TIGHT EFFICIENCY GUARANTEES FOR STRATEGYPROOF LINEAR REGRESSION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the trade-off between squared-error accuracy and incentive compatibility in linear regression. Agents report private labels associated with publicly known features and prefer predictions close to their true labels. Ordinary least squares (OLS) need not elicit truthful reports.

For regression with d parameters, we design a deterministic group-strategyproof mechanism achieving a $(d+1)$-approximation to the least-squares optimum and prove optimality even among universally strategyproof randomized mechanisms, answering an open question of Chen et al. (2018). Relaxing universal strategyproofness to strategyproofness in expectation reveals a sharp separation: squared individual loss retains the factor $d+1$, while absolute individual loss admits the tight ratio $2 - 1/(\lceil d/2 \rceil + 1)$.

1 INTRODUCTION

Machine learning usually starts with a dataset and asks how to build an accurate predictor from it. But when a deployed model informs decisions that affect the people supplying its data, the learning rule can also influence what they report (Perdomo et al., 2020; Kleine Buening et al., 2025). In such settings, where people may have misaligned preferences over the predictors, we must ask not only how accurately it predicts, but also whether people benefit from manipulating its inputs.

One prominent example is linear regression, a widely used model in many deployed prediction systems such as electricity-demand forecasting (Hong et al., 2014; PJM, 2025), residential property valuation (MPAC, 2016), and medical-cost prediction (CMS, 2024). When predictions guide resource allocation, the people supplying information may have a reason to distort it. For example, Zara's demand forecasts incorporated store managers' shipment requests, which managers frequently inflated when popular items were scarce (Caro et al., 2010).

Rather than taking honest inputs for granted, strategyproof mechanism design asks how a learning rule can make truthful reporting a best response. In *strategyproof linear regression*, features are public, each agent reports one private response, and the output is a linear predictor. Each agent wants the prediction at their features to be close to their true response. A strategyproof rule makes honesty optimal whatever others report, without payments, label verification, a prior over agents' information, or equilibrium coordination. In a laboratory study of regression, Perote et al. (2015) find more misreporting under OLS than under a strategyproof resistant line estimator.

Truthfulness alone does not ensure a useful predictor: a rule that ignores every report is strategyproof. We also want low squared error, the standard regression objective (Hastie et al., 2009; Gneiting, 2011) and a social objective in mechanism design (Feldman & Wilf, 2013). Even fitting a constant without public features illustrates the tension. Least squares returns the mean, which an agent can move toward their true value by misreporting. A suitably tie-broken median prevents manipulation but can increase squared error. The approximation ratio against the least-squares fit to the true labels therefore measures the *price of strategyproofness*.

For linear regression over general feature spaces, this price has remained poorly understood. Prior work establishes truthful rules, including absolute-loss minimization and generalized resistant hyperplane mechanisms (Dekel et al., 2010; Chen et al., 2018), but leaves a substantial gap in their squared-error guarantees: the known deterministic upper bound is n, the number of agents, while

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the general lower bound is only 2 (Chen et al., 2018). It is natural to ask for the best accuracy guarantee achievable under incentive compatibility.

Randomization offers a potential way to further reduce this price. With a carefully designed randomized mechanism, a beneficial deviation under some realizations may hurt the agent under other realizations, leaving truthful reporting optimal in expectation. This flexibility has yielded better guarantees in settings such as facility location (Feldman & Wilf, 2013), but its role in strategyproof linear regression has not been studied. Balancing accuracy and incentives is more delicate here: truthfulness in expectation depends on the precise form of agents' losses, not just their preference for smaller individual errors. We therefore ask whether randomization can lower the price of strategyproofness, and how the answer depends on agents' preferences.

Our results. We resolve these questions and show that the relevant price is governed by the dimension. Throughout the paper, *d is the number of free coefficients*, including an intercept if one is fitted. Our social objective is always squared error; an agent's individual loss is the p-th power of their absolute error, for some $p > 0$ common to every agent. Table 1 summarizes the results.

First, we give a deterministic $(d+1)$-approximation (Theorem 3.1) by minimizing a weighted sum of absolute residuals. The weights depend only on public features and capture how strongly the model can move each prediction. This preserves truthfulness while replacing sample-size dependence by dimension dependence. For closed convex hypothesis sets, the guarantee becomes $k + 1$, where $k \leq d$ is the affine dimension of the hypothesis set. This approximation ratio is tight, even among universally strategyproof randomized mechanisms.[1]

Next, we consider strategyproof-in-expectation (SPIE) mechanisms for the complete hypothesis set. Unlike deterministic truthfulness, truthfulness in expectation depends on agents' loss. The most natural choices are arguably absolute error ($p = 1$) and squared error ($p = 2$).

For $p = 1$, randomization gives a substantially better guarantee (Theorem 4.3). We sample a geometrically diverse set of agents, fit their reports exactly, and minimize squared error over all observations using the remaining freedom. Selecting about half as many agents as there are coefficients balances two effects: a selected agent cannot improve on an exact fit, while an unselected agent has limited influence. The resulting ratio is $2-1/(\lceil d/2 \rceil+1)$, and a matching lower bound shows its optimality.

In contrast, for $p = 2$, individual and social losses are both squared error, and the factor $d + 1$ cannot be improved (Theorem 4.5). Randomization therefore offers no improvement in the optimal worst-case guarantee.

Finally, the benefit of randomization is not confined to absolute loss. For every positive exponent $p \neq 2$, we give a finite lottery over combinations of independently sampled interpolating models that strictly beats $d + 1$ (Theorem 4.6). Averaging improves accuracy, while additional outcomes protect truthful reporting.

Our results give a relatively complete picture of the accuracy–incentive trade-off in linear regression. The underlying concern extends across machine learning whenever the people supplying data or feedback have a stake in the model they help train (Ouyang et al., 2022; Chen et al., 2023; Wang et al., 2024; Kleine Buening et al., 2025). We hope this work inspires broader research on trade-offs between statistical accuracy and incentive compatibility in learning systems.

1.1 Related work

Strategic regression and learning. Perote & Perote-Peña (2004) study strategyproof estimators for simple regression. Dekel et al. (2010) develop a learning framework for strategic data and establish truthfulness of absolute-loss minimization in the one-point-per-agent setting. Chen et al. (2018) extend this approach to weighted objectives and convex regularization, develop high-dimensional mechanisms and characterizations, and pose the squared-error efficiency question. In a complementary direction, Hossain & Shah (2021) analyze equilibria of non-strategyproof regression rules. Other work considers strategic behavior in classification models (Meir et al., 2012; Hardt et al.,

[1] A randomized mechanism is universally strategyproof (USP) if it is a distribution over deterministic strategyproof mechanisms, and strategyproof in expectation (SPIE) if no agent can reduce their *expected* loss by misreporting. See Section 2.

Table 1: Worst-case guarantees with d coefficients. p refers to the power in the individual loss. Deterministic and USP guarantees do not depend on p in the individual loss function. The SPIE upper bounds for $p \neq 2$ assume the complete hypothesis set. All but the last row are tight.

Incentive requirement	Individual loss	Approximation ratio
Deterministic / universally SP	–	$d+1$
SP in expectation	$p=1$	$2-1/(\lceil d/2 \rceil + 1)$
	$p=2$	$d+1$
	$p>0,\ p\neq 2$	$< d+1$

2016) or learning with external predictions (Balkanski & Zhu, 2025). Payment-based models address different incentives, including privacy costs (Cummings et al., 2015; Zhu et al., 2024) and costly data acquisition (Cai et al., 2015).

Facility location and estimation. Fitting a constant without public features is equivalent to single-facility location on the line.[2] Generalized medians connect its incentive properties to classical social choice (Moulin, 1980), and approximate mechanism design without money studies the accompanying optimization tradeoffs (Procaccia & Tennenholtz, 2013). For the sum of squared distances and expected-distance preferences, Feldman & Wilf (2013) give an optimal randomized ratio of $3/2$ on the line, compared with the deterministic optimum of 2.

Accuracy under additional requirements. Beyond incentive compatibility, researchers study trade-offs between predictive accuracy and other desirable properties, such as fairness (Menon & Williamson, 2018), differential privacy (Chaudhuri et al., 2011), and adversarial robustness (Tsipras et al., 2019). Related performance trade-offs arise for interpretable clustering (Dasgupta et al., 2020) and explainable information design (Chen et al., 2026).

2 Model and Preliminaries

Data and dimension. There are n agents and $d \geq 1$ coefficients. Agent i has a public row vector $x_i^\top$ and a private label $y_i \in \mathbb{R}$, and controls only the report of that label. Let $X \in \mathbb{R}^{n\times d}$ have rows $x_i^\top$, and let $y \in \mathbb{R}^n$ be the label vector. We assume X has full column rank, so $n \geq d$. A coefficient vector $\beta \in \Theta \subseteq \mathbb{R}^d$ produces the predictions $X\beta$, where Θ is a public nonempty closed convex set fixed independently of reports. Unless a constraint is specified, $\Theta = \mathbb{R}^d$.

Our dimension convention counts *free scalar coefficients*, not the dimension of an augmented data point. For example, $f(u) = a^\top u + b$ on $u \in \mathbb{R}^{d-1}$ is a d-parameter model, represented by $x = (u, 1)$ and $\beta = (a, b)$. We use e_i for the i-th standard basis vector and $\mathbf{1}$ for the all-ones vector.

Social objective. For $z = X\beta$, write $C(z; y) = \|z - y\|_2^2$ and $\mathrm{OPT}_\Theta(y) = \min_{\beta\in\Theta} C(X\beta; y)$. A mechanism maps reports to feasible coefficients, possibly at random; $Z(y)$ denotes its random prediction vector. It is *c-efficient*, or a *c-approximation*, if

$$\mathbb{E}C(Z(y); y) \leq c\ \mathrm{OPT}_\Theta(y) \qquad \text{for every } y \in \mathbb{R}^n.$$

The inequality also applies when the optimum is zero: any finite approximation must then return the exact prediction vector almost surely.

Individual incentives. Each agent's preference over the mechanism's prediction for them is single-peaked at their true value. Here, each agent evaluates a prediction using $\ell_p(z_i, y_i) = |z_i - y_i|^p$, for a fixed finite exponent $p > 0$ shared by all agents. The cases $p = 1$ and $p = 2$ are absolute and squared individual loss, respectively. A randomized mechanism is *strategyproof in expectation* (SPIE) for loss ℓ_p if, for every agent i, every label vector y, and every alternative report $t \in \mathbb{R}$,

$$\mathbb{E}|Z_i(y) - y_i|^p \leq \mathbb{E}|Z_i(t, y_{-i}) - y_i|^p,$$

[2]The true value of an individual corresponds to their location and the fitted value corresponds to the selected location of the facility.

where (t, y_{-i}) replaces only coordinate i of y by t.

A deterministic mechanism is *strategyproof* if the same inequality holds without expectations. It is *group-strategyproof* if no coalition can change its reports so that every member weakly reduces their loss and at least one member strictly reduces it. A randomized mechanism is *universally strategyproof* if it is a report-independent distribution over deterministic strategyproof mechanisms.

All positive powers of absolute error induce the same preferences over deterministic outcomes. Consequently, deterministic and universal strategyproofness do not depend on p.

3 An optimal deterministic mechanism

The deterministic mechanism keeps the incentive protection of absolute-loss minimization, but changes the relative weight of observations. A naive ℓ_1 risk minimizer with equal weights compares an ℓ_1 objective with an ℓ_2 objective in the entire n-dimensional observation space. Our weights instead recognize that feasible predictions move in at most d independent directions.

Geometry and weights. We first introduce the notation. Let $K = X\Theta$ be the feasible prediction set. To describe how predictions can change, define its direction space $L = \operatorname{span}(K - K)$, where $K - K = \{z - z' : z, z' \in K\}$. Thus L contains all linear combinations of feasible prediction differences and ignores any fixed offset. Set $k = \dim L$; because X has full column rank, k is also the affine dimension of Θ.

Let P be the orthogonal projector onto L: for every $v \in \mathbb{R}^n$, Pv is the unique vector in L closest to v in Euclidean distance. To compute it explicitly when $k > 0$, choose $A \in \mathbb{R}^{d\times k}$ whose columns form a basis of $\operatorname{span}(\Theta - \Theta)$. Then $L = \operatorname{col}(XA)$ and

$$P = XA(A^\top X^\top XA)^{-1}A^\top X^\top.$$

The diagonal entries $\tau_i = P_{ii}$ are the *leverage scores* of L; they will determine how much weight each observation receives.

Mechanism: leverage-weighted ℓ_1-ERM. Choose

$$\widehat{\beta}(y) \in \arg\min_{\beta\in\Theta} \sum_{i=1}^{n} \sqrt{\tau_i}\,|y_i - x_i^\top \beta|, \quad \text{breaking ties by minimizing } \|\beta\|_2^2. \tag{1}$$

This is a weighted ℓ_1 risk minimizer, with weights determined by the leverage scores. Intuitively, an observation should receive more weight when feasible models can change its prediction more strongly. Precisely, for $u \in L$, Cauchy–Schwarz gives $|u_i| \le \sqrt{\tau_i}\,\|u\|_2$. This motivates the square root in (1). The leverage scores also satisfy $\sum_i \tau_i = k$, so their total contribution is controlled by dimension. As an example, for constant fitting, every leverage score is $1/n$; the mechanism reduces to a median, with the specified fixed tie-break, and the bound becomes 2.

Theorem 3.1 (Leverage-weighted absolute deviations)**.** *The mechanism in (1) is well defined and group-strategyproof. Its squared-error approximation ratio is at most $k + 1$, where $k \le d$ is the affine dimension of Θ.*

Proof. Absolute-loss minimization with public nonnegative, report-independent weights and a fixed strictly convex tie-break is group-strategyproof over a convex prediction class (Chen et al., 2018, Theorem 1). It remains to prove the approximation guarantee.

Let $z^\star$ be the least-squares minimizer in K, let $\widehat{z} = X\widehat{\beta}$, and set $r^\star = y - z^\star$ and $\widehat{r} = y - \widehat{z}$. Since $P\widehat{r}$ and $(I-P)\widehat{r}$ are orthogonal, we split $\|\widehat{r}\|_2^2 = \|P\widehat{r}\|_2^2 + \|(I-P)\widehat{r}\|_2^2$ and bound them separately.

Both predictions lie in K, so their difference lies in L and $(I - P)\widehat{r} = (I - P)r^\star$. In words, every feasible model has the same residual component perpendicular to the direction space; only the parallel component can change. Thus,

$$\|(I - P)\widehat{r}\|_2^2 = \|(I - P)r^\star\|_2^2 \le \|r^\star\|_2^2.$$

It remains to bound $\|P\widehat{r}\|_2^2$. Define the weighted absolute residual $W(a) = \sum_{i=1}^n \sqrt{\tau_i}|a_i|$. We first show that $W(a)$ bounds the length of the projection of a onto L. Since P is an orthogonal projector, it satisfies $P^\top = P$ and $P^2 = P$. Consequently,

$$\|Pe_i\|_2^2 = e_i^\top P^\top P e_i = e_i^\top P e_i = P_{ii} = \tau_i.$$

Writing $a = \sum_i a_i e_i$ and applying the triangle inequality gives

$$\|Pa\|_2 = \left\| \sum_i a_i P e_i \right\|_2 \le \sum_i |a_i| \, \|Pe_i\|_2 = \sum_i \sqrt{\tau_i}|a_i| = W(a).$$

Thus the weighted absolute residual controls the part of the residual that lies in the prediction direction space.

By definition, our mechanism minimizes $W(y-z)$ over feasible predictions $z \in K$. Since the least-squares prediction $z^\star$ is also feasible, optimality gives $W(\widehat{r}) \le W(r^\star)$. Moreover, the leverage scores sum to the dimension of L: $\sum_i \tau_i = \mathrm{tr}(P) = k$. Cauchy–Schwarz therefore implies

$$W(r^\star) = \sum_i \sqrt{\tau_i}|r_i^\star| \le \left(\sum_i \tau_i \right)^{1/2} \left(\sum_i (r_i^\star)^2 \right)^{1/2} = \sqrt{k} \, \|r^\star\|_2 .$$

Combining these inequalities yields $\|P\widehat{r}\|_2 \le W(\widehat{r}) \le W(r^\star) \le \sqrt{k} \, \|r^\star\|_2$. We conclude

$$\|\widehat{r}\|_2^2 = \|P\widehat{r}\|_2^2 + \|(I-P)\widehat{r}\|_2^2 \le k \, \|r^\star\|_2^2 + \|r^\star\|_2^2 = (k+1) \, \|r^\star\|_2^2 = (k+1) \, \mathrm{OPT}_\Theta(y). \quad \square$$

The lower bound in Section 4.2 shows that the unrestricted factor $d+1$ is optimal even among universally strategyproof randomized mechanisms.

4 Strategyproofness in Expectation

We now introduce randomness and let each agent minimize expected p-th power error, $\mathbb{E}|Z_i - y_i|^p$, for some $p > 0$. The social objective remains squared error. Throughout this section, the hypothesis class is *complete*: $\Theta = \mathbb{R}^d$, so every vector in $L = \mathrm{col}(X)$ is an admissible prediction. We first show a sharp bound for the perhaps most natural choice, $p = 1$, then demonstrate that randomization cannot help for $p = 2$, when the individual loss and the social loss coincide, and finally show that every other positive exponent permits a strict improvement over $d+1$. Together, these results give an exact characterization of when randomization helps.

4.1 Absolute Individual Loss: A Sharp Upper Bound at $p = 1$

A natural randomized approach is a form of *random dictatorship*: randomly select a group of agents and interpolate their reported data. Full interpolation of d independent observations protects the sampled agents but leaves no freedom to fit the remaining data. Interpolating fewer reports preserves useful freedom, but requires extra care to ensure truthfulness.

Mechanism: Partial interpolation. Let $L = \mathrm{col}(X)$ and $H = X(X^\top X)^{-1}X^\top$, the orthogonal projector onto L. The ordinary least-squares (OLS) predictions are Hy, and the residual is $r^\star = (I-H)y$. Thus $r^\star \perp L$ and $\mathrm{OPT}(y) = \|r^\star\|_2^2$. For $B \subseteq [n]$, H_B denotes its principal submatrix, $H_{:,B}$ its columns, and y_B the restricted vector.

For $b \in \{0, \dots, d\}$, draw a set $B \subseteq [n]$ of size b with the probability below and return

$$\mathbb{P}(B) = \frac{\det(H_B)}{\binom{d}{b}}, \qquad Z^B(y) = \operatorname*{arg\,min}_{z \in L:\ z_B = y_B} \|z - y\|_2^2 . \tag{2}$$

The probabilities depend only on public features. Geometrically, the distribution favors observations spanning different directions in the prediction space. We may randomize b independently of the reports. When $b = 0$, the mechanism is the (non-strategyproof) OLS rule; when $b = d$, it is the $(d+1)$-efficient volume sampling mechanism of Dereziński & Warmuth (2017).

We give the probabilities explicitly for ease of interpretation, but sampling does not require enumerating subsets. An iterative sampling scheme, adapted from the projection determinantal point process (DPP) sampler presented by Kulesza & Taskar (2012), takes $O(nd^2)$ arithmetic operations; see Appendix A.1.

Efficiency and strategyproofness. We first note that (2) indeed defines a probability distribution. Since H is a rank-d orthogonal projector, its principal-submatrix determinants satisfy

$$\sum_{|B|=b} \det(H_B) = \binom{d}{b}, \tag{3}$$

as shown by the determinant expansion in Appendix A. These determinants are nonnegative, so dividing by $\binom{d}{b}$ gives probabilities that sum to one.

The following two lemmas establish the approximation guarantee and truthfulness.

Lemma 4.1 (Partial-interpolation efficiency)**.** *For the complete hypothesis class $\Theta = \mathbb{R}^d$, the fixed-size rule (2) with parameter b has squared-error approximation ratio at most $(d+1)/(d-b+1)$.*

Proof. The case $b = 0$ returns OLS, so suppose $b \geq 1$. Write $r^\star = (I - H)y$. We first show that least squares subject to interpolating the reports on B produces the prediction $Hy + q^B$, where $q^B := H_{:,B} H_B^{-1} r^\star_B$.

To interpolate the sampled reports, we add q^B to the OLS prediction Hy. This correction belongs to L and satisfies $(q^B)_B = r^\star_B$. It is also orthogonal to every $v \in L$ with $v_B = 0$, since $\langle q^B, v\rangle = (r^\star_B)^\top H_B^{-1} v_B = 0$. Every other feasible correction has the form $q^B + v$, where $v \in L$ and $v_B = 0$. Orthogonality gives $\|q^B + v\|_2^2 = \|q^B\|_2^2 + \|v\|_2^2$, so q^B has minimum norm. Because $r^\star \perp L$ and $H^2 = H$, we obtain

$$Z^B(y) = Hy + q^B, \qquad \|Z^B(y) - y\|_2^2 = \|r^\star\|_2^2 + (r^\star_B)^\top H_B^{-1} r^\star_B. \tag{4}$$

We now bound the average of $(r^\star_B)^\top H_B^{-1} r^\star_B$, weighted by $\det(H_B)$. A rank-one perturbation captures this product: for each sampled B and $t > 0$, the determinant lemma (e.g. Horn & Johnson (2013, p. 26, Eq. (0.8.5.11))) gives

$$\det(H_B + t r^\star_B (r^\star_B)^\top) = \det(H_B) + t \det(H_B)(r^\star_B)^\top H_B^{-1} r^\star_B. \tag{5}$$

The coefficient of t is the extra error. To sum these determinants, recall that $r^\star \perp L$: adding $t r^\star (r^\star)^\top$ leaves the d unit eigenvalues of H unchanged and adds the eigenvalue $t\|r^\star\|_2^2$ in the residual direction when $r^\star \neq 0$. Each nonzero product of b eigenvalues either uses only unit eigenvalues or uses this additional eigenvalue and $b-1$ unit eigenvalues. Thus,

$$\sum_{|B|=b} \det(H_B + t r^\star_B (r^\star_B)^\top) = \binom{d}{b} + t \binom{d}{b-1} \|r^\star\|_2^2 . \tag{6}$$

The formula also holds when $r^\star = 0$. Summing (5) over sampled sets gives $\binom{d}{b} + t\binom{d}{b}\mathbb{E}_B[(r^\star_B)^\top H_B^{-1} r^\star_B]$. This is at most the full sum in (6): the remaining minors are nonnegative because $H + t r^\star (r^\star)^\top$ is positive semidefinite. Subtracting $\binom{d}{b}$ and dividing by $t\binom{d}{b}$ therefore gives

$$\mathbb{E}_B[(r^\star_B)^\top H_B^{-1} r^\star_B] \leq \frac{\binom{d}{b-1}}{\binom{d}{b}} \|r^\star\|_2^2 = \frac{b}{d-b+1} \|r^\star\|_2^2 .$$

Combining this with (4) and $\|r^\star\|_2^2 = \mathrm{OPT}(y)$ proves the claim. □

Lemma 4.2 (Partial-interpolation strategyproofness)**.** *A report-independent distribution of b with $\mathbb{E}b \geq d/2$ makes the mechanism SPIE for absolute individual loss.*

Proof Sketch. Fix an agent i, the other reports, and a deviation from their true label y_i to $y_i + \delta$. When $i \in B$, the mechanism interpolates their report, so the deviation increases their loss from zero to $|\delta|$. When $i \notin B$, the agent can influence only the part of the fit that remains free after interpolation.

To quantify this remaining influence, let H^B be the orthogonal projector onto $\{z \in L : z_B = 0\}$, and set $h_i^B = (H^B)_{ii} \geq 0$. The constrained least-squares formula in (4) shows that an unselected agent's prediction changes by $h_i^B \delta$. Their absolute loss can therefore decrease by at most $h_i^B |\delta|$.

For each fixed sample size b, we prove in Appendix A that the sampling rule gives

$$\mathbb{P}(i \in B) = \frac{b}{d} H_{ii}, \qquad \mathbb{E}_B[\mathbf{1}_{\{i \notin B\}} h_i^B] = \frac{d-b}{d} H_{ii}. \tag{7}$$

Intuitively, the first formula allocates the b interpolation slots in proportion to each agent's leverage H_{ii}, whose total is d. For the second formula, the b interpolation constraints fix b of the d prediction directions. The determinant weights leave, on average, a fraction $(d-b)/d$ of each agent's original influence in the remaining fit. Thus larger samples make an agent more likely to be selected while reducing the expected influence remaining outside the sampled set.

Let Δ_i^B be the agent's loss after the deviation minus their truthful loss. Since the sampling distribution is independent of reports, the two cases above give

$$\mathbb{E}_B \Delta_i^B \geq |\delta| \left(\mathbb{P}(i \in B) - \mathbb{E}_B[\mathbf{1}_{\{i \notin B\}} h_i^B] \right) = \frac{2b-d}{d} H_{ii} |\delta|.$$

Averaging over the report-independent choice of b replaces b by $\mathbb{E}b$. The expected loss cannot decrease when $\mathbb{E}b \geq d/2$, which proves strategyproofness in expectation. □

To obtain the final bound, take $b = d/2$ when d is even, and mix $b = (d-1)/2$ and $b = (d+1)/2$ equally when d is odd. In both cases $\mathbb{E}b = d/2$, so Lemma 4.2 gives SPIE. Averaging the corresponding bounds in Lemma 4.1 gives the ratio below.

Theorem 4.3 (A uniform sub-2 upper bound)**.** *For absolute individual loss and the complete hypothesis class $\Theta = \mathbb{R}^d$, an SPIE mechanism achieves squared-error approximation ratio*

$$\rho_d = 2 - \frac{1}{\lceil d/2 \rceil + 1}.$$

As an example, in constant fitting ($d = 1$), the rule returns the mean with probability $1/2$ and a uniformly chosen report otherwise, recovering the optimal $3/2$ mechanism of Feldman & Wilf (2013).

The absolute-loss bound is tight. Finally, we prove tightness of the above result.

Theorem 4.4 (Optimality under absolute individual loss)**.** *For every $d \geq 1$, there is a full-column-rank design with $\Theta = \mathbb{R}^d$ on which every absolute-loss SPIE mechanism has squared-error approximation ratio at least ρ_d.*

Proof Sketch. Set $n = d+1$ and take X with first d rows I_d and last row $-\mathbf{1}^\top$, so predictions sum to zero. For labels summing to $t > 0$, residuals therefore sum to t. Least squares spreads this residual equally, giving $\mathrm{OPT}(y) = t^2/n$. For a c-approximation, the normalized residual $V = (y - Z(y))/t$ satisfies $\sum_i V_i = 1$ and $\mathbb{E}\,\|V\|_2^2 \leq c/n$.

Accuracy favors spreading this unit residual over many agents. Let P count the positive coordinates of V, and let A be its total negative mass. The positive mass is $1 + A$, so Cauchy–Schwarz gives $\|V\|_2^2 \geq (1+A)^2/P$. Thus few positive coordinates are costly, and negative residuals require still more positive mass. Accounting for the fact that P is an integer, Appendix B shows that beating ρ_d would require

$$\mathbb{E}[P - A] \geq \frac{n+1}{2} + \Delta$$

for some constant $\Delta > 0$, uniformly over all profiles with $t > 0$.

Incentives constrain this spreading: truthfulness gives a lower bound $2\mathbb{P}(V_i > 0) - 1$ on the derivative of agent i's expected absolute loss with respect to their own label, wherever it exists. If we could sum these rates along a common path of profiles with total t, the total expected absolute loss $F(t)$ would satisfy

$$F'(t) \geq 2\mathbb{E}P - n \geq \frac{F(t)}{t} + 2\Delta,$$

since $F(t)/t = \mathbb{E}\,\|V\|_1 = 1 + 2\mathbb{E}A$. This gives $(F(t)/t)' \geq 2\Delta/t$, forcing normalized loss to grow without bound, whereas accuracy requires $F(t)/t \leq \sqrt{c}$.

The agents' report paths do differ, but averaging over a growing ball in the zero-sum prediction space makes the common-path calculation accurate up to a vanishing error on each fixed t-interval. Appendix B justifies this step, completing the contradiction and proving $c \geq \rho_d$. □

4.2 SQUARED INDIVIDUAL LOSS: A LOWER BOUND AT $p = 2$

Next, we consider what happens when the individual loss has the same form as the social loss and each agent aims to minimize their expected squared loss. Perhaps surprisingly, unlike with absolute individual loss, randomization provides no improvement over deterministic mechanisms in this case.

Theorem 4.5 (Optimality under squared individual loss)**.** *For each $d \geq 1$, there is a full-column-rank design with $\Theta = \mathbb{R}^d$ such that every SPIE mechanism for squared individual loss has approximation ratio at least $d + 1$.*

Proof Sketch. Use the same design as in Theorem 4.4, with $n = d + 1$ and $L = \mathbf{1}^\perp$. For a profile y with $\sum_i y_i = \tau > 0$, the optimum is τ^2/n.

Suppose the mechanism has finite approximation ratio c. Reporting $y_i - \tau$ makes the profile $y - \tau e_i$ realizable, so it must be fitted exactly. Starting from this report and increasing it to y_i, the squared-loss incentive inequalities give

$$C_i(y) = 2 \int_0^\tau r_i\big(y - (\tau - s)e_i\big)\, ds,$$

where $C_i(v) = \mathbb{E}[(Z_i(v) - v_i)^2]$ and $r_i(v) = v_i - \mathbb{E}Z_i(v)$. At parameter s, agent i reports $y_i - \tau + s$, so the sum of all reports is s.

If the integrands were evaluated at a common profile $v(s)$ with $\sum_j v_j(s) = s$, feasibility would give $\sum_i r_i(v(s)) = s$. Their sum would therefore be

$$2 \int_0^\tau \sum_i r_i(v(s))\, ds = 2 \int_0^\tau s\, ds = \tau^2,$$

which is n times the optimum.

Again, different agents follow different paths, but the same translation averaging from Theorem 4.4 justifies summing: the paths differ by bounded shifts within L, whose contribution vanishes when averaged over larger balls. Thus the averaged social cost tends to τ^2, while remaining at most $c\tau^2/n$, giving $c \geq n = d + 1$. Unlike absolute loss, squared loss determines this sum at each fixed τ, and no growth argument over τ is needed. Appendix C supplies the details. □

The same worst-case lower bound holds for affine predictors on $\mathbb{R}^{d-1}$ with a free intercept as well, as detailed in Appendix C.1. Thus, the optimal worst-case ratio is $d + 1$ for squared-loss SPIE. Since all positive powers induce the same deterministic preferences, every universally strategyproof mechanism is also squared-loss SPIE, demonstrating the optimality of Theorem 3.1.

4.3 OTHER POWERS: AN UPPER BOUND FOR EVERY $p \neq 2$

In general, do SPIE mechanisms improve on deterministic mechanisms? In other words, is absolute loss special, or is squared loss the exception? The next construction shows that it is the latter. It combines two independent interpolating fits through a small lottery. Averaging the fits improves accuracy; additional outcomes make truthful reporting optimal. The hypothesis class is again complete, $\Theta = \mathbb{R}^d$.

Theorem 4.6 (Randomization helps at every positive power except two)**.** *For every $p > 0$ and every full-column-rank public design with d free coefficients, the complete hypothesis class $\Theta = \mathbb{R}^d$ admits an SPIE mechanism with squared-error approximation ratio at most $1 + dq_p$, where, for $\kappa = 1 + \sqrt{2}$,*

$$q_p = \begin{cases} \dfrac{2 + 2^{p-1}}{2 + 2^p}, & 0 < p < 1, \\ \dfrac{1}{2} + \lambda_p, \quad \lambda_p = \dfrac{2}{2 + \kappa^{p-2} + \kappa^{2-p}}, & p \geq 1. \end{cases} \tag{8}$$

Here $q_2 = 1$ and $q_p < 1$ for every $p \neq 2$.

Proof Sketch. First, draw B and C independently using (2) with $b = d$. The interpolating fits are $U = XX_B^{-1}y_B$ and $V = XX_C^{-1}y_C$. Singular sets have probability zero. These fits satisfy

$$\mathbb{E}U = Hy, \qquad \sigma^2 := \mathbb{E}\,\|U - Hy\|_2^2 \le d\ \mathrm{OPT}(y). \tag{9}$$

The same statements hold for V. The second-moment bound follows from Lemma 4.1 at $b = d$ and the orthogonality of $U - Hy$ and $y - Hy$. The unbiasedness formula $\mathbb{E}U = Hy$ is proved in Lemma D.1. For $b = d$, these properties were also proved by Dereziński & Warmuth (2017) in the context of learning from small samples.

We consider $p \ge 1$, where the loss is convex, and $0 < p < 1$, where subadditivity replaces convexity.

First suppose $p \ge 1$. Put $M = (U + V)/2$ and return

$$Z = \begin{cases} M, & \text{with probability } 1 - \lambda_p, \\ M + (U - V)/\sqrt{2}, & \text{with probability } \lambda_p/2, \\ M - (U - V)/\sqrt{2}, & \text{with probability } \lambda_p/2. \end{cases} \tag{10}$$

The midpoint reduces variance, while the two extrapolated outcomes enforce incentives. To see the incentive balance, write $Z = (1-T)U+TV$, with $T \in \{1/2, (1-\sqrt{2})/2, (1+\sqrt{2})/2\}$. Conditional on B, C, an agent in neither sample has no influence, while an agent in both is fitted exactly. An agent in exactly one sample trades off the fit that interpolates their report against a fit independent of it. Their expected p-th power loss is convex in their report. The chosen λ_p makes its derivative at truth zero when the two truthful predictions at that agent differ; if they coincide, truth has zero loss. Independence and unbiasedness give squared social cost $\mathrm{OPT}(y) + (\frac{1}{2} + \lambda_p)\sigma^2$, proving the convex-range bound.

For $0 < p < 1$, convexity no longer holds. Instead return U and V each with probability $w_p = 1/(2 + 2^p)$, and their midpoint with probability $2^p w_p$. If only one fit interpolates the agent, writing s for the other fit's error and δ for their deviation gives expected loss $w_p(|\delta|^p + |s|^p + |s + \delta|^p)$. The inequality $|s|^p \le |\delta|^p + |s + \delta|^p$ makes truth optimal. The same variance calculation gives q_p in (8). □

Why is $p = 2$ so different? Squared loss is special because the moment that balances incentives is exactly the moment that controls the benefit of averaging. To see this in the two-fit construction, condition on B, C and consider an agent in $B \setminus C$; the other case is symmetric. If the other fit's prediction differs from their true value by $s \ne 0$, their expected squared loss after a deviation δ is $F(\delta) = \mathbb{E}[(Ts + (1 - T)\delta)^2]$. For truthfulness conditional on the samples, we need $F'(0) = 2s\,\mathbb{E}[T(1 - T)] = 0$. But the variance multiplier of the combined fit is

$$\mathbb{E}[(1 - T)^2 + T^2] = 1 - 2\mathbb{E}[T(1 - T)] = 1.$$

Thus averaging cannot reduce variance under this conditional incentive balance: the extrapolation needed for truthfulness exactly cancels the gain from the midpoint. In our lottery, this balance gives $\lambda_2 = 1/2$ and $q_2 = 1$.

5 Conclusion and Discussion

The optimal worst-case deterministic and universally strategyproof price is $d + 1$ for all single-peaked preferences, where d counts free coefficients rather than observations. For SPIE, squared individual loss is the unique power that retains this price. Absolute loss has the exact optimum $\rho_d = 2 - 1/(\lceil d/2 \rceil + 1) < 2$, while each $p > 0$ with $p \notin \{1, 2\}$ also admits a strict improvement over $d + 1$.

Our SPIE mechanisms assume the complete hypothesis set. An immediate open problem is to obtain similar guarantees for arbitrary convex and compact hypothesis sets. A more ambitious question is to give a complete characterization of strategyproof mechanisms. Other interesting directions include going beyond linear regression, such as to non-linear models or deep learning, or incorporating external predictions and advice.

AI USE STATEMENT

ChatGPT 5.6-Sol and ChatGPT 6-Astra were used to assist with exploring proof sketches, identifying and checking references, and drafting the manuscript. For Section 3, the authors first considered the complete hypothesis set and proposed weighted L_1-ERM, with weights reflecting how strongly changes in the model can affect each agent's prediction. GPT derived the $(d+1)$ approximation guarantee and subsequently generalized the argument to arbitrary convex sets. For Section 4, the authors proposed volume sampling for the special case $b = d$, and GPT subsequently suggested selecting smaller values of b. For the lower bound, the authors first constructed a $(d+1)$ lower bound for all deterministic strategyproof mechanisms, and GPT extended it to SPIE mechanisms with squared loss at the authors' request. The authors are responsible for the correctness, originality, and accuracy of the final submission, including all AI-assisted text and mathematical arguments.

REFERENCES

Eric Balkanski and Cherlin Zhu. Strategyproof learning with advice. In *Proceedings of the 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, pp. 163–166. PMLR, 2025. URL https://proceedings.mlr.press/v272/balkanski25a.html.

Yang Cai, Constantinos Daskalakis, and Christos H. Papadimitriou. Optimum statistical estimation with strategic data sources. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pp. 280–296. PMLR, 2015. URL https://proceedings.mlr.press/v40/Cai15.html.

Felipe Caro, Jérémie Gallien, Miguel Díaz, Javier García, José Manuel Corredoira, Marcos Montes, José Antonio Ramos, and Juan Correa. Zara uses operations research to reengineer its global distribution process. *Interfaces*, 40(1):71–84, 2010. doi: 10.1287/inte.1090.0472. URL https://doi.org/10.1287/inte.1090.0472.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(29):1069–1109, 2011. URL https://jmlr.org/papers/v12/chaudhuri11a.html.

Yiding Chen, Xiaojin Zhu, and Kirthevasan Kandasamy. Mechanism design for collaborative normal mean estimation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49365–49402. Curran Associates, Inc., 2023. doi: 10.52202/075280-2148. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/9af2b1d6acf561af9c4cf70d52c7a49d-Abstract-Conference.html.

Yiling Chen, Chara Podimata, Ariel D. Procaccia, and Nisarg Shah. Strategyproof linear regression in high dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 9–26. ACM, 2018. doi: 10.1145/3219166.3219175. URL https://doi.org/10.1145/3219166.3219175.

Yiling Chen, Tao Lin, Wei Tang, and Jamie Tucker-Foltz. The price and complexity of explainable information design. In *Proceedings of the 27th ACM Conference on Economics and Computation*, 2026. URL https://arxiv.org/abs/2508.14196. One-page abstract. Full version: Explainable Information Design, arXiv:2508.14196.

CMS. Report to Congress: Risk adjustment in Medicare Advantage. Technical report, Centers for Medicare & Medicaid Services, December 2024. URL https://www.cms.gov/medicare/health-plans/medicareadvtgspecratestats/risk-adjustors-items/reporttocongress.

Rachel Cummings, Stratis Ioannidis, and Katrina Ligett. Truthful linear regression. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pp. 448–483. PMLR, 2015. URL https://proceedings.mlr.press/v40/Cummings15.html.

Sanjoy Dasgupta, Nave Frost, Michal Moshkovitz, and Cyrus Rashtchian. Explainable k-means and k-medians clustering. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7055–7065. PMLR, 2020. URL https://proceedings.mlr.press/v119/moshkovitz20a.html.

Ofer Dekel, Felix Fischer, and Ariel D. Procaccia. Incentive compatible regression learning. *Journal of Computer and System Sciences*, 76(8):759–777, 2010. doi: 10.1016/j.jcss.2010.03.003. URL https://doi.org/10.1016/j.jcss.2010.03.003.

Michał Dereziński and Manfred K. Warmuth. Unbiased estimates for linear regression via volume sampling. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3084–3093, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/54e36c5ff5f6a1802925ca009f3ebb68-Abstract.html.

Michal Feldman and Yoav Wilf. Strategyproof facility location and the least squares objective. In *Proceedings of the 14th ACM Conference on Electronic Commerce*, pp. 873–890. ACM, 2013. doi: 10.1145/2492002.2482543. URL https://doi.org/10.1145/2492002.2482543.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011. doi: 10.1198/jasa.2011.r10138. URL https://doi.org/10.1198/jasa.2011.r10138.

Moritz Hardt, Nimrod Megiddo, Christos H. Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 111–122. ACM, 2016. doi: 10.1145/2840728.2840730. URL https://doi.org/10.1145/2840728.2840730.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition, 2009. doi: 10.1007/978-0-387-84858-7. URL https://doi.org/10.1007/978-0-387-84858-7.

Tao Hong, Jason Wilson, and Jingrui Xie. Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, 5(1):456–462, 2014. doi: 10.1109/TSG.2013.2274373. URL https://doi.org/10.1109/TSG.2013.2274373.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2013. doi: 10.1017/CBO9781139020411. URL https://doi.org/10.1017/CBO9781139020411.

Safwan Hossain and Nisarg Shah. The effect of strategic noise in linear regression. *Autonomous Agents and Multi-Agent Systems*, 35(2):21, 2021. doi: 10.1007/s10458-021-09502-0. URL https://doi.org/10.1007/s10458-021-09502-0.

Thomas Kleine Buening, Jiarui Gan, Debmalya Mandal, and Marta Kwiatkowska. Strategyproof reinforcement learning from human feedback. In *Advances in Neural Information Processing Systems*, volume 38, pp. 101431–101464. Curran Associates, Inc., 2025. doi: 10.52202/085713-3396. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/92f43b1d33fae4aa1958f75317f0cec1-Abstract-Conference.html.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2–3):123–286, 2012. doi: 10.1561/2200000044. URL https://doi.org/10.1561/2200000044.

Reshef Meir, Ariel D. Procaccia, and Jeffrey S. Rosenschein. Algorithms for strategyproof classification. *Artificial Intelligence*, 186:123–156, 2012. doi: 10.1016/j.artint.2012.03.008. URL https://doi.org/10.1016/j.artint.2012.03.008.

Aditya Krishna Menon and Robert C. Williamson. The cost of fairness in binary classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pp. 107–118. PMLR, 2018. URL https://proceedings.mlr.press/v81/menon18a.html.

Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980. doi: 10.1007/BF00128122. URL https://doi.org/10.1007/BF00128122.

MPAC. Methodology guide: Valuing residential properties in Ontario, 2016 base year. Technical report, Municipal Property Assessment Corporation, March 2016. URL https://www.mpac.ca/sites/default/files/docs/pdf/ResidentialProperties.pdf.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744, 2022. URL https://proceedings.neurips.cc/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Javier Perote and Juan Perote-Peña. Strategy-proof estimators for simple regression. *Mathematical Social Sciences*, 47(2):153–176, 2004. doi: 10.1016/S0165-4896(03)00085-4. URL https://doi.org/10.1016/S0165-4896(03)00085-4.

Javier Perote, Juan Perote-Peña, and Marc Vorsatz. Strategic behavior in regressions: an experimental study. *Theory and Decision*, 79(3):517–546, 2015. doi: 10.1007/s11238-014-9473-9. URL https://doi.org/10.1007/s11238-014-9473-9.

PJM. 2025 long-term load forecast supplement. Technical report, PJM Interconnection, January 2025. URL https://www.pjm.com/-/media/DotCom/planning/res-adeq/load-forecast/2025-long-term-load-forecast-supplement.pdf.

Ariel D. Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. *ACM Transactions on Economics and Computation*, 1(4):18:1–18:26, 2013. doi: 10.1145/2542174.2542175. URL https://doi.org/10.1145/2542174.2542175.

Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. Robustness may be at odds with accuracy. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=SyxAb30cY7.

Jiongxiao Wang, Junlin Wu, Muhao Chen, Yevgeniy Vorobeychik, and Chaowei Xiao. RLHFPoison: Reward poisoning attack for reinforcement learning with human feedback in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2551–2570. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.140. URL https://aclanthology.org/2024.acl-long.140/.

Liyang Zhu, Amina Manseur, Meng Ding, Jinyan Liu, Jinhui Xu, and Di Wang. Truthful high dimensional sparse linear regression. In *Advances in Neural Information Processing Systems*, volume 37, pp. 66995–67033. Curran Associates, Inc., 2024. doi: 10.52202/079017-2139.

A Implementation and analysis of partial interpolation

We first give an iterative implementation of the sampling rule, then explain how an unselected report affects the fit and prove the two formulas in (7). Together, these calculations complete the proof of Lemma 4.2. Throughout, H is the rank-d orthogonal projector onto L, and a size-b set B has probability $\det(H_B)/\binom{d}{b}$.

A.1 Iterative sampling and running time

We use the Gram–Schmidt projection-DPP sampler presented by Kulesza & Taskar (2012, Section 2.4.4, Algorithm 1) and stop after b selections. The calculation below shows that this yields the distribution in (2).

Compute a thin QR factorization $X = UR$, where $U \in \mathbb{R}^{n\times d}$ has orthonormal columns, so $H = UU^\top$. Write $u_i^\top$ for row i of U. Initialize $v_i = u_i$ for every i and $B = \varnothing$. For $t = 0, \dots, b-1$, perform the following steps:

1. Select an index i_t with probability $\|v_{i_t}\|_2^2/(d-t)$ and add it to B.
2. Set $q_t = v_{i_t}/\|v_{i_t}\|_2$, and update every vector by $v_j \leftarrow v_j - \langle v_j, q_t\rangle q_t$.

Return the unordered set B. For $b = 0$, return the empty set. Each update removes the direction just selected, so subsequent selections favor observations that contribute a new direction. Previously selected rows have zero residual and cannot be selected again.

Correctness. The vectors $q_0, \dots, q_{t-1}$ are orthonormal. At the start of step t, let $P_t = I_d - \sum_{s<t} q_s q_s^\top$, so $v_i = P_t u_i$. Since P_t is a rank-$(d-t)$ projector and $U^\top U = I_d$, the probabilities at that step sum to one:

$$\sum_{i=1}^{n} \|v_i\|_2^2 = \operatorname{tr}(P_t U^\top U P_t) = \operatorname{tr}(P_t) = d - t.$$

Let U_B contain the selected rows in their selection order. Their coordinates in the orthonormal directions $q_0, \dots, q_{b-1}$ form a lower-triangular matrix, with the residual lengths at selection on its diagonal. Thus $\det(U_B U_B^\top) = \det(H_B)$ is the product of the squared residual lengths. Multiplying the stepwise selection probabilities therefore gives, for $b \geq 1$,

$$\mathbb{P}(i_0, \dots, i_{b-1}) = \frac{\det(H_B)}{d(d-1)\cdots(d-b+1)}.$$

There are $b!$ orderings, so their probabilities sum to $\det(H_B)/\binom{d}{b}$, as required.

Computational cost. The QR factorization costs $O(nd^2)$ arithmetic operations and can be reused because X is public and fixed. Each sampling step updates n vectors of length d, so drawing B costs $O(ndb)$ after preprocessing. The fit can also be computed without forming the $n \times n$ matrix H: set $a = U^\top y$, solve $(U_B U_B^\top)c = y_B - U_B a$, and return $U(a + U_B^\top c)$. This is the prediction in (4) and costs $O(nd + db^2 + b^3)$ operations; when $b = 0$, simply return Ua. Since $b \leq d \leq n$, the total cost is $O(nd^2)$. The two independent full-size draws used in Section 4.3 have the same asymptotic running time.

A.2 How an unselected report affects the fit

For a sampled nonempty set B, the projector onto the prediction directions that vanish on B is

$$H^B = H - H_{:,B} H_B^{-1} H_{B,:}, \qquad h_i^B = (H^B)_{ii}.$$

To see this, $H^2 = H$ gives $H_{B,:}H_{:,B} = H_B$, so the subtracted matrix is the orthogonal projector onto the span of the columns He_j, $j \in B$. Within L, being orthogonal to these columns is equivalent to having $z_j = 0$ for all $j \in B$, since $\langle He_j, z\rangle = z_j$. Thus H^B retains exactly the allowed directions of change after interpolation, and $h_i^B \geq 0$.

For $i \notin B$, changing y_i by δ leaves y_B unchanged. Substituting into the prediction formula in (4) gives $Z^B(y + \delta e_i) - Z^B(y) = \delta H^B e_i$. In particular, agent i's prediction changes by $h_i^B \delta$. For $b = 0$, take $H^{\varnothing} = H$: the agent is never selected and retains influence H_{ii}, proving both formulas in (7) in this case. Below, assume $1 \le b \le d$.

A.3 Selection probabilities and remaining influence

A determinant expansion. For a symmetric $m \times m$ matrix A, let A_C retain the rows and columns indexed by the same set C; its determinant is called a *principal minor*, with $\det(A_{\varnothing}) = 1$. If $\lambda_1, \dots, \lambda_m$ are the eigenvalues of A, then

$$\sum_{\substack{C \subseteq [m] \\ |C| = s}} \det(A_C) = \sum_{1 \le j_1 < \cdots < j_s \le m} \lambda_{j_1} \cdots \lambda_{j_s}, \qquad s \ge 1. \tag{11}$$

For $1 \le s \le m$, this is Horn & Johnson (2013, Theorem 1.2.16, p. 54); for $s > m$, both sums are empty and equal zero. To prove it, expand $\det(I_m + tA)$ by linearity in each column. Choosing columns from tA exactly at the indices in C contributes $t^{|C|} \det(A_C)$, since the remaining columns come from I_m. On the other hand, the eigenvalues give $\det(I_m + tA) = \prod_{j=1}^{m} (1 + t\lambda_j)$. Comparing coefficients of t^s proves (11). For $A = H$, only products of s of its d unit eigenvalues are nonzero, giving the normalization in (3); the case $b = 0$ has the single empty set of weight one.

Selection probability. Using the matrix U from Appendix A.1, let U_{-i} denote U with row i removed. The matrix obtained by deleting row and column i from H is $U_{-i}U_{-i}^{\top}$. Its nonzero eigenvalues coincide with those of

$$U_{-i}^{\top} U_{-i} = I_d - u_i u_i^{\top}.$$

The latter has $d - 1$ eigenvalues equal to 1 and one equal to $1 - \|u_i\|_2^2 = 1 - H_{ii}$. Zero eigenvalues do not affect the sums in (11). A product of b eigenvalues therefore either uses only unit eigenvalues or uses $1 - H_{ii}$ together with $b - 1$ unit eigenvalues. Consequently,

$$\sum_{\substack{|B| = b \\ i \notin B}} \det(H_B) = \binom{d-1}{b} + (1 - H_{ii}) \binom{d-1}{b-1},$$

$$\sum_{\substack{|B| = b \\ i \in B}} \det(H_B) = \binom{d}{b} - \sum_{\substack{|B| = b \\ i \notin B}} \det(H_B) = \binom{d-1}{b-1} H_{ii}.$$

Here a binomial coefficient is zero when its lower index exceeds its upper index. Dividing the second equality by $\binom{d}{b}$ proves $\mathbb{P}(i \in B) = (b/d) H_{ii}$.

Expected remaining influence. For a sampled set B with $i \notin B$, order the rows and columns of $H_{B \cup \{i\}}$ with i last. Subtracting $H_{i,B} H_B^{-1}$ times its first b rows from its last row gives

$$\det(H_{B \cup \{i\}}) = \det(H_B) \big(H_{ii} - H_{i,B} H_B^{-1} H_{B,i} \big) = \det(H_B) h_i^B.$$

Thus the sampling weight times the remaining influence is the determinant for the enlarged set $C = B \cup \{i\}$. If $\det(H_C) > 0$, then H_C and its principal submatrix $H_{C \setminus \{i\}}$ are positive definite. Hence enlarged sets not arising from a positive-weight B contribute zero, and we may sum over all C. For $b < d$, the selection calculation for sets of size $b + 1$ gives

$$\begin{aligned} \mathbb{E}_B[\mathbf{1}_{\{i \notin B\}} h_i^B] &= \frac{1}{\binom{d}{b}} \sum_{\substack{|C| = b+1 \\ i \in C}} \det(H_C) \\ &= \frac{\binom{d-1}{b}}{\binom{d}{b}} H_{ii} = \frac{d - b}{d} H_{ii}. \end{aligned}$$

For $b = d$, every enlarged determinant is zero because H has rank d, giving the same formula.

Finally, a deviation by δ costs $|\delta|$ when the agent is selected and can reduce their loss by at most $h_i^B |\delta|$ otherwise. The two formulas in (7) therefore give an expected loss increase of at least $(2\mathbb{E}b - d) H_{ii} |\delta| / d$. This is nonnegative under the condition of Lemma 4.2, completing the proof.

B THE LOWER BOUND FOR ABSOLUTE INDIVIDUAL LOSS

We prove Theorem 4.4 by summing the agents' incentive constraints. Set $n = d + 1$ and take the design with first d rows I_d and last row $-\mathbf{1}^\top$, whose prediction space is $L = \mathbf{1}^\perp$. Write $y(w,t) = w + (t/n)\mathbf{1}$ for $w \in L$ and $t > 0$. Then $\mathrm{OPT}(y(w,t)) = t^2/n$.

The argument has two parts. A geometric inequality shows how many positive residuals a sufficiently accurate mechanism must produce. Truthfulness then forces its absolute loss to grow too quickly. Averaging over translations w lets us compare the agents' report paths at a common t.

B.1 ACCURACY REQUIRES MANY POSITIVE RESIDUALS

For a vector v with $\sum_i v_i = 1$, define $P(v) = |\{i : v_i > 0\}|$ and $A(v) = \sum_{i:v_i<0}(-v_i)$. Thus $1 \le P(v) \le n$, its positive coordinates sum to $1 + A(v)$, and $\|v\|_1 = 1 + 2A(v)$.

Lemma B.1 (A geometric residual bound)**.** *Let $k = \lfloor (n+1)/2 \rfloor$. Every $v \in \mathbb{R}^n$ with $\sum_i v_i = 1$ satisfies*

$$P(v) - A(v) \ge 2k + 1 - k(k+1)\,\|v\|_2^2 .$$

Proof. Cauchy–Schwarz applied to the positive coordinates gives $\|v\|_2^2 \ge (1 + A(v))^2/P(v)$. To sharpen this bound, note that for every positive integer p,

$$\frac{1}{p} - \frac{2k+1-p}{k(k+1)} = \frac{(p-k)(p-k-1)}{pk(k+1)} \ge 0.$$

The last inequality holds because no integer lies strictly between k and $k+1$. Also, $p \le n \le 2k(k+1)$ implies $2/p \ge 1/[k(k+1)]$. Consequently,

$$\|v\|_2^2 \ge \frac{1 + 2A(v)}{P(v)} \ge \frac{2k + 1 - P(v) + A(v)}{k(k+1)}.$$

Rearranging proves the claim. □

B.2 ALIGNING THE AGENTS' REPORT PATHS

Let Z be a measurable absolute-loss SPIE mechanism with approximation ratio $c < \infty$. Write $r(y) = y - Z(y)$ for its random residual and $g_i(y) = \mathbb{E}|r_i(y)|$ for agent i's truthful expected loss. For $R > 0$, let $B_R = \{w \in L : \|w\|_2 \le R\}$ and let W_R be uniform on B_R. For $t > 0$, define

$$Y_R(t) = W_R + \frac{t}{n}\mathbf{1}, \qquad F_R(t) = \mathbb{E}\,\|r(Y_R(t))\|_1 , \qquad N_R(t) = \sum_{i=1}^{n} \mathbb{P}(r_i(Y_R(t)) > 0).$$

These expectations and probabilities include both W_R and the mechanism's randomness. Since every residual sums to t, the triangle inequality and the approximation guarantee give

$$t \le F_R(t) \le \sqrt{n\,\mathbb{E}\,\|r(Y_R(t))\|_2^2} \le \sqrt{c}\,t. \tag{12}$$

For a single agent, truthfulness controls how g_i changes with their own report. We would like to add these changes at the same profile. The next lemma makes this step precise: averaging over B_R aligns the paths up to an error that vanishes as R grows.

Lemma B.2 (Averaged incentive inequality)**.** *The function F_R is locally Lipschitz on $(0,\infty)$, and there is a constant $C_{n,c}$ depending only on n and c such that*

$$F_R'(t) \ge 2N_R(t) - n - \frac{C_{n,c}\,t}{R} \quad \textit{for almost every } t > 0. \tag{13}$$

Proof. **One agent's report.** Fix the other reports and compare reports a and b. Truthfulness and the triangle inequality give

$$g_i(a, y_{-i}) \le \mathbb{E}|Z_i(b, y_{-i}) - a| \le g_i(b, y_{-i}) + |a - b|.$$

Exchanging a and b shows that the truthful expected loss is 1-Lipschitz in the agent's own report.

We also need a one-sided bound. Apply truthfulness at the profile $y - he_i$, with agent i deviating upward to y_i, where $h > 0$. The deviating prediction is drawn from $Z(y)$, so

$$g_i(y) - g_i(y - he_i) \geq \mathbb{E}\big[|r_i(y)| - |r_i(y) - h|\big]. \tag{14}$$

Dividing by h and letting $h \downarrow 0$ would give the contribution $2\mathbb{P}(r_i(y) > 0) - 1$. The issue is that $y - he_i$ depends on i.

Replacing shifted profiles by common profiles. Put $p_i = e_i - \mathbf{1}/n \in L$. The two profiles in (14) satisfy

$$y(w, t) - he_i = y(w - hp_i, t - h).$$

Thus all agents move between the same values t and $t - h$, but their translations within L differ. Averaging over a large ball makes this difference small.

To quantify it, write vol_L for volume in the $(n-1)$-dimensional space L. For $a \in L$ with $\|a\|_2 \leq R$, the symmetric difference of B_R and $B_R + a$ lies in the annulus with radii $R - \|a\|_2$ and $R + \|a\|_2$. Hence, with $\kappa_n = 2^{n-1}(n-1)$,

$$\frac{\mathrm{vol}_L(B_R \triangle (B_R + a))}{\mathrm{vol}_L(B_R)} \leq \frac{(R + \|a\|_2)^{n-1} - (R - \|a\|_2)^{n-1}}{R^{n-1}} \leq \kappa_n \frac{\|a\|_2}{R}.$$

Consequently, for any bounded measurable function $f : L \to \mathbb{R}$,

$$|\mathbb{E}f(W_R + a) - \mathbb{E}f(W_R)| \leq \kappa_n \sup_{w \in L} |f(w)| \, \frac{\|a\|_2}{R} \qquad (\|a\|_2 \leq R). \tag{15}$$

The approximation guarantee supplies the required uniform bound:

$$0 \leq g_i(y(w, s)) \leq \sqrt{\mathbb{E}r_i(y(w, s))^2} \leq \sqrt{\frac{c}{n}}\, s \qquad (w \in L,\ s > 0).$$

Let $G_{R,i}(t) = \mathbb{E}_{W_R} g_i(y(W_R, t))$, so $F_R = \sum_i G_{R,i}$. Since $\|p_i\|_2 \leq 1$, replacing $W_R - hp_i$ by W_R at parameter $t - h$ costs at most $\kappa_n \sqrt{c/n}\, th/R$ when $0 < h < \min\{t, R\}$. Combining this estimate with own-report Lipschitzness yields

$$|G_{R,i}(t) - G_{R,i}(t - h)| \leq h + \kappa_n \sqrt{\frac{c}{n}}\, \frac{th}{R}.$$

Thus each $G_{R,i}$, and therefore F_R, is locally Lipschitz and absolutely continuous on compact subintervals of $(0, \infty)$.

Summing the incentive constraints. Average (14) at $y = y(W_R, t)$ and use (15) to obtain

$$\frac{G_{R,i}(t) - G_{R,i}(t - h)}{h} \geq \mathbb{E}\left[\frac{|r_i(Y_R(t))| - |r_i(Y_R(t)) - h|}{h}\right] - \kappa_n \sqrt{\frac{c}{n}}\, \frac{t}{R}.$$

For every real x, the quotient $(|x| - |x - h|)/h$ lies in $[-1, 1]$ and converges to $2\mathbf{1}_{\{x>0\}} - 1$ as $h \downarrow 0$. Dominated convergence therefore gives, wherever $G'_{R,i}(t)$ exists,

$$G'_{R,i}(t) \geq 2\mathbb{P}(r_i(Y_R(t)) > 0) - 1 - \kappa_n \sqrt{\frac{c}{n}}\, \frac{t}{R}.$$

Summing over agents proves (13) with $C_{n,c} = \kappa_n \sqrt{nc}$. □

B.3 Completing the lower bound and transferring it to affine regression

Proof of Theorem 4.4. Set $k = \lfloor (n+1)/2 \rfloor$. The target ratio can be written as

$$\rho_{n-1} = \frac{n(2k + 1 - (n+1)/2)}{k(k+1)} = \begin{cases} 2 - \dfrac{2}{n+1}, & n \text{ odd}, \\ 2 - \dfrac{2}{n+2}, & n \text{ even}. \end{cases}$$

Suppose for contradiction that the mechanism has ratio $c < \rho_{n-1}$, and define $\Delta = k(k+1)(\rho_{n-1} - c)/n > 0$.

At each profile $y(w,t)$, apply Lemma B.1 to the normalized residual $v = r(y(w,t))/t$. Since $\sum_i v_i = 1$ and $\mathbb{E}\,\|v\|_2^2 \le c/n$, we get

$$\mathbb{E}[P(v) - A(v)] \ge 2k + 1 - \frac{k(k+1)c}{n} = \frac{n+1}{2} + \Delta.$$

After averaging in w, the expected positive count is $N_R(t)$ and the expected normalized negative mass is $(F_R(t)/t - 1)/2$. Therefore

$$2N_R(t) - n \ge \frac{F_R(t)}{t} + 2\Delta. \tag{16}$$

This is the conflict between accuracy and incentives: (16) requires many positive residuals, while each positive residual contributes to the loss growth in (13).

Combining these two bounds, for almost every $t > 0$,

$$\left(\frac{F_R(t)}{t}\right)' = \frac{F_R'(t)}{t} - \frac{F_R(t)}{t^2} \ge \frac{2\Delta}{t} - \frac{C_{n,c}}{R}.$$

Integrating from 1 to any fixed $T > 1$ gives

$$\frac{F_R(T)}{T} - F_R(1) \ge 2\Delta \log T - \frac{C_{n,c}(T-1)}{R}.$$

The left-hand side is at most $\sqrt{c} - 1$ by (12). First letting $R \to \infty$ with T fixed, and then letting $T \to \infty$, gives a contradiction. Hence $c \ge \rho_{n-1} = \rho_d$, proving the claimed bound.

The same lower bound holds for ordinary affine regression. Use the construction in Appendix C.1: choose d affinely independent feature points $x_1, \dots, x_d \in \mathbb{R}^{d-1}$, place one agent at each, and place d^2 additional agents at their centroid $\bar{x}$. The coefficient space is the full $\mathbb{R}^d$, including the intercept. Map a label vector $y \in \mathbb{R}^{d+1}$ of the auxiliary design $\widetilde{X}$ to the vertex reports $y_1, \dots, y_d$ and the common centroid report $u = -y_{d+1}/d$. As shown there, for every coefficient vector, the affine sum of squared residuals equals $\left\|\widetilde{X}\beta - y\right\|_2^2$. The optima and approximation ratios therefore agree.

To transfer absolute-loss SPIE, deviations of the first d auxiliary agents correspond to unilateral deviations of the vertex agents. For the last agent, fix true auxiliary label t and alternative report t', and put $u = -t/d$, $u' = -t'/d$. Change the d^2 centroid reports from u to u' one at a time. If Q_j is the random prediction at the centroid after j changes, individual truthfulness for the next agent, whose true value is u, gives

$$\mathbb{E}|Q_{j-1} - u| \le \mathbb{E}|Q_j - u| \quad (1 \le j \le d^2), \qquad |-dQ_j - t| = d|Q_j - u|.$$

All centroid agents receive the same prediction. Hence these inequalities chain, and multiplication by d gives the incentive inequality for the last auxiliary agent. Any affine mechanism with ratio below ρ_d would therefore induce an absolute-loss SPIE mechanism with that ratio on $\widetilde{X}$, whose column space is $\mathbf{1}^\perp$, contradicting the bound already proved. □

Where equality comes from. On the balanced design, $H = I - \mathbf{1}\mathbf{1}^\top/n$. Conditional on interpolating a set B of size b, the least-squares residual has value zero on B and $\tau/(n-b)$ elsewhere, where $\tau = \sum_i y_i$. For $\tau > 0$, its normalized version therefore has $A = 0$, $P = n - b$, and squared norm $1/(n-b)$. When d is even, our mechanism uses $P = (n+1)/2$; when d is odd, it mixes $P = n/2$ and $P = n/2 + 1$ equally. Thus the geometric bound in Lemma B.1 is tight on this design. Moreover, $F_R(t) = t$ and $2N_R(t) - n = 1$, so the incentive-growth calculation is also tight once the boundary error vanishes. This explains the exact parity-dependent factor.

C THE LOWER BOUND FOR SQUARED INDIVIDUAL LOSS

We first prove the squared-loss lower bound for the balanced prediction space and then transfer it to affine regression with d free coefficients, including the intercept. Randomized mechanisms are assumed measurable. The proof starts with an envelope formula for quadratic loss.

Lemma C.1 (Quadratic envelope formula)**.** *Fix the reports of all agents other than agent i. For a report $r \in \mathbb{R}$, let $Z(r)$ be the random prediction received by agent i, and suppose that $\mathbb{E}[Z(r)^2] < \infty$ for every $r \in \mathbb{R}$. Define $\mu(r) := \mathbb{E}[Z(r)]$ and $q(r) := \mathbb{E}[Z(r)^2]$.*

If the mechanism is strategyproof in expectation for quadratic individual loss, then μ is nondecreasing. Moreover, if $Z(a) = a$ almost surely for some $a \in \mathbb{R}$, then, for every $b \geq a$,

$$\mathbb{E}[(Z(b) - b)^2] = 2 \int_a^b \big(s - \mu(s)\big)\, ds.$$

Proof. For a true label t and an alternative report r, strategyproofness in expectation gives

$$q(t) - 2t\mu(t) \leq q(r) - 2t\mu(r). \tag{17}$$

Let $\Phi(t) := 2t\mu(t) - q(t)$. Equation (17), together with the fact that reporting $r = t$ is feasible, shows that

$$\Phi(t) = \sup_{r \in \mathbb{R}} \{2t\mu(r) - q(r)\}.$$

Thus Φ is a finite convex function. In addition, for all $r, t \in \mathbb{R}$,

$$\Phi(t) \geq \Phi(r) + 2(t - r)\mu(r), \tag{18}$$

so $2\mu(r)$ is a subgradient of Φ at r. Monotonicity of one-dimensional subgradients implies that μ is nondecreasing.

A finite convex function on $\mathbb{R}$ is absolutely continuous on every compact interval and differentiable almost everywhere. At every point s where Φ is differentiable, (18) implies $\Phi'(s) = 2\mu(s)$. Integrating gives

$$\Phi(b) - \Phi(a) = 2 \int_a^b \mu(s)\, ds.$$

Exactness at a gives $\mu(a) = a$, $q(a) = a^2$, and $\Phi(a) = a^2$. Finally, $\mathbb{E}[(Z(b) - b)^2] = b^2 - \Phi(b)$, so

$$\mathbb{E}[(Z(b) - b)^2] = b^2 - a^2 - 2 \int_a^b \mu(s)\, ds = 2 \int_a^b (s - \mu(s))\, ds. \qquad \square$$

We next prove the lower bound for a balanced d-coefficient prediction space. This is the auxiliary instance used in the reduction to affine regression.

Theorem C.2 (Balanced prediction-space lower bound)**.** *Let $n = d + 1$, and let $X \in \mathbb{R}^{n \times d}$ have full column rank with*

$$\mathrm{col}(X) = L := \left\{ z \in \mathbb{R}^n : \sum_{i=1}^n z_i = 0 \right\}.$$

Every randomized mechanism whose outputs lie in L and that is strategyproof in expectation for quadratic individual loss has squared-loss approximation ratio at least $d + 1$.

Proof. Suppose that such a mechanism has a finite approximation ratio c. For a report vector $y \in \mathbb{R}^n$, let $Z(y) \in L$ be its random prediction vector, and write $\mathrm{OPT}(y) := \min_{z \in L} \|z - y\|_2^2$.

If $\sigma(y) := \sum_i y_i$, orthogonal projection onto L gives $\mathrm{OPT}(y) = \sigma(y)^2/n$. In particular, if $y \in L$, then finite approximability forces $\mathbb{E}[\|Z(y) - y\|_2^2] = 0$, and therefore $Z(y) = y$ almost surely.

For each agent i, define $C_i(y) := \mathbb{E}[(Z_i(y) - y_i)^2]$ and $\mu_i(y) := \mathbb{E}[Z_i(y)]$.

Fix $\tau > 0$ and a report vector y with $\sigma(y) = \tau$. If agent i changes only their report from y_i to $y_i - \tau$, the resulting vector $y - \tau e_i$ belongs to L, so the mechanism is exact there. Applying Lemma C.1 while holding y_{-i} fixed, and then putting $s = r - (y_i - \tau)$, gives

$$C_i(y) = 2 \int_0^\tau \big(y_i - \tau + s - \mu_i\big(y - (\tau - s)e_i\big)\big)\, ds. \tag{19}$$

We average (19) over large balls in L. For $R > 0$, let $K_R := \{w \in L : \|w\|_2 \le R\}$, and let W_R be uniform on K_R with respect to the d-dimensional Lebesgue measure on L. Set $Y_R := W_R + (\tau/n)\mathbf{1}$; every realization of Y_R has coordinate sum τ and optimum τ^2/n. For $w \in L$ and $s \in [0, \tau]$, define

$$\Delta_i(w, s) := w_i + \frac{s}{n} - \mu_i\left(w + \frac{s}{n}\mathbf{1}\right), \qquad p_i := e_i - \frac{1}{n}\mathbf{1} \in L.$$

The profile on the path in (19) decomposes as

$$Y_R - (\tau - s)e_i = \big(W_R - (\tau - s)p_i\big) + \frac{s}{n}\mathbf{1}.$$

Consequently, the integrand in (19) is $\Delta_i(W_R - (\tau - s)p_i, s)$.

We now justify replacing the translated ball by the original ball. At a profile $v = w + (s/n)\mathbf{1}$, where $w \in L$, we have $\mathrm{OPT}(v) = s^2/n$. The approximation guarantee and Jensen's inequality give

$$|\Delta_i(w, s)| \le \sqrt{\mathbb{E}[(Z_i(v) - v_i)^2]} \le \sqrt{\frac{c}{n}}\, s. \tag{20}$$

Apply the ball-shift estimate (15) to $f(w) = \Delta_i(w, s)$ with shift $-(\tau - s)p_i$. By (20) and $\|p_i\|_2 \le 1$, for $R > \tau$ and $s \in [0, \tau]$,

$$|\mathbb{E}[\Delta_i(W_R - (\tau - s)p_i, s)] - \mathbb{E}[\Delta_i(W_R, s)]| \le \kappa_n \sqrt{\frac{c}{n}}\, \frac{s(\tau - s)}{R}.$$

By Fubini's theorem and (19),

$$\mathbb{E}[C_i(Y_R)] = 2\int_0^\tau \mathbb{E}[\Delta_i(W_R, s)]\, ds + \varepsilon_{i,R}, \qquad |\varepsilon_{i,R}| \le \kappa_n \sqrt{\frac{c}{n}}\, \frac{\tau^3}{3R}. \tag{21}$$

For fixed $w \in L$ and $s \in [0, \tau]$, let $v = w + (s/n)\mathbf{1}$. Every realization $Z(v)$ belongs to L, so $\sum_i \mu_i(v) = 0$. Since $\sum_i v_i = s$, we therefore have the pointwise formula

$$\sum_{i=1}^n \Delta_i(w, s) = s.$$

Summing (21) over the agents and letting $R \to \infty$ now yields

$$\lim_{R\to\infty} \mathbb{E}\left[\sum_{i=1}^n C_i(Y_R)\right] = 2\int_0^\tau s\, ds = \tau^2.$$

On the other hand, pointwise c-approximability and $\mathrm{OPT}(Y_R) = \tau^2/n$ imply

$$\mathbb{E}\left[\sum_{i=1}^n C_i(Y_R)\right] \le c\frac{\tau^2}{n}$$

for every R. Taking $R \to \infty$ gives $c \ge n = d + 1$. □

C.1 Reduction to ordinary affine regression

We finally transfer the balanced instance to ordinary affine regression.

Theorem C.3 (Affine regression with an intercept)**.** *For every $d \ge 1$, there is an unweighted public design for affine regression on $\mathbb{R}^{d-1}$, with hypothesis class*

$$f_\beta(x) = \beta_0 + \sum_{j=1}^{d-1} \beta_j x_j, \qquad \beta \in \mathbb{R}^d,$$

such that every randomized mechanism that is strategyproof in expectation for quadratic individual loss has squared-loss approximation ratio at least $d + 1$.

Proof. Choose d affinely independent points $x_1, \dots, x_d \in \mathbb{R}^{d-1}$, and let $\bar{x} := d^{-1} \sum_{i=1}^{d} x_i$. Write $a_i^\mathsf{T} := (1, x_i^\mathsf{T})$, let $A \in \mathbb{R}^{d \times d}$ have rows a_i^T, and put $\bar{a} := d^{-1} \sum_i a_i = (1, \bar{x}^\mathsf{T})^\mathsf{T}$. Affine independence makes A invertible. For $d = 1$, this construction is constant regression.

Consider first the auxiliary $(d+1)$-agent homogeneous design

$$\widetilde{X} := \begin{pmatrix} A \\ -d\bar{a}^\mathsf{T} \end{pmatrix} = \begin{pmatrix} I_d \\ -\mathbf{1}^\mathsf{T} \end{pmatrix} A.$$

It has d free coefficients and $\mathrm{col}(\widetilde{X}) = \mathbf{1}^\perp$. Hence Theorem C.2 applies to this design.

Now form an affine instance with one agent at each x_i, $i \in [d]$, and d^2 additional agents at $\bar{x}$. Suppose, toward a contradiction, that there is a mechanism M, strategyproof in expectation, with approximation ratio $c < d + 1$ for this fixed affine design. We use M to construct a mechanism $\widetilde{M}$ for the auxiliary design.

Given auxiliary reports $y = (y_1, \dots, y_{d+1})$, give the affine agents at x_i the reports y_i, and give every one of the d^2 agents at $\bar{x}$ the common report $u := -y_{d+1}/d$.

Run M on this affine report profile and let $\widetilde{M}$ return the same coefficient vector $\beta \in \mathbb{R}^d$. For every deterministic value of β, the auxiliary and affine sums of squared residuals agree:

$$\begin{aligned} \left\| \widetilde{X}\beta - y \right\|_2^2 &= \sum_{i=1}^{d} (a_i^\mathsf{T}\beta - y_i)^2 + (-d\bar{a}^\mathsf{T}\beta - y_{d+1})^2 \\ &= \sum_{i=1}^{d} (f_\beta(x_i) - y_i)^2 + d^2 \left(f_\beta(\bar{x}) + \frac{y_{d+1}}{d} \right)^2. \end{aligned}$$

The second line is exactly the affine social cost after the transformation of reports. Minimizing over β shows that the optima also agree. Therefore $\widetilde{M}$ inherits the approximation ratio c.

It remains to verify strategyproofness in expectation. For any of the first d auxiliary agents, a unilateral change of report changes only the report of the corresponding affine agent, and both agents have the same prediction and quadratic loss. Strategyproofness of M therefore gives the desired incentive inequality.

For auxiliary agent $d + 1$, fix a true label t and an alternative report t', and set $u := -t/d$ and $u' := -t'/d$. Enumerate the $m := d^2$ co-located agents. For $j = 0, \dots, m$, let ρ^j be the affine report profile in which the first j co-located agents report u', the remaining $m - j$ report u, and all other reports are fixed. Let P_j denote the random prediction at $\bar{x}$ when M is run on ρ^j.

Applying strategyproofness to agent j, with true label u and all other reports fixed, gives

$$\mathbb{E}[(P_{j-1} - u)^2] \le \mathbb{E}[(P_j - u)^2].$$

All co-located agents receive the same prediction, so these inequalities chain from $j = 1$ to m. Multiplying the resulting inequality by d^2 and using

$$(-dP_j - t)^2 = d^2(P_j - u)^2$$

shows precisely that auxiliary agent $d + 1$ cannot lower their expected quadratic loss by changing their report from t to t'. Thus $\widetilde{M}$ is strategyproof in expectation. Theorem C.2 then gives $c \ge d + 1$, a contradiction. □

D Randomization for general power losses

We prove Theorem 4.6 for the complete hypothesis class $\Theta = \mathbb{R}^d$. The proof uses two independent full-size volume-sampled interpolants. We first establish the moments in (9), then verify the incentives and squared-error guarantees for the two lotteries in Section 4.3.

D.1 Moments of the two interpolating fits

For a set $B \subseteq [n]$ of size d, let X_B be the corresponding square row submatrix. Full-size volume sampling draws B according to

$$\mathbb{P}(B) = \frac{\det(X_B)^2}{\det(X^\top X)}.$$

The probabilities sum to one by Cauchy–Binet. Singular sets have probability zero; on every supported set, define the interpolating prediction vector $U(y) = XX_B^{-1}y_B$. Define $V(y)$ from an independent set C drawn by the same law. Since $\det(H_B) = \det(X_B)^2/\det(X^\top X)$ when $|B| = d$, this is exactly (2) with $b = d$. The interpolation constraints then determine the fit uniquely, so $U = Z^B(y)$.

The following lemma is due to Dereziński & Warmuth (2017). We give a direct proof.

Lemma D.1 (Moments of a volume-sampled interpolant)**.** *Let $z^\star = Hy$ and $r = (I - H)y$. Then*

$$\mathbb{E}U = \mathbb{E}V = z^\star, \qquad \sigma^2 := \mathbb{E}\,\|U - z^\star\|_2^2 = \mathbb{E}\,\|V - z^\star\|_2^2 \le d\,\|r\|_2^2 = d\,\mathrm{OPT}(y).$$

Proof. For each sampled set B, put $u_B = X_B^{-1}r_B$, so $U - Hy = Xu_B$. For each $j \in [d]$, let $Y^{(j)}$ be the matrix obtained from X by replacing column j by r. Cramer's rule gives $\det(X_B)(u_B)_j = \det(Y_B^{(j)})$. Cauchy–Binet therefore gives

$$\begin{aligned}\mathbb{E}(u_B)_j &= \frac{1}{\det(X^\top X)} \sum_{|B|=d} \det(X_B)\det(Y_B^{(j)}) \\ &= \frac{\det(X^\top Y^{(j)})}{\det(X^\top X)} = 0.\end{aligned}$$

Singular sets contribute zero to the sum, and the final determinant vanishes because its j-th column is $X^\top r = 0$. Hence $\mathbb{E}U = Hy$.

For the second moment, $U - Hy \in L$ and $r = y - Hy \perp L$, so

$$\mathbb{E}\,\|U - y\|_2^2 = \|r\|_2^2 + \mathbb{E}\,\|U - Hy\|_2^2.$$

Applying Lemma 4.1 with $b = d$ bounds the left-hand side by $(d+1)\,\|r\|_2^2$, giving $\mathbb{E}\,\|U - Hy\|_2^2 \le d\,\|r\|_2^2$. The independent copy V has the same moments. □

D.2 The convex range $p \ge 1$

Let $\kappa = 1 + \sqrt{2}$, and let λ_p be as in Theorem 4.6. Put

$$a = \frac{\sqrt{2}-1}{2} = \frac{1}{2\kappa}, \qquad b = \frac{\sqrt{2}+1}{2} = \frac{\kappa}{2}.$$

The three-outcome lottery in (10) can be written $Z = (1-T)U + TV$, where $T = 1/2$ with probability $1 - \lambda_p$, and $T = -a$ and $T = b$ each with probability $\lambda_p/2$. Draw T independently of the sampled sets and reports. Notice that $b = 1 + a$, so T and $1 - T$ have the same distribution.

For $x \ne 0$, define $\psi_p(x) = \mathrm{sgn}(x)|x|^{p-1}$. Direct substitution gives the incentive balance

$$\mathbb{E}[(1-T)\psi_p(T)] = (1-\lambda_p)2^{-p} - \frac{\lambda_p}{2}\left(ba^{p-1} + ab^{p-1}\right) = 0, \tag{22}$$

because

$$2^{p-1}\left(ba^{p-1} + ab^{p-1}\right) = \frac{\kappa^{p-2} + \kappa^{2-p}}{2} = \frac{1-\lambda_p}{\lambda_p}.$$

Lemma D.2 (Incentives for the convex-power lottery)**.** *For every $p \ge 1$, the corresponding lottery in (10) is SPIE for individual loss $|z_i - y_i|^p$.*

Proof. Condition on the sampled sets B, C, fix agent i, and fix all other reports. If $i \notin B \cup C$, their report affects neither fit. If $i \in B \cap C$, both fits interpolate their report, and hence every lottery outcome does too; truthful reporting gives zero loss.

Suppose $i \in B \setminus C$. Let t be their true label and let v be the prediction at i of the fit based on C, which is independent of their report. If they report $t + \delta$, their conditional expected loss is

$$F(\delta) = \mathbb{E}_T \left| T(v - t) + (1 - T)\delta \right|^p .$$

This is a convex function of δ. If $v = t$, then $F(0) = 0$. Otherwise, $T(v-t) \neq 0$ for every outcome, so F is differentiable at zero even when $p = 1$, and

$$F'(0) = p\,\psi_p(v - t)\,\mathbb{E}[(1 - T)\psi_p(T)] = 0$$

by (22). Convexity makes $\delta = 0$ a global minimizer. The case $i \in C \setminus B$ follows from the equality in law of T and $1 - T$. The incentive inequality therefore holds for every fixed pair (B, C), and averaging over their report-independent distribution proves SPIE. □

Lemma D.3 (Squared error of the convex-power lottery)**.** *For $p \geq 1$, the expected squared error is at most $[1 + d(1/2 + \lambda_p)]\,\mathrm{OPT}(y)$.*

Proof. Let $z^\star = Hy$, $\xi = U - z^\star$, and $\eta = V - z^\star$. By Lemma D.1, ξ and η are independent, mean-zero, and have the same second moment σ^2. They lie in L, while $y - z^\star \perp L$. Since T is independent of both fits,

$$\begin{aligned} \mathbb{E}\,\|Z - y\|_2^2 &= \mathrm{OPT}(y) + \mathbb{E}\,\|(1 - T)\xi + T\eta\|_2^2 \\ &= \mathrm{OPT}(y) + \mathbb{E}\big[(1 - T)^2 + T^2\big]\sigma^2. \end{aligned} \tag{23}$$

The cross term vanishes because $\mathbb{E}\langle \xi, \eta \rangle = \langle \mathbb{E}\xi, \mathbb{E}\eta \rangle = 0$. At $T = 1/2$, the multiplier is $1/2$; at either $T = -a$ or $T = b$, it is $a^2 + b^2 = 3/2$. Consequently

$$\mathbb{E}\big[(1 - T)^2 + T^2\big] = (1 - \lambda_p)\frac{1}{2} + \lambda_p \frac{3}{2} = \frac{1}{2} + \lambda_p.$$

The result follows from $\sigma^2 \leq d\,\mathrm{OPT}(y)$. □

D.3 Powers below one: $0 < p < 1$

Set $w_p = (2 + 2^p)^{-1}$. Independently of the sampled sets and reports, return U and V with probability w_p each, and their midpoint with probability $2^p w_p$.

Lemma D.4 (Sublinear-power incentives and efficiency)**.** *For $0 < p < 1$, this lottery is SPIE for individual loss $|z_i - y_i|^p$, and its expected squared error is at most*

$$\left[1 + d\frac{2 + 2^{p-1}}{2 + 2^p}\right] \mathrm{OPT}(y).$$

Proof. Condition again on (B, C). The cases where agent i belongs to neither set or to both sets are immediate. If $i \in B \setminus C$, use the notation from the proof of Lemma D.2 and set $s = v - t$. After a report $t + \delta$, their expected loss is

$$F(\delta) = w_p\big(|\delta|^p + |s|^p + |s + \delta|^p\big).$$

For $0 < p < 1$, the triangle inequality followed by subadditivity of $x \mapsto x^p$ on $\mathbb{R}_+$ gives

$$|s|^p \leq |s + \delta|^p + |\delta|^p.$$

Hence $F(\delta) \geq 2w_p|s|^p = F(0)$. The case $i \in C \setminus B$ is symmetric, so the mechanism is SPIE.

For the squared-error calculation, write the lottery as $Z = (1 - T)U + TV$, with $T = 0, 1, 1/2$ having probabilities $w_p, w_p, 2^p w_p$, respectively. The variance decomposition (23) remains valid, and

$$\mathbb{E}[(1 - T)^2 + T^2] = 2w_p + 2^{p-1}w_p = \frac{2 + 2^{p-1}}{2 + 2^p}.$$

Apply Lemma D.1 to finish the proof. □

Proof of Theorem 4.6. For $p \geq 1$, combine Lemma D.2 and Lemma D.3; for $0 < p < 1$, use Lemma D.4. These give exactly the factor $1 + dq_p$ in (8).

Each outcome is a linear combination of U and V, so it belongs to $L = \mathrm{col}(X)$ and is feasible for the complete hypothesis class $\Theta = \mathbb{R}^d$.

Finally, when $0 < p < 1$, the numerator defining q_p is strictly smaller than its denominator. When $p \geq 1$, the arithmetic–geometric mean inequality gives $\kappa^{p-2} + \kappa^{2-p} \geq 2$, with equality exactly at $p = 2$. Therefore $\lambda_p \leq 1/2$, again with equality exactly at $p = 2$, so $q_2 = 1$ and $q_p < 1$ for every $p \neq 2$. □

Behavior near quadratic loss. For our lottery with $p \geq 1$, putting $x = \kappa^{p-2}$ gives

$$\lambda_p = \frac{2x}{(x+1)^2}, \qquad 1 + dq_p = d + 1 - \frac{d}{2}\left(\frac{x-1}{x+1}\right)^2.$$

The improvement in this bound vanishes quadratically as $p \to 2$:

$$\frac{d}{2}\left(\frac{\kappa^{p-2}-1}{\kappa^{p-2}+1}\right)^2 = \frac{d\log^2\kappa}{8}(p-2)^2 + O\big(d(p-2)^4\big).$$