Byzantine Machine Learning: MultiKrum and an optimal notion of robustness

G. Bareilles*[1], W. Bouaziz*[2], J. Fageot*[1, 3], E.M. El Mhamdi[1]

[1]CMAP, École Polytechnique, [2]Mistral AI, [3]Télécom Paris

TLDR We introduce the *optimal robustness coefficient* $\kappa^\star$, as **quantitative measure** of an aggregation rule's robustness, and use it to prove that **MultiKrum is robust**.

Problem of interest

$$\min_{x\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n \mathrm{g}_i(x)$$

in Fed. Learn., each g_i is on a device

Typical algorithm:

$$v_i \leftarrow \begin{cases} \nabla \mathrm{g}_i(x^k) & \text{if } i \in \mathrm{Hon} \\ v_i^\dagger & \text{if } i \in \mathrm{Adv} \end{cases}$$

$$x^{k+1} = x^k - \eta \underbrace{\tfrac{1}{n}\textstyle\sum_{i=1}^n v_i}_{\mathrm{Agg}}$$

Byzantine setup: there are $|\mathrm{Adv}| = f$ adversaries

Adversaries's goal: manipulate $\left(x^k\right)_k$

e.g. system breakdown, plant backdoor

Defender's goal:

- ideally: $x^k \to \mathrm{argmin} \sum_{i\in Hon} \mathrm{g}_i$
- less demanding:

$$\sum_{i\in \mathrm{Hon}} \nabla \mathrm{g}_i\left(x^k\right) = 0$$

The **aggregation rule** is crucial

$$\mathrm{Agg}\ (v_1, ..., v_n) = \frac{1}{n}\sum_{i=1}^n v_i$$

e.g. mean breaks-down at $f \geq 1$.

Krum and MultiKrum

Consider the score function $s^{\mathcal{V}}(v) = \sum_{i\in\ \mathcal{N}(v)} \| v - v_i \|^2$, with $\mathcal{N}(v)$ the $n-f$ points of $\mathcal{V} = (v_1, ..., v_n)$ closest to v.

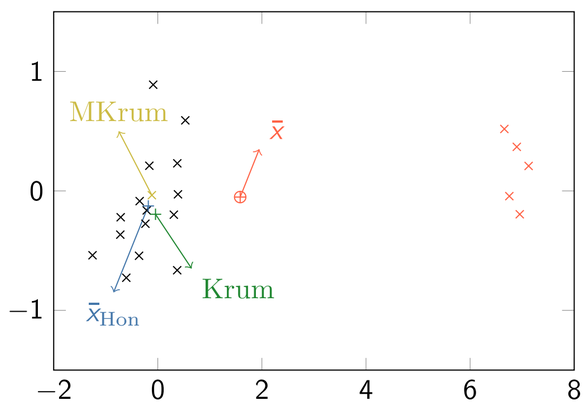
Then,

$$\mathrm{Krum}(\mathcal{V}) = \operatorname*{argmin}_{v\in\{v_1,...,v_n\}} s^{\mathcal{V}}(v)$$

$$\mathrm{MKrum}_m(\mathcal{V}) = \frac{1}{m}\sum_{i\in S_m^\star(\mathcal{V})} v_i$$

where $S_m^\star$ collects the m points of $\mathcal{V}$ with smallest score $s^{\mathcal{V}}$.

e.g. $s^{\mathcal{V}}(v_i) \leq s^{\mathcal{V}}(v_j)$ for all $i \in S_m^\star(\mathcal{V}), j \in \overline{S_m^\star(\mathcal{V})}$



×: honest points

×: adversarial points

Prior notions of robustness

▷ **Break-down point**: min. number f of adv. that can manipulate $\mathrm{Agg}(\mathcal{V})$ *arbitrarily*. Fact: the largest breakdown point is ${}^{f}/_{n} > {}^{1}/_{2}$ [3].

Above, adversaries have the majority, hopeless.

▷ **(f, κ)-robustness** [1]: for any $\mathcal{V} = (v_1, ..., v_n)$, $S \in \mathcal{P}_{n-f}^n$,

$$\| \mathrm{Agg}(\mathcal{V}) - \overline{v}_S \|^2 \leq \kappa \Sigma_S(\mathcal{V})$$

with $\overline{v}_S = \frac{1}{|S|}\sum_{i\in S} v_i$, $\Sigma_S(\mathcal{V}) = \frac{1}{|S|}\sum_{i\in S} \| v_i - \overline{v}_S \|^2$.

Fact: Gradient Descent with (f, κ)-robust Agg with smooth g_i and good step size converges near **honest critical points**:

$$\| \nabla \mathrm{g}_{\mathrm{Hon}}(\hat{x}^T) \|^2 \leq \mathcal{O}\left(\frac{1}{T}\right) + 4\kappa G^2$$

with $\mathrm{g}_{\mathrm{Hon}} = \frac{1}{n-f}\sum_{i\in\ \mathrm{Hon}} \mathrm{g}_i$, G^2 bound on variance of honest gradients.

→ κ captures the neighborhood, but may be loose.

The robustness coefficient

Definition. The *robustness coefficient* $\kappa^\star$ of $\mathrm{Agg} : \left(\mathbb{R}^d\right)^n \to \mathbb{R}^d$ is

$$\kappa^\star(\mathrm{Agg}) = \sup_{\substack{\mathcal{V}=(v_1,...,v_n)\in(\mathbb{R}^d)^n \\ S\in\mathcal{P}_{n-f}^n}} \frac{\| \mathrm{Agg}(\mathcal{V}) - \overline{v}_S \|^2}{\Sigma_S(\mathcal{V})}.$$

Agg is *robust* when $\kappa^\star(\mathrm{Agg}) < \infty$. Convention: $0/0 = -\infty$

▷ Thus for $S = \mathrm{Hon}$ and any $\mathcal{V}$

$$\| \mathrm{Agg}(\mathcal{V}) - \overline{v}_{\mathrm{Hon}} \|^2 \leq \kappa^\star \Sigma_{\mathrm{Hon}}(\mathcal{V})$$

→ optimal control on deviation relative to honest mean, scaled by honest variance.

▷ Computing $\kappa^\star$ is **tough** (symmetries, variance scaling, bad properties of Agg).

We seek

- **lower bounds** evaluate at specific $\mathcal{V}$, S
- **upper bounds** requires global arguments.

▷ [2] implies the following direct bounds:

Aggregation	GeomMedian	Krum	MKrum
Upper bound	$4\left(\frac{1-f/n}{1-2f/n}\right)^2$	$6\frac{1-f/n}{1-2f/n}$	$\emptyset$
Lower bound	$\frac{f/n}{1-2f/n}$	$\frac{f/n}{1-2f/n}$	$\frac{f/n}{1-2f/n}$

MultiKrum's robustness

Theorem (Upper bound). Assume ${}^{f}/_{n} < {}^{1}/_{2}$ and $0 < m \leq n - f$. Then,

$$\kappa_m^\star \leq \frac{1-f/n}{1-2f/n}\min\left(\sqrt{2}+1, \frac{\sqrt{n-2f}}{\sqrt{m}} + \frac{\sqrt{2f}}{\sqrt{m}} + \frac{f}{m}\right)^2$$

→ $\kappa_m^\star$ has a finite upper bound, hence MKrum is robust.

Lemma (Lower bound, $m = 1$).

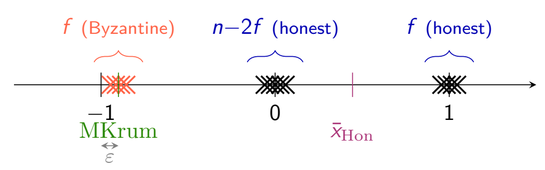
Assume ${}^{f}/_{n} < {}^{1}/_{2}$. Then,

$$\kappa_1^\star \geq \begin{cases} \frac{1-2/n}{1-2f/n+2/n} & \text{if } n \text{ even} \\ \frac{1-1/n}{1-2f/n+1/n} & \text{if } n \text{ odd} \end{cases}$$

Lemma (Lower bound, $m = n - f$).

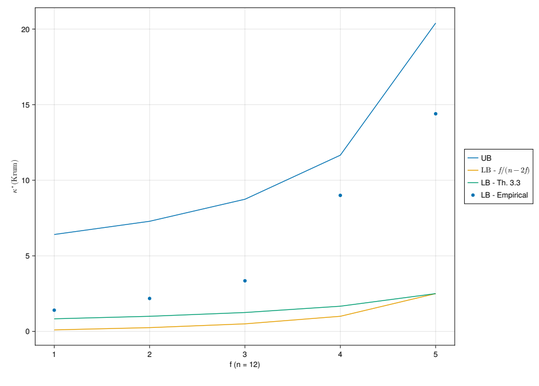
Assume ${}^{f}/_{n} < {}^{1}/_{3}$. Then,

$$\kappa_{n-f}^\star \geq 4\frac{f/n}{1-2f/n}.$$

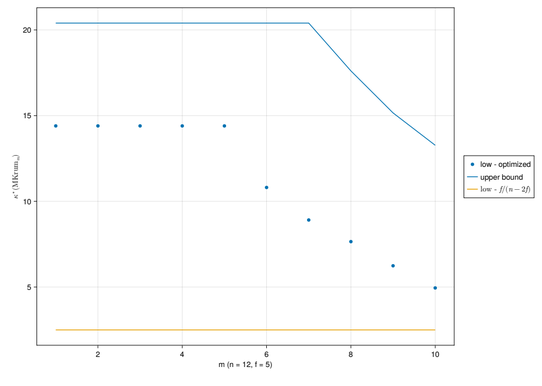


Numerical estimation of $\kappa^\star$

How tight are the lower bounds? Dots: optimized lower bounds on $\kappa^\star$ (valid but not proved optimal).



Krum, $n = 12$



MultiKrum, $n = 12$, $f = 5$

→ Empirical optima $\kappa^\star$ sit between our LB and UB; room for tighter bounds.

References. [1] Blanchard, El Mhamdi, et al. Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent, NeurIPS (2017). [2] Allouah, Farhadkhani et al. Fixing by Mixing: A Recipe for Optimal Byzantine ML under Heterogeneity, ICML (2023). [3] P. Rousseeuw. Multivariate estimation with high breakdown point, Math. Stats. and Apps. (1985).