Scale-Independent Robust Multivariate Polynomial Regression

Abstract

We resolve the sample-versus-scale tradeoff posed by Arora, Bhattacharyya, Boban, Guruswami, and Kelman [1, Remark 1.9] in robust multivariate polynomial regression, giving scale-independent recovery with optimal degree exponents in an exact-real LP-oracle model. The unknown polynomial p has degree at most d in each of n coordinates. Covariates are i.i.d. product-Chebyshev or uniform, and outlier indicators are mutually independent, independent of the covariates, and have probabilities at most a known $\rho < 1/2$. Inlier errors are bounded by an unknown σ; admissible labels may otherwise be adversarial. For $0 < \delta \leq \varepsilon \leq 1/2$, our algorithm returns $\widehat{p}$ with $\|\widehat{p} - p\|_\infty \leq (2 + \varepsilon)\sigma$ with probability at least $1 - \delta$, using $O_{n,\rho}((d/\varepsilon)^n \log(d/\delta))$ product-Chebyshev samples or $O_{n,\rho}((d/\varepsilon)^{2n} \log(d/\delta))$ uniform samples. It does not use σ, recovers p exactly at zero noise, and makes a predetermined $O_{n,\rho}(1 + \log(d/\varepsilon))$ number of LP calls independent of signal magnitude. For fixed dimension, accuracy, positive contamination rate, and success probability, the Chebyshev rate matches the known dependence on d, while the uniform rate has the optimal degree exponent up to an occupancy logarithm. The key ingredient is an arbitrary-representative product-Chebyshev L^1 norming inequality, which enables a Chebyshev-mass-weighted least-absolute-deviations initializer followed by fixed-budget median refinements.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction and main result

Robust polynomial regression asks whether a bounded-degree polynomial can be recovered in supremum norm when a minority of responses are arbitrary. We formalize this question under product-Chebyshev and uniform random designs, compare with the prior sample-versus-scale tradeoff, and state our guarantee.

Fix integers $n, d \geq 1$, and let

$$\mathcal{P}_{d,n} := \left\{ q(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in \{0,\ldots,d\}^n} c_{\boldsymbol{a}} \boldsymbol{x}^{\boldsymbol{a}} : c_{\boldsymbol{a}} \in \mathbb{R} \right\}.$$

Here $\boldsymbol{x}^{\boldsymbol{a}} := \prod_{k=1}^n x_k^{a_k}$, so the degree is at most d in each coordinate and $\dim \mathcal{P}_{d,n} = (d+1)^n$. The two design laws are the normalized uniform probability measure $d\mu_U(\boldsymbol{x}) = 2^{-n} d\boldsymbol{x}$ and the product-Chebyshev probability measure

$$d\nu(\boldsymbol{x}) = \prod_{k=1}^n \frac{dx_k}{\pi\sqrt{1 - x_k^2}}. \tag{1.1}$$

An unknown $p \in \mathcal{P}_{d,n}$ is observed through a batch $(X_i, Y_i)_{i=1}^M \in ([-1,1]^n \times \mathbb{R})^M$, where the X_i are i.i.d. from one of these laws. Let $O_i \in \{0, 1\}$ be mutually independent outlier indicators, independent of $X_{1:M}$, with $\mathbb{P}(O_i = 1) \le \rho$ for every i, where $0 \le \rho < 1/2$. If $O_i = 0$, then

$$|Y_i - p(X_i)| \le \sigma.$$

Conditional on $(X_{1:M}, O_{1:M})$, the inlier errors and outlier labels may be chosen arbitrarily, and may depend on p, subject only to this inlier bound. The algorithm observes the batch and is given $n, d, \varepsilon, \delta$, the design law, and the valid upper bound ρ, but it does not observe the O_i and is not given p or σ. Every probability statement is over $(X_{1:M}, O_{1:M})$ and holds uniformly over all admissible conditional choices of the labels. We write $\|q\|_\infty := \sup_{\boldsymbol{x}\in[-1,1]^n} |q(\boldsymbol{x})|$.

The unresolved tradeoff in Arora et al. [1] is summarized below. Their Theorem 3.1 (ESA version) achieves the target degree exponents but returns error $(2+\varepsilon)\sigma + \eta$ for $\eta > 0$, with a call bound that can be written uniformly as $O(1 + \log_{1/\varepsilon}(1 + \|p\|_\infty/\eta))$. Their randomized scale-independent result, ESA Theorem 5.1, corresponds to Theorem 6.1 of arXiv:2403.09465v1; its deterministic core, ESA Theorem 5.2, corresponds to v1 Theorem 6.4.

Table 1: The sample-versus-scale tradeoff. The grid has m^n cells; its least cell mass determines the sample count under each design.

Method	Grid scale m	L^∞ error	Optimization dependence
Arora et al., ESA Thm. 3.1	$m = O_n(d/\varepsilon)$	$(2+\varepsilon)\sigma + \eta$	$O(1 + \log_{1/\varepsilon}(1 + \|p\|_\infty/\eta))$
Arora et al., ESA Thm. 5.1	$m = \Omega_n(d^{2n+1}/\varepsilon)$	$(2+\varepsilon)\sigma$	independent of signal scale
This work	$m = O_{n,\rho}(d/\varepsilon)$	$(2+\varepsilon)\sigma$	at most $1 + T$ LP-oracle calls

Remark 1.9 of the ESA version (Remark 1.6 of arXiv v1) asks whether the optimal-degree sample rates and the pure $(2+\varepsilon)\sigma$ error can be combined without scale-dependent refinement. The improvement requires a different weighting and norming primitive, not merely a retuning of constants.

We work in an *exact-real LP-oracle model*. Real arithmetic and comparisons, exact access to and comparison with the prescribed algebraic Chebyshev endpoints, and calls to an exact LP optimizer are primitives. Any optimizer returned by the oracle is admissible. We count oracle calls, LP combinatorial description size, and arithmetic work outside the oracle separately; we make no end-to-end finite-precision or bit-complexity claim for arbitrary real-valued continuous-design inputs.

The following theorem resolves the stated tradeoff in this model. The algorithm and the predetermined refinement budget are specified in Section 2.

Theorem 1.1 (Main theorem)**.** *Fix integers $n, d \ge 1$, $0 \le \rho < 1/2$, $0 < \delta \le \varepsilon \le 1/2$, and $\sigma \ge 0$. Under the statistical model above, there is an algorithm that is not given σ and has the following guarantees.*

1. *Under product-Chebyshev locations it uses*

$$M_C = O_{n,\rho}((d/\varepsilon)^n \log(d/\delta))$$

samples. Under normalized uniform locations it uses

$$M_U = O_{n,\rho}\big((d/\varepsilon)^{2n} \log(d/\delta)\big)$$

samples.

2. With probability at least $1-\delta$, it outputs $\widehat{p} \in \mathcal{P}_{d,n}$ satisfying

$$\|\widehat{p} - p\|_\infty \le (2+\varepsilon)\sigma.$$

In particular, when $\sigma = 0$, the same event gives $\widehat{p} = p$ identically.

3. On every input, the algorithm makes at most $1+T$ exact-LP oracle calls, where the predetermined integer $T = O_{n,\rho}(\log(d/\varepsilon))$ is independent of p and σ. On the no-empty-cell branch it makes exactly $1+T$ calls. For fixed n, ρ, the LP descriptions and all work outside the oracle calls are polynomial in the corresponding sample count $M = M_C$ or $M = M_U$.

Here optimality refers only to the dependence on d for fixed dimension, fixed accuracy, fixed positive contamination rate, and constant success probability. We do not claim jointly optimal dependence on $n, \varepsilon, \delta, \rho$; under uniform sampling the upper bound retains the displayed occupancy logarithm. The case $d = 0$ is recorded in Appendix A.

Proof overview. Use the Chebyshev partition for both location laws, but assign every cell its product-Chebyshev mass m^{-n} in the LAD objective. An arbitrary-point $L^1(\nu)$ norming lemma turns the strict inlier majority in each cell into a coercive LAD lower bound with a positive L^1 coefficient. Its multivariate proof tensorizes the reproducing kernel and cell-supremum sum, not a representative-selection operator: each representative coordinate may depend on the full cell index. This yields $\left\|p^{(0)} - p\right\|_\infty \le O_{n,\rho}(d^n\sigma)$. A fixed number of median refinements then reaches $(2+\varepsilon)\sigma$. Randomness is used only to ensure that every cell is occupied and has an inlier majority; the deterministic argument thereafter applies simultaneously to every iterate.

2 Partition and algorithm

We now specify the common product grid, weighted-LAD initializer, and fixed-budget median refinement used for both designs. The design law affects only the batch size; the deterministic optimization steps are shared.

For a positive integer r, write $[r] := \{1, \dots, r\}$. For a positive integer m, partition $[0, \pi]$ into intervals $J_j = [(j-1)\pi/m, j\pi/m]$, $j \in [m]$. Under $x = \cos\theta$, these induce the Chebyshev intervals

$$I_j = \left[\cos\frac{j\pi}{m}, \cos\frac{(j-1)\pi}{m}\right]$$

and the product cells $C_{\boldsymbol{j}} = I_{j_1} \times \cdots \times I_{j_n}$, indexed by $\boldsymbol{j} \in [m]^n$. Let $\mathcal{C}_m := \{C_{\boldsymbol{j}} : \boldsymbol{j} \in [m]^n\}$. Closed cells are used in the analysis. A fixed half-open convention assigns any location on a common boundary to one cell, so the algorithm is defined everywhere. Every cell in $\mathcal{C}_m$ has ν-mass m^{-n}.

Set

$$\alpha = \frac{1+2\rho}{4}, \qquad g = 1 - 2\alpha = \frac{1-2\rho}{2}, \qquad \zeta = \alpha - \rho = \frac{g}{2}, \qquad \beta = \frac{\varepsilon}{12}. \tag{2.1}$$

Thus $1/4 \le \alpha < 1/2$, $0 < g \le 1/2$, $0 < \zeta \le 1/4$, and $0 < \beta \le 1/24$. Choose once and for all a constant $A_n \ge 1$, depending only on n, as guaranteed by Corollary 3.4, and take m to be the least positive integer satisfying

$$m \ge \max\left\{1, \frac{A_n d}{\zeta}, \frac{6\pi n d}{\beta}\right\}. \tag{2.2}$$

Since m is the least such integer and $0 < \varepsilon \le 1/2$,

$$m = \Theta_{n,\rho}(d/\varepsilon). \tag{2.3}$$

Set

$$\gamma := \frac{\pi n d}{m}. \tag{2.4}$$

The choice of m gives $\gamma \le \beta/6 < 1$. Use

$$M = \begin{cases} \lceil C_\rho m^n \log(2m^n/\delta) \rceil, & \text{for product-Chebyshev locations,} \\ \lceil C_\rho m^{2n} \log(2m^n/\delta) \rceil, & \text{for uniform locations,} \end{cases} \tag{2.5}$$

where $C_\rho := 16/(1-2\rho)^2$, as justified in Lemma 6.1. Thus $M = M_C$ for the product-Chebyshev design and $M = M_U$ for the uniform design.

For $C \in \mathcal{C}_m$, let $S_C = \{i : X_i \in C\}$, with membership interpreted according to the fixed boundary convention, and let $N_C = |S_C|$. For the analysis, also set

$$B_C := \{i \in S_C : O_i = 1\}, \qquad G_C := S_C \setminus B_C.$$

The algorithm does not observe these two sets. Call the sample set *good* if every N_C is positive and $|B_C|/N_C \le \alpha$ for every $C \in \mathcal{C}_m$.

The algorithm first checks for an empty cell and, if one exists, returns $\widehat{p} := 0$ without calling the LP oracle. Otherwise it defines the weighted LAD initializer by

$$p^{(0)} \in \arg\min_{q \in \mathcal{P}_{d,n}} \mathcal{L}(q), \qquad \mathcal{L}(q) := \frac{1}{m^n} \sum_{C \in \mathcal{C}_m} \frac{1}{N_C} \sum_{i \in S_C} |q(X_i) - Y_i|. \tag{2.6}$$

The cell weight is m^{-n}, regardless of the location law.

For each cell fix a representative $a_C \in C$, for instance the observed location of its smallest-index sample. Let T be the least nonnegative integer such that

$$\beta^T \frac{4(2d+1)^n}{g} \le \frac{\varepsilon}{2}. \tag{2.7}$$

For $t = 0, \dots, T-1$, let $\widetilde{Y}_{C,t}$ be any median of the residual labels

$$\{Y_i - p^{(t)}(X_i) : i \in S_C\}.$$

Here a median is any number for which at least half the observations lie on each weak side. Compute

$$r_t \in \arg\min_{r \in \mathcal{P}_{d,n}} \max_{C \in \mathcal{C}_m} |r(a_C) - \widetilde{Y}_{C,t}|, \qquad p^{(t+1)} = p^{(t)} + r_t. \tag{2.8}$$

On this branch it returns $\widehat{p} := p^{(T)}$. Both objectives attain their minima, and any optimizer returned by the LP oracle is admissible: Corollary 3.4 makes the full-sample evaluation functional a norm, while Lemma 5.1 with $\gamma < 1$ does the same for representative evaluation. Details and LP sizes appear in Proposition 6.2. Consequently every input uses at most $1 + T$ LP-oracle calls, and the no-empty-cell branch uses exactly $1 + T$.

3 Product-Chebyshev L^1 norming with arbitrary cell representatives

This section proves the deterministic norming inequality underlying the initializer. We lift to the torus, establish localized kernel bounds, and pass from arbitrary representatives to arbitrary finite multisets. The proofs are included here because this norming chain is the main new analytic ingredient.

For $q \in \mathcal{P}_{d,n}$, write

$$Q(\boldsymbol{\theta}) := q(\cos\theta_1, \ldots, \cos\theta_n). \tag{3.1}$$

Then Q is an even trigonometric polynomial of degree at most d in each coordinate. Set $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$ and let $d\mu_{\mathbb{T}}(t) = dt/(2\pi)$ be normalized Haar measure; product torus norms use $\mu_{\mathbb{T}}^{\otimes n}$. We use the conventions

$$\widehat{f}(\boldsymbol{k}) = \int_{\mathbb{T}^n} f(\boldsymbol{\theta}) e^{-i\boldsymbol{k}\cdot\boldsymbol{\theta}}\, d\mu_{\mathbb{T}}^{\otimes n}(\boldsymbol{\theta}), \qquad (f * g)(\boldsymbol{\theta}) = \int_{\mathbb{T}^n} f(\boldsymbol{\theta} - \boldsymbol{u}) g(\boldsymbol{u})\, d\mu_{\mathbb{T}}^{\otimes n}(\boldsymbol{u}).$$

After even 2π-periodic extension, this normalization gives the exact identity

$$\|Q\|_{L^1(\mathbb{T}^n)} = \frac{1}{\pi^n} \int_{[0,\pi]^n} |Q(\boldsymbol{\theta})|\, d\boldsymbol{\theta} = \|q\|_{L^1(\nu)}\,. \tag{3.2}$$

We first construct the localized reproducing kernel used in the arbitrary-representative estimate.

Lemma 3.1 (One-dimensional kernel bounds)**.** *For every integer $N \geq 1$, let*

$$F_N(t) = \frac{1}{N}\left(\frac{\sin(Nt/2)}{\sin(t/2)}\right)^2 = \sum_{|k|<N}\left(1 - \frac{|k|}{N}\right) e^{ikt}$$

be the Fejér kernel, with the quotient defined by continuous extension at $2\pi\mathbb{Z}$, so $F_N(2\pi\ell) = N$. For every integer $d \geq 1$, set $V_d = 2F_{2d} - F_d$. Then

$$\widehat{V_d}(k) = 1 \quad (|k| \leq d), \qquad \|V_d\|_1 \leq 3, \qquad \left\|V_d'\right\|_1 \leq Cd \tag{3.3}$$

for an absolute constant C.

Proof. We first verify reproduction from the Fourier multipliers and then estimate the derivative directly in the spatial variable. For $|k| < d$, the multiplier of V_d is

$$2\left(1 - \frac{|k|}{2d}\right) - \left(1 - \frac{|k|}{d}\right) = 1,$$

and for $|k| = d$ it is $2(1 - 1/2) = 1$. This proves reproduction on all frequencies $|k| \leq d$, including the endpoints. Since $F_N \geq 0$ and $\|F_N\|_1 = 1$, the triangle inequality gives $\|V_d\|_1 \leq 3$.

It remains to prove the L^1 derivative bound without a logarithmic loss. On $|t| \leq N^{-1}$, differentiating the finite Fourier expansion gives

$$|F_N'(t)| \leq \sum_{|k|<N} |k| \left(1 - \frac{|k|}{N}\right) \leq CN^2.$$

For $N^{-1} \leq |t| \leq \pi$, differentiating the sine quotient and using $|\sin(t/2)| \geq |t|/\pi$ gives

$$|F_N'(t)| \leq C\left(|t|^{-2} + N^{-1}|t|^{-3}\right).$$

Both regions contribute $O(N)$ after integration, so $\|F_N'\|_1 = O(N)$. Applying this estimate to $V_d = 2F_{2d} - F_d$ completes the proof. $\square$

Lemma 3.2 (Local supremum and endpoint Bernstein bounds)**.** *There exists $B_n < \infty$, depending only on n, with the following property. Let d, m be integers with $m \geq d \geq 1$, set $h = \pi/m$, and let H be a trigonometric polynomial on $\mathbb{T}^n$ of degree at most d in each coordinate. If $\mathcal{R}_m$ is the product partition of $\mathbb{T}^n$ with $2m$ intervals of length h in each coordinate, then*

$$\left(\frac{h}{2\pi}\right)^n \sum_{R \in \mathcal{R}_m} \sup_R |H| \leq B_n \|H\|_{L^1(\mathbb{T}^n)}, \tag{3.4}$$

$$\|\partial_k H\|_{L^1(\mathbb{T}^n)} \leq B_n d \|H\|_{L^1(\mathbb{T}^n)} \qquad (k \in [n]). \tag{3.5}$$

Proof. We first obtain a shift-uniform one-dimensional cell-supremum bound, then tensorize it, and finally differentiate one kernel factor. For every absolutely continuous periodic function v and every interval I of length h,

$$h \sup_I |v| \leq \int_I |v(t)|\, dt + h \int_I |v'(t)|\, dt.$$

Summing over the $2m$ torus intervals and applying Lemma 3.1 shows, uniformly in u, that

$$\frac{h}{2\pi} \sum_I \sup_{t \in I} |V_d(t-u)| \leq \|V_d\|_1 + h \left\|V_d'\right\|_1 \leq C_0, \tag{3.6}$$

where C_0 is an absolute constant and we used $hd \leq \pi$.

Let $W_d(\boldsymbol{\theta}) = \prod_{k=1}^n V_d(\theta_k)$. The multiplier identity gives $H = H * W_d$. Hence Tonelli's theorem implies

$$\begin{aligned} &\left(\frac{h}{2\pi}\right)^n \sum_{R \in \mathcal{R}_m} \sup_{\boldsymbol{\theta} \in R} |H(\boldsymbol{\theta})| \\ &\leq \int_{\mathbb{T}^n} |H(\boldsymbol{u})| \left[\left(\frac{h}{2\pi}\right)^n \sum_{R \in \mathcal{R}_m} \sup_{\boldsymbol{\theta} \in R} |W_d(\boldsymbol{\theta} - \boldsymbol{u})|\right] d\mu_{\mathbb{T}}^{\otimes n}(\boldsymbol{u}). \end{aligned}$$

The bracket factors by coordinate, and each factor is bounded by C_0 in (3.6). This proves (3.4) with a constant depending only on n. Notice that the factorization applies to the kernel, not to a rule selecting representatives.

Similarly,

$$\partial_k H = H * \left(V_d' \otimes \bigotimes_{\ell \neq k} V_d\right).$$

Young's inequality and Lemma 3.1 give $\|\partial_k H\|_1 \leq C d\, 3^{n-1} \|H\|_1$. Enlarging B_n, if necessary, proves (3.5) simultaneously with (3.4). □

Lemma 3.3 (Arbitrary representatives)**.** *There exists $A_n \geq 1$, depending only on n, such that the following holds for all integers $m \geq d \geq 1$. For every $q \in \mathcal{P}_{d,n}$, choose an arbitrary representative $\xi_C \in C$ for each $C \in \mathcal{C}_m$, and define $q^\sharp$ on the fixed half-open partition to be constant on C, with value $q(\xi_C)$. Then*

$$\left\|q - q^\sharp\right\|_{L^1(\nu)} \leq A_n \frac{d}{m} \|q\|_{L^1(\nu)}. \tag{3.7}$$

The representatives may be adversarial; in particular, their coordinates may depend jointly on the full cell index.

Proof. We transfer the representatives to angular coordinates, reflect them to the torus, and integrate a coordinatewise path estimate. For $C = C_{\boldsymbol{j}}$, map its representative to

$$\boldsymbol{\eta}_C := (\arccos \xi_{C,1}, \dots, \arccos \xi_{C,n}) \in J_{j_1} \times \cdots \times J_{j_n}.$$

Reflect the angular partition and these representatives evenly to $\mathbb{T}^n$. On each reflected product interval R, let $Q^\sharp$ have the constant value $Q(\boldsymbol{\eta}_R)$. For $\boldsymbol{\theta} \in R$, a coordinatewise line-segment path within R gives

$$|Q(\boldsymbol{\theta}) - Q(\boldsymbol{\eta}_R)| \le h \sum_{k=1}^{n} \sup_R |\partial_k Q|.$$

Integrating over all cells, using (3.4) for each $\partial_k Q$, and then applying (3.5), we obtain

$$\begin{aligned} \left\| Q - Q^\sharp \right\|_1 &\le h \sum_{k=1}^{n} \left(\frac{h}{2\pi} \right)^n \sum_{R \in \mathcal{R}_m} \sup_R |\partial_k Q| \\ &\le B_n h \sum_{k=1}^{n} \|\partial_k Q\|_1 \le \pi n B_n^2 \frac{d}{m} \|Q\|_1 . \end{aligned}$$

Choose $A_n \ge \max\{1, \pi n B_n^2\}$. Up to measure-zero boundaries, $\mathbb{T}^n$ decomposes into 2^n reflected copies of $[0, \pi]^n$; hence the normalized identity (3.2) applies equally to $Q - Q^\sharp$ and $q - q^\sharp$. This proves (3.7). □

Corollary 3.4 (Arbitrary multisets in every cell)**.** *Let A_n be as in Lemma 3.3. For every $0 < \zeta < 1$, every integer $m \ge A_n d/\zeta$, and every family $A = (A_C)_{C \in \mathcal{C}_m}$ of nonempty finite multisets $A_C \subset C$, define*

$$\Phi_A(q) := \frac{1}{m^n} \sum_{C \in \mathcal{C}_m} \frac{1}{|A_C|} \sum_{z \in A_C} |q(z)|.$$

Here cardinalities and sums count multiplicity. Simultaneously for all $q \in \mathcal{P}_{d,n}$ and all such families,

$$(1 - \zeta) \|q\|_{L^1(\nu)} \le \Phi_A(q) \le (1 + \zeta) \|q\|_{L^1(\nu)} . \tag{3.8}$$

Proof. For a fixed cell, the average of $|q|$ over A_C lies between the minimum and maximum of $|q|$ on that cell. Since the cell is connected and $|q|$ is continuous, the average equals $|q(\xi_C)|$ for some $\xi_C \in C$. Because each cell has ν-mass m^{-n}, these representatives satisfy

$$\Phi_A(q) = \frac{1}{m^n} \sum_{C \in \mathcal{C}_m} |q(\xi_C)| = \left\| q^\sharp \right\|_{L^1(\nu)} .$$

Thus $\left| \Phi_A(q) - \|q\|_{L^1(\nu)} \right| \le \left\| q^\sharp - q \right\|_{L^1(\nu)}$, and Lemma 3.3 proves (3.8). □

Remark 3.5 (Why a sufficiently fine grid is necessary). In one dimension, take $m = d$ and let T_d be the first-kind Chebyshev polynomial, defined by $T_d(\cos\theta) = \cos(d\theta)$. For $j \in [d]$, let $\theta_j = (j - 1/2)\pi/d$. The point $\xi_j = \cos\theta_j$ belongs to I_j and satisfies $T_d(\xi_j) = 0$. Choosing these representatives makes every discrete value vanish although $T_d \not\equiv 0$. Thus the constant oversampling required for a relative norming estimate cannot be replaced by the bare condition $m = d$.

4 Signal-scale-independent weighted-LAD initialization

On the good event, arbitrary-multiset norming turns cellwise inlier majorities into a coercive LAD lower bound. An L^1-to-L^∞ inequality then gives an initializer whose tuning does not use the unknown signal or noise scale.

Proposition 4.1 (Coercive LAD bound and initialization)**.** *On a good sample set, for every $e \in \mathcal{P}_{d,n}$,*

$$\mathcal{L}(p+e) - \mathcal{L}(p) \geq \frac{g}{2} \|e\|_{L^1(\nu)} - 2\sigma. \tag{4.1}$$

Consequently, $\mathcal{L}$ is coercive and attains its minimum. Every minimizer $p^{(0)}$ satisfies

$$\left\|p^{(0)} - p\right\|_{L^1(\nu)} \leq \frac{4\sigma}{g}. \tag{4.2}$$

Combining this with the L^1-to-L^∞ bound below gives

$$E_0 := \left\|p^{(0)} - p\right\|_\infty \leq \frac{4(2d+1)^n}{g}\,\sigma. \tag{4.3}$$

When $\sigma = 0$, every LAD minimizer equals p identically.

Proof. We compare the objective at $p+e$ and p, apply Corollary 3.4 separately to the inlier and outlier locations, and then convert the resulting $L^1(\nu)$ bound to L^∞. Fix $e \in \mathcal{P}_{d,n}$. For an inlier i, the reverse triangle inequality gives

$$|p(X_i) + e(X_i) - Y_i| - |p(X_i) - Y_i| \geq |e(X_i)| - 2\sigma.$$

For an outlier, with no restriction on its label,

$$|p(X_i) + e(X_i) - Y_i| - |p(X_i) - Y_i| \geq -|e(X_i)|.$$

The accumulated noise term is at most

$$\frac{2\sigma}{m^n} \sum_{C \in \mathcal{C}_m} \frac{|G_C|}{N_C} \leq 2\sigma.$$

Thus summing the samplewise inequalities gives

$$\mathcal{L}(p+e) - \mathcal{L}(p) \geq \frac{1}{m^n} \sum_{C \in \mathcal{C}_m} \frac{1}{N_C} \left(\sum_{i \in G_C} |e(X_i)| - \sum_{i \in B_C} |e(X_i)| \right) - 2\sigma. \tag{4.4}$$

Every G_C is nonempty and $|G_C|/N_C \geq 1 - \alpha$. Since $N_C^{-1} = (|G_C|/N_C)|G_C|^{-1}$, applying Corollary 3.4 to the family of inlier-location multisets gives

$$\frac{1}{m^n} \sum_{C \in \mathcal{C}_m} \frac{1}{N_C} \sum_{i \in G_C} |e(X_i)| \geq (1-\alpha)(1-\zeta) \|e\|_{L^1(\nu)}. \tag{4.5}$$

For each cell, let A_C be the multiset of its outlier locations when $B_C \neq \varnothing$, and otherwise let A_C consist of one arbitrary dummy point in C. Cellwise,

$$\frac{1}{N_C} \sum_{i \in B_C} |e(X_i)| \leq \alpha \frac{1}{|A_C|} \sum_{z \in A_C} |e(z)| :$$

for a nonempty B_C this factors out $|B_C|/N_C \le \alpha$, while for an empty B_C the left side is zero. Applying Corollary 3.4 to the A_C's yields

$$\frac{1}{m^n}\sum_{C\in\mathcal{C}_m}\frac{1}{N_C}\sum_{i\in B_C}|e(X_i)| \le \alpha(1+\zeta)\,\|e\|_{L^1(\nu)}\,. \tag{4.6}$$

Combining (4.4)–(4.6),

$$\mathcal{L}(p+e)-\mathcal{L}(p) \ge \big((1-\alpha)(1-\zeta)-\alpha(1+\zeta)\big)\,\|e\|_{L^1(\nu)} - 2\sigma.$$

The coefficient is $1-2\alpha-\zeta = g/2 > 0$, proving (4.1). Since $\|\cdot\|_{L^1(\nu)}$ is a norm on the finite-dimensional space $\mathcal{P}_{d,n}$, its closed bounded sets are compact. Thus (4.1) makes every sublevel set of $\mathcal{L}$ compact, and continuity gives existence of a minimizer. For any minimizer $p^{(0)}$, substitute $e = p^{(0)} - p$ into (4.1) and use $\mathcal{L}(p^{(0)}) \le \mathcal{L}(p)$ to obtain (4.2).

To convert (4.2) to a uniform bound, reproduce the even periodic lift of e with the product Dirichlet kernel $D_d^{\otimes n}$, where $D_d(t) = \sum_{|k|\le d} e^{ikt}$. Hence

$$\|q\|_\infty \le \left\|D_d^{\otimes n}\right\|_\infty \|q\|_{L^1(\nu)} = (2d+1)^n\,\|q\|_{L^1(\nu)} \qquad (q\in\mathcal{P}_{d,n}). \tag{4.7}$$

Applying (4.7) to (4.2) proves (4.3). If $\sigma = 0$, the positive coefficient $g/2$ forces every minimizer to have $\left\|p^{(0)}-p\right\|_{L^1(\nu)} = 0$, so $p^{(0)} = p$ identically. □

The following example shows that cellwise majority alone is insufficient. Take $n = 1, d = 2, m = 2, p = 0, \sigma = 0$. Put two inliers at each of $-1/2$ and $1/2$, and one outlier at each of -1 and 1. For $K > 0$, give the outliers labels from $q(x) = K(x^2 - 1/4)$. Each cell has outlier fraction 1/3, but the weighted LAD objective is zero at q and equals $K/4$ at p. The condition (2.2) rules out exactly this type of concentration.

More generally, at zero noise the obstruction to weighted LAD is a nonzero direction whose weighted evaluation mass on the corrupted observations is at least its mass on the inliers: scaling the corrupted labels along that direction makes the truth non-unique or suboptimal. The coercive bound (4.1) excludes every such direction simultaneously: its right side is independent of the outlier-label magnitudes, while N_C^{-1} normalizes each cell separately.

5 Median refinement and contraction

We next convert the initializer bound into the target error. A cell-oscillation bound and a median-witness lemma yield a uniform one-step contraction, which is iterated a predetermined number of times.

Lemma 5.1 (Cell oscillation)**.** *For every $q \in \mathcal{P}_{d,n}$ and every $C \in \mathcal{C}_m$,*

$$\sup_{\boldsymbol{x},\boldsymbol{z}\in C} |q(\boldsymbol{x}) - q(\boldsymbol{z})| \le \gamma\,\|q\|_\infty\,. \tag{5.1}$$

Proof. For the angular lift Q, the trigonometric Bernstein inequality in each coordinate gives $\|\partial_k Q\|_\infty \le d\,\|Q\|_\infty$. Two points in the same angular product cell differ by at most π/m in each coordinate. Integrating along coordinatewise line segments gives (5.1). □

Lemma 5.2 (Median witness)**.** *Let I be a nonempty finite index set, let $X_i \in C$ and $Z_i \in \mathbb{R}$ for $i \in I$, where $C \in \mathcal{C}_m$, and let $G \subseteq I$ satisfy $|G| > |I|/2$. If $e \in \mathcal{P}_{d,n}$ and $|Z_i - e(X_i)| \le \sigma$ for every $i \in G$, then any median $\widetilde{Y}$ of $(Z_i)_{i\in I}$ has a witness $z_C \in C$ such that*

$$|e(z_C) - \widetilde{Y}| \le \sigma. \tag{5.2}$$

The witness need not be one of the sample locations X_i.

Proof. Let $I_- := \{i \in I : Z_i \le \widetilde{Y}\}$ and $I_+ := \{i \in I : Z_i \ge \widetilde{Y}\}$. The median property gives $|I_-|, |I_+| \ge |I|/2 > |I \setminus G|$, so choose $i_- \in I_- \cap G$ and $i_+ \in I_+ \cap G$. Write $Z_- = Z_{i_-}$, $Z_+ = Z_{i_+}$, $x_- = X_{i_-}$, and $x_+ = X_{i_+}$. The noise bounds give the one-sided consequences

$$e(x_-) \le \widetilde{Y} + \sigma, \qquad e(x_+) \ge \widetilde{Y} - \sigma.$$

If either value already lies in $[\widetilde{Y} - \sigma, \widetilde{Y} + \sigma]$, that location proves the claim. Otherwise $e(x_-) < \widetilde{Y} - \sigma$ and $e(x_+) > \widetilde{Y} + \sigma$. The line segment from x_- to x_+ lies in the rectangular cell, and continuity supplies a point on it at which $e = \widetilde{Y}$. This continuity witness need not equal any X_i. □

Proposition 5.3 (One-step contraction)**.** *On a good sample set, let $e_t = p - p^{(t)}$, $E_t = \|e_t\|_\infty$, and define r_t by (2.8). If $\gamma \le \beta/(2+\beta)$, then*

$$E_{t+1} \le (2+\beta)\sigma + \beta E_t. \tag{5.3}$$

Proof. The median witness first controls the target residual at an unknown point in each cell. Oscillation moves this control to the fixed representatives; minimax optimality and a second oscillation step then give a term γE_{t+1}, which we absorb.

For an inlier,

$$\left| Y_i - p^{(t)}(X_i) - e_t(X_i) \right| = |Y_i - p(X_i)| \le \sigma.$$

The good event and Lemma 5.2 therefore give, in every cell, a point z_C such that $|e_t(z_C) - \widetilde{Y}_{C,t}| \le \sigma$. By Lemma 5.1,

$$|e_t(a_C) - \widetilde{Y}_{C,t}| \le \sigma + \gamma E_t.$$

Using e_t as the feasible comparison polynomial in (2.8) shows that its optimal value λ_t is at most $\sigma + \gamma E_t$. For every $C \in \mathcal{C}_m$, the triangle inequality therefore gives

$$|(r_t - e_t)(a_C)| \le |r_t(a_C) - \widetilde{Y}_{C,t}| + |e_t(a_C) - \widetilde{Y}_{C,t}| \le 2(\sigma + \gamma E_t).$$

Hence

$$\max_{C\in\mathcal{C}_m} |(r_t - e_t)(a_C)| \le 2\sigma + 2\gamma E_t. \tag{5.4}$$

Put $q_t = r_t - e_t$ and $R = \|q_t\|_\infty = E_{t+1}$. Choose a maximizer $\boldsymbol{x}_\star \in C_\star$. By Lemma 5.1 and (5.4),

$$R = |q_t(\boldsymbol{x}_\star)| \le |q_t(a_{C_\star})| + \gamma R \le 2\sigma + 2\gamma E_t + \gamma R.$$

Thus

$$R \le \frac{2}{1-\gamma}\sigma + \frac{2\gamma}{1-\gamma} E_t.$$

The condition $(2+\beta)\gamma \le \beta$ bounds the two coefficients by $2+\beta$ and β, respectively, proving (5.3). □

The choice (2.2) gives $\gamma \le \beta/6 \le \beta/(2+\beta)$. Iterating (5.3) and using (4.3),

$$E_T \le \frac{2+\beta}{1-\beta}\sigma + \beta^T E_0. \tag{5.5}$$

Since $\beta = \varepsilon/12 \le 1/24$,

$$\frac{2+\beta}{1-\beta} = 2 + \frac{3\beta}{1-\beta} \le 2 + \frac{\varepsilon}{2}. \tag{5.6}$$

By (2.7) and (4.3),

$$\beta^T E_0 \le \beta^T \frac{4(2d+1)^n}{g}\sigma \le \frac{\varepsilon}{2}\sigma. \tag{5.7}$$

Equations (5.5)–(5.7) prove the target $(2+\varepsilon)\sigma$ error on every good sample set. No step divides by σ; at $\sigma = 0$, Proposition 4.1 already gives $E_0 = 0$. Moreover, after the good event is fixed the contraction is deterministic and uniform over e_t, so reusing the same samples introduces no adaptive probability issue.

6 Probability, sample complexity, and computational scope

We now show that the good event holds with high probability, derive the two design-specific sample bounds, and delimit the exact-real LP-oracle complexity. The only design-dependent quantity is the least cell mass.

Lemma 6.1 (A good product grid)**.** *Let m, M be positive integers and $0 < \delta \le 1$. Suppose the locations are i.i.d., the indicators satisfy the independence assumptions of Section 1, and every $C \in \mathcal{C}_m$ has probability π_C, where $\pi_{\min} := \min_{C\in\mathcal{C}_m} \pi_C > 0$. With $C_\rho = 16/(1-2\rho)^2$,*

$$M \ge C_\rho \pi_{\min}^{-1} \log \frac{2m^n}{\delta} \tag{6.1}$$

makes the sample set good with probability at least $1-\delta$.

Proof. For a fixed cell, a Chernoff bound gives

$$\mathbb{P}[N_C < M\pi_C/2] \le e^{-M\pi_C/8}.$$

Conditional on the selected locations and on $N_C = s$, the outlier indicators in the cell are independent Bernoulli variables with means at most ρ. Hoeffding's inequality therefore gives

$$\mathbb{P}[|B_C| > \alpha s \mid N_C = s] \le e^{-2(\alpha-\rho)^2 s}.$$

Consequently,

$$\mathbb{P}[|B_C| > \alpha N_C,\ N_C \ge M\pi_C/2] \le e^{-(\alpha-\rho)^2 M\pi_C}.$$

Let $L = \log(2m^n/\delta)$. Since $\alpha - \rho = (1-2\rho)/4$, (6.1) bounds the two failure terms for each cell by e^{-2L} and e^{-L}, respectively. A union bound over m^n cells gives total failure at most $m^n(e^{-L} + e^{-2L}) \le \delta$. □

For product-Chebyshev locations, every cell has probability exactly

$$\pi_{\min} = m^{-n}. \tag{6.2}$$

For the normalized uniform law, the smallest one-dimensional Chebyshev cell is an endpoint cell and has probability

$$\frac{1-\cos(\pi/m)}{2} = \sin^2 \frac{\pi}{2m} \geq m^{-2}. \tag{6.3}$$

Consequently, under the n-dimensional uniform product distribution,

$$\pi_{\min} \geq m^{-2n}. \tag{6.4}$$

Combining Lemma 6.1 with (6.2) and (6.4) gives

$$M_C = O_\rho(m^n \log(m^n/\delta)), \qquad M_U = O_\rho\big(m^{2n} \log(m^n/\delta)\big). \tag{6.5}$$

For uniform designs, the conditional distribution inside a Chebyshev cell is not the Chebyshev law. No such assumption is used: Corollary 3.4 is deterministic and holds for every finite collection of locations in the cell. The design law affects only the occupancy probability.

Because $m = O_{n,\rho}(d/\varepsilon)$, $d \geq 1$, and $\delta \leq \varepsilon$,

$$\log(m^n/\delta) = O_{n,\rho}(\log(d/\delta)).$$

Substitution in (6.5), followed by the deterministic analysis of Sections 4 and 5, proves the statistical and error claims in Theorem 1.1.

Proposition 6.2 (LP size and computational scope)**.** *Let $D = (d+1)^n$. On the no-empty-cell branch, the initializer LP has $D + M$ variables, $2M$ residual constraints, and $O(MD)$ nonzero coefficients. Each of the T refinement LPs has $D+1$ variables, $2m^n$ constraints, and $O(m^n D)$ nonzero coefficients. Every input uses at most $1 + T$ LP-oracle calls. For fixed n, ρ, these combinatorial description sizes and all work outside the oracle calls are polynomial in M.*

Proof. Introduce one nonnegative slack for each absolute residual in (2.6). For (2.8), introduce a variable λ and the two linear inequalities $-\lambda \leq r(a_C) - \widetilde{Y}_{C,t} \leq \lambda$ for each cell. Each residual constraint contains at most $D + 1$ potentially nonzero coefficients in the dense monomial representation, which gives the stated description bounds.

The norming estimate proves coercivity for the first LP. For the second, suppose $q(a_C) = 0$ for every $C \in \mathcal{C}_m$, and choose a maximizer $x_\star \in C_\star$. By Lemma 5.1, $\|q\|_\infty \leq \gamma \|q\|_\infty$; since $\gamma < 1$, this forces $q = 0$. Thus representative evaluation defines a norm on $\mathcal{P}_{d,n}$. Finite-dimensional norm equivalence makes the translated minimax objective coercive, so both LP objectives attain their minima.

Since $m \geq d$ and $d \geq 1$,

$$(d+1)^n \leq 2^n m^n = O_n(M).$$

Moreover,

$$T = \left\lceil \frac{\log(8(2d+1)^n/(g\varepsilon))}{\log(12/\varepsilon)} \right\rceil \leq 1 + \frac{\log(8(2d+1)^n/(g\varepsilon))}{\log(12/\varepsilon)} = O_{n,\rho}(\log(d/\varepsilon)) = O_{n,\rho}(\log M).$$

Assigning the M samples to cells takes $O(nM \log m)$ comparisons after the grid is formed. In each refinement, naive evaluation of all residuals takes $O(M(d+1)^n)$ arithmetic operations, and all cell medians can be found in $O(M \log M)$ comparisons, or in linear time by selection. Thus the work outside the LP-oracle calls is polynomial in M.

The call budget and combinatorial LP structure contain no σ, signal magnitude, or additive tolerance. When $\sigma > 0$, the call count is in particular independent of $\|p\|_\infty / \sigma$ and $1/\sigma$. Numerical LP coefficients include the observed labels Y_i, however, so their values and encoding lengths may depend on label scale. As stated in Section 1, no end-to-end finite-precision guarantee is claimed. $\square$

This completes the proof of Theorem 1.1.

7 Significance, related work, and limitations

We compare the theorem with prior upper and lower bounds, identify the source of the improved degree exponent, and delimit the result's scope. Extensions and boundary cases are collected in Appendix A.

In the stated exact-real LP-oracle model, Theorem 1.1 attains the target rates posed in Remark 1.9 of the ESA version of Arora et al. [1]. Their optimal-degree scheme retains an additive accuracy parameter and a scale-dependent refinement count, whereas their scale-independent initializer uses a much finer grid. Product-Chebyshev cell weighting instead gives

$$E_0 \le \frac{4(2d+1)^n}{g}\sigma,$$

so the refinement budget can be fixed in advance. The same coercive argument applies directly at $\sigma = 0$; exact recovery is not obtained by taking a limit as $\sigma \downarrow 0$.

The quantitative improvement comes from changing the norm targeted by the initializer. Here dx denotes unnormalized Lebesgue measure on $[-1,1]^n$, and $\lesssim_n, \gtrsim_n$ suppress positive constants that depend only on n. With Lebesgue-volume cell weights, ESA Theorems 1.11–1.12 of Arora et al. [1] (Theorems 1.8–1.9 in arXiv v1) give

$$\left\| q - q^\sharp \right\|_{L^1(dx)} \lesssim_n \frac{d}{m} \left\| q \right\|_\infty \lesssim_n \frac{d^{2n+1}}{m} \left\| q \right\|_{L^1(dx)}. \tag{7.1}$$

Thus a relative Lebesgue-L^1 proxy requires $m \gtrsim_n d^{2n+1}/\varepsilon$. In contrast, Lemma 3.3 directly gives

$$\left\| q - q^\sharp \right\|_{L^1(\nu)} \lesssim_n \frac{d}{m} \left\| q \right\|_{L^1(\nu)}. \tag{7.2}$$

The remaining d^n loss occurs only in the initializer's $L^1(\nu)$-to-L^∞ conversion. It therefore affects the logarithmic refinement count, not the grid size or the sample exponent.

The proof therefore separates deterministic norming from random occupancy. In angular coordinates, arbitrary cell representatives discretize $L^1(\nu)$ with relative error $O_n(d/m)$, independently of the location law or the conditional distribution within a cell. The law enters only through the least cell probability: m^{-n} for Chebyshev sampling and $\Theta(m^{-2n})$ for uniform sampling. This produces the d^n and d^{2n} degree exponents, respectively.

For the lower-bound comparison, set $C = 2 + \varepsilon$. At constant success probability exceeding $2/3$, fixed ε, and fixed $\rho > 0$, ESA Theorem 1.7 of Arora et al. [1] gives a constant $c = c(C, \rho) > 0$, independent of d, and the distribution-free lower bound $\Omega((cd)^n \log d)$. Thus the product-Chebyshev rate matches the known d-dependence. Under uniform sampling, ESA Theorem 1.6 gives $c = c(C) > 0$ and a lower bound $\Omega((cd)^{2n})$ for every outlier rate, including zero. The uniform upper bound therefore has the optimal polynomial exponent in d, up to the occupancy logarithm. These fixed-parameter comparisons do not assert joint optimality in the other parameters.

Earlier work on robust algebraic curve fitting and polynomial regression includes Arora and Khot [2], Daltrophe et al. [5], Guruswami and Zuckerman [9]. The closest algorithmic precedents are Kane et al. [11], who combine arbitrary Chebyshev-cell representatives, weighted LAD, and median refinement in one dimension, and Arora et al. [1], who extend robust regression to several variables. The latter's scale-independent Lebesgue-weighted initializer requires $m \gtrsim_n d^{2n+1}/\varepsilon$. Our new step is the product-Chebyshev L^1 norming inequality for jointly chosen multivariate representatives, which gives the cellwise LAD bound at $m = O_{n,\rho}(d/\varepsilon)$.

Adjacent robust-regression methods use different corruption models and loss criteria. Klivans et al. [12] allow adaptive replacement of complete covariate–label pairs under hypercontractivity and control population L^2 prediction loss. Diakonikolas et al. [7] treat strong contamination for general stochastic objectives and obtain approximate population critical points under gradient-regularity assumptions. Diakonikolas et al. [8] study Gaussian linear regression with full-pair replacement and Euclidean parameter error, while Prasad et al. [15] use robust gradient estimation for population-risk minimization under Huber contamination or heavy-tailed gradients. These settings and targets do not directly yield uniform recovery of a bounded-noise polynomial on a fixed cube under the present clean-design, label-corruption model.

These norming estimates are part of the Marcinkiewicz–Zygmund and Plancherel–Polya literature. Bounds for nonuniform grids and perturbed nodes are given in [14, 16]. De Marchi and Kroó [6] treat multivariate domains, using selected nodes and weights rather than arbitrary points in prescribed product cells. Other sampling-discretization results include [3, 4, 13], and Kämmerer [10] studies structured Chebyshev grids for noiseless reconstruction.

The argument does not settle the analogous arbitrary-representative question for unweighted Lebesgue $L^1(dx)$; in angular coordinates that norm carries the non-translation-invariant weight $\prod_k \sin\theta_k$. This issue is unnecessary for Theorem 1.1, because the initializer may target $L^1(\nu)$ even under uniform locations. The theorem also assumes independent random outlier support and an exact-real LP oracle; it does not provide an end-to-end finite-precision guarantee. The source-version detail behind the median witness is recorded in Appendix A.2.

In summary, weighting cells by their product-Chebyshev mass separates deterministic norming from distribution-dependent occupancy. It yields the d^n and d^{2n} degree exponents without using σ as an algorithmic input, under the statistical and computational model stated above.

References

[1] V. Arora, A. Bhattacharyya, M. Boban, V. Guruswami, and E. Kelman. Outlier robust multivariate polynomial regression. In *32nd Annual European Symposium on Algorithms (ESA 2024)*, volume 308 of *LIPIcs*, pages 12:1–12:17, 2024. doi:10.4230/LIPIcs.ESA.2024.12. Extended version: arXiv:2403.09465v1.

[2] S. Arora and S. Khot. Fitting algebraic curves to noisy data. *Journal of Computer and System Sciences*, 67(2):325–340, 2003. doi:10.1016/S0022-0000(03)00012-6.

[3] F. Dai, E. Kosov, and N. Murasko. Empirical approximation of L_p norms. arXiv:2606.00347, 2026.

[4] F. Dai, E. Kosov, and V. Temlyakov. Some improved bounds in sampling discretization of integral norms. *Journal of Functional Analysis*, 285(4):109951, 2023. doi:10.1016/j.jfa.2023.109951.

[5] H. Daltrophe, S. Dolev, and Z. Lotker. Big data interpolation using functional representation. *Acta Informatica*, 55:213–225, 2018. doi:10.1007/s00236-016-0288-8.

[6] S. De Marchi and A. Kroó. Marcinkiewicz–Zygmund type results in multivariate domains. *Acta Mathematica Hungarica*, 154(1):69–89, 2018. doi:10.1007/s10474-017-0769-4.

[7] I. Diakonikolas, G. Kamath, D. Kane, J. Li, J. Steinhardt, and A. Stewart. SEVER: A robust meta-algorithm for stochastic optimization. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1596–1606, 2019. https://proceedings.mlr.press/v97/diakonikolas19a.html.

[8] I. Diakonikolas, W. Kong, and A. Stewart. Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2745–2754, 2019. doi:10.1137/1.9781611975482.170.

[9] V. Guruswami and D. Zuckerman. Robust Fourier and polynomial curve fitting. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 751–759, 2016. doi:10.1109/FOCS.2016.75.

[10] L. Kämmerer. Constructing efficient spatial discretizations of spans of multivariate Chebyshev polynomials. arXiv:2406.03281, 2024.

[11] D. Kane, S. Karmalkar, and E. Price. Robust polynomial regression up to the information theoretic limit. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 391–402, 2017. doi:10.1109/FOCS.2017.43.

[12] A. Klivans, P. K. Kothari, and R. Meka. Efficient algorithms for outlier-robust regression. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1420–1430, 2018. https://proceedings.mlr.press/v75/klivans18a.html.

[13] E. D. Kosov. Marcinkiewicz-type discretization of L^p-norms under the Nikol'skii-type inequality assumption. *Journal of Mathematical Analysis and Applications*, 504(1):125358, 2021. doi:10.1016/j.jmaa.2021.125358.

[14] J. Ortega-Cerdà and J. Saludes. Marcinkiewicz–Zygmund inequalities. *Journal of Approximation Theory*, 145(2):237–252, 2007. doi:10.1016/j.jat.2006.09.001.

[15] A. Prasad, A. S. Suggala, S. Balakrishnan, and P. Ravikumar. Robust estimation via robust gradient estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):601–627, 2020. doi:10.1111/rssb.12364.

[16] K. V. Runovski and W. Sickel. Marcinkiewicz–Zygmund-type inequalities, trigonometric interpolation on non-uniform grids and unconditional Schauder bases in Besov spaces on the torus. *Zeitschrift für Analysis und ihre Anwendungen*, 16(3):669–687, 1997. doi:10.4171/ZAA/784.

A Boundary cases, source note, and extensions

This appendix records results that are useful for scope and reproducibility but are not needed in the main proof. None changes the parameter regime of Theorem 1.1.

A.1 Constant polynomials

The convention $d \geq 1$ in Theorem 1.1 avoids the expression $\log(d/\delta)$. When $d = 0$, the kernel construction is unnecessary: every polynomial is constant, and both the representative-norming error and cell-oscillation constant are zero. Taking $m = 1$, any weighted-LAD minimizer is a median of the labels. On the strict-majority event it lies in $[p - \sigma, p + \sigma]$, and it equals p when $\sigma = 0$. A Chernoff bound makes this event hold with probability $1 - \delta$ using $O_\rho(\log(1/\delta))$ samples under either design.

A.2 Source-version note for the median witness

Claim B.1 of arXiv:2403.09465v1 in Arora et al. [1] is correct, but the displayed proof in that version reverses the signs of the one-sided noise terms in its interval endpoints. The valid implications are precisely those used in Lemma 5.2; continuity then supplies a witness that need not be an observed sample location or correspond to an observed inlier residual.

A.3 Other design laws and polynomial spaces

For any i.i.d. design under which every Chebyshev cell has probability at least $\pi_{\min} > 0$, the same estimator succeeds under the independent outlier model with $O_\rho(\pi_{\min}^{-1} \log(m^n/\delta))$ samples. Conditioned on the good-grid event, the deterministic proof also permits any cellwise corruption pattern with outlier fraction at most α; independent indicators are used only to establish that event.

The theorem remains valid when $\mathcal{P}_{d,n}$ is replaced by the space of polynomials of total degree at most d. This is a linear subspace of $\mathcal{P}_{d,n}$, so every analytic and probabilistic estimate restricts to it; the sample bounds are unchanged and the coefficient dimension becomes $\binom{n+d}{n}$.

For coordinate degrees $d_k \in \mathbb{Z}_{\geq 0}$ and positive integers $m_k \geq \max\{1, d_k\}$, the product-kernel proof gives

$$\left\| q - q^\sharp \right\|_{L^1(\nu)} \leq A_n \left(\sum_{k=1}^{n} \frac{d_k}{m_k} \right) \|q\|_{L^1(\nu)} ,$$

and the cell-oscillation factor is at most $\pi \sum_{k=1}^n d_k/m_k$. Use V_{d_k} in coordinate k, with $V_0 \equiv 1$; the path argument contributes $(\pi/m_k) \left\|\partial_k Q\right\|_1$, and the Bernstein bound contributes $O_n(d_k) \left\|Q\right\|_1$.

More abstractly, the proof separates four modules: arbitrary-multiset weighted L^1 norming, an L^1-to-L^∞ inequality, a compatible cell-oscillation bound, and a lower bound on cell probabilities. A global empirical L^1 margin or an algebraic error-locator could provide other routes, but neither is analyzed or used in Theorem 1.1.