Distributed Trace Ratio Optimization in Fully-Connected Sensor Networks

Cem Ates Musluoglu and Alexander Bertrand

Presentation by Cem Ates Musluoglu

cemates.musluoglu@esat.kuleuven.be

The Trace Ratio Optimization (TRO) Problem

- Various uses in signal processing and machine learning such as:
  - Dimensionality reduction
  - Signal enhancement
  - Pattern detection
- Optimization problem consisting of finding a filter or projection matrix X^* which solves:

$$\begin{aligned} \underset{X}{\text{maximize}} \quad & \varrho(X) \triangleq \frac{\operatorname{tr}(X^T A X)}{\operatorname{tr}(X^T B X)} \\ \text{subject to} \quad & X^T X = I \iff X \in \mathcal{S} \end{aligned}$$

- Similar, **but different** to the generalized eigenvalue decomposition (GEVD) problem:

$$\begin{aligned} \underset{X}{\text{maximize}} \quad & \frac{\operatorname{tr}(X^T A X)}{\operatorname{tr}(X^T B X)} \\ \text{subject to} \quad & X^T B X = I \end{aligned}$$

Iterative Algorithm for TRO

- Iterative methods have been described for solving the TRO problem [Wang et al., 2007], [Ngo et al., 2012]:

Initialize X^0 randomly, $\rho^0 = \varrho(X^0)$

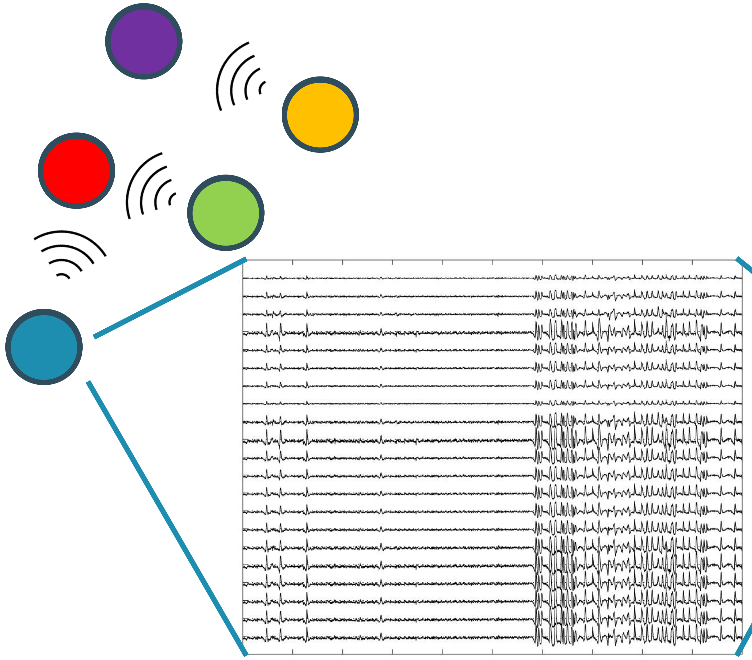
- $X^{i+1} \leftarrow \mathrm{EVC}(A - \rho^i B)$
- $\rho^{i+1} \leftarrow \varrho(X^{i+1})$

Repeated until convergence

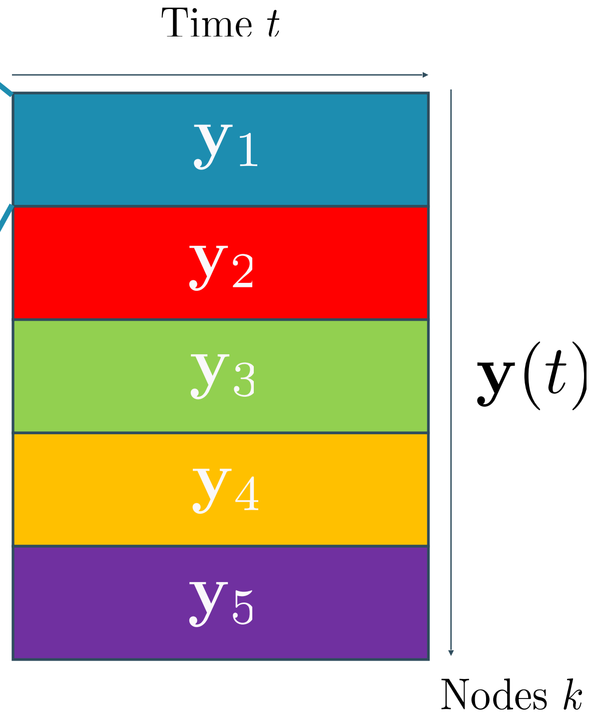
[Wang et al., 2007] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, “Trace ratio vs. ratio trace for dimensionality reduction,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007, pp. 1–8.
[Ngo et al., 2012] T. T. Ngo, M. Bellalij, and Y. Saad, “The trace ratio optimization problem,” SIAM review, vol. 54, no. 3, pp. 545–569, 2012.

Wireless Sensor Networks (WSNs)

- Node k measures multi-channel signals $\boldsymbol{y}_k, \ \boldsymbol{v}_k$ and communicates *linearly compressed* signals $F_k^T \boldsymbol{y}_k, \ F_k^T \boldsymbol{v}_k$ with all other nodes



- The total available data $\boldsymbol{y}$ is a stacking of individual observations



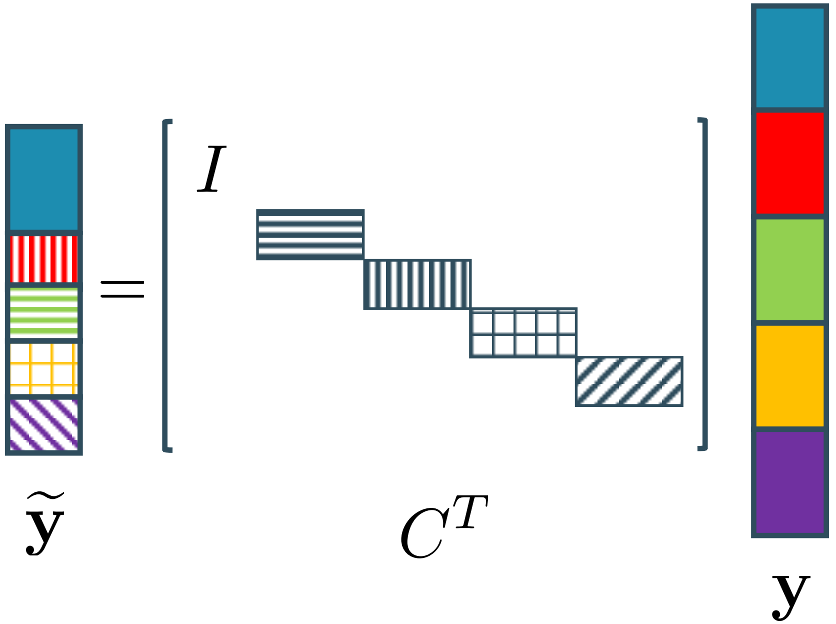
$$\begin{aligned} \underset{X}{\text{maximize}} \quad & \frac{\operatorname{tr}(X^T R_{\mathbf{yy}} X)}{\operatorname{tr}(X^T R_{\mathbf{vv}} X)} \\ \text{subject to} \quad & X^T X = I \end{aligned}$$

- **Goal:** Compute TRO solution on network-wide covariance matrices

Global and local data

TRO

- $X^{i+1} \leftarrow \mathrm{EVC}(R_{\mathbf{yy}} - \rho^i R_{\mathbf{vv}})$
- $\rho^{i+1} \leftarrow \varrho(X^{i+1})$



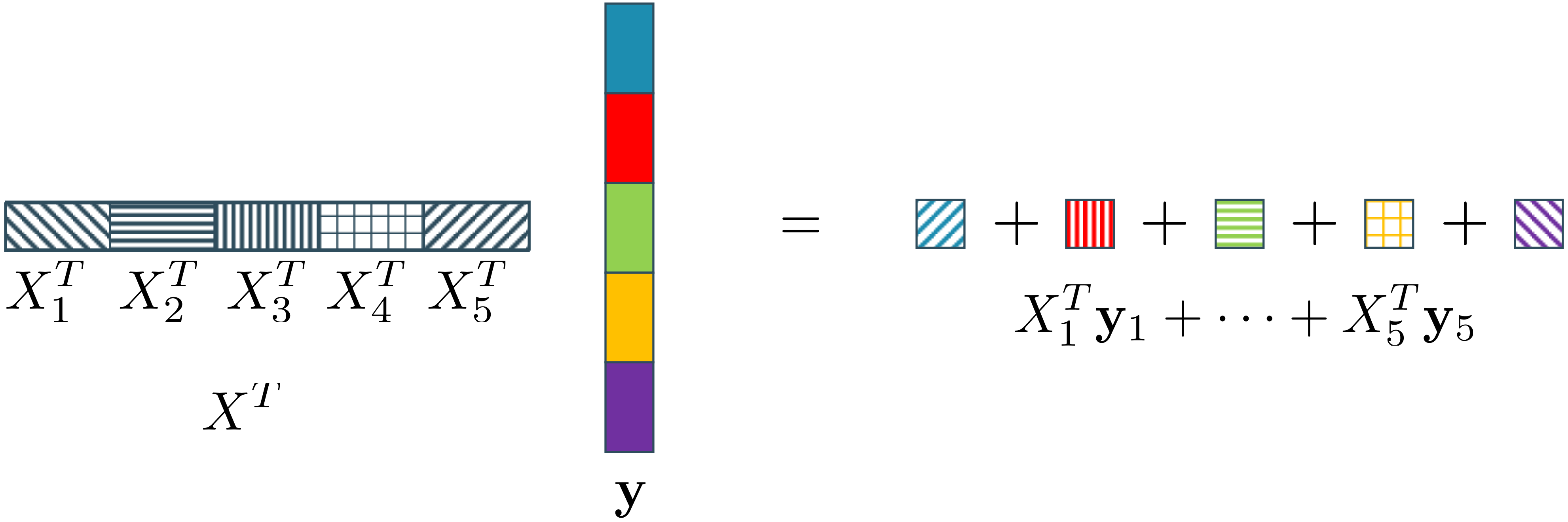
Since nodes do not have access to the full data, they cannot construct the full covariance matrices.

Instead, they are able to construct the covariance matrices $R_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$, $R_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$ of the locally available data.

Observation

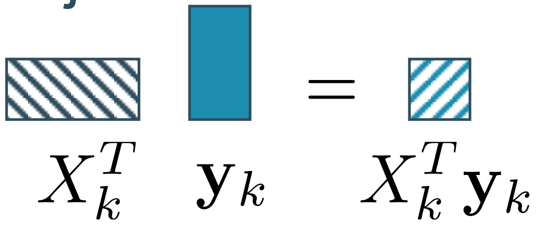
- We observe that:

$$\mathrm{tr}(X^T R_{\mathbf{yy}} X) = \mathrm{tr}(X^T E[\mathbf{y}\mathbf{y}^T] X) = \mathrm{tr}(E[X^T \mathbf{y}\mathbf{y}^T X])$$



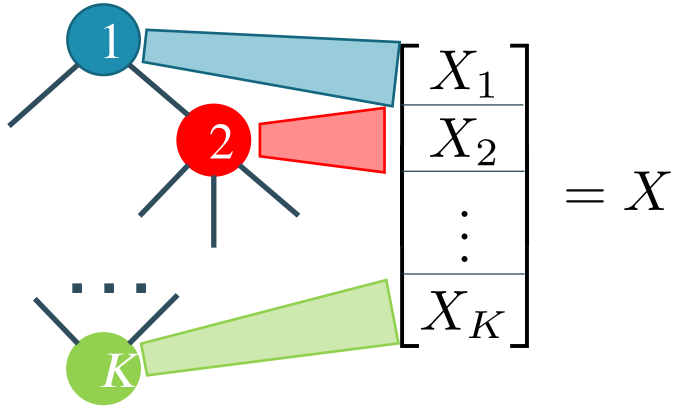
Local compression matrix

- We choose $F_k = X_k$, so that every node can reconstruct the objective value $\varrho(X)$ locally.



- Every node is responsible of updating its own local variable.

- Remark: X_k's act both as a compressor and the variable to be estimated.

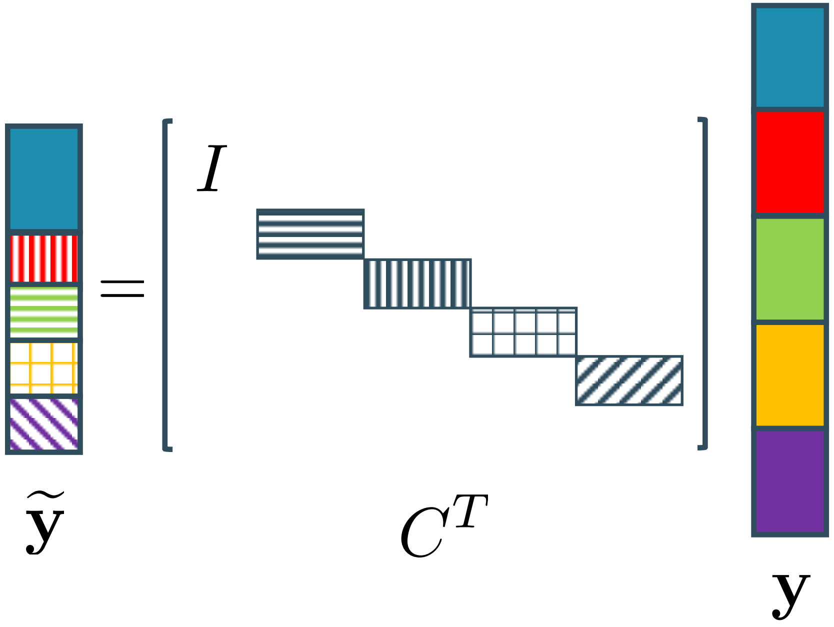


From EVD to GEVD

$$X\Lambda = (R_{\mathbf{yy}} - \rho^i R_{\mathbf{vv}})X \Leftarrow X \leftarrow \mathrm{EVC}(R_{\mathbf{yy}} - \rho^i R_{\mathbf{vv}}) \qquad R_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = C^T R_{\mathbf{yy}} C$$

$$= (R_{\mathbf{yy}} - \rho^i R_{\mathbf{vv}})(C\tilde{X})$$

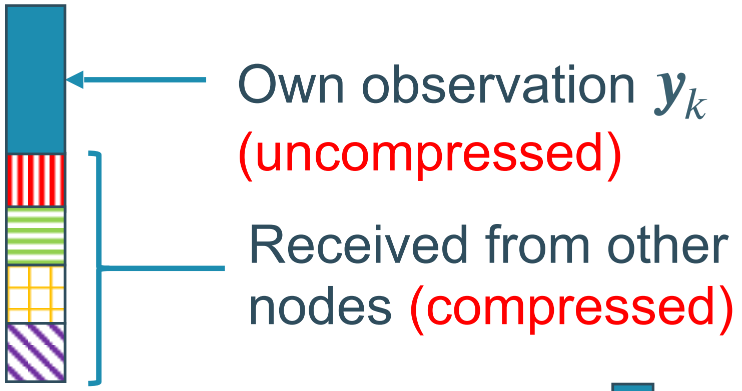
$$\overset{C^T\cdot}{\Rightarrow} (R_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} - \rho^i R_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}})\tilde{X} = C^T C \tilde{X}\tilde{\Lambda} \Rightarrow \tilde{X} \leftarrow \mathrm{GEVC}(R_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} - \rho^i R_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}, C^T C)$$



$$C^T C = \begin{bmatrix} I & & & & \\ & X_2^T X_2 & & & \\ & & X_3^T X_3 & & \\ & & & X_4^T X_4 & \\ & & & & X_5^T X_5 \end{bmatrix}$$

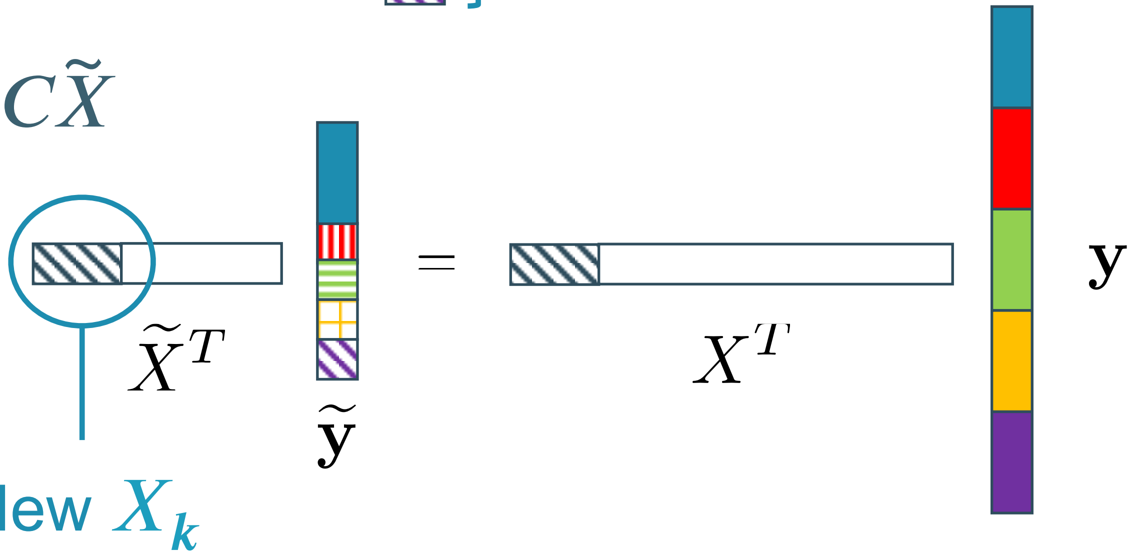
Strategy

- Choose one node k (updating node). Every other node sends their compressed observation to the updating node.



- At the updating node, define new data $\widetilde{\mathbf{y}} = C^T \boldsymbol{y}$.
- Solve GEVD for the variable $\widetilde{X}$: $X = C\widetilde{X}$

  and compute $\rho = \varrho(C\widetilde{X})$



- Repeat for other nodes.

DTRO Algorithm

Full solution of the network-wide EVD problem at each iteration before updating ρ. Can be solved in a distributed setting but would require nested iterations.

Only partial solution of the EVD problem before updating ρ:

- Non-trivial but provable convergence
- Single time scale.

TRO

Initialize X^0 randomly, $\rho^0 = \varrho(X^0)$

- $X^{i+1} \leftarrow \mathrm{EVC}(R_{\mathbf{yy}} - \rho^i R_{\mathbf{vv}})$
- $\rho^{i+1} \leftarrow \varrho(X^{i+1})$

Repeated until convergence

DTRO

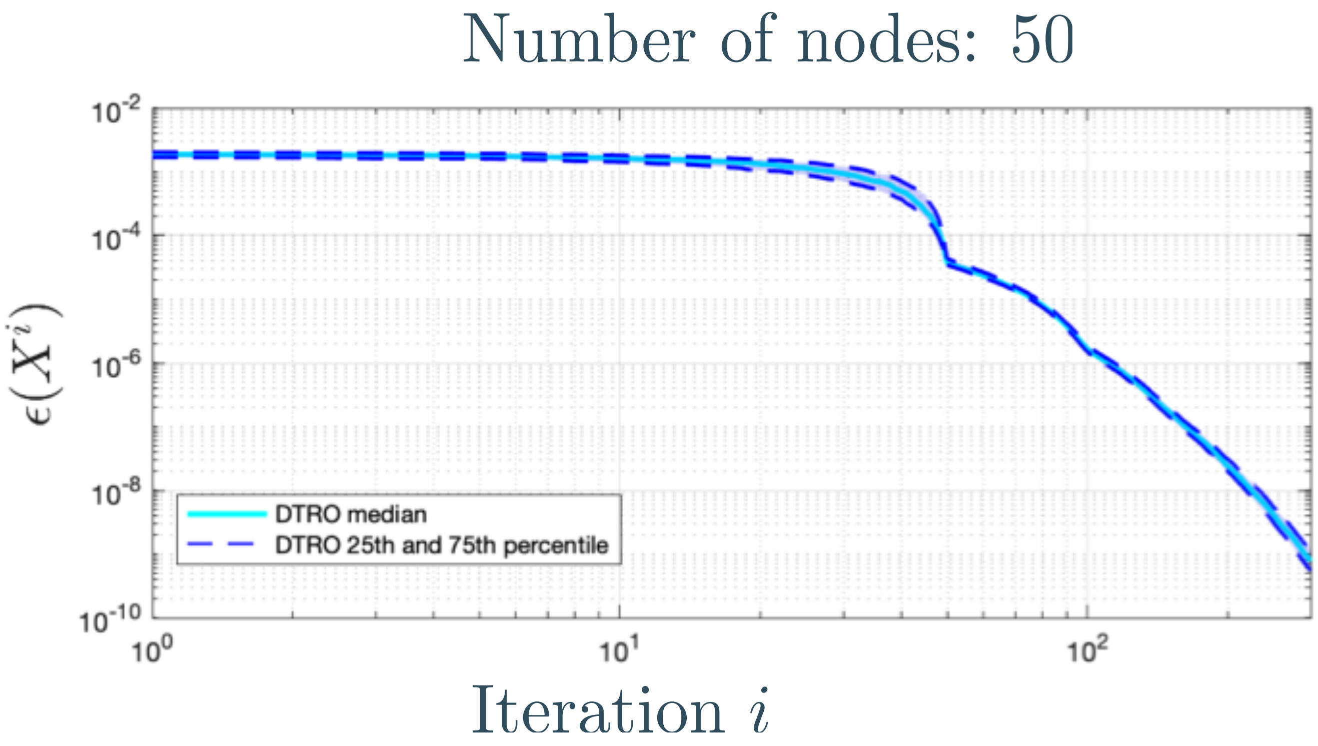
Initialize X^0 randomly, $\rho^0 = \varrho(X^0)$

- $k \leftarrow (i \bmod K) + 1$
- $\widetilde{X}_k^{i+1} \leftarrow \mathrm{GEVC}(R_{\widetilde{\mathbf{y}}_k\widetilde{\mathbf{y}}_k} - \rho^i R_{\widetilde{\mathbf{v}}_k\widetilde{\mathbf{v}}_k}, C_k^T C_k)$
- $\rho^{i+1} \leftarrow \dfrac{\mathrm{tr}(\widetilde{X}_k^{(i+1)T} R_{\widetilde{\mathbf{y}}_k\widetilde{\mathbf{y}}_k} \widetilde{X}_k^{i+1})}{\mathrm{tr}(\widetilde{X}_k^{(i+1)T} R_{\widetilde{\mathbf{v}}_k\widetilde{\mathbf{v}}_k} \widetilde{X}_k^{i+1})}$

Repeated until convergence

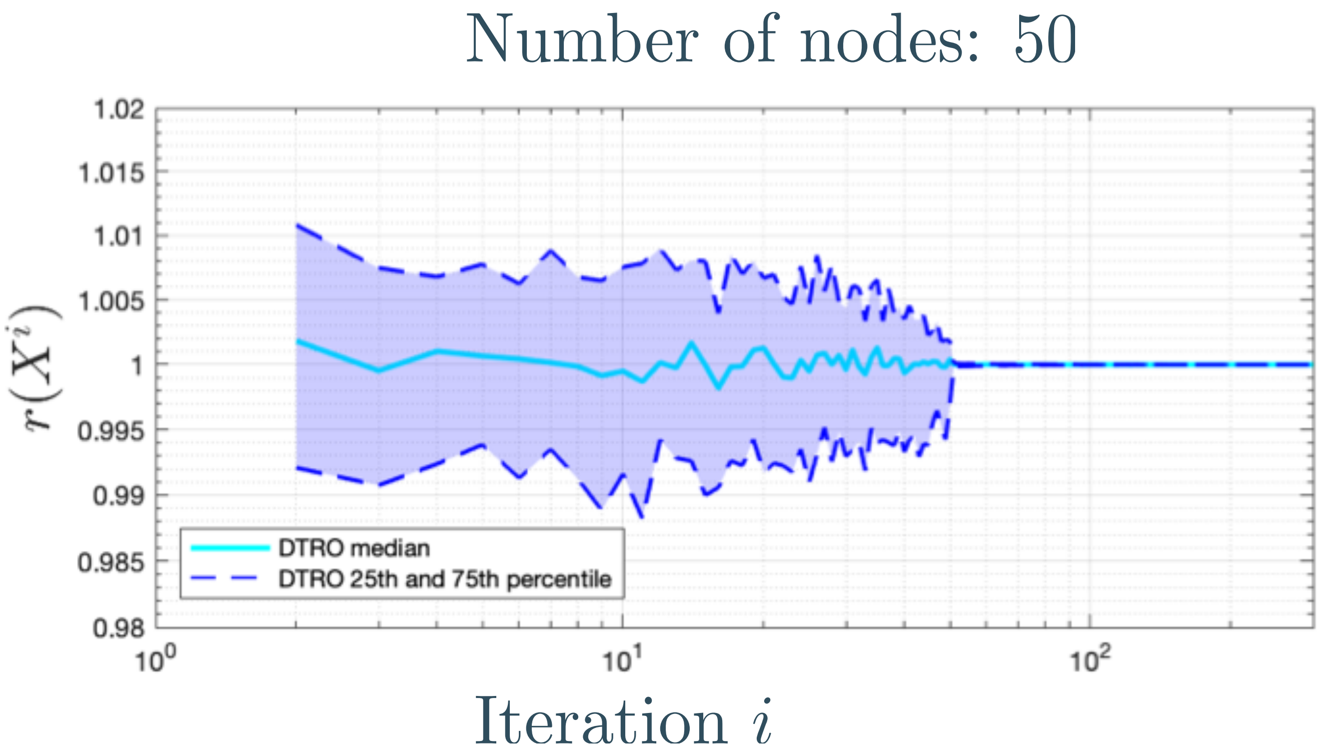
Experimental Results – Mean Squared Error

$$\epsilon(X^i) = \frac{1}{MQ}||X^i - X^*||_F^2$$



Experimental Results – Asymptotic Convergence Rate

$$r(X^i) = \frac{||X^i - X^*||_F^2}{||X^{i+1} - X^*||_F^2}$$



Conclusion and future work

- A distributed algorithm for the TRO problem
- Provable convergence, provided in an extended manuscript currently under review
- DTRO in connected networks
- Study of asynchronous updating schemes