Pure Nash Equilibria in Linear Regression

Safwan Hossain
University of Toronto, Vector Institute
safwan.hossain@mail.utoronto.ca

Nisarg Shah
University of Toronto
nisarg@cs.toronto.edu

Abstract

We build on an emerging line of work which studies strategic manipulations in training data provided to machine learning algorithms. Specifically, we focus on the ubiquitous task of linear regression. Prior work focused on the design of strategyproof algorithms, which aim to prevent such manipulations altogether by aligning the incentives of data sources. However, algorithms used in practice are often not strategyproof, which induces a strategic game among the agents. We focus on a broad class of non-strategyproof algorithms for linear regression, namely ℓ_p norm minimization ($p > 1$) with convex regularization. We show that when manipulations are bounded, every algorithm in this class admits a unique pure Nash equilibrium outcome. We also shed light on the structure of this equilibrium by uncovering a surprising connection between strategyproof algorithms and pure Nash equilibria of non-strategyproof algorithms in a broader setting, which may be of independent interest. Finally, we analyze the quality of equilibria under these algorithms both theoretically and empirically.

1 Introduction

Linear regression aims to find a linear relationship between explanatory variables and response variables. Under certain assumptions, it is known that minimizing a suitable loss function on training data generalizes well to unseen test data (Bousquet, Luxburg, and Gunnar 2004). However, traditional analysis assumes that the algorithm has access to untainted data drawn from the underlying distribution. Relaxing this assumption, a significant body of recent work has focused on making machine learning algorithms robust to stochastic or adversarial noise; the former is too benign (Littlestone 1988; Goldman and Sloan 1995; Frénay and Verleysen 2013; Natarajan et al. 2013), while the latter is too pessimistic (Kearns and Li 1993; Bshouty, Eiron, and Kushilevitz 2002; Chen, Caramanis, and Mannor 2013; Gu and Rigazio 2014). A third, more recent and prescient model is that of *strategic noise*, which is a game-theoretic modeling of noise that sits in between the two. Here, it is assumed that the training set is provided by self-interested agents, who may manipulate to minimize loss on their own data.

We focus on strategic noise in linear regression. Dekel, Fischer, and Procaccia (2010) provide an example of retailer Zara, which uses regression to predict product demand at each store, partially based on self-reported data provided by the stores. Given limited supply of popular items, store managers may engage in strategic manipulation to ensure the distribution process benefits them, and there is substantial evidence that this is widespread (Caro et al. 2010). Strategic behavior by even a small number of agents can significantly affect the overall system, including agents who have not participated in such behavior. Prior work has focused on designing *strategyproof* algorithms for linear regression (Perote and Perote-Pena 2004; Dekel, Fischer, and Procaccia 2010; Chen et al. 2018), under which agents provably cannot benefit by misreporting their data. While strategyproofness is a strong guarantee, it is only satisfied by severely restricted algorithms. Indeed, as we observe later in the paper, most practical algorithms for linear regression are *not* strategyproof.

When strategic agents with competing interests manipulate the input data under a non-strategyproof algorithm, a game is induced between them. Game theory literature offers several tools to analyze such behaviour, such as Nash equilibria and the price of anarchy (Nisan et al. 2007). We use these tools to answer three key questions:

- Does the induced game always admit a pure Nash equilibrium?
- What are the characteristics of these equilibria?
- Is there a connection between strategyproof algorithms and equilibria of non-strategyproof algorithms?

We focus on linear regression algorithms which minimize the ℓ_p-norm of residuals (where $p > 1$) with convex regularization. This class includes most popular linear regression algorithms, including the ordinary least squares (OLS), lasso, group lasso, ridge regression, and elastic net regression. Our key result is that the game induced by an algorithm in this class has three properties: a) it always has a pure Nash equilibrium, b) all pure Nash equilibria result in the same regression hyperplane, and c) there exists a strategyproof algorithm which returns this equilibrium regression hyperplane given non-manipulated data. We also analyze the quality of this equilibrium outcome, measured by the pure price of an-

archy. Our theoretical results show that the worst-case pure price of anarchy is unbounded. Experiments with synthetic and real datasets reveal a less pessimistic average-case scenario and uncover a surprising effect of several parameters on the pure price of anarchy.

1.1 Related Works

A special case of linear regression is facility location in one dimension (Moulin 1980), where each agent i is located at some y_i on the real line. An algorithm elicits the preferred locations of the agents (who can misreport) and chooses a location $\overline{y}$ to place a facility. A significant body of literature in game theory is devoted to understanding strategyproof algorithms in this domain (Moulin 1980; Caragiannis, Procaccia, and Shah 2016), which includes placing the facility at the median of the reported locations. A more recent line of work studies equilibria of non-strategyproof algorithms such as placing the facility at the average of the reported locations (Renault and Trannoy 2005; 2011; Yamamura and Kawasaki 2013). Similarly, in the more general linear regression setting, prior work has focused on strategyproof algorithms (Perote and Perote-Pena 2004; Dekel, Fischer, and Procaccia 2010; Chen et al. 2018). We complete the picture by studying equilibria of non-strategyproof algorithms for linear regression.

We use a standard model of strategic manipulations in linear regression (Perote and Perote-Pena 2004; Dekel, Fischer, and Procaccia 2010; Chen et al. 2018). Perote and Perote-Pena (2004) designed a strategyproof algorithm in two dimensions. Dekel, Fischer, and Procaccia (2010) proved that least absolute deviations (LAD), which minimizes the ℓ_1-norm of residuals without regularization, is strategyproof. Chen et al. (2018) extended their result to include regularization, and designed a new family of strategyproof algorithms in high dimensions. They also analyzed the loss in mean squared error (MSE) under a strategyproof algorithm as compared to the OLS, which minimizes MSE. They showed that any strategyproof algorithm has at least twice as much MSE as the OLS in the worst case, and that this ratio is $\Theta(n)$ for LAD. Our result (Theorem 6) shows that this ratio is unbounded under equilibria of algorithms we study. Through a connection we establish to strategyproof algorithms (Theorem 5), this also implies unbounded loss for the corresponding strategyproof algorithms.

Finally, we mention that strategic manipulations have been studied in various other machine learning contexts, e.g., manipulations of feature vectors (Hardt et al. 2016; Dong et al. 2018), strategic classification (Meir, Procaccia, and Rosenschein 2012; Hardt et al. 2016; Dong et al. 2018), competition among different algorithms (Mansour, Slivkins, and Wu 2017; Immorlica et al. 2011; Ben-Porat and Tennenholtz 2017; 2019), or manipulations due to privacy concerns (Cummings, Ioannidis, and Ligett 2015; Cai, Daskalakis, and Papadimitriou 2015).

2 Model

In linear regression, we are given n data points of the form $(\boldsymbol{x_i}, y_i)$, where $\boldsymbol{x_i} \in \mathbb{R}^d$ are the explanatory variables, and $y_i \in \mathbb{R}$ is the response variable.[1] Let $\boldsymbol{X}$ be the matrix with $\boldsymbol{x_i}$ as its i^{th} column, and $\boldsymbol{y} = (y_1, \ldots, y_n)$. The goal of a linear regression algorithm is to find a hyperplane with normal vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}^T \boldsymbol{x_i}$ is a good estimate of y_i. The residual of point i is $r_i = |y_i - \boldsymbol{\beta}^T x_i|$.

Algorithms: We focus on a broad class of algorithms parametrized by $p > 1$ and function $R : \mathbb{R}^d \to \mathbb{R}$. The (p, R)-regression algorithm minimizes the following loss function over $\boldsymbol{\beta}$:

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}) = \textstyle\sum_{i=1}^{n} |y_i - \boldsymbol{\beta}^T \boldsymbol{x_i}|^p + R(\boldsymbol{\beta}). \qquad (1)$$

We assume that R is convex and differentiable. For $p > 1$, this is strictly convex, admitting a unique optimum $\boldsymbol{\beta}^*$. When there is no regularization, we refer to it as the $(p, 0)$-regression algorithm.

Strategic model: We follow a standard model of strategic interactions studied in the literature (Perote and Perote-Pena 2004; Dekel, Fischer, and Procaccia 2010; Chen et al. 2018). Data point $(\boldsymbol{x_i}, y_i)$ is provided by an agent i. $N = [n] := \{1, \ldots, n\}$ denotes the set of all agents. $\boldsymbol{x_i}$ is public information, which is non-manipulable, but y_i is held private by agent i. We assume a subset of agents $H \subset N$ (with $h = |H|$) are honest and always report $\widetilde{y}_i = y_i$. The remaining agents in $M = N \setminus H$ (with $m = |M|$) are strategic and may report $\widetilde{y}_i \neq y_i$. For convenience, we assume that $M = [m]$ and $H = \{m + 1, \ldots, n\}$. However, we emphasize that our algorithms do not distinguish between strategic and honest agents. Given a set of reports $\widetilde{\boldsymbol{y}}$, honest agents' reports are denoted by $\widetilde{\boldsymbol{y}}_H$ (note that $\widetilde{\boldsymbol{y}}_H = \boldsymbol{y}_H$) and strategic agents' reports by $\widetilde{\boldsymbol{y}}_M$.

The (p, R)-regression algorithm takes as input $\boldsymbol{X}$ and $\widetilde{\boldsymbol{y}}$, and returns $\boldsymbol{\beta^*}$ minimizing the loss in Equation (1). We say that $\overline{y}_i = (\boldsymbol{\beta^*})^T \boldsymbol{x_i}$ is the *outcome* for agent i. Since $\boldsymbol{X}$ and $\boldsymbol{y}_H$ are non-manipulable, we can treat them as fixed. Hence, $\widetilde{\boldsymbol{y}}_M$ is the only input which matters, and similarly, $\overline{\boldsymbol{y}}_M$ is the output which matters. For an algorithm f, we use the notation $f(\widetilde{\boldsymbol{y}}_M) = \overline{\boldsymbol{y}}_M$, and let f_i denote the function returning agent i's outcome $\overline{y}_i$. A strategic agent i manipulates to ensure this outcome is as close to her true response variable y_i as possible. Formally, agent i has *single-peaked preferences* $\succeq_i$ (with strict preference denoted by $\succ_i$) over $\overline{y}_i$ with peak at y_i. That is, for all $a < b \leq y_i$ or $a > b \geq y_i$, we have $b \succ_i a$. Agent i is *perfectly happy* when $\overline{y}_i = y_i$. In this work, we assume that for each agent i, both y_i and $\widetilde{y}_i$ are bounded (WLOG, say they belong to $[0, 1]$).

Nash equilibria: This strategic interaction induces a game amongst agents in M, and we are interested in the pure Nash equilibria (PNE) of this game. We say that $\widetilde{\boldsymbol{y}}_M$ is a *Nash equilibrium* (NE) if no strategic agent $i \in M$ can strictly gain by changing her report, i.e., if $\forall i, \forall \widetilde{y}'_i$, $f_i(\widetilde{\boldsymbol{y}}_M) \succeq_i f_i(\widetilde{y}'_i, \widetilde{\boldsymbol{y}}_{M \setminus \{i\}})$. We say that $\widetilde{\boldsymbol{y}}_M$ is a *pure Nash equilibrium* (PNE) if it is a NE and each $\widetilde{y}_i$ is deterministic. Let $\mathrm{NE}_f(\boldsymbol{y})$ denote the set of pure Nash equilibria under

[1] Following standard convention, we assume the last component of each $\boldsymbol{x_i}$ is a constant, say 1.

f when the peaks of agents' preferences are given by $\boldsymbol{y}$.[2] For $\widehat{\boldsymbol{y}}_M \in \mathrm{NE}_f(\boldsymbol{y})$, let $f(\widehat{\boldsymbol{y}}_M)$ be the corresponding *PNE outcome*.

Strategyproofness: We say that an algorithm f is *strategyproof* if no agent can benefit by misreporting her true response variable regardless of the reports of the other agents, i.e., $\forall i, \forall \widetilde{\boldsymbol{y}}_M, f_i(y_i, \widetilde{\boldsymbol{y}}_{M\setminus\{i\}}) \succeq_i f_i(\widetilde{\boldsymbol{y}}_M)$. Note that strategyproofness implies that each agent reporting her true value (i.e. $\widetilde{\boldsymbol{y}}_M = \boldsymbol{y}_M$) is a pure Nash equilibrium.

Pure price of anarchy (PPoA): It is natural to measure the cost of selfish behavior on the overall system. A classic notion is the *pure price of anarchy* (PPoA) (Koutsoupias and Papadimitriou 1999; Nisan et al. 2007), which is defined as the ratio between the maximum social cost under any PNE and the optimal social cost under honest reporting, for an appropriate measure of social cost. Here, social cost is a measure of the overall fit. In regression, it is typical to measure fit using the ℓ_q norm of absolute residuals for some q. While we study the equilibrium of ℓ_p regression mechanism for different p values, we need to evaluate them using a single value of q, so that the results are comparable. For our theoretical analysis, we use mean squared error (which corresponds to $q = 2$) since it is the standard measure of fit in literature (Chen et al. 2018). One way to interpret our results is: *If our goal were to minimize the MSE, which ℓ_p regression mechanism would we choose, assuming that the strategic agents would achieve equilibrium?* In our empirical simulations, we also present results for other values of q. Slightly abusing the notation by letting f map all reports to all outcomes (not just for agents in M), we formally write:

$$\mathrm{PPoA}(f) = \max_{\boldsymbol{y}\in[0,1]^n} \frac{\max_{\overline{\boldsymbol{y}}\in \mathrm{NE}_f(\boldsymbol{y})} \sum_{i=1}^n |y_i - \overline{y}_i|^2}{\sum_{i=1}^n |y_i - \overline{y}_i^{\mathrm{OLS}}|^2},$$

where $\overline{\boldsymbol{y}}^{\mathrm{OLS}}$ is the outcome of OLS (i.e. the $(2,0)$-regression algorithm) under honest reporting, which minimizes mean squared error. Note that the PPoA, as we have defined it, measures the impact of the behavior of strategic agents on all agents, including on the honest agents.

3 Warm-Up: The 1D Case

As a warm-up, we briefly review the more restricted facility location setting in one dimension. Here, each agent i has an associated scalar value $y_i \in [0,1]$ and the algorithm must produce the same outcome for all agents (i.e. $\overline{y}_i = \overline{y}_j \, \forall i,j \in N$). Hence, the algorithm is a function $f : [0,1]^m \to \mathbb{R}$. This is a special case of linear regression where agents have identical independent variables.

We provide a detailed overview of prior work in this 1D setting in Appendix A. Briefly, in this setting, the (p,R)-regression algorithm described in Section 2 reduces to $f(\widetilde{y}_1,\ldots,\widetilde{y}_m) = \arg\min_{\overline{y}\in\mathbb{R}} \sum_{i=1}^m |\widetilde{y}_i - \overline{y}|^p + \sum_{i=m+1}^n |y_i - \overline{y}|^p + R(\overline{y})$. For $p = 1$, this is known to be strategyproof (Chen et al. 2018). However, for $p > 1$, which is the focus of our work, this is not strategyproof. Yamamura and Kawasaki (2013) show that for a wide family of facility location algorithms, including the $(p,0)$-regression algorithm for $p > 1$ with no honest agents or regularization, there is always a pure Nash equilibrium, the PNE outcome is unique, and the outcome matches with that of a strategyproof algorithm. Below, we extend this to all (p,R)-regression algorithm with $p > 1$ and convex regularizer R (and with the possibility of honest agents). We omit the proof because, in the next section, we prove this more generally for the linear regression setting (Theorems 3, 4, and 5).

Theorem 1. *Consider facility location with n agents, of which a subset of agents M are strategic and have single-peaked preferences with peaks at $\boldsymbol{y}_M \in [0,1]^m$. Let f denote the (p,R)-regression algorithm with $p > 1$ and convex regularizer R. Then, the following statements hold for f.*

1. *For each $\boldsymbol{y}_M$, there is a pure Nash equilibrium $\widehat{\boldsymbol{y}}_M \in \mathrm{NE}_f(\boldsymbol{y}_M)$.*
2. *For each $\boldsymbol{y}_M$, all pure Nash equilibria $\widehat{\boldsymbol{y}}_M \in \mathrm{NE}_f(\boldsymbol{y}_M)$ have the same outcome $f(\widehat{\boldsymbol{y}}_M)$.*
3. *There exists a strategyproof algorithm h such that for all $\boldsymbol{y}_M$ and all pure Nash equilibria $\widehat{\boldsymbol{y}}_M \in \mathrm{NE}_f(\boldsymbol{y}_M)$, $f(\widehat{\boldsymbol{y}}_M) = h(\boldsymbol{y}_M)$.*

Theorem 1 guarantees the existence of a pure Nash equilibrium and highlights an interesting structure of the equilibrium. The next immediate question is to analyze the quality of this equilibrium. We show that the PPoA of any $(p,0)$-regression algorithm (i.e. without regularization) is $\Theta(n)$. Interestingly, this holds even if only a single agent is strategic, and the bound is independent of p. The proof is given in Appendix A.

Theorem 2. *Consider facility location with n agents, of which a subset of agents M are strategic. Let f denote the $(p,0)$-regression algorithm with $p > 1$. When $|M| \geq 1$, $\mathrm{PPoA}(f) = \Theta(n)$.*

We remark that Theorems 1 and 2, due to their generality, are novel results in the facility location setting.

4 Linear Regression

We now turn to the more general linear regression setting, which is the focus of our work, and highlight interesting similarities and differences to the facility location setting. Recall that for linear regression, the (p,R)-regression algorithm finds the optimal $\boldsymbol{\beta}^*$ minimizing the loss function:

$$\mathcal{L}(\widetilde{\boldsymbol{y}}, \boldsymbol{X}, \boldsymbol{\beta}) = \sum_{i=1}^m |\widetilde{y}_i - \boldsymbol{\beta}^T \boldsymbol{x_i}|^p + \sum_{i=m+1}^n |y_i - \boldsymbol{\beta}^T \boldsymbol{x_i}|^p + R(\boldsymbol{\beta})$$

For a strategic agent $i \in M$, recall that we denote her outcome by $\overline{y}_i = (\boldsymbol{\beta}^*)^T \boldsymbol{x_i}$. Let $\mathrm{br}_i(\widetilde{\boldsymbol{y}}_{-i}) = \{\widetilde{y}_i \in [0,1] : f_i(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i}) \succeq_i f_i(\widetilde{y}'_i, \widetilde{\boldsymbol{y}}_{-i}) \, \forall \widetilde{y}'_i \in [0,1]\}$ denote the set of her best responses as a function of the reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents. Informally, it is the set of reports that agent i can submit to induce her most preferred outcome.

4.1 Properties of the Algorithm, Best Responses, and Pure Nash Equilibria

We begin by establishing intuitive properties of (p,R)-regression algorithms. We first derive the following result,

[2] Equilibria can generally depend on the full preferences, but results in Section 4 show only peaks matter.

whose proof is in Appendix B.

Lemma 1. *The outcome $\overline{y}_i$ of agent i is continuous in $\widetilde{\boldsymbol{y}}$, and strictly increasing in her own report $\widetilde{y}_i$ for any fixed reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents.*

This property demonstrates that (p, R)-regression is not strategyproof. Consider an instance where each strategic agent i has $y_i \notin \{0, 1\}$ and their true data points do not lie on a hyperplane. Then under honest reporting, not all strategic agents can be perfectly happy, and any agent i with $\overline{y}_i > y_i$ (or $\overline{y}_i < y_i$) can slightly decrease (or increase) her report to achieve a strictly more preferred outcome. Next, we show that the best response of an agent is always unique and continuous in the reports of the other agents. The slightly intricate proof of this result is provided in Appendix B.

Lemma 2. *For each strategic agent i, the following hold about the best response function br_i.*

1. *The best response is unique, i.e., $|\mathrm{br}_i(\widetilde{\boldsymbol{y}}_{-i})| = 1$ for any reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents.*
2. *br_i is a continuous function of $\widetilde{\boldsymbol{y}}_{-i}$.*

We remark that part 1 of Lemma 2 is a strong result: it establishes a unique best response for every possible single-peaked preferences that the agent may have (in fact, our proof shows that this best response depends only on the peak and not on the full preferences). This allows us to avoid further assumptions on the structure of the agent preferences.

Finally, we derive a simple characterization of pure Nash equilibria in our setting. We show that under a PNE, each strategic agent i must be in one of three states: either she is perfectly happy ($\overline{y}_i = y_i$), or wants to decrease her outcome ($\overline{y}_i > y_i$) but is already reporting $\widetilde{y}_i = 0$, or wants to increase her outcome ($\overline{y}_i < y_i$) but is already reporting $\widetilde{y}_i = 1$. The proof is provided in Appendix B.

Lemma 3. *$\widetilde{\boldsymbol{y}}_M$ is a pure Nash Equilibrium if and only if $(\overline{y}_i < y_i \wedge \widetilde{y}_i = 1) \vee (\overline{y}_i > y_i \wedge \widetilde{y}_i = 0) \vee (\overline{y}_i = y_i)$ holds for all $i \in M$.*

Note that Lemma 3 immediately implies a naïve but simple algorithm to find a pure Nash equilibrium. Since $\widetilde{y}_i \in \{0, y_i, 1\}$ for each i, this induces 3^m possible $\widetilde{\boldsymbol{y}}_M$ vectors. For each such vector, we can compute the outcome of the mechanism $\overline{\boldsymbol{y}}$, and check whether the conditions of Lemma 3 are satisfied. This might lead one to believe that the strategic game that we study is equivalent to the finite game induced by the 3^m possible strategy profiles. However, this is not true because limiting the strategy set of the agents can give rise to new equilibria which are not equilibria of the original game; we give an example illustrating this in Appendix D. We further discuss the issue of computing a PNE in Section 4.4.

4.2 Analysis of Pure Nash Equilibria

We are now ready to prove the main results of our work. We begin by showing that a PNE always exists, generalizing the first statement of Theorem 1 from 1D facility allocation to linear regression.

Theorem 3. *For $p > 1$ and convex regularizer R, the (p, R)-regression algorithm admits a pure Nash Equilibrium.*

Proof. Consider the mapping T from the reports of strategic agents to their best responses, i.e., $T(\widetilde{y}_1, \ldots, \widetilde{y}_m) = (\mathrm{br}_1(\widetilde{\boldsymbol{y}}_{-1}), \ldots, \mathrm{br}_m(\widetilde{\boldsymbol{y}}_{-m}))$. Recall that best responses are unique due to Lemma 2. Also, note that pure Nash equilibria are precisely fixed points of this mapping.

Brouwer's fixed point theorem states that any continuous function from a convex compact set to itself has a fixed point (Pugh 2003). Note that T is a function from $[0, 1]^m$ to $[0, 1]^m$, and $[0, 1]^m$ is a convex compact set. Further, T is a continuous function since each br_i is a continuous function (Lemma 2). Hence, by Brouwer's fixed point theorem, T has a fixed point (i.e. pure Nash equilibrium). □

Next, we show that there is a unique pure Nash equilibrium outcome, generalizing the second statement in Theorem 1. Note that this does not imply uniqueness of PNE. The proof is in Appendix C.

Theorem 4. *For $p > 1$ and convex regularizer R, the (p, R)-regression algorithm has a unique pure Nash equilibrium outcome.*

The non-uniqueness of equilibrium strategy stems from different sets of reports mapping to the same regression hyperplane. In the simplest case, consider the ordinary least squares (OLS) with no regularization, i.e., the $(2, 0)$-regression, where all n agents are strategic. Given $\boldsymbol{X} \in \mathbb{R}^{d \times n}$, the OLS produces a linear mapping from the reports $\widetilde{\boldsymbol{y}}$ to the outcomes $\overline{\boldsymbol{y}}$ given by $\boldsymbol{H}\widetilde{\boldsymbol{y}} = \overline{\boldsymbol{y}}$, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \in \mathbb{R}^{n \times n}$ is a symmetric idempotent matrix of rank d (known as the hat matrix). When $n > d$, $\boldsymbol{H}$ is singular, leading to infinitely many $\widetilde{\boldsymbol{y}}$ which map to the same $\overline{\boldsymbol{y}}$. Of course, they need to still satisfy the conditions of being a PNE (Lemma 3). For a concrete example, suppose that the n data points lie on a hyperplane. Then, any of the infinitely many reports $\widetilde{\boldsymbol{y}}$ under which OLS returns this hyperplane — making all n agents perfectly happy — is a PNE.

Given the linear structure of OLS, one wonders if our results can be extended to all *linear mappings*. We say a game is induced by a linear mapping if a matrix $\boldsymbol{H}$ relates the agents' outcomes $\overline{\boldsymbol{y}}$ to their reports $\widetilde{\boldsymbol{y}}$ by the equation $\boldsymbol{H}\widetilde{\boldsymbol{y}} = \overline{\boldsymbol{y}}$. When $\boldsymbol{H}$ is a hat matrix arising from OLS, Theorems 3 and 4 show that the induced game admits a PNE with a unique outcome. Interestingly, it is easy to show that the proof of Theorem 3 (existence of PNE) can be extended to all matrices $\boldsymbol{H}$. However, in Appendix C, we show an example matrix $\boldsymbol{H}$ for which the corresponding game has multiple PNE outcomes. It is an interesting open question to identify the precise conditions on $\boldsymbol{H}$ for the induced game to satisfy Theorem 4 and have a unique PNE outcome.

4.3 Connection to Strategyproofness

A social choice rule maps true preferences of the agents ($\boldsymbol{y}$) to a socially desirable outcome ($\overline{\boldsymbol{y}}$ or $\boldsymbol{\beta}^*$). Strategyproofness is a strong requirement: when f is strategyproof, honest reporting is a *dominant strategy* for each agent (i.e., it is an optimal strategy regardless of the strategies of other agents). We say that rule f is implementable in dominant strategies if there exists a rule g such that $f(\boldsymbol{y})$ is a dominant strategy

outcome under g. Although a seemingly weaker requirement (since for a strategyproof rule f, one can set $g = f$), the classic revelation principle argues otherwise: if f can be implemented in dominant strategies, then directly eliciting agents' preferences and implementing f must be strategyproof.

A truly weaker requirement is that f be *Nash-implementable*, i.e., that there be a rule g such that $f(\boldsymbol{y})$ is a Nash equilibrium under g.[3] Generally, not every Nash-implementable rule is strategyproof.

However, in restricted domains, this may be true. A classic line of work in economics (Roberts 1979; Dasgupta, Hammond, and Maskin 1979; Laffont and Maskin 1982) proves this for "rich" preference domains. It is easy to check that our domain with single-peaked preferences does not satisfy their "richness" condition. For single-peaked preferences, we noted in Section 3 that Yamamura and Kawasaki (2013) proved such a result in 1D facility location for a family of algorithms with unique PNE outcomes. We extend this to the more general linear regression setting. At this point, we make two remarks. First, the result we establish is stronger than the revelation principle (albeit in this specific domain) as it "converts" Nash-implementability (rather than the stronger dominant-strategy-implementability) into strategyproofness. Second, the result of Yamamura and Kawasaki (2013) for 1D facility location relied on the analytical form of the PNE outcome, so strategyproofness could be explicitly checked. However, the analytical form of the PNE outcome is unknown in the linear regression setting, requiring an indirect argument to establish strategyproofness.

We note that our result actually applies to a even broader setting than linear regression: specifically, it applies to any function $f : [0,1]^m \to \mathbb{R}^m$ which has a unique PNE outcome and satisfies an additional condition. We believe that this could have further implications in the theory about implementability of rules, and may be of independent interest.

Theorem 5. *Let M be a set of agents with $|M| = m$. Each agent i holds a private $y_i \in [0,1]$. Let f be a function which elicits agent reports $\widetilde{\boldsymbol{y}} \in [0,1]^m$ and returns an outcome $\overline{\boldsymbol{y}} \in \mathbb{R}^m$. Each agent i has single-peaked preferences over $\overline{y}_i$ with peak at y_i. Suppose the following are satisfied:*

1. *For each $i \in M$ and each $\widetilde{\boldsymbol{y}}_{-i} \in [0,1]^{m-1}$, $\overline{y}_i = f_i(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i})$ is continuous and strictly increasing in $\widetilde{y}_i$.*
2. *For each $\boldsymbol{y} \in [0,1]^m$ and each $T \subseteq M$, f has a unique pure Nash equilibrium outcome when agents in T report honestly and agents in $M \setminus T$ strategize.*

For $\boldsymbol{y} \in [0,1]^m$, let $h(\boldsymbol{y})$ denote the unique pure Nash equilibrium outcome under f when all agents strategize. Then, h is strategyproof.

Proof. Let $\boldsymbol{y}$ denote the true peaks of agent preferences. To show that h is strategyproof, we need to show that each agent i weakly prefers reporting her true y_i to any other y'_i, regardless of the reports $\boldsymbol{y}'_{-i}$ submitted to h by the other agents. Fix $\boldsymbol{y}'_{-i}$. Let h_i denote the outcome of h for agent i. We want to show that $h_i(y_i, \boldsymbol{y}'_{-i}) \succeq_i h_i(y'_i, \boldsymbol{y}'_{-i})$ for all $y'_i \in [0,1]$.

[3]This is weaker because for a strategyproof rule f, $f(\boldsymbol{y})$ is a dominant strategy equilibrium (and thus also a Nash equilibrium).

Note that $h(y'_i, \boldsymbol{y}'_{-i})$ finds the unique PNE outcome under f in the hypothetical scenario where the agents' preferences have peaks at $\boldsymbol{y}'$, as opposed to the real scenario in which the peaks are at $\boldsymbol{y}$. Let us introduce a helper function $g_i : [0,1] \to \mathbb{R}$ such that $g_i(\lambda)$ returns the unique PNE outcome for agent i under f, when the report of agent i is fixed to λ and the other agents strategize according to their preferences $\boldsymbol{y}'_{-i}$ (this is well defined due to condition 2 of the theorem). Note that this is independent of agent i's preferences as we fixed her report. Let $\widehat{\boldsymbol{y}}_{-i}$ be an equilibrium strategy of the other agents in this case. Then, $(\lambda, \widehat{\boldsymbol{y}}_{-i})$ is a PNE under f for all m agents with preferences $\boldsymbol{y}'$ if and only if agent i is happy with reporting λ. The other agents are already happy given agent i's report. Using condition 1 of the theorem and an argument similar to Lemma 3, this is equivalent to

$$(g_i(\lambda) > y'_i \wedge \lambda = 0) \ \vee (g_i(\lambda) < y'_i \wedge \lambda = 1) \ \vee (g_i(\lambda) = y'_i) \tag{2}$$

By condition 2 of the theorem, we know that for each $y'_i \in [0,1]$, there exists a unique $\lambda^*(y'_i)$ satisfying Equation (2). Note that $h_i(y'_i, \boldsymbol{y}'_{-i}) = g_i(\lambda^*(y'_i))$. Using this, we can derive three key properties of the function g_i. Let $a = g_i(0)$ and $b = g_i(1)$.

- $\boldsymbol{a \leq b}$ **:** Assume for contradiction that $a > b$. Choose $y'_i \in (b, a)$. Note that $\lambda = 0$ implies $g_i(\lambda) = a > y'_i$, which satisfies the first clause of Equation (2), while $\lambda = 1$ implies $g_i(\lambda) = b < y'_i$, which satisfies the second clause of Equation (2). Hence, both $\lambda = 0$ and $\lambda = 1$ satisfy Equation (2), which is a contradiction, since λ^* is unique.
- $\boldsymbol{\forall \lambda \in [0,1], g_i(\lambda) \in [a,b]}$ **:** Assume for contradiction that there exists $\widehat{\lambda} \in [0,1]$ such that $g_i(\widehat{\lambda}) \notin [a,b]$. WLOG, assume $g_i(\widehat{\lambda}) = k < a$ (hence, $\widehat{\lambda} \neq 0$). Choose $y'_i = k$. Note that $\lambda = 0$ implies $g(\lambda) = a > k = y'_i$, which satisfies the first clause of Equation (2). Similarly, for $\lambda = \widehat{\lambda}$, we have $g_i(\widehat{\lambda}) = k = y'_i$, which satisfies the third clause of Equation (2). Hence, both $\lambda = 0$ and $\lambda = \widehat{\lambda} \neq 0$ satisfy Equation (2), which is a contradiction.
- $\boldsymbol{g_i : [0,1] \to [a,b]}$ **is surjective/onto:** Assume for contradiction that there exists $\exists c \in (a,b)$ such that $g(\lambda) \neq c$ for any $\lambda \in [0,1]$. Choose $y'_i = c$. Hence, there is no λ satisfying the third clause in Equation (2). We see that for $\lambda = 0$, we have $g_i(\lambda) = a < c$, which violates the first clause. Similarly, for $\lambda = 1$, we have $g_i(\lambda) = b > c$, which violates the second clause. Hence, there is no λ satisfying Equation (2), which is again a contradiction.

We are now ready to show that $h_i(y_i, \boldsymbol{y}'_{-i}) = g_i(\lambda^*(y_i)) \succeq_i g_i(\lambda^*(y'_i)) = h_i(y'_i, \boldsymbol{y}'_{-i})$ for all $y'_i \in [0,1]$. If $y_i \in [a,b]$, then it is easy to see that $\lambda^*(y_i)$ is the unique value which satisfies $g_i(\lambda^*(y_i)) = y_i$ (this exists because g_i is onto). That is, in the equilibrium where agent i reports her true preference, she is perfectly happy. If $y_i < a$, then it is easy to check that $\lambda^*(y_i) = 0$ satisfies Equation (2), and we have $g_i(\lambda^*(y_i)) = a$. Since $g_i(\lambda^*(y'_i)) \in [a,b]$ for any y'_i, she will not strictly prefer this outcome. A symmetric argument holds for the $y_i > b$ case as well. Hence, we have established strategyproofness of h. □

Corollary 1. *Let f denote the (p, R)-regression algorithm with $p > 1$ and convex regularizer R. Then, there exists a strategyproof algorithm h such that $\forall \boldsymbol{y} \in [0,1]^m$ and $\widehat{\boldsymbol{y}} \in \mathrm{NE}_f(\boldsymbol{y})$, $f(\widehat{\boldsymbol{y}}) = h(\boldsymbol{y})$.*

Proof. We already established that the (p, R)-regression algorithm satisfies the conditions of Theorem 5. Specifically, f_i is continuous and strictly increasing in the report of agent i (Lemma 1). The second condition follows from Theorems 3 and 4, which hold irrespective of which agents are strategic and which are honest. Hence, the result follows immediately from Theorem 5. □

4.4 Computation of Pure Nash Equilibria

So far, our results in general linear regression draw similar conclusions with the 1D facility location setting. We proved that in both cases, a PNE exists, the PNE outcome is unique, and it coincides with the outcome of a strategyproof algorithm. However, there are fundamental differences between the two settings, which we now highlight.

The first deals with the computation of pure Nash equilibria. In facility location, a fully constructive characterization of strategyproof algorithms is known (Moulin 1980). This, along with Theorem 1 and the formula of Yamamura and Kawasaki (2013) (see Appendix A), allows an easy computation of the PNE outcome of any (p, R)-regression. However, characterizing strategyproof algorithms is a challenging open question for the linear regression setting (Chen et al. 2018). Thus, while Theorem 5 demonstrates that the PNE outcome is also the outcome of a strategyproof algorithm, it does not allow us to derive an analytic expression.

In Section 4.1, we outlined an exponential-time approach that follows immediately from Lemma 3. However, this is impractical unless there are very few agents. Turning elsewhere, a standard approach to computing Nash equilibria is through best-response updates (Ben-Porat and Tennenholtz 2017; 2019; Yamamura and Kawasaki 2013). Specifically, we start from an (arbitrary) profile of reports by the agents, and in each step, allow an agent not already playing her best response, to switch to her best response. If this process terminates, it must do so at a PNE, regardless of initial conditions. For 1D facility location, it is easy to show that this terminates at a PNE in finitely many steps (see Appendix E for details). For linear regression, however, we provide an example in Appendix E in which the process does not terminate in finitely many steps.

Proposition 1. *For the OLS (i.e. $(2, 0)$-regression algorithm), there exists a family of instances in which no best-response path starting from honest reporting terminates in finite steps.*

In this example, although best-response updates do not terminate in finite steps, they do *converge* to a PNE in the limit. We conjecture that this is true in general. In our experiments in Section 5, to find the unique PNE outcome, we used best-response updates, found the outcome they converged to, and verified that it was a PNE (and it always was). We leave further theoretical exploration of the convergence of best-response dynamics for future work.

4.5 Pure Price of Anarchy

Another contrast between 1D facility location and linear regression is the pure price of anarchy. For this analysis, we focus on the $(p, 0)$-regression algorithm. In the 1D case, we showed that the PPoA is $\Theta(n)$ when at least one agent is strategic. While high, this is still bounded. In linear regression, we show that with $n \geq 4$ agents and $m \geq 2$ strategic agents, the PPoA is already unbounded. In other words, strategic behavior can make the overall system boundlessly worse-off. We emphasize, however, that this is a worst-case result; the loss in our experiments (Section 5) is not nearly as pessimistic. The proof of the next result is in Appendix E.

Theorem 6. *In the linear regression setting with $|N| \geq 4$ agents of which $|M| \geq 2$ are strategic, the PPoA of the $(p, 0)$-regression algorithm with $p > 1$ is unbounded.*

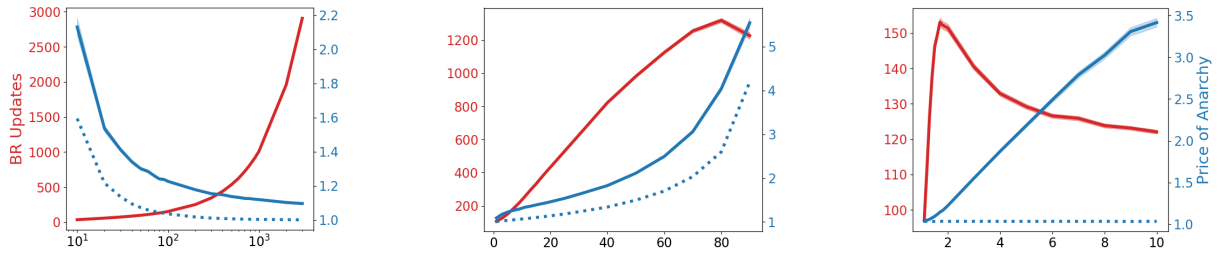
5 Experiments

We conduct experiments with both synthetic data and real data to measure two aspects of strategic manipulation: the number of best-response updates needed to reach a pure Nash Equilibrium (red line) and the average PPoA (with $q = 2$) of a (p, R)-regression algorithm (solid blue line), which we compare against the average PPoA of the strategyproof LAD (i.e. $(1, 0)$-regression) algorithm (dotted blue line). We focus on four key parameters: the number of agents n, the dimension of independent variables d, the norm value p, and the fraction of agents who are strategic, denoted $\alpha = m/n \in [0, 1]$.

Synthetic experiments: In each experiment, we vary one parameter, while using default values for the others. The default values are $n = 100$, $d = 6$, $p = 2$, and $\alpha = 1$. We plot the average results over $1,000$ random instances along with 95% confidence bounds (although they are too narrow to be visible in most plots). The data generation process is as follows. First, we sample $\boldsymbol{\beta^*} \in [-1, 1]^{d+1}$ uniformly at random. Next, we sample each entry in $\boldsymbol{X} \in \mathbb{R}^{d \times n}$ iid from the standard normal distribution and set each $y_i = (\boldsymbol{\beta^*})^T x_i + \epsilon_i$, where ϵ_i is Gaussian noise with zero mean and s.d. 0.5. Finally, we normalize $\boldsymbol{y}$ to lie in $[0, 1]^n$.

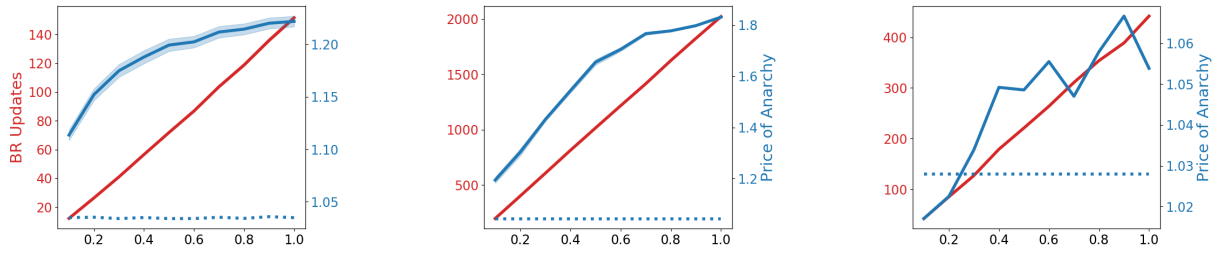
Real experiments: We also conduct experiments with two real-world housing datasets: the California Housing Prices dataset from Kaggle with $n \approx 2000$ and $d = 9$ (Figure 2b) and the real estate valuation dataset from UCI with $n \approx 400$ and $d = 7$ (Figure 2c) (of California 1990; Yeh and Hsu 2018). In these experiments, we also normalize $\boldsymbol{y}$ to lie in $[0, 1]^n$.

Figures 1a, 1b, 1c and 2a show the effect of varying n, d, p, and α, respectively, in our synthetic experiments. With a higher number of agents n, the best-response process takes longer, but the PPoA decreases quickly. The dependence on d is more interesting. For $d < n$, the number of best-response steps and the PPoA increase with d (with a slight decrease in the former and a quicker increase in the latter as d approaches $n = 100$). Of course, when $d = n$, the only PNE is where all agents are perfectly happy, which means the number of best-response steps drop to zero and PPoA drop to 1. Hence, for $d < n$, there is a curse of dimensionality, even though $d = n$ is an ideal scenario.



(a) $n \in [10, 3000]$ (b) $d \in [1, 90]$ (c) $p \in [1.1, 10]$

Figure 1: The effect of varying n, d, and p on synthetic data with 95% confidence intervals



(a) $\alpha \in [0.1, 1]$, synthetic data (b) $\alpha \in [0.1, 1]$, Kaggle dataset (c) $\alpha \in [0.1, 1]$, UCI dataset

Figure 2: The effect of varying α on synthetic and real data. Plots with synthetic data have 95% confidence intervals.

The effect of p is also surprising. With $p \in (1, 2]$, intuitively, one would expect a tradeoff. Mechanisms with p closer to 1 may be less vulnerable to manipulation than the OLS ($p = 2$); indeed, $p = 1$ is known to be strategyproof. But given the equilibrium reports, OLS at least minimizes the MSE, which is the objective underlying our PPoA definition, whereas mechanisms with $p < 2$ optimize a different objective. Given this, we find it surprising that, not only does $p < 2$ result in a lower PPoA than $p = 2$, but PPoA seems to increase monotonically with p (Figure 4 in Appendix F shows that this is also true when PPoA is measured using the q-norm for other values of q). We also note that the strategyproof $(1, 0)$-regression algorithm performs no worse than the PNE of the $(p, 0)$-regression algorithm for any $p > 1$ in terms of MSE. Another observation of note is that the number of best-response updates increases until $p \approx 2$ and then decreases. In our synthetic and real experiments, both the number of best-response updates and the PPoA generally increase with α, which is expected. However, it is worth noting that in Fig. 2b, even as few as 10% of the agents strategizing leads to a 27% increase in the overall MSE, and with all agents strategizing, the MSE doubles. In Fig 2c, the effect of strategizing is more restrained. Surprisingly, in this case, the OLS equilibrium outperforms the $(1, 0)$-regression algorithm for small α.

6 Discussion and Future Work

This work focused on the role of *strategic noise* in linear regression, where data sources manipulate their inputs to minimize their own loss. We established that a popular class of linear regression algorithms — minimizing the ℓ_p loss with a convex regularizer — has a unique pure Nash equilibrium outcome. Our theoretical results show that in the worst case, strategic behavior can cause a significant loss of efficiency, but experiments highlight a less pessimistic average case, which future work can focus on rigorously analyzing.

It is also interesting to ponder the implications of our general result connecting strategyproof algorithms to the unique PNE of non-strategyproof algorithms beyond linear regression. Similar results are known in other domains (Roberts 1979; Dasgupta, Hammond, and Maskin 1979; Laffont and Maskin 1982), including unique equilibria of first-price auctions (Chawla and Hartline 2013). This indicates the possibility of a more general result along these lines.

Lastly, the study of strategic noise in machine learning environments is still in its infancy. We view our work as not only advancing the state-of-the-art, but also as a stepping stone to more realistic analysis. For example, future work can move past assuming that agents have complete information about others' strategies — a common assumption in the literature (Dekel, Fischer, and Procaccia 2010; Ben-Porat and Tennenholtz 2017; 2019) — and consider Bayes-Nash equilibria. Other extensions include studying non-strategyproof algorithms in environments such as classification or generative modeling, and investigating generalization of equilibria (i.e. whether the equilibrium with many agents can be approximated by sampling a few agents).

References

Ben-Porat, O., and Tennenholtz, M. 2017. Best response regression. In *Advances in Neural Information Processing Systems*, 1499–1508.

Ben-Porat, O., and Tennenholtz, M. 2019. Regression equilibrium. In *Proc. of 20th EC*, 173–191.

Bousquet, O.; Luxburg, U. v.; and Gunnar, R. 2004. *Introduction to Statistical Learning Theory*. Springer.

Bshouty, N. H.; Eiron, N.; and Kushilevitz, E. 2002. PAC learning with nasty noise. *Theoretical Computer Science* 288(2):255–275.

Cai, Y.; Daskalakis, C.; and Papadimitriou, C. H. 2015. Optimum statistical estimation with strategic data sources. In *Proc. of 28th COLT*, 280–296.

Caragiannis, I.; Procaccia, A.; and Shah, N. 2016. Truthful univariate estimators. In *International Conference on Machine Learning*, 127–135.

Caro, F.; Gallien, J.; Díaz, M.; García, J.; Corredoira, J. M.; Montes, M.; Ramos, J. A.; and Correa, J. 2010. Zara uses operations research to reengineer its global distribution process. *Interfaces* 40(1):71–84.

Chawla, S., and Hartline, J. D. 2013. Auctions with unique equilibria. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, 181–196. ACM.

Chen, Y.; Podimata, C.; Procaccia, A. D.; and Shah, N. 2018. Strategyproof linear regression in high dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, 9–26. ACM.

Chen, Y.; Caramanis, C.; and Mannor, S. 2013. Robust sparse regression under adversarial corruption. In *International Conference on Machine Learning*, 774–782.

Cummings, R.; Ioannidis, S.; and Ligett, K. 2015. Truthful linear regression. In *Proc. of 28th COLT*, 448–483.

Dasgupta, P.; Hammond, P.; and Maskin, E. 1979. The implementation of social choice rules: Some general results on incentive compatibility. *The Review of Economic Studies* 46(2):185–216.

Dekel, O.; Fischer, F.; and Procaccia, A. D. 2010. Incentive compatible regression learning. *Journal of Computer and System Sciences* 76(8):759–777.

Dong, J.; Roth, A.; Schutzman, Z.; Waggoner, B.; and Wu, Z. S. 2018. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, 55–70.

Frénay, B., and Verleysen, M. 2013. Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems* 25(5):845–869.

Goldman, S. A., and Sloan, R. H. 1995. Can PAC learning algorithms tolerate random attribute noise? *Algorithmica* 14(1):70–84.

Gu, S., and Rigazio, L. 2014. Towards deep neural network architectures robust to adversarial examples. arXiv:1412.5068.

Hardt, M.; Megiddo, N.; Papadimitriou, C. H.; and Wootters, M. 2016. Strategic classification. In *Proc. of 7th ITCS*, 111–122.

Immorlica, N.; Kalai, A. T.; Lucier, B.; Moitra, A.; Postlewaite, A.; and Tennenholtz, M. 2011. Dueling algorithms. In *Proc. of 43rd STOC*, 215–224.

Kearns, M., and Li, M. 1993. Learning in the presence of malicious errors. *SIAM Journal on Computing* 22(4):807–837.

Koutsoupias, E., and Papadimitriou, C. 1999. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, 404–413. Springer.

Laffont, J.-J., and Maskin, E. 1982. Nash and dominant strategy implementation in economic environments. *Journal of Mathematical Economics* 10(1):17–47.

Littlestone, N. 1988. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning* 2:285–318.

Mansour, Y.; Slivkins, A.; and Wu, Z. S. 2017. Competing bandits: Learning under competition. arXiv:1702.08533.

Meir, R.; Procaccia, A. D.; and Rosenschein, J. S. 2012. Algorithms for strategyproof classification. *Artificial Intelligence* 186:123–156.

Moulin, H. 1980. On strategy-proofness and single peakedness. *Public Choice* 35(4):437–455.

Natarajan, N.; Dhillon, I. S.; Ravikumar, P. K.; and Tewari, A. 2013. Learning with noisy labels. In *Advances in neural information processing systems*, 1196–1204.

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V. 2007. *Algorithmic game theory*. Cambridge university press.

of California, S. 1990. California housing prices. data retrieved from Kaggle, https://www.kaggle.com/camnugent/california-housing-prices.

Perote, J., and Perote-Pena, J. 2004. Strategy-proof estimators for simple regression. *Mathematical Social Sciences* 47(2):153–176.

Pugh, C. 2003. *Real Mathematical Analysis*. Undergraduate Texts in Mathematics. Springer New York.

Renault, R., and Trannoy, A. 2005. Protecting minorities through the average voting rule. *Journal of Public Economic Theory* 7(2):169–199.

Renault, R., and Trannoy, A. 2011. Assessing the extent of strategic manipulation: the average vote example. *SERIEs* 2(4):497–513.

Roberts, K. 1979. The characterization of implementable choice rules. *Aggregation and revelation of preferences* 12(2):321–348.

Rockafellar, R. T., and Wets, R. J.-B. 2009. *Variational analysis*, volume 317. Springer Science & Business Media.

Yamamura, H., and Kawasaki, R. 2013. Generalized average rules as stable nash mechanisms to implement generalized median rules. *Social Choice and Welfare* 40(3):815–832.

Yeh, I.-C., and Hsu, T.-K. 2018. Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing* 65:260–271.

Appendix

A 1D Facility Location

A.1 Detailed overview of the 1D case

Much of the literature on facility location has focused on strategyproof algorithms. Moulin (1980) showed that an algorithm f is strategyproof and anonymous[4] if and only if it is a *generalized median* given by $f(y_1, \ldots, y_m) = \text{med}(y_1, \ldots, y_m, \alpha_0, \ldots, \alpha_m)$, where med denotes the median and α_k is a fixed constant (called a *phantom*) for each k. Caragiannis, Procaccia, and Shah (2016) focused on a notion of worst-case statistical efficiency, and provided a characterization of generalized medians which exhibit optimal efficiency. In particular, they showed that the *uniform generalized median* given by $f(y_1, \ldots, y_m) = \text{med}(y_1, \ldots, y_m, 0, 1/m, 2/m, \ldots, 1)$ is has optimal statistical efficiency.

A more recent line of literature has focused on manipulations under non-strategyproof rules. Recall that under a non-strategyproof rule f, each strategic agent $i \in M$ reports a value $\widetilde{y}_i$, which may be different from y_i. For the facility location setting, the (p, R)-regression algorithm described in Section 2 reduces to $f(\widetilde{y}_1, \ldots, \widetilde{y}_m) = \arg\min_y \sum_{i=1}^m |\widetilde{y}_i - y|^p + \sum_{i=m+1}^n |y_i - y|^p + R(y)$. For $p = 1$, this is known to be strategyproof (Chen et al. 2018). When $p > 1$, which is the focus of our work, this rule is not strategyproof, as we observe in Section 4.

In this family, the most natural rule is the *average rule* given by $f(\widetilde{y}_1, \ldots, \widetilde{y}_m) = (1/m)\sum_{i=1}^m \widetilde{y}_i$. This corresponds to $p = 2$ with no honest agents or regularization. For this rule, Renault and Trannoy (2005) showed that there is always a pure Nash equilibrium, and the pure Nash equilibrium outcome is unique. This outcome is given by $\text{med}(y_1, \ldots, y_m, 0, 1/m, \ldots, 1)$, which coincides with the outcome of the uniform generalized median, which is strategyproof.

Generalizing this result, Yamamura and Kawasaki (2013) proved that any algorithm f satisfying four natural axioms has a unique PNE outcome, which is given by the generalized median $\text{med}(y_1, \ldots, y_m, \alpha_0, \ldots, \alpha_m)$, where $\alpha_k = f(0, \ldots, 0, \underbrace{1, \ldots, 1}_{k \text{ times}})$ for each k.

It is easy to show that the 'vanilla' ℓ_p-norm algorithm with no honest agents or regularization satisfies the axioms of Yamamura and Kawasaki (2013). Using the result of Yamamura and Kawasaki (2013), this algorithm has a unique PNE outcome given by the generalized median $\text{med}(y_1, \ldots, y_m, \alpha_0, \ldots, \alpha_m)$, where $\alpha_k = \frac{k^{\frac{1}{p-1}}}{(n-k)^{\frac{1}{p-1}} + k^{\frac{1}{p-1}}}$ for each $k \in \{0, 1, \ldots, m\}$. The derivation is as follows.

It is easy to see that $\alpha_0 = 0$ and $\alpha_n = 1$. For $k \in \{1, \ldots, n-1\}$, α_k is the minimizer $\arg\min_{\overline{y} \in \mathbb{R}} k|1 - \overline{y}|^p + (n-k)|\overline{y}|^p$. Taking the derivative w.r.t. $\overline{y}$, we can see that the optimal solution is given by

$$-k(1-\alpha_k)^{p-1} + (-k)\alpha_k^{p-1} = 0$$
$$\implies \alpha_k = \frac{k^{\frac{1}{p-1}}}{(n-k)^{\frac{1}{p-1}} + k^{\frac{1}{p-1}}} \tag{3}$$

Theorem 2 (Section 3)**.** *Consider facility location with n agents, of which a subset of agents M are strategic. Let f denote the $(p, 0)$-regression algorithm with $p > 1$. When $|M| \geq 1$, $\text{PPoA}(f) = \Theta(n)$.*

Proof. Define $a = \min_i y_i$ and $b = \max_i y_i$. Let $\overline{y}_h = (1/n)\sum_i y_i$, and let $\overline{y}_{ne}$ denote the unique PNE outcome of the algorithm. Note that $\overline{y}_h, \overline{y}_{ne} \in [a, b]$. For $\overline{y}_h$, this holds by definition. To see this for $\overline{y}_{ne}$, WLOG let $\overline{y}_{ne} < a$. Then by Lemma 3, all manipulating agents must be reporting 1, and the honest agents maintain their honest reports in $[a, b]$. However, then ℓ_p loss optimal outcome on this input cannot be $\overline{y}_{ne} < a$ as a would have a lower loss. A symmetric argument holds for $\overline{y}_{ne} > b$. Thus, $\overline{y}_{ne} \in [a, b]$.

Lower bound of $\Omega(n)$: Suppose a strategic agent $j \in M$ has preference with peak at $\alpha_{n-1} = \frac{(n-1)^{\frac{1}{p-1}}}{1+(n-1)^{\frac{1}{p-1}}}$ and the remaining agents have preferences with peak at 1. Note that $a = \alpha_k$ and $b = 1$. We claim that a PNE equilibrium is given by $\widetilde{y}_j = 0$ and $\widetilde{y}_i = 1 \, \forall i \neq j$, regardless of which agents other than j are strategic. By Equation (3), the outcome on this input is $a = \alpha_k$. Now, we have that the MSE in the equilibrium is

$$MSE_{eq} = \sum_i |y_i - \overline{y}_{ne}|^2 = (n-1)(b-a)^2,$$

whereas the optimal MSE under honest reports is

$$\begin{aligned} MSE_h &= \sum_i |y_i - \overline{y}_h|^2 \\ &= \left(b - \frac{(n-1)b + a}{n}\right)^2 (n-1) + \left(\frac{(n-1)b + a}{n} - a\right)^2 \\ &= \left(\frac{b-a}{n}\right)^2 (n-1) + \left(\frac{(n-1)(b-a)}{n}\right)^2 \\ &= \frac{(b-a)^2(n-1) + (n-1)^2(b-a)^2}{n^2} \\ &= \frac{n(n-1)(b-a)^2}{n^2} = \frac{(n-1)(b-a)^2}{n} \end{aligned}$$

Hence, we have that

$$\text{PPoA} = \frac{MSE_{eq}}{MSE_h} = n = \Omega(n)$$

Upper bound: $O(n)$ Since the MSE is a strictly convex function with a minimum at the sample mean $\overline{y}_h$, the maximum allowable value of MSE_{eq} is achieved at one of the end-points a or b. Hence, we have

$$\text{PPoA} = \frac{\sum_i |y_i - \overline{y}_{ne}|^2}{\sum_i |y_i - \overline{y}|^2} \leq \max\left\{\frac{\sum_i |y_i - a|^2}{\sum_i |y_i - \overline{y}|^2}, \frac{\sum_i |y_i - b|^2}{\sum_i |y_i - \overline{y}|^2}\right\}$$

We show that each quantity on the right hand side is $O(n)$. Let us prove this for a. The argument is symmetric for b.

[4] This is a mild condition which requires treating the agents symmetrically.

Note that for each i and each $y \in \mathbb{R}$, we have

$$|y_i - y|^2 + |a - y|^2 \geq |y_i - (y_i + a)/2|^2 + |a - (y_i + a)/2|^2$$
$$= \frac{|y_i - a|^2}{2}$$

Hence, we have that for each i,

$$|y_i - a|^2 \leq 2 \cdot |y_i - \overline{y}|^2 + |a - \overline{y}|^2 \leq 2 \sum_i |y_i - \overline{y}|^2.$$

Summing this over all i, we get

$$\frac{\sum_i |y_i - a|^2}{\sum_i |y_i - \overline{y}|^2} \leq 2n,$$

as desired. □

B Proofs of Key Lemmas

Lemma 4. *Fix strategic agent $i \in M$ and reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents. Let $\widetilde{y}_i^1$ and $\widetilde{y}_i^2$ be two possible reports of agent i, and let $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ be the corresponding optimal regression coefficients, respectively. Then, $\widetilde{y}_i^1 \neq \widetilde{y}_i^2$ implies $\boldsymbol{\beta}^1 \neq \boldsymbol{\beta}^2$.*

Proof. Suppose for contradiction that $\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = \boldsymbol{\beta}^*$. We note that at the optimal regression coefficients, the gradient of our strictly convex loss function must vanish. Let the loss functions on the two instances be given by $\mathcal{L}^1$ and $\mathcal{L}^2$, respectively. So for $k \in \{1, 2\}$,

$$\mathcal{L}^k(\boldsymbol{\beta}) = |\widetilde{y}_i^k - \boldsymbol{x_i}^T \boldsymbol{\beta}|^p + \sum_{j \neq i} |\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}|^p + R(\boldsymbol{\beta}).$$

Since $\boldsymbol{\beta}^*$ is optimal for $\mathcal{L}^1$, we have

$$-p|\widetilde{y}_i^1 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*|^{p-2}(\widetilde{y}_i^1 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*)\boldsymbol{x_i} -$$
$$\sum_{j \neq i} p|\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^*|^{p-2}(\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^*)\boldsymbol{x_j} +$$
$$\nabla R(\boldsymbol{\beta}^*) = 0.$$

Hence,

$$\nabla R(\boldsymbol{\beta}^*) - \sum_{j \neq i} p|\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^*|^{p-2}(\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^*)\boldsymbol{x_j}$$
$$= p|\widetilde{y}_i^1 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*|^{p-2}(\widetilde{y}_i^1 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*)\boldsymbol{x_i}$$
$$\neq p|\widetilde{y}_i^2 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*|^{p-2}(\widetilde{y}_i^2 - \boldsymbol{x_i}^T \boldsymbol{\beta}^*)\boldsymbol{x_i},$$

where the last inequality follows because $\widetilde{y}_i^1 \neq \widetilde{y}_i^2$ and $\boldsymbol{x_i}$ is not the $\mathbf{0}$ vector (its last element is a non-zero constant). Hence, the gradient of $\mathcal{L}^2$ at $\boldsymbol{\beta}^*$ is not zero, which is a contradiction. □

Lemma 5. *For $a_1 \geq a_2$, $b_1 \geq b_2$, and $p \geq 1$, we have*

$$|a_1 - b_1|^p + |a_2 - b_2|^p \leq |a_1 - b_2|^p + |a_2 - b_1|^p.$$

Proof. Note that vector $(a_1 - b_2, a_2 - b_1)$ majorizes the vector $(a_1 - b_1, a_2 - b_2)$. For $p \geq 1$, $f(x) = |x|^p$ is a convex function. Hence, by the Karamata majorization inequality, the result follows. □

Lemma 1 (Section 4)**.** *The outcome $\overline{y}_i$ of agent i is continuous in $\widetilde{\boldsymbol{y}}$, and strictly increasing in the her own report $\widetilde{y}_i$ for any fixed reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents.*

Proof. For **continuity**, we refer to Corollary 7.43 in (Rockafellar and Wets 2009), which states that for a function $F(\widetilde{\boldsymbol{y}}) = \arg\min_{\boldsymbol{\beta}} \mathcal{L}(\widetilde{\boldsymbol{y}}, \boldsymbol{\beta})$, where $\mathcal{L} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ is proper[5], strictly convex, lower semi-continuous, and has $\mathcal{L}^\infty(\mathbf{0}, \beta) > 0 \, \forall \boldsymbol{\beta} \neq 0$,[6] F is single-valued and continuous on its domain.

It is easy to check that our loss function given in Equation (1) satisfies these conditions. Hence, its minimizer $\boldsymbol{\beta}^*$ is continuous in $\widetilde{\boldsymbol{y}}$. Since $\overline{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}^*$, it follows that $\overline{\boldsymbol{y}}$ is also continuous in $\widetilde{\boldsymbol{y}}$.

For **strict monotonicity**, first note that $\overline{y}_i = \boldsymbol{x_i}^T \boldsymbol{\beta}^*$. Now consider two instances of (p, R)-linear regression, u and w, that differ only in agent i's reported response, denoted $\widetilde{y}_i^u$ and $\widetilde{y}_i^w$, respectively, in the two instances. Hence, $\widetilde{y}_i^u \neq \widetilde{y}_i^w$. Let $\boldsymbol{\beta}^u$ and $\boldsymbol{\beta}^w$ be the corresponding optimal regression parameters. WLOG assume $\widetilde{y}_i^u > \widetilde{y}_i^w$, and for contradiction, suppose that $\boldsymbol{x_i}^T \boldsymbol{\beta}^w \geq \boldsymbol{x_i}^T \boldsymbol{\beta}^u$. Using Lemma 4, we get that $\boldsymbol{\beta}^u \neq \boldsymbol{\beta}^w$. Because our strictly convex loss function has a unique minimizer, we have $\mathcal{L}(\widetilde{\boldsymbol{y}}^u, \boldsymbol{\beta}^u) < \mathcal{L}(\widetilde{\boldsymbol{y}}^u, \boldsymbol{\beta}^w)$ and $\mathcal{L}(\widetilde{\boldsymbol{y}}^w, \boldsymbol{\beta}^w) < \mathcal{L}(\widetilde{\boldsymbol{y}}^w, \boldsymbol{\beta}^u)$. Let us define $\mathcal{C}^u = \sum_{j \neq i} |\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^u|^p + R(\boldsymbol{\beta}^u)$ and $\mathcal{C}^w = \sum_{j \neq i} |\widetilde{y}_j - \boldsymbol{x_j}^T \boldsymbol{\beta}^w|^p + R(\boldsymbol{\beta}^w)$, we get

$$|\widetilde{y}_i^u - \boldsymbol{x_i}^T \boldsymbol{\beta}^u|^p + \mathcal{C}^u < |\widetilde{y}_i^u - \boldsymbol{x_i}^T \boldsymbol{\beta}^w|^p + \mathcal{C}^w. \quad (4)$$

$$|\widetilde{y}_i^w - \boldsymbol{x_i}^T \boldsymbol{\beta}^w|^p + \mathcal{C}^w < |\widetilde{y}_i^w - \boldsymbol{x_i}^T \boldsymbol{\beta}^u|^p + \mathcal{C}^u. \quad (5)$$

Adding Equations (5) and (4), we have:

$$|\widetilde{y}_i^u - \boldsymbol{x_i}^T \boldsymbol{\beta}^u|^p + |\widetilde{y}_i^w - \boldsymbol{x_i}^T \boldsymbol{\beta}^w|^p <$$
$$|\widetilde{y}_i^u - \boldsymbol{x_i}^T \boldsymbol{\beta}^w|^p + |\widetilde{y}_i^w - \boldsymbol{x_i}^T \boldsymbol{\beta}^u|^p \quad (6)$$

Note that because we assumed $\widetilde{y}_i^u > \widetilde{y}_i^w$ and $\boldsymbol{x_i}^T \boldsymbol{\beta}^u \leq \boldsymbol{x_i}^T \boldsymbol{\beta}^w$, using Lemma 5, we get

$$|\widetilde{y}_i^u - \boldsymbol{x_i^t} \boldsymbol{\beta}^w|^p + |\widetilde{y}_i^w - \boldsymbol{x_i^t} \boldsymbol{\beta}^u|^p \leq |\widetilde{y}_i^u - \boldsymbol{x_i^t} \boldsymbol{\beta}^u|^p + |\widetilde{y}_i^w - \boldsymbol{x_i^t} \boldsymbol{\beta}^w|^p,$$

which contradicts Equation 6. □

Lemma 2 (Section 4)**.** *For each strategic agent i, the following hold about the best response function br_i.*

1. *The best response is unique, i.e., $|\mathrm{br}_i(\widetilde{\boldsymbol{y}}_{-i})| = 1$ for any reports $\widetilde{\boldsymbol{y}}_{-i}$ of the other agents.*
2. *br_i is a continuous function of $\widetilde{\boldsymbol{y}}_{-i}$.*

Proof of Lemma 2. We first show **uniqueness** of the best response. By Lemma 1, f_i is continious and strictly increasing in $\widetilde{y}_i$. Consider the minimization problem: $\arg\min_{\widetilde{y}_i \in [0,1]} |y_i - f_i(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i})|^p$, where $\widetilde{\boldsymbol{y}}_{-i}$ is constant. So for now, let us consider f_i to be a function of only $\widetilde{y}_i$. Since $\widetilde{y}_i \in [0, 1]$, it achieves a minimum at $a = f_i(0)$ and a maximum at $b = f_i(1)$. If $a \leq b \leq y_i$, then the minimum of the problem is achieved at $\widetilde{y}_i = 1$. Symmetric

[5] A function is proper if the domain on which it is finite is non-empty.

[6] $\mathcal{L}^\infty(\mathbf{0}, \beta)$ is known as the horizon function of $\mathcal{L}$.

case holds for $y_i \leq a \leq b$ where minimum is achieved at $\widetilde{y}_i = 0$. Lastly, if $y_i \in [a, b]$, by intermediate value theorem, $\exists\, \widetilde{y}_i$ s.t $f_i(\widetilde{y}_i) = y_i$, which is then the minimum. In all cases, the minimum is unique since f_i is strictly increasing. We now show that this unique minimum $\widetilde{y}_i^*$ is indeed the unique best response. If $y_i \in [a, b]$ then reporting $\widetilde{y}_i^*$ makes agent i perfectly happy as her outcome matches the peak of her preference, which is clearly best response. If $y_i > b$, then $\widetilde{y}_i^* = 1$ and her outcome is $\overline{y}_i = b$. Under any other report, her outcome would be $\overline{y}_i \leq b$, which cannot be more preferred. A symmetric argument holds for $y_i < a$ case.

Now we can use the uniqueness of the best response to argue its **continuity**. More specifically, we want to show that $br_i(\widetilde{\boldsymbol{y}}_{-i}) = \arg\min_{\widetilde{y}_i \in [0,1]} g(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i})$ is continuous, where $g(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i}) = |y_i - f_i(\widetilde{y}_i, \widetilde{\boldsymbol{y}}_{-i})|^p$ is jointly continious due to the continuity of f_i. We use the sequence definition of continuity. Fix a convergent sequence $\{\widetilde{\boldsymbol{y}}_{-i}^n\} \to \widetilde{\boldsymbol{y}}_{-i}$. Since there is always a unique minimum, the sequence $\{br_i(\widetilde{\boldsymbol{y}}_{-i}^n)\}$ is well-defined. We want to show $\{br_i(\widetilde{\boldsymbol{y}}_{-i}^n)\} \to br_i(\widetilde{\boldsymbol{y}}_{-i})$. By the Bolzano-Weirstrass theorem, every bounded sequence in $\mathbb{R}$ has a convergent sub-sequence. Therefore, this has a convergent sub-sequence $\{br_i(\widetilde{\boldsymbol{y}}_{-i}^{n_k})\}$ that converges to some θ. Let $br_i(\widetilde{\boldsymbol{y}}_{-i}) = \theta^*$. We want to first show $\theta = \theta^*$. By the continuity of g, $\{g(\theta^*, \widetilde{\boldsymbol{y}}_{-i}^{n_k})\} \to g(\theta^*, \widetilde{\boldsymbol{y}}_{-i})$. Also by the minimum, for every individual element of the subsequence n_k, we have that $g(\theta^*, \widetilde{\boldsymbol{y}}_{-i}^{n_k}) \geq g(br_i(\widetilde{\boldsymbol{y}}_{-i}^{n_k}), \widetilde{y}^{n_k})$. Now again by continuity of g, both the above sequences converge and we have: $g(\theta^*, \widetilde{\boldsymbol{y}}_{-i}) \geq g(\theta, \widetilde{\boldsymbol{y}}_{-i}))$. Since θ^* is the unique minimizer for $\widetilde{\boldsymbol{y}}_{-i}$, we have that $\theta = \theta^*$. So, every convergent sub-sequence of $br_i(\widetilde{\boldsymbol{y}}_{-i}^n)$ converges to $br_i(\widetilde{\boldsymbol{y}}_{-i})$. Since this is a bounded sequence, we have that if $\{\widetilde{\boldsymbol{y}}_{-i}^n\} \to \widetilde{\boldsymbol{y}}_{-i}$, then $\{br_i(\widetilde{\boldsymbol{y}}_{-i}^n)\} \to br_i(\widetilde{\boldsymbol{y}}_{-i})$. Thus, br_i is continuous. □

Lemma 3 (Section 4). *$\widetilde{\boldsymbol{y}}_M$ is a pure Nash Equilibrium if and only if $(\overline{y}_i < y_i \wedge \widetilde{y}_i = 1) \vee (\overline{y}_i > y_i \wedge \widetilde{y}_i = 0) \vee (\overline{y}_i = y_i)$ holds for all $i \in M$.*

Proof. For the 'if' direction, we check that in each case, agent $i \in M$ cannot change her report to attain a strictly better outcome. When $\overline{y}_i < y_i$ and $\widetilde{y}_i = 1$, every other report $\widetilde{y}_i' < \widetilde{y}_i = 1$ will result in an outcome $\overline{y'}_i < \overline{y}_i < y_i$ (Lemma 1), which the agent prefers even less. A symmetric argument holds for the $\overline{y}_i > y_i$ and $\widetilde{y}_i = 0$ case. Finally, when $\overline{y}_i = y_i$, the agent is already perfectly happy.

For the 'only if' direction, suppose $\widetilde{\boldsymbol{y}}_M$ is a PNE. Consider agent $i \in M$. The only way the condition is violated is if $\overline{y}_i < y_i$ and $\widetilde{y}_i \neq 1$ or $\overline{y}_i > y_i$ and $\widetilde{y}_i \neq 0$. In the former case, Lemma 1 implies that for a sufficiently small $\epsilon > 0$, agent i increasing her report to $\widetilde{y}_i' = 1 + \epsilon$ must result in an outcome $\overline{y'}_i \in (\overline{y}_i, y_i]$, which the agent strictly prefers over $\overline{y}_i$. This contradicts the assumption that $\widetilde{\boldsymbol{y}}_M$ is a PNE. A symmetric argument holds for the second case. □

C Uniqueness of the PNE outcome

Theorem 4 (Section 4.2). *For $p > 1$ and convex regularizer R, the (p, R)-regression algorithm has a unique pure Nash equilibrium outcome.*

Proof. Assume by contradiction that there are two equilibria $\widetilde{\boldsymbol{y}}^1$ and $\widetilde{\boldsymbol{y}}^2$, which result in distinct outcomes $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^1$, respectively. By Lemma 3, any agent whose preference is strictly above or below both hyperplanes must have the same report in both cases. Similarly, any agent i whose preference is strictly below $\overline{y}_i^1$ and above $\overline{y}_i^2$, had $\widetilde{y}_i^1 = 0$ and $\widetilde{y}_i^2 = 1$. A symmetric case holds for between preferences $\in (\overline{y}_i^2, \overline{y}_i^1)$. Lastly, any agent i whose preference $= \overline{y}_i^2$ but is below $\overline{y}_i^1$ had $\widetilde{y}_i^2 \in [0, 1]$ and $\widetilde{y}_i^1 = 0$. A similar argument holds for the symmetric case. In all such instances, we note that agents change their reports weakly in the opposite direction as their respective projections. If only one agent did this, Lemma 1 shows that it leads to a contradiction. We rely on a similar technique to show that multiple agents doing this also leads to contradictions. Note, the only exception to this are agents $k \in \mathcal{B}$, whose preference lies on both hyperplanes (i.e. at their intersection)

Let $\mathcal{A}$ be the set of points who change their reports weakly in the opposite direction as their projections, $\mathcal{B}$ as defined above, and $\mathcal{S}$, the remaining agents who either don't change or are honest. Recall $\overline{y}_i = \boldsymbol{x_i}^T \boldsymbol{\beta}$. Note by above, $\forall\, i \in \mathcal{A}$:

$$\forall\, i \in \mathcal{A} \quad \widetilde{y}_i^1 \geq \widetilde{y}_i^2 \implies \boldsymbol{x_i}^T\boldsymbol{\beta}^2 \geq \boldsymbol{x_i}^T\boldsymbol{\beta}^1 \quad \bigwedge \quad \widetilde{y}_i^2 \geq \widetilde{y}_i^1 \implies \boldsymbol{x_i}^T\boldsymbol{\beta}^1 \geq \boldsymbol{x_i}^T\boldsymbol{\beta}^2 \tag{7}$$

$$\forall\, k \in \mathcal{B} \quad \boldsymbol{x_k}^T\boldsymbol{\beta}^1 = \boldsymbol{x_k}^T\boldsymbol{\beta}^2 \tag{8}$$

Defining $\mathcal{C}^u = \sum_{j \in \mathcal{S}} |\widetilde{y}_j - \boldsymbol{x_j}^T\boldsymbol{\beta}^1|^p + R(\boldsymbol{\beta}^1)$ and $\mathcal{C}^2 = \sum_{j \in \mathcal{S}} |\widetilde{y}_j - \boldsymbol{x_j}^T\boldsymbol{\beta}^2|^p + R(\boldsymbol{\beta}^2)$ and noting that $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^1$ uniquely minimizes the loss for instances 1 and 2 respectively with $\boldsymbol{\beta}^1 \neq \boldsymbol{\beta}^1$, we have:

$$\sum_{i \in \mathcal{A}} |\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p + \sum_{k \in \mathcal{B}} |\widetilde{y}_k^1 - \boldsymbol{x_k}^T\boldsymbol{\beta}^1|^p + \mathcal{C}^1 < \sum_{i \in \mathcal{A}} |\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p + \sum_{k \in \mathcal{B}} |\widetilde{y}_k^1 - \boldsymbol{x_k}^T\boldsymbol{\beta}^2|^p + \mathcal{C}^2$$

and

$$\sum_{i \in \mathcal{A}} |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p + \sum_{k \in \mathcal{B}} |\widetilde{y}_k^2 - \boldsymbol{x_k}^T\boldsymbol{\beta}^2|^p + \mathcal{C}^2 < \sum_{i \in \mathcal{A}} |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p + \sum_{k \in \mathcal{B}} |\widetilde{y}_k^2 - \boldsymbol{x_k}^T\boldsymbol{\beta}^1|^p + \mathcal{C}^1$$

Adding the above two equations and noting Equation 8, we have:

$$\sum_{i \in \mathcal{A}} \left\{ |\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p \right\} < \sum_{i \in \mathcal{A}} \left\{ |\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p \right\} \tag{9}$$

Due to equation 7, when we apply lemma 5 to each $i \in \mathcal{A}$, where:

$$|\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p \leq |\widetilde{y}_i^1 - \boldsymbol{x_i}^T\beta_1^*|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p \tag{10}$$

Thus adding this up for all i, we have:

$$\sum_{i\in\mathcal{A}}\{|\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p\} \leq \sum_{i\in\mathcal{A}}\{|\widetilde{y}_i^1 - \boldsymbol{x_i}^T\boldsymbol{\beta}^1|^p + |\widetilde{y}_i^2 - \boldsymbol{x_i}^T\boldsymbol{\beta}^2|^p\} \tag{11}$$

which contradicts equation 9. □

D Interesting Examples

D.1 Finite game leading to different equilibria

While Lemma 3 suggests that we may be able to treat our problem as a finite game with each agent i only allowed to report $\widetilde{y}_i = 0$, $\widetilde{y}_i = 1$, or $\widetilde{y}_i = y_i$ (her true response variable), we show this is erroneous as it can lead to a game with different equilibria. We use a 1D facility location example — recall that this is a special case of linear regression — to illustrate this point.

Consider two agents 1 and 2, whose true points are $y_1 = 0.4$ and $y_2 = 0.5$, respectively. Suppose each agent i's preferences are such that she prefers an outcome $\overline{y}$ strictly more when it has lower $|\overline{y} - y_i|$. If the agents must report values in the range $[0, 1]$, then it is easy to check that the unique PNE of the game is agent 1 reporting $\widetilde{y}_1 = 0$ and agent 2 reporting $\widetilde{y}_2 = 1$, and the PNE outcome is $\overline{y} = 0.5$.

Now, consider the version with finite strategy spaces, where each agent i must report $\widetilde{y}_i \in \{0, 1, y_i\}$. Suppose the agents report honestly, i.e., $\widetilde{\boldsymbol{y}} = \boldsymbol{y} = (0.4, 0.5)$. Then, the outcome is $\overline{y} = 0.45$. The only way agent 1 could possibly improve is by reporting 0, but in that case the outcome would switch to $\overline{y} = 0.25$, increasing $|\overline{y} - y_1|$. A similar argument holds for agent 2. Hence, one can check that honest reporting is a PNE of the finite game, but not of the original game.

D.2 Multiple PNE Outcomes in General Linear Mappings

Consider a linear mapping with two agents given by the following matrix $\boldsymbol{H}$.

$$H = \begin{bmatrix} 0.8 & -1 \\ -1.2 & 1 \end{bmatrix}$$

Recall that the inputs and the outputs are related by the equation $\boldsymbol{H}\widetilde{\boldsymbol{y}} = \overline{\boldsymbol{y}}$. Suppose the agents' preferred values are given by $\boldsymbol{y} = [0\ 0]^T$. Then, when they report $\widetilde{\boldsymbol{y}} = (0, 0)$, the outcome is $(0, 0)$. This is clearly a PNE as both agents are perfectly happy. When they report $\widetilde{\boldsymbol{y}} = (1, 1)$, the outcome is $\overline{\boldsymbol{y}} = (-0.2, -0.2)$. While neither agent is perfectly happy as the outcome is below their preferred value, neither can increase their outcome because they are already reporting 1. Hence, this is also a PNE with a different outcome. It is easy to see that this is not a pathological case. Small perturbations in $\boldsymbol{H}$ will still result in this phenomenon.

E PPoA and Best Response

E.1 Best response converges in finite iterations for 1d

We give an informal argument that under the average rule in 1D, starting from any reports, there is always a best response path that terminates at a PNE in finitely many iterations. For n agents (of which m are strategic), to move the mean by an amount Δ, an agent has to move their report by an amount $n\Delta$. Now fix an initial set of reports. Consider only the 2 strategic agents with the lowest and the highest preferred values, say these are y_1 and y_m, respectively. Consider best response updates by only one of these two agents. If initially $\overline{y} \notin [y_1, y_m]$, both agents increase their reports until $\overline{y} \in [y_1, y_m]$. The only case where this does not happen is if both agents become saturated by reporting 1. If they do bring $\overline{y} \in [y_1, y_m]$, then after each move of agent 1: (a) she is perfectly happy, causing the agent m to move up by $n(y_m - y_1)$ or become saturated at $\widetilde{y}_m = 1$, or (b) she goes to 0 and becomes saturated. Hence, in each iteration, either one agent moves (in a constant direction) by at least $n(y_m - y_1)$, or one agent becomes saturated. Hence, in finitely many steps, either agent 1 is saturated at 0 with $\overline{y} \geq y_1$ or agent m is saturated at 1 with $\overline{y} \leq y_m$. It is easy to see that this agent will never move again. We can now ignore the saturated agent, and repeat the process with the remaining $m - 1$ strategic agents. Using this approach inductively, it follows that an equilibrium will be reached in finitely many iterations.

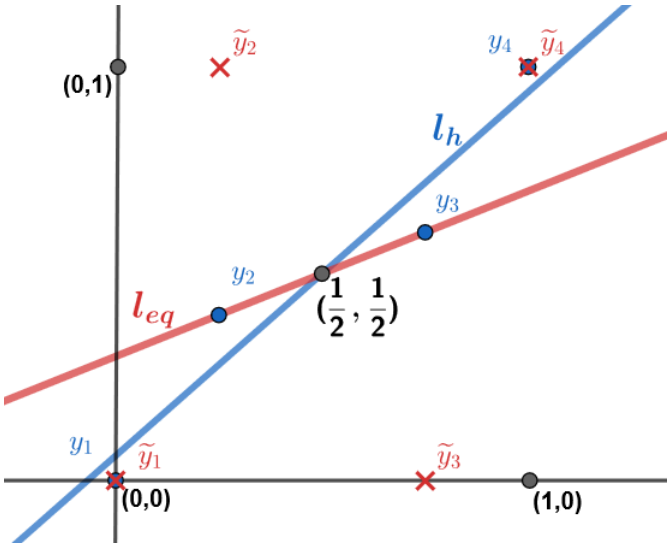


Figure 3: Diagram for Theorem 6 and Proposition 1 with $p = 2$ and $R = 0$. Blue denotes the honest points and the corresponding line, and red denotes the points at a pure nash equilibrium and the corresponding equilibrium line.

E.2 Best Response and PPoA for Linear Regression

Proposition 1 (Section 4.4)**.** *For OLS (i.e. $(2, 0)$-regression algorithm), there exists a family of instances in which no best-response path starting from honest reporting terminates in finite steps.*

Proof. Consider the 4 agent setting (also used in Theorem 6) illustrated in Figure 3. That is, let the preferred values be: $(0, 0), (\frac{1-\epsilon}{2}, y_2), (\frac{1+\epsilon}{2}, y_3), (1, 1)$, where y_2 and y_3 are such that when $\widetilde{\boldsymbol{y}} = [0, 1, 0, 1]$, the corresponding projections are:

$\overline{y}_2 = y_2$ and $\overline{y}_3 = y_3$. Thus, $\widetilde{\boldsymbol{y}} = [0, 1, 0, 1]$ is an equilibrium strategy. Let agents 2 and 3 be strategic.[7] Since $p = 2$, we have a linear mapping characterized by $\boldsymbol{H}\widetilde{\boldsymbol{y}} = \overline{\boldsymbol{y}}$. H_{ij} reflects the effect $\widetilde{y}_i$ has on $\overline{y}_j$, and $\boldsymbol{H}$ is symmetric. By strong monotonicity (Lemma 1), H_{ii} is always positive. It is easy to compute that $H_{23} = H_{32} = \frac{1-\epsilon^2}{4(1+\epsilon^2)} > 0$. Let the agents initially start by reporting honestly, and as such $\overline{y}_2 < y_2 = \widetilde{y}_2$ and $\overline{y}_3 > y_3 = \widetilde{y}_3$.

Since there are only 2 strategic agents, they take turns playing best response alternatively. Consider a round in which agent 2 plays best response, and at the start of the round, the following hold: (1) $\widetilde{y}_2 \geq y_2$, $\widetilde{y}_3 \leq y_3$, and (2) $\overline{y}_2 \leq y_2$ and $\overline{y}_3 \geq y_3$. Since agent 2 is playing best response, she is not perfectly happy. Hence, $\overline{y}_2 < y_2$. Thus, by Lemma 1, agent 2 must increase $\widetilde{y}_2$ by some $a > 0$. Since $H_{23} > 0$, this maintains $\overline{y}_3 > y_3$. Similarly, when agent 3 plays a best response, it maintains $\overline{y}_2 < y_2$. Since the initial conditions (honest reporting) satisfy (1) and (2), they will always be satisfied. That is, player 2 will always report less than 1 and have $\overline{y}_2 < y_2$, and player 3 will always report greater than 0 and have $\overline{y}_3 > y_3$. Thus, the PNE will never be reached in finitely many steps.

To see this formally, consider a stage satisfying (1) and (2) wherein the best response of agent 2 is $\widetilde{y}_2 = 1$ and $\widetilde{y}_3 \neq 0$. Since this is a best response, $\overline{y}_2 \leq y_2$ (¡ in case she isn't perfectly happy) and thus $\overline{y_3} > y_3$. If agent 3 now under-reports and plays $\widetilde{y}_3 = 0$, then since $H_{32} > 0$, $\overline{y}_2 < y_2$. However, we now have $\widetilde{\boldsymbol{y}} = (0,\ 1,\ 0,\ 1)$ where we know to have the outcome: $\overline{y}_2 = y_2$ and $\overline{y}_3 = y_3$. Since the regression outcome is unique, this is a contradiction. A similar situation hold for agent 3. Thus if $\widetilde{y_3} \neq 0$, best response of agent 2, $\neq 1$ and if $\widetilde{y}_2 \neq 1$, the best response of agent 3, $\neq 0$. Since these conditions hold initially, they hold in all rounds.

Thus starting from honest values, agent 2 always over-reports and 3 under-reports and the outcome is never the unique equilibrium outcome. Moreover, at no round does agent 2 or 3 ever reach their equilibrium strategy. Thus at this initial value, no possible best response sequence will terminate in finite iterations. □

Theorem 6 (Section 4.5)**.** *In the linear regression setting with $|N| \geq 4$ agents of which $|M| \geq 2$ are strategic, the PPoA of the $(p, 0)$-regression algorithm with $p > 1$ is unbounded.*

Proof of Theorem 6. We will be using $\overline{y}_i^p$ to denote the projection of the $(p, 0)$-regression equilibrium plane at some x_i and $\overline{\boldsymbol{y}}^p$ for the vector of all projections. $\overline{\mathcal{Y}}_i$ denotes the projection of the $(2, 0)$-regression line using the honest points and $\overline{\boldsymbol{\mathcal{Y}}}$ for the vector of all such projections. Thus, $\text{PPoA} = MSE_{eq}/MSE_h$, where $MSE_{eq} = \sum_i (y_i - \overline{y}_i^p)^2$ and $MSE_h = \sum_i (y_i - \overline{\mathcal{Y}}_i)^2$.

Consider the example in Figure 3. There are four agents with reported values $(0, 0), (\frac{1-\epsilon}{2}, 1), (\frac{1+\epsilon}{2}, 0), (1, 1)$. That is, $\widetilde{\boldsymbol{y}} = (0, \frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}, 1)$. Let the $(p, 0)$-regression line for these points pass through $(0, \overline{y}_1^p), (\frac{1-\epsilon}{2}, \overline{y}_2^p), (\frac{1+\epsilon}{2}, \overline{y}_3^p), (1, \overline{y}_4^p)$. By the symmetry of the problem this line must also pass through $(\frac{1}{2}, \frac{1}{2})$. For $p = 1$, we have that $\overline{\boldsymbol{y}}^1 = [0, \frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}, 1]$. Note that the residuals for points 2 and 3 are higher than for points 1 and 4, and observe that for $p > 1$, the $(p, 0)$-linear regression algorithm progressively tries to minimize the larger residuals. One can check that for $p > 1$, $\overline{y}_2^p = \overline{y}_2^1 + a = \frac{1-\epsilon}{2} + a$ and $\overline{y}_3^p = \overline{y}_3^1 - a = \frac{1+\epsilon}{2} - a$ for some $a > 0$. Since all ℓ_p-regression lines pass through $(\frac{1}{2}, \frac{1}{2})$, by similar triangles we have that for $p > 1$, $\overline{y}_1^p = \overline{y}_1^1 + \frac{a}{\epsilon} = \frac{a}{\epsilon}$. Now if the preferred values of the 4 agents are: $y^p = [0, \overline{y}_2^p, \overline{y}_3^p, 1]$, the reported values above are a pure Nash Equilibrium, and the projection values are unique (by Theorem 4). Note this is regardless of whether agents 1 and 4 are strategic or honest. As such, we have $MSE_{eq} = 2\left(\frac{a}{\epsilon}\right)^2$.

For MSE_h, note that the hat matrix for $(2, 0)$-regression depends only on $\boldsymbol{X}$, and has the form $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ and $\overline{\boldsymbol{\mathcal{Y}}} = \boldsymbol{H}\boldsymbol{y}$. The symmetry of the honest points for any p means that the $(2, 0)$-regression line always passes through $(\frac{1}{2}, \frac{1}{2})$ as well. For $p = 1$, the honest points are co-linear, meaning the $(2, 0)$-regression line of these points have 0 residual for all points (in fact, it's the same as the equilibrium line). For $p > 1$, as we mentioned above, honest points 2 and 3 adjust by some a and we have $y^p = [0, \overline{y}_2^1 + a, \overline{y}_3^1 - a, 1]^T$. We now consider the affect of these two changed honest points on the residual at x_1 and x_2. That is, we consider $r_1^h = |y_1 - \overline{\mathcal{Y}}_1|$ and $r_2^h = |y_2 - \overline{\mathcal{Y}}_2|$ respectively - noting a symmetric case exists for r_3^h and r_4^h. First, we have the following values for the matrix H, which we computed using Wolfram Mathematica.

$$\begin{aligned} \boldsymbol{H}_{21} &= \frac{(1+\epsilon)^2}{4(1+\epsilon^2)} \quad \boldsymbol{H}_{31} = \frac{(\epsilon-1)^2}{4(1+\epsilon^2)} \\ \boldsymbol{H}_{22} &= \frac{3\epsilon^2+1}{4(1+\epsilon^2)} \quad \boldsymbol{H}_{32} = \frac{1-\epsilon^2}{4(1+\epsilon^2)} \end{aligned} \tag{12}$$

$r_1^h = r_2^h = 0$ when $p = 1$, and only y_2 and y_3 have changed (by $+a$ and $-a$ respectively) for $p \neq 1$. Recall, $y_1 = 0$ and $y_2 = \overline{y}_2^p = \overline{y}_2^1 + a$, and $\overline{y}_2^1 = \overline{\mathcal{Y}}_2$. Denote the i^{th} row of H by $\boldsymbol{h_i}$. Then we have:

$$r_1^h = \boldsymbol{h_1} \cdot \begin{bmatrix} 0 \\ \overline{y}_2^1 + a \\ \overline{y}_3^1 - a \\ 1 \end{bmatrix} - 0 = \boldsymbol{h_1} \cdot \begin{bmatrix} 0 \\ \overline{y}_2^1 \\ \overline{y}_3^1 \\ 1 \end{bmatrix} + \boldsymbol{h_1} \cdot \begin{bmatrix} 0 \\ a \\ -a \\ 0 \end{bmatrix} = \boldsymbol{h_1} \cdot \begin{bmatrix} 0 \\ a \\ -a \\ 0 \end{bmatrix}$$

$$\therefore r_1^h = a\frac{(1+\epsilon)^2}{4(1+\epsilon^2)} - a\frac{(\epsilon-1)^2}{4(1+\epsilon^2)} = \frac{a\epsilon}{(1+\epsilon^2)}$$

Similarly, for r_2^h, we have that:

$$r_2^h = \left(\frac{1-\epsilon}{2} + a\right) - \boldsymbol{h_2} \cdot \begin{bmatrix} 0 \\ \overline{y}_2^1 + a \\ \overline{y}_3^1 - a \\ 1 \end{bmatrix}$$

$$= \left(\frac{1-\epsilon}{2} + a\right) - \left(\frac{1-\epsilon}{2} + \boldsymbol{h_2} \cdot \begin{bmatrix} 0 \\ a \\ -a \\ 0 \end{bmatrix}\right)$$

[7] Whether agents 1 and 4 are strategic or honest does not matter in this example.

$$\therefore r_2 = a - a\frac{3\epsilon^2+1}{4(1+\epsilon^2)} + a\frac{1-\epsilon^2}{4(1+\epsilon^2)} = \frac{a}{1+\epsilon^2}$$

By symmetry, $r_1 = r_4$ and $r_2 = r_3$. Now, we can express the price of anarchy this $(p, 0)$-regression equilibrium as:

$$\text{PPoA} = \frac{2\left(\frac{a}{\epsilon}\right)^2}{2\left[\left(\frac{a\epsilon}{(1+\epsilon^2)}\right)^2 + \left(\frac{a}{1+\epsilon^2}\right)^2\right]} = \frac{\frac{1}{\epsilon^2}}{\frac{1}{1+\epsilon^2}} = 1 + \frac{1}{\epsilon^2}$$

As $\epsilon \to 0$, the PPoA becomes unbounded. □

F Experiment - PPoA with different q

In the main text, we consider PPoA measured with respect to mean squared error ($q = 2$), which is the squared ℓ_2 norm of residuals. In this section, we experimentally evaluate PPoA measured with respect to other values of q, as defined below:

$$\text{PPoA}_q(f) = \max_{\boldsymbol{y}\in[0,1]^n} \frac{\max_{\overline{\boldsymbol{y}}\in \text{NE}_f(\boldsymbol{y})} \sum_{i=1}^n |y_i - \overline{y}_i|^q}{\sum_{i=1}^n |y_i - \overline{y}_i^{q\text{-opt}}|^q},$$

where $\overline{y}_i^{q\text{-opt}}$ is the outcome of the mechanism minimizing ℓ_q norm of residuals with honest reports.

Figure 4 shows PPoA_q for different ℓ_p regression algorithms. Once again, we notice the same pattern for each value of q as we did in Figure 1c for $q = 2$: the PPoA increases monotonically with p.

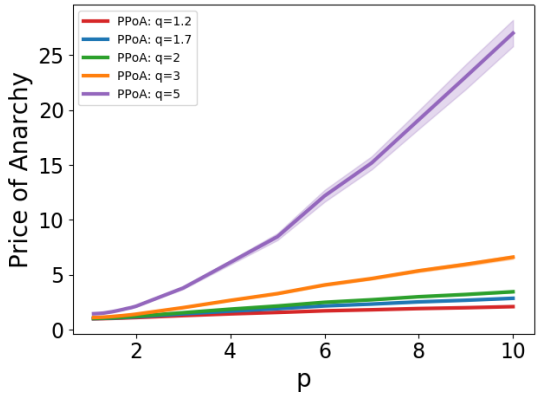


Figure 4: Varying p between 1.1 and 10 and graphing the PPoA using different values of q. The same defaults are used as in other synthetic experiments ($n = 100$, $d = 6$, $\alpha = 1$) and the average of 1000 random instances are plotted with 95% confidence intervals (though too narrow to be visible on some curves).