Centered Clipping with Per-client Thresholds given Side Information

1 Byzantine Impact Bound via Side Information

We give a bound on the *aggregate Byzantine impact* that holds under omniscient attackers. The bound depends only on quantities that are either (i) directly controlled by side information and optimisation tolerance, or (ii) intrinsic to the honest cohort.

Setup and Notation. Let $\mathcal{G}$ and $\mathcal{B}$ be the honest and Byzantine (client) index sets, with $|\mathcal{B}| = \delta n$ and $|\mathcal{G}| = (1-\delta)n$. At the current round, the initial guess for the aggregation is $\boldsymbol{x}_0 \in \mathbb{R}^d$ and client proposals are $\boldsymbol{x}_i \in \mathbb{R}^d$. Define

$$\boldsymbol{d}_i := \frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|} \quad (\boldsymbol{x}_i \neq \boldsymbol{x}_0), \qquad \alpha_i(\boldsymbol{\nu}) := \min\Big(1, \frac{\nu_i}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|}\Big) \in [0,1].$$

The one–step clipped aggregate is

$$\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} \;=\; \boldsymbol{x}_0 + \frac{1}{n}\sum_{i=1}^{n} \alpha_i(\boldsymbol{\nu})\,(\boldsymbol{x}_i - \boldsymbol{x}_0).$$

Let $\boldsymbol{g}_{\mathcal{V}}$ be the validation gradient (side information). Assume the $\boldsymbol{\nu}$–selection is solved up to tolerance ε_ν such that

$$\big\|\boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}\big\| \;\le\; \varepsilon_\nu. \tag{1}$$

Let the honest mean be $\bar{\boldsymbol{x}} := {}^1\!/_{|\mathcal{G}|} \sum_{i\in\mathcal{G}} \boldsymbol{x}_i$ and define the validation bias w.r.t. the honest mean

$$\varepsilon_{\mathcal{V}} \;:=\; \big\|\boldsymbol{g}_{\mathcal{V}} - \bar{\boldsymbol{x}}\big\|.$$

We also define the (average) honest dispersion

$$\bar{\zeta}_h \;:=\; \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|.$$

Byzantine Aggregate. We denote the aggregate Byzantine contribution by

$$\boldsymbol{B} \;:=\; \frac{1}{n}\sum_{j\in\mathcal{B}} \alpha_j(\boldsymbol{\nu})\,(\boldsymbol{x}_j - \boldsymbol{x}_0), \qquad \text{so that} \quad \hat{\boldsymbol{x}}_{\boldsymbol{\nu}} - \boldsymbol{x}_0 = \boldsymbol{B} + \underbrace{\frac{1}{n}\sum_{i\in\mathcal{G}} \alpha_i(\boldsymbol{\nu})\,(\boldsymbol{x}_i - \boldsymbol{x}_0)}_{=:\,\boldsymbol{G}}.$$

Theorem 1.1 (Byzantine impact bound given side information)**.** *Under the setup above, for* arbitrary *Byzantine choices* $\{\boldsymbol{x}_j\}_{j\in\mathcal{B}} \subset \mathbb{R}^d$ *and the one–step* $\boldsymbol{\nu}$ *selected to satisfy* (1)*, the Byzantine aggregate obeys*

$$\|\boldsymbol{B}\| \le (2-\delta)\,\big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| \;+\; \varepsilon_\nu \;+\; (1-\delta)\,(\varepsilon_{\mathcal{V}} + \bar{\zeta}_h). \tag{2}$$

Consequently, for any $\delta \in [0,1)$*, the Byzantine impact can be made arbitrarily small by driving the controllable quantities* $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$*,* ε_ν*, and* $\varepsilon_{\mathcal{V}}$ *via curation of validation dataset and tighter optimisation for the robust aggregation.*

Remark 1.2 (Behaviour as $\delta \to 1$). When $\delta \to 1$ (i.e., $|\mathcal{G}| \to 0$), the honest term vanishes and the bound reduces to

$$\|\boldsymbol{B}\| \;\le\; \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| \;+\; \varepsilon_\nu.$$

Formally, $\bar{\boldsymbol{x}}$ and $\bar{\zeta}_h$ are defined only when $|\mathcal{G}| \ge 1$; the display should be read as the $\delta \to 1$ *limit* of (2). In words: even if the attacker set occupies (almost) all clients, the Byzantine impact remains controlled by *side-information alignment* $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$ and the *optimisation tolerance* ε_ν.

Proof. Starting from the decomposition $\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0 = (\boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}) + (\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} - \boldsymbol{x}_0) = \boldsymbol{e} + (\boldsymbol{B} + \boldsymbol{G})$, where $\boldsymbol{e} := \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}$, we have

$$\boldsymbol{B} = \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0 - \boldsymbol{e} - \boldsymbol{G}.$$

Taking norms and using (1),

$$\|\boldsymbol{B}\| \;\le\; \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| + \|\boldsymbol{e}\| + \|\boldsymbol{G}\| \;\le\; \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| + \varepsilon_\nu + \|\boldsymbol{G}\|. \tag{3}$$

Since $\alpha_i(\boldsymbol{\nu}) \in [0,1]$, we have

$$\|\boldsymbol{G}\| \;=\; \Big\|\frac{1}{n}\sum_{i\in\mathcal{G}} \alpha_i(\boldsymbol{\nu})\,(\boldsymbol{x}_i - \boldsymbol{x}_0)\Big\| \;\le\; \frac{1}{n}\sum_{i\in\mathcal{G}} \alpha_i(\boldsymbol{\nu})\,\|\boldsymbol{x}_i - \boldsymbol{x}_0\| \;\le\; \frac{1}{n}\sum_{i\in\mathcal{G}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|,$$

which substituted into (3) yields

$$\|\boldsymbol{B}\| \le \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| \;+\; \varepsilon_\nu \;+\; \frac{1}{n}\sum_{i\in\mathcal{G}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|.$$

For (2), observe that

$$\frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \;\le\; \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \big(\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\| + \|\bar{\boldsymbol{x}} - \boldsymbol{x}_0\|\big) \;=\; \bar{\zeta}_h + \|\bar{\boldsymbol{x}} - \boldsymbol{x}_0\|.$$

Moreover,

$$\|\bar{\boldsymbol{x}} - \boldsymbol{x}_0\| \;\le\; \|\bar{\boldsymbol{x}} - \boldsymbol{g}_{\mathcal{V}}\| + \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| \;=\; \varepsilon_{\mathcal{V}} + \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|.$$

Combining the two displays gives

$$\frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \;\le\; \bar{\zeta}_h + \varepsilon_{\mathcal{V}} + \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|.$$

Multiplying by $|\mathcal{G}|/n = (1-\delta)$ and substituting into (3) yields

$$\|\boldsymbol{B}\| \;\le\; \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| + \varepsilon_\nu + (1-\delta)\,\big(\bar{\zeta}_h + \varepsilon_{\mathcal{V}} + \|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|\big) \;=\; (2-\delta)\,\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| + \varepsilon_\nu + (1-\delta)(\varepsilon_{\mathcal{V}} + \bar{\zeta}_h),$$

which is (2). □