Sparse and Low-Rank Decomposition for Big Data Systems via Smoothed Riemannian Optimization

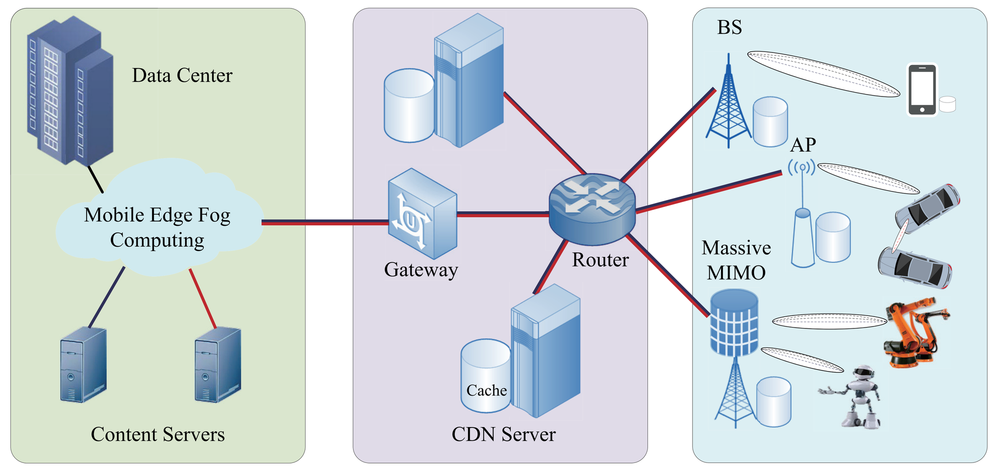
Yuanming Shi, ShanghaiTech University; and Bamdev Mishra, Amazon Development Centre India

Sparse and Low-Rank Decomposition

- Optimize over communication, computation and storage resources in mobile big data systems

$$\mathscr{P}_1: \underset{\boldsymbol{X}\in\mathbb{R}^{K\times K}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{X}) + \lambda\mathcal{R}(\boldsymbol{X}) \quad \text{subject to } \operatorname{rank}(\boldsymbol{X}) = r.$$

–$\mathcal{L}$ is a smooth convex loss function; $\mathcal{R}$ is a non-smooth (possibly non-convex) sparsity inducing function

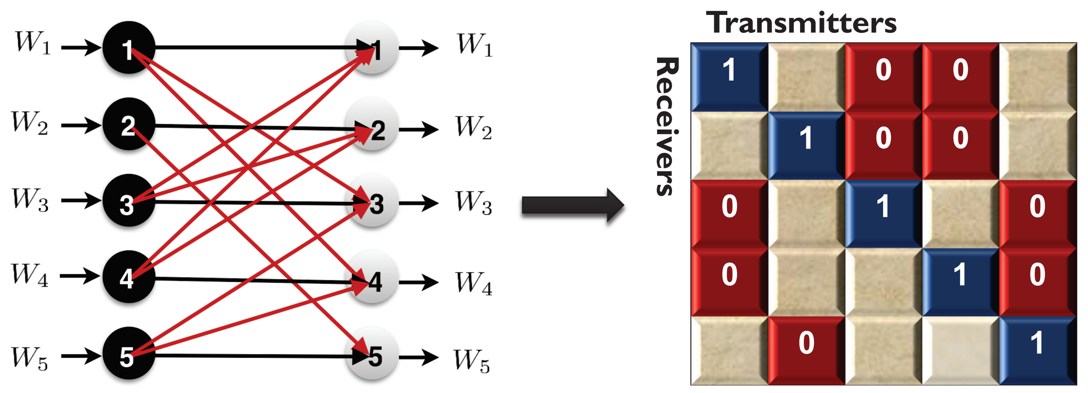


- We will focus on the family that the conventional convex relaxation approaches (e.g., nuclear-norm relaxation) are inapplicable!

Example I: Topological Interference Alignment

- Topological interference alignment problem in partially connected K-user interference network [1]

$$\mathscr{P}_1: \underset{\boldsymbol{X}\in\mathbb{R}^{K\times K}}{\text{minimize}}\ \operatorname{rank}(\boldsymbol{X}) \quad \text{subject to } X_{ii} = 1, \forall i = 1,\ldots,K, \quad X_{ij} = 0, \forall i \neq j, (i,j) \in \mathcal{V}.$$



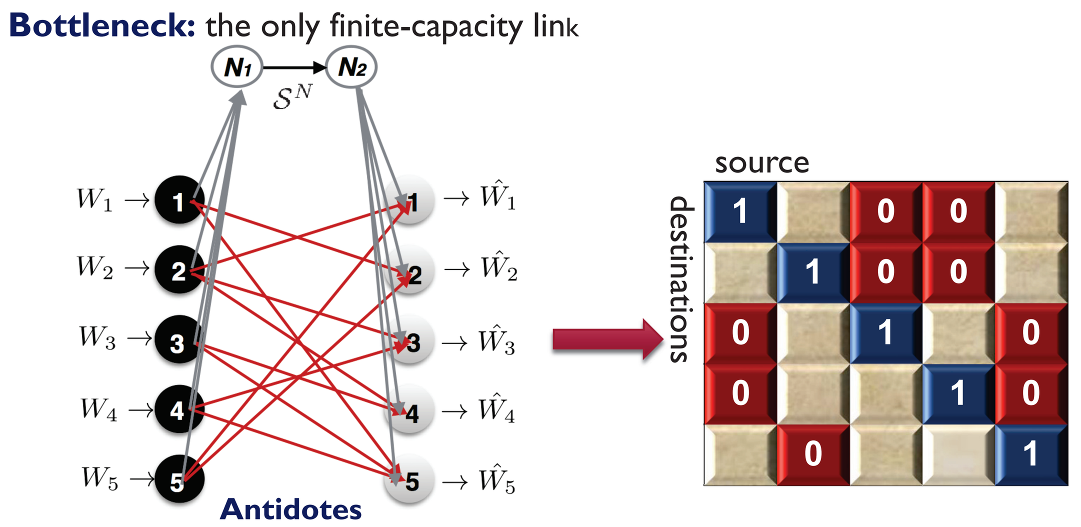
–$\mathcal{V}$ is the index set of connected transceiver pairs

- The well-known nuclear norm relaxation approach fails: always return identity matrix.
- Key conclusion: DoF $= 1/\operatorname{rank}(\boldsymbol{X})$

Example II: Network Topology Control

- Network topology control problem in the partially connected K-user interference network [2]

$$\mathscr{P}_2: \underset{\boldsymbol{X}\in\mathbb{R}^{K\times K}}{\text{minimize}}\ \|\boldsymbol{X}\|_0 \quad \text{subject to } X_{ii} = 1, \forall i = 1,\ldots,K, \quad \operatorname{rank}(\boldsymbol{X}) = r.$$



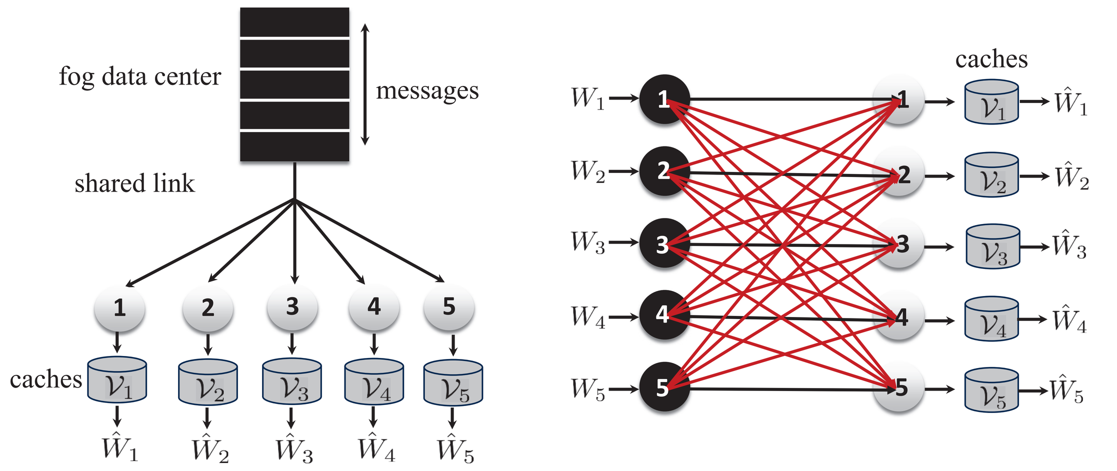
–$\|\boldsymbol{X}\|_0$: # side information for network connectivity, storage and computation load

- The widely used mixed ℓ_1-norm and nuclear norm convex penalty relaxation approach is inapplicable!

Example III: User Admission Control

- User admission control problem in the communication, caching and distributed computing networks [3]

$$\mathscr{P}_3: \underset{\boldsymbol{X}\in\mathbb{R}^{K\times K}}{\text{maximize}}\ \|\operatorname{diag}(\boldsymbol{X})\|_0 \quad \text{subject to } X_{ij} = 0, \forall i \neq j, (i,j) \in \mathcal{V}, \quad \operatorname{rank}(\boldsymbol{X}) = r.$$

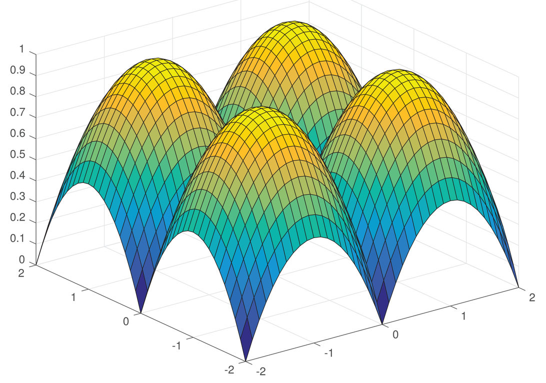


(a) (b)

- A simple ℓ_1-norm relaxation approach yields the objective unbounded and non-convex

Smoothed Regularized Sparsity Inducing Functions

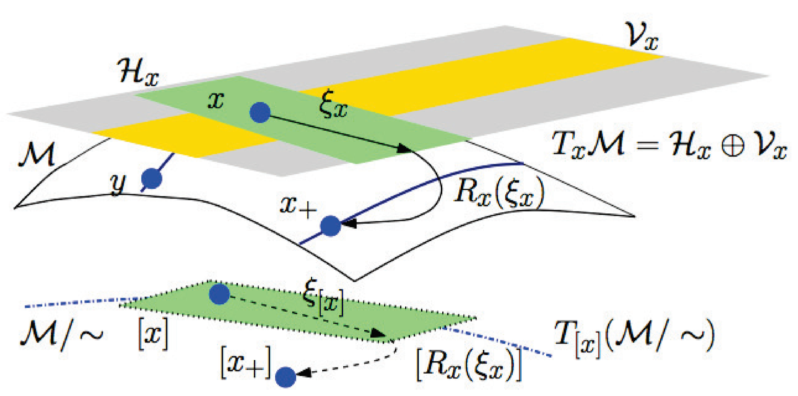
- To find a good sparsity pattern, we propose a smoothed regularized version of the problems $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$ as a smooth fixed-rank optimization problem $\mathscr{P}$.
  - For problem $\mathscr{P}_1$, we have $\mathcal{L}(\boldsymbol{X}) = \sum_i (X_{ii}-1)^2 + \sum_{(i,j)\in\mathcal{V}} X_{ij}^2$ and $\lambda = 0$.
    –Riemannian pursuit: alternatively perform fixed-rank optimization and rank updating
  - For problem $\mathscr{P}_2$, we have $\mathcal{L}(\boldsymbol{X}) = \sum_i (X_{ii}-1)^2$, $\mathcal{R}_\epsilon(\boldsymbol{X}) = \|\boldsymbol{X}\|_{1,\epsilon}$, and $\lambda \geq 0$.
  - For problem $\mathscr{P}_3$, we have $\mathcal{L}(\boldsymbol{X}) = \sum_{(i,j)\in\mathcal{V}} X_{ij}^2$, $\mathcal{R}_\epsilon(\boldsymbol{X}) = \rho\|\operatorname{diag}(\boldsymbol{X})\|_2^2 - \|\operatorname{diag}(\boldsymbol{X})\|_{1,\epsilon}$
    –$\lambda, \rho \geq 0$, $\rho\|\operatorname{diag}(\boldsymbol{X})\|_2^2$ serves the purpose of bounding the overall objective function



The non-convex regularized sparsity inducing function: $f(z) = \|z\|_1 - 0.5\|z\|_2^2$.

Algorithm: Matrix Manifold Optimization

- Solve fixed-rank problem $\mathscr{P}$ by Riemannian optimization
  –Generalize Euclidean gradien (Hessian) to Riemannian gradient (Hessian)



$$\nabla_{\mathcal{M}} f(\mathbf{X}^{(k)}) = P_{\mathbf{X}^{(k)}}(\nabla f(\mathbf{X}^{(k)}))$$

Riemannian Gradient — Euclidean Gradient

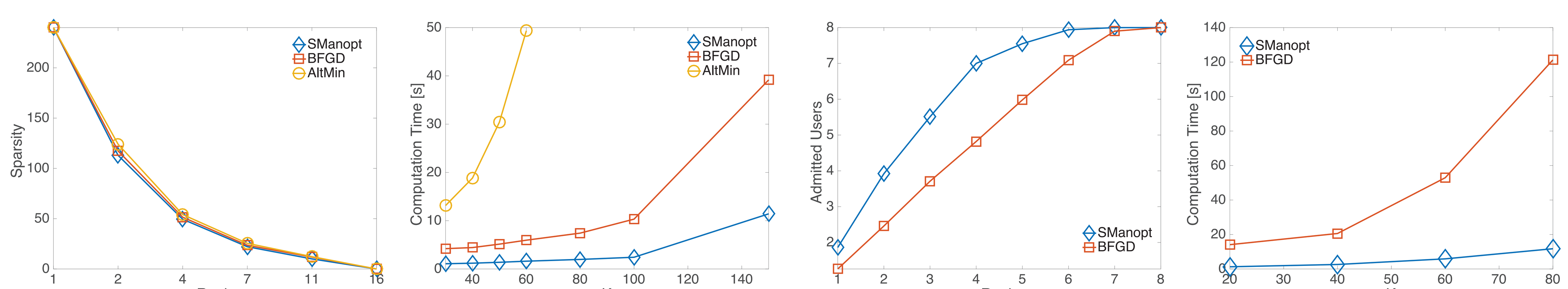
$$\mathbf{X}^{(k+1)} = \mathcal{R}_{\mathbf{X}^{(k)}}(-\alpha^{(k)}\nabla_{\mathcal{M}} f(\mathbf{X}^{(k)}))$$

Retraction Operator

$$[\mathbf{X}] = \{(\mathbf{U}\mathbf{M}^{-1}, \mathbf{V}\mathbf{M}^T) : \mathbf{M} \in \mathrm{GL}(r)\}$$

Experiment Results

- Performance of algorithms on problems $\mathscr{P}_2$ and $\mathscr{P}_2$



(a) Network control (lower is better). (b) Computation time. (c) User control (higher is better). (d) Computation time.

References

[1] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by Riemannian pursuit," IEEE Trans. Wireless Commun., vol. 15, no. 7, Jul. 2016.
[2] Y. Shi and B. Mishra, "A sparse and low-rank optimization framework for index coding via Riemannian optimization," tech. rep., arXiv preprint arXiv:1604.04325, 2016.
[3] Y. Shi and B. Mishra, "Topological interference management with user admission control via Riemannian optimization," arXiv preprint arXiv:1607.07252, 2016.
[4] Y. Shi and B. Mishra, "Sparse and low-rank decomposition for big data systems via smoothed Riemannian optimization," 9th NIPS Workshop on Optimization for Machine Learning (OPT 2016), Barcelona, Spain.

shiym@shanghaitech.edu.cn bamdevm@amazon.com