Byzantine-Resilient Federated Alternating Gradient Descent and Minimization for Partly-Decoupled Low Rank Matrix Learning

Problem Setting

Learn a low rank $r \ll n, q$ matrix $\boldsymbol{\Theta}^* \in \Re^{n\times q}$ from measurements of the form

$$\boldsymbol{y}_k := \boldsymbol{X}_k \boldsymbol{\theta}_k^*, k \in [q].$$

In a federated setting, while being resilient to **Byzantine Attacks**.

Factor $\boldsymbol{\Theta} = \boldsymbol{U}\boldsymbol{B}$. Solving this problem requires solving (AltGDmin [Nayer and Vaswani 2022] and FedRep [Collins et al. 2021]),

$$\min_{\substack{\tilde{\boldsymbol{U}}\in\Re^{n\times r}\\ \tilde{\boldsymbol{B}}\in\Re^{r\times q}}} f(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{B}}) = \min_{\substack{\tilde{\boldsymbol{U}}\in\Re^{n\times r}\\ \tilde{\boldsymbol{B}}\in\Re^{r\times q}}} \sum_{k=1}^{q} \|\boldsymbol{y}_k - \boldsymbol{X}_k \tilde{\boldsymbol{U}} \tilde{\boldsymbol{b}}_k\|^2$$

Theorem 1: Byz-Fed-AltGDmin-Learn

Bounded heterogeneity:
$\max_{\ell,\ell'\in[L]} \|\boldsymbol{B}_\ell^* - \boldsymbol{B}_{\ell'}^*\|_F^2 \le G_B^2 \sigma_{\max}^{*2}$

Assume RSV incoherence, Bounded heterogeneity Assumption holds, and $\frac{L_{byz}}{L} < 0.4$. *If*

$$n\tilde{q}p \ge C\tilde{\kappa}^{10}\mu^2\tilde{q}r^2 \log\tilde{q}\log\left(\frac{1}{\epsilon}\right)$$

then, w.h.p. after $T = C\tilde{\kappa}^2\log\left(\frac{1}{\epsilon}\right)$ *iterations,*

$$\boldsymbol{SD}_F(\boldsymbol{U}^*, \boldsymbol{U}_T) \le \max(\epsilon, 14C\tilde{\kappa}^2 G_B)$$

Experiments

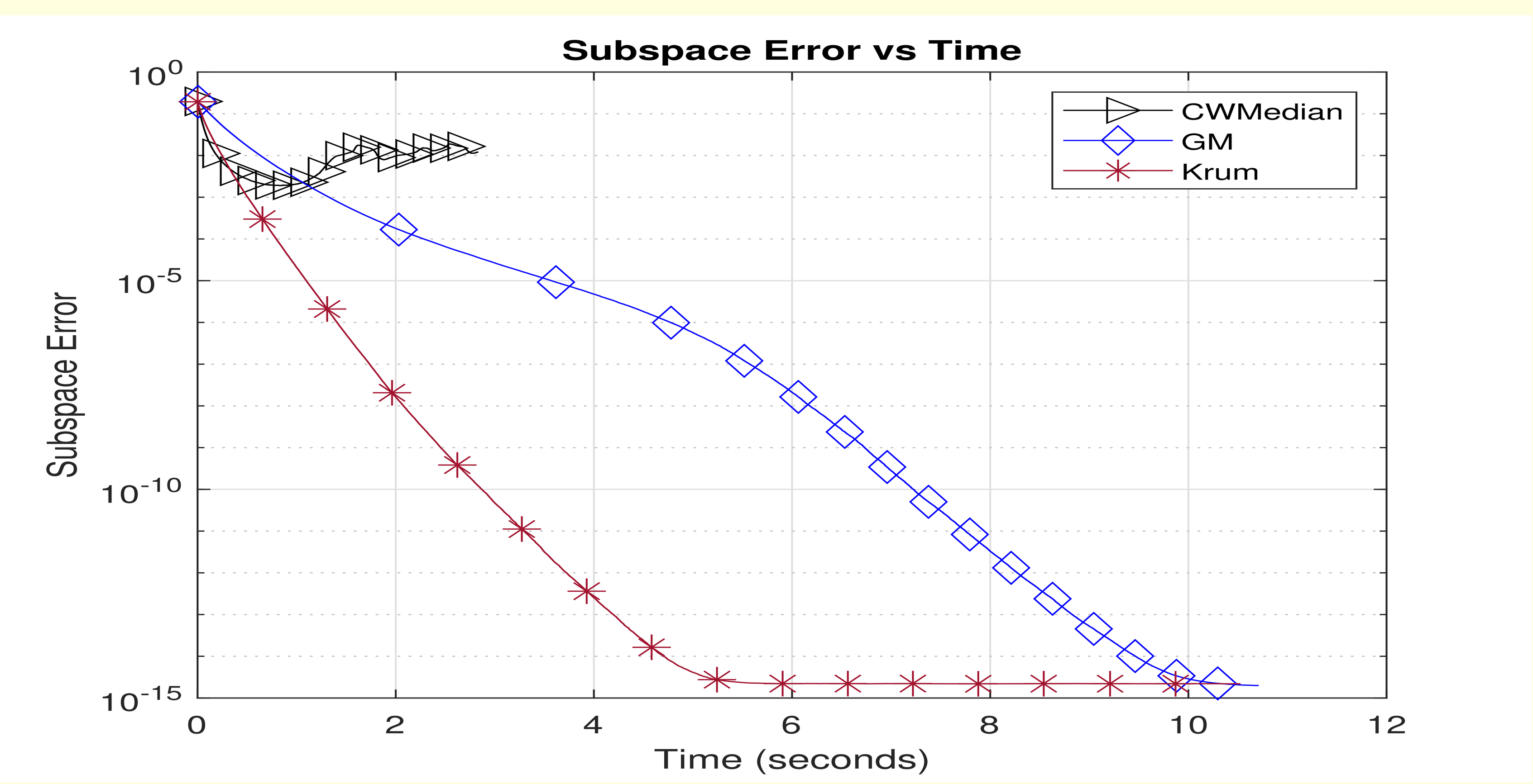

Figure 1. LRMC experiment.

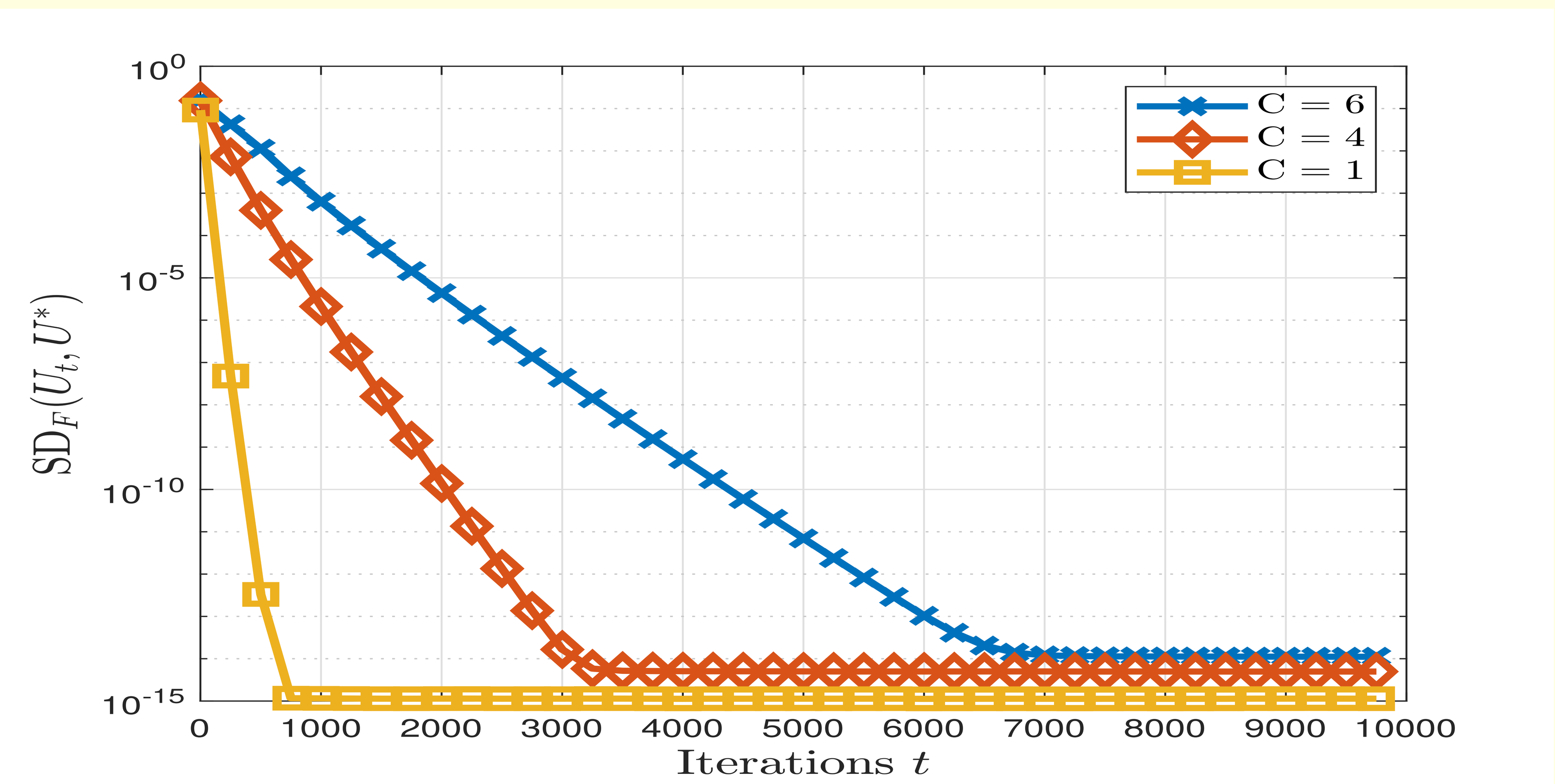

Figure 2. **Heterogeneity Effect.**

Applications

The way matrix $\boldsymbol{X}_k$ is defined it gives variety of differently problems:

- **LoRA** technique for **large language models** PEFT, $\boldsymbol{X}_k$ being identity matrices.
- For **collaborative filtering in recommendation systems** a.k.a (LRMC) problem it is a diagonal 1-0 matrix.
- For **linear multi task learning** a.k.s LRCS problem it is a random Gaussian matrix.
- For **nonlinear multi task representation** a.k.s LRPR problem it is a random Gaussian matrix, but only the magnitudes of the measurements are observed i.e., $\boldsymbol{z}_k = |\boldsymbol{y}_k|$.
- **Initialization:** All these problems require initialization, which reduces to a subspace estimation or PCA problem.

Byz-AltGDmin-LRMC algorithm

AltGDmin Initialization:
Nodes $\ell = 1, ..., L$
Calculate and Push $\boldsymbol{U}_{0\ell}$ to center
Central Server
Define set $\mathcal{I}_0 = \{\}$
Key Idea 1: Subspace Median with filtering
for $\ell = 1$ to L **do**
 if $\|\boldsymbol{u}_{0\ell}^j\| \le 1.5\mu\sqrt{\frac{r}{n}}$ for all $j \in [n]$ **then**
  Add ℓ to set $\mathcal{I}_0$
end for
$\boldsymbol{U}_0 \longleftarrow \mathrm{Byz-SubspaceEstimation}\{\boldsymbol{U}_{0\ell}\}_{\ell\in\mathcal{I}_0}$
Push $\boldsymbol{U}_0$ to nodes.
AltGDmin Iterations:
for $t = 1$ to T **do**
 Nodes $\ell = 1, ..., L$
 Calculate and Push ∇_ℓ to center
 Central Server
 Define set $\mathcal{I}_t = \{\}$
 Key Idea 2: Filtering with robust aggregation
 for $\ell = 1$ to L **do**
  Compute $\boldsymbol{U}_{temp} \leftarrow \boldsymbol{U}_{t-1} - \eta\nabla_\ell$
  if $\|\boldsymbol{u}_{temp}^j\| \le (1 - \frac{0.4}{\tilde{\kappa}^2})\|\boldsymbol{u}_{t-1}^j\| + 1.4\mu\sqrt{\frac{r}{n}}$ for all $j \in [n]$ **then**
   Add ℓ to set $\mathcal{I}_t$
 end for
 $\nabla_{Kr/GM} = \mathrm{Krum/GM}\{\nabla_\ell\}_{\ell\in\mathcal{I}_t}$
 Compute $\boldsymbol{U}_t \leftarrow QR(\boldsymbol{U}_{t-1} - \eta\nabla_{Kr/GM})$
 Push $\boldsymbol{U}_t$ to nodes.
end for
Output $\boldsymbol{U}_T$.

sankit@iastate.edu