Byzantine Machine Learning: MultiKrum and an optimal notion of robustness

Gilles Bareilles

CMAP École Polytechnique

Brief Announcement – PODC, Egham
8 July 2026

Joint work with

Wassim (Wes) Bouaziz
Mistral AI, work done at CMAP

Julien Fageot
Télécom Paris

El Mahdi El Mhamdi
CMAP, École Polytechnique

Byzantine Machine Learning

- Task:

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

- Specificities
  - *distributed*: i-th device holds g_i
  - *adversarial*: devices in $\mathrm{Adv} \subset [n]$ are not trustworthy; $|\mathrm{Adv}| = f$
- D-GD, iteration k

$$\begin{cases} v_i \leftarrow \begin{cases} \nabla g_i(x^k) & \text{if } i \in \mathrm{Hon} \\ v_i^\dagger & \text{if } i \in \mathrm{Adv} \end{cases} \\ x^{k+1} = x^k - \eta \mathrm{Agg}(v_1, \ldots, v_n) \end{cases}$$

- Attacker's goal: manipulate $(x^k)_k$
- Defender's goal: find $\bar{x} \in \arg\min \sum_{i \in \mathrm{Hon}} g_i$
- Everything hinges on the aggregation rule

Examples

- Distributed / Federated Learning
- Beyond gradients, data poisoning

Plan

1. Characterize robustness?
2. Is MultiKrum robust?

Prior notions of robustness (non-exhaustive)

▷ **Break-down point**: how many adversaries are enough to manipulate $\mathrm{Agg}(\mathcal{V})$ *arbitrarily*?

Largest breakdown point we can hope is $f/n > 1/2$. Above, adversaries have the majority, hopeless ◇ Rousseeuw '85

Prior notions of robustness (non-exhaustive)

▷ **Break-down point**: how many adversaries are enough to manipulate $\mathrm{Agg}(\mathcal{V})$ *arbitrarily*?

Largest breakdown point we can hope is $f/n > 1/2$. Above, adversaries have the majority, hopeless ◇ Rousseeuw '85

▷ (f, κ)-**robustness**: for any $\mathcal{V} = (v_1, \ldots, v_n)$, $S \in \mathcal{P}^n_{n-f}$ ◇ Fixing by Mixing, AFGGGS, '23

$$\|\mathrm{Agg}(\mathcal{V}) - \bar{v}_S\|^2 \leq \kappa \Sigma_S(\mathcal{V}) \qquad \text{with} \quad \begin{cases} \bar{v}_S & = \frac{1}{|S|} \sum_{i \in S} v_i \\ \Sigma_S(\mathcal{V}) & = \frac{1}{|S|} \sum_{i \in S} \|v_i - \bar{v}_S\|^2 \end{cases}$$

For any subgroup, bound the distance of aggregate to subgroup mean, up to subgroup variance.

Prior notions of robustness (non-exhaustive)

▷ **Break-down point**: how many adversaries are enough to manipulate $\mathrm{Agg}(\mathcal{V})$ *arbitrarily*?

Largest breakdown point we can hope is $f/n > 1/2$. Above, adversaries have the majority, hopeless ◇ Rousseeuw '85

▷ (f, κ)-**robustness**: for any $\mathcal{V} = (v_1, \ldots, v_n)$, $S \in \mathcal{P}^n_{n-f}$ ◇ Fixing by Mixing, AFGGGS, '23

$$\|\mathrm{Agg}(\mathcal{V}) - \bar{v}_S\|^2 \leq \kappa \Sigma_S(\mathcal{V}) \qquad \text{with} \quad \begin{cases} \bar{v}_S & = \frac{1}{|S|} \sum_{i \in S} v_i \\ \Sigma_S(\mathcal{V}) & = \frac{1}{|S|} \sum_{i \in S} \|v_i - \bar{v}_S\|^2 \end{cases}$$

For any subgroup, bound the distance of aggregate to subgroup mean, up to subgroup variance.

▷ Good notion:

- guarantees **convergence** of D-GD $\quad \|\nabla g_{\mathrm{Hon}}(\hat{x}^T)\|^2 \leq \mathcal{O}\left(\frac{1}{T}\right) + 4\kappa G^2$
- generalizes previous robustness notions

But, κ may be loose, thus **preventing quantitative comparison** of aggregation rules

The robustness coefficient

The **robustness coefficient** $\kappa^\star$ of an aggregation rule $\mathrm{Agg} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ is

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

Agg is (n,f)-**robust** when $\kappa^\star_{n,f}(\mathrm{Agg}) < \infty$. convention: $0/0 = -\infty$

$\to$ $\kappa^\star_{n,f}$ captures the **worst-case deviation** between aggregate and honest mean, scaled by honest variance.

▹ Computing $\kappa^\star$ is **tough** (symmetries, variance scaling, bad properties of Agg)
$\to$ we seek matching **lower** and **upper** bounds

Krum & MultiKrum ◇ Blanchard, El Mhamdi, Guerraoui, Steiner '17

▹ Basic notions

- $\mathcal{N}(v)$: $n-f$ points of $\mathcal{V}$ closest to v
- $s^{\mathcal{V}}(v) = \sum_{i \in \mathcal{N}(v)} \|v - v_i\|^2$
- ordering permutation: $s^{\mathcal{V}}(v_{(1)}) \leq \cdots \leq s^{\mathcal{V}}(v_{(n)})$

▹ $\mathrm{Krum}^{n,f}(\mathcal{V}) = v_{(1)}$

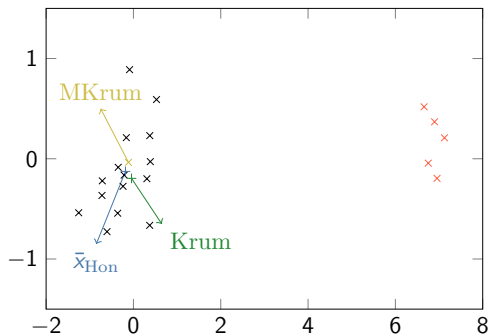
✓ When $f/n < 1/2$, cannot be arbitrarily manipulated

✗ High variance discards all-but-one

▹ $\mathrm{MKrum}_m^{n,f}(\mathcal{V}) = \frac{1}{m}\sum_{i=1}^{m} v_{(i)}$

✓ Reduced variance, preferred in practice

✗ No theoretical guarantees



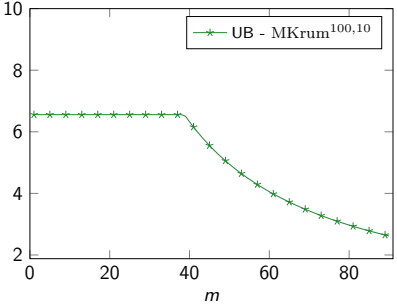
MultiKrum is robust

Theorem (BBFEM, '26)

Assume that $f/n < 1/2$ *and* $0 < m \leq n - f$. *Then,*

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

finite upper bound $\Rightarrow \mathrm{MKrum}^{n,f}_{m}$ is robust



Take-home messages

- the **robustness coefficient** $\kappa^\star$ captures precisely the worst-case scenario
- the popular **MultiKrum** is robust

Perspectives

- compute or bound $\kappa^\star_{n,f}$ of usual aggregation rules, and **compare them quantitatively**

G. Bareilles*, W. Bouaziz*, J. Fageot*, E.M. El Mhamdi: *Byzantine Machine Learning: MultiKrum and an optimal notion of robustness*, 2026.

Thank you!