Stationarity of Clientwise Centered Clipping and Bound on Byzantine Weights

See Section 1. Worst-Case Byzantine Impact Bound via Side Information.

Proposition 0.1 (No δ-threshold required)**.** *Assume the setup of Theorem 1.1 and that $|\mathcal{H}| = (1-\delta)n \geq 1$ (i.e., at least one honest client). Then, for* any *attacker fraction $\delta \in [0,1)$, the worst–case Byzantine aggregate obeys the side–information bound*

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

which contains no denominator in δ *and thus imposes* no threshold *(e.g., no condition like $\delta < 1/2$) for validity.*

Proof. Equation (1) is exactly Theorem 1.1 (Equation (7)) restated. The right-hand side depends on δ only via linear coefficients $(2-\delta)$ and $(1-\delta)$, and does not place δ in any denominator. Hence the bound holds uniformly for all $\delta \in [0,1)$, requiring no threshold such as $\delta < \delta_0$. $\square$

Remark 0.2 (Behaviour as $\delta \to 1$). When $\delta \to 1$ (i.e., $|\mathcal{H}| \to 0$), the honest term vanishes and the bound reduces to

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

Formally, $\bar{\boldsymbol{x}}$ and $\bar{\zeta}_h$ are defined only when $|\mathcal{H}| \geq 1$; the display should be read as the $\delta \to 1$ *limit* of (1). 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.

Remark 0.3 (Operational knobs unaffected by δ). The quantities $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$ and ε_ν are *algorithmically controllable*: recentering $\boldsymbol{x}_0 \leftarrow \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}$ drives $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| \downarrow 0$, and tightening the $\boldsymbol{\nu}$-selection drives $\varepsilon_\nu \downarrow 0$. The validation bias ε_V decreases with better/larger $\mathcal{V}$. Thus, (1) yields an *arbitrarily small* Byzantine impact for fixed honest dispersion $\bar{\zeta}_h$, *independently of any threshold on δ.*

1 Worst-Case Byzantine Impact Bound via Side Information

We give a bound on the *aggregate Byzantine impact* that holds even under a fully adversarial choice of Byzantine vectors (omniscient attackers who know $\boldsymbol{x}_0$), without using any coherence or norm–separation assumptions on $\mathcal{B}$. 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 $\mathcal{H}$ and $\mathcal{B}$ be the honest and Byzantine index sets, with $|\mathcal{B}| = \delta n$ and $|\mathcal{H}| = (1-\delta)n$. At the current round, the centre 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) so that

$$\big\|\boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}\big\| \;\le\; \varepsilon_\nu. \tag{2}$$

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

$$\varepsilon_V \;:=\; \big\|\boldsymbol{g}_{\mathcal{V}} - \bar{\boldsymbol{x}}\big\|. \tag{3}$$

We also define the (average) honest dispersion

$$\bar{\zeta}_h \;:=\; \frac{1}{|\mathcal{H}|}\sum_{i\in\mathcal{H}} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|. \tag{4}$$

Byzantine aggregate. We denote the aggregate Byzantine contribution by

$$\boldsymbol{B} \;:=\; \frac{1}{n}\sum_{j\in\mathcal{B}} \alpha_j(\boldsymbol{\nu})\,(\boldsymbol{x}_j - \boldsymbol{x}_0), \qquad \text{so that} \quad \hat{\boldsymbol{x}}_{\boldsymbol{\nu}} - \boldsymbol{x}_0 = \boldsymbol{B} + \underbrace{\frac{1}{n}\sum_{i\in\mathcal{H}} \alpha_i(\boldsymbol{\nu})\,(\boldsymbol{x}_i - \boldsymbol{x}_0)}_{=:\,\boldsymbol{H}}. \tag{5}$$

Theorem 1.1 (Worst–case Byzantine impact bound without coherence)**.** *Under the setup above, for* arbitrary *Byzantine choices* $\{\boldsymbol{x}_j\}_{j\in\mathcal{B}}$ *and the one–step* $\boldsymbol{\nu}$ *selected to satisfy* (2)*, the Byzantine aggregate obeys the following bounds:*

$$\|\boldsymbol{B}\| \le \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| \;+\; \varepsilon_\nu \;+\; \frac{1}{n}\sum_{i\in\mathcal{H}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \quad \text{(exact triangle bound)}, \tag{6}$$

$$\|\boldsymbol{B}\| \le (2-\delta)\,\big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| \;+\; \varepsilon_\nu \;+\; (1-\delta)\,(\,\varepsilon_V + \bar{\zeta}_h\,) \qquad \text{(side–information bound)}. \tag{7}$$

Consequently, for fixed $(\delta, \bar{\zeta}_h)$*, the Byzantine impact can be made arbitrarily small by driving the controllable quantities* $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$*,* ε_ν*, and* ε_V *to zero (via iteration, tighter optimisation, and larger validation).*

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

$$\boldsymbol{B} = \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0 - \boldsymbol{e} - \boldsymbol{H}. \tag{8}$$

Taking norms and using (2),

$$\|\boldsymbol{B}\| \;\le\; \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| + \|\boldsymbol{e}\| + \|\boldsymbol{H}\| \;\le\; \big\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\big\| + \varepsilon_\nu + \|\boldsymbol{H}\|. \tag{9}$$

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

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

which substituted into (9) yields (6).
For (7), observe that

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

Moreover,

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

Combining the two displays gives

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

Multiplying by $|\mathcal{H}|/n = (1-\delta)$ and substituting into (9) yields

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

which is (7). $\square$

Interpretation and tunable knobs. The bound (7) is *worst–case* in that it imposes *no constraints whatsoever* on the geometry or norms of Byzantine proposals; the adversary may choose directions and magnitudes adversarially knowing $\boldsymbol{x}_0$. Yet the Byzantine impact is upper–bounded entirely in terms of:

- **Side–information alignment** $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$: can be made arbitrarily small by iterating the centred clipping update and recentring at $\hat{\boldsymbol{x}}_{\boldsymbol{\nu}}$.
- **Optimisation tolerance** ε_ν: directly controlled by how tightly (2) is solved each round.
- **Validation bias** ε_V: reduced by enlarging or improving the validation set $\mathcal{V}$.
- **Honest dispersion** $\bar{\zeta}_h$ (intrinsic): a property of the honest cohort; independent of attackers.

Thus, for fixed $(\delta, \bar{\zeta}_h)$, increasing validation quality and optimisation accuracy, and recentering iterates toward $\boldsymbol{g}_{\mathcal{V}}$ jointly drive $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| \downarrow 0$, $\varepsilon_\nu \downarrow 0$, and $\varepsilon_V \downarrow 0$, making the Byzantine impact $\|\boldsymbol{B}\|$ *arbitrarily small* (up to the honest dispersion envelope).

Variant using supremum dispersion. If one prefers a supremum heterogeneity parameter $\zeta_h^{\max} := \max_{i\in\mathcal{H}} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|$ instead of (4), the same proof yields

$$\|\boldsymbol{B}\| \;\le\; (2-\delta)\,\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| + \varepsilon_\nu + (1-\delta)\,(\varepsilon_V + \zeta_h^{\max}).$$

This is looser but may be convenient when only a uniform heterogeneity bound is available.

Q1. 이 접근으로 “얻어지는 바” (What you get)

핵심 한 줄

한 라운드 클리핑 업데이트의 Byzantine 총 기여 벡터 $\boldsymbol{B}$에 대해

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

를 얻습니다.

- $\boldsymbol{g}_{\mathcal{V}}$: 검증 셋(Validation set)에서 얻은 **검증 그라디언트**(side information).
- $\boldsymbol{x}_0$: 이번 라운드의 **센터**(current center).
- ε_ν: **$\boldsymbol{\nu}$-선택(반경 최적화)**을 얼마나 정밀하게 풀었는지의 **최적화 허용오차**(optimisation tolerance).
- $\varepsilon_V = \|\boldsymbol{g}_{\mathcal{V}} - \bar{\boldsymbol{x}}\|$: **검증 신호 vs. honest 평균**의 **편차**(validation bias).
- $\bar{\zeta}_h = \frac{1}{|\mathcal{H}|}\sum_{i\in\mathcal{H}}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|$: **honest 분산**(honest dispersion)의 평균 크기.
- $\delta = |\mathcal{B}|/n$: Byzantine 비율.

이게 의미하는 것

- $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\|$, ε_ν, ε_V는 **우리가 줄일 수 있는 노브(knobs)**입니다.
  - 반복적으로 **리센터링(recentering)**: 매 라운드 $\boldsymbol{x}_0 \leftarrow \hat{\boldsymbol{x}}_\nu$로 옮기면 $\|\boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0\| \downarrow 0$.
  - $\boldsymbol{\nu}$-최적화를 더 **정밀**하게 풀면 $\varepsilon_\nu \downarrow 0$.
  - **검증 셋 품질/규모**를 키우면 $\varepsilon_V \downarrow 0$.
- 따라서 δ와 $\bar{\zeta}_h$가 주어지면, **공격자 기여 상한 $\|\boldsymbol{B}\|$**을 **임의로 작게** 만들 수 있습니다.
- 특히, 이 상한은 **공격자의 기하(방향 정렬, 노름 크기)**에 어떤 제약도 두지 않습니다. 공격자가 $\boldsymbol{x}_0$을 알고, 방향을 맞추고, 노름을 크게/작게 조작해도 **위 상한은 유효**합니다.
- 요컨대, worst-case에서도 side information + **최적화 정밀도**만으로 **공격 영향의 상한을 우리가 직접 컨트롤**합니다.

Q2. 이 접근의 "가정"과 그 현실성

아래는 위 상한을 얻는 데 실제로 쓰인 가정들만 **명시적으로** 정리한 것입니다. (일부는 선택적 대안도 병기)

(A) 검증 신호 가정 (Side information)

- **가정**: 서버가 공격자가 건드릴 수 없는 **검증 데이터셋**($\mathcal{V}$)을 가지고, 그로부터 $g_{\mathcal{V}}$를 계산합니다.
- **현실성**: 연합학습/분산학습에서 서버가 중앙 검증셋을 보유하거나, 라운드별 **프라이빗 샘플링/부트스트랩**으로 $g_{\mathcal{V}}$를 만드는 건 **일반적**입니다. 공격자가 $g_{\mathcal{V}}$를 "안다고" 해도, 본 상한은 유효합니다(비밀성 없이도 성립). 중요한 것은 공격자가 **검증 데이터 자체를 조작할 수 없다**는 점입니다.

(B) 반경 최적화 정밀도 (Optimisation tolerance)

- **가정**: ν-선택 문제를 풀어 $\|g_{\mathcal{V}} - \hat{x}_{\nu}\| \le \varepsilon_{\nu}$를 달성합니다.
- **현실성**: 목표 함수 $\psi_0(\nu) = \frac{1}{2}\|g_{\mathcal{V}} - \hat{x}_{\nu}\|^2$는 **연속/구간별 매끄러운**(piecewise smooth) 구조이고, 실무적으로 **그리디/좌표강하/선형탐색**으로 원하는 ε_{ν}까지 쉽게 내려갑니다. 즉, ε_{ν}는 **엔지니어링 가능한 노브**입니다.

(C) honest 분산의 유한성 (Honest dispersion)

- **가정**: $\bar{\zeta}_h = \frac{1}{|\mathcal{H}|}\sum_{i\in\mathcal{H}}\|x_i - \bar{x}\|$ (또는 $\zeta_h^{\max} = \max\|\cdot\|$)가 **유한**합니다.
- **현실성**: 표준 데이터 전처리/정규화(예: 스케일링, 그래디언트 클리핑) 하에서는 자연스럽게 성립합니다. $\bar{\zeta}_h$는 **데이터 특성**이므로 우리가 "0으로 만드는" 값은 아니지만, 바운드에서 **선형 항**으로만 등장합니다.

(D) 공격자 비율 δ (Attacker fraction)

- **가정**: $|\mathcal{B}| = \delta n$ (혹은 상계 $\delta_{\max}$)를 씁니다.
- **현실성**: 실제 시스템에서는 δ를 정확히 모를 수 있지만, **상계**만 알아도 위 식에서 $\delta \to \delta_{\max}$로 대체해 **보수적 상한**을 즉시 얻습니다. (바운드 사용에는 충분)

(E) 불필요한 가정들 (Not assumed)

- **필요 없음**: Byzantine 사이의 **정렬도/코히어런스**($\kappa_{\mathcal{B}}$) 가정 **불요.**
- **필요 없음**: Byzantine 노름의 **하한/상한**($R_{\mathcal{B}}, M_{\mathcal{B}}$) 가정 **불요.**
- **필요 없음**: 확률적/무작위성 가정 **불요.**
  → 즉, **적대적 최악(adversarial worst-case)**에서도 성립하는 상한입니다.

정리: 제한적인가? 현실적인가?

- **제한적이지 않음**: 위 상한은 **공격자 기하에 대한 제약을 전혀 요구하지 않기 때문**에, 특히 $(|\mathcal{B}| - 1)\kappa_{\mathcal{B}} < 1$ 같은 구조적 가정 없이도 통합니다. 최악의 협공/정렬/노름조작에도 그대로 적용됩니다.
- **현실적**: 필요한 건
  1. 서버가 **검증 신호**($g_{\mathcal{V}}$)를 만들 수 있고,
  2. ν-선택을 **원하는 정밀도**(ε_{ν})로 풀 수 있으며,
  3. honest 분산($\bar{\zeta}_h$)을 **측정/상계**할 수 있다는 것.
     이 셋은 현대 분산/연합 세팅에서 **일반적**으로 달성 가능한 운영 가정입니다.
- **컨트롤 가능 노브**(Controllable knobs): $\|g_{\mathcal{V}} - x_0\|$, ε_{ν}, $\varepsilon_{\mathcal{V}}$는 **우리가 설계로 줄일 수 있는 항**입니다. 따라서 $\delta, \bar{\zeta}_h$가 고정되어도, **반복**(recentering)·**정밀 최적화**·**검증 고도화**로 $\|B\|$의 상한을 **임의로 작게** 만들 수 있습니다.

Graveyard

Setup and Definitions

Data. We have n client vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^d$, a center (initial point) $\boldsymbol{x}_0 \in \mathbb{R}^d$, and a validation gradient (side information) $\boldsymbol{g}_{\mathcal{V}} \in \mathbb{R}^d$. Let $\mathcal{H}$ be the index set of honest clients and $\mathcal{B}$ that of Byzantine clients; $\mathcal{H} \cup \mathcal{B} = \{1, \dots, n\}$ and $\mathcal{H} \cap \mathcal{B} = \emptyset$.

Directions and clipping ratios. For each i with $\boldsymbol{x}_i \neq \boldsymbol{x}_0$ define the unit direction

$$\boldsymbol{d}_i \;:=\; \frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|}. \tag{10}$$

Given radii $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n) \in \mathbb{R}^n$, the (client-wise) centered clipping ratio is

$$\alpha_i(\boldsymbol{\nu}) \;:=\; \min\!\Big(1,\; \frac{\nu_i}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|}\Big) \quad \in [0,1]. \tag{11}$$

One-step aggregate. The clipped aggregate produced in a single step is

$$\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} \;:=\; \boldsymbol{x}_0 \;+\; \frac{1}{n}\sum_{i=1}^{n} \alpha_i(\boldsymbol{\nu})\,\big(\boldsymbol{x}_i - \boldsymbol{x}_0\big). \tag{12}$$

Validation fitting objective. We choose $\boldsymbol{\nu}$ by minimizing the validation mismatch

$$\psi_0(\boldsymbol{\nu}) \;:=\; \frac{1}{2}\,\big\|\boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}\big\|^2. \tag{13}$$

Step 1. Stationarity in $\boldsymbol{\nu_j}$

Differentiate (13) with respect to ν_j. By the chain rule,

$$\frac{\partial \psi_0}{\partial \nu_j} = -\Big\langle \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}},\; \frac{\partial \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}}{\partial \nu_j}\Big\rangle. \tag{14}$$

From (12) and (11),

$$\frac{\partial \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}}{\partial \nu_j} = \frac{1}{n}\,(\boldsymbol{x}_j - \boldsymbol{x}_0)\,\frac{\partial \alpha_j}{\partial \nu_j} = \begin{cases} \frac{1}{n}\,\boldsymbol{d}_j, & \text{if } \nu_j < \|\boldsymbol{x}_j - \boldsymbol{x}_0\| \quad \text{(unsaturated)},\\ 0, & \text{if } \nu_j \geq \|\boldsymbol{x}_j - \boldsymbol{x}_0\| \quad \text{(saturated)}. \end{cases}$$

Hence, at any (first-order) stationary point with $\nu_j < \|\boldsymbol{x}_j - \boldsymbol{x}_0\|$,

$$\boxed{\Big\langle \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}},\; \boldsymbol{d}_j \Big\rangle = 0.} \tag{15}$$

Step 2. Exact Projection Identity

Introduce the error vector

$$\boldsymbol{e} \;:=\; \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}. \tag{16}$$

Plugging (12) into (16) and taking inner product with $\boldsymbol{d}_j$,

$$\begin{aligned}\langle \boldsymbol{e},\ \boldsymbol{d}_j\rangle &= \Big\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0 - \frac{1}{n}\sum_{i=1}^{n}\alpha_i(\boldsymbol{\nu})\,(\boldsymbol{x}_i - \boldsymbol{x}_0),\ \boldsymbol{d}_j\Big\rangle \\ &= \underbrace{\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0,\ \boldsymbol{d}_j\rangle}_{\text{term (A)}} - \frac{1}{n}\sum_{i=1}^{n}\alpha_i(\boldsymbol{\nu})\,\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\ \langle \boldsymbol{d}_i,\ \boldsymbol{d}_j\rangle. \end{aligned} \tag{17}$$

By (15), the left-hand side of (17) equals 0 when ν_j is unsaturated.

Step 3. Splitting the Sum and Bounding Honest Cross Terms

Split the sum in (17) over $\mathcal{H}$ and $\mathcal{B}$:

$$\sum_{i=1}^{n}\alpha_i\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle \;=\; \underbrace{\sum_{i\in\mathcal{H}}\alpha_i\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle}_{S_{\mathcal{H}}(j)} + \underbrace{\sum_{i\in\mathcal{B}}\alpha_i\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle}_{S_{\mathcal{B}}(j)}. \tag{18}$$

Assumptions for honest dispersion. Let $\bar{\boldsymbol{x}} := \frac{1}{|\mathcal{H}|}\sum_{k\in\mathcal{H}}\boldsymbol{x}_k$ denote the honest mean. Assume there exist finite constants

$$\|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}\| \le \varepsilon_0, \qquad \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\| \le \zeta_h \quad \forall\, i\in\mathcal{H}, \qquad \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \ge R_H > 0 \quad \forall\, i\in\mathcal{H}. \tag{19}$$

These say: the current center $\boldsymbol{x}_0$ and honest client vectors stay in a bounded neighborhood of the honest mean, and honest displacements are not degenerate.

Bounding a single honest inner product. Write $\boldsymbol{x}_i - \boldsymbol{x}_0 = (\bar{\boldsymbol{x}} - \boldsymbol{x}_0) + (\boldsymbol{x}_i - \bar{\boldsymbol{x}})$. Then

$$\begin{aligned}|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| &= \frac{|\langle \boldsymbol{x}_i - \boldsymbol{x}_0,\ \boldsymbol{x}_j - \boldsymbol{x}_0\rangle|}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\,\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \\ &= \frac{|\langle \bar{\boldsymbol{x}} - \boldsymbol{x}_0,\ \boldsymbol{x}_j - \boldsymbol{x}_0\rangle + \langle \boldsymbol{x}_i - \bar{\boldsymbol{x}},\ \boldsymbol{x}_j - \boldsymbol{x}_0\rangle|}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\,\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \\ &\le \frac{\|\bar{\boldsymbol{x}} - \boldsymbol{x}_0\|\ \|\boldsymbol{x}_j - \boldsymbol{x}_0\| + \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|\ \|\boldsymbol{x}_j - \boldsymbol{x}_0\|}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\,\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \le \frac{\varepsilon_0 + \zeta_h}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|} \le \frac{\varepsilon_0 + \zeta_h}{R_H}. \end{aligned} \tag{20}$$

Bounding the honest sum $S_{\mathcal{H}}(j)$. Using $0 \le \alpha_i \le 1$ and (20),

$$\begin{aligned}|S_{\mathcal{H}}(j)| &\le \sum_{i\in\mathcal{H}}\alpha_i\,\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\ |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| \le \sum_{i\in\mathcal{H}}\|\boldsymbol{x}_i - \boldsymbol{x}_0\|\ \frac{\varepsilon_0 + \zeta_h}{R_H} \\ &\le |\mathcal{H}|\,M_H\,\frac{\varepsilon_0 + \zeta_h}{R_H} \;=:\; \rho_H, \end{aligned} \tag{21}$$

where $M_H := \max_{i\in\mathcal{H}}\|\boldsymbol{x}_i - \boldsymbol{x}_0\|$.

Step 4. Bounding Byzantine Cross Terms Except j

Write $S_{\mathcal{B}}(j) = \alpha_j \|\boldsymbol{x}_j - \boldsymbol{x}_0\| \langle \boldsymbol{d}_j, \boldsymbol{d}_j \rangle + \sum_{i \in \mathcal{B},\, i \neq j} \alpha_i \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle$. Since $\langle \boldsymbol{d}_j, \boldsymbol{d}_j \rangle = 1$,

$$S_{\mathcal{B}}(j) = \alpha_j \|\boldsymbol{x}_j - \boldsymbol{x}_0\| + S_{\mathcal{B}}^{-j}(j), \qquad S_{\mathcal{B}}^{-j}(j) := \sum_{\substack{i \in \mathcal{B} \\ i \neq j}} \alpha_i \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle. \tag{22}$$

Incoherence among Byzantine directions. Assume there exists $\kappa_{\mathcal{B}} \in [0, 1)$ such that

$$|\langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle| \leq \kappa_{\mathcal{B}} \qquad \text{for all distinct } i, j \in \mathcal{B}. \tag{23}$$

Let $M_{\mathcal{B}} := \max_{i \in \mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|$ and define the maximal Byzantine clipping ratio

$$\alpha_{\mathcal{B}}^{\max} := \max_{i \in \mathcal{B}} \alpha_i. \tag{24}$$

Then from (22) and (23),

$$|S_{\mathcal{B}}^{-j}(j)| \leq \sum_{\substack{i \in \mathcal{B} \\ i \neq j}} \alpha_i \, \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \; |\langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle| \leq (|\mathcal{B}| - 1) \, \alpha_{\mathcal{B}}^{\max} \, M_{\mathcal{B}} \, \kappa_{\mathcal{B}}. \tag{25}$$

Step 5. Solving Stationarity for α_j and a Fixed-Point Bound

Insert the split (18) into (17), use (21), (22), (25), and recall that $\langle \boldsymbol{e}, \boldsymbol{d}_j \rangle = 0$ by (15). We obtain

$$0 = \langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0, \ \boldsymbol{d}_j \rangle - \frac{1}{n} \Big(\alpha_j \|\boldsymbol{x}_j - \boldsymbol{x}_0\| + S_{\mathcal{B}}^{-j}(j) + S_{\mathcal{H}}(j) \Big). \tag{26}$$

Rearranging (26) yields the exact identity

$$\alpha_j = \frac{n \, \langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0, \ \boldsymbol{d}_j \rangle - S_{\mathcal{B}}^{-j}(j) - S_{\mathcal{H}}(j)}{\|\boldsymbol{x}_j - \boldsymbol{x}_0\|}. \tag{27}$$

Taking absolute values and using (21) and (25),

$$|\alpha_j| \ \leq \ \frac{n \, |\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0, \ \boldsymbol{d}_j \rangle|}{\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \ + \ \frac{\rho_H}{\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \ + \ \frac{(|\mathcal{B}| - 1) \, M_{\mathcal{B}} \, \kappa_{\mathcal{B}}}{\|\boldsymbol{x}_j - \boldsymbol{x}_0\|} \ \alpha_{\mathcal{B}}^{\max}. \tag{28}$$

Let $R_{\mathcal{B}} := \min_{i \in \mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|$ and

$$\eta := \max_{j \in \mathcal{B}} |\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0, \ \boldsymbol{d}_j \rangle|. \tag{29}$$

Then from (28),

$$\max_{j \in \mathcal{B}} |\alpha_j| \ \leq \ \underbrace{\frac{n \, \eta + \rho_H}{R_{\mathcal{B}}}}_{=: \, \tau} \ + \ \underbrace{\frac{(|\mathcal{B}| - 1) \, \kappa_{\mathcal{B}}}{R_{\mathcal{B}}} \ M_{\mathcal{B}}}_{=: \, \beta} \ \alpha_{\mathcal{B}}^{\max}. \tag{30}$$

Since the left side equals $\alpha_{\mathcal{B}}^{\max}$ by definition (24), (30) is a *fixed-point inequality*:

$$\alpha_{\mathcal{B}}^{\max} \ \leq \ \tau + \beta \, \alpha_{\mathcal{B}}^{\max}. \tag{31}$$

If the incoherence factor satisfies $\beta < 1$ (i.e., $R_{\mathcal{B}} > (|\mathcal{B}| - 1) M_{\mathcal{B}} \kappa_{\mathcal{B}}$), then (31) implies the explicit bound

$$\boxed{\alpha_{\mathcal{B}}^{\max} \ \leq \ \frac{\tau}{1 - \beta} \ = \ \frac{n \, \eta + \rho_H}{R_{\mathcal{B}} - (|\mathcal{B}| - 1) M_{\mathcal{B}} \kappa_{\mathcal{B}}}.} \tag{32}$$

Interpretation. The quantity η in (29) measures how well the validation direction $\boldsymbol{g}_{\mathcal{V}}$ suppresses any Byzantine direction $\boldsymbol{d}_j$ via the inner product; ρ_H from (21) is the (controlled) leakage from honest cross terms; β captures mutual alignment among Byzantine directions. When η and ρ_H are small (strong validation hint, tight honest dispersion) and $\beta < 1$ (no near-collinearity among Byzantine directions), (45) forces every Byzantine clipping ratio to be small.

$$
\begin{aligned}
\eta &:= \max_{j\in\mathcal{B}} |\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0,\ \boldsymbol{d}_j\rangle|. \\
\rho_H &:= |\mathcal{H}|\, M_H \, \frac{\varepsilon_0 + \zeta_h}{R_H}, \\
M_H &:= \max_{i\in\mathcal{H}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|, \\
\varepsilon_0 &:= \|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}\| \\
\zeta_h &:= \max_{i\in\mathcal{H}} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|, \\
\beta &:= \frac{(|\mathcal{B}|-1)\,\kappa_{\mathcal{B}}}{R_{\mathcal{B}}}\ M_{\mathcal{B}}, \\
\kappa_{\mathcal{B}} &:= \max_{\substack{i\neq j,\\ i,j\in\mathcal{B}}} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|, \\
R_{\mathcal{B}} &:= \min_{j\in\mathcal{B}} \|\boldsymbol{x}_j - \boldsymbol{x}_0\|, \\
M_{\mathcal{B}} &:= \max_{j\in\mathcal{B}} \|\boldsymbol{x}_j - \boldsymbol{x}_0\|.
\end{aligned}
$$

Three values at the bottom are under full control of omniscient Byzantines.

Step 6. Consequence for the One-Step Aggregate

From (12),

$$
\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} = \boldsymbol{x}_0 + \frac{1}{n}\sum_{i\in\mathcal{H}} \alpha_i(\boldsymbol{\nu})(\boldsymbol{x}_i - \boldsymbol{x}_0) + \frac{1}{n}\sum_{j\in\mathcal{B}} \alpha_j(\boldsymbol{\nu})(\boldsymbol{x}_j - \boldsymbol{x}_0). \tag{33}
$$

Hence the Byzantine contribution is bounded by

$$
\Big\|\frac{1}{n}\sum_{j\in\mathcal{B}} \alpha_j(\boldsymbol{\nu})(\boldsymbol{x}_j - \boldsymbol{x}_0)\Big\| \ \le\ \frac{|\mathcal{B}|}{n}\ \alpha_{\mathcal{B}}^{\max} M_{\mathcal{B}} \ =\ \delta\ \alpha_{\mathcal{B}}^{\max} M_{\mathcal{B}}, \tag{34}
$$

where $\delta = |\mathcal{B}|/n$. Combining (34) with (45) gives a fully explicit upper bound on the Byzantine distortion of the one-step aggregate in terms of *observable or design* constants $(\varepsilon_0, \zeta_h, R_H)$, $(M_{\mathcal{B}}, R_{\mathcal{B}}, \kappa_{\mathcal{B}})$, and the validation alignment η.

A Byzantine Contribution Bound Without Using β

This section provides a bound on the *actual Byzantine contribution* that does not depend on the norm parameters $R_{\mathcal{B}} = \min_{j\in\mathcal{B}} \|\boldsymbol{x}_j - \boldsymbol{x}_0\|$ and $M_{\mathcal{B}} = \max_{j\in\mathcal{B}} \|\boldsymbol{x}_j - \boldsymbol{x}_0\|$. The bound is stated directly in terms of: (i) the *directional coherence* among Byzantine directions, (ii) the *validation alignment* with Byzantine directions, and (iii) an a priori bound on the honest cross term.

Notation and standing assumptions. Let $\mathcal{H}$ and $\mathcal{B}$ denote the honest and Byzantine index sets, respectively, with $|\mathcal{B}| = \delta n$. At the current round, the center is $\boldsymbol{x}_0 \in \mathbb{R}^d$ and the client proposals are $\boldsymbol{x}_i \in \mathbb{R}^d$. Define unit directions

$$\boldsymbol{d}_i \;:=\; \frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|} \quad (\,\boldsymbol{x}_i \neq \boldsymbol{x}_0\,), \qquad \alpha_i(\boldsymbol{\nu}) \;:=\; \min\!\Big(1,\; \frac{\nu_i}{\|\boldsymbol{x}_i - \boldsymbol{x}_0\|}\Big) \in [0,1]. \tag{35}$$

The one-step clipped aggregate is $\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} = \boldsymbol{x}_0 + \frac{1}{n}\sum_{i=1}^{n} \alpha_i(\boldsymbol{\nu})(\boldsymbol{x}_i - \boldsymbol{x}_0)$. As in the main text, we assume the *stationarity* condition holds for every *unsaturated* coordinate (i.e. $\nu_j < \|\boldsymbol{x}_j - \boldsymbol{x}_0\|$):

$$\big\langle \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}},\; \boldsymbol{d}_j \big\rangle \;=\; 0. \tag{36}$$

We will only invoke (36) for indices $j \in \mathcal{B}$ that are unsaturated. [1]

Finally, define:

$$\eta \;:=\; \max_{j\in\mathcal{B}} \big|\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0,\; \boldsymbol{d}_j\rangle\big|, \qquad \kappa_{\mathcal{B}} \;:=\; \max_{\substack{i,j\in\mathcal{B}\\ i\neq j}} \big|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle\big|, \tag{37}$$

and let ρ_H be any (uniform-in-j) bound on the honest cross term

$$\Big|\, S_{\mathcal{H}}(j) \Big| \;:=\; \Big| \sum_{i\in\mathcal{H}} \alpha_i(\boldsymbol{\nu})\; \|\boldsymbol{x}_i - \boldsymbol{x}_0\|\; \langle \boldsymbol{d}_i,\, \boldsymbol{d}_j\, \rangle \Big| \;\leq\; \rho_H, \qquad \forall\, j \in \mathcal{B}. \tag{38}$$

(For example, one may take the explicit ρ_H from the honest-dispersion bound in the main text.)

We now work with the *Byzantine contribution magnitudes*

$$C_j \;:=\; \alpha_j(\boldsymbol{\nu})\; \|\boldsymbol{x}_j - \boldsymbol{x}_0\|, \qquad C_{\max} \;:=\; \max_{j\in\mathcal{B}} C_j. \tag{39}$$

Lemma A.1 (Projection identity for a fixed Byzantine index)**.** *Fix $j \in \mathcal{B}$ such that $\nu_j < \|\boldsymbol{x}_j - \boldsymbol{x}_0\|$. Then, using* (36)*,*

$$n\; \big\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0,\; \boldsymbol{d}_j \big\rangle \;=\; \underbrace{\sum_{i\in\mathcal{H}} \alpha_i \|\boldsymbol{x}_i - \boldsymbol{x}_0\|\, \langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle}_{S_{\mathcal{H}}(j)} \;+\; \underbrace{\sum_{i\in\mathcal{B}} \alpha_i \|\boldsymbol{x}_i - \boldsymbol{x}_0\|\, \langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle}_{S_{\mathcal{B}}(j)}. \tag{40}$$

Moreover,

$$S_{\mathcal{B}}(j) \;=\; C_j \;+\; \sum_{\substack{i\in\mathcal{B}\\ i\neq j}} C_i\, \langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle. \tag{41}$$

Proof. From $\hat{\boldsymbol{x}}_{\boldsymbol{\nu}} = \boldsymbol{x}_0 + \frac{1}{n}\sum_i \alpha_i(\boldsymbol{\nu})(\boldsymbol{x}_i - \boldsymbol{x}_0)$ and (36), we get

$$0 = \big\langle \boldsymbol{g}_{\mathcal{V}} - \hat{\boldsymbol{x}}_{\boldsymbol{\nu}}, \boldsymbol{d}_j \big\rangle = \big\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0, \boldsymbol{d}_j \big\rangle - \frac{1}{n}\sum_{i=1}^{n} \alpha_i \|\boldsymbol{x}_i - \boldsymbol{x}_0\| \langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle,$$

which rearranges to (40). The decomposition (41) is the definition of C_i plus separating the $i = j$ term. $\square$

[1] If a particular $j \in \mathcal{B}$ is saturated, then $\alpha_j = 1$ and the bound below trivially controls its contribution through the $\kappa_{\mathcal{B}}$ term; alternatively, one can work with subgradient KKT conditions.

Theorem A.2 (Byzantine magnitude bound independent of norms)**.** *Assume* (38) *holds and define* $\eta, \kappa_{\mathcal{B}}$ *as in* (37)*. Then*

$$\big(1 - (|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}}\big)\; C_{\max} \;\le\; n\,\eta + \rho_H. \tag{42}$$

In particular, if $(|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}} < 1$*, then*

$$\boxed{C_{\max} \;\le\; \frac{n\,\eta + \rho_H}{1 - (|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}}}.} \tag{43}$$

Proof. Fix $j \in \mathcal{B}$ unsaturated and start from (40). Taking absolute values and using (37) and (38),

$$C_j \;\le\; n\,\eta \;+\; |S_{\mathcal{H}}(j)| \;+\; \sum_{\substack{i\in\mathcal{B}\\ i\neq j}} C_i\,|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| \;\le\; n\,\eta \;+\; \rho_H \;+\; (|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}}\, C_{\max}.$$

Now take the maximum over $j \in \mathcal{B}$ on the left to obtain

$$C_{\max} \;\le\; n\,\eta \;+\; \rho_H \;+\; (|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}}\, C_{\max},$$

which rearranges to (42) and yields (43) when $(|\mathcal{B}| - 1)\kappa_{\mathcal{B}} < 1$. □

Corollary A.3 (Aggregate Byzantine contribution)**.** *The total Byzantine contribution to the one-step aggregate satisfies*

$$\left\|\frac{1}{n}\sum_{j\in\mathcal{B}} \alpha_j(\boldsymbol{\nu})\,(\boldsymbol{x}_j - \boldsymbol{x}_0)\right\| \;\le\; \frac{|\mathcal{B}|}{n}\, C_{\max} \;=\; \delta\, C_{\max} \;\le\; \boxed{\frac{\delta}{1 - (|\mathcal{B}| - 1)\,\kappa_{\mathcal{B}}}\;\big(n\,\eta + \rho_H\big).} \tag{44}$$

Remarks. (i) The bounds (43)–(44) do *not* involve $R_{\mathcal{B}}$ or $M_{\mathcal{B}}$; hence they are robust even if an attacker knows $\boldsymbol{x}_0$ and attempts to manipulate vector norms. (ii) The only structural requirement is $(|\mathcal{B}| - 1)\kappa_{\mathcal{B}} < 1$, i.e. Byzantine directions are not nearly collinear; this condition is typically mild in moderate/high dimension (and can be enforced with tiny dithering). (iii) The *numerators* $n\eta + \rho_H$ are *algorithmically controllable*: validation fitting and iteration can drive $\eta{\downarrow}0$, and the honest bias/dispersion bound ρ_H can be reduced by iteration and data curation. Consequently, (44) shows the Byzantine impact can be made arbitrarily small under a mild directional incoherence condition.

Sufficient Conditions for $\beta < 1$

To make this supplement self-contained and directly pluggable into `res.tex`, we recall the key quantities and the bound we reference. Let $\mathcal{H}$ and $\mathcal{B}$ be the honest and Byzantine index sets, respectively. For each client i, let $\boldsymbol{x}_i \in \mathbb{R}^d$ and let the iteration center be $\boldsymbol{x}_0 \in \mathbb{R}^d$. Define directions $\boldsymbol{d}_i := (\boldsymbol{x}_i - \boldsymbol{x}_0)/\|\boldsymbol{x}_i - \boldsymbol{x}_0\|$ whenever $\boldsymbol{x}_i \neq \boldsymbol{x}_0$. Let

$$M_{\mathcal{B}} := \max_{i\in\mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|, \quad R_{\mathcal{B}} := \min_{i\in\mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|, \quad \kappa_{\mathcal{B}} := \max_{\substack{i,j\in\mathcal{B}\\ i\neq j}} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|.$$

Write $\alpha_{\mathcal{B}}^{\max} := \max_{i\in\mathcal{B}} \alpha_i$ for the maximal Byzantine clipping ratio, and

$$\eta := \max_{j\in\mathcal{B}} |\langle \boldsymbol{g}_{\mathcal{V}} - \boldsymbol{x}_0,\; \boldsymbol{d}_j\rangle|, \qquad \rho_H := |\mathcal{H}|\, M_H\,(\varepsilon_0 + \zeta_h)/R_H,$$

where $(\varepsilon_0, \zeta_h, R_H, M_H)$ summarize the honest dispersion/bias constants defined in the main text. The fixed-point inequality derived earlier yields the explicit bound

$$\boxed{\alpha_{\mathcal{B}}^{\max} \;\le\; \frac{n\,\eta + \rho_H}{R_{\mathcal{B}} - (|\mathcal{B}| - 1)\, M_{\mathcal{B}}\,\kappa_{\mathcal{B}}} \quad \text{whenever} \quad R_{\mathcal{B}} > (|\mathcal{B}| - 1)\, M_{\mathcal{B}}\,\kappa_{\mathcal{B}}.} \tag{45}$$

We now state clean sufficient conditions ensuring the denominator in (45) is positive, i.e., $\beta < 1$ below.

Sufficient Conditions for $\beta < 1$

Recall (from (45)) that

$$\beta \;=\; \frac{(|\mathcal{B}|-1)\, M_{\mathcal{B}}\, \kappa_{\mathcal{B}}}{R_{\mathcal{B}}}, \qquad M_{\mathcal{B}} := \max_{i\in\mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|, \quad R_{\mathcal{B}} := \min_{i\in\mathcal{B}} \|\boldsymbol{x}_i - \boldsymbol{x}_0\|, \quad \kappa_{\mathcal{B}} := \max_{\substack{i,j\in\mathcal{B}\\ i\neq j}} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|.$$

Lemma A.4 (Equivalences)**.** *The following are equivalent:*

$$\begin{aligned} &\text{(i)} && \beta < 1,\\ &\text{(ii)} && R_{\mathcal{B}} > (|\mathcal{B}|-1)\, M_{\mathcal{B}}\, \kappa_{\mathcal{B}},\\ &\text{(iii)} && \kappa_{\mathcal{B}} < \frac{R_{\mathcal{B}}}{(|\mathcal{B}|-1)\, M_{\mathcal{B}}},\\ &\text{(iv)} && |\mathcal{B}|-1 < \frac{R_{\mathcal{B}}}{M_{\mathcal{B}}\, \kappa_{\mathcal{B}}} \quad (\textit{when } \kappa_{\mathcal{B}} > 0). \end{aligned}$$

Proof. *All statements are immediate rearrangements of $\beta = (|\mathcal{B}|-1)M_{\mathcal{B}}\kappa_{\mathcal{B}}/R_{\mathcal{B}}$.* □

Corollary A.5 (Equal Byzantine magnitudes)**.** *If all Byzantine displacements have the same norm $\|\boldsymbol{x}_i - \boldsymbol{x}_0\| \equiv r_{\mathcal{B}}$ (thus $R_{\mathcal{B}} = M_{\mathcal{B}} = r_{\mathcal{B}}$), then*

$$\beta \;=\; (|\mathcal{B}|-1)\, \kappa_{\mathcal{B}} \quad \textit{and hence} \quad \beta < 1 \iff \kappa_{\mathcal{B}} < \frac{1}{|\mathcal{B}|-1}.$$

Corollary A.6 (Bound on attacker count for given geometry)**.** *Suppose $\kappa_{\mathcal{B}} \le \bar{\kappa} < 1$ and $R_{\mathcal{B}} \ge r > 0$, $M_{\mathcal{B}} \le m < \infty$. If*

$$|\mathcal{B}| \;\le\; 1 + \left\lfloor \frac{r}{m\,\bar{\kappa}} \right\rfloor,$$

then $\beta < 1$. Proof. *By Lemma A.4(iv), $|\mathcal{B}|-1 < R_{\mathcal{B}}/(M_{\mathcal{B}}\kappa_{\mathcal{B}}) \ge r/(m\bar{\kappa})$.* □

Corollary A.7 (Angular separation)**.** *Let $\theta_{\min}$ be the minimal pairwise angle between Byzantine directions:*

$$\theta_{\min} \;:=\; \min_{\substack{i,j\in\mathcal{B}\\ i\neq j}} \arccos\big(|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|\big).$$

Then $\kappa_{\mathcal{B}} = \max_{i\neq j} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| \le \cos\theta_{\min}$, and a sufficient condition for $\beta < 1$ is

$$\theta_{\min} \;>\; \arccos\Big(\frac{R_{\mathcal{B}}}{(|\mathcal{B}|-1)\, M_{\mathcal{B}}}\Big).$$

In particular, if Byzantine directions are pairwise orthogonal ($\theta_{\min} = 90^\circ$, so $\kappa_{\mathcal{B}} = 0$), then $\beta = 0$ automatically.

Corollary A.8 (Norm separation)**.** *If Byzantine norms are not too imbalanced, i.e., $R_{\mathcal{B}}/M_{\mathcal{B}} \ge \gamma$ for some $\gamma \in (0,1]$, and the pairwise coherence satisfies $\kappa_{\mathcal{B}} \le c$, then*

$$\beta \;\le\; \frac{|\mathcal{B}|-1}{\gamma}\, c.$$

Hence a simple sufficient condition is

$$c \;<\; \frac{\gamma}{|\mathcal{B}|-1} \quad (\textit{equivalently, } \kappa_{\mathcal{B}} < \gamma/(|\mathcal{B}|-1)).$$

Practical reading.

- **Few attackers or weak alignment.** For fixed geometry $(R_{\mathcal{B}}, M_{\mathcal{B}})$, either keep $|\mathcal{B}|$ small or ensure Byzantine directions are poorly aligned ($\kappa_{\mathcal{B}}$ small).
- **Equal-norm case is sharp.** When $R_{\mathcal{B}} = M_{\mathcal{B}}$, the clean threshold is $\kappa_{\mathcal{B}} < 1/(|\mathcal{B}| - 1)$ (Cor. A.5).
- **Angle or coherence budgets.** If you can certify a minimum inter-attacker angle $\theta_{\min}$, then Cor. A.7 turns it into a direct check.
- **Robust to scaling.** If attackers cannot make one vector arbitrarily small ($R_{\mathcal{B}}$ not tiny) while growing the others ($M_{\mathcal{B}}$ not huge), Cor. A.8 gives an immediate margin.

Example: High–dimensional random directions ("extremely mild" coherence). Let $m := |\mathcal{B}|$ and suppose the Byzantine directions $\{\boldsymbol{d}_i\}_{i\in\mathcal{B}}$ are independent and uniformly distributed on the unit sphere $\mathbb{S}^{d-1}$ (or sufficiently "random-looking" so that spherical concentration applies).

Lemma (Spherical concentration + union bound). For any $t \in (0,1)$,

$$\mathbb{P}\Big(\max_{i\neq j\in\mathcal{B}} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| \;\geq\; t\Big) \;\leq\; 2\,m(m-1)\,\exp\Big(-\tfrac{(d-1)\,t^2}{2}\Big).$$

Sketch. For fixed $i \neq j$, $|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|$ is sub-Gaussian with tail $\leq 2\exp(-(d-1)t^2/2)$; take a union bound over $m(m-1)$ pairs.

Dimension threshold for the norm–separation condition. Target the sufficient condition of Cor. A.8 with $c = t = \gamma/(m-1)$. Given failure probability $\delta \in (0,1)$, it suffices to pick

$$d \;\geq\; 1 \;+\; \frac{2\,(m-1)^2}{\gamma^2}\,\log\Big(\frac{2\,m(m-1)}{\delta}\Big) \tag{46}$$

so that with probability at least $1-\delta$ one has $\kappa_{\mathcal{B}} = \max_{i\neq j} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle| \leq \gamma/(m-1)$ and hence $\beta \leq \frac{m-1}{\gamma}\,\kappa_{\mathcal{B}} < 1$ by Cor. A.8.

Numerical illustrations.

- *Moderate d, few attackers.* Take $\gamma = 1$ (equal Byzantine norms), $m = 5$ (four attackers), and $\delta = 0.01$. Then (46) gives

  $$d \;\geq\; 1 + 2\cdot 4^2\,\log\Big(\tfrac{2\cdot 5\cdot 4}{0.01}\Big) \;=\; 1 + 32\,\log(4000) \;\approx\; 267.$$

  Thus in dimension $d \geq 267$, with probability at least 99% we have $\kappa_{\mathcal{B}} \leq 1/4$ and the condition $\beta < 1$ holds.

- *Larger m, still mild in high d.* Let $\gamma = 1$, $m = 8$ and $\delta = 10^{-6}$. Then (46) yields

  $$d \;\geq\; 1 + 2\cdot 7^2\,\log\Big(\tfrac{2\cdot 8\cdot 7}{10^{-6}}\Big) \;\approx\; 1 + 98\cdot 18.53 \;\approx\; 1818.$$

  So in $d \geq 1818$ (well within common model dimensions), with probability at least $1 - 10^{-6}$ we have $\kappa_{\mathcal{B}} \leq 1/7$ and hence $\beta < 1$.

Two–attacker special case (equal norms). When $m = 2$ and $R_{\mathcal{B}} = M_{\mathcal{B}}$ (equal norms), the condition is simply $\kappa_{\mathcal{B}} < 1$, i.e., the two directions are not perfectly collinear. This holds with probability 1 under any continuous-noise model and is thus *extremely mild.* For a robust quantitative margin, one may enforce $\kappa_{\mathcal{B}} \leq 1 - \xi$ ($\xi \in (0,1)$), which yields the well-conditioned bound $\alpha_{\mathcal{B}}^{\max} \leq (n\,\eta + \rho_H)/(R_{\mathcal{B}}\,\xi)$ from the fixed-point inequality.