Strategyproof Linear Regression in High Dimensions

YILING CHEN, Harvard University
CHARA PODIMATA, Harvard University
ARIEL D. PROCACCIA, Carnegie Mellon University
NISARG SHAH, University of Toronto

This paper is part of an emerging line of work at the intersection of machine learning and mechanism design, which aims to avoid noise in training data by correctly aligning the incentives of data sources. Specifically, we focus on the ubiquitous problem of *linear regression*, where *strategyproof* mechanisms have previously been identified in two dimensions. Our main contribution is the discovery of a family of *group strategyproof* linear regression mechanisms in any number of dimensions, which we call *generalized resistant hyperplane* mechanisms. The game-theoretic properties of these mechanisms — and, in fact, their very existence — are established through a connection to a discrete version of the Ham Sandwich Theorem.

1 INTRODUCTION

Designing machine learning algorithms that are robust to noise in training data is a topic of intense research. A large body of work addresses stochastic noise [Goldman and Sloan, 1995, Littlestone, 1991]. On the other extreme, another branch of the literature focuses on adversarial noise [Bshouty et al., 2002, Kearns and Li, 1993], that is, errors are introduced by an adversary with the explicit purpose of sabotaging the algorithm. The latter approach is often too pessimistic, and generally leads to negative results.

More recently, some researchers have taken a game-theoretic viewpoint; it suggests a model of *strategic noise* that can be seen as occupying the middle ground of noise models. Specifically, training data is provided by strategic sources — hereinafter *agents* — that may intentionally introduce errors *to maximize their own benefit*. Compared to adversarial noise, the advantage of this model (when its underlying assumptions hold true) is that, if we aligned the agents' incentives correctly, it would be possible to obtain uncontaminated data. From this viewpoint, the ideal is the design of learning algorithms that are *strategyproof*, i.e., where supplying pristine data is a dominant strategy for each agent.

We subscribe to this agenda, and advance it in the context of the ubiquitous problem of linear regression, i.e., fitting a hyperplane through given data. We consider agents who can manipulate their dependent variables in order to minimize their vertical distance from the output hyperplane, and design strategyproof regression mechanisms without payments.

When does this type of strategic regression problem arise? Dekel et al. [2010] give the real-world example of the global fashion chain Zara, whose distribution process relies on regression [Caro and Gallien, 2010]. Specifically, the demand for each product at each store is predicted based on historical data, as well as information provided by store managers. Since the supply of popular items is limited, store managers may strategically manipulate requested quantities so that the output of the regression process would better fit their needs, and, indeed, there is ample evidence that many of them have done so [Caro et al., 2010]. More generally, as discussed in detail by Perote and Perote-Peña [2004], this type of setting is relevant whenever "data could come from surveys composed by agents interested in not being perceived as real outliers if the estimation results could be used in the future to change the economic situation of the agents that generate the sample."

1.1 Our Model and Results

A bit more formally, we study a linear regression setting in which the task is to fit a hyperplane through data points $(\boldsymbol{x}_i, y_i)$ for $i \in \{1, \ldots, n\}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ are the independent variables and $y_i \in \mathbb{R}$ is the dependent variable. Following Dekel et al. [2010] and Perote and Perote-Peña [2004], we assume that the independent variables are public information, but dependent variable y_i is held privately by agent i. A mechanism elicits the private information of the agents, and returns a hyperplane represented by vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_0) \in \mathbb{R}^{d+1}$. Under this outcome, the residual for agent i is $r_i = y_i - \boldsymbol{\beta}_1^T \boldsymbol{x}_i - \boldsymbol{\beta}_0$, and, loosely speaking, agents wish to minimize $|r_i|$ (see Section 2 for a precise description of agent preferences).

Our starting point is the work of Dekel et al. [2010], who show that empirical risk minimization (ERM) with the L_1 loss (in short, L_1-ERM), coupled with a specific tie-breaking rule, is group strategyproof, that is, no coalition of agents can be weakly better off by misreporting. We extend this result and show that replacing the L_1 loss by a weighted L_1 loss and adding convex regularization to the risk function preserves group strategyproofness. But this still gives a relatively restricted family of strategyproof mechanisms, and we seek a broader understanding of what is possible in our setting.

To that end, we look to the work of Perote and Perote-Peña [2004], who focus on the two-dimensional case (known as *simple regression*), i.e., fitting a line through points on a plane. They propose a wide family of strategyproof mechanisms, which they call *clockwise repeated median* (CRM) mechanisms. These mechanisms are parametrized by two subsets of agents S and S'. Perote and Perote-Peña [2004] establish conditions on S and S' under which they claim that CRM mechanisms are strategyproof. We identify a mistake in this result, present counterexamples showing violation of strategyproofness under their conditions, and identify three stricter conditions under which we can recover strategyproofness — in fact, we prove group strategyproofness. Under one of our conditions, CRM mechanisms coincide with a family of mechanisms from the statistics literature known as *resistant line mechanisms* [Johnstone and Velleman, 1985]. Our work therefore establishes the group strategyproofness of these mechanisms.

Our main result is that we generalize the CRM family to higher dimensions, thereby justifying the title of this paper. We introduce the family of *generalized resistant hyperplane* (GRH) mechanisms, which, to the best of our knowledge, is the first extension of resistant line mechanisms beyond the plane. In $d + 1$ dimensions, GRH mechanisms are parametrized by $d + 1$ subsets of agents. Through a surprising connection to the literature on the Ham Sandwich Theorem, we find a condition on the subsets under which GRH mechanisms are group strategyproof. Strikingly, our proof of this general *group strategyproofness* result in *any number* of dimensions is much shorter than the (incorrect) proof of Perote and Perote-Peña [2004] for the *strategyproofness* of CRM mechanisms in *two dimensions*.

We also study a property called impartiality, which is stricter than strategyproofness. We establish the existence of a wide family of impartial mechanisms, which, unlike our generalized L_1-ERM and generalized resistant hyperplane mechanisms, are strategyproof but not group strategyproof (except for constant functions). Building upon the work of Moulin [1980], we also provide two non-constructive characterizations of strategyproof mechanisms for linear regression.

Finally, we compare (families of) strategyproof mechanisms in terms of their approximation of the optimal squared loss, leveraging our characterization. Most importantly, we establish a lower bound of 2 on the approximation ratio of any strategyproof mechanism, which means that any mechanism that is even close to *ordinary least squares* regression must be manipulable .

1.2 Related Work

As discussed above, our work is most closely related to that of Perote and Perote-Peña [2004] and Dekel et al. [2010]. Here we try to give a broader picture of the state of research on machine learning algorithms that are robust to strategic noise. This research can be categorized using three key axes: (i) manipulable information, (ii) goal of the agents, and (iii) use of payments and incentive guarantees.

On the first axis, like us, most papers assume that independent variables (or *feature vectors* in the language of classification) are public information, and dependent variables (labels) are private, manipulable information [Dekel et al., 2010, Meir et al., 2012, Perote and Perote-Peña, 2003, 2004], though some papers also design algorithms robust to strategic feature vectors [Dong et al., 2017, Hardt et al., 2016]. Meir et al. [2012] provide strong positive results for designing strategyproof classifiers when there are either only two classifiers, or the agents are interested in a shared set of input points. On the other hand, Hardt et al. [2016] study the problem of constructing classifiers that are robust to agents strategically misreporting their *feature vector*, in order to trick the algorithm into misclassifying them. Their setting is modeled as a one-shot Stackelberg game. The more recent work of Dong et al. [2017] models the same problem in an online setting; they provide guarantees that ensure that the problem is convex, and, therefore, they are able to derive a computationally efficient learning algorithm that has diminishing *Stackelberg regret*.

On the second axis, one line of research focuses on agents motivated by privacy concerns, with a tradeoff between accuracy and privacy [Cai et al., 2015, Cummings et al., 2015]; another focuses on agents who want the algorithm to make accurate assessment on their own sample, even if this reduces the overall accuracy. This form of strategic manipulation has been studied for estimation [Caragiannis et al., 2016], classification [Meir et al., 2011, 2010, 2012], and regression [Dekel et al., 2010, Perote and Perote-Peña, 2004] problems. Our problem falls squarely into the second category.

Finally, on the third axis, various papers differ on whether monetary payments to agents are allowed [Cai et al., 2015], and on how strongly to guarantee truthful reporting: the stronger strategyproofness requirement [Meir et al., 2012, Perote and Perote-Peña, 2003, 2004] versus the weaker Bayes-Nash incentive compatibility [Cummings et al., 2015, Ioannidis and Loiseau, 2013]. Our work falls into the literature of mechanism design without money; we study linear regression mechanisms that enforce strategyproofness without paying the agents, or asking the agents to pay.

2 MODEL

Let $[k] \triangleq \{1, \ldots, k\}$ be the set of first k natural numbers, and $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ be the extended real line. Given numbers $t_1, \ldots, t_k \in \overline{\mathbb{R}}$, let $\min(t_1, \ldots, t_k)$ denote the smallest value, and $\min^j(t_1, \ldots, t_k)$ denote the j^{th} smallest value. Let $\text{med}(t_1, \ldots, t_k)$ denote their median: when k is odd, this is equal to $\min^{(k+1)/2}(t_1, \ldots, t_k)$, but when k is even, this could be either $\min^{k/2}(t_1, \ldots, t_k)$ (the "left median") or $\min^{k/2+1}(t_1, \ldots, t_k)$ (the "right median").[1]

Our work focuses on the problem of linear regression, i.e., fitting a hyperplane through given data. Let $N = [n]$. We are given a collection of data points $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ are called the *independent* and *dependent* variables of point i, respectively. Let $\overline{\boldsymbol{x}_i} = (\boldsymbol{x}_i, 1)$. Our goal is to find a vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_0) \in \mathbb{R}^{d+1}$ such that $\boldsymbol{\beta}^T \overline{\boldsymbol{x}_i} = \boldsymbol{\beta}_1^T \boldsymbol{x}_i + \beta_0$ is a good approximation of y_i for each $i \in N$. The quantity $r_i = y_i - \boldsymbol{\beta}^T \overline{\boldsymbol{x}_i}$ is called the residual of point i.

[1] This is different from the standard definition, which takes the average of the left and right medians, but necessary to ensure incentive guarantees.

Strategic setting. We study a setting in which each data point $p_i = (\boldsymbol{x}_i, y_i)$ is provided by a strategic agent i. We also denote the set of agents by N. Following Perote and Perote-Peña [2004] and Dekel et al. [2010], we assume that the independent variables $\boldsymbol{x} = (\boldsymbol{x}_i)_{i \in N}$ constitute *public* information, which the agents cannot manipulate. Each agent i holds the dependent variable y_i as private information, and may report a different value $\widetilde{y}_i$ in order to receive a more preferred outcome. Thus, the principal observes the reported data points $\widetilde{\mathcal{D}} = (\boldsymbol{x}_i, \widetilde{y}_i)_{i \in N}$. Let us denote $\boldsymbol{y} = (y_i)_{i \in N}$ and $\widetilde{\boldsymbol{y}} = (\widetilde{y}_i)_{i \in N}$.

Mechanisms. Because the agents cannot change $\boldsymbol{x}$, we can effectively treat it as fixed. A mechanism for linear regression $M^{\boldsymbol{x}}$ is therefore defined for given public information $\boldsymbol{x}$, takes as input reported private information $\widetilde{\boldsymbol{y}}$, and returns a vector $\boldsymbol{\beta}$. We omit $\boldsymbol{x}$ when it is clear from the context.

Agent preferences. When a mechanism returns $\boldsymbol{\beta}$, we say that the outcome for agent i is $\widehat{y}_i(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \boldsymbol{x}_i$. We omit $\boldsymbol{\beta}$ when it is clear from the context. The agent only cares about her own outcome $\widehat{y}_i$, and would like it to be as close to y_i as possible. Formally, we assume that agent i has *single-peaked preferences* [Black, 1958, Moulin, 1980] over $\widehat{y}_i$ with peak at y_i. We represent the weak preference relation by $\succcurlyeq_i$ and the strict preference relation by $\succ_i$. Formally, for all $a, b \in \mathbb{R}$, $y_i > a \geq b$ or $y_i < a \leq b$ must imply $y_i \succ_i a \succcurlyeq_i b$.

Game-theoretic desiderata. Our goal is to prevent agents from misreporting their private information. The game theory literature offers a strong desideratum under which agents have no incentive to manipulate even if they have know what the other agents would report.

Definition 2.1 (Strategyproofness). A mechanism $M^{\boldsymbol{x}}$ is called *strategyproof* (SP) if each agent weakly prefers truthfully reporting her private information to misreporting it, regardless of the reports of the other agents. Formally, for each $i \in N$, $y_i \in \mathbb{R}$, and $\widetilde{\boldsymbol{y}} \in \mathbb{R}^n$, we need $\widehat{y}_i(M^{\boldsymbol{x}}(y_i, \widetilde{\boldsymbol{y}}_{-i})) \succcurlyeq_i \widehat{y}_i(M^{\boldsymbol{x}}(\widetilde{\boldsymbol{y}}))$. Note that this must hold for any possible single-peaked preferences the agent may have.

While no individual agent can benefit from misreporting under a strategyproof mechanism, a group of agents may still be able to collude, and benefit by simultaneously misreporting. This can be prevented by imposing a stronger desideratum.

Definition 2.2 (Group Strategyproofness). A mechanism $M^{\boldsymbol{x}}$ is called *group strategyproof* (GSP) if no coalition of agents can simultaneously misreport in a way that no agent in the coalition is strictly worse off and some agent in the coalition is strictly better off, irrespective of the reports of the other agents. Formally, for each $S \subseteq N$, $\boldsymbol{y}_S = (y_i)_{i \in S} \in \mathbb{R}^{|S|}$, and $\widetilde{\boldsymbol{y}} \in \mathbb{R}^n$, it should not be the case that $\widehat{y}_i(M^{\boldsymbol{x}}(\widetilde{\boldsymbol{y}})) \succcurlyeq_i \widehat{y}_i(M^{\boldsymbol{x}}(\boldsymbol{y}_S, \widetilde{\boldsymbol{y}}_{N \setminus S}))$ for every $i \in S$, and the preference is strict for at least one $i \in S$.

The game theory literature also considers a weaker notion of group strategyproofness in which not all the agents in a manipulating coalition should be strictly better off. We do not consider this notion because our group strategyproof mechanisms are able to satisfy the stronger notion.

Note that we do not assume that the data points are generated by an underlying statistical process. Our results are independent of how the data points were generated.

3 FAMILIES OF STRATEGYPROOF MECHANISMS

In this section, we analyze families of (group) strategyproof mechanisms for linear regression. Our results generalize existing families of mechanisms, and propose novel families.

3.1 Empirical Risk Minimization with the L_1 Loss

Consider a single dimensional setting, in which each agent i has a private value y_i, reports a possibly different value $\widetilde{y}_i$, and the mechanism returns a single value $\widehat{y}$. Each agent i has single-peaked

preferences over $\widehat{y}$ with peak at y_i. This corresponds to the special case of our setting in which $\boldsymbol{x}_i = \boldsymbol{x}_j$ for all $i, j \in N$, or alternatively, the dimension $d = 0$. In this setting, it has long been known that choosing the *median* of the reported values achieves group strategyproofness [Dummett and Farquharson, 1961]. It can be shown that the median minimizes the sum of absolute (L_1) losses with respect to the reports, i.e., given $\boldsymbol{y}$, it chooses $\arg\min_{y \in \mathbb{R}} \sum_{i=1}^{n} |y - y_i|$, with an appropriate tie-breaking when n is even. In the machine learning terminology, the median is the empirical risk minimizer (ERM) with the L_1 loss.

Inspired by this, Dekel et al. [2010] study ERM with the L_1 loss in a more general regression setting, and show that it remains group strategyproof. Specifically, they focus on finding a (potentially non-linear) regression function $f : \mathbb{R}^d \to \mathbb{R}$ from a given convex set $\mathcal{F}$. Given $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$, define the empirical L_1 risk of a regression function $f \in \mathcal{F}$ as $\widehat{R}(f, \mathcal{D}) = \sum_{i \in N} |y_i - f(\boldsymbol{x}_i)|$. Let $\|\cdot\| : \mathcal{F} \to \mathbb{R}$ be a strictly convex function. They show that minimizing the empirical L_1 risk, and breaking ties among the optimal solutions by minimizing $\|\cdot\|$ is group strategyproof. We refer to this mechanism by L_1-ERM, which is presented as Algorithm 1. For linear regression, this approach is known by various names in the literature, such as Least Absolute Deviations (LAD), Minimum Sum of Absolute Errors (MSAE), or Least Absolute Value (LAV). The tie-breaking step is crucially required because the empirical L_1 risk may have multiple minimizers.

ALGORITHM 1: L_1-ERM (ERM with the L_1 loss)

Input: Data points $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$, convex set $\mathcal{F}$ of regression functions, strictly convex function $\|\cdot\| : \mathcal{F} \to \mathbb{R}$.
Output: Function $f^* \in \mathcal{F}$.

$\forall f \in \mathcal{F},\ \widehat{R}(f, \mathcal{D}) \triangleq \sum_{i \in N} |y_i - f(\boldsymbol{x}_i)|$;
$r^* \leftarrow \inf_{f \in \mathcal{F}} \widehat{R}(f, \mathcal{D})$;
return $f^* \leftarrow \arg\min_{f \in \mathcal{F} : \widehat{R}(f, \mathcal{D}) = r^*} \|f\|$;

We present a generalization of their mechanism while retaining group strategyproofness. In particular, we extend the objective function $\widehat{R}$ in two ways: i) we allow a weighted L_1 loss, in which the loss of each agent i is multiplied by a weight $w_i^{\boldsymbol{x}}$, and ii) we allow adding a convex regularizer $h : \mathcal{F} \to \mathbb{R}$. Note that regularization is widely used in machine learning to prevent ERM from overfitting. Our generalization, which we term *generalized L_1-ERM*, is presented as Algorithm 2. While we are only interested in linear regression, we note that generalized L_1-ERM works for the general regression setting of Dekel et al. [2010].

ALGORITHM 2: Generalized L_1-ERM (Regularized ERM with a weighted L_1 loss)

Input: Data points $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$, convex hypothesis space $\mathcal{F}$, constants $(w_i^{\boldsymbol{x}})_{i \in N}$, convex regularizer $h : \mathcal{F} \to \mathbb{R}$, strictly convex function $\|\cdot\| : \mathcal{F} \to \mathbb{R}$.
Output: Function $f^* \in \mathcal{F}$.

$\forall f \in \mathcal{F},\ \widehat{R}(f, \mathcal{D}) \triangleq \sum_{i \in N} w_i^{\boldsymbol{x}} \cdot |y_i - f(\boldsymbol{x}_i)| + h(f)$;
$r^* \leftarrow \inf_{f \in \mathcal{F}} \widehat{R}(f, \mathcal{D})$;
return $f^* \leftarrow \arg\min_{f \in \mathcal{F} : \widehat{R}(f, \mathcal{D}) = r^*} \|f\|$;

Theorem 3.1. *Generalized L_1-ERM is a group strategyproof regression mechanism.*

Our proof, presented in Appendix A for completeness, essentially mirrors the proof of Dekel et al. [2010]; we identify three steps in their proof where they use the structure of the risk function $\widehat{R}$, and observe that these steps follow through with our more general risk function.

There are several potential advantages of generalized L_1-ERM over the vanilla L_1-ERM. First, generalized L_1-ERM allows eliminating the tie-breaking step if the new risk function is guaranteed to have a unique minimizer. For instance, adding a *strictly convex* regularizer would achieve this.

Second, for the aforementioned single dimensional setting, Moulin [1980] proved that every strategyproof[2] and anonymous[3] mechanism is a *generalized median*: for every $\alpha_1, \ldots, \alpha_{n+1} \in \overline{\mathbb{R}}$, the corresponding generalized median returns $\text{med}\{y_1, \ldots, y_n, \alpha_1, \ldots, \alpha_{n+1}\}$. Here, $\{\alpha_j\}_{j\in[n+1]}$ are called "phantoms". We can alternatively view this as returning $\arg\min_{y\in\mathbb{R}} \sum_{i\in[n]} |y - y_i| + h(y)$, where $h(y) = \sum_{j\in[n+1] \text{ s.t. } \alpha_j\in\mathbb{R}} |y - \alpha_j| + (k_{-\infty} - k_{\infty}) \cdot y$, and for $t \in \{-\infty, \infty\}$, $k_t = |\{j : \alpha_j = t\}|$.[4] Since $h(y)$ is a convex function, we can view it as a regularizer in our generalized L_1-ERM. Hence, for the single dimensional setting, generalized L_1-ERM covers all generalized medians. In contrast, L_1-ERM reduces to a specific mechanism in this family, the median.

Finally, algorithms that add convex regularization to L_1-ERM have been studied in the machine learning literature [Wang, 2013, Wang et al., 2006]; our generalization establishes group strategyproofness of these algorithms.

We also note that in the statistics literature, the vanilla L_1-ERM is treated as a member of the more general family of *quantile regression* mechanisms [Koenker and Bassett, 1978], which, given $q \in [0, 1]$, minimize the following empirical risk function:

$$\widehat{R}_q(f, \mathcal{D}) = \sum_{i\in N: y_i \geq f(\boldsymbol{x}_i)} q \cdot |y_i - f(\boldsymbol{x}_i)| + \sum_{i\in N: y_i < f(\boldsymbol{x}_i)} (1 - q) \cdot |y_i - f(\boldsymbol{x}_i)|. \tag{1}$$

L_1-ERM corresponds to the choice of $q = 0.5$. In the one-dimensional setting, other values of q correspond to different quantiles (i.e., correspond to $\min^k$ for various k), and thus induce strategyproof mechanisms. One might wonder if quantile regression remains strategyproof in higher dimensions. We answer this *negatively* by providing an example in Appendix C, in which the quantile regression mechanism for $q = 0.4$ is shown to violate strategyproofness. It is an interesting question to discover a strategyproof version of quantiles for linear regression.

3.2 Generalized Resistant Hyperplane Mechanisms

In this section, we introduce a novel family of strategyproof mechanisms for linear regression. Our family extends the known family of resistant line mechanisms from the statistics literature [Johnstone and Velleman, 1985], which were only defined for simple linear regression ($d = 1$), to higher dimensions. We first take a slight detour through a previous approach in the literature.

3.2.1 A Detour Through Clockwise Repeated Median Mechanisms. Perote and Perote-Peña [2004] introduced a novel family of mechanisms, which they termed *Clockwise Repeated Median* (CRM) mechanisms. CRM mechanisms are only defined for the special case of *simple linear regression*, i.e., for fitting a straight line through a set of points on a plane. In describing these mechanisms, we use scalar notations where possible. For instance, we use x_i to denote the x-coordinate of agent i,

[2]Moulin [1980] shows that for the single dimensional setting, strategyproofness is equivalent to group strategyproofness.
[3]A mechanism is anonymous if permuting the reports of the agents does not change the output of the mechanism. This is a reasonable desideratum in the single dimensional setting due to the absence of public information that distinguishes agents naturally.
[4]When all phantoms are finite, $h(y) = \sum_{j\in[n+1]} |y - \alpha_j|$. The term $|y - \alpha_j|$ has derivative 1 when $y > \alpha_j$, and -1 when $y < \alpha_j$. For $\alpha_j = -\infty$ (resp. ∞), we can mimic this effect by adding a different term whose derivative is always -1 (resp. 1).

and β_1 to denote the slope of the regression line. For CRM mechanisms to be well defined, we also need to assume that the set of points is "admissible".

Definition 3.2 (Admissible Set). A collection of data points $\mathcal{D} = (x_i, y_i)_{i \in N}$ is called *admissible* if $x_i \neq x_j$ for all distinct $i, j \in N$.

The CRM family is parametrized by two subsets of agents, $S, S' \subseteq N$. These subsets must be chosen based on the public information $\boldsymbol{x}$, and therefore can be treated as fixed. Informally, given $S, S' \subseteq N$, the (S, S')-CRM mechanism first computes the median *clockwise angle* (CWA), defined below, from each point $i \in S$ to points in S'. Then, it chooses the point $i^* \in S$ whose median CWA is the median of the median CWAs from all points in S. If the median CWA from point i^* is towards point $j^* \in S'$, then the mechanism returns the straight line passing through points i^* and j^*. Formally, the mechanism is defined as follows. Perote and Perote-Peña [2004] established the equivalence of this formal definition and the aforementioned informal description.

Definition 3.3 (CRM Mechanisms). Define the *clockwise angle* (CWA) from (x_i, y_i) to (x_j, y_j) as:

$$\mathrm{CWA}((x_i, y_i), (x_j, y_j)) = \pi + \mathrm{sign}(x_j - x_i) \cdot \frac{\pi}{2} + \mathrm{sign}\left(\frac{y_j - y_i}{x_j - x_i}\right) \left|\arctan\left(\frac{y_j - y_i}{x_j - x_i}\right)\right|. \tag{2}$$

Given $\mathcal{D} = (x_i, y_i)_{i \in N}$ and $S, S' \subseteq N$, let the *directing angle* be defined as:

$$\mathrm{DA}(S, S') = \operatorname*{med}_{i \in S} \operatorname*{med}_{j \in S' : j \neq i} \mathrm{CWA}((x_i, y_i), (x_j, y_j)). \tag{3}$$

Then, the (S, S')-CRM mechanism returns the line $\boldsymbol{\beta} = (\beta_1, \beta_0)$ given by:

$$\begin{aligned} \beta_1 &= \tan\left[\mathrm{DA}(S, S') - \pi - \frac{\pi}{2} \cdot \mathrm{sign}\left(\mathrm{DA}(S, S') - \pi\right)\right], \\ \beta_0 &= \operatorname*{med}_{i \in S} (y_i - \beta_1 \cdot x_i). \end{aligned} \tag{4}$$

First, we notice that the definition of the CRM family uses three medians: two to define the directing angle $DA(S, S')$, and one to define the y-intercept β_0. Each median, when taken over an even number of values, can be the left median or the right median. While Perote and Perote-Peña [2004] do not mention how these choices should be made, it is easy to check that in order to achieve the desired incentive properties, these choices cannot be made independently of each other. Later, we present a generalization which captures the different feasible choices in a simpler form.

Perote and Perote-Peña [2004] claimed that the (S, S')-CRM mechanism is strategyproof when $S \subseteq S'$ or $S \cap S' = \emptyset$, and provided an involved, geometric proof. However, we have identified a mistake in their proof. In fact, we have found two counterexamples, one with $S \subseteq S'$ and one with $S \cap S' = \emptyset$, for which the corresponding (S, S')-CRM mechanisms violate strategyproofness, thus disproving their claim. These counterexamples are presented in Figure 1,

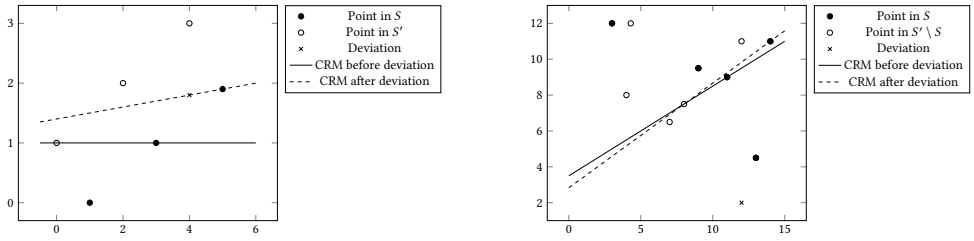
Example 3.4 (Example with $S \cap S' = \emptyset$.). This example is shown in Figure 1a. Points in filled dots are in S, while points in empty dots are in S'. The coordinates of these points are as follows.

$$S = \{(1, 0), (3, 1), (5, 1.9)\}, S' = \{(0, 1), (2, 2), (4, 3)\}.$$

Notice that $S \cap S' = \emptyset$. Also, $|S|$ and $|S'|$ are odd, alleviating the need to choose between left and right medians in the CRM definition.

When the agents truthfully report, one can check that CRM returns the line connecting points $(3, 1)$ from S and $(0, 1)$ from S'. This line is given by the equation $y = 1$.

Suppose that the agent i controlling the point at $x = 4$ manipulates, and reports $\widetilde{y_i} = 1.8$ instead of $y_i = 3$. The new point is depicted with a cross. One can check that this causes the CRM mechanism to switch to the dashed line ($y = 0.1 \cdot x + 1.4$), which makes agent i strictly better off, and violates strategyproofness.



(a) $S \cap S' = \emptyset$ (b) $S \subseteq S'$

Fig. 1. Counterexamples showing violation of strategyproofness of (S, S')-CRM mechanisms. Figure 1a shows a case with $S \cap S' = \emptyset$, while Figure 1b shows a case with $S \subseteq S'$.

Example 3.5 (Example with $S \subseteq S'$.). This example is shown in Figure 1b. Points in S (thus also in S') are depicted with filled dots, while points in $S' \setminus S$ are depicted with empty dots. The coordinates of these points are as follows.

$$S = \{(3, 12), (9, 9.5), (11, 9), (13, 4.5), (14, 11)\}, S' = S \cup \{(4, 8), (4.3, 12), (7, 6.5), (8, 7.5), (12, 11)\}.$$

Notice that $S \subseteq S'$. Further, $|S|$ is odd, and $|S'|$ is even (thus, for each $i \in S$, $|S' \setminus \{i\}|$ is odd), once again eliminating the need to choose between the left and the right medians in the CRM definition.

When all points are reported truthfully, one can check that the CRM mechanism chooses the solid line ($3y = 2x + 8$). Suppose now that agent i with point $(12, 11)$ reports $\widetilde{y}_i = 0$, instead of $y_i = 11$. Then, the CRM mechanism chooses the dashed line, which makes agent i strictly better off, again violating strategyproofness.

Nevertheless, we have been able to identify a subset of the CRM family, for which we can establish strategyproofness (in fact, group strategyproofness). In particular, we replace $S \subseteq S'$ with the more restrictive condition $S = S'$, and for $S \cap S' = \emptyset$, we either add $|S| = 1$ or $|S'| = 1$, or replace it with a stricter condition that we define below.

Definition 3.6 (Separable Sets of Points in a Plane). Let S, S' be two sets of points in $\mathbb{R}^2$. We say that S and S' are *separable* if $\max_{i \in S} x_i < \min_{j \in S'} x_j$ or $\max_{j \in S'} x_j < \min_{i \in S} x_i$. In other words, it should be possible to separate them by a vertical line.

Note that separability of S and S' implies $S \cap S' = \emptyset$. We now present a corrected version of the result of Perote and Perote-Peña [2004], and claim the stronger guarantee of group strategyproofness. We do not present a proof as we later introduce a much broader family of mechanisms, and prove their group strategyproofness directly.

Theorem 3.7. *Given $S, S' \subseteq N$, the (S, S')-CRM mechanism is group strategyproof if one of the following conditions holds.*

(1) $S = S'$.
(2) S and S' are separable.
(3) $S \cap S' = \emptyset$ and $\min(|S|, |S'|) = 1$.

The third condition partially resembles dictatorship as the agent in the singleton set is guaranteed to have zero residual (i.e., be on the regression line).

3.2.2 Generalized Resistant Line Mechanisms on a Plane. In this section, our goal is to introduce a novel family of group strategyproof mechanisms that include, as special cases, the mechanisms covered in the three cases of Theorem 3.7. Our starting point is the family of *resistant line* (RL)

mechanisms from the statistics literature [Johnstone and Velleman, 1985], which Perote and Perote-Peña [2004] showed to be equivalent to the case of separable S and S'.

The standard formulation of the RL mechanism involves three sets $L, M, R \subseteq N$ such that $\max_{i\in L} x_i < \min_{i\in M} x_i$ and $\max_{i\in M} x_i < \min_{i\in R} x_i$, and returns a line $\boldsymbol{\beta} = (\beta_1, \beta_0)$ given by

$$\text{med}_{i\in L}\, y_i - \beta_1 \cdot x_i - \beta_0 = \text{med}_{i\in R}\, y_i - \beta_1 \cdot x_i - \beta_0 = 0.$$

That is, the line makes the median residuals in L and R zero. It is known that this equation yields a unique solution [Johnstone and Velleman, 1985]. Perote and Perote-Peña [2004] showed that this is identical to the (L, R)-CRM mechanism. Indeed, separability of L and R makes clockwise angles from points in L to points in R monotonic in (and thus replaceable by) slopes, yielding the following formulation for the (L, R)-CRM mechanism.

$$\beta_1 = \text{med}_{i\in L}\, \text{med}_{j\in R}\, \tfrac{y_j - y_i}{x_j - x_i},$$
$$\beta_0 = \text{med}_{i\in L}\, y_i - \beta_1 x_i = \text{med}_{j\in R}\, y_j - \beta_1 x_j.$$

The alternative definition of $\beta_0 = \text{med}_{j\in R}(y_j - \beta_1 \cdot x_j)$ follows from the fact that if the line passes through $i^* \in L$, it is directed towards the point in R which is at the median angle or slope, and thus bisects R in addition to bisecting L.

Along with Theorem 3.7, this observation establishes group strategyproofness of all resistant line mechanisms. Two popular mechanisms from this family are the Brown-Mood mechanism [Brown and Mood, 1951], in which L and R each contain half of the points while M is empty, and the Tukey mechanism [Tukey, 1977], in which L, M, and R each contain a third of the points.

Our next step is to extend this family. A natural idea is that instead of making the *median* residuals from S and S' zero, we make the k^{th} smallest residual in S and the $(k')^{\text{th}}$ smallest residual in S' zero, for fixed $k \in [|S|]$ and $k' \in [|S'|]$.

Definition 3.8 (Generalized Resistant Line (GRL) Mechanisms). Given separable sets $S, S' \subseteq N$, $k \in [|S|]$, and $k' \in [|S'|]$, the (S, S', k, k')-generalized resistant line (GRL) mechanism returns the line $\boldsymbol{\beta} = (\beta_1, \beta_0)$ given by

$$\min^{k}_{i\in S} y_i - \beta_1 x_i - \beta_0 = \min^{k'}_{j\in S'} y_i - \beta_1 x_j - \beta_0 = 0. \tag{5}$$

We show that these mechanisms are well defined (i.e., there is a unique solution to Equation (5)), and they are group strategyproof. Once again, we omit the proof because we later introduce an even broader family of mechanisms, for which we prove these results directly.

Theorem 3.9. *For separable sets $S, S' \subseteq N$, $k \in [|S|]$ and $k' \in [|S'|]$, the (S, S', k, k')-generalized resistant line mechanism is well defined and group strategyproof.*

While it is clear that generalized resistant line mechanisms cover the second case of Theorem 3.7 (i.e., separable S and S'), we surprisingly find that they also cover the first case ($S = S'$) and the third case ($S \cap S' = \emptyset$ and $\min(|S|, |S'|) = 1$). That is, Theorem 3.9 strictly generalizes Theorem 3.7. The proof of the next result is in Appendix A.

Lemma 3.10. *For $S \subseteq N$, the (S, S)-CRM mechanism is a generalized resistant line mechanism. For $S, S' \subseteq N$ with $S \cap S' = \emptyset$ and $\min(|S|, |S'|) = 1$, the (S, S')-CRM mechanism is also a generalized resistant line mechanism.*

3.2.3 Generalized Resistant Hyperplane Mechanisms in High Dimensions. Surprisingly, the statistics literature does not offer an extension of resistant line mechanisms to higher dimensions. In our efforts to do so, we quickly realized that this is a non-trivial task. In two dimensions, a generalized resistant line mechanism takes two subsets of data points separable by a vertical line,

and returns the regression line which makes prescribed percentiles of residuals in each set zero. In $d+1$ dimensions (recall that $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$), it seems natural to take $d+1$ "separable" subsets of data points, and return the regression hyperplane which makes prescribed percentiles of residuals in each set zero. However, the separability condition must now ensure existence of a unique hyperplane with this property, even if we ignore our game-theoretic desiderata.

In resolving this issue, we make a connection to the literature on the *Ham Sandwich Theorem* and its generalizations. Hereinafter, given a hyperplane H, we denote by H^+ and H^- its positive and negative closed half-spaces, respectively. A basic version of the ham sandwich theorem due to Stone and Tukey [1942] states that given k continuous measures $\mu_1, \ldots, \mu_k$ on $\mathbb{R}^k$, there exists a hyperplane H such that $\mu_i(H^+) = 1/2$ for each $i \in [k]$. A discrete version of this result due to Elton and Hill [2011] states that given k finite sets $S_1, \ldots, S_k \subseteq \mathbb{R}^k$, there exists a hyperplane H such that for each $i \in [k]$, H "bisects" S_i and $H \cap S_i \neq \emptyset$. Here, we say that a hyperplane H bisects a set of points S if each *closed* half-space of H contains at least $\lceil |S|/2 \rceil$ points.

For linear regression, this implies that given $S_1, \ldots, S_{d+1} \subseteq \mathcal{D}$, there exists a "resistant hyperplane" which makes the median residual from S_t zero, for each $t \in [d+1]$. While this seems like a natural generalization of resistant line mechanisms, it is easy to check that such a hyperplane is not always unique, even in two dimensions. Further, if the median is replaced by other percentiles, the existence is no longer guaranteed.[5]

Steiger and Zhao [2010] provide a generalization that *almost* perfectly fits our needs. They show that under certain conditions on $S_1, \ldots, S_{d+1}$, we can find a unique hyperplane H which contains a given number of points from each set in its negative half-space. This discrete result builds upon previous continuous variants of the result [Bárány et al., 2008, Breuer, 2010]. We first define a condition they require, which also plays a key role in our result.

Definition 3.11 (Well Separable Sets [Kermer and Németh, 1973]). Given $t \in [k+1]$, finite sets $S_1, \ldots, S_t$ of points in $\mathbb{R}^k$ are called *well separable* if for all disjoint $I, J \subseteq [t]$, there exists a hyperplane H such that $S_i \subset H^+ \setminus H$ for each $i \in I$ and $S_j \subset H^- \setminus H$ for each $j \in J$, i.e., H separates $\cup_{i \in I} S_i$ from $\cup_{j \in J} S_j$ by putting them in different *open* half-spaces.

Well separable sets are sometimes called *affinely independent* sets [Breuer, 2010]. Well separability is equivalent to various other conditions [Breuer, 2010, Steiger and Zhao, 2010]. In what follows, $\mathrm{Conv}(\cdot)$ denotes the convex hull.

Proposition 3.12. *For $t \in [k+1]$, finite sets $S_1, \ldots, S_t$ in $\mathbb{R}^k$ are well separable if and only if:*

(1) For all choices of $(x_i \in \mathrm{Conv}(S_i))_{i \in [t]}$, the affine hull of $x_1, \ldots, x_t$ is a $(t-1)$-dimensional flat.
(2) No $(t-2)$-dimensional flat has a nonempty intersection with $\mathrm{Conv}(S_i)$ for each $i \in [t]$.
(3) $\mathrm{Conv}(S_1), \ldots, \mathrm{Conv}(S_t)$ are well separable.

Steiger and Zhao [2010] impose an additional condition, which we eliminate in our work.

Definition 3.13 (Weak General Position). Finite sets $S_1, \ldots, S_k \subset \mathbb{R}^k$ are said to have *weak general position* if for every choice of $(x_i \in S_i)_{i \in [k]}$, the affine hull of $x_1, \ldots, x_k$ is a $(k-1)$-dimensional flat which contains no other point of $\cup_{i \in [k]} S_i$.

Theorem 3.14 ([Steiger and Zhao, 2010]). *If finite sets $S_1, \ldots, S_k \subset \mathbb{R}^k$ are well separable and have weak general position, then given any choice of $k_i \in [|S_i|]$ for $i \in [k]$, there exists a unique hyperplane H such that for each $i \in [k]$, $H \cap S_i \neq \emptyset$ and $|H^- \cap S_i| = k_i$.*

This result gives us *almost* what we want for linear regression in $\mathbb{R}^{d+1}$. Given a family of sets $S_1, \ldots, S_{d+1} \subseteq \mathcal{D}$ that are well separable and have weak general position, and $k_t \in [|S_t|]$ for

[5] Recall that even in two dimensions, we needed an additional condition on the sets S and S': separability by a vertical line.

$t \in [d+1]$, it ensures the existence of a unique hyperplane which makes the k_t^{th} smallest residual in each set S_t zero. However, it falls short of our requirements in two key aspects.

- Theorem 3.14 allows the assignment of points in $\mathcal{D}$ to sets $S_1, \ldots, S_{d+1}$ to depend on the private information $\boldsymbol{y}$. For strategyproofness, we need this assignment based solely on the public information $\boldsymbol{x}$. Recall that in two dimensions, we chose sets S and S' separable by a *vertical* line. We choose the $d+1$ sets so that they are well separable in the d-dimensional public information space,[6] and establish group strategyproofness using a technical lemma, which may be of independent interest.
- While we only want to make the k_t^{th} smallest residual in each S_t zero, Steiger and Zhao [2010] aim for something stronger: they want the number of points from each S_t in the negative closed halfspace to be exactly k_t. This necessitates their weak general position assumption, which we relax.

We are now ready to present our results. They closely mirror, but do not make use of, the results of Steiger and Zhao [2010]. We revert to using notation of our linear regression setting. Recall that a hyperplane $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_0)$ passes through $(\boldsymbol{x}_i, \boldsymbol{\beta}^T\ \overline{\boldsymbol{x}_i})$ for each $i \in N$, where $\overline{\boldsymbol{x}_i} = (\boldsymbol{x}_i, 1)$.

Definition 3.15. Given a family $\mathcal{S} = (S_1, \ldots, S_k)$ of nonempty disjoint subsets of N, and a set of points $P = (p_i)_{i \in N}$, define the partition function $\mathcal{P}(P, \mathcal{S}) = (P_t)_{t \in [k]}$, where $P_t = (p_i)_{i \in S_t}$ for each $t \in [k]$. That is, $\mathcal{P}(P, \mathcal{S})$ partitions the set of points P based on index sets from $\mathcal{S}$.

Definition 3.16 (Publicly Separable Sets of Agents). We say that a family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of nonempty disjoint subsets of N is *publicly separable* if $\mathcal{P}(\boldsymbol{x}, \mathcal{S})$ is well separable.

Definition 3.17 (Generalized Resistant Hyperplane (GRH) Mechanisms). Given a family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of publicly separable sets of agents, and $\boldsymbol{k} = (k_t)_{t \in [d+1]}$ with $k_t \in [|S_t|]$ for $t \in [d+1]$, the $(\mathcal{S}, \boldsymbol{k})$-generalized resistant hyperplane (GRH) mechanism returns a hyperplane $\boldsymbol{\beta}$ such that $\min_{i \in S_t}^{k_t}(r_i \triangleq y_i - \boldsymbol{\beta}^T\ \overline{\boldsymbol{x}_i}) = 0$ for each $t \in [d+1]$. That is, it makes the k_t^{th} smallest residual from every set $S_t \in \mathcal{S}$ zero.

We first need to establish that the GRH mechanisms are well defined, i.e., the hyperplane they seek is guaranteed to exist and be unique. To that end, we prove a useful technical lemma, which may be of independent interest.

Lemma 3.18 (Hyperplane Comparison Lemma). *Given a family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of publicly separable sets of agents, and two distinct hyperplanes $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$, there exists a set $S_t \in \mathcal{S}$ such that either $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{x}_i} < (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{x}_i}$ for all $i \in S_t$, or $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{x}_i} > (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{x}_i}$ for all $i \in S_t$.*

Proof. Consider the intersection of the two hyperplanes in $\mathbb{R}^{d+1}$, and let W be its projection on $\mathbb{R}^d$ (the public information space). Note that W is a $(d-1)$-dimensional hyperplane in $\mathbb{R}^d$. Given an *open* half-space of W (say W^+), let Z be the set of points $\mathbb{R}^{d+1}$ whose projection on $\mathbb{R}^d$ lies in W^+. Then, either $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{p}} > (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{p}}$ for all $\boldsymbol{p} \in Z$, or $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{p}} < (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{p}}$ for all $\boldsymbol{p} \in Z$, where $\overline{\boldsymbol{p}} = (\boldsymbol{p}, 1)$.

Let $\mathcal{P}(\boldsymbol{x}, \mathcal{S}) = (X_1, \ldots, X_{d+1})$. Because $\mathcal{S}$ is publicly separable, $X_1, \ldots, X_{d+1}$ are well separable. By Proposition 3.12, no $(d-1)$-dimensional flat has a nonempty intersection with $\text{Conv}(X_t)$ for each $t \in [d+1]$. Because W is a $(d-1)$-dimensional flat, there exists $t \in [d+1]$ such that W does not intersect $\text{Conv}(X_t)$, i.e., X_t lies entirely in an *open* half-space of W. Using the previous argument, either $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{x}_i} < (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{x}_i}$ for all $i \in S_t$, or $(\boldsymbol{\beta}^1)^T\ \overline{\boldsymbol{x}_i} > (\boldsymbol{\beta}^2)^T\ \overline{\boldsymbol{x}_i}$ for all $i \in S_t$. ■

Proposition 3.19. *Generalized resistant hyperplane mechanisms are well defined. That is, given a family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of publicly separable sets of agents, and $\boldsymbol{k} = (k_1, \ldots, k_{d+1})$ with $k_t \in [|S_t|]$ for $t \in [d+1]$, there exists a unique hyperplane $\boldsymbol{\beta}$ for which $\min_{i \in S_t}^{k_t} y_i - \boldsymbol{\beta}^T\ \overline{\boldsymbol{x}_i} = 0$ for each $t \in [d+1]$.*

[6] While Theorem 3.14 uses $d+1$ well separable sets in $\mathbb{R}^{d+1}$, even $\mathbb{R}^d$ allows up to $d+1$ well separable sets.

Proof. First, we show that *if* such a hyperplane exists, it must be unique. Suppose for contradiction that there are two distinct hyperplanes $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ which make the k_t^{th} smallest residual from every $S_t \in \mathcal{S}$ zero. By the hyperplane comparison lemma (Lemma 3.18), there exists $S_t \in \mathcal{S}$ such that either $(\boldsymbol{\beta}^1)^T \overline{\boldsymbol{x}_i} < (\boldsymbol{\beta}^2)^T \overline{\boldsymbol{x}_i}$ for all $i \in S_t$, or $(\boldsymbol{\beta}^1)^T \overline{\boldsymbol{x}_i} > (\boldsymbol{\beta}^2)^T \overline{\boldsymbol{x}_i}$ for all $i \in S_t$. Without loss of generality, suppose it is the former. Then, at least k_t points in S_t which have a non-positive residual under $\boldsymbol{\beta}^2$ have a negative residual under $\boldsymbol{\beta}^1$, contradicting the fact that $\boldsymbol{\beta}^1$ makes the k_t^{th} smallest residual from S_t zero.

For proving existence, we use a counting technique. Create two bipartite graphs $G = (V \cup W, E)$ and $G' = (V' \cup W, E')$. Let V (resp. V') contain a vertex $v_{\boldsymbol{k}}$ (resp. $v'_{\boldsymbol{k}}$) corresponding to each $\boldsymbol{k} = (k_1, \ldots, k_{d+1})$ such that $k_t \in [|S_t|]$ for each $t \in [d+1]$. Thus, $|V| = |V'| = \prod_{t=1}^{d+1} |S_t|$. Let W contain a vertex $w_{\boldsymbol{\beta}}$ corresponding to every *traversal* hyperplane $\boldsymbol{\beta}$, i.e., every hyperplane that passes through at least one point from each set $S_t \in \mathcal{S}$.

In graph G, we draw an edge between $v_{\boldsymbol{k}}$ and $w_{\boldsymbol{\beta}}$ if $\boldsymbol{\beta}$ makes the k_t^{th} smallest residual zero in each $S_t \in \mathcal{S}$. For constructing graph G', we fix an arbitrary ordering of points in each set, so that we can write $S_t = \{i_1^t, \ldots, i_{|S_t|}^t\}$. Then, we draw an edge in G' between $v'_{\boldsymbol{k}}$ and $w_{\boldsymbol{\beta}}$ if $\boldsymbol{\beta}$ passes through point $i_{k_t}^t$ for each $t \in [d+1]$.

Our goal is to show that each vertex $v_{\boldsymbol{k}} \in V$ has exactly one incident edge in graph G. We prove this through a sequence of claims. First, we argue that each vertex $v'_{\boldsymbol{k}} \in V'$ has exactly one incident edge in graph G'. The fact that it has *at least* one incident edge follows from the fact that any set of $d+1$ points in $\mathbb{R}^{d+1}$ (in particular, $T = \{i_{k_t}^t\}_{t \in [d+1]}$) lie on a hyperplane. If $v'_{\boldsymbol{k}}$ has two or more incident edges, then there exist two distinct hyperplanes $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ which pass through all points in T. Then, their intersection $\boldsymbol{\beta}^*$, which is a $(d-1)$-dimensional flat in $\mathbb{R}^{d+1}$, must also pass through all points in T. Let $\mathcal{P}(\boldsymbol{x}, \mathcal{S}) = (X_1, \ldots, X_{d+1})$. Then, the projection of $\boldsymbol{\beta}^*$ on the public information space $\mathbb{R}^d$ is a $(d-1)$-dimensional hyperplane in $\mathbb{R}^d$ which intersects each X_t (and thus each $\text{Conv}(X_t)$). However, $\mathcal{S}$ is a publicly separable family, i.e., $X_1, \ldots, X_{d+1}$ are well separable in $\mathbb{R}^d$. This violates the first condition of Proposition 3.12.

Since each vertex in V' has exactly one incident edge, we have $|E'| = |V'| = \prod_{t=1}^{d+1} |S_t|$. We next argue that $|E| = |E'|$. Take a vertex $w_{\boldsymbol{\beta}} \in W$. Note that if hyperplane $\boldsymbol{\beta}$ passes through a_t points from each $S_t \in \mathcal{S}$, then it has degree $\prod_{t=1}^{d+1} a_t$ in both G and G'. Since each vertex in W has the same degree in both graphs, we have $|E| = |E'| = |V'| = |V|$.

Finally, we already established that if there is a hyperplane which makes the k_t^{th} smallest residual in each S_t zero, then it must be unique. Thus, each vertex in V has *at most* one incident edge in G. Together with $|E| = |V|$, this implies that each vertex in V has *exactly* one incident edge in G. ■

We are now ready to present our main contribution.

Theorem 3.20. *Every generalized resistant hyperplane mechanism is group strategyproof.*

Proof. Consider an $(\mathcal{S}, \boldsymbol{k})$-generalized resistant hyperplane mechanism. Consider a set of data points $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$. Suppose a coalition $S \subseteq N$ of agents manipulates, reports $(\widetilde{y}_i)_{i \in S}$, and changes the resulting hyperplane from $\boldsymbol{\beta}$ to $\widetilde{\boldsymbol{\beta}}$. Set $\widetilde{y}_i = y_i$ for $i \in N \setminus S$, and let $\widetilde{\mathcal{D}} = (\boldsymbol{x}_i, \widetilde{y}_i)_{i \in N}$.

By the hyperplane comparison lemma (Lemma 3.18), there exists $S_t \in \mathcal{S}$ such that either $\boldsymbol{\beta}^T \overline{\boldsymbol{x}_i} < \widetilde{\boldsymbol{\beta}}^T \overline{\boldsymbol{x}_i}$ for all $i \in S_t$, or $\boldsymbol{\beta}^T \overline{\boldsymbol{x}_i} > \widetilde{\boldsymbol{\beta}}^T \overline{\boldsymbol{x}_i}$ for all $i \in S_t$.

Without loss of generality, suppose it is the former. The k_t^{th} smallest residual from S_t is zero under $\boldsymbol{\beta}$ in $\mathcal{D}$, and under $\widetilde{\boldsymbol{\beta}}$ in $\widetilde{\mathcal{D}}$. If $S \cap S_t = \emptyset$, or if every manipulator in $S \cap S_t$ has a positive residual under $\boldsymbol{\beta}$ in $\mathcal{D}$, then at least k_t non-manipulators in $N \setminus S$ have a non-positive residual under $\boldsymbol{\beta}$ in $\mathcal{D}$, and thus a strictly negative residual under $\widetilde{\boldsymbol{\beta}}$ in $\widetilde{\mathcal{D}}$, which contradicts the fact that $\widetilde{\boldsymbol{\beta}}$ makes the k_t^{th} smallest residual in S_t zero in $\widetilde{\mathcal{D}}$.

In other words, there must exist a manipulator $i \in S \cap S_t$ who has a non-positive residual under $\boldsymbol{\beta}$ in $\mathcal{D}$. Thus, $\widetilde{\boldsymbol{\beta}}^T \overline{\boldsymbol{x_i}} > \boldsymbol{\beta}^T \overline{\boldsymbol{x_i}} \geq y_i$, implying that the manipulator is strictly worse off after the manipulation. Hence, the mechanism is group strategyproof. ■

For two dimensions ($d = 1$), we already argued that our sub-family of group strategyproof CRM mechanisms given by Theorem 3.7 is part of the larger family of GRL mechanisms (Lemma 3.10). It is easy to see that GRL mechanisms are precisely GRH mechanisms in two dimensions. Indeed, GRH mechanisms would require two subsets of agents S_1, S_2 that are publicly separable, i.e., well separable on the x-axis. Note that this coincides with the separability definition used by GRL mechanisms (Definition 3.6). Hence, the (S, S', k, k')-GRL mechanism is precisely the $(\mathcal{S}, \boldsymbol{k})$-GRH mechanism with $\mathcal{S} = (S, S')$ and $\boldsymbol{k} = (k, k')$. In three or more dimensions, we do not know if, given $\boldsymbol{x}$, one can always construct a family $\mathcal{S}$ of publicly separable sets of agents such that each set $S_t \in \mathcal{S}$ contains at least a constant fraction of the agents.

3.3 Strategyproofness vs Group Strategyproofness

In the single dimensional setting ($d = 0$), Moulin [1980] proved that all strategyproof mechanisms are also group strategyproof. This alternatively follows from a result by Barberà et al. [2010], who gave a sufficient condition on the underlying domain for the sets of strategyproof and group strategyproof mechanisms to coincide.

Interestingly, all known strategyproof mechanisms for the multidimensional linear regression setting (including generalized L_1-ERM and generalized resistant hyperplane mechanisms) are group strategyproof as well. However, it is easy to check that the linear regression setting does not satisfy the sufficient condition of Barberà et al. [2010]. Is it still true that all strategyproof mechanisms for linear regression are also group strategyproof? We answer this question *negatively*.

Example 3.21. Consider the simple linear regression setting ($d = 1$) with $n = 2$ agents. Fix the public information $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$, and consider the mechanism M that, on input $\boldsymbol{y} = (y_1, y_2)$, returns the line passing through points (x_1, y_2) and (x_2, y_1). Under this mechanism, the outcome for each agent is independent of the agent's report: indeed, the outcome for agent 1 (resp. agent 2) is $\widehat{y}_1 = y_2$ (resp. $\widehat{y}_2 = y_1$). Hence, the mechanism is clearly strategyproof. However, group strategyproofness is violated because when $y_1 \neq y_2$, the two agents can collude, and report $\widetilde{\boldsymbol{y}} = (y_2, y_1)$. This makes the resulting line pass through both agents, making both strictly better off.

The requirement that the outcome for each agent be independent of the agent's report, called *impartiality* in mechanism design, is stricter than (i.e., logically implies) strategyproofness, and has been studied for aggregating opinions or dividing rewards [de Clippel et al., 2008, Fischer and Klimm, 2015, Holzman and Moulin, 2013, Kurokawa et al., 2015, Tamura and Ohseto, 2014].

Definition 3.22 (Impartial Mechanisms). A mechanism M is called *impartial* if the outcome for each agent is independent of the agent's report. Formally, for every agent $i \in N$, reports $\boldsymbol{y}$, and alternative report y_i' by agent i, we require that $\widehat{y}_i(M(\boldsymbol{y})) = \widehat{y}_i(M(y_i', \boldsymbol{y}_{-i}))$.

In linear regression, when the number of agents is $n = d + 1$, we can easily characterize all impartial mechanisms because we can set $\widehat{y}_i$ to be an arbitrary function of $\boldsymbol{y}_{-i}$, and return a hyperplane passing through the resulting $d + 1$ points $(\boldsymbol{x}_i, \widehat{y}_i)_{i \in N}$.

Proposition 3.23. *For $n = d + 1$, mechanism M is impartial if and only if there exist functions $f_1, \ldots, f_n : \mathbb{R}^{n-1} \to \mathbb{R}$ such that given $\boldsymbol{y}$, M returns a hyperplane passing through $(\boldsymbol{x}_i, f_i(\boldsymbol{y}_{-i}))_{i \in N}$.*

Note that functions f_i can even be discontinuous, which can make the regression hyperplane discontinuous in the input $\boldsymbol{y}$. However, we later show (Theorem 4.3) that under any strategyproof

mechanism, the outcome $\widehat{y}_i$ for agent i must be a continuous function of y_i (it is a constant function of y_i in case of impartial mechanisms).

With $n > d + 1$ points, the question of whether impartial mechanisms even exist is non-trivial. While we still need to set each $\widehat{y}_i$ as a function of $\boldsymbol{y}_{-i}$, it cannot be done arbitrarily as the resulting points $(vx(i), \widehat{y}_i)_{i \in N}$ may no longer lie on a hyperplane. In other words, setting $\widehat{y}_i$ as a function of $\boldsymbol{y}_{-i}$ for $d + 1$ agents already determines the hyperplane, and thus $\widehat{y}_j$ for all remaining agents j. The mechanism must ensure that these $\widehat{y}_j$ are also independent of y_j. At first glance, this may seem impossible, except in the trivial case where a constant hyperplane is returned regardless of $\boldsymbol{y}$.

Nonetheless, we show that there exists a wide family of non-trivial impartial mechanisms for linear regression. Our family provides a full characterization of impartial mechanisms for $d = 1$ (i.e., for simple linear regression). In the result below, we use the notation $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ instead of $\boldsymbol{a}^T \boldsymbol{b}$ for the sake of simplicity. Its proof is in Appendix A.

Theorem 3.24. *Given $\boldsymbol{x}$, mechanism $M^{\boldsymbol{x}}$ for linear regression is impartial if there exist functions $\{g_i^{\boldsymbol{x}} : \mathbb{R} \to \mathbb{R}^d\}_{i \in N}$ and constant $c^{\boldsymbol{x}} \in \mathbb{R}$ such that for all $\boldsymbol{y}$, we have $M^{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_0)$, where*

$$\boldsymbol{\beta}_1 = \textstyle\sum_{i \in N} g_i^{\boldsymbol{x}}(y_i), \quad \beta_0 = c^{\boldsymbol{x}} - \sum_{i \in N} \left\langle g_i^{\boldsymbol{x}}(y_i), \boldsymbol{x}_i \right\rangle. \tag{6}$$

For $d = 1$ and an admissible set of points, this characterizes all impartial mechanisms.

Impartial mechanisms are not compelling from a statistical viewpoint. For instance, in the standard two-dimensional stochastic model where the data points are assumed to be generated by taking points on an underlying line and i.i.d. errors in the dependent variables, it is easy to show that no impartial mechanism can produce an unbiased estimator of the underlying regression line. Nonetheless, impartial mechanisms help us establish the existence of a rather wide family of strategyproof mechanisms that are *not* group strategyproof. In fact, the next result shows that almost all impartial mechanisms violate group strategyproofness; its proof is in Appendix A.

Proposition 3.25. *For simple linear regression ($d = 1$) with an admissible set of points, an impartial mechanism is group strategyproof if and only if it is a constant function (i.e., it returns a fixed regression line regardless of its input).*

4 CHARACTERIZING STRATEGYPROOF MECHANISMS

As mentioned in Section 3.1, Moulin [1980] studied the one-dimensional setting ($d = 0$), and analytically characterized all strategyproof mechanisms for n agents. While we are unable to provide an analytical characterization for multidimensional linear regression, we provide two non-constructive characterizations, and discuss their implications.

Interestingly, to characterize strategyproof mechanisms for linear regression with n agents, we use the characterization of strategyproof mechanisms for the one-dimensional setting with a single agent. In this case, Moulin [1980] shows that a mechanism is strategyproof if and only if there exist constants $\alpha^1, \alpha^2 \in \overline{\mathbb{R}}$ such that when the agent reports y, the mechanism returns $\widehat{y} = \text{med}(y, \alpha^1, \alpha^2)$. Constants α^1 and α^2 are called *phantoms*. First, we extend this result by providing an alternative characterization, which uses the following definition. The proof of the next result is in Appendix A.

Definition 4.1 (Locally Constant Function). For $A, B \subseteq \mathbb{R}$, function $f : A \to B$ is called locally constant at $x \in A$ if there exists $\epsilon > 0$ such that for all $x' \in [x - \epsilon, x + \epsilon]$, $f(x') = f(x)$.

Lemma 4.2. *Suppose mechanism $\pi : \mathbb{R} \to \mathbb{R}$ for the one-dimensional setting with a single agent elicits private value y from the agent and return $\pi(y)$. Then, π being strategyproof is equivalent to each of the following conditions.*

(a) There exist constants $\alpha^1, \alpha^2 \in \overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$ such that for all $y \in \mathbb{R}$, $\pi(y) = \text{med}(y, \alpha^1, \alpha^2)$.

(b) π is continuous, and for every $y \in \mathbb{R}$, either $\pi(y) = y$ or π is locally constant at y.

In the one-dimensional setting, Moulin [1980] observed that a mechanism is strategyproof if and only if its outcome is strategyproof in the report of each individual agent when other agents' reports are fixed. That is, a mechanism $\pi : \mathbb{R}^n \to \mathbb{R}$ for n agents is strategyproof if and only if

$$\forall i \in [n], \quad \exists \alpha_i^1, \alpha_i^2 \in \overline{\mathbb{R}} \text{ independent of } y_i \text{ s.t. } \pi(y_1, \ldots, y_n) = \text{med}(y_i, \alpha_i^1, \alpha_i^2). \tag{7}$$

Moulin [1980] solved Equation (7) to derive an elegant analytical expression for π in terms of $\{y_i\}_{i \in [n]}$. Note that in this equation, the outcome $\widehat{y} = \pi(y_1, \ldots, y_n)$ is common to all agents.

In contrast, in linear regression each agent i has a potentially different outcome $\widehat{y}_i$. Like before, strategyproofness requires that each $\widehat{y}_i$ obey the conditions in Lemma 4.2, when seen as a function of y_i, when other agents' reports are fixed. However, the outcomes for different agents are now constrained so that $(\boldsymbol{x}_i, \widehat{y}_i)_{i \in N}$ lie on a hyperplane. This added complexity prevented us from solving the equations to derive an analytical characterization, despite significant effort. The only exception was the special case of *impartial* mechanisms, where we further restrict $\widehat{y}_i$ to be independent of y_i (Theorem 3.24). This corresponds to the case where $\alpha_i^1 = \alpha_i^2$ for each agent i. Nonetheless, by simply applying Lemma 4.2 for every agent i, we obtain the following non-constructive characterization of strategyproof mechanisms for linear regression.

Theorem 4.3. *Given public information $\boldsymbol{x}$, mechanism $M^{\boldsymbol{x}}$ for linear regression being strategyproof is equivalent to each of the following conditions.*

(a) For every $\boldsymbol{y}_{-i} \in \mathbb{R}^{n-1}$ and $i \in N$, there exist $\ell_i, h_i \in \overline{\mathbb{R}}$ such that $\widehat{y}_i(M^{\boldsymbol{x}}(\boldsymbol{y})) = \text{med}(y_i, \ell_i, h_i)$ for all $y_i \in \mathbb{R}$;

(b) For every $\boldsymbol{y}_{-i} \in \mathbb{R}^{n-1}$ and $i \in N$, function $f_i(\cdot) = \widehat{y}_i(M(\cdot, \boldsymbol{y}_{-i}))$ is continuous, and for every $y_i \in \mathbb{R}$, either $f_i(y_i) = y_i$ or f_i is locally constant at y_i.

The first condition provides an analytical form of $\widehat{y}_i$ in terms of y_i, and is perhaps the more useful characterization. For instance, we crucially use this characterization in the next section to give a lower bound on the efficiency of strategyproof mechanisms. Our earlier (more complex) proof of group strategyproofness of GRH mechanisms (Theorem 3.20) was also based on this condition, and identified the precise ℓ_i and h_i for each agent i.

Note that for fixed $\boldsymbol{y}_{-i}$, we have $\widehat{y}_i = y_i$ when $y_i \in [\ell_i, h_i]$. For $y_i \le \ell_i$, $\widehat{y}_i = \ell_i$ is fixed, and for $y_i \ge h_i$, $\widehat{y}_i = h_i$ is fixed. We therefore say that agent i is *influential* over the interval (ℓ_i, h_i), and call ℓ_i and h_i the *lower* and *upper influence bounds*, respectively. Analysis of influence bounds has received attention in the statistics literature, where it is called *sensitivity analysis*. For instance, Narula and Wellington [1985] observed that under L_1-ERM, the regression hyperplane is unaffected when the dependent variable of a point is changed so that the point still lies on the same side of the hyperplane as before. From Theorem 4.3, we can see that for every strategyproof mechanism, doing so should at least keep the outcome for agent i unchanged. Narula and Wellington [1985] also focused on computing the influence bounds. Theorem 4.3 lends a simple algorithm to compute influence bounds (see Appendix B). Finally, note that while $\widehat{y}_i$ must be continuous in y_i, it need not be continuous in $\boldsymbol{y}$ (see our discussion on Proposition 3.23).

5 EFFICIENCY OF STRATEGYPROOF MECHANISMS

Insofar, we studied families of strategyproof mechanisms for linear regression. In the absence of strategic considerations, a popular mechanism for linear regression is the OLS (ordinary least squares), which is the empirical risk minimizer for the squared loss. Under this loss function, which is also called the *residual sum of squares* (RSS), the loss when choosing hyperplane $\boldsymbol{\beta}$ given data points $\mathcal{D}$ is $\text{RSS}(\mathcal{D}, \boldsymbol{\beta}) = \sum_{i \in N} \left(y_i - \boldsymbol{\beta}^T \overline{\boldsymbol{x}_i}\right)^2$. A classic justification for the OLS is due to

the Gauss-Markov theorem, which states that when the errors (deviations of data points from an underlying hyperplane we wish to identify) are stochastic, zero in expectation, uncorrelated, and of equal variance, the OLS is the *best linear unbiased estimator*.

However, in our strategic setting, the OLS is not strategyproof [Dekel et al., 2010]. This raises an important question: *Is there a strategyproof mechanism that is close to the OLS?* We assess this by the worst-case approximation ratio of a mechanism for the optimal squared loss.

Definition 5.1 (Efficiency). Given $\boldsymbol{x}$, we say that mechanism $M^{\boldsymbol{x}}$ for linear regression is c-efficient if for every $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$, we have $\text{RSS}(\mathcal{D}, M^{\boldsymbol{x}}(\boldsymbol{y})) \leq c \cdot \inf_{\boldsymbol{\beta}} \text{RSS}(\mathcal{D}, \boldsymbol{\beta})$.

We show that no strategyproof mechanism that is too close to the OLS can be strategyproof. The proof of the next result leverages our characterization of strategyproof mechanisms (Theorem 4.3).

Theorem 5.2. *For $n \geq 4$, there exist $\boldsymbol{x}$ for which no strategyproof mechanism is $(2-\epsilon)$-efficient for any $\epsilon > 0$.*

Proof. For simplicity of notation, we use $n+1$ agents instead of n agents (and assume $n+1 \geq 4$, i.e., $n \geq 3$). We also consider simple linear regression ($d = 1$); the proof easily extends to higher dimensions by simply setting all other coordinates to zero. Fix $n \geq 3$. Consider a setting with $n+1$ agents where $x_i = i$ for $i \in [n]$, and $x_{n+1} = X$, where X is the solution of the following equation:

$$\frac{n^3 - n}{2(1 + 3n + 2n^2 + 6X^2 - 6Xn - 6X)} = 1. \tag{8}$$

Interested readers may note that $X = \Theta(n^{1.5})$. Let T denote the LHS in Equation (8).

Consider a strategyproof mechanism $M^{\boldsymbol{x}}$. Suppose $M^{\boldsymbol{x}}$ is c-efficient. We want to show that $c \geq 2$. We consider a family of inputs $\boldsymbol{y}$, in which we fix $y_i = 0$ for $i \in [n]$, and vary $y_{n+1} = Y$. First, we note that the optimal RSS, as a function of Y, is given by

$$f_0(Y) = Y^2 \cdot \frac{n^3 - n}{2 + 5n + 4n^2 + n^3 - 12X - 12nX + 12X^2} = Y^2 \cdot \frac{T}{T+1} = \frac{Y^2}{2},$$

where the first transition is obtained by minimizing $(Y - X \cdot \beta_1 - \beta_0)^2 + \sum_{i=1}^{n} (i \cdot \beta_1 + \beta_0)^2$ over all (β_1, β_0), the second transition follows through simple algebra, and the final transition follows from Equation (8). For verification of these claims through Mathematica, see Figure 2 in Appendix A.

Recall that we fixed y_i for $i \in [n]$. Due to our characterization result (Theorem 4.3), there exist $\ell, h \in \overline{\mathbb{R}}$ with $\ell \leq h$ such that the line returned by the mechanism passes through $(X, \text{med}(Y, \ell, h))$ for all Y. We take two cases.

Case 1: $h > 0$. Set $Y = h$. Then, the line returned by the mechanism passes through (X, h). In this case, we can show that the RSS of the mechanism is at least

$$f_1 = h^2 \cdot \frac{n^3 - n}{2(1 + 3n + 2n^2 + 6X^2 - 6Xn - 6X)} = h^2 \cdot T = h^2,$$

where the first transition is obtained by minimizing $(Y - \beta_1 \cdot X - \beta_0)^2 + \sum_{i=1}^{n} (\beta_1 \cdot i + \beta_0)^2$ over all (β_1, β_0) which satisfy $\beta_1 \cdot X + \beta_0 = Y$, and the rest follows from Equation (8). For verification of these claims through Mathematica, see Figure 2 in Appendix A. This implies $c \geq f_1 / f_0(h) = 2$.

Case 2: $h \leq 0$. Set $Y = 1$. Then, the line returned by the mechanism passes through (X, h). In this case, the RSS of the mechanism is at least $f_2 = 1$ because agent $n+1$ contributes $(1-h)^2 \geq 1$ to the squared loss. Once again, we have $c \geq f_2 / f_0(1) = 2$.

The proof is complete as we have $c \geq 2$ in each case. ∎

For $n = 2$ agents (or $n = d + 1$ agents in $d + 1$ dimensions), there is an obvious 1-efficient strategyproof mechanism which returns a hyperplane passing through all input points. Theorem 4.3 leaves open the case of $n = 3$ in two dimensions.

6 DISCUSSION

Our work leaves several open questions. Perhaps the most ambitious one is to find a constructive characterization of all strategyproof or group strategyproof mechanisms for linear regression, which may allow us to pinpoint the most efficient strategyproof mechanism; Caragiannis et al. [2016] provide a similar analysis in the one-dimensional setting. It is easy to show that L_1-ERM is n-efficient (see Proposition A.2 in Appendix A). Does there exist a more efficient strategyproof mechanism? It would also be interesting to analyze efficiency in a stochastic setting where the data points are drawn from an underlying distribution.

The characterization result of Moulin [1980] for strategyproof and anonymous mechanisms in the one-dimensional setting extends the median to generalized medians by adding fixed phantom values, and then taking the median. It is also shown that adding $n + 1$ phantoms is sufficient to obtain full generality. We can extend all our proposed families of mechanisms by adding a certain number of "phantom points" in $\mathbb{R}^{d+1}$, and then applying the mechanisms to the union of data points and phantom points. The resulting mechanism retains the incentive guarantees.[7] Given n data points, how many phantoms are sufficient to obtain full generality? Do the phantoms play a role in obtaining the elusive constructive characterization?

Another interesting observation is that our generalized resistant hyperplane mechanisms are guaranteed pass through $d + 1$ input points in $d + 1$ dimensions. It is known that at least one minimizer of the $L1$ loss also has this property. It would be interesting to identify a generic family of conditions, which, when imposed in addition to the requirement of making $d + 1$ residuals zero, yield group strategyproofness.

Finally, Dekel et al. [2010] study a regression setting in which a single agent may control multiple data points, show that L_1-ERM is no longer strategyproof, and provide novel strategyproof mechanisms. It would be useful to see if our ideas can be used to design additional strategyproof mechanisms in this model. Another interesting variant is when only a small number of data points are held by strategic agents, but the mechanism does not know which ones. A similar setting was studied by Charikar et al. [2017], but for classification and with adversarial manipulations. On a high level, we view our work as a stepping stone to studying incentives in more realistic machine learning environments.

REFERENCES

I. Bárány, A. Hubard, and J. Jerónimo. 2008. Slicing convex sets and measures by a hyperplane. *Discrete & Computational Geometry* 39, 1-3 (2008), 67–75.

S. Barberà, D. Berga, and B. Moreno. 2010. Individual versus group strategy-proofness: When do they coincide? *Journal of Economic Theory* 145, 5 (2010), 1648–1674.

D. Black. 1958. *Theory of Committees and Elections.* Cambridge University Press.

F. Breuer. 2010. Uneven splitting of ham sandwiches. *Discrete & Computational Geometry* 43, 4 (2010), 876–892.

G. W. Brown and A. M. Mood. 1951. On Median Tests for Linear Hypotheses. In *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*. 159–166.

N. H. Bshouty, N. Eiron, and E. Kushilevitz. 2002. PAC Learning with Nasty Noise. *Theoretical Computer Science* 288, 2 (2002), 255–275.

Y. Cai, C. Daskalakis, and C. H. Papadimitriou. 2015. Optimum Statistical Estimation with Strategic Data Sources. In *Proceedings of the 28th Conference on Computational Learning Theory (COLT)*. 280–296.

I. Caragiannis, A. D. Procaccia, and N. Shah. 2016. Truthful Univariate Estimators. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. 127–135.

F. Caro and J. Gallien. 2010. Inventory Management of a Fast-Fashion Retail Network. *Operations Research* 58, 2 (2010), 257–273.

[7] We also considered adding phantom values directly in the equations where a median is used. However, most such attempts violated strategyproofness.

F. Caro, J. Gallien, M. D. Miranda, J. C. Torralbo, J. M. C. Corras, M. M. Vazquez, J. A. R. Calamonte, and J. Correa. 2010. Zara Uses Operations Research to Reengineer its Global Distribution Process. *Interfaces* 40, 1 (2010), 71–84.

M. Charikar, J. Steinhardt, and G. Valiant. 2017. Learning from untrusted data. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC)*. 47–60.

R. Cummings, S. Ioannidis, and K. Ligett. 2015. Truthful Linear Regression. In *Proceedings of the 28th Conference on Computational Learning Theory (COLT)*. 448fi??–483.

G. de Clippel, H. Moulin, and N. Tideman. 2008. Impartial division of a dollar. *Journal of Economic Theory* 139 (2008), 176–191.

O. Dekel, F. Fischer, and A. D. Procaccia. 2010. Incentive Compatible Regression Learning. *J. Comput. System Sci.* 76, 8 (2010), 759–777.

J. Dong, A. Roth, Z. Schutzman, B. Waggoner, and Z. S. Wu. 2017. Strategic Classification from Revealed Preferences. arXiv:1710.07887. (2017).

M. Dummett and R. Farquharson. 1961. Stability in voting. *Econometrica* 29, 1 (1961), 33–43.

J. H. Elton and T. P. Hill. 2011. A stronger conclusion to the classical ham sandwich theorem. *European Journal of Combinatorics* 32, 5 (2011), 657–661.

F. Fischer and M. Klimm. 2015. Optimal impartial selection. *SIAM J. Comput.* 44, 5 (2015), 1263–1285.

S. A. Goldman and R. H. Sloan. 1995. Can PAC Learning Algorithms Tolerate Random Attribute Noise? *Algorithmica* 14, 1 (1995), 70–84.

M. Hardt, N. Megiddo, C. H. Papadimitriou, and M. Wootters. 2016. Strategic Classification. In *Proceedings of the 7th Innovations in Theoretical Computer Science Conference (ITCS)*. 111–122.

R. Holzman and H. Moulin. 2013. Impartial nominations for a prize. *Econometrica* 81, 1 (2013), 173–196.

S. Ioannidis and P. Loiseau. 2013. Linear regression as a non-cooperative game. In *Proceedings of the 9th Conference on Web and Internet Economics (WINE)*. 277–290.

I. M. Johnstone and P. F. Velleman. 1985. The resistant line and related regression methods. *J. Amer. Statist. Assoc.* 80, 392 (1985), 1041–1054.

M. Kearns and M. Li. 1993. Learning in the Presence of Malicious Errors. *SIAM J. Comput.* 22, 4 (1993), 807–837.

H. Kermer and A. B. Németh. 1973. Supporting spheres for families of independent convex sets. *Archiv der Mathematik* 24, 1 (1973), 91–96.

R. Koenker and Gilbert Bassett, Jr. 1978. Regression quantiles. *Econometrica* 46, 1 (1978), 33–50.

D. Kurokawa, O. Lev, J. Morgenstern, and A. D. Procaccia. 2015. Impartial Peer Review.. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*. 582–588.

N. Littlestone. 1991. Redundant noisy attributes, attribute errors, and linear-threshold learning using winnow. In *Proceedings of the 4th Conference on Computational Learning Theory (COLT)*. 147–156.

R. Meir, S. Almagor, A. Michaely, and J. S. Rosenschein. 2011. Tight bounds for strategyproof classification. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 319–326.

R. Meir, A. D. Procaccia, and J. S. Rosenschein. 2010. On the limits of dictatorial classification. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 609–616.

R. Meir, A. D. Procaccia, and J. S. Rosenschein. 2012. Algorithms for Strategyproof Classification. *Artificial Intelligence* 186 (2012), 123–156.

H. Moulin. 1980. On strategy-proofness and single-peakedness. *Public Choice* 35 (1980), 437–455.

S. C. Narula and J. F. Wellington. 1985. Interior analysis for the minimum sum of absolute errors regression. *Technometrics* 27, 2 (1985), 181–188.

J. Perote and J. Perote-Peña. 2003. The impossibility of strategy-proof clustering. *Economics Bulletin* 4, 23 (2003), 1–9.

J. Perote and J. Perote-Peña. 2004. Strategy-proof estimators for simple regression. *Mathematical Social Sciences* 47 (2004), 153–176.

W. Steiger and J. Zhao. 2010. Generalized ham-sandwich cuts. *Discrete & Computational Geometry* 44, 3 (2010), 535–545.

A. H. Stone and J. W. Tukey. 1942. Generalized "sandwich" theorems. *Duke Mathematical Journal* 9, 2 (1942), 356–359.

S. Tamura and S. Ohseto. 2014. Impartial nomination correspondences. *Social Choice and Welfare* 43 (2014), 47–54.

J. W. Tukey. 1977. *Exploratory Data Analysis*. Addison-Wesley.

L. Wang. 2013. The L1 penalized LAD estimator for high dimensional linear regression. *Journal of Multivariate Analysis* 120 (2013), 135–151.

L. Wang, M. D. Gordon, and J. Zhu. 2006. Regularized least absolute deviations regression and an efficient algorithm for parameter tuning. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM)*. 690–700.

APPENDIX

A MISSING RESULTS AND PROOFS

In this section, we present the results and proofs missing from the main body of the paper.

A.1 Generalized L_1-ERM is Group Strategyproof

Proof of Theorem 3.1. In order to show that the weighted L_1-ERM with convex regularizer is group strategyproof we first need to show that Proposition 4.2 from Dekel et al. [2010] still holds. Let $\widehat{S} = \{(\boldsymbol{x}_i, \widehat{y}_i)\}_{i=1}^m$ and $\widetilde{S} = \{(\boldsymbol{x}_i, \widetilde{y}_i)\}_{i=1}^m$ be two training sets on the same set of points and let $\widehat{f} = \texttt{w-ERM-reg}(\mathcal{F}, \ell, \widehat{S})$ and $\widetilde{f} = \texttt{w-ERM-reg}(\mathcal{F}, \ell, \widetilde{S})$, where by `w-ERM-reg` we denote the weighted L_1-ERM with convex regularizer and by ℓ the L_1 loss function. If $\widehat{f} \neq \widetilde{f}$ then, there exists $i \in N$, such that $\widehat{y}_i \neq \widetilde{y}_i$ and

$$\ell(\widehat{f}(\boldsymbol{x}_i), \widehat{y}_i) < \ell(\widetilde{f}(\boldsymbol{x}_i), \widehat{y}_i) \tag{9}$$

Let $U = \{i : \widehat{y}_i \neq \widetilde{y}_i\}$ and assume that $\ell(\widehat{f}(\boldsymbol{x}_i), \widehat{y}_i) \geq \ell(\widetilde{f}(\boldsymbol{x}_i), \widehat{y}_i)$ for all $i \in U$. First, we will consider functions of the form $f_\alpha(\boldsymbol{x}) = \alpha\widetilde{f}(\boldsymbol{x}) + (1-\alpha)\widehat{f}(\boldsymbol{x})$ and prove that there exists $\alpha \in (0, 1]$ such that:

$$\widehat{R}\left(\widehat{f}, \widetilde{S}\right) - \widehat{R}\left(\widehat{f}, \widehat{S}\right) = \widehat{R}\left(f_\alpha, \widetilde{S}\right) - \widehat{R}\left(f_\alpha, \widehat{S}\right) \tag{10}$$

For all $i \in U$ from Equation (9) we get that either of the four inequalities below holds:

$$\widetilde{f}(\boldsymbol{x}_i) \leq \widehat{y}_i < \widehat{f}(\boldsymbol{x}_i), \quad \widetilde{f}(\boldsymbol{x}_i) \geq \widehat{y}_i > \widehat{f}(\boldsymbol{x}_i) \tag{11}$$

$$\widehat{y}_i \leq \widetilde{f}(\boldsymbol{x}_i)\widehat{f}(\boldsymbol{x}_i), \quad \widehat{y}_i \geq \widetilde{f}(\boldsymbol{x}_i) \geq \widehat{f}(\boldsymbol{x}_i) \tag{12}$$

Observe now, that similarly to Dekel et al. [2010] since $\widetilde{y}_i \neq \widetilde{f}(\boldsymbol{x}_i)$ produces the least sum of the weighted loss and the convex regularizer, then assuming that $\widetilde{y}_i = \widetilde{f}(\boldsymbol{x}_i)$ will cause greater risk reduction for $\widetilde{f}$, and therefore, $\widetilde{f}$ will still minimize the risk. If one of the two inequalities in Equation (11) holds:

$$\alpha_i = \frac{\widehat{y}_i - \widehat{f}(\boldsymbol{x}_i)}{\widetilde{f}(\boldsymbol{x}_i) - \widehat{f}(\boldsymbol{x}_i)} \tag{13}$$

where $\alpha_i \in (0, 1]$ and $f_{\alpha_i}(\boldsymbol{x}_i) = \widehat{y}_i$. By substituting, for every $\alpha \in (0, \alpha_i]$ it holds that:

$$\widetilde{y}_i \leq \widehat{y}_i \leq f_{\alpha_i}(\boldsymbol{x}_i) < \widehat{f}(\boldsymbol{x}_i) \quad \text{or} \quad \widetilde{y}_i \geq \widehat{y}_i \geq f_{\alpha_i}(\boldsymbol{x}_i) > \widehat{f}(\boldsymbol{x}_i)$$

Based on the above and if we set $c_i = |\widehat{y}_i - \widetilde{y}_i|$, we have that for all $\alpha \in (0, \alpha_1]$:

$$\ell(\widehat{f}(\boldsymbol{x}_i), \widetilde{y}_i) - \ell(\widehat{f}(\boldsymbol{x}_i), \widehat{y}_i) = c_i \quad \text{and} \quad \ell(f_\alpha(\boldsymbol{x}_i), \widetilde{y}_i) - \ell(f_\alpha(\boldsymbol{x}_i), \widehat{y}_i) = c_i \tag{14}$$

By using Equation (12) and setting $\alpha_i = 1$ and $c_i = -|\widetilde{y}_i - \widehat{y}_i|$ one ends up again with Equation (14). Equation (14) holds for every $i \in U$ if we set $\alpha = \min_{i \in U} \alpha_i$ and it trivially holds for all $i \neq U$ with $c_i = 0$. Multiplying with the appropriate weights $w_i^{\boldsymbol{x}}$ the equalities for each i and summing them all, one gets to Equation (10). Note that in this step, the regularizer $h(f)$ can be ignored, since it cancels out from each side of the equation.

Since $\mathcal{F}$ is a convex set, $f_\alpha \in \mathcal{F}$. Since $\widehat{f}$ minimizes the empirical risk with respect to $\widehat{S}$ over $\mathcal{F}$ we have that $\widehat{R}(\widehat{f}, \widehat{S}) \leq \widehat{R}(f_\alpha, \widehat{S})$ and combining with Equation (9) we get that $\widehat{R}(\widehat{f}, \widetilde{S}) \leq \widehat{R}(f_\alpha, \widetilde{S})$. The emprical risk function is convex in its first argument (we are using a strictly convex regularizer) we have that:

$$\widehat{R}(\widehat{f}, \widetilde{S}) \leq \widehat{R}(f_\alpha, \widetilde{S}) \leq \alpha\widehat{R}(\widetilde{f}, \widetilde{S}) + (1-\alpha)\widehat{R}(\widehat{f}, \widetilde{S}) \tag{15}$$

However, since $\widetilde{f}$ minimizes the loss with respect to $\widetilde{S}$: $\widehat{R}(\widetilde{f},\widetilde{S}) \leq \widehat{R}(\widehat{f},\widetilde{S})$ and thus

$$\widehat{R}(\widehat{f},\widetilde{S}) = \widehat{R}(\widetilde{f},\widetilde{S}) = \min_{f\in\mathcal{F}} \widehat{R}(f,\widetilde{S}) \tag{16}$$

In other words we have shown that both $\widehat{f}$ and $\widetilde{f}$ minimize the empirical risk with respect to $\widetilde{S}$. The only thing that is left to be shown for the contradiction argument is that the tie breaking step of the algorithm does not distinguish between two functions that are risk minimizers. In order words, we need to show that both functions attain the minimum norm over all empirical risk minimizers.

Combining Equation (16) with (15) we get that $\widehat{R}(f_\alpha,\widetilde{S}) \leq \widehat{R}(\widehat{f},\widetilde{S})$. From (10) we have that $\widehat{R}(f_\alpha,\widetilde{S}) \leq \widehat{R}(\widehat{f},\widehat{S})$ and thus $\widehat{R}(f_\alpha,\widehat{S}) = \widehat{R}(\widehat{f},\widehat{S})$. However, $\widehat{f}$ was chosen to miminimize the empirical risk with respect to $\widehat{S}$ and therefore, $||\widehat{f}|| \leq ||f_\alpha||$. Using convexity of the norm, we get $||\widehat{f}|| \leq ||\widetilde{f}||$. Also, for the case of sample $\widetilde{S}$, algorithm chose function $\widetilde{f}$ and therefore $||\widehat{f}|| \geq ||\widetilde{f}||$. This concludes our contradiction argument, since

$$||\widehat{f}|| = ||\widetilde{f}|| = \min_{f\in\mathcal{F}:\widehat{R}(f,\widetilde{S})=\widehat{R}(\widetilde{f},\widetilde{S})} ||f|| \tag{17}$$

■

A.2 CRM Mechanisms are Also GRL Mechanisms

In the CRM mechanism, we refer to the point in S which has the median of all median CWAs (i.e., DA) as the "directing point", and the point in S' to which this DA is pointing as the "directed point".

Proof of Lemma 3.10. First, we show that for any $S \subseteq N$, the (S,S)-CRM mechanism is (L,R,k,k')-GRL mechanism for some L,R,k,k'. Without loss of generality, we can assume $S = N$ as the other points are simply ignored. Thus, we will refer to the (N,N)-CRM mechanism.

First, consider the case where n is even. Let L (resp. R) be the set of $n/2$ points with the smallest (resp. largest) x coordinates. We show equivalence of the (N,N)-CRM mechanism to the (L,R,k,k')-GRL mechanism for appropriate k and k'. Let (β_1,β_0) be the line returned by the CRM mechanism.

Choose $x^* \in (\max_{i\in L} x_i, \min_{i\in R} x_i)$, and define the following sets.

- $A = \{i : x_i < x^*, y_i \geq \beta_1 x_i + \beta_0\}$
- $B = \{i : x_i > x^*, y_i > \beta_1 x_i + \beta_0\}$
- $C = \{i : x_i < x^*, y_i < \beta_1 x_i + \beta_0\}$
- $D = \{i : x_i > x^*, y_i \leq \beta_1 x_i + \beta_0\}$

Note that $A \cup C = L$ and $B \cup D = R$. For $i \in N$, let $MCWA_i$ denote the median CWA from i to points in $N \setminus \{i\}$. Note that for each $i \in L$, there are strictly more points in $N \setminus \{i\}$ to the right of it, than to the left of it, implying that $MCWA_i \in [\pi, 2\pi]$. Similarly, for each $i \in R$, we have $MCWA_i \in [0,\pi]$.

Let DA be the directing angle under the CRM mechanism. Then, $DA = \min_{i\in L} MCWA_i$ or $DA = \max_{i\in R} MCWA_i$ based on whether the outer median in the directing angle definition uses the right median or the left median. Let us assume it uses the left median, so $DA = \max_{i\in R} MCWA_i$. The proof for the other case is symmetric.

We now show that in this case, $B = C = \emptyset$. This would imply that the mechanism is equivalent to $(L,R,|L|,1)$-GRL because every point in L has a non-positive residual while every point in R has a non-negative residual.

Suppose for contradiction that $B \neq \emptyset$. Take a point $i_B \in B$. Note that $MCWA_{i_B} \leq \max_{i\in R} MCWA_i = DA$. Note that the directing point i^* is on the regression line, and hence $i^* \in D$. Then, one can check that if $x_{i_B} < x_{i^*}$, then x_{i_B} has strictly less number of points to which its angle is less than

$MCWA_{i_B}$ than x_{i^*} has to which its angle is less than $MCWA_{i^*} = DA$. In the case $x_{i_B} > x_{i^*}$, the same happens but for points with angle greater than MCWA. This is a contradiction because each point has exactly $(n-2)/2$ points with angle more or less than its MCWA. Hence, $B = \emptyset$. Using a symmetric argument, we can establish $C = \emptyset$, which completes the proof.

We now consider the case where n is odd. In this case, let L (resp. R) be the set of $(n-1)/2$ points with the smallest (resp. largest) x-coordinate, and let i^* be the point with the median x-coordinate. Once again, we have that $MCWA_i \in [\pi, 2\pi]$ for each $i \in L$, and $MCWA_i \in [0, \pi]$ for each $i \in R$. We add i^* to L if $MCWA_{i^*} \in [\pi, 2\pi]$, and to R otherwise. Suppose we add it to R, and let $R' = R \cup \{i^*\}$. Then using an argument similar to above, we can check that the CRM mechanism is equivalent to (L, R', k, k') for appropriate k, k'.

The case where $S \cap S' = \emptyset$ and $\min(|S|, |S'|) = 1$ is much simpler. Again, without loss of generality, we can consider $S \cup S' = N$, and for simplicity, consider the case where n is even and $|S| = 1$. The other cases are similar. Let $S = \{i^*\}$. Without loss of generality, suppose there are more points to the right of i^* than to the left of it. Let R be the set of points to the right of i^*, and L be the set of points to the left of i^*. Then, it is easy to see that when we take the median CWA from i^* (say, the left median, i.e., the $(n/2-1)^{\text{th}}$ smallest CWA), it will always be towards a point in R. Moreover, it will be the $(n/2-1-|S|)^{\text{th}}$ smallest CWA towards points in R. However, CWAs towards points in R are monotonic in slopes to points in R. Hence, the regression line will make the $(n/2-1-|S|)^{\text{th}}$ smallest residual in R zero. In other words, the mechanism is equivalent to $(\{i^*\}, R, 1, n/2-1-|S|)$-GRL. ■

A.3 Impartial Mechanisms

We now present the proof of Theorem 3.24. First, we need the following definition.

Definition A.1 (Completely Additively Separable). Function $f : \mathbb{R}^k \to \mathbb{R}$ is called *completely additively separable* if there exist functions $\{g_i\}_{i=1}^k$ such that $f(t_1, \ldots, t_k) = \sum_{i=1}^k g_i(t_i)$ for all $\boldsymbol{t} = (t_1, \ldots, t_k) \in \mathbb{R}^k$.

It is well known that f is completely additively separable if and only if for all $\boldsymbol{t} \in \mathbb{R}^k$, $i \in [k]$, and $t_i' \in \mathbb{R}$, $f(t_i, \boldsymbol{t}_{-i}) - f(t_i', \boldsymbol{t}_{-i})$ is independent of $\boldsymbol{t}_{-i}$.

Proof of Theorem 3.24. We omit $\boldsymbol{x}$ from all superscripts for simplicity. Suppose mechanism M is given by Equation (6). Then:

$$\begin{aligned}\widehat{y}_i(\boldsymbol{\beta}) = \langle \boldsymbol{\beta}, \overline{\boldsymbol{x}_i} \rangle &= \left\langle \textstyle\sum_{j \in N} g_j(y_j), \boldsymbol{x}_i \right\rangle + c - \sum_{j \in N} \left\langle g_j(y_j), \boldsymbol{x}_j \right\rangle \\ &= c + \textstyle\sum_{j \in N \setminus \{i\}} \left\langle g_j(y_j), \boldsymbol{x}_i - \boldsymbol{x}_j \right\rangle.\end{aligned}$$

Note that $\widehat{y}_i(\boldsymbol{\beta})$ is independent of y_i, which implies that M is impartial.

We now prove the converse for simple linear regression ($d = 1$) with an admissible set of points. Suppose mechanism M is impartial. Given $\boldsymbol{y}$, let $\beta_1(\boldsymbol{y})$ be the slope of the line returned by M, and $f_i(\boldsymbol{y}) = \widehat{y}_i(M(\boldsymbol{y}))$ be the outcome for agent i. Because M is impartial, f_i is independent of y_i. Hence, we denote the outcome for agent i by $f_i(\boldsymbol{y}_{-i})$.

We want to show that h is completely additively separable. Equivalently, for every $\boldsymbol{y}$ and $\widetilde{\boldsymbol{y}}$ such that $\boldsymbol{y}_{-i} = \widetilde{\boldsymbol{y}}_{-i}$, we want to show that $\beta_1(\boldsymbol{y}) - \beta_1(\widetilde{\boldsymbol{y}})$ is independent of $\boldsymbol{y}_{-i}$. Choose $j \in N \setminus \{i\}$ arbitrarily. By the definition of the slope of a line, we have

$$\beta_1(\boldsymbol{y}) = \frac{f_j(\boldsymbol{y}_{-j}) - f_i(\boldsymbol{y}_{-i})}{x_j - x_i},\ \beta_1(\widetilde{\boldsymbol{y}}) = \frac{f_j(\widetilde{\boldsymbol{y}}_{-j}) - f_i(\widetilde{\boldsymbol{y}}_{-i})}{x_j - x_i}.$$

Taking the difference, and noting that $\boldsymbol{y}_{-i} = \widetilde{\boldsymbol{y}}_{-i}$, we get

$$\beta_1(\boldsymbol{y}) - \beta_1(\widetilde{\boldsymbol{y}}) = \frac{f_j(\boldsymbol{y}_{-j}) - f_j(\widetilde{\boldsymbol{y}}_{-j})}{x_j - x_i}.$$

Note that the RHS is independent of y_j. Since we chose $j \in N \setminus \{i\}$ arbitrarily, it follows that $\beta_1(\boldsymbol{y}) - \beta_1(\widetilde{\boldsymbol{y}})$ is independent of $\boldsymbol{y}_{-i}$, implying that h is completely additively separable. Thus, there must exist functions $\{g_i\}_{i \in N}$ such that $\beta_1(\boldsymbol{y}) = \sum_{i \in N} g_i(\boldsymbol{y})$.

We now want to calculate β_0. Recall that for every $i \in N$, the outcome for agent i is

$$f_i(\boldsymbol{y}_{-i}) = \beta_1(\boldsymbol{y}) \cdot x_i + \beta_0 = g_i(y_i) \cdot x_i + \sum_{j \in N \setminus \{i\}} g_j(y_j) \cdot x_i + \beta_0.$$

Since the LHS is independent of y_i, so must be the RHS. Hence, $\beta_0 + g_i(y_i) \cdot x_i$ must be independent of y_i for each $i \in N$. This implies $\beta_0 = c - \sum_{i \in N} g_i(y_i) \cdot x_i$ for some constant c, as desired. ■

Proof of Proposition 3.25. By Theorem 3.24, an impartial mechanism for simple linear regression with an admissible set of points must be of the form given in Equation (6). We want to show that function $g_i^{\boldsymbol{x}}$ is constant for each $i \in N$. Suppose for contradiction that for some agent $i \in N$, function $g_i^{\boldsymbol{x}}$ is not constant. Thus, there exist y_i^1 and y_i^2 such that $g_i^{\boldsymbol{x}}(y_i^1) \neq g_i^{\boldsymbol{x}}(y_i^2)$. Fix an agent $j \in N \setminus \{i\}$ and $\boldsymbol{y}_{-\{i,j\}} \in \mathbb{R}^{n-2}$. Let $\widehat{y_j^1}$ and $\widehat{y_j^2}$ denote the outcomes for agent j under the impartial mechanism when agent i reports y_i^1 and y_i^2, respectively, and agents in $N \setminus \{i,j\}$ report $\boldsymbol{y}_{-\{i,j\}}$. That is,

$$\widehat{y_j^t} = g_i^{\boldsymbol{x}}(y_i^t) \cdot (x_j - x_i) + \textstyle\sum_{k \in N \setminus \{i,j\}} g_k^{\boldsymbol{x}}(y_k) \cdot (x_j - x_k) + c^{\boldsymbol{x}}, \forall t \in \{1,2\}.$$

Note that $g_i^{\boldsymbol{x}}(y_i^1) \neq g_i^{\boldsymbol{x}}(y_i^2)$ and $x_i \neq x_j$ imply that $\widehat{y_j^1} \neq \widehat{y_j^2}$. Now, suppose that the private values of the agents are $(y_i^1, \widehat{y_j^2}, \boldsymbol{y}_{-\{i,j\}})$. In this case, the outcome for agent j is $\widehat{y_j^1}$, which is different from her private value $\widehat{y_j^2}$. If agent i changes her report to y_i^2, her own outcome would not change, but the outcome for agent j would change to $\widehat{y_j^2}$, making agent j strictly better off. Thus, the coalition $\{i,j\}$ successfully manipulates, showing a violation of group strategyproofness.

For the reverse direction, note that all constant functions are trivially group strategyproof. ■

A.4 Characterization of Strategyproof Mechanisms

Proof of Lemma 4.2. Part (a) is precisely the characterization of strategyproof mechanisms due to Moulin [1980, Proposition 3], applied to the case of a single agent.[8]

We would like to show that part (b) is equivalent to part (a). It is easy to check that a function π of the form given in part (a) satisfies the conditions of part (b). We now show the converse.

Suppose that π is continuous, and for every $y \in \mathbb{R}$, either $\pi(y) = y$ or π is locally constant at y. Let $O = \{y \in \mathbb{R} : \pi \text{ is locally constant at } y\}$. We first show that O is an open set. That is, if $y \in O$, there must exist a $\delta > 0$ such that $(y - \delta, y + \delta) \subseteq O$. Indeed, fix a $y \in O$. Because π is locally constant at y, there must exist an $\epsilon > 0$ such that π is constant in $[y - \epsilon, y + \epsilon]$. Set $\delta = \epsilon/2$, and pick an arbitrary $y' \in (y - \delta, y + \delta)$. We want to show that $y' \in O$. Note that for $\epsilon' = \epsilon/2$, $[y' - \epsilon', y' + \epsilon'] \subseteq [y - \epsilon, y + \epsilon]$. Hence, π is constant in $[y' - \epsilon', y' + \epsilon']$, implying that $y' \in O$. This concludes the proof that O is an open set.

Next, we use the well-known fact that any open subset of $\mathbb{R}$ is a countable union of pairwise disjoint open intervals. That is, we can write $O = \cup_{k \in \mathbb{N}} (a_k, b_k)$, where $a_k, b_k \in \overline{\mathbb{R}}$. For $k \in \mathbb{N}$, because π is locally constant over (a_k, b_k), and an open interval is a connected metric space, it

[8] Equivalently, one can use Proposition 2, which characterizes strategyproof and anonymous mechanisms, as anonymity becomes trivial in case of a single agent.

follows that π is globally constant over (a_k, b_k). That is, there exists a value $t_k \in \mathbb{R}$ such that $\pi(y) = t_k$ for all $y \in (a_k, b_k)$.

We now show that for any $k \in \mathbb{N}$ with $a_k \neq b_k$ (i.e., the interval (a_k, b_k) is non-empty), it cannot be the case that both a_k and b_k are finite. Suppose for contradiction that both are finite. Note that continuity of π implies that $\pi(a_k) = \pi(b_k) = t_k$. However, since $a_k, b_k \notin O$, we have $\pi(a_k) = a_k$ while $\pi(b_k) = b_k$, which is a contradiction because $a_k \neq b_k$. Hence, for every $k \in \mathbb{N}$ with $a_k \neq b_k$, at least one of the two must lie in $\{-\infty, \infty\}$.

This leaves precisely five possibilities for the set O: $\emptyset$, $\mathbb{R}$, $(-\infty, a)$ for $a \in \mathbb{R}$, (b, ∞) for $b \in \mathbb{R}$, and $(-\infty, a) \cup (b, \infty)$ for $a, b \in \mathbb{R}$ with $b \geq a$. We know that π is constant over each interval in O, and the identity function for every point outside O. For each of these five cases, we show that π must be of the form given in part (a) by identifying the corresponding constants α^1 and α^2.

(1) $O = \emptyset$: π is the identity function everywhere, i.e., $\alpha^1 = -\infty$ and $\alpha^2 = \infty$.
(2) $O = \mathbb{R}$: There exists $t \in \mathbb{R}$ such that $\pi(y) = t$ for all $y \in \mathbb{R}$. This corresponds to $\alpha^1 = \alpha^2 = t$.
(3) $O = (-\infty, a)$ for $a \in \mathbb{R}$: Then $\pi(y) = y$ for all $y \geq a$. In particular, $\pi(a) = a$. Because π is continuous and constant over $(-\infty, a)$, we have $\pi(y) = a$ for $y \in (-\infty, a)$. This corresponds to $\alpha^1 = a$ and $\alpha^2 = \infty$.
(4) $O = (b, \infty)$ for $b \in \mathbb{R}$: Similarly to case (3), this corresponds to $\alpha^1 = -\infty$ and $\alpha^2 = b$.
(5) $O = (-\infty, a) \cup (b, \infty)$ for finite $b \geq a$: As argued in the previous two cases, for $y \in (-\infty, a)$ we have $\pi(y) = \pi(a) = a$, and for $y \in (b, \infty)$ we have $\pi(y) = \pi(b) = b$. For $y \in [a, b]$, we have $\pi(y) = y$. This corresponds to $\alpha^1 = a$ and $\alpha^2 = b$.

This concludes our proof. ■

A.5 Efficiency of Strategyproof Mechanisms

Figure 2 below verifies several claims made in the proof of Theorem 5.2 using Mathematica.

```
Minimize[{∑_{i=1}^{n} (a*i + (1 - a*X))^2, n > 2 && X > 2}, a]

Minimize[{(∑_{i=1}^{n} (a*i + b))^2 + (a*X + b - 1)^2, n > 2 && X > 2}, {a, b}]

{{(-n+n^3)/(2 (1+3 n+2 n^2-6 X-6 n X+6 X^2)), X > 2 && n > 2}, {a → {Indeterminate, True}}}

{{(-n+n^3)/(2+5 n+4 n^2+n^3-12 X-12 n X+12 X^2), X > 2 && n > 2}, {a → {Indeterminate, True}, b → {Indeterminate, True}}}

FullSimplify[(-n + n^3)/(2 + 5 n + 4 n^2 + n^3 - 12 X - 12 n X + 12 X^2) == (n^3-n)/(2 (1+3 n+2 n^2+6 X (X-n-1)))/(1 + (n^3-n)/(2 (1+3 n+2 n^2+6 X (X-n-1))))]

True
```

Fig. 2. Verification of various claims through Mathematica

We remark that none of the strategyproof mechanisms we study achieve a constant approximation. For instance, it is easy to show that L_1-ERM is n-efficient.

Proposition A.2. *The L_1-ERM mechanism is n-efficient.*

Proof. Fix $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i \in N}$. Let $\boldsymbol{\beta^1}$ and $\boldsymbol{\beta^*}$ be the outputs of L_1-ERM and OLS, respectively. Then, we have

$$\text{RSS}(\mathcal{D}, \boldsymbol{\beta^1}) \leq \left(\sum_{i \in N} \left|y_i - (\boldsymbol{\beta^1})^T \overline{\boldsymbol{x}_i}\right|\right)^2 \leq \left(\sum_{i \in N} \left|y_i - (\boldsymbol{\beta^*})^T \overline{\boldsymbol{x}_i}\right|\right)^2 \leq n \cdot \text{RSS}(\mathcal{D}, \boldsymbol{\beta^*}),$$

where the first inequality follows from the power mean inequality, the second inequality holds because $\boldsymbol{\beta}^{\mathbf{1}}$ minimizes the sum of absolute losses, and the third inequality follows from the Cauchy-Schwarz inequality. This concludes the proof. ■

B COMPUTING INFLUENCE BOUNDS

ALGORITHM 3: Computing Influence Bounds

Input: Data points $\mathcal{D} = (\boldsymbol{x}_j, y_j)_{j \in N}$, agent $i \in N$.
Output: ℓ_i, h_i
$Z \leftarrow$ set of hyperplanes $\boldsymbol{\beta}$ which pass through $d+1$ agents from $N \setminus \{i\}$;
$t_{\boldsymbol{\beta}} \leftarrow \boldsymbol{\beta}^T \overline{\boldsymbol{x}_i}, \forall \boldsymbol{\beta} \in Z$;
$L \leftarrow \min_{\boldsymbol{\beta} \in Z} t_{\boldsymbol{\beta}} - 1$;
$H \leftarrow \max_{\boldsymbol{\beta} \in Z} t_{\boldsymbol{\beta}} + 1$;
$V_L \leftarrow M^{\boldsymbol{x}}(L, \boldsymbol{y}_{-i})$;
$V_H \leftarrow M^{\boldsymbol{x}}(H, \boldsymbol{y}_{-i})$;
if $V_L = L$ **then**
  $\ell_i \leftarrow -\infty$;
else
  $\ell_i \leftarrow V_L$;
end
if $V_H = H$ **then**
  $h_i \leftarrow \infty$;
else
  $h_i \leftarrow V_H$;
end
return ℓ_i, h_i;

Our characterization result (Theorem 4.3) establishes existence of influence bounds $\ell_i, h_i \in \overline{\mathbb{R}}$ for each agent i as a function of the reports of the other agents. In this section, we address the problem of computing these influence bounds for a given strategyproof mechanism.

Fix $\boldsymbol{y}_{-i}$. We being from the simple observation that if ℓ_i is finite, then for a sufficiently low value of y_i (any $y_i \le \ell_i$), we have that the outcome for agent i will be $\widehat{y}_i = \text{med}(y_i, \ell_i, h_i) = \ell_i$. If $\ell_i = -\infty$, then for all $y_i < h_i$, the outcome for agent i will be $\widehat{y}_i = y_i$. Thus, if we can identify a *sufficiently low* value of y_i, we can check if $\widehat{y}_i$ is equal to y_i (in which case $\ell_i = -\infty$), or $\widehat{y}_i$ is equal to some other value (in which case this value must be ℓ_i). A symmetric observation holds for h_i.

While it is difficult to pin down a sufficiently low value for an arbitrary strategyproof mechanism, we can do so for the class of strategyproof mechanisms which are guaranteed to pass through $d+1$ data points in $d+1$ dimensions (e.g., the generalized resistant hyperplane mechanisms).

In this case, note that ℓ_i, if finite, must be the point where a hyperplane containing *some* $d+1$ agents (excluding agent i) intersects the vertical line at $\boldsymbol{x}_i$. Thus, if we iterate through all hyperplanes passing through $d+1$ agents except agent i, and find their intersections with the vertical line at $\boldsymbol{x}_i$, then any value lower than the lowest intersection point will work as a sufficiently low value. Once again, a symmetric observation can be made for h_i.

This provides an algorithm that runs in time that is polynomial in n, but exponential in d, and makes two calls to the strategyproof mechanism (one to identify ℓ_i and one for h_i). This is presented as Algorithm 3.

C QUANTILE REGRESSION IS NOT STRATEGYPROOF

In this section, we show that quantile regression is not guaranteed to be strategyproof. In particular, we show that quantile regression with $q = 0.4$ violates strategyproofness. The coordinates for the 20 data points shown in Figure 3 are as follows.

(-79.3, -45.8)	(-77.3, 89.5)	(-74.8, -87.4)	(-58.5, 14.3)	(-33.2, -28.4)
(-31.5, 5.2)	(-8.0, -73.1)	(-1.7, -52.8)	(10.0, 88.6)	(13.0, 13.3)
(13.9, 7.4)	(15.4, 39.4)	(18.5, -2.0)	(23.0, 6.6)	(23.8, -33.0)
(24.2, -60.3)	(26.0, 49.5)	(39.5, 49.5)	(45.3, 88.9)	(71.2, 33.2)

If the agents report truthfully, then the quantile regression mechanism with $q = 0.4$ returns the solid line. If agent with data point $(13.9, 7.4)$ reports a very large value of y (e.g., 2000), then the output line becomes the dashed one, which is clearly beneficial for the manipulating agent.

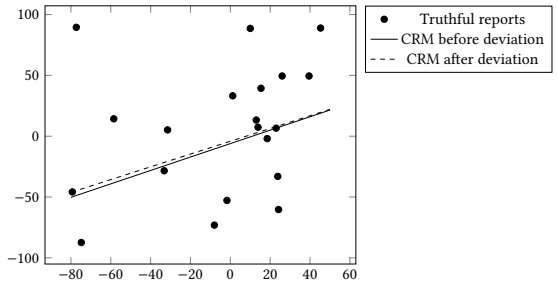


Fig. 3. Example of a beneficial manipulation under quantile regression with $q = 0.4$.