Robust Sparse Voting

Youssef Allouah
EPFL

Rachid Guerraoui
EPFL

Lê-Nguyên Hoang
Calicarpa, Tournesol Association

Oscar Villemaud
EPFL

Abstract

Many applications, such as content moderation and recommendation, require reviewing and scoring a large number of alternatives. Doing so robustly is however very challenging. Indeed, voters' inputs are inevitably *sparse*: most alternatives are only scored by a small fraction of voters. This sparsity amplifies the effects of biased voters introducing *unfairness*, and of malicious voters seeking to hack the voting process by reporting *dishonest* scores.

We give a precise definition of the problem of *robust sparse voting*, highlight its underlying technical challenges, and present a novel voting mechanism addressing the problem. We prove that, using this mechanism, no voter can have more than a small parameterizable effect on each alternative's score; a property we call *Lipschitz resilience*. We also identify conditions of voters comparability under which any unanimous preferences can be recovered, even when each voter provides sparse scores, on a scale that is potentially very different from any other voter's score scale. Proving these properties required us to introduce, analyze and carefully compose novel aggregation primitives which could be of independent interest.

1 INTRODUCTION

Voting has proven to be an effective way to reach collective decisions despite irreconcilable preferences. However, traditional voting schemes have been designed to handle a tractable set of alternatives. Mechanisms like the *majority judgment* [Balinski and Laraki, 2011], *Borda*'s count [Emerson, 2013], *Kemeny-Young*'s scheme [Kemeny and Snell, 1962], *randomized Condorcet* [Hoang, 2017] and *Schulze method*

[Schulze, 2003], among others, typically require voters to provide ballots whose size is linear in the number of alternatives, and whose computation time is polynomial. Such approaches are inapplicable when the number of alternatives is very large, e.g. when electing the best movie of the year, the best paper of a conference, or the best text of law to implement. In such a context, voting is inherently *sparse*: as voters can only judge a small fraction of all alternatives. Sparsity amplifies two major issues: heterogeneity in *expression styles* and vulnerability to *malicious voters*.

Expression styles. On the one hand, different reviewers may adopt very distinct expression styles [Wang and Shah, 2019]. For example, in the case of scientific peer review, junior reviewers might use only modest judgments, e.g. *weak* accept/reject, while other reviewers may frequently use definitive judgments, e.g. *strong* accept/reject. Meanwhile, some may be systematically positive and rarely suggest rejection, while others may be consistently harsh and almost always recommend rejection. Thus, the resulting acceptance decision of a paper may depend more on the expression styles of the assigned reviewers, than on the actual quality of the paper. In addition, this phenomenon is exacerbated by the fact that the assignment of papers to reviewers is rarely uniformly random. In practice, a paper is more likely to be reviewed by someone whose expertise is close to the paper's focus. Some reviewers may also prefer reviewing top quality papers only, while other reviewers may focus on papers that are easy to reject. Thus, sparsity may not be random: it may be *adversarial* for some papers and raises a risk of *systematic unfairness*. *Robust sparse voting* requires protections against diverging expression styles.

Malicious voters. On the other hand, especially when the number of alternatives far exceeds what honest voters can collaboratively score, as for social media content, we must expect the existence of alternatives that no honest voter has scored. This makes classical solutions for the robust statistics toolbox, like the median, ill-suited to protect the security of the vote, as such alternatives with no honest voter's assessment will be arbitrarily manipulable by a single malicious voter.

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, Oscar Villemaud

Robust sparse voting also requires protection against malicious voters targeting such alternatives.

1.1 Contributions

- We propose a formalization of the *robust sparse voting* problem, by identifying two formal properties tackling the aforementioned issues of voting biases and malicious voters. The first property, which we call *sparse unanimity*, stipulates that any unanimous preference can be recovered when sufficiently many voters participate, despite diverging expression styles and sparse score reporting. The second property, which we call *L-Lipschitz resilience*, is a security property guaranteeing that any contributor can have at most a parameterizable impact L on the scoring of alternatives.

- We introduce and analyze two novel Lipschitz-resilient aggregation primitives: *Quadratically Regularized Median* is a generalization of the median, and *Lipschitz-Robustified Mean* successfully outputs the mean under the right conditions. We believe our new aggregation primitives to be of independent interest.

- We propose a new voting algorithm, which we call MEHESTAN[1], and formally prove that it solves the *robust sparse voting* problem. Underlying MEHESTAN lie our novel aggregation primitives mentioned above, which we carefully compose to conduct two crucial operations: transform the different scores to bring them to a common scale, thus diluting voting biases, and perform robust aggregation on the transformed scores.

- We empirically compare MEHESTAN to natural baselines for robust sparse voting, under various adversarial settings. In particular, we evaluate MEHESTAN under high sparsity (voters score only a few alternatives), biased sparsity (voters only score a biased subset of alternatives), and in the presence of malicious voters sending random scores.

1.2 Applications

Large-scale algorithms routinely address a massive number of ethical dilemmas. Namely, whenever a user searches "climate change" or "vaccines" on YouTube, Facebook or Amazon, the algorithmic answers have potential life-or-death consequences transcending geographical boundaries. These algorithms, heavily reliant on (implicit) voting mechanisms using upvotes or star ratings, struggle with ethical decision-making due to the inherent sparsity of evaluations. Indeed, most alternatives receive scrutiny from only a small fraction of users, if any at all, creating a complex challenge in addressing these ethical quandaries.

In the context of online *content recommendation*, resilience to arbitrary behavior has become critical. Social media have become information battlegrounds [Atallah, 2019, Bradshaw et al., 2021], and their recommendation algorithms have been weaponized by all sorts of private and public actors [Satariano, 2021, Yue, 2019], many of which leverage troll farms to fabricate misleading online activities [Bradshaw and Howard, 2019, Neudert et al., 2019, Woolley, 2020], or even simply exploit the vulnerabilities of the social media's advertisement systems [Edelson et al., 2020]. The sheer scale of disinformation campaigns is staggering, as exemplified by Facebook's removal of *15 billion* fake accounts in just two years [Dolden, 2021]. Unfortunately, the algorithms that underpin content moderation, recommendations, and ad-targeting remain opaque, providing fertile ground for an extensive industry of fake accounts that continuously manipulate online content [Moore, 2023]. This ongoing manipulation underscores the critical need for resilient, transparent, and accountable algorithms to ensure the integrity and trustworthiness of online information.

It is also worth mentioning low-stake applications such as *online surveys*. There, robust sparse voting algorithms can handle incomplete and potentially biased responses in online surveys and polls, providing accurate representations of public opinions even when faced with malicious attempts to influence the results or self-selection biases [Bethlehem, 2010, Schaurer and Weiß, 2020]. Besides, in *peer-to-peer platforms* like Airbnb and Uber, reputation is vital for trust and safety. Malicious users might attempt to damage the reputation of others or artificially boost their own [Dellarocas, 2000, Xiong and Liu, 2004]. A robust sparse voting algorithm can reduce the influence of such behavior while accommodating real-world feedback.

1.3 Structure of the Paper

Section 2 proposes a formalization of the *robust sparse voting* problem. Section 3 introduces our new robust aggregation primitives. Section 4 introduces MEHESTAN, and proves its *resilience* and *sparse unanimity* properties. Section 5 presents our empirical evaluation[2] of MEHESTAN under adversarial settings. Section 6 reviews related work in social choice theory and

---

[1]Mehestan is the name of one of the earliest proto-parliaments in Asia.

[2]Our code is available at `https://github.com/ysfalh/robust-voting`.

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, Oscar Villemaud

robust statistics. Additional related work on recommender systems and robust voting is in Appendix A. Missing proofs are in appendices C to F. Appendix J presents additional experiments on MEHESTAN. Appendix I exposes the difficulty of sparse voting by proving the impossibility of sparse unanimity for individually scaled preferences. Appendix K extends MEHESTAN to guarantee desirable properties such as differential privacy [Dwork et al., 2014].

2 ROBUST SPARSE VOTING

We consider a set $[N] = \{1, \ldots, N\}$ of voters, and a set $[A] = \{1, \ldots, A\}$ of alternatives to score. Each voter $n \in [N]$ is asked to provide score $\theta_{na} \in \mathbb{R}$ for each alternative $a \in [A]$. Moreover, we allow their vote to be *sparse*: voters may fail to score an alternative a, in which case we denote $\theta_{na} \triangleq \perp$. Denote $\tilde{\mathbb{R}} \triangleq \mathbb{R} \cup \{\perp\}$. Then each voter's input is a vector $\theta_n \in \tilde{\mathbb{R}}^A$. We denote $\boldsymbol{\theta} \triangleq (\theta_1, \ldots, \theta_N)$ the tuple of voters' scores. Note that assuming that input scores are bounded has no incidence on, nor is needed in, our theory. For the sake of exposition, we assume that all voters are given a unit voting right. The appendix details the more general case of continuous voting rights.

Following the classical Von Neumann-Morgenstern rationality framework [Von Neumann and Morgenstern, 1953], we assume that each voter n's cardinal preference θ_n is defined up to a positive affine transformation. That is, we consider θ_n, θ'_n to be *equivalent*, and denote $\theta_n \sim \theta'_n$, if and only if the two vectors score the same subset A_n of alternatives, and there exist $s > 0, \tau \in \mathbb{R}$ such that $\theta_{na} = s\theta'_{na} + \tau$ for every alternative $a \in A_n$.

Our goal is to aggregate the voters' partial scores for all alternatives. In other words, we aim to construct a voting algorithm $\text{VOTE} : \left(\tilde{\mathbb{R}}^A\right)^N \to \mathbb{R}^A$, which maps tuple $\boldsymbol{\theta}$ to a score vector $\rho \in \mathbb{R}^A$. We denote by $\text{VOTE}_a(\boldsymbol{\theta})$ the score given to alternative a using VOTE.

2.1 The Lipschitz Resilience Property

We now introduce Lipschitz resilience, our security property against malicious voters. Classical solutions sometimes demand that the output of the algorithm be insensitive to malicious voters [Lamport et al., 1982]. But this is ill-suited to sparse voting where, on some alternatives, the majority of voters may be malicious. Instead, Lipschitz resilience demands that the maximal impact of any (malicious) voter be bounded.

To formalize this definition, let us define $\perp_A \in \tilde{\mathbb{R}}^A$ the empty vector, i.e. defined by $(\perp_A)_a = \perp$ for all $a \in [A]$.

Definition 1 (Lipschitz resilience). VOTE guarantees

L-Lipschitz resilience if, for all inputs $\boldsymbol{\theta}$ and any voter $n \in [N]$, discarding voter n's inputs can affect each output of the vote by at most L, i.e.

$$\forall \boldsymbol{\theta}, n, \forall a, \ |\text{VOTE}_a(\boldsymbol{\theta}) - \text{VOTE}_a(\boldsymbol{\theta}_{-n}, \perp_A)| \leq L,$$

where $\boldsymbol{\theta}_{-n}$ is the tuple $\boldsymbol{\theta}$ deprived of voter n's inputs. VOTE is simply said to be resilient if VOTE is L-Lipschitz resilient for some $L > 0$.

Interpretation. Lipschitz resilience can be naturally interpreted as Lipschitz continuity, if we consider the ℓ_0-norm for the input tuple $\boldsymbol{\theta}$, and the ℓ_∞-norm for the output vector $\text{VOTE}(\boldsymbol{\theta})$. In the appendix, we generalize this definition to continuous voting rights, and show that the Lipschitz continuity interpretation still holds with the ℓ_1-norm on the voting rights. The variable L can be interpreted as a resilience measure: F malicious voters cannot deviate the final score of an alternative by more than $F \cdot L$.

Note that this is a nontrivial condition to guarantee. In particular, algorithms based on identifying a subset of reference/anchor alternatives to scale all users' preferences will usually fail to provide Lipschitz resilience, as they often feature a discontinuity when the subset of reference alternatives or users is changed.

Use cases. Lipschitz resilience is particularly useful in three scenarios. First, it is critical for sparse voting where only a few honest voters score some alternatives. Without Lipschitz resilience, malicious voters could manipulate the scores of such alternatives. This is especially harmful in applications where low scores lead to censorship, while high scores lead to celebration. Second, Lipschitz resilience is important for *stability* in systems where high volatility may be discreditable. Last, L-Lipschitz resilience is desirable for privacy-sensitive applications, such as healthcare and finance. Our voting algorithm MEHESTAN guarantees ε-differential privacy [Dwork et al., 2014] when additionally injecting Laplacian noise proportional to L. This adaptation is possible for any L-Lipschitz resilient voting algorithm and is explained in Appendix K.1.

2.2 The Sparse Unanimity Property

Our second desirable property is *sparse unanimity*, which guarantees that the voting algorithm recovers unanimous preferences despite sparsity. More precisely, consider the situation where an algorithm is given N (positive affine) transformations of the same ground-truth scores vector $\theta_* \in \mathbb{R}^A$, where each transformation has some hidden coordinates (i.e., is sparsified). Sparse unanimity guarantees that, once enough voters participate, the algorithm recovers θ_*.

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, Oscar Villemaud

We first introduce notation: for any voter score $\theta_n \in \mathbb{\dot{R}}^A$, denote A_n the subset of alternatives a for which $\theta_{na} \neq \perp$ has been reported. For any subset $B \subset [A]$, let $\theta_{|B} \in \mathbb{R}^B$ be the (partial) score vector obtained by selecting only the entries $a \in B$ from the partial vector θ. Moreover, let $N_a \triangleq \{n \in N \,|\, a \in A_n\}$ be the set of voters who scored alternative a. Define $C_{nm}(\boldsymbol{\theta}) \triangleq \{(a,b) \in (A_n \cap A_m)^2 \,|\, a < b, \theta_{na} \neq \theta_{nb}$ and $\theta_{ma} \neq \theta_{mb}\}$ the set of couples of alternatives that both voters n and m scored, each providing distinct scores to the two alternatives. We can now formalize *sparse unanimity* whose definition will be clarified right after.

Definition 2 (Sparse unanimity)**.** VOTE is *sparsely unanimous* if, for all $\theta_* \in \mathbb{R}^A$, there exists $N_0 \geq 0$ such that, whenever voters' scores are θ_*-*unanimous*, *comparable* and N_0-*scored*, VOTE retrieves the unanimous preferences. More precisely, the assumptions

θ_*-*unanimity:* $\quad \forall n \in [N],\ \theta_n \sim \theta_{*|A_n},$
comparability: $\quad \forall n \neq m \in [N],\ C_{nm}(\boldsymbol{\theta}) \neq \varnothing,$
N_0-*scored:* $\qquad \forall a \in [A],\ |N_a| \geq N_0$

must imply $\text{VOTE}(\boldsymbol{\theta}) \sim \theta_*$.

We start by clarifying the conditions of sparse unanimity. First, θ_*-*unanimity* requires the reported scores of each voter n to be equivalent to unanimous preferences θ_* (i.e., ground-truth scores), but only on the subset A_n of alternatives scored by voter n. Second, *comparability* requires, for every distinct voters $n, m \in [N]$, the existence of at least one pair of alternatives they both scored distinctly. Intuitively, comparability limits the sparsity of the inputs. Finally, voters' scores are N_0-*scored* if every alternative was scored by at least N_0 voters. One difficulty is that N_0 must only depend on θ_*, and cannot be made to depend on which voters scored which alternatives.

Let us consider a numerical example to illustrate the sparse unanimity property:

Example. *Consider a situation with $N = 2K$ voters and $A = 4$ alternatives. The unanimous preferences are given by $\theta_* = [-2, -1, 1, 2]$. Odd voter $2k + 1$ scores the first three alternatives $A_1 = \{1, 2, 3\}$ with $\theta_1 = [-3, -2, 0, \perp]$, and even voter $2k$ scores the last three alternatives $A_2 = \{2, 3, 4\}$ with $\theta_2 = [\perp, 0, 4, 6]$. Voters' scores are θ_*-unanimous: we have $\theta_1 = \theta_{*|A_1} - 1$, and $\theta_2 = 2\,\theta_{*|A_2} + 2$. Also, they verify comparability since both voters scored alternatives 2 and 3 differently. Any sparse unanimity guarantee on VOTE implies the existence of a value $N_0(\theta_*)$, for which we guarantee $\text{VOTE}(\boldsymbol{\theta}) \sim \theta_*$. For $K \geq N_0(\theta_*)/2$, VOTE would then be guaranteed to recover θ_*.*

Failure of naive solutions. Sparse unanimity is a minimally desirable *correctness* property that any

sparse voting algorithm ought to satisfy. Yet, it is surprisingly nontrivial to guarantee because of sparsity. Indeed, if all voters score all alternatives, a natural solution is to aggregate voters' scores after any basic normalization of each voter's score (e.g., using min-max normalization). Such solutions unfortunately fail for sparse voting. More precisely, in Appendix I, we prove an impossibility result: (reasonable) aggregations of voters' scores with individual-based normalizations (i.e., the normalization does not depend on other voters) fail to be *sparsely unanimous*.

Strengthening sparse unanimity. We leave open the question of strengthening sparse unanimity, and of constructing a Lipschitz resilient algorithm that satisfies this strengthened condition. We believe this problem to be very challenging. For instance, we conjecture that no Lipschitz resilient algorithm can guarantee what could be called *sparse majority*: informally, if we require the recovery of any preference that is consensual among a *majority* of sufficiently active voters, then Lipschitz resilience cannot hold.

2.3 Nontriviality

We introduce a third property, which we call *nontriviality*, requiring that the diameter of the output of the vote be at least 1, under the assumptions of sparse unanimity.

Definition 3 (Nontriviality)**.** VOTE is *nontrivial* if, whenever voters' scores are θ_*-*unanimous*, *comparable* and N_0-*scored*, there exist $a, b \in [A]$ such that $\text{VOTE}_a(\boldsymbol{\theta}) - \text{VOTE}_b(\boldsymbol{\theta}) \geq 1$.

The nontriviality property ensures that the vote's result is placed on an informative scale. Although beneficial, this feature is not essential, and its removal does not simplify the challenge of robust sparse voting.

3 ROBUST PRIMITIVES

We now introduce robust score aggregation functions used in MEHESTAN to guarantee Lipschitz resilience. For simplicity, in this section, we assume that there is only one alternative; i.e. $A = 1$ and thus each θ_n is a *scalar* (or non-reported value).

3.1 Weighted Averaging and Median

We first show that classical (robust) statistics operators, averaging and median, fail to be Lipschitz resilient.

Averaging. A widely used algorithm, e.g. for *collaborative filtering* algorithms for group recommender systems [Felfernig et al., 2018a], is the averaging of

available data, i.e. $\text{MEAN}(\boldsymbol{\theta}) \triangleq \left(\sum_{n \in N^*} \theta_n\right) / |N^*|$, where N^* is the set of voters n who reported a score, i.e. $\theta_n \neq \perp$. It was proved to satisfy several desirable voting properties [Pennock et al., 2000].

Median. A popular robust mean estimator is the median, which we denote MED, and is the main ingredient of the popular *majority judgement* [Balinski and Laraki, 2011] voting algorithm. A median $M \triangleq \text{MED}(\boldsymbol{w}, \boldsymbol{\theta})$ must divide voters with reported scores into two sets of equal sizes, i.e. $|\{n \in N^* : \theta_n < M\}| \leq \frac{1}{2} |N^*| \geq |\{n \in N^* : \theta_n > M\}|$.

Proposition 1 below shows that weighted averaging and median fail to satisfy our resilience property.

Proposition 1. *Neither* MEAN *nor* MED *is Lipschitz resilient.*

The proof, which can be found in Appendix C, instantiates a simple situation where the malicious voter can manipulate the outcome more than what is allowed by L-Lipschitz resilience. One can similarly prove the failure of other classical aggregators, like trimmed/winsorized mean, to be Lipschitz resilient.

3.2 Quadratically Regularized Median

We now introduce the *quadratically regularized median*, denoted QRMED. QRMED_L is parameterized by $L > 0$, and is defined as follows:

$$\text{QRMED}_L(\boldsymbol{\theta}) \triangleq \underset{z \in \mathbb{R}}{\arg\min} \ \frac{1}{2L} z^2 + \sum_{n : \theta_n \neq \perp} |z - \theta_n|. \quad (1)$$

In practice, we approximate QRMED_L by solving (1), which is an exponentially fast operation (in the approximation error) using gradient descent. Note also that QRMED_L corresponds to the median when $L \to \infty$ [Minsker, 2015]. Theorem 1 guarantees that QRMED_L is L-Lipschitz resilient. The full proof is deferred to Appendix D.

Theorem 1. QRMED_L *is well-defined and L-Lipschitz resilient.*

Sketch of proof. The objective (1) minimized by QRMED_L is $\frac{1}{L}$-strongly convex, which implies that its minimizer is unique and that QRMED_L is well-defined. L-Lipschitz resilience follows from strong convexity, and the fact that the derivative of each summand $|z - \theta_n|$ is bounded by 1. □

An additional property, which is used in the proof for MEHESTAN's resilience, is the Lipschitz continuity of QRMED, as stated by Proposition 2 in Appendix G.

3.3 Lipschitz-Robustified Mean

We now introduce *Lipschitz-Robustified Mean*, which we denote LRMEAN, a primitive that returns the mean of any bounded inputs when sufficiently many voters participate, while satisfying L-Lipschitz resilience. It builds upon the *clipped mean* CLMEAN centered on μ and of radius Δ defined as:

$$\text{CLMEAN}(\boldsymbol{\theta}|\mu, \Delta) \triangleq \text{MEAN}(\text{CLIP}(\boldsymbol{\theta}|\mu, \Delta))$$
$$= \frac{1}{|N^*|} \sum_{n \in N^*} \text{CLIP}(\theta_n|\mu, \Delta),$$

where $\text{CLIP}(x|\mu, \Delta) \triangleq \max\{\mu - \Delta, \min\{\mu + \Delta, x\}\}$ clips x within the interval $[\mu - \Delta, \mu + \Delta]$, and where $N^* \triangleq \{n : \theta_n \neq \emptyset\}$. LRMEAN is then obtained by executing CLMEAN, centered on QRMED, with a radius that grows linearly with the number of users N^*:

$$\text{LRMEAN}_L(\boldsymbol{\theta}) \triangleq \text{CLMEAN}\left(\boldsymbol{\theta} \ \middle| \ \text{QRMED}_{L/4}(\boldsymbol{\theta}), \frac{L|N^*|}{4}\right).$$

As stated by Theorem 2 below, LRMEAN verifies several properties. The full proof is in Appendix E.

Theorem 2. LRMEAN_L *is L-resilient. Moreover, if there exists $\Delta > 0$ such that $|N^*| \geq 8\Delta/L$ and $\theta_n \in [-\Delta, \Delta]$ for all n, then $\text{LRMEAN}_L(\boldsymbol{\theta}) = \text{MEAN}(\boldsymbol{\theta})$.*

Sketch of proof. In the proof, we show that CLMEAN is 1-Lipschitz in the center μ and in the radius Δ, and is also sufficiently resilient for small radii, when the number of voters is large enough. The guarantee $\text{LRMEAN} = \text{MEAN}$ then holds once enough voters participate, so that the radius of CLMEAN could safely grow large enough to contain all voters' inputs. □

Remarkably, LRMEAN eventually returns the mean of bounded inputs, provided that sufficiently many voters participate, despite being oblivious to the input bounds. This is a critical property that will be at the heart of the sparse unanimity guarantee of MEHESTAN. Designing a resilient aggregation with this feature turned out to be the most challenging aspect of our algorithm design. An additional property, which we use in the proof for MEHESTAN's resilience, is the Lipschitz continuity of LRMEAN with respect to its inputs, as stated by Proposition 3 in Appendix G.

4 OUR ALGORITHM: MEHESTAN

In this section, we first introduce our algorithm MEHESTAN, and conclude with theoretical guarantees.

4.1 Description of MEHESTAN

MEHESTAN proceeds in four principal steps:

1. Local normalization. First, MEHESTAN normalizes every score vector using min-max normalization, so that the minimal and maximal scores are respectively 0 and 1:

$$\tilde{\theta}_n \triangleq \frac{\theta_n - \min_{a \in A_n} \theta_{na}}{\max_{a \in A_n} \theta_{na} - \min_{a \in A_n} \theta_{na}}. \qquad (2)$$

Note that min-max normalization is well-defined only if θ_n has at least two distinct reported scores. If not, the scores are non-informative with respect to (multiplicative) scaling since they are equivalent to the vector of zeros, so we simply set $\tilde{\theta}_n$ to be the vector of zeros, without any incidence on the theoretical guarantees.

2. Scaling factor search. For any voter $n \in [N]$, we define $N_n^C \triangleq \{m \in [N] \mid C_{nm}(\boldsymbol{\theta}) \neq \varnothing\}$ the set of voters comparable to n. In words, $m \in N_n^C$ if and only if there exist two alternatives that n and m both scored differently. For each voter $m \in N_n^C$, we compute the comparative scaling s_{nm} of voters n and m defined as

$$s_{nm} \triangleq \frac{1}{|C_{nm}|} \sum_{(a,b) \in C_{nm}} \frac{|\tilde{\theta}_{ma} - \tilde{\theta}_{mb}|}{|\tilde{\theta}_{na} - \tilde{\theta}_{nb}|}. \qquad (3)$$

From a high-level perspective, each comparative scaling s_{nm} is an implicit vote by voter m for the scaling factor of voter n. In this step, MEHESTAN aggregates the comparative scalings s_{nm} via LRMEAN:

$$s_n \triangleq 1 + \text{LRMEAN}_{L/7}\big(\{s_{nm} - 1 \mid m \in N_n^C\}\big). \qquad (4)$$

Recall from the definition of LRMEAN in Section 3, that the step of Equation 4 implicitly clips the comparative scalings to be close to their quadratically regularized median (QRMED), and then takes the plain average. Above, we do not directly take the LRMEAN of the comparative scaling ratios s_{nm} so that the default value equals 1 in the absence of comparable voters. Moreover, since each s_{nm} corresponds to a vote for the scaling factor of s_n, the computation of s_n can be interpreted as a search for a *common scale*.

3. Translation factor search. While the previous step computes a common multiplicative scaling factor, the current step searches for a *common translation*. More precisely, define $N_n^A \triangleq \{m \in [N] \mid A_{nm} \neq \emptyset\}$ the set of translation-comparable voters. For each voter $m \in N_n^A$, we compute

$$\tau_{nm} \triangleq \frac{1}{|A_{nm}|} \sum_{a \in A_{nm}} \left(s_m \tilde{\theta}_{ma} - s_n \tilde{\theta}_{na} \right). \qquad (5)$$

In this step, MEHESTAN aggregates the comparative translation factors τ_{nm} via LRMEAN:

$$\tau_n \triangleq \text{LRMEAN}_{L/7}\left(\{\tau_{nm} \mid m \in N_n^A\}\right). \qquad (6)$$

Algorithm 1 MEHESTAN

Input: Voters' scores $\boldsymbol{\theta}$, Lipschitz resilience L
Output: The aggregate scores $\text{MEHESTAN}_L(\boldsymbol{w}, \boldsymbol{\theta})$

1: $\forall n$, compute $\tilde{\theta}_n$, the min-max normalization of θ_n
   ▷ Local normalization
2: $\forall n, m \in [N]$, compute s_{nm} following Equation (3), if n, m are comparable
3: $\forall n, \; s_n \leftarrow 1 + \text{LRMEAN}_{L/7}\big(s_{nm} - 1 \mid m \in N_n^C\big)$
4: $\forall n, m \in [N]$, compute τ_{nm} following Equation (5), if n, m are comparable
5: $\forall n, \; \tau_n \leftarrow \text{LRMEAN}_{L/7}\big(w_m, \tau_{nm} \mid m \in N_n^A\big)$
6: $\forall n, \; \hat{\theta}_n \leftarrow s_n \tilde{\theta}_n + \tau_n$ ▷ Global normalization
7: $\forall a, \; \rho_a \leftarrow \text{QRMED}_{L/7}(\hat{\theta}_{na} \mid n \in [N])$
8: **return** ρ

---

4. Alternative-wise score aggregation. Finally, MEHESTAN linearly transforms the scores vectors with the obtained scaling and translation factors:

$$\hat{\theta}_n := s_n \tilde{\theta}_n + \tau_n. \qquad (7)$$

We refer to this step as *global* normalization, as opposed to *local* normalization (2), because the transformation of each vector is dependent on all input scores. Then, MEHESTAN aggregates the transformed scores along each alternative $a \in [A]$ via QRMED:

$$\text{MEHESTAN}_{La}(\boldsymbol{\theta}) = \text{QRMED}_{L/7}\left(s_n \tilde{\theta}_{na} + \tau_n \;\middle|\; n \in [N] \right).$$

The full procedure of MEHESTAN is summarized in Algorithm 1.

4.2 Theoretical Guarantees

Finally, we state the main result of our paper in Theorem 3 below, showing that MEHESTAN satisfies both properties of the robust sparse voting problem. We defer the full proof to Appendix F.

Theorem 3. MEHESTAN$_L$ *is L-Lipschitz resilient, sparsely unanimous and nontrivial, with*

$$N_0(\theta_*) \triangleq \frac{8}{L} \left(\frac{\max_{a,b} |\theta_{*a} - \theta_{*b}|}{\min_{a,b:\theta_{*a} \neq \theta_{*b}} |\theta_{*a} - \theta_{*b}|} \right)^2. \qquad (8)$$

Sketch of proof. Let us first address *sparse unanimity*. Placing ourselves in the situation where voters' scores are θ_*-unanimous, we can write each input scores vector as $\theta_n = s_n^* \theta_{*|A_n} + \tau_n^*$, where $s_n^* > 0$ and τ_n^* are *unknown* to the algorithm. The main challenge of the proof is to show that the scores $\hat{\theta}_n$, obtained after global normalization (7), can be written in the form $s^* \theta_{*|A_n} + \tau^*$, where $s^* > 0, \tau^* \in \mathbb{R}$ are voter-independent. The latter quantities are exactly what defines the "common scale"

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, Oscar Villemaud

found by MEHESTAN, before applying an alternative-wise aggregation. To do so, the proof proceeds as follows: once enough voters participate, the properties of LRMEAN, especially that it returns the mean in specific conditions (see Section 3), allow showing that each comparative scaling factor (see Equation 3) is in fact $s_{nm} = s_m^*/s_n^*$. We then show the scaling factor s_n (see Equation 4) to equal s^*/s_n^*, where s^* is voter-independent. Similarly, we show the translation factor τ_n (see Equation 6) to equal $\tau^* - s^*\tau_n^*/s_n^*$, where τ^* is voter-independent. Overall, for every n, we obtain $\hat{\theta}_n = s_n \tilde{\theta}_{na} + \tau_n = s^*\theta_{*|A_n} + \tau^*$. Then, the final aggregation performed with QRMED_L returns $s^*\theta_* + \tau^*$, which satisfies sparse unanimity.

We now address *Lipschitz resilience*. Since we show LRMEAN_L and QRMED_L to be L-Lipschitz resilient, all aggregation operations made in MEHESTAN are resilient, and the proof mainly composes the bounds guaranteed by L-Lipschitz resilience. Note that we set the resilience parameter of LRMEAN and QRMED to $L/7$ so that the bounds of resilience can be correctly composed to yield the final L-Lipschitz resilience.

Finally, *nontriviality* from the fact that scaled individual scores are necessarily blown up, compared to min-max normalization of the full score vector θ_*, which already satisfies nontriviality. □

Trade-off discussion. The analysis of MEHESTAN underlying Theorem 3 raises a tension between L-Lipschitz resilience and sparse unanimity. To see this, recall that verifying sparse unanimity requires enough voters to participate, which means that alternatives should be N_0-scored (see Definition 2). In fact, as shown in Equation 8, it is sufficient for N_0 to be proportional to $\frac{1}{L}$ and a function of θ_* bounding the difference between voters' scalings; i.e. to what extent voters express the same preferences θ_* differently. Therefore, stronger Lipschitz resilience (lower L) implies that more participation is needed (larger N_0) to recover unanimous preferences. This trade-off raises an interesting research question: whether this trade-off is fundamental or algorithm-dependent. Answering this question can lead to more efficient algorithms or tighter theoretical bounds for MEHESTAN.

Complexity. The collaborative scaling of voters' scores (steps 2 and 3 in Section 4.1) is the computational bottleneck of MEHESTAN. In the worst case, for all pairs of voters, it requires going through all pairs of alternatives, thereby yielding a $\mathcal{O}(A^2 N^2)$ time complexity. In practice, this heavy workload can be mitigated, by performing collaborative scaling in an asynchronous rare manner, and by assuming that the scaling factors do not vary much over time [Beylerian et al., 2022].

5 EMPIRICAL EVALUATION

We report experiments[3] testing the performance of MEHESTAN under sparsity and malicious attacks.
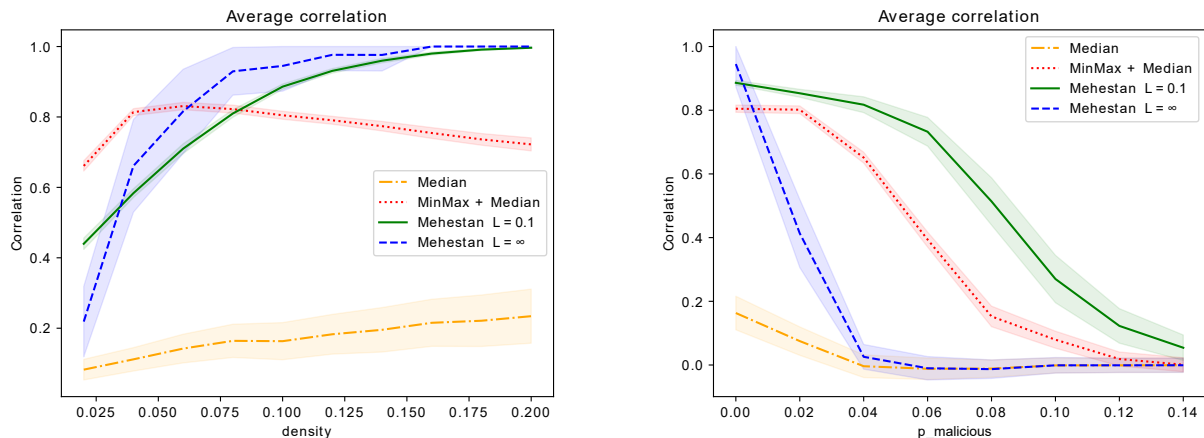
5.1 Setting

Data generation. We generate synthetic data for $N = 150$ voters and $A = 300$ alternatives. We randomly draw a ground-truth score vector $\theta_* \in \mathbb{R}^A$, by independently sampling the coordinates from the standard Gaussian distribution. Each honest voter n is assigned the score vector $\theta_n = s_n^*\theta_{*|A_n} + \tau_n^*$, where $s_n^* > 0$ and τ_n^* are randomly drawn from the log-normal and normal distributions, respectively. To simulate sparsity, each alternative is scored by each voter following a Bernoulli trial with probability `density`. Additionally, to simulate biased sparsity, we remove the votes of half of the voters for the top 20% alternatives (sorted according to θ_*), and the votes of the other half of voters for the bottom 20% alternatives.

Description of experiments. Each experiment measures Pearson's correlation between the algorithms' output and the ground-truth preferences θ_*, which we know since we synthetically generate the data. Our experiments compare four voting algorithms: *Median* (alternative-wise median), *MinMax+Median* (alternative-wise median of min-max normalized scores), MEHESTAN with resilience parameters $L = 0.1$ and $L \to \infty$. The first two algorithms serve as natural baselines (as discussed in Section 2.2): Median is a naive baseline, highlighting the difficulty of the problem, and Min-max +Median is a natural baseline performing well in the absence of adversaries. The latter is representative of a large class of voting methods, based on robust aggregation after local normalization, all of which we expect to perform similarly following the negative results of Appendix I.

In the experiment reported in Figure 1b, we add malicious voters, whose votes are all the same, randomly drawn from the standard Gaussian. In Figure 1b, the x-axis is `p_malicious`, which denotes the fraction of malicious voters. Note that this parameter takes reasonably large values in Figure 1b. Indeed, the malicious voters scores every alternative, as opposed to the honest voters who score $\approx 8\%$ of them (when `density = 0.1`). Each experiment is repeated 20 times, with the seeds 1 to 20 for reproducibility. The average correlation values and the 95% confidence intervals are plotted.

---

[3]Our code is available at `https://github.com/ysfalh/robust-voting`.

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, Oscar Villemaud

(a) Performance under sparsity without malicious voters.

(b) Performance with malicious voters (`density` $= 0.1$).

Figure 1: Performance of MEHESTAN under sparsity, with and without malicious voters.

5.2 Results

Our experimental results are shown in Figure 1. Observe on both plots that *Median* fails to recover the ground-truth scores. This is expected, as (alternative-wise) *Median* is not designed to tackle voting biases. Below, we compare MEHESTAN with *MinMax+Median*, which puts the focus on the global normalization procedure specific to MEHESTAN (see Algorithm 1). Next, we compare MEHESTAN with $L \to \infty$ and $L = 0.1$, highlighting the impact of the resilience parameter L.

Impact of global normalization. Figure 1a shows that MEHESTAN performs well under sparsity, and correctly recovers the unanimous preferences when the density is large enough (as guaranteed by sparse unanimity). *MinMax+Median* fails to do so, even for larger densities. However, we observe that for the highest levels of sparsity (i.e. low density), MEHESTAN is less effective than *MinMax+Median*. This can be explained by the fact that, at these densities, the search for scaling and translation factors may fail, because of the absence of comparable voters (see Section 4).

Impact of resilience parameter L. On the one hand, setting $L = 0.1$ only slightly hinders the performance of MEHESTAN in the absence of malicious voters (Figure 1a). This is expected, as the non-malicious case does not require L-Lipschitz resilience. On the other hand, setting $L = 0.1$ enables better tolerance to malicious voters (Figure 1b). Indeed, MEHESTAN with $L \to \infty$ fares poorly in Figure 1b, as setting $L \to \infty$ leaves the global normalization vulnerable to malicious manipulation. This confirms the trade-off between resilience and sparse unanimity, as discussed in Section 4.2.

6 RELATED WORK

We cover closely related work below, and defer additional related work to Appendix A.

Social choice theory. The problem of *sparse voting*, without malicious voters, has mainly been addressed in ordinal voting, i.e. voters provide incomplete *rankings* or pairwise comparisons [Negahban et al., 2012, Lu and Boutilier, 2013, Moreno-Centeno and Escobedo, 2016, Fotakis et al., 2021, Imber et al., 2022]. There, voting seeks to retrieve a central ranking despite sparsity, which is very similar to sparse unanimity, except that the latter also recovers cardinal preferences, which may be more suitable for some applications. Additionally, intriguingly, [Wang and Shah, 2019] proves that cardinal inputs allow constructing estimators that strictly outperform ordinal inputs, even when considering arbitrary order-preserving input miscalibrations, which suggests that leveraging cardinal inputs may be valuable. The cardinal (robust) sparse voting setting has however been relatively understudied. In this setting, the work of Meir et al. [Meir et al., 2022] is closest to ours, but we argue that our theoretical approach is more general. Namely, [Meir et al., 2022] (i) does not tackle the general cardinal case, where voters provide scores for multiple alternatives; (ii) assumes honest voters to be either passive (do not vote at all), or active (vote for all alternatives), while we allow honest voters to vote for a (strict) subset of alternatives; and (iii) assumes the existence of a distinguished alternative, called "reality", which serves as an anchor to their safety and liveness properties.

Robust statistics. A lot of prior work has provided a wide range of robust statistical estimators [Morgenthaler, 2007]. However, the theory of ro-

bust statistics has usually relied on majority-based principles. To the best of our knowledge, our paper is the first to study Lipschitz resilience, i.e. bounding the maximal impact of any data source, which is arguably more adapted to a sparse setting. Note that our algorithms rely on regularization to stabilize the estimation. A similar idea was previously used in signal processing [Liu and Palomar, 2019], in order to achieve robust mean and covariance estimation with incomplete data considering a monotone missing-data pattern.

7 CONCLUSION

This paper introduces the robust sparse voting problem, highlights its technical challenges, and presents MEHESTAN, a novel algorithm to solve it. Our work opens several research directions. Particularly appealing are analyzing the strategyproofness of the system, and exploring connections with ordinal voting. Another interesting direction is investigating properties such as order consistency and independence of irrelevant alternatives. Overall, we regard our work as merely a first, hopefully inspiring, step towards understanding how a group of individuals should collaborate to securely evaluate an overwhelming amount of alternatives.

Acknowledgments

References

[Andersen et al., 2008] Andersen, R., Borgs, C., Chayes, J. T., Feige, U., Flaxman, A. D., Kalai, A., Mirrokni, V. S., and Tennenholtz, M. (2008). Trust-based recommendation systems: an axiomatic approach. In *WWW 2008*, pages 199–208. ACM.

[Arrow, 1950] Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of political economy*, 58(4):328–346.

[Atallah, 2019] Atallah, N. M. (2019). How internet has become a battleground in the lebanese revolution. *Le Commerce du Levant*, 26.

[Balinski and Laraki, 2011] Balinski, M. and Laraki, R. (2011). *Majority judgment: measuring, ranking, and electing.* MIT press.

[Bethlehem, 2010] Bethlehem, J. (2010). Selection bias in web surveys. *International statistical review*, 78(2):161–188.

[Beylerian et al., 2022] Beylerian, R., Colbois, B., Faucon, L., Hoang, L. N., Jungo, A., Noac'h, A. L., and Matissart, A. (2022). Tournesol: Permissionless collaborative algorithmic governance with security guarantees. *arXiv preprint arXiv:2211.01179*.

[Bradshaw et al., 2021] Bradshaw, S., Bailey, H., and Howard, P. N. (2021). Industrialized disinformation: 2020 global inventory of organized social media manipulation.

[Bradshaw and Howard, 2019] Bradshaw, S. and Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation.* Project on Computational Propaganda.

[Chauhan and Garg, 2013] Chauhan, H. and Garg, V. K. (2013). Democratic elections in faulty distributed systems. In *International Conference on Distributed Computing and Networking*, pages 176–191. Springer.

[Dellarocas, 2000] Dellarocas, C. (2000). Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, pages 150–157.

[Dolden, 2021] Dolden, L. (2021). Facebook removed over 15 billion fake accounts in 2 years. *Tech Round*.

[Dwork et al., 2014] Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

[Edelson et al., 2020] Edelson, L., Lauinger, T., and McCoy, D. (2020). A security analysis of the facebook ad library. In *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*, pages 661–678. IEEE.

[Emerson, 2013] Emerson, P. (2013). The original borda count and partial voting. *Social Choice and Welfare*, 40(2):353–358.

[Endriss, 2017] Endriss, U. (2017). *Trends in computational social choice.* Lulu. com.

[Fang et al., 2018] Fang, M., Yang, G., Gong, N. Z., and Liu, J. (2018). Poisoning attacks to graph-based recommender systems. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pages 381–392.

[Felfernig et al., 2018a] Felfernig, A., Boratto, L., Stettinger, M., and Tkalčič, M. (2018a). Algorithms for group recommendation. In *Group recommender systems*, pages 27–58. Springer.

[Felfernig et al., 2018b] Felfernig, A., Boratto, L., Stettinger, M., and Tkalčič, M. (2018b). *Group recommender systems: An introduction*. Springer.

[Fotakis et al., 2021] Fotakis, D., Kalavasis, A., and Stavropoulos, K. (2021). Aggregating incomplete and noisy rankings. In *International Conference on Artificial Intelligence and Statistics*, pages 2278–2286. PMLR.

[Fouss et al., 2007] Fouss, F., Pirotte, A., Renders, J.-M., and Saerens, M. (2007). Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on knowledge and data engineering*, 19(3):355–369.

[Hoang, 2017] Hoang, L. N. (2017). Strategy-proofness of the randomized condorcet voting system. *Soc. Choice Welf.*, 48(3):679–701.

[Imber et al., 2022] Imber, A., Israel, J., Brill, M., and Kimelfeld, B. (2022). Approval-based committee voting under incomplete information. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5076–5083.

[Kemeny and Snell, 1962] Kemeny, J. G. and Snell, L. (1962). Preference ranking: an axiomatic approach. *Mathematical models in the social sciences*, pages 9–23.

[Lamport et al., 1982] Lamport, L., Shostak, R. E., and Pease, M. C. (1982). The byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401.

[Lemire, 2005] Lemire, D. (2005). Scale and translation invariant collaborative filtering systems. *Information Retrieval*, 8(1):129–150.

[List, 2013] List, C. (2013). Social choice theory.

[Liu and Palomar, 2019] Liu, J. and Palomar, D. P. (2019). Regularized robust estimation of mean and covariance matrix for incomplete data. *Signal Processing*, 165:278–291.

[Lu and Boutilier, 2013] Lu, T. and Boutilier, C. (2013). Multi-winner social choice with incomplete preferences. In *IJCAI*.

[Masthoff, 2003] Masthoff, J. (2003). Modeling the multiple people that are me. In *International Conference on User Modeling*, pages 258–262. Springer.

[Meir et al., 2022] Meir, R., Talmon, N., Shahaf, G., and Shapiro, E. (2022). Sybil-resilient social choice with low voter turnout. In *European Conference on Multi-Agent Systems*, pages 257–274. Springer.

[Melnyk et al., 2018] Melnyk, D., Wang, Y., and Wattenhofer, R. (2018). Byzantine preferential voting. In *International Conference on Web and Internet Economics*, pages 327–340. Springer.

[Minsker, 2015] Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli*, 21(4):2308 – 2335.

[Moore, 2023] Moore, M. (2023). Fake accounts on social media, epistemic uncertainty and the need for an independent auditing of accounts. *Internet Policy Review*, 12(1).

[Moreno-Centeno and Escobedo, 2016] Moreno-Centeno, E. and Escobedo, A. R. (2016). Axiomatic aggregation of incomplete rankings. *IIE Transactions*, 48(6):475–488.

[Morgenthaler, 2007] Morgenthaler, S. (2007). A survey of robust statistics. *Statistical Methods and Applications*, 15(3):271–293.

[Negahban et al., 2012] Negahban, S., Oh, S., and Shah, D. (2012). Iterative ranking from pair-wise comparisons. *Advances in neural information processing systems*, 25.

[Neudert et al., 2019] Neudert, L.-M., Howard, P., and Kollanyi, B. (2019). Sourcing and automation of political news and information during three european elections. *Social Media+ Society*, 5(3):2056305119863147.

[Pennock et al., 2000] Pennock, D. M., Horvitz, E., Giles, C. L., et al. (2000). Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *AAAI/IAAI*, pages 729–734.

[Resnick et al., 1994] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186.

[Resnick and Varian, 1997] Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):56–58.

[Satariano, 2021] Satariano, A. (2021). Inside a pro-huawei influence campaign. *The New York Times*.

[Schaurer and Weiß, 2020] Schaurer, I. and Weiß, B. (2020). Investigating selection bias of online surveys on coronavirus-related behavioral outcomes. In *survey research methods*, volume 14, pages 103–108.

[Schulze, 2003] Schulze, M. (2003). A new monotonic and clone-independent single-winner election method. *Voting matters*, 17(1):9–19.

[Sen, 1999] Sen, A. (1999). The possibility of social choice. *American economic review*, 89(3):349–378.

[Si and Li, 2020] Si, M. and Li, Q. (2020). Shilling attacks against collaborative recommender systems: a review. *Artificial Intelligence Review*, 53(1):291–319.

[Smith, 2000] Smith, W. D. (2000). Range voting. *The paper can be downloaded from the author's homepage at http://www. math. temple. edu/˜ wds/homepage/works. html.*

[Su and Khoshgoftaar, 2009] Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.

[Tran et al., 2009] Tran, D. N., Min, B., Li, J., and Subramanian, L. (2009). Sybil-resilient online content voting. In Rexford, J. and Sirer, E. G., editors, *USENIX, NSDI 2009*, pages 15–28. USENIX Association.

[Tseng, 2017] Tseng, L. (2017). Voting in the presence of byzantine faults. In *PRDC*, pages 1–10. IEEE.

[Von Neumann and Morgenstern, 1953] Von Neumann, J. and Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton university press.

[Wang and Shah, 2019] Wang, J. and Shah, N. B. (2019). Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings. In Elkind, E., Veloso, M., Agmon, N., and Taylor, M. E., editors, *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, pages 864–872. International Foundation for Autonomous Agents and Multiagent Systems.

[Woolley, 2020] Woolley, S. (2020). *The Reality Game: A gripping investigation into deepfake videos, the next wave of fake news and what it means for democracy*. Hachette UK.

[Xie and Lui, 2015] Xie, H. and Lui, J. C. (2015). Mathematical modeling and analysis of product rating with partial information. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(4):1–33.

[Xiong and Liu, 2004] Xiong, L. and Liu, L. (2004). Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE transactions on Knowledge and Data Engineering*, 16(7):843–857.

[Yan et al., 2013] Yan, S., Zheng, X., Chen, D., and Wang, Y. (2013). Exploiting two-faceted web of trust for enhanced-quality recommendations. *Expert systems with applications*, 40(17):7080–7095.

[Yu et al., 2009] Yu, H., Shi, C., Kaminsky, M., Gibbons, P. B., and Xiao, F. (2009). Dsybil: Optimal sybil-resistance for recommendation systems. In *2009 30th IEEE Symposium on Security and Privacy*, pages 283–298. IEEE.

[Yue, 2019] Yue, N. (2019). The "weaponization" of facebook in myanmar: A case for corporate criminal liability. *Hastings LJ*, 71:813.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [**Yes**/No/Not Applicable]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [**Yes**/No/Not Applicable]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [**Yes**/No/Not Applicable]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [**Yes**/No/Not Applicable]

   (b) Complete proofs of all theoretical results. [**Yes**/No/Not Applicable]

   (c) Clear explanations of any assumptions. [**Yes**/No/Not Applicable]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [**Yes**/No/Not Applicable]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [**Yes**/No/Not Applicable]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [**Yes**/No/Not Applicable]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [**Yes**/No/Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Yes/No/**Not Applicable**]

   (b) The license information of the assets, if applicable. [Yes/No/**Not Applicable**]

   (c) New assets either in the supplemental material or as a URL, if applicable. [**Yes**/No/Not Applicable]

   (d) Information about consent from data providers/curators. [Yes/No/**Not Applicable**]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Yes/No/**Not Applicable**]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Yes/No/**Not Applicable**]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Yes/No/**Not Applicable**]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Yes/No/**Not Applicable**]

Organization of the Appendix

Additional related work on recommender systems and robust voting is in Appendix A. Missing proofs are in appendices C to F. Appendix J presents additional experiments on MEHESTAN. Appendix I exposes the difficulty of sparse voting by proving the impossibility of sparse unanimity for individually scaled preferences. Appendix K extends MEHESTAN to guarantee desirable properties such as differential privacy [Dwork et al., 2014].

A   Additional Related Work

Social choice theory.   This area [List, 2013, Endriss, 2017] is concerned with combining individual preferences into a collective aggregate, to facilitate collective decisions. An important result from social choice theory is Arrow's Impossibility [Arrow, 1950], stating that dictatorship is the only ordinal voting mechanism (i.e., voters only provide rankings instead of scores) that satisfies desirable fairness criteria. Fortunately, it is possible to escape this pessimistic result by including additional information from voters [Sen, 1999] (e.g., range voting [Smith, 2000]). As such, in our work, we assume that voters provide real-valued scores to alternatives (i.e., *cardinal* voting). However, we do not require the voters to score all alternatives, nor to be all honest (i.e., non-malicious).

Group recommender systems. An important use case of our work is that of group recommender systems (GRS). In contrast to a single-user recommender system (SRS) [Resnick and Varian, 1997], a GRS [Felfernig et al., 2018b] provides recommendations aimed at a group of people, rather than an individual. In practice, any algorithm used for group recommendation faces the problem of preference aggregation at some point. A categorization of the aggregation functions used in GRS, all inspired from social choice theory, was proposed in [Felfernig et al., 2018a]. When the inputs of a GRS are sparse real-valued lists of ratings, especially when there could be malicious users, then our voting system can be used effectively for GRS.

In fact, there are situations where the preference aggregation techniques used in GRS can be employed to solve problems encountered in SRS. Two such situations were discussed in [Masthoff, 2003]: (i) the cold-start problem, that is when new users join the system and the SRS has no previous data on them, and (ii) when there are multiple criteria rated for each alternative. Interestingly, single-user recommendation is another potential application of our mechanism.

The scaling problem.   When voters provide numerical ratings, preference aggregation functions need to consider the difference in users' internal rating scale [Pennock et al., 2000, Lemire, 2005, Yan et al., 2013]. Scale invariance is then a desirable property for preference aggregation functions. Interestingly, it was shown in [Lemire, 2005] that some scale-invariant collaborative filtering algorithms outperform their non-invariant counterparts. The attempt to verify the scale invariance property may explain why collaborative filtering based recommender systems usually use weighted averaging to aggregate preferences [Pennock et al., 2000]. This clearly makes such scaling-invariant solutions very vulnerable to Byzantine attacks. A significant amount of previous work proposed solutions to address the scaling problem. Scale-invariant similarity metrics such as Pearson's correlation were leveraged in [Lemire, 2005, Su and Khoshgoftaar, 2009]. Various normalization techniques were used in [Resnick et al., 1994, Lemire, 2005, Xie and Lui, 2015] to bring users' ratings to a single global scale. Section I however shows that individual-based normalization techniques fail to solve the scaling problem in a sparse setting.

Content recommendation with adversaries. Vanilla recommender systems are not immune to adversaries. For example, collaborative filtering is known to be vulnerable to shilling attacks [Si and Li, 2020], namely fake profiles and bogus ratings injection. Meanwhile, graph-based recommender systems [Fouss et al., 2007] were shown in [Fang et al., 2018] to recommend a target item a hundred times more, when only 1% of the total users were injected fake profiles. Most prior work investigated ways to defend against such attacks by leveraging trust (between voters). For instance, trust network was used in [Tran et al., 2009] to limit the amount of votes collected from fake users via the max-flow concept. However, it is not clear how the quality of the vote can be evaluated. A trust-based SRS was developed in [Andersen et al., 2008] (with binary recommendations) based on random walks that are incentive compatible [Yu et al., 2009].

Distributed Byzantine voting. Voting in the presence of a limited fraction of Byzantine voters has been addressed recently in the distributed setting. This line of research tackles a generalization of the Byzantine agreement problem [Lamport et al., 1982] in distributed computing. For example, the algorithms proposed

in [Chauhan and Garg, 2013, Tseng, 2017] yield the single winner of an election held in the presence of a fraction of Byzantine voters, using the plurality voting rule. Voters' rankings were used in [Melnyk et al., 2018] to output an aggregated ranking leveraging the Kemeny rule [Kemeny and Snell, 1962], while verifying a correctness condition. Sparse unanimity (Definition 2) is closely related to the correctness properties sought in the aforementioned works. However, in contrast, our voting mechanism tolerates sparse inputs and can utilize scores (and not just rankings). Finally, instead of considering a constant fraction of Byzantine players, we seek L-Lipschitz resilience (Definition 1).

B Preliminaries

B.1 Mathematical Reminders

In this section, we recall the notions of Lipschitz continuity and strong convexity. We define these notions for a general multidimensional space, although we use them for the particular one-dimensional case in the paper. Let $d \in \mathbb{N}$ be the dimension of the space $\mathbb{R}^d$. Denote by $\langle \cdot, \cdot \rangle$ the Euclidean scalar product in $\mathbb{R}^d$, and by $\|\cdot\|_2$ the Euclidean norm.

Definition 4 (Lipschitz continuity)**.** Let $L \geq 0$, and consider two metric spaces (X, d_X) and (Z, d_Z). A function $g \colon X \to Z$ is L-Lipschitz continuous, if

$$\forall x, y \in X, \ d_Z(g(x), g(y)) \leq L d_X(x, y).$$

Definition 5 (Strong convexity)**.** Let $\mu \geq 0$. A subdifferentiable function $g \colon \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if for every $x, y \in \mathbb{R}^d$, and any subgradients $h_x \in \nabla g(x)$ and $h_y \in \nabla g(y)$, it holds that

$$\langle h_x - h_y, x - y \rangle \geq \mu \|x - y\|_2^2.$$

B.2 Generalization to Continuous Voting Rights

In this section, we generalize the setting of the main part of the paper to include continuous voting rights. In addition to address a more general setting, which can be useful in applications [Beylerian et al., 2022], accounting for continuous voting rights yields an arguably more natural insight into the concept of Lipschitz resilience.

We start by formalizing voting rights by an assignment of a nonnegative weight $w_n \geq 0$ to each voter n. We denote $\boldsymbol{w} = (w_1, \ldots, w_N)$ the tuple of voting rights. Interesting Lipschitz resilience is then defined, for fixed inputs $\boldsymbol{\theta}$, a Lipschitz continuity of the map $\boldsymbol{w} \mapsto \text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta})$, from $(\mathbb{R}^N, \|\cdot\|_1)$ to $(\mathbb{R}^A, \|\cdot\|_\infty)$. Put differently, we define Lipschitz resilience as follows.

Definition 6. A vote $\text{VOTE} \colon \mathbb{R}^N \times X^N \to \mathbb{R}^A$ is L-Lipschitz resilient if,

$$\forall \boldsymbol{\theta} \in X^N, \ \forall \boldsymbol{w}, \boldsymbol{w}' \in \mathbb{R}^N, \ \|\text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta}) - \text{VOTE}(\boldsymbol{w}', \boldsymbol{\theta})\|_\infty \leq L \|\boldsymbol{w} - \boldsymbol{w}'\|_1. \tag{9}$$

Note that, assuming that not reporting a score is tantamount to having zero voting right (which is the case for all our algorithms), this definition clearly generalizes the one in the main part of the paper, which amounts to flipping $w_n = 1$ to $w_n = 0$.

Let us also formalize the generalization of sparse unanimity to varying voting rights. Essentially, the only modified condition is the N_0-scored, which we replace by a W_0-scored condition that essentially say that the set of voters that scored an alternative must have cumulative voting right of at least W_0.

Definition 7. An input $(\boldsymbol{w}, \boldsymbol{\theta})$ is W_0-scored if, for any alternative $a \in [A]$, we have $\sum_{n \in N_a} w_n \geq W_0$.

Definition 8. A vote is sparsely unanimous if, for all $\theta_* \in \mathbb{R}^A$, there exists $W_0 \geq 0$ such that, whenever the inputs $(\boldsymbol{w}, \boldsymbol{\theta})$ are θ_*-unanimous, comparable and W_0-scored, we have $\text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta}) \sim \theta_*$.

C Proof of Proposition 1: The Mean and Median are not Resilient

The mean and median are well-known to be generalizable to varying voting rights, yielding the weighted mean and the weighted median. Note that the median may be ill-defined, as there may be multiple real numbers satisfying the corresponding properties. In this case, we define $\text{MED}(\boldsymbol{w}, \boldsymbol{\theta})$ to be the one closest to zero, which can be easily proven to be unique. We recall the statement of Proposition 1 below for convenience.

Proposition 1. MEAN *and* MED *are not Lipschitz resilient.*

Proof. First, we prove that MEAN is arbitrarily manipulable by any voter with a positive voting right. More precisely, for any voter $f \in [N]$ with voting right $w_f > 0$, for any voting right w_n and score θ_n for voters $n \neq f$, and for any target score $x \in \mathbb{R}$, there exists a malicious score reporting $\theta_f \in \mathbb{R}$ such that $\text{MEAN}(\boldsymbol{w}, \boldsymbol{\theta}) = x$. Indeed, it suffices to consider

$$\theta_f = \frac{1}{w_f} \sum_{n \in [N]} w_n x - \frac{1}{w_f} \sum_{n \neq f} w_n \theta_n.$$

Arbitrare manipulability then clearly implies that MEAN cannot be Lipschitz resilient.

Second, we prove the assertion for MED. The proposition is trivial in the case $N = 1$, as MED is then arbitrarily manipulable by the single voter. But to provide further insight into Lipschitz resilience, we prove it in the case where the number of voters is very large. Let $L > 0$ and $N_0 \in \mathbb{N}$. Consider $N \triangleq 2N_0 + 1$, with $w_n = 1$ for all $n \in [N]$. Assume moreover that $\theta_n \triangleq 0$ for $n \in [N_0]$, and $\theta_n \triangleq 2L$ for $n \in \{N_0 + 1, N_0 + 2, \ldots, 2N_0\}$. Now define $f \triangleq 2N_0 + 1$, i.e. the malicious voter is the last voter. Then the median without accounting for voter f is 0. But by reporting $\theta_f \triangleq 2L$, then $\text{MED}(\boldsymbol{w}, \boldsymbol{\theta}) = 2L$. Thus, the intervention of the malicious voter modifies the output score by $2L$, which is strictly larger than L. Thus MED fails to be L-Lipschitz resilient, for any value of L. This proves the proposition. □

D   Proof of Theorem 1: QRMED is Lipschitz Resilient

Let $L > 0$. We generalize QRMED_L to varying voting rights as follows:

$$\text{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \underset{z \in \mathbb{R}}{\arg\min} \left\{ \mathcal{L}_{\text{QRMED}_L}(z|\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \frac{1}{2L} z^2 + \sum_{n \in N^*} w_n |z - \theta_n| . \right\}. \tag{10}$$

We recall Theorem 1 below for convenience.

Theorem 1. QRMED_L *is well-defined and L-Lipschitz resilient.*

Proof. Consider any inputs $(\boldsymbol{w}, \boldsymbol{\theta})$ and $(\boldsymbol{w}', \boldsymbol{\theta})$. Denote $q \triangleq \text{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})$ and $q' \triangleq \text{QRMED}_L(\boldsymbol{w}', \boldsymbol{\theta})$, and let $\Delta_w \triangleq \|w - w'\|_1$ and $\Delta_q \triangleq |q' - q|$. We aim to prove that we must have $\Delta_q \leq L\Delta_w$.

Denote $\ell(z) \triangleq \mathcal{L}_{\text{QRMED}_L}(z|\boldsymbol{w}, \boldsymbol{\theta})$ and $\ell'(z) \triangleq \mathcal{L}_{\text{QRMED}_L}(z|\boldsymbol{w}', \boldsymbol{\theta})$, Note that their difference is

$$\ell(z) - \ell'(z) = \sum_{n \in N^*} (w_n - w'_n) |z - \theta_n|. \tag{11}$$

The subderivatives of this difference can thus be bounded by

$$\sup |\partial(\ell - \ell')(z)| \leq \sum_{n \in N^*} |w_n - w'_n| \sup \text{sign}(z - \theta_n) \leq \sum_{n \in N^*} |w_n - w'_n| = \|w - w'\|_1 = \Delta_w. \tag{12}$$

Thus, for any $g' \in \partial\ell'(z)$, there exists $g \in \partial\ell(z)$ such that $|g - g'| \leq \Delta_w$.

Now, without loss of generality, assume that $q' > q$ (the case $q^{-F} < q$ can be treated similarly). Now note that, since q minimizes ℓ, we must have $0 \in \partial\ell(q)$. Thus, in particular, $\sup \partial\ell(q) \geq 0$.

Similarly, we have $\inf \partial\ell'(q') \leq 0$. Using what we have seen above, this implies that there must be $g \in \partial\ell(q')$ such that $|g - \inf \partial\ell'(q')| \leq \Delta_w$, which implies $g \leq \Delta_w$. In particular, $\inf \partial\ell(q') \leq \Delta_w$.

Now, since ℓ is a sum of a quadratic term with a coefficient $(1/2L)$, and of a convex term, it is clearly $(1/L)$-strongly convex (see Definition 5). Therefore we have $(\inf \partial\ell(q') - \sup \partial\ell(q))(q' - q) \geq (q' - q)^2/L$, which implies $\inf \partial\ell(q') \geq \sup \partial\ell(q) + \Delta_q/L \geq \Delta_q/L$. Since we already showed that the left-hand side is at most Δ_w, it follows that $\Delta_q \leq L\Delta_w$, which is what was needed. □

E Proof of Theorem 2: LRMEAN is Resilient

In this section, we provide the complete proof of the L-Lipschitz resilience of LRMEAN. First, we prove auxiliary properties on the clipped mean CLMEAN operator, which we generalize to varying voting rights as follows:

$$\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) \triangleq \text{MEAN}(\boldsymbol{w}, \text{CLIP}(\boldsymbol{\theta}|\mu, \Delta)) = \frac{1}{\|\boldsymbol{w}\|_1} \sum_{n \in [N]} w_n \text{CLIP}(\theta_n|\mu, \Delta),$$

where $\text{CLIP}(x|\mu, \Delta) \triangleq \max\{\mu - \Delta, \min\{\mu + \Delta, x\}\}$ clips x within the interval $[\mu - \Delta, \mu + \Delta]$.

Lemma 1. CLIP *is 1-Lipschitz continuous with respect to the radius, i.e.*

$$\forall x, \mu, \Delta, \Delta' \in \mathbb{R}, \quad |\text{CLIP}(x|\mu, \Delta) - \text{CLIP}(x|\mu, \Delta')| \le |\Delta - \Delta'|. \tag{13}$$

Proof. Let $x, \mu, \Delta, \Delta' \in \mathbb{R}$. Without loss of generality, assume $\Delta' \ge \Delta$. If $x \in [\mu - \Delta, \mu + \Delta]$, then $x \in [\mu - \Delta', \mu + \Delta']$, and hence $\text{CLIP}(x|\mu, \Delta) = x = \text{CLIP}(x|\mu, \Delta')$. Thus, the statement clearly holds. Otherwise, if $x > \mu + \Delta$, then $\text{CLIP}(x|\mu, \Delta) = \mu + \Delta$ and $\text{CLIP}(x|\mu, \Delta') = \min\{\mu + \Delta', x\} \in [\mu + \Delta, \mu + \Delta']$. We then have $|\text{CLIP}(x|\mu, \Delta) - \text{CLIP}(x|\mu, \Delta')| \le (\mu + \Delta') - (\mu + \Delta) = |\Delta - \Delta'|$. The case $x < \mu + \Delta$ is treated similarly. This concludes the proof. $\square$

Lemma 2. CLMEAN *is 1-Lipschitz continuous with respect to the radius, i.e.*

$$\forall \boldsymbol{w}, \boldsymbol{\theta}, \ \forall \mu, \Delta, \Delta' \in \mathbb{R}, \quad |\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta')| \le |\Delta - \Delta'|. \tag{14}$$

Proof. Let $\mu, \Delta, \Delta' \in \mathbb{R}$. By triangle inequality and Lemma 1, we have

$$|\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta')| =$$

$$\left| \frac{1}{\|\boldsymbol{w}\|_1} \sum_{n \in [N]} w_n \text{CLIP}(\theta_n|\mu, \Delta) - \frac{1}{\|\boldsymbol{w}\|_1} \sum_{n \in [N]} w_n \text{CLIP}(\theta_n|\mu, \Delta') \right|$$

$$\le \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\text{CLIP}(\theta_n|\mu, \Delta) - \text{CLIP}(\theta_n|\mu, \Delta')|$$

$$\le \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\Delta - \Delta'| = |\Delta - \Delta'|.$$

This concludes the proof. $\square$

Lemma 3. CLIP *is 1-Lipschitz continuous with respect to the center, i.e.*

$$\forall x, \Delta, \mu, \mu' \in \mathbb{R}, \quad |\text{CLIP}(x|\mu, \Delta) - \text{CLIP}(x|\mu', \Delta)| \le |\mu - \mu'|. \tag{15}$$

Proof. Let $x, \mu, \mu', \Delta \in \mathbb{R}$. Without loss of generality, assume $\mu \le \mu'$.

Suppose for now that $\mu + \Delta \le \mu' - \Delta$. We then consider the five (possibly empty) intervals $(-\infty, \mu - \Delta]$, $[\mu - \Delta, \mu + \Delta]$, $[\mu + \Delta, \mu' - \Delta]$, $[\mu' - \Delta, \mu' + \Delta]$ and $[\mu' + \Delta, +\infty)$, which cover $\mathbb{R}$. If x is in the first interval, then $\text{CLIP}(x|\mu, \Delta) = \mu - \Delta$ and $\text{CLIP}(x|\mu', \Delta) = \mu' - \Delta$, and their absolute difference is equal to $|\mu - \mu'|$. If x is in the second or third interval, we have $\text{CLIP}(x|\mu, \Delta) \ge \mu - \Delta$ and $\text{CLIP}(x|\mu', \Delta) = \mu' - \Delta$, and their absolute difference is thus at most $|\mu - \mu'|$. If x is in the fourth interval, then we have $\text{CLIP}(x|\mu, \Delta) = \mu + \Delta$ and $\text{CLIP}(x|\mu', \Delta) \le \mu' + \Delta$, and their absolute difference is thus at most $|\mu - \mu'|$. Finally, the fifth interval is akin to the first interval.

Now assume that $\mu + \Delta \ge \mu' - \Delta$. We now consider the five (possibly empty) intervals $(-\infty, \mu - \Delta]$, $[\mu - \Delta, \mu' - \Delta]$, $[\mu' - \Delta, \mu + \Delta]$, $[\mu + \Delta, \mu' + \Delta]$ and $[\mu' + \Delta, +\infty)$, which cover $\mathbb{R}$. The first and fifth intervals are still treated similarly as before. Now assume that x is in the second interval. Then $\text{CLIP}(x|\mu, \Delta) \ge \mu - \Delta$ and $\text{CLIP}(x|\mu', \Delta) = \mu' - \Delta$, which implies that their absolute difference is at most $|\mu - \mu'|$. The fourth interval is treated symmetrically. Finally, when x is in the third interval, then $\text{CLIP}(x|\mu, \Delta) = \text{CLIP}(x|\mu', \Delta)$, and their absolute difference is zero. In all cases, the inequality of the lemma holds. This concludes the proof. $\square$

Lemma 4. CLMEAN *is 1-Lipschitz continuous with respect to the center, i.e.*

$$\forall \boldsymbol{w}, \boldsymbol{\theta}, \ \forall \Delta, \mu, \mu' \in \mathbb{R}, \ |\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu', \Delta)| \leq |\mu - \mu'| . \tag{16}$$

Proof. Let $\Delta, \mu, \mu' \in \mathbb{R}$. By triangle inequality and Lemma 3, we have

$$|\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu', \Delta)| =$$

$$\left| \frac{1}{\|\boldsymbol{w}\|_1} \sum_{n \in [N]} w_n \text{CLIP}(\theta_n|\mu, \Delta) - \frac{1}{\|\boldsymbol{w}\|_1} \sum_{n \in [N]} w_n \text{CLIP}(\theta_n|\mu', \Delta) \right|$$

$$\leq \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\text{CLIP}(\theta_n|\mu, \Delta) - \text{CLIP}(\theta_n|\mu', \Delta)|$$

$$\leq \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\mu - \mu'| = |\mu - \mu'| .$$

This concludes the proof. $\qquad\square$

Lemma 5. *If $\boldsymbol{v} \geq 0$, then $|\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w} + \boldsymbol{v}, \boldsymbol{\theta}|\mu, \Delta)| \leq 2\frac{\|\boldsymbol{v}\|_1}{\|\boldsymbol{w}\|_1}\Delta$.*

Proof. Denote $\boldsymbol{y} \triangleq \text{CLIP}(\boldsymbol{\theta}|\mu, \Delta)$, and $\bar{y} \triangleq \text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta)$. Then, we have

$$|\text{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta}|\mu, \Delta) - \text{CLMEAN}(\boldsymbol{w} + \boldsymbol{v}, \boldsymbol{\theta}|\mu, \Delta)|$$

$$= \left| (\bar{y} - \mu) - \frac{1}{\|\boldsymbol{w} + \boldsymbol{v}\|_1} \sum_{n \in [N]} (w_n + v_n)(y_n - \mu) \right|$$

$$= \left| (\bar{y} - \mu) - \frac{\|\boldsymbol{w}\|_1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} (\bar{y} - \mu) - \frac{1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \sum_{n \in [N]} v_n(y_n - \mu) \right|$$

$$\leq \left(1 - \frac{\|\boldsymbol{w}\|_1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \right) |\bar{y} - \mu| + \frac{1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \sum_{n \in [N]} v_n |y_n - \mu|$$

$$\leq \left(1 - \frac{\|\boldsymbol{w}\|_1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \right) \Delta + \frac{1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \sum_{n \in [N]} v_n \Delta$$

$$\leq \frac{2\|\boldsymbol{v}\|_1}{\|\boldsymbol{w}\|_1 + \|\boldsymbol{v}\|_1} \Delta \leq \frac{2\|\boldsymbol{v}\|_1}{\|\boldsymbol{w}\|_1}\Delta,$$

which concludes the proof. $\qquad\square$

Let $L > 0$. We generalize LRMEAN_L to varying voting rights as follows:

$$\text{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \text{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} \left| \text{QRMED}_{L/4}(\boldsymbol{w}, \boldsymbol{\theta}), \frac{\|L\boldsymbol{w}\|_1}{4} \right. \right).$$

We finally recall and prove Theorem 2 below.

Theorem 2. LRMEAN_L *is L-resilient. Moreover, if there exists $\Delta > 0$ such that $\|\boldsymbol{w}\|_1 \geq 8\Delta/L$ and $\theta_n \in [-\Delta, \Delta]$ for all n, then $\text{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta}) = \text{MEAN}(\boldsymbol{w}, \boldsymbol{\theta})$.*

Proof. Denote $q_{-F} \triangleq \text{QRMED}_{L/4}(\boldsymbol{w}_{-F}, \boldsymbol{\theta})$ and $q \triangleq \text{QRMED}_{L/4}(\boldsymbol{w}, \boldsymbol{\theta})$. Then, using the triangle inequality, we

have

$$
\begin{aligned}
&\left| \text{LrMean}_L(\boldsymbol{w}_{-F}, \boldsymbol{\theta}) - \text{LrMean}_L(\boldsymbol{w}, \boldsymbol{\theta}) \right| \\
&= \left| \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q_{-F}, \frac{L\|\boldsymbol{w}_{-F}\|_1}{4} \right) - \text{ClMean}\left(\boldsymbol{w}, \boldsymbol{\theta} \,\middle|\, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq \left| \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q_{-F}, \frac{L\|\boldsymbol{w}_{-F}\|_1}{4} \right) - \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q_{-F}, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\quad + \left| \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q_{-F}, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\quad + \left| \text{ClMean}\left(\boldsymbol{w}_{-F}, \boldsymbol{\theta} \,\middle|\, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \text{ClMean}\left(\boldsymbol{w}, \boldsymbol{\theta} \,\middle|\, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq \frac{L\|\boldsymbol{w} - \boldsymbol{w}_{-F}\|_1}{4} + |q_{-F} - q| + 2\frac{\|\boldsymbol{w}_F\|_1}{\|\boldsymbol{w}\|_1}\frac{L\|\boldsymbol{w}\|_1}{4} \leq L\|\boldsymbol{w}_F\|_1\,,
\end{aligned}
$$

where we used lemmas 2, 4 and 5 successively.

Now, we prove the second part of the theorem. Assume that there exists $\Delta > 0$ such that $\|\boldsymbol{w}\|_1 \geq 8\Delta/L$ and $\theta_n \in [-\Delta, \Delta]$ for all n. By the latter assumption, observe that we have $q \in [-\Delta, \Delta]$, as one can check that the subderivatives of the loss minimized by QrMed are positive at Δ and negative at $-\Delta$ (similar to the proof of Theorem 1). Therefore, using the fact that $\|\boldsymbol{w}\|_1 \geq 8\Delta/L$, the clipping interval $[q - \frac{L\|\boldsymbol{w}\|_1}{4}, q + \frac{L\|\boldsymbol{w}\|_1}{4}]$ then contains all of $[-\Delta, \Delta]$, and thus also all voter scores. Therefore, under these conditions, LrMean_L returns the mean. This concludes the proof. $\qquad\square$

F   Proof of Theorem 3

For convenience, we restate Theorem 3, proof of which is decomposed in lemmas in the next section.

Theorem 3. Mehestan_L *is L-Lipschitz resilient and sparsely unanimous.*

F.1   Overview of the Proof

Below, we sketch the proof of Theorem 3 in a guided proof with high-level intuitions. We first discuss resilience, which requires controlling the worst-case impact of malicious voters at each step of Mehestan. We then prove the sparse unanimity guarantee. The missing full proofs can be found in Appendix F.2.

F.1.1 Mehestan is Resilient

We now sketch the proof of the resilience of Mehestan. First, note that LrMean ensures resilience in the search of the scaling and translation factors (steps 2 and 3 in Section 4). The guarantee a priori depends on the maximum score of each input vector $\|\theta_n\|_\infty$, but the prior use of min-max pre-normalization allows to remove the dependency on this quantity.

Lemma 6. *For any subset $F \subset [N]$, denote s_n^{-F} the scaling obtained by involving only the voters $n \notin F$. Then, for any alternative $a \in [A]$, we have $\left| s_n \tilde{\theta}_{na} - s_n^{-F} \tilde{\theta}_{na} \right| \leq \frac{L\|\boldsymbol{w}^F\|_1}{7}$.*

Proof. By the $(L/7)$-resilience of $\text{LrMean}_{L/7}$ shown in Theorem 2, we have $\left| s_n - s_n^{-F} \right| \leq L\|\boldsymbol{w}^{-F}\|_1/7 \left\|\tilde{\theta}_n\right\|_\infty$. It follows that, for any alternative $a \in [A]$, we have $\left| s_n \tilde{\theta}_{na} - s_n^{-F} \tilde{\theta}_{na} \right| = \left| s_n - s_n^{-F} \right| \left| \tilde{\theta}_{na} \right| \leq \left| s_n - s_n^{-F} \right| \left\|\tilde{\theta}_n\right\|_\infty \leq \frac{L\|\boldsymbol{w}^F\|_1}{7}$. $\qquad\square$

Lemma 7. *For any subset $F \subset [N]$, denote τ_n^{-F} the translation obtained by involving only the voters $n \notin F$. Then, for any alternative $a \in [A]$, we have $\left| \tau_n - \tau_n^{-F} \right| \leq \frac{5L\|\boldsymbol{w}^F\|_1}{7}$.*

Proof sketch (The full proof is in Appendix F.2). One complication is that malicious voters $f \in F$ affect the comparative translation factors τ_{nm} by affecting the scaling factors s_n and s_m that appear in their computations,

even when $m \notin F$. Fortunately, combining Lemma 6, Proposition 3 allows to bound this impact. Combining this to Theorem 2 for the direct impact of malicious voters through τ_{nf} allows to conclude. □

Combining our two lemmas above, and Theorem 1 on the resilience of QRMED used in step 4 in Section 4, guarantees the resilience of MEHESTAN.

Lemma 8. MEHESTAN$_L$ *is L-Lipschitz resilient.*

Proof sketch (The full proof is in Appendix F.2). The proof is akin to Lemma 7, by leveraging previous bounds, the Lipschitz continuity of QRMED (Proposition 2) and its resilience (Theorem 1). □

F.1.2 MEHESTAN is Sparsely Unanimous

We now sketch our proof of sparse unanimity, whose full proofs are provided in Appendix F.2. For any $\theta_* \in \mathbb{R}^A$, we first define the scaling bound

$$\bar{s}(\theta_*) \triangleq \frac{\max_{a,b} |\theta_{*a} - \theta_{*b}|}{\min_{a,b:\theta_{*a} \neq \theta_{*b}} |\theta_{*a} - \theta_{*b}|}, \tag{17}$$

which is trivially scale-invariant. This quantity bounds the largest possible multiplicative scaling between any voter's pre-normalized scores, and the min-max normalization $\tilde{\theta}_* \triangleq \text{MinMaxNorm}(\theta_*)$ of θ_*. Intuitively, it is an important quantity; the larger it is, the more voters will be needed to re-scale appropriately the voters' pre-normalized scores. Interestingly, as the following lemma states it, $\bar{s}(\theta)$ also bounds the translation discrepancies between the voters' pre-normalized scores and the min-max normalized scores $\tilde{\theta}_*$.

Lemma 9. *Suppose θ_*-unanimity and comparability. Then, for any $n \in [N]$, there must exist s_n^* and τ_n^* such that $\tilde{\theta}_n = s_n^* \tilde{\theta}_{*|A_n} + \tau_n^*$, with $1 \leq s_n^* \leq \bar{s}(\theta_*)$ and $-\bar{s}(\theta_*) \leq \tau_n^* \leq 0$.*

Proof sketch (the full proof is in Appendix F.2). Denoting a_n and b_n the best and worst alternatives scored by voter n, as opposed to the best and worst alternatives a and b according to θ_*, we can see that $s_n^* = \frac{\theta_{*a} - \theta_{*b}}{\theta_{*a_n} - \theta_{*b_n}} \in [1, \bar{s}(\theta_*)]$. The bound on τ_n^* is then obtained by looking at the normalized score of b_n. □

The previous lemma says that the scaling and translation factors s_n^* and τ_n^* that must be learned for each voter are bounded. Thus, this will also be the case of the relative scaling and translations s_{nm} and τ_{nm}. Therefore, assuming sufficiently many voting rights have been allocated, applying LRMEAN to such quantities will return a mean. Combining this observation with the linearity of the mean then allows to guarantee that the adequate scaling and translation will be inferred for all voters, as precisely proved by the following lemma.

Lemma 10. *Suppose θ_*-unanimity, comparability, and that alternatives are $(8\bar{s}(\theta_*)^2/L)$-scored. Then the voters' re-scaled scores are consistent, in the sense that $s_n \tilde{\theta}_{na} + \tau_n = s_m \tilde{\theta}_{ma} + \tau_m$, for all voters $n, m \in [N]$ and alternatives $a \in A_{nm}$ that both voters scored.*

Sketch of proof (the full proof is in Appendix F.2). By comparability, Lemma 9 and Theorem 2, $s_n = \text{MEAN}(\boldsymbol{w}, (s_m^*/s_n^*)_{m \in [N]}) = \text{MEAN}(\boldsymbol{w}, \vec{s^*})/s_n^*$. Thus $s_n s_n^*$, which is the overall multiplicative re-scaling compared to $\tilde{\theta}_*$, is independent of n. Similarly, Theorem 2 guarantees that $\tau_n = \text{MEAN}(\boldsymbol{w}, (s_m \tau_m^*)_{m \in [N]}) - s_n \tau_n^*$. Thus the overall translation, with respect to $\tilde{\theta}_*$, is independent of the voter n. □

It is noteworthy that the global multiplicative rescaling is a weighted average of the values s_n^*, which are known to be at least 1. Thus, under $(8\bar{s}(\theta_*)^2/L)$-scored condition, the global normalization step (see Algorithm 1) expands the scores' multiplicative scales. Aggregating correctly scaled unanimous preferences then allows to recover these preferences.

Lemma 11. *Under θ_*-unanimity, comparability, and that alternatives are $(8\bar{s}(\theta_*)^2/L)$-scored, θ_* is recovered, i.e. MEHESTAN$_L(\boldsymbol{\theta}) \sim \theta_*$. In particular, MEHESTAN$_L$ guarantees sparse unanimity, for $W_0 \triangleq 8\bar{s}(\theta_*)^2/L$.*

Proof. By Lemma 10, under θ_*-unanimity, comparability and $(8\bar{s}(\theta_*)^2/L)$-scored, we know that there exists $\vec{s^*} \triangleq \text{MEAN}(\boldsymbol{w}, \vec{s^*}) \in [1, \bar{s}(\theta_*)]$ and $\tilde{\tau}^* \triangleq \text{MEAN}(\boldsymbol{w}, (s_m \tau_m^*)_{m \in [N]}) \in [-\bar{s}(\theta_*)^2, 0]$ such that $s_n \tilde{\theta}_{na} + \tau_n = \vec{s^*} \tilde{\theta}_{*a} + \tilde{\tau}^* \in [-\bar{s}(\theta_*), \bar{s}(\theta_*)^2]$ for all voters n and alternatives $a \in A_n$. Now any alternative a has received at least $8\bar{s}(\theta_*)^2/L$ votes (measured in voting rights). We know that these votes are all identical and equal to $\vec{s^*} \tilde{\theta}_{*a} + \tilde{\tau}^*$. But then,

the optimality condition of QRMED_L shows that we must have $\text{QRMED}_L(\boldsymbol{w}, (s_n\tilde{\theta}_{na} + \tau_n)_{n\in[N]}) = \bar{s}^*\tilde{\theta}_{*a} + \tilde{\tau}^*$. Thus, for all alternatives $a \in [A]$, we must have $\text{MEHESTAN}_{L,a}(\boldsymbol{w}, \boldsymbol{\theta}) = \bar{s}^*\tilde{\theta}_{*a} + \tilde{\tau}^*$, which is a positive affine transformation of $\tilde{\theta}_*$ (and thus of θ_*). $\qquad\square$

F.2 Missing Proofs

In this section, we provide missing proofs of lemmas from the previous section.

Lemma 7. *Denote τ_n^{+f} the scaling by including inputs from a new voter $f \notin [N]$. Then, for any alternative $a \in [A]$, we have $\left|\tau_n^{+f} - \tau_n\right| \le \frac{5Lw_f}{7}$.*

Proof. Let $f \in F$ and $a \in [A]$. Let us introduce the following variable

$$t_n \triangleq \text{LRMEAN}_{L/7}\left\{w_m, t_{nm} \mid m \in N_n^A\right\}. \tag{18}$$

where $t_{nm} \triangleq \frac{1}{|A_{nm}|}(\sum_{a\in A_{nm}} s_m^{+f}\tilde{\theta}_{ma} - s_n^{+f}\tilde{\theta}_{na})$.

First, we will prove that

$$|\tau_n - t_n| \le \frac{4Lw_f}{7}. \tag{19}$$

For this, consider $\xi = \left[(s_m - s_m^{+f})\tilde{\theta}_{ma} - (s_n - s_n^{+f})\tilde{\theta}_{na}\right]_{\substack{m\in[N]\\a\in A_{nm}}}$.

We deduce from Lemma 6 that $\|\xi\|_\infty \le \frac{2Lw_f}{7}$. Now, by using the previous ξ in Proposition 3 of LRMEAN, we obtain Inequality (19). Second, by Proposition 3, we know that

$$\left|\tau_n^{+f} - t_n\right| \le \frac{Lw_f}{7}. \tag{20}$$

We conclude by using the triangular inequality and inequalities (19) and (20). $\qquad\square$

Lemma 8. MEHESTAN_L *is L-Lipschitz resilient.*

Proof. Let $f \in F$ and $a \in [A]$. Let us introduce the following variable

$$r_a \triangleq \text{QRMED}_{L/7}(\boldsymbol{w}, (s_n^{+f}\tilde{\theta}_n + \tau_n^{+f})_{n\in[N]}). \tag{21}$$

First, we will prove that

$$|r_a - \text{MEHESTAN}_L(\boldsymbol{w}, \boldsymbol{\theta})| \le \frac{6w_f}{7w}. \tag{22}$$

For this, consider $\xi = \left[(s_n - s_n^{+f})\tilde{\theta}_n + (\tau_n - \tau_n^{+f})\right]_{n\in[N]}$.

We deduce from propositions 6 and 7 that $\|\xi\|_\infty \le \frac{6Lw_f}{7}$. Now, by using the previous ξ in Proposition 2 (Lipschitz continuity of QRMED in score inputs), we obtain Inequality (22). Second, by Proposition 2, we know that

$$\left|r_a - \text{MEHESTAN}_L(w, \theta^{+f})\right| \le \frac{Lw_f}{7}. \tag{23}$$

We conclude by using the triangular inequality and inequalities (22) and (23). $\qquad\square$

Lemma 9. *Suppose θ_*-unanimity and comparability. Then, for any $n \in [N]$, there must exist s_n^* and τ_n^* such that $\tilde{\theta}_n = s_n^*\tilde{\theta}_{*|A_n} + \tau_n^*$, with $1 \le s_n^* \le \bar{s}(\theta_*)$ and $-\bar{s}(\theta_*) \le \tau_n^* \le 0$.*

Proof. Let $n \in [N]$. By θ_*-unanimity, we know that $\tilde{\theta}_n \sim \theta_n \sim \theta_{*|A_n} \sim \tilde{\theta}_{*|A_n}$. Thus there exists s_n^* and τ_n^* such that $\tilde{\theta}_n = s_n^*\tilde{\theta}_{*|A_n} + \tau_n^*$. Now consider a_n the best-scored alternative by voter n, and b_n their worst-scored alternative. By comparability, we have $C_{nm} \ne \emptyset$, which implies that there exists two alternatives $a, b \in A_n$ that voter n scored

differently. Therefore, $\theta_{na} > \theta_{nb}$. Then we must have $\tilde{\theta}_{na} > \tilde{\theta}_{nb}$. In particular, by min-max normalization, we then have $1 = \max_{c \in A_n} \tilde{\theta}_{nc} - \min_{c \in A_n} \tilde{\theta}_{nc} = \tilde{\theta}_{na} - \tilde{\theta}_{nb} = (s_n^* \tilde{\theta}_{*a} + \tau_n^*) - (s_m^* \tilde{\theta}_{*b} + \tau_m^*) = s_n^*(\tilde{\theta}_{*a} - \tilde{\theta}_{*b})$. Therefore $s_n^* = \frac{1}{\tilde{\theta}_{*a} - \tilde{\theta}_{*b}} = \frac{\max_{c,c'} |\tilde{\theta}_{*c} - \tilde{\theta}_{*c'}|}{\tilde{\theta}_{*a} - \tilde{\theta}_{*b}} \leq \bar{s}(\tilde{\theta}_*) = \bar{s}(\theta_*)$. Moreover, since by min-max normalization, we must also have $\tilde{\theta}_{*a} - \tilde{\theta}_{*b} \leq 1$, we also conclude that $s_n^* \geq 1$. Finally, observe that $0 = \tilde{\theta}_{nb} = s_n^* \tilde{\theta}_{*b} + \tau_n^*$. Thus $\tau_n = -s_n^* \tilde{\theta}_{*b}$. Since $\tilde{\theta}_{*b} \in [0, 1]$, we must then have $\tau_n^* \in [-s_n^*, 0] \subset [-\bar{s}(\theta_*), 0]$. $\qquad\square$

Lemma 10. *Suppose θ_*-unanimity, comparability, and that alternatives $(8\bar{s}(\theta_*)^2)/L$-scored. Then the voters' re-scaled scores are consistent, in the sense that $s_n \tilde{\theta}_{na} + \tau_n = s_m \tilde{\theta}_{ma} + \tau_m$, for all voters $n, m \in [N]$ and alternatives $a \in A_{nm}$ that both voters scored.*

Proof. Under θ_*-unanimity and comparability, by virtue of Lemma 9, we know that $s_{nm} = \frac{1}{|C_{nm}|} \sum_{(a,b) \in C_{nm}} \frac{|\tilde{\theta}_{ma} - \tilde{\theta}_{mb}|}{|\tilde{\theta}_{*a} - \tilde{\theta}_{*b}|} \frac{|\tilde{\theta}_{*a} - \tilde{\theta}_{*b}|}{|\tilde{\theta}_{na} - \tilde{\theta}_{nb}|} = \frac{1}{|C_{nm}|} \sum_{(a,b) \in C_{nm}} \frac{s_m^*}{s_n^*} = \frac{s_m^*}{s_n^*}$. Using the bounds of Lemma 9, we know that $s_{nm} \in [0, \bar{s}(\theta_*)]$. In particular, $|s_{nm}| - 1 \leq \bar{s}(\theta_*)$ (using also $\bar{s}(\theta_*) \geq 1$). Theorem 2 (combined with comparability) then guarantees that for $\|\boldsymbol{w}\|_1 \geq 8s(\theta_*)/L$, we have $s_n \triangleq 1 + \mathrm{LRMEAN}_L(\boldsymbol{w}, \vec{s}_n - 1) = 1 + \mathrm{MEAN}(\boldsymbol{w}, \vec{s}_n - 1) = \frac{1}{\|\boldsymbol{w}\|_1} \sum_m w_m \frac{s_m^*}{s_n^*} = \frac{1}{s_n^*} \mathrm{MEAN}(\boldsymbol{w}, \vec{s^*})$, where $\vec{s}_n \triangleq (s_{nm})_{m \in [N]}$ and $\vec{s^*} \triangleq (s_n^*)_{n \in [N]}$. Crucially, we then have $s_n s_n^* = \mathrm{MEAN}(\boldsymbol{w}, \vec{s^*})$, which is independent from n. In particular, we then have $s_n s_n^* = s_m s_m^*$ for all voters n, m. In fact, we make the additional remark that $s_n s_n^*$ is an average of values s_n^*, which all belong to $[1, \bar{s}(\theta_*)]$. Thus $s_n s_n^* \in [1, \bar{s}(\theta_*)]$.

Now we note that we have the equality $\tau_{nm} = \frac{1}{A_{nm}} \sum_{a \in A_{nm}} \left(s_m(s_m^* \tilde{\theta}_{*a} + \tau_m^*) - s_n(s_n^* \tilde{\theta}_{*a} + \tau_n^*) \right) = \frac{1}{A_{nm}} \sum_{a \in A_{nm}} (s_m \tau_m^* - s_n \tau_n^*) = s_m \tau_m^* - s_n \tau_n^*$, using the equality $s_n s_n^* = s_m s_m^*$. But given Lemma 9, we know that $-\bar{s}(\theta_*)^2 \leq s_n \tau_n^* \leq 0$, thus $|\tau_{nm}| \leq \bar{s}(\theta_*)^2$. Theorem 2 (combined with comparability) then guarantees that for $\|\boldsymbol{w}\|_1 \geq 8s(\theta_*)^2/L$, we have $\tau_n \triangleq \mathrm{LRMEAN}_L(\boldsymbol{w}, \vec{\tau}_n) = \mathrm{MEAN}(\boldsymbol{w}, \vec{\tau}_n) = \mathrm{MEAN}(\boldsymbol{w}, (s_m \tau_m^*)_{m \in [N]}) - s_n \tau_n^*$. As a result, $\tau_n + s_n \tau_n^* = \mathrm{MEAN}(\boldsymbol{w}, (s_m \tau_m^*)_{m \in [N]})$, which is independent from n.

We conclude by noting that, for any voter $n \in [N]$ and any alternative $a \in A_n$, we then have $s_n \tilde{\theta}_{na} + \tau_n = s_n(s_n^* \tilde{\theta}_{*a} + \tau_n^*) + \tau_n = s_n s_n^* \tilde{\theta}_{*a} + \tau_n + s_n \tau_n^* = \mathrm{MEAN}(\boldsymbol{w}, \vec{s^*})\tilde{\theta}_{*a} + \mathrm{MEAN}(\boldsymbol{w}, (s_m \tau_m^*)_{m \in [N]})$, which is independent from voter n. $\qquad\square$

G   Other Proofs

Proposition 2. QRMED *is 1-Lipschitz continuous with respect to the ℓ_∞-norm. That is, for any $\xi \in \mathbb{R}^N$, we have*

$$|\mathrm{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta} + \xi) - \mathrm{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})| \leq \|\xi\|_\infty.$$

Proof. Let $\xi \in \mathbb{R}^N$. For simplicity, we will use the notation $q \triangleq \mathrm{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})$. Let us also denote for all $a \in [A]$ the loss function as follows:

$$L_\xi(x) = \mathrm{QRMED}_L(x | \boldsymbol{w}, \boldsymbol{\theta} + \xi) \triangleq \sum_{n \in [N]} w_n |x - \theta_n - \xi_n| + \frac{1}{2L} x^2. \tag{24}$$

We will prove that

$$\partial L_\xi(q + \|\xi\|_\infty) \cap \mathbb{R}^+ \neq \varnothing, \text{ and} \tag{25}$$

$$\partial L_\xi(q - \|\xi\|_\infty) \cap \mathbb{R}^- \neq \varnothing. \tag{26}$$

Given that L_ξ is strictly convex, and since it is differentiable almost everywhere, its derivative L'_ξ is increasing. We will only show (25) holds, as (26) is an analogous case.

Now, since q minimizes L, we know that $0 \in \partial L(q)$. This implies that

$$0 = \frac{q}{L} + \sum_{n \in [N]} w_n \operatorname{sign}(q - \theta_n) \leq \frac{1}{L} \|\xi\|_\infty + \frac{q}{L} + \sum_{n \in [N]} w_n \operatorname{sign}(q - \theta_n)$$

$$\leq \frac{1}{L} \cdot (q + \|\xi\|_\infty) + \sum_{n \in [N]} w_n \operatorname{sign}(q + \|\xi\|_\infty - \theta_n - \xi_n) \in \partial L_\xi(q + \|\xi\|_\infty).$$

This proves (25) (and (26) by analogy) and concludes the proof. $\qquad\square$

Lemma 12. CLIP *is 1-Lipschitz continuous, i.e.*

$$\forall x, \xi, \Delta, \mu \in \mathbb{R}, \quad |\operatorname{CLIP}(x + \xi | \mu, \Delta) - \operatorname{CLIP}(x | \mu, \Delta)| \leq |\xi|.$$

Proof. We conclude by remarking that the function $x \mapsto \operatorname{CLIP}(x | \mu, \Delta)$ is piecewise, continuous, and that its subdifferential is a subset of $[0, 1]$ at every point. $\qquad\square$

Lemma 13. CLMEAN *is 1-Lipschitz continuous with respect to the input scores. Formally, for any $\xi \in \mathbb{R}^N$, we have*

$$\forall \boldsymbol{w}, \boldsymbol{\theta}, \ \forall \Delta, \mu \in \mathbb{R}, \quad |\operatorname{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta} + \xi | \mu, \Delta) - \operatorname{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta} | \mu, \Delta)| \leq \|\xi\|_\infty.$$

Proof. By triangle inequality and Lemma 3, we have $|\operatorname{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta} + \xi | \mu, \Delta) - \operatorname{CLMEAN}(\boldsymbol{w}, \boldsymbol{\theta} | \mu, \Delta)| \leq \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\operatorname{CLIP}(x_n + \xi_n | \mu, \Delta) - \operatorname{CLIP}(x_n | \mu, \Delta)| \leq \frac{1}{\|\boldsymbol{w}\|_1} \sum w_n |\xi_n| \leq \|\xi\|_\infty$. $\qquad\square$

Proposition 3. LRMEAN *is 2-Lipschitz continuous with respect to the ℓ_∞-norm. Formally, for any $\xi \in \mathbb{R}^N$, we have*

$$|\operatorname{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta} + \xi) - \operatorname{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta})| \leq \|\xi\|_\infty.$$

Proof. Denote $q_{+\xi} \triangleq \operatorname{QRMED}_{L/4}(\boldsymbol{w}, \boldsymbol{\theta} + \xi)$ and $q \triangleq \operatorname{QRMED}_{L/4}(\boldsymbol{w}, \boldsymbol{\theta})$. Using the triangle inequality, we have

$$\begin{aligned}
&|\operatorname{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta} + \xi) - \operatorname{LRMEAN}_L(\boldsymbol{w}, \boldsymbol{\theta})| \\
&= \left| \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q_{+\xi}, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq \left| \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q_{+\xi}, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\quad + \left| \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq |q_{+\xi} - q| + \left| \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq \|\xi\|_\infty + \left| \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} + \xi \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) - \operatorname{CLMEAN}\left(\boldsymbol{w}, \boldsymbol{\theta} \, \middle| \, q, \frac{L\|\boldsymbol{w}\|_1}{4} \right) \right| \\
&\leq 2\|\xi\|_\infty,
\end{aligned}$$

where the last three steps are successively due to Lemma 4, Proposition 2 and Lemma 13. $\qquad\square$

Lemma 14. $\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})$ *has the same sign as* $\operatorname{MED}(\boldsymbol{w}, \boldsymbol{\theta})$ *and* $|\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})| \leq |\operatorname{MED}(\boldsymbol{w}, \boldsymbol{\theta})|$.

Proof. Recall the notation from Equation 10. If $\operatorname{sign}(\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})) \neq \operatorname{sign}(\operatorname{MED}(\boldsymbol{w}, \boldsymbol{\theta}))$, then $\mathcal{L}_{\operatorname{QRMED}_L}(0) < \mathcal{L}_{\operatorname{QRMED}_L}(\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta}))$. This contradicts the fact that $\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{x})$ minimizes $\mathcal{L}_{\operatorname{QRMED}_L}$. This proves the first assertion.

If $|\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})| > |\operatorname{MED}(\boldsymbol{w}, \boldsymbol{\theta})|$, then we have $\mathcal{L}_{\operatorname{QRMED}_L}(\operatorname{MED}(\boldsymbol{w}, \boldsymbol{\theta})) < \mathcal{L}_{\operatorname{QRMED}_L}(\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta}))$. This contradicts the fact that $\operatorname{QRMED}_L(\boldsymbol{w}, \boldsymbol{\theta})$ minimizes $\mathcal{L}_{\operatorname{QRMED}_L}$. This proves the second assertion and concludes the proof. $\qquad\square$

H   Technical lemmas

Lemma 15. *Let $(u_n)_{n \geq 1}$ a sequence of real numbers. Consider the sequence $(S_n = \frac{1}{n} \sum_{k=1}^{n} u_k)_{n \geq 1}$. If $(u_n)_{n \geq 1}$ converges to $l \in \mathbb{R}$, then $(S_n)_{n \geq 1}$ converges to l as well.*

Proof. Assume $(u_n)_{n \geq 1}$ converges to $l \in \mathbb{R}$. Let $\varepsilon > 0$.

By the convergence of $(u_n)_{n \geq 1}$ to l, we know that there exists N_0 such that for all $n \geq N_0, |u_n - l| \leq \epsilon/2$. Also, we know that there exists $N_1 \geq N_0$ such that for all $n \geq N_1$, we have $\left| \frac{1}{n} \sum_{k=1}^{N_1} u_k \right| \leq \epsilon/2$. As a consequence, we have for all $n > N_1$

$$|S_n - l| \leq \left| \frac{1}{n} \sum_{k=1}^{N_1} u_k \right| + \left| \frac{1}{n} \sum_{k=N_1+1}^{n} (u_k - l) \right| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

$\square$

Lemma 16. *Let $(u_n)_{n \geq 1}$ a sequence of real numbers. Consider the sequence $(M_n = \text{MED}((u_k)_{1 \leq k \leq n}))$. If $(u_n)_{n \geq 1}$ converges to $l \in \mathbb{R}$, then $(M_n)_{n \geq 1}$ converges to l as well.*

Proof. Assume $(u_n)_{n \geq 1}$ converges to $l \in \mathbb{R}$. Let $\varepsilon > 0$. By the convergence of $(u_n)_{n \geq 1}$ to l, we know that there exists N_0 such that for all $n \geq N_0, |u_n - l| \leq \epsilon$. This implies that for all $n \geq 2N_0 + 1$, we have $|M_n - l| \leq \epsilon$. Indeed, when $n \geq 2N_0 + 1$ there is a majority (at least $N_0 + 1$ among $2N_0 + 1$) of terms of the sequence in the interval $[l - \epsilon, l + \epsilon]$. $\square$

I   Sparse Unanimity Requires Collaborative Preference Normalization

In this section, we present an impossibility theorem, which roughly says that any *coordinate-wise* vote with *individually normalized* scores must violate sparse unanimity. Our result highlights a central and nontrivial challenge for *sparse voting*, even in the absence of disagreeing voters. In spirit, we essentially prove that any scale-resilient sparse voting algorithm must leverage *collaborative* preference scaling, as MEHESTAN does. To formalize our impossibility theorem, we first need to introduce some assumptions on what seem to be reasonable *individual-based normalizations* and *score aggregations*.

I.1   Individual-based Normalization

Intuitively, any *robust sparse voting* algorithm must make sure that its output will not be affected by the scaling used by voters when they report their scores. The simplest way to guarantee this is to perform a score *normalization* on voters' reported scores. In this section, we define what a score normalization is, and what desirable properties it ought to have. First we define a **normalizer** as a function $\overrightarrow{\text{NORM}} : (\mathbb{R}^{\leq A})^N \to (\mathbb{R}^{\leq A})^N$ that preserves voters' preferences, i.e. such that, for any voter $n \in [N]$, we have $\overrightarrow{\text{NORM}}_n(\theta) \sim \theta_n$. Below, we list other desirable properties.

Definition 9.   1. A normalizer $\overrightarrow{\text{NORM}}$ is **individual-based** if a voter's normalized scores only depend on the voters' reported scores, i.e.

$$\exists \text{NORM} : \mathbb{R}^{\leq A} \to \mathbb{R}^{\leq A}, \ \forall \boldsymbol{\theta} \in (\mathbb{R}^{\leq A})^N, \ \forall n \in [N], \ \overrightarrow{\text{NORM}}_n(\boldsymbol{\theta}) = \text{NORM}(\theta_n), \tag{27}$$

2. A normalizer $\overrightarrow{\text{NORM}}$ is **scale-invariant** if the normalized scores are independent of the preference scaling of the reported scores, i.e.

$$\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in (\mathbb{R}^{\leq A})^N, \ \forall n \in [N], \ \theta_n \sim \theta'_n \implies \overrightarrow{\text{NORM}}(\boldsymbol{\theta}) = \overrightarrow{\text{NORM}}(\boldsymbol{\theta}'). \tag{28}$$

3. A normalizer $\overrightarrow{\text{NORM}}$ is **neutral** if it treats all alternatives symmetrically. More precisely, denote $S(A)$ the set of permutations of $[A]$. For any $\theta \in \mathbb{R}^{\leq A}$ and $\sigma \in S(A)$, we define $(\sigma \cdot \theta)_a \triangleq \theta_{\sigma(a)}$ if the entry $\sigma(a)$ of partial vector θ exists (otherwise $(\sigma \cdot \theta)_a$ is not defined). Similarly, we define $(\sigma \cdot \boldsymbol{\theta})_n \triangleq \sigma \cdot \theta_n$. **Neutrality** then demands that

$$\forall \boldsymbol{\theta} \in (\mathbb{R}^{\leq A})^N, \ \forall \sigma \in S(A), \ \overrightarrow{\text{NORM}}(\sigma \cdot \boldsymbol{\theta}) = \sigma \cdot \overrightarrow{\text{NORM}}(\boldsymbol{\theta}). \tag{29}$$

4. An individual-based normalizer $\overrightarrow{\text{NORM}}$ is **stable** if the function $x \mapsto \text{NORM}(0, x, 1)$ is Lipschitz continuous on $[0, 1]$.

As an example of a (single-voter) normalizer, standardization is given by $\text{STDNORM}_a(x) \triangleq \frac{x_a - \text{MEAN}(x)}{\text{SD}(x)}$, where A_x is the subset of alternatives scored by the score vector $x \in \mathbb{R}^{\leq A}$, $\text{MEAN}(x) \triangleq \frac{1}{|A_x|} \sum_{a \in A_x} x_a$ is the mean of the scores and $\text{SD}^2(x) \triangleq \frac{1}{|A_x|-1} \sum_{a \in A_x} (x_a - \text{MEAN}(x))^2$ is their standard deviation, assuming $\text{SD}(x) > 0$. If $\text{SD}(x) = 0$, then we may simply set $\text{STDNORM}_a(x) \triangleq 0$ for all scored alternatives $a \in A_x$. Another popular normalizer is min-max normalization, given by $\text{MINMAXNORM}_a(x) \triangleq \frac{x_a - \min_{b \in A_x} x_b}{\max_{b \in A_x} x_b - \min_{b \in A_x} x_b}$, assuming $\max_{b \in A_x} x_b > \min_{b \in A_x} x_b$ (otherwise, we return $\text{MINMAXNORM}_a(x) \triangleq 0$ for all scored alternatives $a \in A_x$). Applying such single-voter normalizers to all voters clearly yield normalizers $\overrightarrow{\text{STDNORM}}$ and $\overrightarrow{\text{MINMAXNORM}}$.

Proposition 4. *Standardization and min-max normalizers are individual-based, scale-invariant, neutral and stable.*

Proof. They are clearly individual-based, scale-invariant and neutral normalizers. Plus, min-max normalizer is clearly stable. To show that standardization is stable, consider $g \colon x \mapsto \text{STDNORM}(0, x, 1) = \sqrt{2}(-x - 1, 2x - 1, 2 - x)/\sqrt{(x+1)^2 + (2x-1)^2 + (x-2)^2}$. It follows that g is continuously differentiable, and therefore Lipschitz continuous on $[0, 1]$. $\qquad\square$

I.2 Score Aggregation

A score aggregation is a function $\text{AGG} : (\mathbb{R}_+ \times \mathbb{R}^{\leq A})^N \to \mathbb{R}^A$. Below, we identify properties that score aggregations may have.

Definition 10. 1. A score aggregation AGG is **coordinate-wise** if the score computed by an alternative only depends on the reported scores for this alternative, i.e., for any alternative $a \in [A]$,

$$\exists \text{AGG}_a : (\mathbb{R}_+ \times \mathbb{R})^{\leq N} \to \mathbb{R}, \ \forall \boldsymbol{w}, \boldsymbol{\theta}, \ (\text{AGG}(\boldsymbol{w}, \boldsymbol{\theta}))_a = \text{AGG}_a ((w_n)_{n \in N_a}, (\theta_{na})_{n \in N_a}). \tag{30}$$

2. A score aggregation AGG is **anonymous** if it treats alternatives symmetrically, i.e.

$$\forall \boldsymbol{w} \in \mathbb{R}_+^N, \ \forall \boldsymbol{\theta} \in (\mathbb{R}^{\leq A})^N, \ \forall \sigma \in S(A), \ \text{AGG}(\boldsymbol{w}, \sigma \cdot \boldsymbol{\theta}) = \sigma \cdot \text{AGG}(\boldsymbol{w}, \boldsymbol{\theta}). \tag{31}$$

3. A score aggregation AGG is **neutral** if it treats voters symmetrically, i.e.

$$\forall \boldsymbol{w} \in \mathbb{R}_+^N, \ \forall \boldsymbol{\theta} \in (\mathbb{R}^{\leq A})^N, \ \forall \sigma \in S(N), \ \text{AGG}(\sigma \cdot \boldsymbol{w}, \sigma \cdot \boldsymbol{\theta}) = \text{AGG}(\boldsymbol{w}, \boldsymbol{\theta}), \tag{32}$$

where the action of σ on $\boldsymbol{\theta}$ is defined by $(\sigma \cdot \boldsymbol{\theta})_n \triangleq \theta_{\sigma(n)}$.

4. A coordinate-wise score aggregation AGG is **max-dominated** if each of its coordinates is dominated by the max aggregation, i.e.

$$\exists \lambda > 0, \ \forall \boldsymbol{w} \in \mathbb{R}_+^N, \ \forall \theta \in \mathbb{R}^{\leq A}, \ \|\text{AGG}(w, \theta)\|_\infty \leq \lambda \max_{n \in [N], a \in [A]} |\theta_{na}|. \tag{33}$$

5. A coordinate-wise score aggregation AGG is **locally Lipschitz continuous** if each of its coordinates is locally Lipschitz continuous with respect to the ℓ_∞-norm.

6. A coordinate-wise score aggregation AGG is **asymptotically correct** if each of its coordinates can recover any score θ_*, once the input is a sufficiently large sequence converging to θ_*, i.e. $\lim_{n \to \infty} \xi_n = 0$ implies that

$$\forall \theta_* \in \mathbb{R}^A, \ \forall \varepsilon > 0, \ \exists N_0 > 0, \ \left\| \text{AGG}(\vec{1}^{N_0}, (\theta_* + \xi_n)_{n \in [N_0]}) - \theta_* \right\|_\infty \leq \varepsilon, \tag{34}$$

where $\vec{1}^{N_0} \in \mathbb{R}^{N_0}$ is the vector whose entries all equal 1.

Lemma 17. QRMED_L *is asymptotically correct.*

Proof. Let $x \in \mathbb{R}$, and $(\xi_n)_{n \geq 1}$ a sequence of real numbers converging to 0. We will assume that $x \geq 0$, the case $x \leq 0$ being analogous. Now, let $\varepsilon > 0$. By convergence of $(\xi_n)_{n \geq 1}$, we know that there is N_0 such that for all $n \geq N_0 + 1$, we have $|\xi_n| \leq \varepsilon/2$. We then introduce $N_1 \triangleq 2N_0 + \max(0, \frac{1}{L}(x - \varepsilon))$ and the function $L \colon t \mapsto \frac{1}{2L}t^2 + \sum_{1 \leq n \leq N_1} |t - x - \xi_n|$. The function L is strictly convex, subdifferentiable, and its minimum is attained at $q \triangleq \text{QRMED}_L((x + \xi_n)_{1 \leq n \leq N_1})$. Thanks to the aforementioned result about the convergence of $(\xi_n)_{n \geq 1}$, we can show that $\partial L(x + \varepsilon) \subset \mathbb{R}^+$. Similarly, we can show that $\partial L(x - \varepsilon) \subset \mathbb{R}^-$. We conclude by convexity of L that we necessarily have $q \in [x - \varepsilon, x + \varepsilon]$, i.e. $|x - q| \leq \varepsilon$. $\qquad \square$

Proposition 5. *The mean, the median and* QRMED *are all coordinate-wise, anonymous, neutral, max-dominated, locally Lipschitz continuous and asymptotically correct score aggregations.*

Proof. It is straightforward that the mean, the median and QRMED are coordinate-wise, anonymous, and neutral score aggregations. The mean and the median are asymptotically correct by Lemmas 15 and 16 respectively. QRMED is asymptotically correct by Lemma 1. The mean and the median are trivially max-dominated. QRMED is max-dominated since $|\text{QRMED}| \leq |\text{MED}|$ by Lemma 17. The mean is trivially Lipschitz continuous. QRMED is Lipschitz continuous by Proposition 2. Since $\text{MED} = \text{QRMED}_0$, the median is also Lipschitz continuous. They are thus locally Lipschitz continuous. $\qquad \square$

I.3   The Impossibility Theorem

We now state our impossibility theorem. We stress that the theorem does not assume any Byzantine voter; in fact, as demanded by *sparse unanimity*, it assumes that all voters are honest and express the same preference θ_*, albeit each voter only scores a (potentially small) subset of all alternatives.

Theorem 4. *Given any* **individual-based**, *scale-invariant, neutral and stable normalizer* $\overrightarrow{\text{NORM}}$ *and any* **coordinate-wise**, *anonymous, neutral, max-dominated, locally Lipschitz continuous and asymptotically correct score aggregation* AGG, $\text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \text{AGG}(\boldsymbol{w}, \overrightarrow{\text{NORM}}(\boldsymbol{\theta}))$ *fails to be* **sparsely unanimous**.

Sketch of proof. Our proof assumes $\theta_{*a} \triangleq a$, for all alternatives $a \in [A]$. We consider $N \triangleq K \cdot (A - 2)$ voters, each with a unit voting right, with voters $\{n, \ldots, n + K - 1\}$ reporting the scores of alternatives 1, 2 and $n + 2$. The assumptions on $\overrightarrow{\text{NORM}}$ and AGG then imply that most alternatives a will receive roughly the same score, especially for a large enough. This then implies that the vote outputs scores that are hardly correlated with θ_*. In fact, the correlation goes to 0 in the limit $A, K \to \infty$. Appendix I.4 provides the full proof. $\qquad \square$

Theorem 4 suggests that *sparse unanimity* cannot be achieved with individual score normalization. Instead, Robust Sparse voting seems to require adapting a voter's score normalization based on other voters' scores, i.e. the score normalization must be *collaborative*. It is critical to note that this may create a vulnerability in practice, as Byzantine voters may leverage their impact on other voters' scores to scale these scores as best fits their purposes. Typically, whenever a voter n prefers a to b, a disagreeing Byzantine voter may want to make voter n's preference scale vanish, so that the vote essentially considers that voter n is nearly indifferent between a and b. Instead, our solution to Robust Sparse voting builds upon new provably L-Lipschitz-resilient primitives.

I.4   Proof of our Impossibility Theorem

Theorem 4. *Given any* **individual-based**, *scale-invariant, neutral and stable normalizer* $\overrightarrow{\text{NORM}}$ *and any* **coordinate-wise**, *anonymous, neutral, max-dominated, locally Lipschitz continuous and asymptotically correct score aggregation* AGG, *the vote* $\text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \text{AGG}(\boldsymbol{w}, \overrightarrow{\text{NORM}}(\boldsymbol{\theta}))$ *fails to be* **sparsely unanimous**.

Proof. Consider θ_* defined by $\theta_{*a} = a$, for all alternatives $a \in [A]$. Assume that there are $N \triangleq K \cdot (A - 2)$ voters, with voters $\{n, \ldots, n + K - 1\}$ reporting the scores of alternatives 1, 2 and $n + 2$. Note that any alternative $a \geq 3$ is then scored only by the group of voters $\{a - 2, \ldots, a + K - 3\}$. We then give the same voting right $w_n \triangleq w_0$ to all voters n, which we choose to be large enough to guarantee the conditions of sparse unanimity. We will then consider the limit of this setting, as $A, K \to \infty$.

By scale invariance, we know that $\text{NORM}(\theta_n) = \text{NORM}\left(\frac{\theta_n - 1}{n + 1}\right) = \text{NORM}(0, \frac{1}{n+1}, 1)$, where the entries of $(0, \frac{1}{n+1}, 1)$ correspond to the alternatives 1, 2 and $n + 2$. Since NORM is stable, we have $\text{NORM}(\theta_n) = \text{NORM}(0, 0, 1) + \mathcal{O}(1/n) =$

$(x, x, z) + \mathcal{O}(1/n)$, where $x, z \in \mathbb{R}$ are fully determined by the normalization function NORM (note that we use the neutrality of NORM to guarantee that x and z do not depend on which alternatives' scores have been reported by voter n). Moreover, we necessarily have $x \neq z$ because NORM is a normalizer, and thus preserves strict order. Since F is coordinate-wise, for any alternative a, there exists a function f_a such that $F_a(\boldsymbol{\theta}) = f_a(\boldsymbol{\theta}_a)$. Since F is neutral, we must have $f_a = f_b \triangleq f$ for any two alternatives $a, b \in [A]$. Since F is anonymous and neutral, we must have $f(\boldsymbol{\theta}_a) = f(\boldsymbol{\theta}_b)$ if both $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ contain the same entries (but potentially from different voters). By local Lipschitz continuity of F, if these entries are equal to $z + \mathcal{O}(1/a)$, then $F_a(\boldsymbol{\theta}) = f(z, \dots, z) + \mathcal{O}(1/a)$.

Now, recall that for all $x, y \in \mathbb{R}^A$, $\text{CORREL}(x, y) = \langle \frac{\pi(x)}{\|\pi(x)\|_2}, \frac{\pi(y)}{\|\pi(y)\|_2} \rangle$, where for all $x \in \mathbb{R}^A$, $\pi(x) = x - \text{MEAN}(x)$.

Denote $\rho \triangleq \text{VOTE}(w, \boldsymbol{\theta})$. We then have $\rho_1 = f(\vec{x}_1)$ and $\rho_2 = f(\vec{x}_2)$, where $x_{1n} = x + \mathcal{O}(1/n)$ and $x_{2n} = x + \mathcal{O}(1/n)$. Since F is max-dominated, it directly follows that $f(\vec{x}_1), f(\vec{x}_2) = \mathcal{O}(1)$. For $a \geq 3$, we then have $\rho_a = f(z, \dots, z) + \mathcal{O}(1/a)$. Moreover, since F is asymptotically correct, once K is large enough, we have $|f(z, \dots, z) - z| \leq |z - x|/2$. In particular, since $z \neq x$, this means that there exists a constant $c_1 > 0$ such that $|f(z, \dots, z) - x| > c_1$.

Thus, on one hand

$$\text{MEAN}(\rho) = \frac{\mathcal{O}(1) + \mathcal{O}(1) + A f(z, \dots, z) + \sum_a \mathcal{O}(1/a)}{A} = f(z, \dots, z) + \mathcal{O}\left(\frac{\ln A}{A}\right).$$

And on the other hand, we have that there exists a constant $c_2 > 0$ such that

$$\|\rho - \text{MEAN}(\rho)\|_2^2 > c_2.$$

Also, thanks to the choice of θ_*, by computing $\frac{\pi(\theta_*)}{\|\pi(\theta_*)\|_2}$ we have

$$\left\| \frac{\pi(\theta_*)}{\|\pi(\theta_*)\|_2} \right\|_\infty = \mathcal{O}(\frac{1}{\sqrt{A}}).$$

We can now conclude that

$$\begin{aligned}
|\text{CORREL}(\rho, \theta_*)| &= \left| \langle \frac{\pi(\theta_*)}{\|\pi(\theta_*)\|_2}, \frac{\pi(\rho)}{\|\pi(\rho)\|_2} \rangle \right| \\
&\leq \left\| \frac{\pi(\theta_*)}{\|\pi(\theta_*)\|_2} \right\|_\infty \cdot \left(\mathcal{O}(1) + \mathcal{O}(1) + A\mathcal{O}(\frac{\ln A}{A}) + \sum_a \mathcal{O}(\frac{1}{a}) \right) \\
&= \mathcal{O}(\frac{\ln A}{\sqrt{A}}).
\end{aligned}$$

$\square$

J   Additional Experiments

In this section, we present the results of additional experiments aiming at testing MEHESTAN's ability to tolerate sparsity. Recall the experimental setup from Section 5 in the main body. We conduct the following additional experiments:

- the same experiments as Section 5 with θ_* following a Uniform (Figure 2), Gaussian (Figure 3) and Cauchy (Figure 4) distribution. All were performed both with and without biased sparsity.

- the measurement of the performance depending on the sparsity bias, with θ_* following a Uniform, Gaussian, and Cauchy distribution (Figure 5).
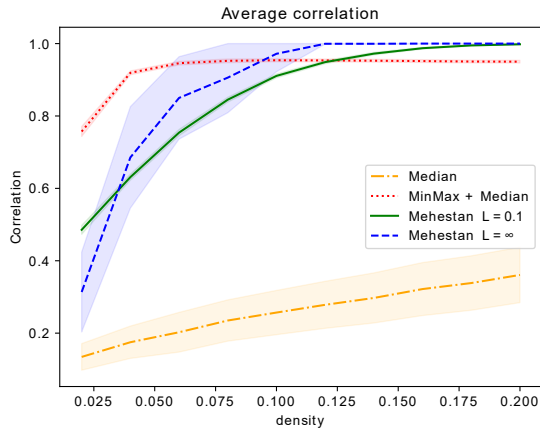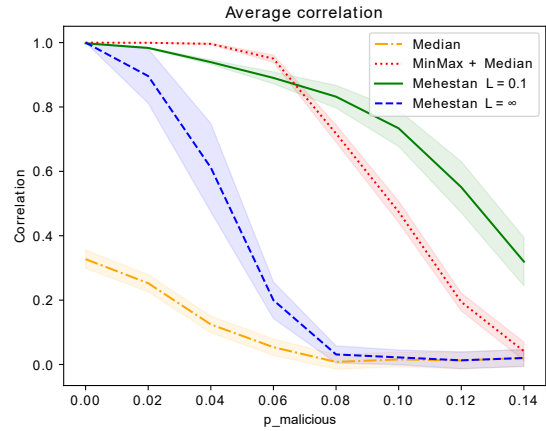
(a) Influence of the **density** (with bias in sparsity)
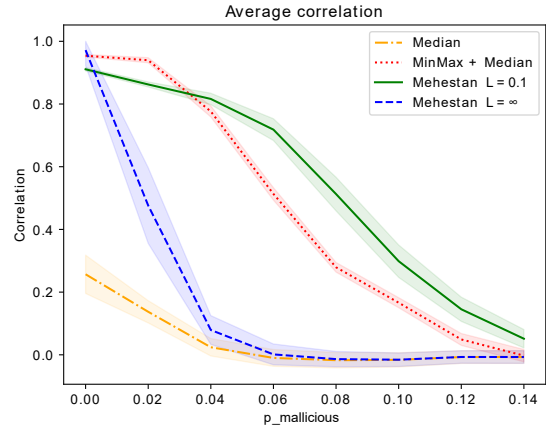
(b) Influence of the **fraction of malicious voters** (with bias in sparsity)
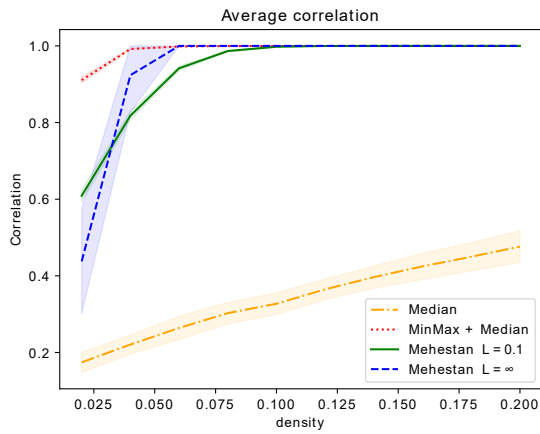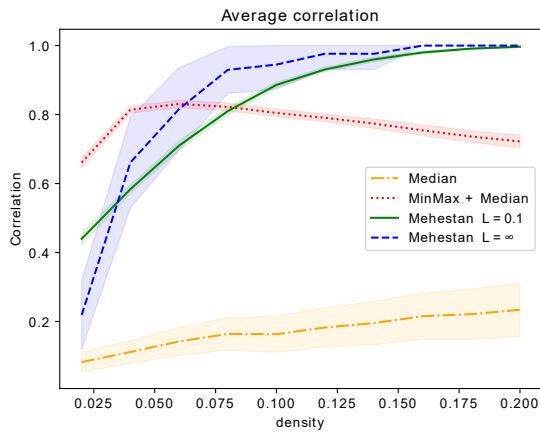
(c) Influence of the **density** (without bias in sparsity)
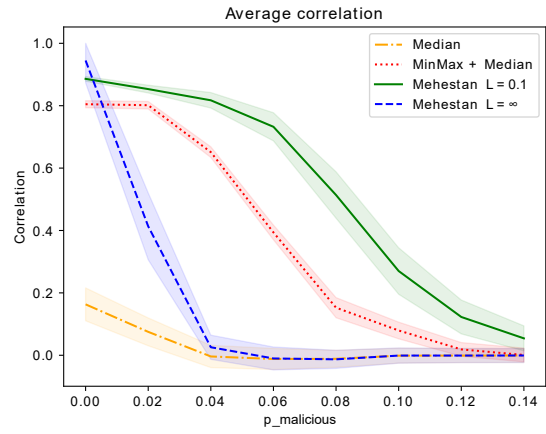
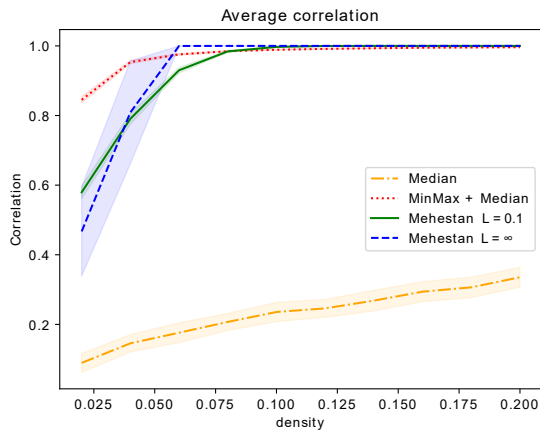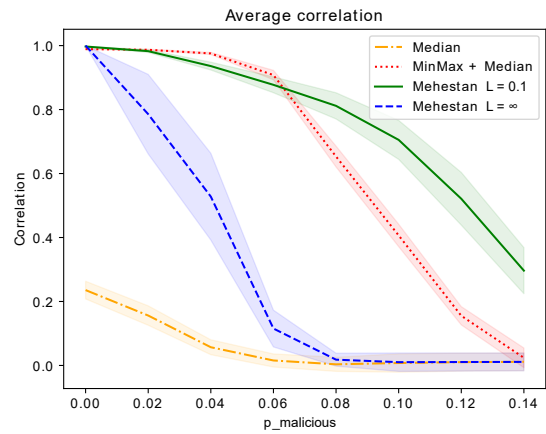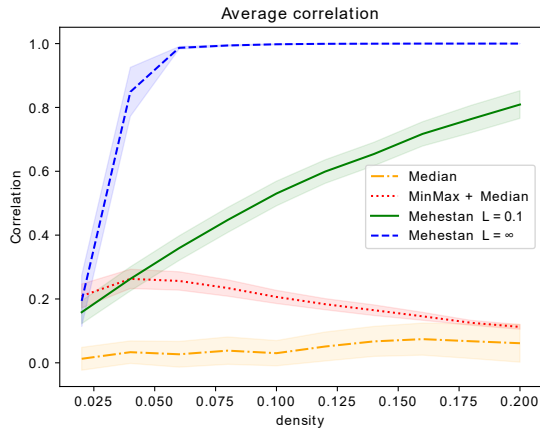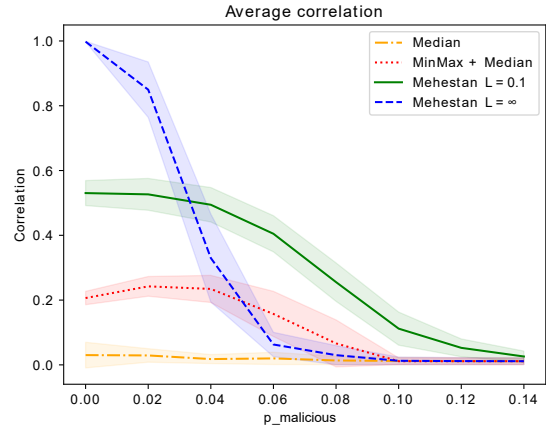(d) Influence of the **fraction of malicious voters** (without bias in sparsity)

Figure 2: Performance of MEHESTAN under sparsity, with and without malicious voters (**Uniform** distribution of θ_*)

(a) Influence of the **density** (with bias in sparsity)

(b) Influence of the **fraction of malicious voters** (with bias in sparsity)

(c) Influence of the **density** (without bias in sparsity)

(d) Influence of the **fraction of malicious voters** (without bias in sparsity)
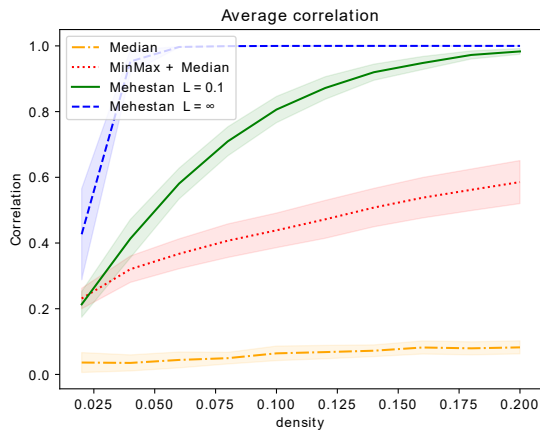
Figure 3: Performance of MEHESTAN under sparsity, with and without malicious voters. (**Gaussian** distribution of θ_*)
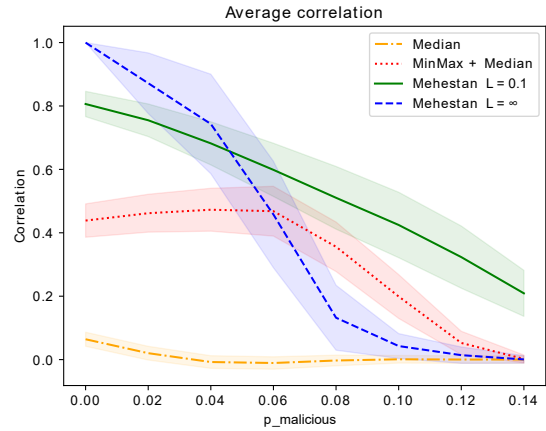
(a) Influence of the **density** (with bias in sparsity)

(b) Influence of the **fraction of malicious voters** (with bias in sparsity)

(c) Influence of the **density** (without bias in sparsity)

(d) Influence of the **fraction of malicious voters** (without bias in sparsity)

Figure 4: Performance of MEHESTAN under sparsity, with and without malicious voters. (**Cauchy** distribution of θ_*)
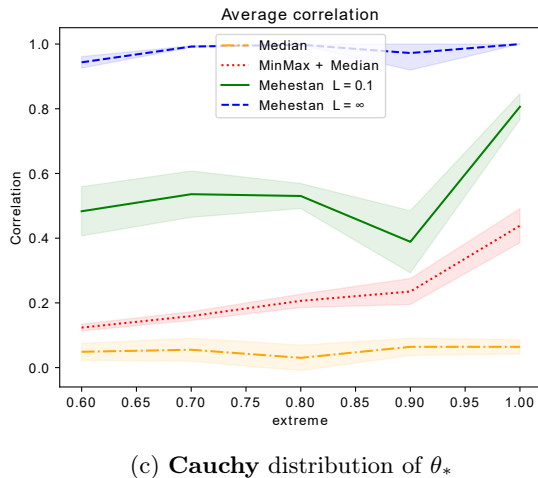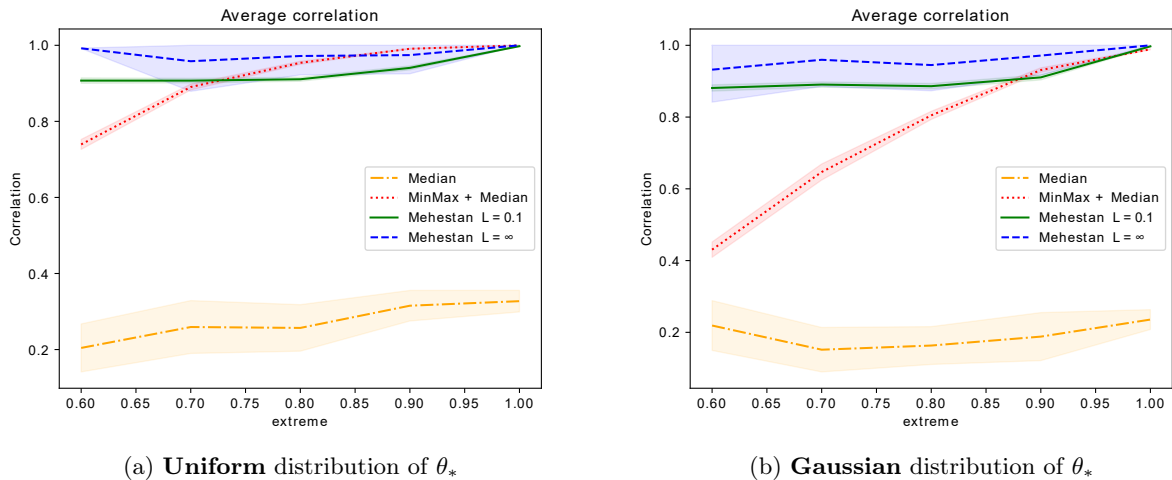
(a) **Uniform** distribution of θ_*

(b) **Gaussian** distribution of θ_*

(c) **Cauchy** distribution of θ_*

Figure 5: Performance of Mehestan depending on the **sparsity bias**.

The `extreme` parameter indicates the proportion of alternatives each voter has access to (either the top or bottom `extreme`). Out of these, the voter rates each of them with probability 0.1.

Influence of the distribution. From these additional experiments, we observe that the distribution of the ground-truth preferences θ_* can significantly influence the performance of the voting algorithms. More specifically, it appears that the more the distribution is heavy-tailed, the harder it is for the voting algorithms to recover the preferences. This is particularly visible for the Cauchy distribution. Indeed, the baseline *MinMax+Median* performs very poorly on Figure 4 (*Median* alone is even worse), whereas Mehestan is less affected by the change. A possible explanation for this is that the more the distribution is heavy-tailed, the more there are strong outliers in θ_*. This leads to a concentration of the rest of the scores after the min-max normalization, thus showing that global normalization (see Section 4) is critical.

Influence of the sparsity bias. Figure 5 displays the performance of the voting algorithm as the bias in sparsity decreases. As expected, Mehestan is more robust to high levels of bias in the sparsity (left of the plots) than the baseline *MinMax+Median*. For the highest level of bias in sparsity on Figure 5, each voter only rates either the top 60% or the bottom 60% of the alternatives. This shows that the global normalization step (see Section 4) is crucial, and we can see that the algorithms not performing it (*Median* and *MinMax+Median*) fare poorly in this setting.

K Extensions

We discuss here how additional desirable properties can be guaranteed for *robust sparse voting*, by tweaking MEHESTAN.

K.1 Differential Privacy

In several applications, guaranteeing the privacy of *robust sparse voting* may be critical to prevent voter coercion. In this section, we prove that MEHESTAN can be easily made differentially private. Let us first recall the definition of (voter-level) differential privacy.

Definition 11 (Voter-level differential privacy). Let $\epsilon > 0$. A (randomized) vote VOTE is ϵ-differentially private if, for any $f \in [N]$ and all subsets $\mathcal{S} \subset \mathbb{R}^A$, we have

$$\mathbb{P}\left[\text{VOTE}(\boldsymbol{w}, \boldsymbol{\theta}) \in \mathcal{S}\right] \leq e^{\epsilon} \mathbb{P}\left[\text{VOTE}(\boldsymbol{w}_{-f}, \boldsymbol{\theta}_{-f}) \in \mathcal{S}\right]. \tag{35}$$

For any parameter $\varepsilon > 0$, we define the ε-differentially private MEHESTAN voting algorithm by simply adding a Laplacian noise to each returned global score, whose scale is proportional to L/ε. More formally, for any alternative $a \in [A]$, we define

$$\text{DP-MEHESTAN}_{L\varepsilon a}(\boldsymbol{w}, \boldsymbol{\theta}) \triangleq \text{MEHESTAN}_{La}(\boldsymbol{w}, \boldsymbol{\theta}) + \text{LAP}\left(0, \frac{L\|\boldsymbol{w}\|_{\infty}}{\epsilon}\right), \tag{36}$$

where $\text{LAP}(0, b)$ is a random variable drawn from the Laplace distribution with mean 0 and scale b.

Theorem 5. *For all $\epsilon > 0$, $\text{DP-MEHESTAN}_{L\varepsilon}$ is ϵ-differentially private.*

Proof. This follows directly from the ε-differential privacy of the Laplace mechanism, combined with the resilience guarantee of MEHESTAN. □

K.2 Uncertainty-aware Voting

In practice, reported scores are noisy, with potentially different levels of noise. Here, we show how MEHESTAN can be enhanced to account for uncertainty in the input. Our key solution is to leverage a new operator called the *mean-risk distance* MRDIST. Essentially, MRDIST simulates the fact that a voter's uncertainty-aware vote is its expected vote, when the voter's score is drawn from our Bayesian prior on their actual score. More formally, given a prior probability distribution $\mathcal{D}$ of finite expectation and a point $z \in \mathbb{R}$, MRDIST is defined as:

$$\text{MRDIST}(z \mid \mathcal{D}) \triangleq \mathbb{E}_{\theta \sim \mathcal{D}}\left[|z - \theta|\right]. \tag{37}$$

QRMED can then be easily made uncertainty-aware, by replacing the absolute values as follows:

$$\text{QRMED}_L(\boldsymbol{w}, \mathcal{D}) \triangleq \underset{z \in \mathbb{R}}{\arg\min} \, \frac{1}{2L} z^2 + \sum_{n \in [N]} \text{MRDIST}(z|\mathcal{D}_n). \tag{38}$$

Crucially, since MRDIST is a mean of functions whose subderivatives are always of absolute value at most 1, the subderivatives of MRDIST are also always of absolute value at most 1. Intuitively, this means that if z falls into a voter n's uncertainty set, then the voter n will only slightly pull z towards their maximum-a-posterior score. However, if z falls very far from this uncertainty set, the voter n will be pulling with its entire voting rights w_n In any case, the fact that any voter's pull remains bounded by their voting rights guarantees the L-Lipschitz resilience of this generalization of QRMED. Additionally, MEHESTAN can be similarly adapted, though handling collaborative scaling normalization is not straightforward. Interestingly, MRDIST yields a closed form expression for some parameterized priors, like the Laplacian prior.

Proposition 6. *For a Laplacian prior $\mathcal{D} = \text{LAP}(\mu, \delta)$, $\text{MRDIST}(z \mid \mathcal{D}) = |z - \mu| - \mu + \delta e^{-\frac{|z-\mu|}{\delta}}$.*

Proof. This is a straightforward integral computation. □

K.3   Measuring Preference Polarization

We now propose a resilient polarization measure on a given alternative a. To do this, we divide the voters $n \in N_a$ who scored a in two equal subsets of high and low-scoring voters. More precisely, define $w_a \triangleq \sum_{n \in N_a} w_n$ and $m_a \triangleq \text{MED} \left\{ w_n, s_n \tilde{\theta}_{na} + \tau_n \,\middle|\, n \in N_a \right\}$. Now consider $N_a^+ \triangleq \left\{ n \in N_a \,\middle|\, s_n \tilde{\theta}_{na} + \tau_n > m_a \right\}$ and $N_a^- \triangleq \left\{ n \in N_a \,\middle|\, s_n \tilde{\theta}_{na} + \tau_n < m_a \right\}$. Define finally $w_a^+ \triangleq \sum_{n \in N_a^+} w_n$ and $w_a^- \triangleq \sum_{n \in N_a^-} w_n$. By definition of the median, we must have $w_a^- \geq \frac{1}{2} w_a$ and $w_a^+ \geq \frac{1}{2} w_a$.

Now denote $\rho_a \triangleq \text{MEHESTAN}_{La}(w, \boldsymbol{\theta})$. For any voter $n \in N_a$ who scored a, we define $\eta_{na}^+ \triangleq \max(0, s_n \tilde{\theta}_{na} + \tau_n - \rho_a)$. Similarly, denote $\eta_{na}^- \triangleq \max(0, \rho_a - s_n \tilde{\theta}_{na} - \tau_n)$. We define the positive polarization ψ_a^+ and the negative polarization ψ_a^-, on alternative a by

$$\psi_a^\star \triangleq 1 + \text{QRMED}_L \left\{ \{w_n, \eta_{na}^\star - 1 \mid n \in N_a^\star\} \cup \left\{ \frac{1}{2} w_a - w_a^\star, \max\{m_a - \rho_a, -1\} \right\} \right\}, \tag{39}$$

for $\star \in \{-, +\}$. In other words, the polarization is initially assumed to be 1. Voters who believe that ρ_a is underestimated will then pull ψ_a^+ towards larger values; but they can only do so with a unit force. This prevents malicious voters from hacking polarization measures.