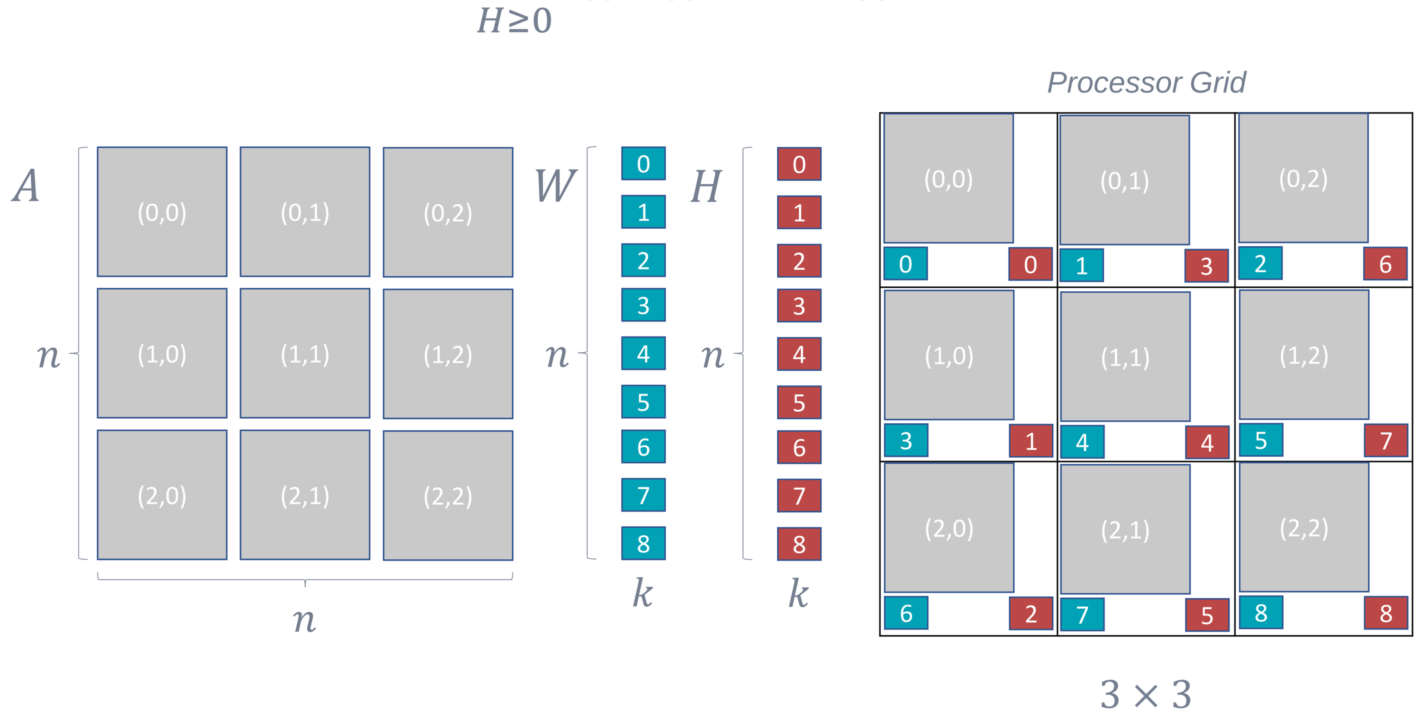


PLANC and general strategy

PLANC is a software library for nonnegative matrix and tensor factorization.

1. **2D parallel matrix multiply** for NMF.
2. **Alternating update** strategy for factor matrices.
3. Includes **popular solvers** for nonnegative least squares.

$$\min_{W \geq 0} \|A - WH^T\|_F$$
$$\min_{H \geq 0} \|A - WH^T\|_F$$
$$\min_{H \geq 0} W^TWH^T = W^TA$$



*Eswar et al. "PLANC: Parallel Low Rank Approximation with Non-negativity Constraints." arXiv preprint (2019).

ANLS based SymNMF

1. **Drop** the symmetry constraints and **regularize!**
2. Modify NLS formulation and **reuse PLANC** adding pairwise communication.

$$\min_{W,H \geq 0} \|A - WH^T\|_F + \gamma\|W - H\|_F$$

$$\min_{H \geq 0} \left\| \begin{bmatrix} W \\ \sqrt{\gamma} I_k \end{bmatrix} H^T - \begin{bmatrix} A \\ \sqrt{\gamma} W^T \end{bmatrix} \right\|_F^2$$

$$\min_{H \geq 0} (W^TW + \gamma I_k)H^T = W^TA + \gamma W^T$$

*Kuang et al. "SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering." JOGO (2015)

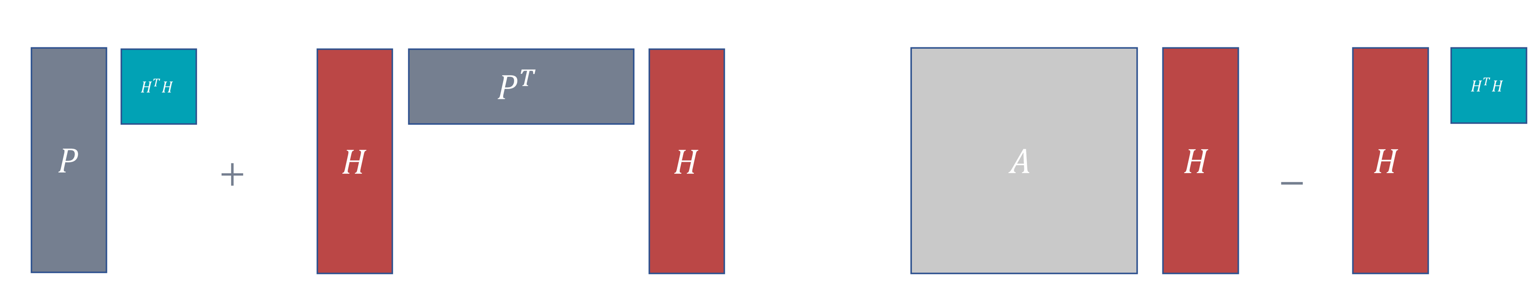
Gauss-Newton based SymNMF

$$H_{(t+1)} = \left[H_{(t)} - P_{(t)}\right]_+$$
$$P_{(t)} = \text{reshape}\left(\underset{p \in \mathbb{R}^{nk}}{\text{argmin}} \|J_{(t)}p - r_{(t)}\|_F^2\right)$$
$$J_{(t)}^T J_{(t)} p = J_{(t)}^T r_{(t)}$$

1. Iteratively update guess along **step direction**.
2. Step direction is an unconstrained least squares problem using the **Jacobian** and **residual**.
3. Solve using **Conjugate Gradient** and project back to nonnegative orthant.

$$J_{(t)} = -(H_{(t)} \otimes I_n) - P_{n,n}(H \otimes I_n)$$
$$J_{(t)}^T r_{(t)} = J_{(t)}^T \text{vec}(A - H_{(t)}H_{(t)}^T) = -2\text{vec}(AH_{(t)} - H_{(t)}H_{(t)}^T H_{(t)})$$
$$J_{(t)}^T J_{(t)} p_{(t)} = J_{(t)}^T J_{(t)} \text{vec}(P_{(t)}) = 2\text{vec}(P_{(t)}H_{(t)}^T H_{(t)} + H_{(t)}P_{(t)}^T H_{(t)})$$

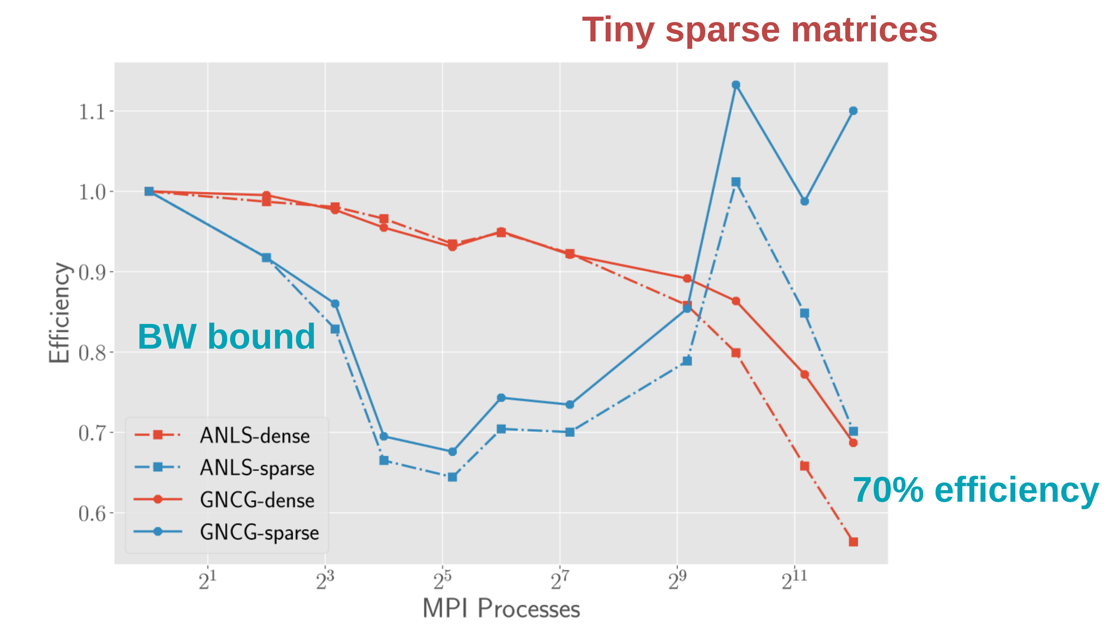
1. Exploit the **structure** of the Jacobian to reduce prohibitive memory and computation costs of applying it.
2. Conjugate Gradient needs only **matrix application!**

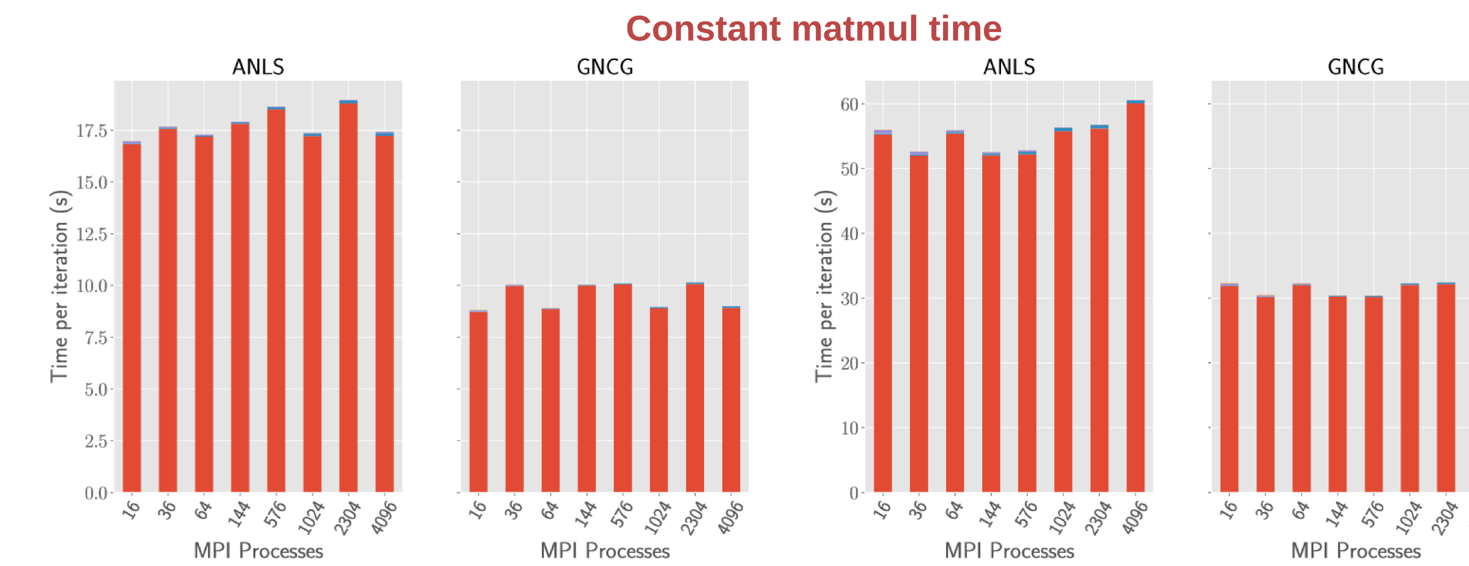


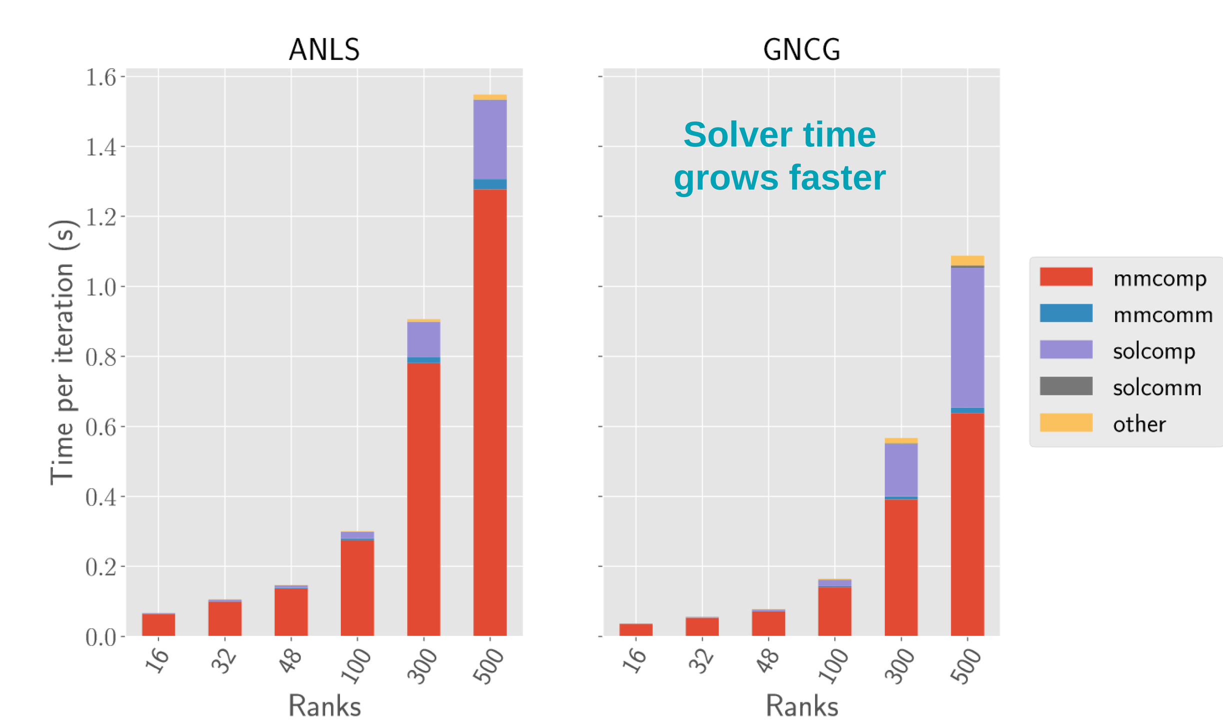
Check out our SC paper and presentation!

Experiments

1. Scaling experiments conducted on Summit.
2. Flat MPI setting – 1 MPI process per core.

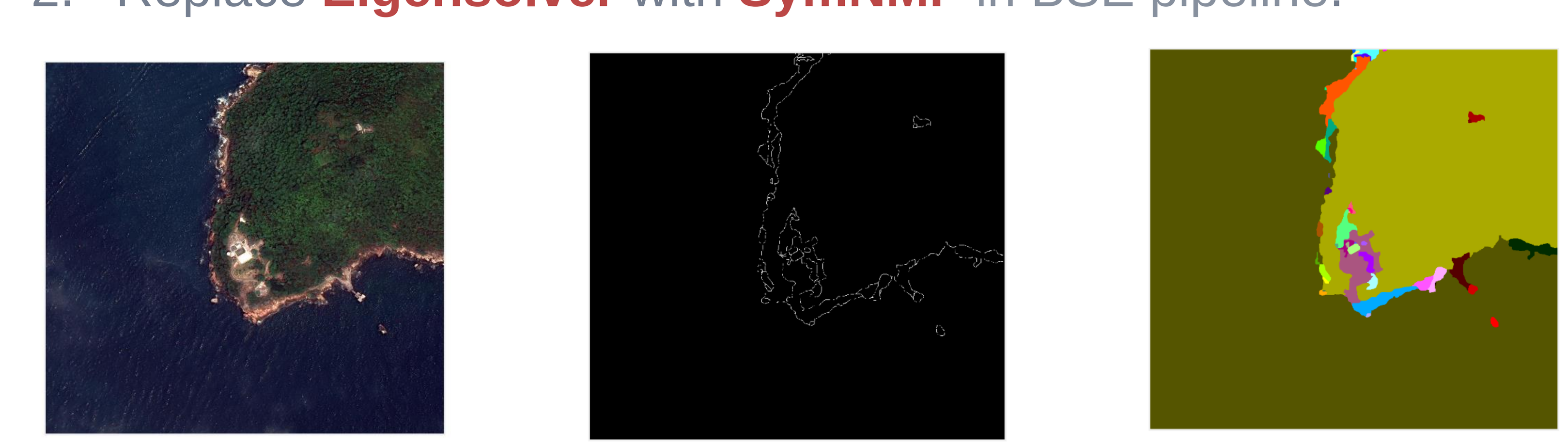






Applications

1. Segmenting large **satellite images** with **2M-20M pixels**.
2. Replace **Eigensolver** with **SymNMF** in BSE pipeline.



*Arbelaez et al. "Contour detection and hierarchical image segmentation." TPAMI (2010)