Byzantine-resilient Federated Low Rank Column-wise Compressive Sensing

Ankit Pratap Singh and Namrata Vaswani

Iowa State University

Problem Setting

Recover an $n \times q$ rank-r matrix $\mathbf{X}^* = [\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_q^*]$, with $r \ll \min(q, n)$, from

$$\mathbf{y}_k := \mathbf{A}_k \mathbf{x}_k^*, \;\; k \in [q]$$

- $\mathbf{y}_k$ is an m-length vector with $m < n$ that is **given** (Undersampled measurements)
- Measurement matrices $\mathbf{A}_k$'s are $m \times n$ that is **given**
- The matrices $\mathbf{A}_k$s are independent and identically distributed (i.i.d.) over k
- We assume that each $\mathbf{A}_k$ is a "random Gaussian" matrix, i.e., entry of it is i.i.d. standard Gaussian.

Federated Pipeline

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_q] = [\mathbf{A}_1\mathbf{x}_1^*, \mathbf{A}_2\mathbf{x}_2^*, ..., \mathbf{A}_q\mathbf{x}_q^*].$$

We assume that there are a total of L nodes and each node measures/observes/sketches a disjoint subset of $\widetilde{m}$ rows of $\mathbf{Y}$, thus $m = L\widetilde{m}$.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \\ \vdots \\ \mathbf{Y}^{(\ell)} \\ \vdots \\ \mathbf{Y}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1^{(1)}\mathbf{x}_1^* & \mathbf{A}_2^{(1)}\mathbf{x}_2^* & \ldots & \mathbf{A}_q^{(1)}\mathbf{x}_q^* \\ \mathbf{A}_1^{(2)}\mathbf{x}_1^* & \mathbf{A}_2^{(2)}\mathbf{x}_2^* & \ldots & \mathbf{A}_q^{(2)}\mathbf{x}_q^* \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_1^{(\ell)}\mathbf{x}_1^* & \mathbf{A}_2^{(\ell)}\mathbf{x}_2^* & \ldots & \mathbf{A}_q^{(\ell)}\mathbf{x}_q^* \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_1^{(L)}\mathbf{x}_1^* & \mathbf{A}_2^{(L)}\mathbf{x}_2^* & \ldots & \mathbf{A}_q^{(L)}\mathbf{x}_q^*. \end{bmatrix}$$

We conduct Byzantine attacks while recovering $\mathbf{X}^*$ in a federated setting.

Guarantee: provably resilient to Byzantine attacks and ϵ-accurate recovery possible w.h.p. if

- initialization step of our algorithm is attack-free (explained later)
- right sing vec's of $\mathbf{X}^*$ incoherent;
- sample comp: $mq \gtrsim (n+q)r^2 \log(1/\epsilon)L$;
- $\frac{L_{byz}}{L} < 0.3$

(ϵ: final desired error)

Byzantine Attacks

Byzantine attack is a model update poisoning attack where Byzantine nodes can send arbitrary values. They have white-box access to the model or non-Byzantine node updates.

$$\nabla_\ell^{Broadcast} = \begin{cases} \nabla_\ell, & \text{If node } \ell \text{ is not attacked} \\ *, & \text{If node } \ell \text{ is attacked} \end{cases}$$

Here $*$ can be any function of $\nabla_\ell; \ell \in [L]$, $\mathbf{U}$, $\mathbf{y}_k^{(\ell)}, \mathbf{A}_k^{(\ell)}; k \in [q]$

Non-asymptotic Result [Chen, Yudong & Xu, ACM, 2017][1]

Uses the geometric median (GM) of means to replace the regular mean/sum of the partial gradients from each node. Under standard assumptions (strong convexity, Lipschitz gradients, sub-exponential-ity of sample gradients, and an upper bound on the fraction of Byzantine nodes), it provided an exponentially decaying bound.

[1] Chen, Yudong & Xu, Distributed statistical machine learning in adversarial settings: Byzantine gradient descent

Federated and Byzantine resilient design: altGDmin with geometric median (GM)

We can make the altGDmin iterations Byzantine resilient by replacing the sum over all nodes' gradients in the gradient computation step by geometric median (GM).

Basic altGDmin [Nayer & Vaswani, IEEE Trans. Info. Theory, 2023][2]

- Use sample splitting: new indep set of samples for each update
- Factorize $\mathbf{X} = \mathbf{U}\mathbf{B}$, initialize $\mathbf{U}$ by spectral initialization,
- alternate b/w minimization over $\mathbf{B}$ and (projected) GD for $\mathbf{U}$
- projected GD for $\mathbf{U}$

$$\mathbf{U}^{+} \leftarrow \mathrm{QR}(\mathbf{U} - \eta \nabla_U f(\mathbf{U}, \mathbf{B}))$$

[2] Nayer and Vaswani, Fast and Sample-Efficient Federated Low Rank Matrix Recovery from column-wise Linear and Quadratic Projections

Geometric Median [Minsker, Bernoulli, 2015][3]

Theorem

Let $\mathcal{A} = \{\mathbf{z}_1, ..., \mathbf{z}_L\}$ with each $\mathbf{z}_\ell \subseteq \Re^n$, and let $\mathbf{z}_{gm}$ denote exact Geometric Median. For a $\tau < 0.4$, suppose that, for at least $(1-\tau)L$ $\mathbf{z}_\ell$'s,

$$\Pr\{\|\mathbf{z}_\ell - \tilde{\mathbf{z}}\| \leq \epsilon\|\tilde{\mathbf{z}}\|\} \geq 1 - p$$

Then, w.p. at least $1 - \exp(-L\psi(0.4 - \tau, p))$,

$$\|\mathbf{z}_{gm} - \tilde{\mathbf{z}}\| \leq 6\epsilon\|\tilde{\mathbf{z}}\|$$

where $\psi(a, b) = (1 - a)\log\frac{1-a}{1-b} + a\log\frac{a}{b}$.

[3] Minsker, Geometric median and robust estimation in Banach spaces

Proof Ideas

Recall that $\mathbf{U}^+ = QR(\mathbf{U} - (\eta/\widetilde{m})\nabla f^{GM})$

- Obtain the expression for Subspace Distance error bound between $\mathbf{U}^*$, $\mathbf{U}^+$

$$\mathbf{SD}_F(\mathbf{U}^*, \mathbf{U}^+) \leq \frac{\|\mathbf{I}_r - \eta \mathbf{B}_{\ell_1} \mathbf{B}_{\ell_1}^\top\| \mathbf{SD}_F(\mathbf{U}^*, \mathbf{U}) + \frac{\eta}{\widetilde{m}} \|\mathrm{Err}\|_F}{1 - \frac{\eta}{\widetilde{m}} \|\mathbb{E}[\nabla f_{\ell_1}(\mathbf{U}, \mathbf{B})]\| - \frac{\eta}{\widetilde{m}} \|\mathrm{Err}\|}$$

- Bound

$$\mathrm{Err} = \nabla f^{GM} - \mathbb{E}[\nabla f_{\ell_1}(\mathbf{U}, \mathbf{B})].$$

Lemmas for bounding Err

Consider any $\ell \in \mathcal{J}$ where $\mathcal{J}$ is the set of non-byzantine/good nodes.

1. w.p. at least, $1 - \exp(\log q + r - c\epsilon_1^2 \widetilde{m})$

$$\|\mathbf{B}_\ell - \mathbf{G}\|_F \leq 1.7\epsilon_1 \delta_t \sigma^*_{\max}$$

2. w.p. at least, $1 - \exp(\log q + r - c\epsilon_1^2 \widetilde{m})$

$$\sigma_{\max}(\mathbf{B}_\ell) \leq 1.1\sigma^*_{\max}$$

3. w.p. at least $1 - \exp\left((n + r) - c\epsilon_1^2 \frac{\widetilde{m}q}{r\mu^2}\right) - 2\exp(\log q + r - c\epsilon_1^2 \widetilde{m})$

$$\|\nabla f_\ell - \mathbb{E}[\nabla f_\ell]\|_F \leq 1.5\epsilon_1 \sqrt{r} \delta_t \widetilde{m} \sigma^{*2}_{\max}$$

Bounding Err

- Recall to use Geometric Median theorem we need to bound $\nabla f_\ell - \mathbb{E}[\nabla f_{\ell_1}]$

$$
\begin{aligned}
&\|\nabla f_\ell - \mathbb{E}[\nabla f_{\ell_1}]\|_F \leq \\
&\|\nabla f_\ell - \mathbb{E}[\nabla f_\ell]\|_F + \|\mathbb{E}[\nabla f_\ell] - \mathbb{E}[\nabla f_{\ell_1}]\|_F \leq \\
&1.5\epsilon_1\sqrt{r}\delta_t \widetilde{m}\sigma_{\max}^{*2} \\
&+ \left\|\widetilde{m}(\mathbf{X}_\ell - \mathbf{X}^*)\mathbf{B}_\ell^\top - \widetilde{m}(\mathbf{X}_{\ell_1} - \mathbf{X}^*)\mathbf{B}_{\ell_1}^\top\right\|_F
\end{aligned}
$$

- Note $\mathbb{E}[\nabla f_{\ell_i}] \neq \mathbb{E}[\nabla f_{\ell_j}]$

$$
\begin{aligned}
&\left\|\widetilde{m}(\mathbf{X}_\ell - \mathbf{X}^*)\mathbf{B}_\ell^\top - \widetilde{m}(\mathbf{X}_{\ell_1} - \mathbf{X}^*)\mathbf{B}_{\ell_1}^\top\right\|_F \\
&\leq \widetilde{m}3.2\sigma_{\max}^* \|\mathbf{B}_\ell - \mathbf{B}_{\ell_1} \pm \mathbf{G}\|_F \\
&\leq \widetilde{m}3.2\sigma_{\max}^*(\|\mathbf{B}_\ell - \mathbf{G}\|_F + \|\mathbf{B}_{\ell_1} - \mathbf{G}\|_F) \leq \widetilde{m}11{\sigma_{\max}^*}^2\epsilon_1\sqrt{r}\delta_t
\end{aligned}
$$

AltGDmin Error Decay

Theorem

Assume incoherence of right singular vectors. If, at each iteration t, $\widetilde{m}q \geq C_1\kappa^2\mu^2(n+q)r^2$, $\widetilde{m} > C_2 \max(\log q, \log n)$; if $\frac{L_{byz}}{L} < 0.3$; and if the initial estimate $\mathbf{U}_0$ satisfies $\mathbf{SD}(\mathbf{U}^*, \mathbf{U}_0) \leq \delta_0 = 0.1/\kappa^2$, then w.p. at least $1 - tn^{-4(L-L_{byz})}$,

$$\mathbf{SD}(\mathbf{U}^*, \mathbf{U}_{t+1}) \leq \delta_{t+1} := \left(1 - (\eta{\sigma^*_{\max}}^{*2})\frac{0.31}{\kappa^2}\right)^{t+1} \delta_0$$

and $\|\mathbf{x}_k^* - (\mathbf{x}_k)_{t+1}\| \leq \delta_{t+1}\|\mathbf{x}_k^*\|$ for all $k \in [q]$.

Federated design: Initialization

altGDmin: **Initialize U** using a truncated spectral initialization by computing the top r singular vectors of the following matrix

$$\mathbf{X}_{init} = \sum_k \mathbf{A}_k^\top (\mathbf{y}_k \circ \mathbf{1}_{|\mathbf{y}_k| \leq \sqrt{\alpha}})$$

Federated Power method

Recall that basic PM runs the following iteration: $\mathbf{U} \leftarrow orth(\mathbf{X}_0 \mathbf{X}_0^\top \mathbf{U}_{\tau-1})$ $\tau \in [T_{PM}]$.

In our federated setting, $\tilde{\mathbf{U}} = \mathbf{X}_0 \mathbf{X}_0^\top \mathbf{U}_{\tau-1}$ is computed as

- $\mathbf{V} = \sum_\ell \mathbf{X}_0^{\ell\top} \mathbf{U}_{\tau-1}$
- $\tilde{\mathbf{U}} = \sum_\ell \mathbf{X}_0^\ell \mathbf{V}$

PM is initialized with a random matrix $\mathbf{U}_{rand} \equiv \mathbf{U}_{\tau=0}$ with i.i.d. standard Gaussian entries.

Why our current result needs to assume no attacks during initialization.

Obvious Solution modify power method.

The initialization for the power method is a random Gaussian matrix, $\mathbf{U}_{rand}$. The cosine of the smallest principal angle between a random r-dimensional subspace in $\Re^n$ and a given one is order $1/\sqrt{nr}$ [Rudelson & Vershynin, Communications on Pure and Applied Mathematics, 2009][4]

[4]Rudelson & Vershynin, Smallest singular value of a random rectangular matrix

Improving our result: dealing with attacks in initialization

In ongoing work, we are working on designing and analyzing a Byzantine-resilient subspace estimation algorithm. This can potentially be used to handle Byzantine nodes in the initialization step. Paper on arxiv Byzantine-Resilient Federated PCA and Low Rank Matrix Recovery https://arxiv.org/abs/2309.14512