Low-rank Interaction with Sparse Additive Effects Model for Large Data Frames

Geneviève Robin[1], Hoi-To Wai[2], Julie Josse[1], Olga Klopp[3], Éric Moulines[1]

[1]École Polytechnique, [2]University of Hong Kong, [3]ESSEC Business School

December 6. 2018

Thirty-second Conference on Neural Information Processing Systems

Poster #87
210 & 230 AB
5-7pm

Motivation: species monitoring

White headed duck: endangered

- lead poisoning
- wetland loss

Eurasian curlew: declining

- lead poisoning
- habitat destruction
- disturbances

Waterbirds counts

Y

	2008	2009	2010
site 1	NA	16	32
site 2	299	286	346
site 3	NA	96	151
site 4	NA	NA	NA
site 5	NA	NA	NA
site 6	4647	6054	2442
site 7	16	45	30
site 8	5916	6485	1249

Sites and year covariates

U

Site	Surface	Country	Latitude
1	0.35	Algeria	36.64
2	15.4	Tunisia	34.11
3	1.12	Lybia	35.75
4	0.34	Morocco	35.56
5	2.8	Algeria	34.49
6	2.6	Algeria	35.91
7	0.98	Tunisia	35.75
8	7.2	Morocco	30.36

Year	Spring N/O	Spring N/E	Winter S/O
2008	0,499	1,672	0,505
2009	0,175	2,527	0,215
2010	0,36	-1,453	0,290

1) Characteristics of the data

- ***Mixed***: categorical, real and discrete
- ***Large scale***: 25,000+ survey sites
- ***Incomplete***: missing values
- ***Side information***: row & column covariates

2) Goal: estimate

- ***Main effects***: effect of covariates
- ***Interactions***: the remaining effects

Low-rank Interaction and Sparse main effects

Heterogeneous exponential family parametric model:

$$f_{Y_{ij}}(y) = f_{ij}(y, X_{ij})$$

parameter (unknown)

depends on the entry

Main effects and interactions in parameter space:

$$X_{ij} = \langle u_{ij}, \alpha \rangle + L_{ij}$$

regression term

"residual"

$$X = \sum_{k=1}^{q} \alpha_k U^k + L$$

sparse regression on dictionary

low-rank design

Estimation:

$$(\hat{\alpha}, \hat{L}) \in \operatorname{argmin} \mathcal{L}(Y; X) + \lambda_1 \|L\|_\star + \lambda_2 \|\alpha\|_1$$

Two-fold generalisation of "sparse plus low-rank" matrix recovery

1. **general sparsity pattern**
2. **exponential family noise**

Statistical guarantees

$$(\hat{\alpha}, \hat{L}) \in \operatorname{argmin} \mathcal{L}(Y; X) + \lambda_1 \|L\|_\star + \lambda_2 \|\alpha\|_1$$

↓

Near optimal error bounds for main effects **and** interactions

Theorem 1:

$$\left\|\hat{\alpha} - \alpha^0\right\|_2^2 \leq \frac{\left\|\alpha^0\right\|_1}{\pi} \times \frac{\max_k \|\mathbf{U}(k)\|_1}{\kappa^2} + \mathrm{D}_\alpha$$

$$\|\hat{L} - L^0\|_F^2 \leq \frac{\operatorname{rank}(L^0) \max(n, p)}{\pi} + \mathrm{D}_L$$

Convergence results

Mixed Coordinate Gradient Descent Algorithm:

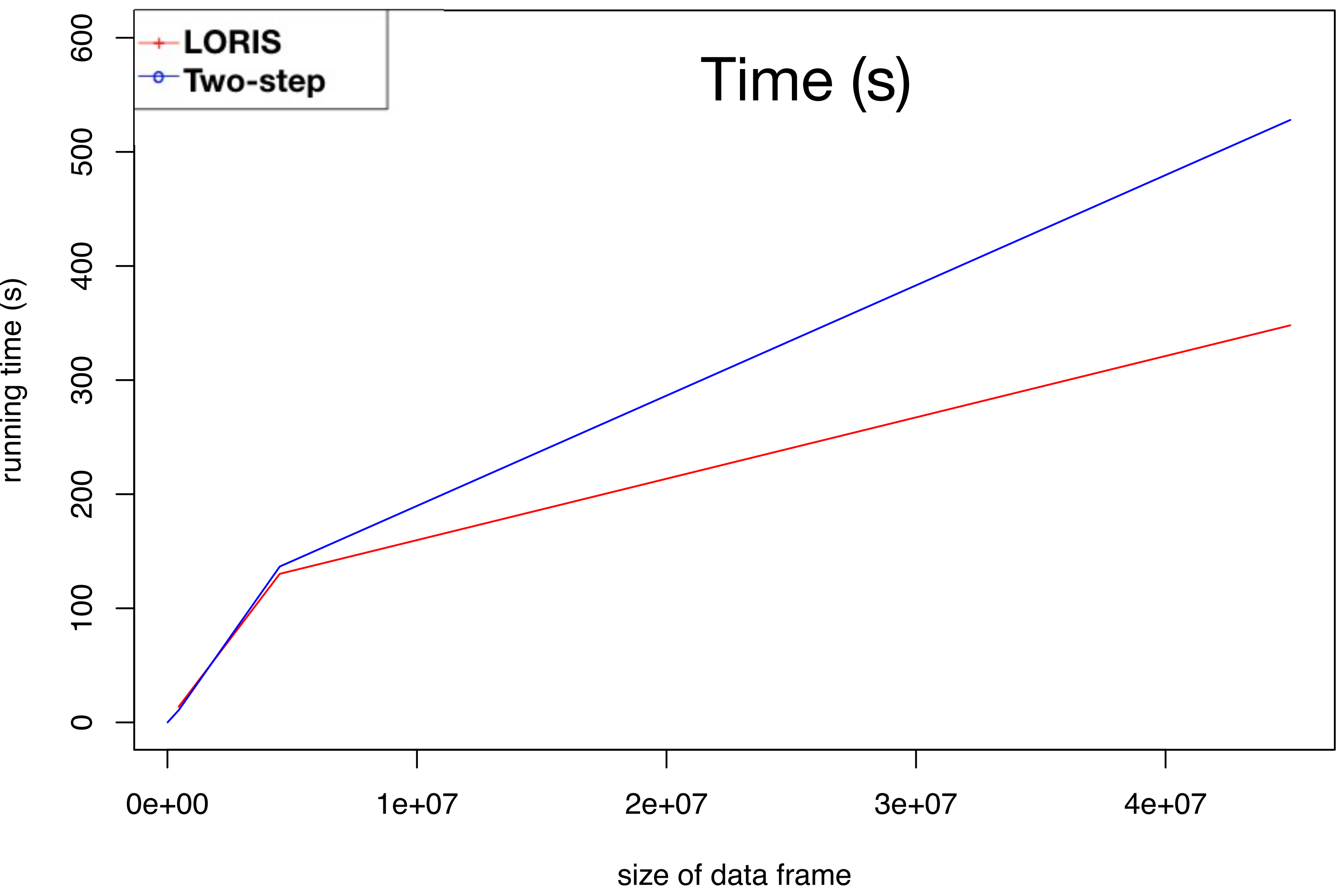
- proximal update for α
- conditional gradient/Franke-Wolfe update for L

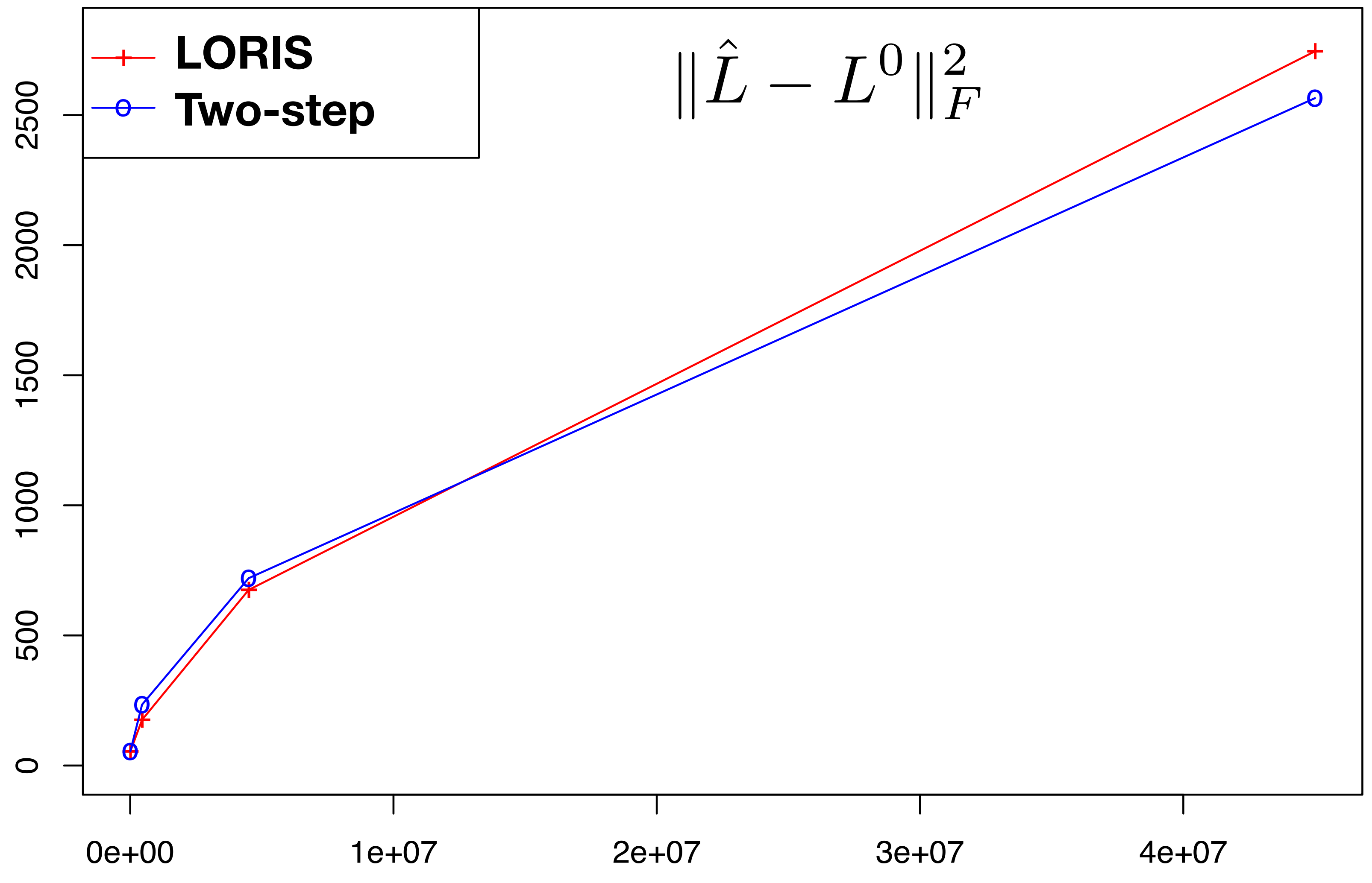
↓

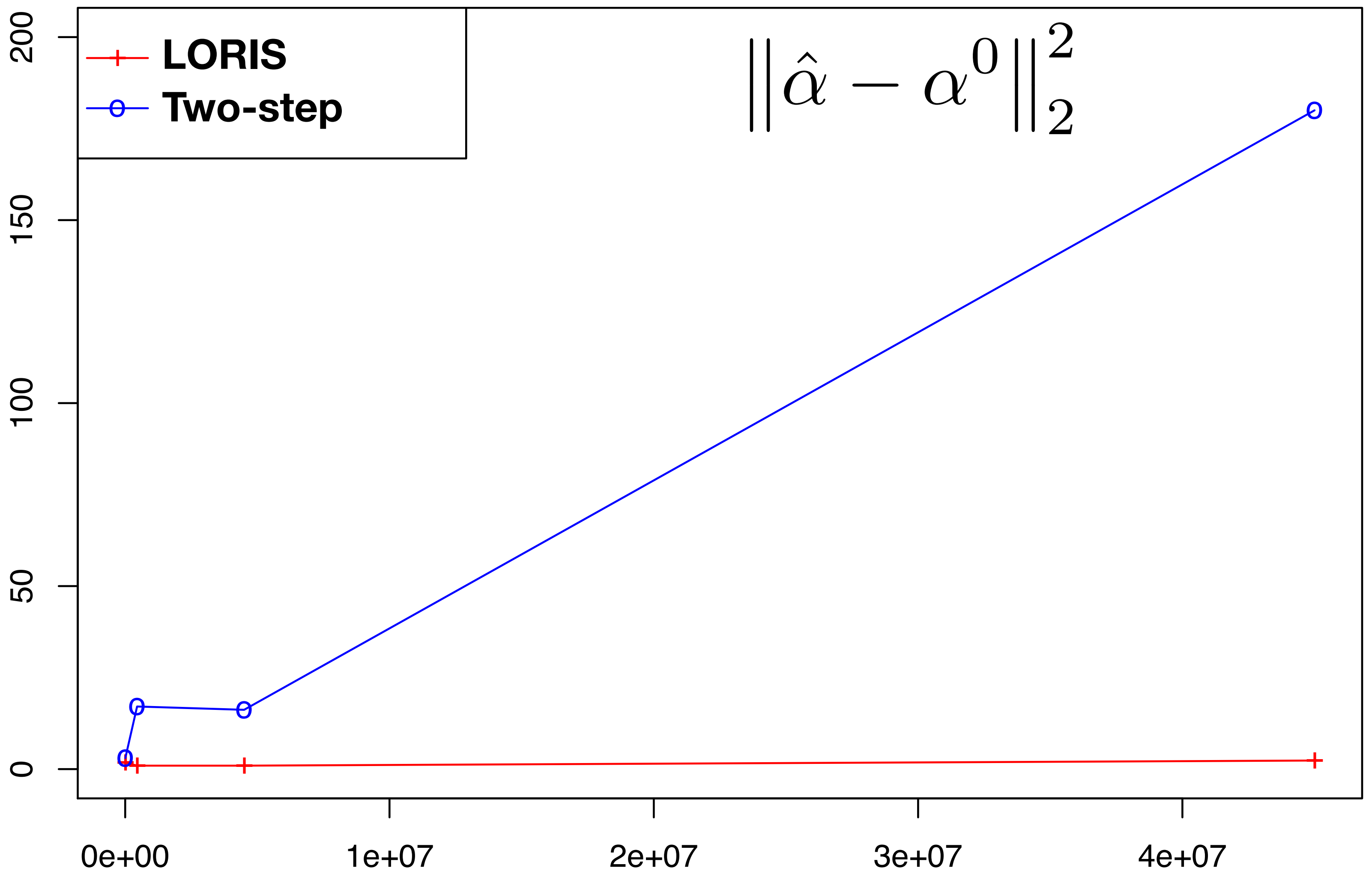
Sublinear convergence and **computationally** efficient

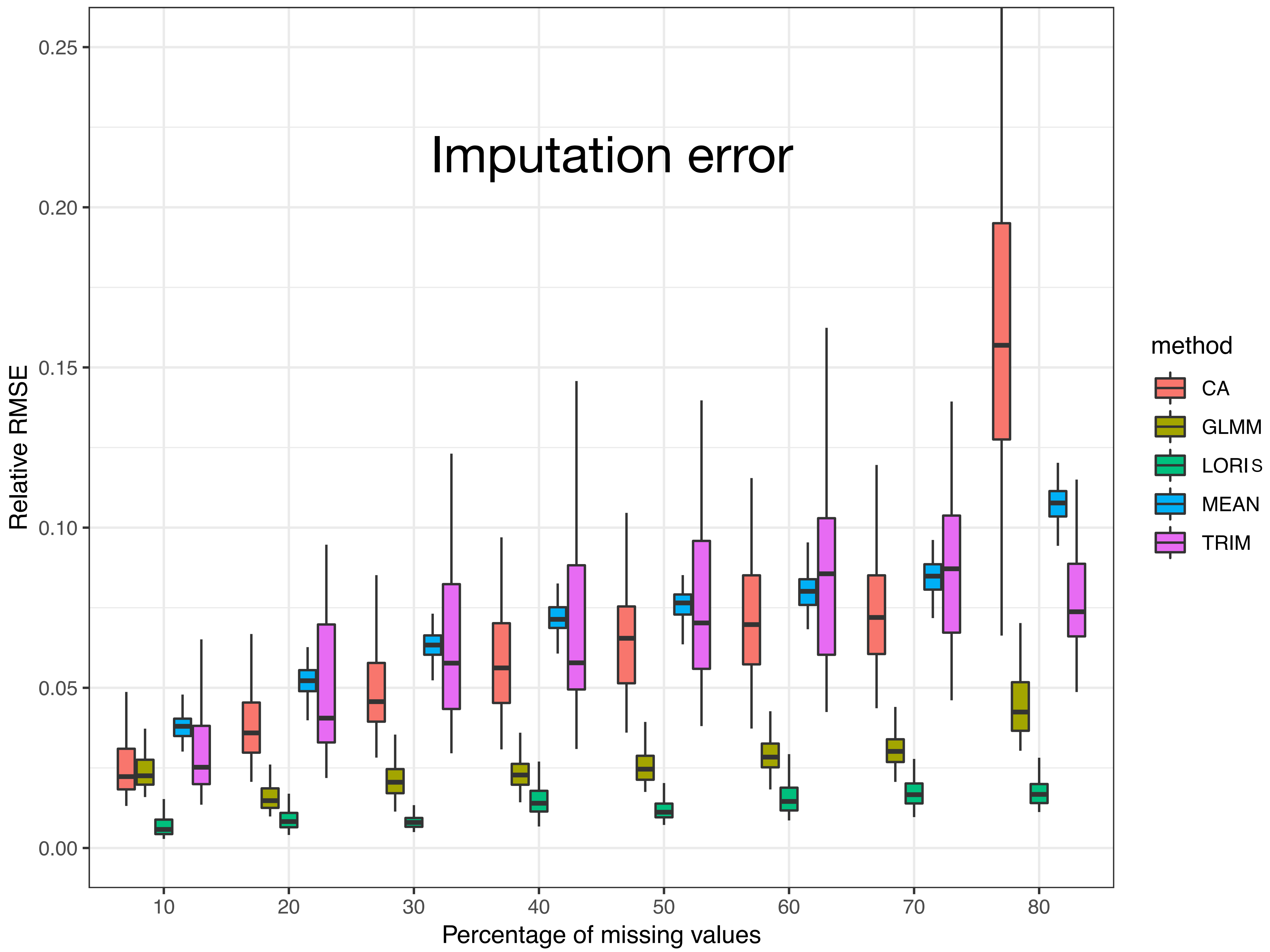
Theorem 2:

The MCGD method converges to an ϵ- solution in $\mathcal{O}(1/\epsilon)$ iterations









Fast in large dimensions

Estimation of main effects constant with dimensions

Robust to large proportions of missing values


Poster #87
210 & 230 AB
5-7pm