arXiv:2602.03899v1 [stat.ML] 3 Feb 2026

拜占庭机器学习：MultiKrum 与一种最优鲁棒性概念

Gilles Bareilles*, †Wassim Bouaziz*, ‡Julien Fageot*, ◇和 El Mahdi El Mhamdi†

摘要。聚合规则是存在对手时分布式（或联邦）学习的基石，这种情形遵循所谓的拜占庭威胁模型。从鲁棒均值估计的角度来看，它们也是值得研究的数学对象。Krum 聚合规则已得到广泛研究，并具备形式化的鲁棒性和收敛保证。然而，尽管此前尚无理论保证，作为 Krum 的自然扩展，MultiKrum 因其更优的实证表现而常在实践中受到青睐。在这项工作中，我们首次证明 MultiKrum 是一种鲁棒聚合规则，并给出其鲁棒性系数的界。为此，我们引入 $\kappa^\star$聚合规则的最优鲁棒性系数，该系数能够以比以往采用的鲁棒性概念更精确的方式，量化存在对手时均值估计的准确性。随后，我们构造了 MultiKrum 鲁棒性系数的上界和下界。作为推论，我们还改进了目前已知的 Krum 鲁棒性系数界。我们证明，MultiKrum 的界从不劣于 Krum 的界，并且在现实情形下更优。最后，我们通过实验研究下界的质量，以说明这一分析。

1. 引言

视觉、自然语言处理和商品推荐领域的现代机器学习任务涉及规模不断扩大的数据集和模型，如今也越来越依赖于在多个计算设备之间分发数据和模型。在这种分布式学习设置中，中央设备（或模拟中央设备的去中心化协议）负责协调学习任务，而其他设备则基于本地数据或模型执行计算。

分布式学习虽然提升了可扩展性，但也引入了诸多漏洞，例如训练期间的对抗攻击 [BEGS17, EGR18, AAZL18, XKG20, EMFG+21, HFJ+21]、针对数据集的攻击 [CGD+23, BUEM25, NQVHN+24]，或针对模型本身的攻击 [WSR+20, LLZ25]。这些威胁催生了拜占庭式机器学习这一领域，其灵感源于分布式计算中的拜占庭容错 [LSP82], ，在这种情况下，一部分设备可能会采取对抗行为。

在分布式、协作式和联邦学习中 [BMMB22, KMA+21], [EMFG+21]确立了基于梯度的鲁棒分布式学习与鲁棒均值估计之间的等价关系，从而可以将学习的多个方面抽象为一个均值估计问题。

这一框架的核心是聚合规则，用于合并诚实工作节点和拜占庭工作节点的输入，而无需知道哪些节点属于哪一类。早期研究包括几何中位数 [Rou85], ，随后是 Krum 和 MultiKrum [BEGS17],发现基于距离的方法存在漏洞，并提出 Bulyan [EGR18], 以及

*同等贡献。
†巴黎综合理工学院 CMAP，法国。
‡Mistral AI，法国——工作在巴黎综合理工学院 CMAP 完成。◇洛桑联邦理工学院，瑞士。电子邮箱：first.last@polytechnique.edu、first.last@epfl.ch。

对逐坐标聚合规则的分析 [YCKB18]。此外， [FGV+22，BUEM25]指出，分布式学习中出现的梯度攻击与更为普遍的数据投毒攻击之间存在等价关系；后者甚至会出现在单机场景中。在大型人工智能模型（如大语言模型）的背景下，这种攻击甚至可以间接发生 [BVUEM26]，且所需样本数近乎恒定 [SRC+25]。因此，提升鲁棒聚合的安全性在机器学习中具有重要意义，其影响超越了分布式场景——在该场景中，机器会交换梯度和模型更新。

伴随这些进展，鲁棒性的形式化定义也不断演变：从对可容忍对手数量的界，到 (f, α)-拜占庭弹性 [BEGS17], ，再到目前公认的 (f, κ)-鲁棒性 [AFG+23]。这一概念采用最坏情况视角，要求聚合输出与诚实向量平均值之间的距离有界，且该界取决于它们的内在异质性。

这一概念颇具吸引力：常见的聚合规则对于某个 κ 都是 (f, κ)-鲁棒的，这一点尤其保证了分布式随机梯度下降能够收敛到一个使诚实设备的损失最小化的点，且性能退化幅度以 κ 为线性界。不过，(f, κ)-鲁棒性无法通过一个唯一且定义明确的量来刻画鲁棒性，因此用户在权衡不同聚合规则所带来的鲁棒性与优化性能取舍时，难以做出明智决策。

为解决这一局限性，我们将聚合规则的鲁棒性系数 $\kappa^\star$ 引入为一个唯一确定的量，并将其定义为某个优化问题的解。这一概念扩展了（f， κ）-鲁棒性的概念。我们还定义了鲁棒性系数的上界和下界这两个互补概念。这表明，先前认为是紧的若干（f， κ）-鲁棒性保证实际上并不紧。我们的定义是最优的，因为它刻画了这样一个最小常数：在诚实参与者身份的最坏情况下，该常数能够根据诚实参与者的异质性进行缩放，从而界定聚合输出与诚实参与者平均值之间的距离。[1]

独立于理论研究，实践者明显更偏好$MultiKrum$ 聚合规则（它是 $Krum$ 聚合的扩展）而不是 Krum [BEGS17]。据我们所知，目前尚无关于 MultiKrum 鲁棒性的证明。我们通过提出首个关于 MultiKrum 鲁棒性系数的结果来弥补这一理论空白，从而证明其鲁棒性。

1.1. 贡献。 我们的主要贡献可概括如下：

- 我们提出了聚合规则的鲁棒性系数 $\kappa^\star$ 这一概念，通过一个定义明确的优化问题量化其鲁棒性。
- 我们为广受欢迎的 MultiKrum 和 Krum 聚合规则的鲁棒性系数给出了更精确的上界和下界。尤其是，我们首次证明了 MultiKrum 具有鲁棒性。

在此过程中，我们还改进了 Krum 的鲁棒性系数，并指出，与以往的看法相反，先前被认为“紧致”的鲁棒性分析实际上并不紧致。

本文其余部分结构如下。第2节回顾MultiKrum估计器，并定义鲁棒性系数。第3节给出关于MultiKrum的鲁棒性系数的主要结果。第4节给出主要结果的证明。第5节对MultiKrum的鲁棒性系数的界进行数值研究。最后，我们以第6节作结。

[1] 请注意，我们并未称其为“唯一最优”，而是称为“一种最优”，因为其他理想的 le
char特性（而不仅仅是接近诚实平均值）也会导向其他判据。

聚合	Krum [非拜占庭梯度下降+23]	Krum（本文工作）	MultiKrum $m = n - f$ (本文工作)
上界	$6\frac{n-f}{n-2f}$	$(\sqrt{2}+1)^2\frac{n-f}{n-2f}$	$\frac{n-f}{n-2f}\left(\frac{\sqrt{n-2f}}{\sqrt{n-f}}+\frac{\sqrt{2f}}{\sqrt{n-f}}+\frac{f}{n-f}\right)^2$
下界	$\frac{f}{n-2f}$	$\begin{cases}\frac{n-2}{n-2f+2} & \text{若 } n \text{ 为偶数}\\ \frac{n-1}{n-2f+1} & \text{若 } n \text{ 为奇数}\end{cases}$	$4\frac{f}{n-2f}$ (对于 $n > 3f$)

表1： 针对鲁棒性系数 $\kappa^\star$ （Krum与MultiKrum）的下界与上界。注意， [AFG+23] 并未计算 $\kappa^\star$，但正如我们在分析中所解释的，若一个聚合器是$(f,\ \kappa)$-鲁棒的，则 κ 是其鲁棒性系数的一个上界。

1.2. **背景与相关工作。攻击。**针对人工智能模型的攻击归根结底是在其参数空间中操纵模型更新，无论是通过梯度攻击——如在分布式学习设定中 [BEGS17], 还是通过数据投毒——后者进而诱发恶意梯度更新 [BUEM25]。尽管有些攻击依赖于直观上可理解的方向，并且已经能够奏效 [BUEM25], ，神经网络极高的维度与非线性使得识别恶意模型更新变得困难，从而为潜在攻击策略提供了广阔空间 [EGR18，BBG19，SH21，XKG20]。

防御方法/聚合规则。Krum [BEGS17] 提出了首个具有韧性保证的聚合规则。Krum 广受欢迎：它出现在许多研究中 [GBLH25、EMGR20、AGG+23、ZSS22a、CDR23、GMRBLH25b、Cam25、EM20、Rou22]。其他防御方法包括几何中位数 [Rou85], 、逐坐标中位数、逐坐标截尾均值。最初的 Krum 论文提到了 MultiKrum 这一扩展，但没有对其进行理论分析。

自 Krum 发表以来，多篇论文表明，实践者倾向于选择 MultiKrum 而非 Krum [SSR+25、HWF+24、TSMM25、ZSS22b、CDR23、GMRBLH25a、TAGA23、ZSWJ22、DEMG+19、LDC+19、Rou22]。更具体地说， [GMRBLH25a], 中的图12、 [KMJT25], 中的表I、 [TSMM25] 中的图2和表II， [以及最初 Krum 论文17] BEGS[2] 中的图2、4、5所给出的 MultiKrum 与 Krum 的实证比较均表明，MultiKrum 优于 Krum，既提高了鲁棒性，也提升了优化速度。

此外，还可以通过使用所谓的元聚合规则来提高鲁棒性，例如 Bulyan [EGR18], Median ofMeans [LL20、TLMC21], 分桶 [KHJ21], 或最近邻混合 [AFG+23]。

鲁棒性的概念。 [BEGS17] 提出了（α, f）-拜占庭韧性的概念，许多后续工作也在此基础上提出了改进，例如 [GMRBLH25a、EGR18、BMMB22、AFG+23、EMFG+21]。特别是， [AFG+23] 后来提出了（f, κ）-鲁棒性。他们证明了（f, κ）-鲁棒性蕴含（α, f）-拜占庭韧性，并且包括 Krum 但不包括 MultiKrum 在内的若干聚合规则，在“紧致”的分析下具有（f, κ）鲁棒性。[3] 该工作没有对 MultiKrum 进行分析，也没有提供任何理由来支持其所提出的 κ 确实是紧的。相比之下，本文澄清并形式化了最优鲁棒性系数的概念，并给出了 MultiKrum 的该系数的上界和下界。 [GDH25] 将（f, κ）-鲁棒性扩展到了去中心化环境。

Krum 鲁棒性的已知结果。

我们总结了 [AG23], 命题3和6见表1 ，并按照相同的鲁棒性概念给出我们的新结果。随后（第3节）我们将根据鲁棒性系数给出结果，这是我们改进后的鲁棒性概念。

[2]请注意，在 [ZSWJ22] 等若干研究中（图 4 和图 5），在测试鲁棒聚合规则时甚至没有提及 Krum，而是只考虑了 MultiKrum。[3]请参见 [AFG+23]中的表 1 标题。

2. 聚合器的鲁棒性

本节中，我们首先回顾MultiKrum聚合规则，然后介绍我们的鲁棒性概念。

记号 我们令 $[p] = \{1, \dots, p\}$ 表示 $p \in \mathbb{N}$，并令 $\mathcal{P}_m^n$ 表示$[n]$ 的所有大小为 m 的子集构成的集合。我们采用约定 $0/0 = -\infty$。我们令 $\mathcal{X} = \{x_1, \dots, x_n\}$ 表示待聚合的 n 个向量构成的集合，并令 f 表示拜占庭向量的数量。

2.1. **MultiKrum**。 我们令 $\mathcal{N}$: $\mathbb{R}^d \to \mathcal{P}_{n-f}^n$，使得 $\mathcal{N}(x)$ 表示 $\mathcal{X}$ 中距离 x 最近的 $n-f$ 个点的下标，若出现并列则任意打破。我们还令 $\mathcal{N}_i = \mathcal{N}(x_i)$ 表示 $i \in [n]$。给定 $\mathcal{X}$，我们令 $s^{\mathcal{X}}$: $\mathbb{R}^n \to \mathbb{R}$，使得 $s^{\mathcal{X}}(x) = (n-f)^{-1} \sum_{i \in \mathcal{N}(x)} \|x - x_i\|^2$；在没有歧义时，我们将 s 记作 $s^{\mathcal{X}}$。

定义 1. *m-MultiKrum* 聚合规则定义为得分最低的 m 个输入的平均值：

$$F_m(x_1, \dots, x_n) \triangleq \frac{1}{m} \sum_{i \in S_m^\star} x_i,$$

其中， $S_m^\star$ 收集了得分最低的 $[n]$ 的 m 个索引：

$$s(x_i) \leq s(x_j) \quad \text{for all } i \in S_m^\star, j \in [n] \setminus S_m^\star.$$

Krum 是 $m = 1$ 时的 m-MultiKrum。

备注 1. 这里，我们选择使 $\mathcal{N}(x_i)$ 包含点 x_i，归一化常数为$(n-f)^{-1}$。虽然这看起来可能与原始定义 [BEGS17]，（该定义排除了点 x_i 并使用归一化常数 $(n-f-1)^{-1}$）有所不同，但很容易验证这两个对象相同，并且系数 $(n-f)^{-1}$ 略微改善了最终界。

2.2. **聚合器的鲁棒性**。 现在我们定义鲁棒性概念。

定义 2. 设 $f \leq n/2$，并考虑一个聚合规则 $F : (\mathbb{R}^d)^n \to \mathbb{R}^d$。我们将 F 的鲁棒性系数 $\kappa^\star \in \mathbb{R} \cup \{+\infty\}$ 定义为

$$\kappa^\star(F) \triangleq \sup_{\substack{x_1, \dots, x_n \in \mathbb{R}^d \\ S \in \mathcal{P}_{n-f}^n}} \frac{\|F(x_1, \dots, x_n) - \bar{x}_S\|^2}{\frac{1}{|S|} \sum_{i \in S} \|x_i - \bar{x}_S\|^2},$$

其中 $\bar{x}_S = \frac{1}{|S|} \sum_{i \in S} x_i$。当 $\kappa^\star < \infty$ 时，我们称 F 是 f-鲁棒的。最后，将满足 $c \geq \kappa^\star$（分别为 $c \leq \kappa^\star$）的任意实数 c 称为鲁棒性系数的上界（分别为下界）。

有几点需要说明：

- 采用约定 $0/0 = -\infty$，可排除所有点在 S 中都相等的退化情形。这种情形意义不大：如果 $n-f$ 个点在 n 个候选值中相同，它们就构成多数，而任何合理的聚合规则都应输出这个共同值。

- 通过规定 [AFG+23] 满足 F 为$(f,\ \kappa)$-鲁棒的，只要$\kappa \geq \kappa^\star$。因此，分布式随机梯度下降的收敛保证 [AFG+23，定理 1] 成立，且最优鲁棒性系数为 $\kappa^\star$。

- 最后，计算聚合规则的鲁棒性系数，归根结底就是推导出相匹配的下界和上界。下界可通过在任意选定的 $x_1, \dots, x_n$和 S处计算目标函数得到。上界则需要构造目标函数的一致界，使其对所有 $x_1, \dots, x_n$和 S均成立，并且依赖于

聚合规则。特别是，如果一个聚合器是 (f,κ)-鲁棒的，那么 κ 就是其鲁棒性系数的一个上界。

与分布式机器学习中的许多观察结果类似，[AFG+23，命题6] 或者与较早的鲁棒均值估计文献中各种不同形式的观察结果类似，[Rou85，DH83，Hub11]：所有鲁棒性系数都高于某个通用值。我们的最优鲁棒性系数 $\kappa^\star$也具有这一性质。

命题1（通用下界）。考虑一个鲁棒聚合规则 F 即，满足 $\kappa^\star < \infty$ 的规则。那么，$n-2f>0$，并且有 $\kappa^\star \geq \frac{f}{n-2f}$.

3. The bounds on m-MultiKrum's $\kappa^\star_m$

本章节中，我们给出关于 m-MultiKrum 聚合规则鲁棒性的主要结果。全文中，我们用 $\kappa^\star_m$ 表示 m-MultiKrum 的鲁棒性系数。

3.1. 上界。 以下结果通过给出其鲁棒性系数的上界，表明 m-MultiKrum 是一种鲁棒聚合规则。对于较小的 m，MultiKrum 的鲁棒性系数上界 $\kappa^\star_m$ 为常数，随后则逐渐减小。

定理1. 考虑整数 f、n、m，满足 $n-2f>0$且 $0<m\leq n-f$。那么，*m-MultiKrum* 具有鲁棒性，其鲁棒性系数 $\kappa^\star_m$ 具有如下上界：

$$\kappa^\star_m \geq \frac{n-f}{n-2f}\min\left((\sqrt{2}+1)^2, \kappa_2^{>}(m)\right) \tag{1}$$

其中 $\kappa_2^{>}(m) = \left(\frac{\sqrt{n-2f}}{\sqrt{m}} + \frac{\sqrt{2f}}{\sqrt{m}} + \frac{f}{m}\right)^2$。

定理1的证明通过取定理5和6 中推导出的两个上界的较小值，将它们结合起来。接下来的结果刻画了过渡点 $m^\dagger$ 的行为，该过渡点分隔了定理1中的两种情形。

定理 2（过渡点的界）。设 $m^\dagger = m^\dagger(n,f) > 0$ 满足 $\kappa_2^{>}(m^\dagger) = (\sqrt{2}+1)^2$。则有

$$m^\dagger(n,f) \underset{f/n\to 0}{=} \frac{n}{(\sqrt{2}+1)^2}\left(1+o(1)\right)$$

当 $f \ll n$ 时，拜占庭工作节点所占比例可以忽略不计。在此情形下，过渡点满足 $m^\dagger \approx n/(\sqrt{2}+1)^2 \simeq 0.17$，$n$，因此 $\kappa_2^{<}(m)$ 情形开始起作用，并且当 m 适中且较大时，$\kappa(m)$ 会减小。这表明，当拜占庭工作节点的数量不是太少时，MultiKrum 可能比 Krum 更有优势。

证明定理 *2*。设 $C=(\sqrt{2}+1)^2$ 且

$$A = A(n,f) \coloneqq \sqrt{n-2f}+\sqrt{2f}.$$

根据 $m^\dagger$ 的定义，方程 $\kappa_2^{>}(m^\dagger) = C$ 在替换 $x = \sqrt{m^\dagger} > 0$ 后等价于

$$Cx^4 = \left(A+\frac{f}{x}\right)^2 = A^2 + 2A\frac{f}{x} + \frac{f^2}{x^2}.$$

两边乘以 x^2，得

$$Cx^4 = A^2x^2 + 2Afx + f^2.$$

首先，我们推导出 $m^\dagger$的下界。由于 $2Afx \geq 0$，从右侧去掉这个正项会使其减小，因此 $Cx^4 \geq A^2x^2 + f^2$. 由于 $x^2 = m^\dagger$，该不等式变为

$$C(m^\dagger)^2 - A^2m^\dagger - f^2 \geq 0,$$

其解满足 $m^\dagger \geq \frac{A^2+\sqrt{A^4+4Cf^2}}{2C}$. 因此，

$$m^\dagger = x^2 \geq \frac{A^2 + \sqrt{A^4 + 4(\sqrt{2}+1)^2 f^2}}{2(\sqrt{2}+1)^2}.$$

其次，我们推导出关于 $m^\dagger$ 的上界。根据算术-几何平均不等式，$2Afx \leq A^2x^2 + f^2$，这意味着 $A^2x^2 + 2Afx + f^2 \leq 2A^2x^2 + 2f^2$.。代入定义方程可得 $Cx^4 \leq 2A^2x^2 + 2f^2$.。同样，我们有

$$C(m^\dagger)^2 - 2A^2m^\dagger - 2f^2 \leq 0,$$

这意味着 $m^\dagger \leq \frac{2A^2+\sqrt{4A^4+8Cf^2}}{2C} = \frac{A^2+\sqrt{A^4+2Cf^2}}{C}$. 因此，

$$m^\dagger = x^2 \leq \frac{A^2 + \sqrt{A^4 + 2(\sqrt{2}+1)^2 f^2}}{(\sqrt{2}+1)^2}.$$

最后，我们推导出 时的渐近行为 $f \ll n$。注意，对于 $f \ll n$，我们有

$$A = \sqrt{n-2f} + \sqrt{2f} = \sqrt{n}(1 + o(1)),$$

因此 $A^2 = n(1 + o(1))$，且 $\sqrt{A^4 + 4Cf^2} = A^2(1 + o(1)) = n(1 + o(1))$，上界中的项也同理。因此，两个界都化简为

$$m^\dagger = \frac{n}{(\sqrt{2}+1)^2}(1 + o(1)),$$

证明至此完成。 ■

3.2. **两个下界**。 我们强调，定理1 给出的是鲁棒性系数 κ_m 的上界，而非该系数本身。在本章节中，我们通过推导 Krum 以及 m-MultiKrum（其中 $m = n - f$）的下界来补充这一结果。后一种设定在所有 m-MultiKrum 聚合器中给出最小的上界（见（1）），因此是最优鲁棒性的最强候选。

下面我们给出 Krum 的鲁棒性系数 $\kappa_1^\star$ 的一个下界。该界相对于通用界 $\frac{f}{n-2f}$ （命题1）呈现出互补的性质。

定理3 （Krum 的下界）。 成立

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

定理的证明 定理3。 令 $e = e_1 \in \mathbb{R}^d$ 表示第一个标准基向量。考虑 n 为偶数，并定义 $\mathcal{X} = (x_i)_{i=1}^n$ 使得 $x_1 = \cdots = x_{n/2-1} = e$，且 $x_{n/2} = \cdots = x_n = 0$。此外，令 $S = \{1, \ldots, n-f\}$。于是容易验证 $\bar{x}_S = \frac{n-2}{2}(n-f)$，且 $\Sigma_S = \frac{(n-2)(n-2f+2)}{4(n-f)^2}$。此外，评分函数表示为 $s(0) = \frac{n/2-f-1}{n-f}$，以及 $s(e) = \frac{n/2-f+1}{n-f}$，因此 Krum聚合为 $x_1^\star = 0$。由此，经过化简可得

$$\kappa_1^\star \geq \frac{\|x_1^\star - \bar{x}_S\|^2}{\Sigma_S} = \frac{n-2}{n-2f+2}.$$

现在考虑 n 为奇数，这次取 $\mathcal{X} = (x_i)_{i=1}^n$ 使得 $x_1 = \cdots = x_{(n-1)/2} = e$，且 $x_{(n+1)/2} = \cdots = x_n = 0$。再次令 $S = \{1, \ldots, n-f\}$，可得 $\bar{x}_S = \frac{n-2f-1}{n-1}(n-f)e$，且 $\Sigma_S = \frac{(n-1)(n-2f+1)}{(n-2f+1)^2}4(n-f)^2$。评分函数表示为 $s(0) = \frac{n-2f-1}{2}(n-f)$，以及 $s(1) = \frac{n-2f+1}{2}(n-f)$，因此Krum聚合为 $x_1^\star = 0$。由此，经过化简可得

$$\kappa_1^\star \geq \frac{\|x_1^\star - \bar{x}_S\|^2}{\Sigma_S} = \frac{n-1}{n-2f+1}$$

■

定理4 ($(n-f)$-MultiKrum 下界)。 设 $f, n \in \mathbb{N}$ 满足 $n > 3f$。那么，$(n-f)$-*MultiKrum* 聚合器的最优鲁棒性系数 $\kappa_{n-f}^\star$ 满足

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

条件 $n > 3f$ 源于我们证明技术的局限性。当 $2f < n \leq 3f$时，通用下界（命题1）已适用于所有聚合规则，而我们的构造无法为 $(n-f)$-MultiKrum 得出更强的界。因此，我们将重点放在 $n > 3f$这一范围内，此时相同的论证可得出一个严格更优、且专用于 MultiKrum 的下界。

我们在 $\kappa_m^\dagger$ 的所有 $1 \leq m \leq n-f$ 取值下的闭式下界见 附录B。

定理4的证明。 令 $e = e_1 \in \mathbb{R}^d$ 为第一个标准基向量，并考虑$\mathbf{x} = (x_i)_{i=1}^n$，其定义为

$$\begin{cases} f \text{ vectors} & = -e, \\ n-2f \text{ vectors} & = 0, \\ f \text{ vectors} & = (1-\varepsilon)e, \end{cases}$$

其中 $\varepsilon > 0$ 任意小。

对于该配置，MultiKrum分数满足：对于 $1 \leq i \leq f$、 $f+1 \leq j \leq n-f$和 $n-f+1 \leq k \leq n$,

$$s(x_i) = \frac{n-2f}{n-f}, \quad s(x_j) = \frac{f}{n-f}, \quad s(x_k) = (1-\varepsilon)\frac{n-2f}{n-f}.$$

由于 $n > 3f$，我们有 $s(x_j) < s(x_k) < s(x_i)$，因此最小的（$n-f$）个分数对应于 $S_{n-f}^\star = \{f+1, \ldots, n\}$，即所有零向量以及所有（$1-\varepsilon$）$e$ 向量。因此，

$$\bar{x}_{S_{n-f}^\star} = (1-\varepsilon)\frac{f}{n-f}e.$$

我们选取参考集 $S = \{1, \ldots, n-f\}$，其经验均值为 $\bar{x}_S = -\frac{f}{n-f}e$。因此平方偏差为

$$\|\bar{x}_{S_{n-f}^\star} - \bar{x}_S\|^2 = \left((2-\varepsilon)\frac{f}{n-f}\right)^2.$$

此外， S 的经验方差为

$$\Sigma_S = \frac{1}{n-f}\sum_{i \in S}\|x_i - \bar{x}_S\|^2 = \frac{f(n-2f)}{(n-f)^2}.$$

令 $\varepsilon \to 0$ 可得 $\frac{\|\bar{x}_{S_{n-f}^\star} - \bar{x}_S\|^2}{\Sigma_S} = 4\frac{f}{n-2f}$，证明完毕。

■

4. Proofs

在本章节中，我们给出MultiKrum的鲁棒性系数上界的证明，定理1。

4.1. 均值-方差关系。 对于任意大小为 $|A| = a$的集合 $A \subset [n]$ ，我们设定

$$\bar{x}_A = \frac{1}{a}\sum_{i\in A} x_i \quad \text{and} \quad \Sigma_A = \frac{1}{a}\sum_{i\in A}\|x_i - \bar{x}_A\|^2.$$

引理1. 设 A、 $B \subset [n]$ 的大小分别为 a、 b。对于任意 A、 B,

$$\frac{1}{ab}\sum_{i\in A}\sum_{j\in B}\|x_i - x_j\|^2 \;=\; \|\bar{x}_A - \bar{x}_B\|^2 + \Sigma_A + \Sigma_B. \tag{2}$$

对于 $A = B$, 因此我们有

$$\Sigma_A \;=\; \frac{1}{2a^2}\sum_{i,j\in A}\|x_i - x_j\|^2.$$

证明. 固定 $i \in A$, $j \in B$，并考虑如下分解

$$x_i - x_j = (x_i - \bar{x}_A) + (\bar{x}_A - \bar{x}_B) + (\bar{x}_B - x_j).$$

将内积平方并展开，得到

$$\begin{aligned}\|x_i - x_j\|^2 = \|x_i - \bar{x}_A\|^2 + \|\bar{x}_A - \bar{x}_B\|^2 + \|\bar{x}_B - x_j\|^2 \\ + 2\langle x_i - \bar{x}_A, \bar{x}_A - \bar{x}_B\rangle + 2\langle \bar{x}_A - \bar{x}_B, \bar{x}_B - x_j\rangle \\ + 2\langle x_i - \bar{x}_A, \bar{x}_B - x_j\rangle.\end{aligned}$$

对 $i \in A$ 和 $j \in B$取平均时，混合内积项会消失，因为$\sum_{i\in A}(x_i - \bar{x}_A) = 0 = \sum_{j\in B}($ $x_j - \bar{x}_B)$。因此

$$\begin{aligned}\frac{1}{ab}\sum_{i\in A}\sum_{j\in B}\|x_i - x_j\|^2 = \frac{1}{a}\sum_{i\in A}\|x_i - \bar{x}_A\|^2 \\ + \frac{1}{b}\sum_{j\in B}\|x_j - \bar{x}_B\|^2 + \|\bar{x}_A - \bar{x}_B\|^2,\end{aligned}$$

这恰好是（2）。 ■

4.2. 两个关键引理。 在本章节中，我们给出两个用于分析 MultiKrum 鲁棒性的关键引理。

引理2。 考虑整数 f、 n 和 m，满足 $n - 2f > 0$、 $0 < m \le n - f$，以及由 $\mathcal{X} = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$ 个 n 向量构成的集合，以及 $[n]$ 的一个大小为 $|S| = n - f$ 的子集 S。令 $S^\star_m \in \mathcal{P}^n_m$ 表示使 $s(x.)$ 最小的 $[n]$ 的索引。则

$$\frac{1}{m}\sum_{i\in S^\star_m} s(x_i) \le 2\,\Sigma_S. \tag{3}$$

证明。 根据定义， $S^\star_m$ 是使 s (x_i) 最小的 m 个索引。由于最小的 m 个元素的平均值不大于 S 中所有元素的平均值，因此我们有

$$\frac{1}{m}\sum_{i\in S^\star_m} s(x_i) \le \frac{1}{|S|}\sum_{i\in S} s(x_i). \tag{4}$$

接下来注意，每个 $\mathcal{N}(x_i)$ 都包含 $\mathcal{X}$ 中 x_i 的 $n - f$个最近邻。因此，对于每个$i \in S$，对 $\mathcal{N}(x_i)$ 求平均总是小于对更大的集合 S 求平均:

$$s(x_i) = \frac{1}{|\mathcal{N}(x_i)|}\sum_{j\in\mathcal{N}(x_i)}\|x_j - x_i\|^2 \le \frac{1}{|S|}\sum_{j\in S}\|x_j - x_i\|^2.$$

对前一个公式在 $i \in S$ 上取平均，然后利用 引理1，可得

$$\frac{1}{|S|}\sum_{i\in S} s_i \le \frac{1}{|S|^2}\sum_{i,j\in S}\|x_j - x_i\|^2 = 2\Sigma_S. \tag{5}$$

结合公式 (4)和(5) 即可完成证明。 ■

引理 3。设两个整数为 f, n，满足 $n-2f>0$；有一组 $\mathcal{X}=(x_1,\dots,x_n)\in(\mathbb{R}^d)^n$ 个 n 向量，以及一个大小为 $|S|=n-f$ 的 $[n]$ 子集 S。那么，对于任意$x\in\mathbb{R}^d$ 以及任意 $\alpha>0$

$$\|x-\bar{x}_S\|^2 \;\le\; \frac{n-f}{n-2f}\left((1+\alpha)\,s(x) \;+\; \left(1+\frac{1}{\alpha}\right)\Sigma_S\right). \tag{6}$$

证明。对于 $x\in\mathbb{R}^d$，定义 $U(x) := S\cap\mathcal{N}(x)$，其中 $\mathcal{N}(x)$ 是 $\mathcal{X}$ 中距离 x 最近的 $n-f$ 个点构成的集合。由于 $|S|=|\mathcal{N}(x)|=n-f$，且 $|S\cap\mathcal{N}(x)|\le n$，我们有

$$|U(x)| = |S\cup\mathcal{N}(x)| = |S|+|\mathcal{N}(x)|-|S\cap\mathcal{N}(x)| \ge n-2f.$$

我们将其分解为

$$\|x-\bar{x}_S\|^2 = \frac{1}{|U(x)|}\sum_{j\in U(x)}\|x-x_j+x_j-\bar{x}_S\|^2.$$

应用杨氏不等式（A.1），取参数 $\alpha>0$ 和点 $x-x_j$ 及 $x_j-\bar{x}_S$，可得

$$\begin{aligned}\|x-\bar{x}_S\|^2 \le (1+\alpha)\frac{1}{|U(x)|}\sum_{j\in U(x)}\|x-x_j\|^2 \\ + \left(1+\frac{1}{\alpha}\right)\frac{1}{|U(x)|}\sum_{j\in U(x)}\|x_j-\bar{x}_S\|^2.\end{aligned} \tag{7}$$

第一项利用 $U(x)\subset\mathcal{N}(x)$ 加以控制：

$$\begin{aligned}\frac{1}{|U(x)|}\sum_{j\in U(x)}\|x-x_j\|^2 &\le \frac{1}{|U(x)|}\sum_{j\in\mathcal{N}(x)}\|x-x_j\|^2 \\ &= \frac{n-f}{|U(x)|}s(x) \le \frac{n-f}{n-2f}s(x).\end{aligned}$$

第二项利用 $U(x)\subset S$ 加以控制：

$$\begin{aligned}\frac{1}{|U(x)|}\sum_{j\in U(x)}\|x_j-\bar{x}_S\|^2 &\le \frac{1}{|U(x)|}\sum_{j\in S}\|x_j-\bar{x}_S\|^2 \\ &= \frac{n-f}{|U(x)|}\Sigma_S \le \frac{n-f}{n-2f}\Sigma_S.\end{aligned}$$

该结果可通过将两项的界代入 式 (7)。 ■

4.3. **MultiKrum 的 $\kappa^\star$ 的两个上界**。 在本章节中，我们证明 m-MultiKrum 的鲁棒性系数的两个上界。 定理5 给出第一个一致上界； 定理 6 给出第二个递减上界。最后，我们给出 定理1 的证明。

定理5. 考虑整数 f、 n 和 m，使得 $n-2f>0$、 $0<m\le n-f$，一个集合$\mathcal{X}=(x_1,\dots,x_n)\in(\mathbb{R}^d)^n$（由 n 个向量组成），以及子集 S（属于 $[n]$，大小为 $|S|=n-f$）。则 m-*MultiKrum* 的鲁棒性系数具有如下上界

$$\kappa_m^\star \le \kappa_1 \triangleq (\sqrt{2}+1)^2\frac{n-f}{n-2f}.$$

证明。 我们首先注意到，根据詹森不等式，

$$\|\bar{x}_{S^\star_m} - \bar{x}_S\|^2 = \left\|\frac{1}{m}\sum_{i\in S^\star_m}(x_i - \bar{x}_S)\right\|^2 \overset{(A.2)}{\leq} \frac{1}{m}\sum_{i\in S^\star_m}\|x_i - \bar{x}_S\|^2$$

对 引理 3的界取平均，对于 i 属于 $S^\star_m$，再应用引理2，可得对任意$\alpha > 0$

$$\begin{aligned}
&\frac{1}{m}\sum_{i\in S^\star_m}\|x_i - \bar{x}_S\|^2 \\
&\overset{(6)}{\leq} \frac{n-f}{n-2f}\left((1+\alpha)\frac{1}{m}\sum_{i\in S^\star_m} s(x_i) \;+\; \left(1+\frac{1}{\alpha}\right)\Sigma_S\right) \\
&\overset{(3)}{\leq} \frac{n-f}{n-2f}\left(2(1+\alpha) + \left(1+\frac{1}{\alpha}\right)\right)\Sigma_S
\end{aligned}$$

鉴于 $\min_{\alpha>0} 2(1+\alpha) + \left(1+\frac{1}{\alpha}\right) = (\sqrt{2}+1)^2$，，有

$$\|\bar{x}_{S^\star_m} - \bar{x}_S\|^2 \leq \frac{n-f}{n-2f}\left(\sqrt{2}+1\right)^2\Sigma_S,$$

由此得到 κ 的所述界。 ■

定理 6。 设整数 f、 n和 m满足 $n-2f>0$、 $0<m\leq n-f$，有一组$\mathcal{X}=(x_1,\ldots,x_n)\in(\mathbb{R}^d)^n$ 个 n 向量，以及一个大小为 $|S|=n-f$ 的 $[n]$ 子集 S。那么， m-*MultiKrum* 的鲁棒性系数满足以下上界

$$\kappa^\star_m \leq \kappa_2(m) = \frac{n-f}{n-2f}\begin{cases}\kappa_2^{\leq}(m) & \text{if } m\leq f\\ \kappa_2^{>}(m) & \text{else}\end{cases}$$

with

$$\kappa_2^{\leq}(m) = \left(\frac{\sqrt{n-2f}}{\sqrt{m}} + \sqrt{2} + 1\right)^2,$$
$$\kappa_2^{>}(m) = \left(\frac{\sqrt{n-2f}}{\sqrt{m}} + \frac{\sqrt{2f}}{\sqrt{m}} + \frac{f}{m}\right)^2.$$

证明。 令 $S^\star_m$ 表示由 MultiKrum 选出的 m 个索引的子集，并记

$$u := |S^\star_m \setminus S|, \qquad v := |S^\star_m \cap S|, \qquad u+v=m.$$

必然地， $0\leq u\leq$最小值（m， f）以及最大值（0， $m-f$） $\leq v\leq m$。对于任意 $\alpha_1>0$，引理 A.1给出了分解

$$\begin{aligned}
\|\bar{x}_{S^\star_m} - \bar{x}_S\|^2 &= \left\|\frac{1}{m}\sum_{i\in S^\star_m}(x_i - \bar{x}_S)\right\|^2 \\
&\overset{(A.1)}{\leq} (1+\alpha_1)\,A + \left(1+\frac{1}{\alpha_1}\right)B, \qquad (8)
\end{aligned}$$

其中 $A = \left\|\frac{1}{m}\sum_{i\in S^\star_m\cap S}(x_i-\bar{x}_S)\right\|^2$ 和 $B = \left\|\frac{1}{m}\sum_{i\in S^\star_m\setminus S}(x_i-\bar{x}_S)\right\|^2$. 为了界定 A，我们先应用詹森不等式，然后利用 $S^\star_m\cap S\subseteq S$:

$$A = \frac{v^2}{m^2}\left\|\frac{1}{v}\sum_{i\in S^\star_m\cap S}(x_i-\bar{x}_S)\right\|^2 \overset{(A.2)}{\leq} \frac{v}{m^2}\sum_{i\in S^\star_m\cap S}\|x_i-\bar{x}_S\|^2$$

$$\leq \frac{v}{m^2}\sum_{i\in S}\|x_i-\bar{x}_S\|^2 = \frac{v(n-f)}{m^2}\,\Sigma_S.$$

为了界定 B，我们先应用詹森不等式，然后利用引理 3：对任意 $\alpha_2>0$

$$\begin{aligned}
B &= \frac{u^2}{m^2}\Big\|\frac{1}{u}\sum_{i\in S_m^\star\setminus S}(x_i-\bar{x}_S)\Big\|^2\\
&\overset{(A.2)}{\leq} \frac{u}{m^2}\sum_{i\in S_m^\star\setminus S}\|x_i-\bar{x}_S\|^2\\
&\overset{(6)}{\leq} \frac{u}{m^2}\sum_{i\in S_m^\star\setminus S}\frac{n-f}{n-2f}\left((1+\alpha_2)\,s(x_i)+\left(1+\frac{1}{\alpha_2}\right)\Sigma_S\right)\\
&= \frac{u}{m^2}\frac{n-f}{n-2f}\left((1+\alpha_2)\sum_{i\in S_m^\star\setminus S}s(x_i)+\left(1+\frac{1}{\alpha_2}\right)u\Sigma_S\right).
\end{aligned}$$

MultiKrum选择挑选局部得分较小的索引：引理2 意味着 $\sum_{i\in S_m^\star}s(x_i)\leq m2\Sigma_S$，因此

$$\sum_{i\in S_m^\star\setminus S}s(x_i)\leq\sum_{i\in S_m^\star}s(x_i)\leq 2m\Sigma_S.$$

将此界代入前面的表达式可得

$$B\leq\frac{u}{m^2}\frac{n-f}{n-2f}\left((1+\alpha_2)\,2m+\left(1+\frac{1}{\alpha_2}\right)u\right)\Sigma_S.$$

对 $\alpha_2>0$ 求最小值可得显式极小值点 $\alpha_2^\star=\sqrt{u/(2m)}$，最小值为$(\sqrt{2m}+\sqrt{u})^2$。因此

$$B\leq\frac{u}{m^2}\frac{n-f}{n-2f}(\sqrt{2m}+\sqrt{u})^2\,\Sigma_S.$$

将 A 和 B 的界代入公式（8），并对 $\alpha_1>0$ 进行优化（最优选择为 $\alpha_1=\sqrt{B/A}$，其最小值为 $(\sqrt{A}+\sqrt{B})^2$），可得逐实例系数

$$\kappa_2(m,u)\triangleq\frac{n-f}{m^2}\left(\sqrt{v}+(\sqrt{2m}+\sqrt{u})\sqrt{\frac{u}{n-2f}}\right)^2,$$

使得$\|\bar{x}_{S_m^\star}-\bar{x}_S\|^2\leq\kappa_2(m,\ u)\,\Sigma_S$。

利用 $v\leq m$ 和 $u\leq u_{\max}\triangleq\min(m,f)$，我们得到上界 $\kappa_2(m)$：

$$\begin{aligned}
\kappa_2(m,u)&\leq\kappa_2(m)\\
&\triangleq\frac{n-f}{m^2}\left(\sqrt{m}+\left(\sqrt{2m}+\sqrt{u_{\max}}\right)\sqrt{\frac{u_{\max}}{n-2f}}\right)^2\\
&=\frac{n-f}{m}\left(1+\left(\sqrt{2}+\sqrt{\frac{u_{\max}}{m}}\right)\sqrt{\frac{u_{\max}}{n-2f}}\right)^2\\
&=\frac{1}{m}\frac{n-f}{n-2f}\left(\sqrt{n-2f}+\sqrt{2}\sqrt{u_{\max}}+\frac{u_{\max}}{\sqrt{m}}\right)^2.
\end{aligned}$$

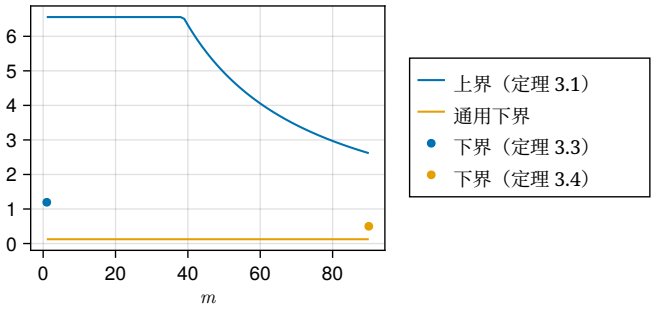

图 1： $\kappa_m^\star$ 的上界和下界示意图，其中 $n = 100$、 $f = 10$ 和 m 在 1 到 $n - f = 90$ 之间变化。

最后请注意，当 $m \leq f$ 时， $u_{\max} = m$，因此 $\kappa_2(m)$ 可简化为

$$\kappa_2^<(m) \triangleq \frac{n-f}{n-2f}\left(\frac{\sqrt{n-2f}}{\sqrt{m}} + \sqrt{2} + 1\right)^2.$$

当 $m \geq f$, $u_{\max} = f$ 时，使得 $\kappa_2(m)$ 化简为

$$\kappa_2^>(m) \triangleq \frac{1}{m}\frac{n-f}{n-2f}\left(\sqrt{n-2f} + \sqrt{2f} + \frac{f}{\sqrt{m}}\right)^2. \qquad \blacksquare$$

现在我们可以证明定理1。

定理*定理1*的证明。 定理5 和6 给出了 κ_1 和 $\kappa_2(m)$ 关于 m-MulitKrum 的最优鲁棒性系数的两个上界。因此，该系数 $\kappa_m^\star$ 受其最小值限制：

$$\begin{aligned}\kappa_m^\star &\leq \min(\kappa_1, \kappa_2(m)) \\ &= \min\left(\kappa_1, \frac{n-f}{n-2f}\begin{cases}\kappa_2^<(m) & \text{if } m \leq f \\ \kappa_2^>(m) & \text{else}\end{cases}\right).\end{aligned}$$

首先注意，对于任意 $m \in [n-f]$,

$$\kappa_1 = (\sqrt{2}+1)^2 < \left(\frac{\sqrt{n-2f}}{\sqrt{m}} + \sqrt{2} + 1\right)^2 = \kappa_2^<(m).$$

接着注意到， $\kappa_2^>$是一个递减函数，因此 $\kappa_2^>(f) = \left(\frac{\sqrt{n-2f}}{\sqrt{f}} + \sqrt{2} + 1\right)^2 \geq (\sqrt{2}+1)^2 = \kappa_1$。

所以，对于 $m \in [f]$， $\kappa_1 < \kappa_2^>(m)$。结论由此得出。 $\blacksquare$

5. 实验示例

我们在图1中报告了 $\kappa_m^\star$ 在第3节中得到的上界和下界的演变： 定理1 给出了对所有 m 均成立的上界，而定理3 和4 给出了上界 $m = 1$ 和 $m = n - f$。

我们在图2中报告了对 $m^\dagger(n, f)$ 的数值研究，即定理3的上界开始下降的第一个值 m。我们使用二分搜索对 $m^\dagger$ 进行数值计算，并展示归一化值 $m^\dagger(n, f)/n$。我们考虑三组 (n, f)，每组的对手比例 f/n 分别为 $\{10\%、1\%、0.1\%\}$。三条相应曲线较为平坦，其值随该比例单调递减。这表明量 $m^\dagger$,

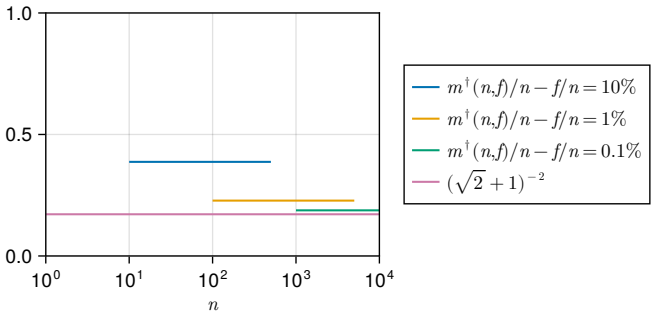

图 2： 对 $m^\dagger(n,f)/n$的示意图，其中 $m^\dagger(n,f)$ 表示满足以下条件的最小 m：$(\sqrt{2}+1)^2 > \kappa_2^>(m)$，以及 $(1+\sqrt{2})^{-2}$ 给出的界，该界由定理 2提供。

仅取决于对手的比例 f/n，而非同时取决于两个变量 f 和 n。这些值似乎收敛于 $(\sqrt{2}+1)^{-2}$ 当 $f/n \to 0$，正如 定理 2。

6. 总结

在这项工作中，我们研究了存在对手时均值估计聚合规则的鲁棒性。我们引入了鲁棒性系数 $\kappa^\star$，这是一种最优度量，可细化现有的鲁棒性概念。

我们通过推导 MultiKrum 鲁棒性系数的上界和下界，首次为 MultiKrum 给出了鲁棒性保证；此外，我们还改进了 Krum 目前已知的最佳界。我们的分析表明，上界会随着聚合参数m的增大而下降，这与经验观察一致：MultiKrum 的鲁棒性更强，因此在实践中比 Krum 更受青睐。

最后，我们提出一个问题：如何计算或精确界定其他经典聚合规则的鲁棒性系数$\kappa^\star$，例如几何中位数、截尾均值和逐坐标中位数。

附录 A. 初等关系回顾

我们回顾两个初等引理：一个是杨氏不等式的一个已知推论，另一个是关于平方范数的詹森不等式。

引理 A.1. 对于任意 x、$y \in \mathbb{R}^n$ 和 $\alpha > 0$，均有

$$\|x+y\|^2 \le (1+\alpha)\|x\|^2 + \left(1+\frac{1}{\alpha}\right)\|y\|^2. \tag{A.1}$$

证明. 杨氏不等式保证，对于任意 x、$y \in \mathbb{R}^n$ 和 $\alpha > 0$，均有$2\langle x$、$y\rangle \le \alpha\|x\|^2 + \frac{1}{\alpha}\|y\|^2$。回顾一下，这可直接由恒等式$\|\sqrt{\alpha}x - \frac{1}{\sqrt{\alpha}}y\|^2 \ge 0$ 展开平方得到。所述界可通过展开平方得出：

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + 2\langle x,y\rangle + \|y\|^2 \\ &\le (1+\alpha)\|x\|^2 + \left(1+\frac{1}{\alpha}\right)\|y\|^2.\end{aligned}$$

■

引理 A.2（詹森不等式）。设 $\mathbb{R}^n$ 中有 p 个点 x_1、……、x_p，以及 p 个正实数 λ_1、……、λ_p。则，

$$\left\|\frac{\sum_i \lambda_i x_i}{\sum_i \lambda_i}\right\|^2 \le \frac{\sum_i \lambda_i \|x_i\|^2}{\sum_i \lambda_i} \tag{A.2}$$

证明。 对平方范数应用权重 $\lambda_i / \sum_j \lambda_j$，由其凸性可得。 ■

附录B. 任意 m 下 m-MultiKrum 的下界

命题B.1 ($\kappa_m^\star$ 的 m-MultiKrum 聚合器的下界)。 设整数f、 n、 m满足 $n-3f>0$和 $0<m\le n-f$。则，鲁棒性参数 $\kappa_m^\star$满足

$$\kappa_m^\star \ge R(n,f,m),$$

其中，令

$$a=a(n,f,m)=\min(m,n-2f)$$

我们有

$$R(n,f,m)\triangleq \frac{(n-f)^2}{f(n-2f)}\left(\frac{a\cdot\frac{f}{n-f}+m-a}{m}\right)^2.$$

证明。 令 $e=e_1\in\mathbb{R}^d$ 为第一个标准基向量。定义 $\mathcal{X}=(x_i)_{i=1}^n$ 如下

$$\begin{cases} f \text{ vectors} & = -e,\\ n-2f \text{ vectors} & = 0,\\ f \text{ vectors} & = (1-\varepsilon)e, \end{cases}$$

且 $\varepsilon>0$可任意小。MultiKrum分数满足：对于 $1\le i\le f$、 $f+1\le j\le n-f$ 和 $n-f+1\le k\le n$

$$s(i)=\frac{n-2f}{n-f},\quad s(j)=\frac{f}{n-f},\quad s(k)=(1-\epsilon)\frac{n-2f}{n-f}.$$

When $(n-2f)\ge f$， m-MultiKrum 首先选出 a 个零向量，然后选出 $(m-a)$ 个向量 $(1+\epsilon)e$.

$$S_m^\star=\underbrace{0,\dots,0}_{a},\underbrace{(1-\varepsilon)e,\dots,(1-\varepsilon)e}_{m-a}\}.$$

考虑参考集

$$S=\{\underbrace{-e,\dots,-e}_{f},\underbrace{0,\dots,0}_{n-2f}\},\quad |S|=n-f.$$

它们的重心满足

$$\bar{x}_{S_m^\star}=\frac{m-a}{m}(1-\varepsilon)e,\quad \bar{x}_S=-\frac{f}{n-f}e.$$

平方偏差为

$$\|\bar{x}_{S_m^\star}-\bar{x}_S\|^2=\left(\frac{a\frac{f}{n-f}+(m-a)(1-\varepsilon)}{m}\right)^2.$$

S 的方差为

$$\Sigma_S=\frac{f(n-2f)}{(n-f)^2}.$$

令 $\varepsilon\to 0$，我们得到

$$R(n,f,m)=\frac{\|\bar{x}_{S_m^\star}-\bar{x}_S\|^2}{\Sigma_S}=\frac{(n-f)^2}{f(n-2f)}\left(\frac{a\frac{f}{n-f}+m-a}{m}\right)^2,$$

由此完成证明。 ■

参考文献

[AAZL18] Dan Alistarh、Zeyuan Allen-Zhu 和 Jerry Li，《拜占庭随机梯度下降》，神经信息处理系统，2018 年，第 4613–4623 页。[AFG+23] Youssef Allouah、Sadegh Farhadkhani、Rachid Guerraoui、Nirupam Gupta、Rafaël Pinot 和 John Stephan，《通过混合修复：异质性条件下实现最优拜占庭机器学习的方法》，人工智能与统计国际会议，PMLR，2023 年，第 1232–1300 页。[AGG+23] Youssef Allouah、Rachid Guerraoui、Nirupam Gupta、Rafael Pinot 和 Geovani Rizk，《鲁棒分布式学习：数据异质性下的紧致误差界与崩溃点》，神经信息处理系统进展 36（2023），45744–45776。[BBG19] Moran Baruch、Gilad Baruch 和 Yoav Goldberg，《少许即足够：绕过分布式学习的防御方法》，第 33 届神经信息处理系统国际会议论文集，第 775 号，Curran Associates Inc.，美国纽约州雷德胡克，2019 年 12 月，第 8635–8645 页。[BEGS17] Peva Blanchard、El-Mahdi El-Mhamdi、Rachid Guerraoui 和 Julien Stainer，《对抗者环境下的机器学习：拜占庭容错梯度下降》，神经信息处理系统，2017 年，第 118–128 页。[BMMB22]Djamila Bouhata、Hamouma Moumen、Jocelyn-Ahmed Mazari 和 Ahcène Bounceur，《分布式机器学习中的拜占庭容错：综述》，《实验与理论人工智能杂志》0（2022），第 0 期，1–59。[BUEM25] Wassim Bouaziz、Nicolas Usunier 和 El-Mahdi El-Mhamdi，《反转梯度攻击可实现强效数据投毒》，《机器学习研究汇刊》（2025），0。[BVUEM26] Wassim Bouaziz、Mathurin Videau、Nicolas Usunier 和 El-Mahdi El-Mhamdi，《冬日战士：通过间接数据投毒，在预训练阶段对语言模型植入后门》，国际学习表征会议，2026 年。[Cam25] Mélanie Cambus，《动态故障网络中的分布式计算：分布式图算法与多维一致性》，博士论文，阿尔托大学，2025 年。[CDR23] Francesco Colosimo 和 Floriano De Rango，《Median-krum：一种用于联邦学习、基于距离与统计的联合拜占庭鲁棒算法》，国际 ACM 移动管理与无线接入研讨会论文集（美国纽约州纽约），MobiWac '23，计算机协会，2023 年，第 61–68 页。[CGD+23] Antonio Emanuele Cinà、Kathrin Grosse、Ambra Demontis、Sebastiano Vascon、Werner Zellinger、Bernhard A Moser、Alina Oprea、Battista Biggio、Marcello Pelillo 和 Fabio Roli，《重探恶意模式：针对训练数据投毒的机器学习安全综述》，ACM计算综论 55（2023），第 13s 期，1–39。[DEMG+19] Georgios Damaskinos、El-Mahdi El-Mhamdi、Rachid Guerraoui、Arsany Guirguis 和 Sébastien Rouault，《Aggregathor：通过鲁棒梯度聚合实现拜占庭机器学习》，《机器学习与系统论文集》1（2019），81–106。[DH83] David L Donoho 和 Peter J Huber，《崩溃点的概念》，《献给 Erich L. Lehmann 的纪念文集》157184（1983），0。[EGR18] El-Mahdi El-Mhamdi、Rachid Guerraoui 和 Sébastien Rouault，《拜占庭环境中分布式学习的隐蔽漏洞》，国际机器学习会议，第 80 卷，2018 年，第 3521–3530 页。[EM20] El-Mahdi El-Mhamdi，《鲁棒分布式学习》，博士论文，洛桑联邦理工学院，2020 年。[EMFG+21] El Mahdi El-Mhamdi、Sadegh Farhadkhani、Rachid Guerraoui、Arsany Guirguis、Lê-Nguyên Hoang 和 Sébastien Rouault，《丛林中的协作学习（去中心化、拜占庭、异质、异步与非凸学习）》，神经信息处理系统进展 34（2021），25044–25057。

[EMGR20] El-Mahdi El-Mhamdi、Rachid Guerraoui 和 Sébastien Rouault，高维快速鲁棒分布式学习，2020 年国际可靠分布式系统研讨会（SRDS），IEEE，2020 年，第 71–80 页。[FGV+22] Sadegh Farhadkhani、Rachid Guerraoui、Oscar Villemaud 等，数据投毒与拜占庭梯度攻击之间的等价性，国际机器学习会议，PMLR，2022 年，第 6284–6323 页。[GBLH25] Mario García-Márquez、Nuria Rodríguez Barroso、María Victoria Luzón 和 Francisco Herrera，Krum 联邦链（KFC）：利用区块链防御联邦学习中的对抗攻击，CoRR abs/2502.06917（2025），0。[GDH25] Renaud Gaucher、Aymeric Dieuleveut 和 Hadrien Hendrikx，拜占庭鲁棒 Gossip 的统一崩溃分析，第四十二届国际机器学习会议，2025 年 6 月。[GMRBLH25a] M. García-Márquez、N. Rodríguez-Barroso、M. V. Luzón 和 F. Herrera，通过逐层聚合和余弦距离提升联邦学习中的 (α, f)-拜占庭韧性，Knowledge-Based Systems 326（2025），114004。[GMRBLH25b] Mario García-Márquez、Nuria Rodríguez-Barroso、M Victoria Luzón 和 Francisco Herrera，通过逐层聚合提升 Krum 在联邦学习中的拜占庭韧性，2025 年国际神经网络联合会议（IJCNN），IEEE，2025 年，第 1–8 页。[HFJ+21] Lê-Nguyên Hoang、Louis Faucon、Aidan Jungo、Sergei Volodin、Dalia Papuc、Orfeas Liossatos、Ben Crulis、Mariame Tighanimine、Isabela Constantin、Anastasiia Kucherenko 等，Tournesol：构建大型、安全且可信的可靠人类判断数据库，arXiv 预印本 arXiv:2107.07334（2021），20。[Hub11] Peter J Huber，稳健统计，《国际统计科学百科全书》，Springer，2011 年，第 1248–1251 页。[HWF+24] Lijuan Huo、Libing Wu、Jiaqi Feng、Xing Fan、Enshu Wang 和 Xinchen Li，Dp-caka：在分布式联邦学习中，基于差分隐私和基于复杂 ACC 的 Multi-Krum 算法防御本地模型投毒攻击，2024 年 IEEE 国际高性能计算与通信会议（HPCC），IEEE，2024 年，第 1409–1418 页。[KHJ21] Sai Praneeth Karimireddy、Lie He 和 Martin Jaggi，利用历史信息实现拜占庭鲁棒优化，《第 38 届国际机器学习会议论文集》（Marina Meila 和 Tong Zhang 编），《机器学习研究论文集》第 139 卷，PMLR，2021 年 7 月 18–24 日，第 5311–5319 页。[KMA+21] Peter Kairouz、H. Brendan McMahan、Brendan Avent、Aurélien Bellet、Mehdi Bennis、Arjun Nitin Bhagoji、Kallista Bonawitz、Zachary Charles、Graham Cormode、Rachel Cummings、Rafael G. L. D’ Oliveira、Hubert Eichner、Salim El Rouayheb、David Evans、Josh Gardner、Zachary Garrett、Adrià Gascón、Badih Ghazi、Phillip B. Gibbons、Marco Gruteser、Zaid Harchaoui、Chaoyang He、Lie He、Zhouyuan Huo、Ben Hutchinson、Justin Hsu、Martin Jaggi、Tara Javidi、Gauri Joshi、Mikhail Khodak、Jakub Konecný、Aleksandra Korolova、Farinaz Koushanfar、Sanmi Koyejo、Tancrède Lepoint、Yang Liu、Prateek Mittal、Mehryar Mohri、Richard Nock、Ayfer Özgür、Rasmus Pagh、Hang Qi、Daniel Ramage、Ramesh Raskar、Mariana Raykova、Dawn Song、Weikang Song、Sebastian U. Stich、Ziteng Sun、Ananda Theertha Suresh、Florian Tramèr、Praneeth Vepakomma、Jianyu Wang、Li Xiong、Zheng Xu、Qiang Yang、Felix X. Yu、Han Yu 和 Sen Zhao，《联邦学习的进展与开放问题》，《机器学习基础与趋势》14（2021），第 1–2 期，1–210。[KMJT25] Emmanouil Kritharakis、Antonios Makris、Dusan Jakovetic 和 Konstantinos Tserpes，Fedgreed：一种用于联邦学习的拜占庭鲁棒基于损失的聚合方法，arXiv 预印本 arXiv:2508.18060（2025），0。

[LDC+19] Xiaoli Li、Erxin Du、Chuan Chen、Zibin Zheng、Ting Cai 和 Qiang Yan，基于区块链的可信且隐私保护的 QoS 感知 Web 服务推荐，区块链与可信系统国际会议，Springer，2019 年，第 621–635 页。[LL20] Guillaume Lecué 和 Matthieu Lerasle，通过均值中位数实现鲁棒机器学习：理论与实践，《统计学年刊》48（2020），第 2 期，906–931。[LLZ25] Xiang Li、Lannan Luo 和 Qiang Zeng，通过单比特翻转对神经网络发起后门攻击，《IEEE/CVF国际计算机视觉会议论文集》，2025 年，第 4328–4338 页。[LSP82] Leslie Lamport、Robert Shostak 和 Marshall Pease，拜占庭将军问题，《编程语言与系统汇刊》4 3，382 1982 401 ( )，第 – 期。[NQVHN+24] Thanh Toan Nguyen、Nguyen Quoc Viet Hung、Thanh Tam Nguyen、Thanh Trung Huynh、Thanh Thi Nguyen、Matthias Weidlich 和 Hongzhi Yin，操纵推荐系统：投毒攻击与对策综述，ACM计算综论 57（2024），第 1 期，1–39。[Rou85] Peter J Rousseeuw，具有高崩溃点的多元估计，《数理统计与应用》8（1985），283–297。[Rou22] Sébastien Rouault，实用的抗拜占庭随机梯度下降，博士论文，洛桑联邦理工学院，2022。[SH21] Virat Shejwalkar 和 Amir Houmansadr，操纵拜占庭行为：优化联邦学习中的模型投毒攻击与防御方法，NDSS，2021。[SRC+25] Alexandra Souly、Javier Rando、Ed Chapman、Xander Davies、Burak Hasircioglu、Ezzeldin Shereen、Carlos Mougan、Vasilios Mavroudis、Erik Jones、Chris Hicks 等，针对大语言模型的投毒攻击所需的投毒样本数量近乎恒定，arXiv 预印本 arXiv:2510.07192（2025），0。[SSR+25] Sajal Saha、Moinul Sayed、Md Motiur Rahman、Miad Faezipour 和 Smrity Bhatt，采用 Multi-Krum 聚合和异常检测的 DDoS 检测鲁棒联邦学习，2025 年 SmartNets 国际智能应用、通信与网络会议，IEEE，2025 年，第 1–6 页。[TAGA23] Rahim Taheri、Farzad Arabikhan、Alexander Gegov 和 Negar Akbari，联邦学习中的鲁棒聚合函数，信息与知识系统国际会议，Springer，2023 年，第 168–175 页。[TLMC21] Jiyuan Tu、Weidong Liu、Xiaojun Mao 和 Xi Chen，用于拜占庭鲁棒分布式推断的方差缩减均值中位数估计量，《机器学习研究期刊》22（2021），第 84 期，1–67。[TSMM25] Aditya Tripathi、Karan Sharma、Rahul Mishra 和 Tapas Kumar Maiti，SpectralKrum：一种用于联邦学习中防御拜占庭攻击的谱几何防御方法，arXiv 预印本 arXiv:2512.11760（2025），0。[WSR+20] Hongyi Wang、Kartik Sreenivasan、Shashank Rajput、Harit Vishwakarma、Saurabh Agarwal、Jy-yong Sohn、Kangwook Lee 和 Dimitris Papailiopoulos，尾部攻击：没错，你确实可以对联邦学习植入后门，《神经信息处理系统进展》33（2020），16070–16084。[XKG20] Cong Xie、Oluwasanmi Koyejo 和 Indranil Gupta，帝国陨落：通过内积操纵攻破拜占庭容忍随机梯度下降，《第 35 届人工智能不确定性会议论文集》，PMLR，2020 年 8 月，第 261–270 页。[YCKB18] Dong Yin、Yudong Chen、Ramchandran Kannan 和 Peter Bartlett，拜占庭鲁棒分布式学习：迈向最优统计速率，《第 35 届国际机器学习会议论文集》，PMLR，2018 年 7 月，第 5650–5659 页。

[ZSS22a] 张志远、苏琦和孙旭，*Dim-krum*：采用基于维度级 *Krum* 的聚合、用于自然语言处理的抗后门联邦学习，载于《计算语言学协会论文集：2022年自然语言处理经验方法会议》（阿布扎比，阿拉伯联合酋长国）（Yoav Goldberg、Zornitsa Kozareva 和张跃编），计算语言学协会，2022年12月，第339–354页。[ZSS22b] ,*Dim-krum*：采用基于维度级 *Krum* 的聚合、用于自然语言处理的抗后门联邦学习，arXiv 预印本 arXiv:2210.06894（2022），0。[ZSWJ22] 赵博、孙鹏、王涛和江科宇，*FedInv*：通过反转本地模型更新实现拜占庭鲁棒联邦学习，载于《AAAI人工智能会议论文集》，第36卷，2022年，第9171–9179页。