# Uniform Expected Risk for Distance-Based Boundary Regression Designs

CausalSmith*

August 25, 2026

**Abstract**

This paper establishes two new unconditional logarithmic minimax lower bounds for distance-compressed boundary regression. In the unsigned experiment, a rule observes outcomes and scalar Euclidean distances from each boundary query point. Over the compact nonparametric law class $\mathcal{P}_{\mathrm{NP}}(L, q)$ in Definition 9, with $q \geq 1$ and $L \geq 4$, both the CTY common-map risk in Definition 12 and the larger point-indexed outer risk in Definition 14 have minimax expected boundary sup-loss bounded below on the scale

$$a_n = \left(\frac{\log n}{n}\right)^{1/4}.$$

The support-boundary hypercube constructs separated boundary perturbations with matching scalar-distance information at the queried point, yielding the same lower-bound scale even when the rule may use point-indexed Borel sections. Economically, this experiment isolates the risk cost of summarizing proximity by scalar distance after angular location and treatment-side information have been removed: a distance-only rule cannot tell which boundary arc or treatment side generated nearby observations, so recovering the entire boundary curve incurs a logarithmic testing penalty. The second unconditional lower bound holds in a signed-distance experiment with known treatment geometry, already on a fixed rectangular subexperiment. Conditional on the three CTY-style analytic inputs collected in Definition 33, the stabilized local-polynomial estimator and that lower bound characterize the signed-distance expected outer-risk rate on the same scale for every fixed polynomial order $p$, moment exponent $\nu \geq 2$, and envelope $L \geq L_0(p)$. Thus the unsigned contribution is lower-bound sharpness, while the signed two-sided frontier is explicitly conditional on $\mathsf{AI}_{p,\nu,L}$.

## 1 Introduction

Regression discontinuity designs use discontinuities in treatment assignment to identify local causal effects near assignment thresholds. The classical scalar-cutoff design begins with the threshold assignment framework of Thistlethwaite and Campbell [1960] and the continuity-based identification analysis of Hahn et al. [2001]; modern econometric treatments develop local comparisons, bandwidth selection, bias correction, and inference around that threshold [Imbens

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; where a result from the literature is a formal dependency, it enters as a published input rather than a Lean-checked step, and cited work otherwise supplies framing and positioning.

and Lemieux, 2008, Lee and Lemieux, 2010, Imbens and Kalyanaraman, 2012, Calonico et al., 2014, 2020, Armstrong and Kolesar, 2020, Kolesár and Rothe, 2018, Imbens and Wager, 2019]. Boundary and geographic RD designs extend the same logic to bivariate or multivariate scores, where treatment changes across a curve or interface rather than at a scalar cutoff [Keele and Titiunik, 2015, Keele et al., 2015, Keele and Titiunik, 2016, Papay et al., 2011, Reardon and Robinson, 2012, Wong et al., 2013, Choi and Lee, 2018, Díaz and Zubizarreta, 2023, Sawada et al., 2024, Merlano, 2025, Kendall et al., 2026].

Distance-to-boundary reductions are central in this setting. They turn a two-dimensional local problem into a one-dimensional running-coordinate problem and are used both as empirical summaries of proximity to the boundary and as coordinates for local-polynomial fitting. Cattaneo et al. [2026] develop distance-based identification and estimation tools for boundary discontinuity designs, including local-polynomial methods for boundary treatment-effect curves. This paper studies the minimax expected-risk consequences of two precise distance experiments: an unsigned experiment in which angular and side information is compressed away, and a known-geometry signed experiment in which the rule retains the treatment side induced by the boundary.

The first experiment is an unsigned distance-compressed boundary regression problem. The law class $\mathcal{P}_{\mathrm{NP}}(L,q)$ in Definition 9 consists of compact bivariate random-design regression laws with bounded continuous design density, Hölder regression smoothness of order $q$, and uniformly nondegenerate conditional variance. At a boundary query point $x$, the data available to the rule are

$$V_{n,P}(x) = \big((Y_i, \|X_i - x\|_2) : 1 \le i \le n\big),$$

as defined in Definition 10. The target is the boundary regression function $x \mapsto \mu_P(x)$, and the loss is expected supremum error over $\mathrm{bd}(\mathcal{X}_P)$.

For this unsigned experiment, Theorem 1 establishes that, for every integer $q \ge 1$ and every $L \ge 4$, the common-map risk $R_n^{\mathrm{NP}}$ in Definition 12 is bounded below on the scale $a_n = (\log n/n)^{1/4}$. Theorem 2 gives the same lower scale for the larger point-indexed class in Definition 13, evaluated with the outer-expectation risk in Definition 14. Thus the logarithmic lower bound is tied to the unsigned distance information set itself: even law-independent Borel sections chosen separately at each boundary point face the same order of uniform expected loss over $\mathcal{P}_{\mathrm{NP}}(L,q)$.

The proof of the unsigned converse combines a direct-product testing inequality with a support-boundary packing. Lemma 1 converts many small coordinatewise overlaps into a lower bound on the probability of at least one decoding error. Lemma 2 builds many separated boundary points on a square support and assigns an independent binary regression perturbation to each local cell. The perturbations are smooth enough to remain in $\mathcal{P}_{\mathrm{NP}}(L,q)$, their boundary values differ by order $a_n$, and the unsigned distance laws at each coordinate have Kullback–Leibler divergence of logarithmic size. The support-boundary packing yields separated boundary signals of order $a_n$ and logarithmic compressed-distance KL control.

The second experiment is a known-geometry signed-distance boundary RD problem. The law class $\mathcal{P}_{12}(p,\nu,L)$ in Definition 22 is the displayed Euclidean-distance, uniform-kernel specialization of the CTY Assumptions 1–2 boundary design, with rectangular support, bounded continuous density, smooth potential-outcome regressions, continuous variances, a $2+\nu$ moment envelope, a rectifiable interior interface, two-sided local mass, and a population Gram lower bound. The rule observes signed-distance data $U^{\pm}_{n,P}(x)$ from Definition 21 and may depend on the known geometry $G_P$ from Definition 20. The target is the interface treatment-effect curve $\tau_P(x) = \mu_{1,P}(x) - \mu_{0,P}(x)$ in Definition 23.

For the signed-distance experiment, Theorem 3 proves a lower bound on the signed-distance minimax risk $R_n^{12,\pm}(p, \nu, L)$ in Definition 26. For every fixed order $p$, there is an envelope threshold $L_0(p) \geq 48$ such that, for every $\nu \geq 2$ and $L \geq L_0(p)$, the risk is bounded below on the $a_n$ scale. The same lower bound is already present in the fixed rectangular subexperiment in Definition 27. The hard family in Lemma 6 places independent local perturbations along the bottom edge of a known treated rectangle and keeps all vertices inside the same known-geometry design.

The signed-distance upper bound uses an explicit local-polynomial rule. Definition 28 defines a degree-$p$ estimator that winsorizes outcomes at $B(h) = h^{-1/3}$, inverts the empirical Gram matrix only on a uniform conditioning event, and clips the resulting treatment-effect estimate to $[-2L, 2L]$. Under the distance-identification, first-order bias, and expected maximal inputs collected in Definition 33, Proposition 1 shows that this estimator with bandwidth $h_n = a_n$ has expected outer risk bounded by a constant multiple of $a_n$ uniformly over $P \in \mathcal{P}_{12}(p, \nu, L)$.

Under $\mathsf{AI}_{p,\nu,L}$, the signed-distance upper bound combines with the lower bound to give the conditional frontier in Theorem 4. For every fixed $p$, $\nu \geq 2$, and $L \geq L_0(p)$, the signed-distance minimax risk satisfies

$$0 < c \leq \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n} \leq \limsup_{n\to\infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n} \leq C < \infty.$$

The same theorem records that the winsorized, Gram-stabilized signed-distance local-polynomial estimator attains the upper side of this bound under the same three analytic inputs. A short reproducibility note records the checked scope for the anchored mathematical statements in Section D.

The comparison between the two experiments is organized by law class, information set, and risk. In the unsigned problem, scalar Euclidean distances from each query point create a boundary-indexed compression in which many angularly separated alternatives remain hard to distinguish, producing unconditional lower bounds. In the signed problem, known boundary geometry and treatment side supply the one-dimensional coordinate used by local-polynomial RD methods; its fixed-geometry lower bound is unconditional, while its two-sided rate characterization uses the displayed analytic inputs. The common normalization $a_n$ therefore records a shared logarithmic lower scale, with a conditional upper side for the signed experiment.

The remainder of the paper proceeds as follows. Section 3 defines the compact nonparametric unsigned distance experiment, its decision classes, and its minimax risks. Section 4 defines the known-geometry signed-distance RD experiment, the A1/A2 law class, the treatment-effect target, and the stabilized local-polynomial estimator. Section 5 states the unsigned lower bounds, the signed-distance lower bound, the conditional upper bound, and the matched signed-distance frontier. Section 6 discusses the interpretation for boundary RD designs. Sections A to E contain the direct-product and packing arguments, the signed-distance empirical-process ingredients, the fixed-geometry signed hypercube, the reproducibility note, and deferred proofs.

## 2 Related Literature

Regression discontinuity designs originate in the threshold assignment framework of Thistlethwaite and Campbell [1960] and the identification analysis of Hahn et al. [2001], with modern econometric treatments emphasizing local comparisons, bandwidth choice, and inferential refinements [Imbens and Lemieux, 2008, Lee and Lemieux, 2010, Imbens and Kalyanaraman, 2012].

Geographic and boundary designs extend this logic to assignment regions in space, where the score is vector-valued and the estimand is indexed by an interface rather than a scalar cutoff [Keele and Titiunik, 2015, Keele et al., 2015, Keele and Titiunik, 2016, Papay et al., 2011, Reardon and Robinson, 2012, Wong et al., 2013, Choi and Lee, 2018, Díaz and Zubizarreta, 2023, Sawada et al., 2024, Merlano, 2025]. Within that literature, Cattaneo et al. [2026] develop distance-based identification and estimation tools for boundary regression designs. Their analysis establishes identification from distance-to-boundary reductions, sequence-scoped bias expansions, and probability-order stochastic control for local-polynomial procedures under their maintained design conditions.

Theorem 6 of Cattaneo et al. [2026] studies the unsigned minimax risk

$$\inf_{T_n} \sup_{P \in \mathcal{P}_{\mathrm{NP}}(L,q)} \mathbb{E}_P \left[ \sup_{x \in \mathrm{bd}(\mathcal{X}_P)} |T_n(U_n(x)) - \mu_P(x)| \right], \qquad U_n(x) = ((Y_i, \|X_i - x\|_2) : i \leq n),$$

where one law-independent measurable map $T_n$ is used at every boundary point. Their theorem proves a positive lower bound after multiplication by $n^{1/4}$. The printed conjecture immediately following that theorem replaces $n^{1/4}$ by $(n/\log n)^{1/4}$, matching the scale of their uniform upper bound. The unsigned common-map result here proves that conjectured strengthening without changing the law class, transformed data, target, decision class, or expected supremum loss. The point-indexed result changes only the decision rule: a law-independent Borel section may now be selected separately for every query point, while the transformed data and regression target remain the same; outer expectation supplies the measurable envelope for the boundary supremum. It proves the same logarithmic lower scale for this larger class.

The signed result concerns a different econometric target and information set. Motivated by CTY's signed-distance identification and local-polynomial analysis, it studies the boundary treatment-effect curve when the assignment geometry and treatment side are available to the rule. The paper proves a new minimax lower bound already on one fixed known rectangular geometry. Its expected-risk upper bound is stated under the displayed identification, bias, and expected maximal hypotheses. Table 1 records these three deliverables in parallel.

This paper studies the uniform expected-risk consequences of those distance reductions. In the unsigned experiment, the relevant object is the compact bivariate nonparametric regression class $\mathcal{P}_{\mathrm{NP}}(L, q)$ in Definition 9, where a rule observes outcomes and unsigned Euclidean distances as in Definition 10. The decision problem uses the common-map class $\mathcal{T}_n^{\mathrm{NP}}$ in Definition 11 and its completed-expectation risk $R_n^{\mathrm{NP}}$ in Definition 12. The resulting converse is a same-class statement: over the law class and information set just described, the minimax expected boundary sup-loss has the logarithmic scale $a_n$ introduced in Definition 8. The point-indexed extension uses the sectionwise-Borel outer-risk setup in Definitions 13 and 14, thereby characterizing the same compression barrier for rules that may choose a law-independent Borel section separately at each boundary point.

This focus differs from the usual role of distance in geographic RD applications. Empirical distance reductions often serve as low-dimensional summaries of proximity to a boundary or as practical coordinates for local fitting [Keele and Titiunik, 2015, Keele et al., 2015, Keele and Titiunik, 2016, Kendall et al., 2026]. Here the distance reduction is part of the statistical experiment: the rule receives the outcome and the scalar unsigned distance from the query point, and the risk is the expected supremum over boundary points. The lower bound therefore speaks to information loss under the compressed observation scheme, rather than to specification choice in a geographic RD regression. Related diagnostic and design tools, including manipulation tests

| Deliverable | Nearest CTY result | Exact increment in this paper |
|---|---|---|
| Unsigned CTY decision problem | Theorem 6 of Cattaneo et al. [2026]: $\mathcal{P}_{\mathrm{NP}}(L,q)$, data $U_n(x) = ((Y_i, \|X_i - x\|_2) : i \le n)$, one common map $T_n$, boundary regression target, and expected boundary sup-loss; lower scale $n^{-1/4}$. | Keeps the law class, information, target, decision class, and risk fixed and proves the conjectured lower scale $a_n$. |
| Point-indexed unsigned rules | CTY's lower-bound theorem uses one law-independent common map across query points. | Allows a separate law-independent Borel section for each query point, uses outer expectation for the boundary supremum, and still proves the $a_n$ lower scale. |
| Signed known-geometry treatment effects | Theorems 1–2 and Supplemental SA-8.3–SA-8.4 provide the nearest identification, bias, and stochastic-process ingredients for signed-distance local-polynomial analysis. | Changes the target to the boundary treatment-effect curve and the information to signed distance with known geometry; proves an unconditional $a_n$ lower bound on a fixed rectangular subexperiment and, under $\mathsf{AI}_{p,\nu,L}$, a stabilized expected outer-risk upper bound at the same scale. |

Table 1: Nearest CTY benchmarks and the exact change in target, information, decision class, risk, or conditioning status.

and geographic balance analyses [McCrary, 2008, Cattaneo and Titiunik, 2022, Cattaneo et al., 2019, 2023, 2025], address complementary features of RD credibility; the present comparison is organized around the target, the available data transformation, and the uniform loss criterion.

The signed-distance part connects to local-polynomial boundary estimation. Local-polynomial methods provide the standard approximation architecture for nonparametric regression and RD estimation [Fan and Gijbels, 1996, Ruppert and Wand, 1994, Calonico et al., 2014, 2020, Imbens and Wager, 2019]. Robust bias correction and optimal inference results refine confidence interval construction and coverage for scalar or boundary-local targets [Calonico et al., 2014, 2020, Armstrong and Kolesar, 2020, Kolesár and Rothe, 2018]. The known-geometry signed-distance experiment fixes the CTY Assumptions 1–2 potential-outcome boundary design through $\mathcal{P}_{12}(p, \nu, L)$ in Definition 22, observes signed-distance data in Definition 21, and evaluates treatment-effect rules in Definitions 25 and 26. Assuming the CTY distance-identification, uniform first-order bias, and expected maximal inputs collected as $\mathsf{AI}_{p,\nu,L}$ in Definition 33, the paper combines the signed-distance lower bound with the winsorized, Gram-stabilized local-polynomial upper bound to obtain the two-sided expected-risk rate.

The minimax structure draws on classical nonparametric lower-bound and empirical-process ideas. Packing and modulus arguments underlie the rate calculations in nonparametric estimation [Stone, 1982, Tsybakov, 2009], while boundary and set-indexed empirical-process tools support the stochastic controls used in local smoothing problems [Mammen and Tsybakov, 1995, Cuevas and Rodriguez-Casal, 2004, Vapnik and Chervonenkis, 1971, Pollard, 1984, van der Vaart and Wellner, 1996]. The lower-bound construction in this paper adapts those ideas to boundary-indexed distance compression: many separated boundary locations carry independent binary perturbations, while scalar distance observations constrain the coordinatewise information available to any admissible rule. The empirical-process component enters the signed-distance upper result through the stated analytic inputs and the bounded-envelope maximal inequality used for the winsorized score. Together, these ingredients place the paper between geographic RD methodology and minimax nonparametric theory: the estimand is a boundary treatment-effect or boundary regression curve, the information set is a distance-compressed experiment, and the criterion is uniform expected loss over the displayed law classes.

# 3 Setup: Distance-Compressed Boundary Regression

This section fixes the statistical experiment used for the unsigned lower bounds. The object of interest is a boundary-indexed regression function under a compact bivariate random-design law, and the information available at a boundary point is the outcome together with the scalar Euclidean distance from that point. The normalization used throughout the converse arguments is introduced first.

**Definition 1.** [`synth_25`] (Law operator $\mathcal{L}$) *For a random element $Z$ taking values in a measurable space $(E, \mathcal{E})$ under a probability law $P$, write $\mathcal{L}_P(Z) = P \circ Z^{-1}$ for the distribution of $Z$; when $P$ is clear, write $\mathcal{L}(Z)$. For an event $A$ with $P(A) > 0$, write $\mathcal{L}_P(Z \mid A)$ for the conditional distribution satisfying $\mathcal{L}_P(Z \mid A)(B) = P(Z \in B, A)/P(A)$ for every $B \in \mathcal{E}$.*

**Definition 2.** [`synth_28`] (Smallest quadratic-form eigenvalue $\lambda_{\min}$) *For a real symmetric matrix $\Psi$, we write $\lambda_{\min}(\Psi) = \inf_{\|v\|_2=1} v^\top \Psi v$, equivalently $\lambda_{\min}(\Psi) \geq c$ means $c\|v\|_2^2 \leq v^\top \Psi v$ for every vector $v$.*

**Definition 3.** [`synth_43`] (Closed Euclidean ball $B_2(x, r)$) *For $x \in \mathbb{R}^2$ and $r \geq 0$, we write $B_2(x, r)$ for the closed Euclidean ball $\{z \in \mathbb{R}^2 : \|z - x\|_2 \leq r\}$.*

**Definition 4.** [`synth_36`] (Euclidean $q$-Hölder ball $\mathcal{H}^q(\mathcal{X}_P, L)$) *We say $g \in \mathcal{H}^q(\mathcal{X}_P, L)$ when $g : \mathbb{R}^2 \to \mathbb{R}$ is $C^{q-1}$ on $\mathcal{X}_P$, satisfies $\|D^j g(x)\| \leq L$ for every $0 \leq j \leq q-1$ and every $x \in \mathcal{X}_P$, and its $(q-1)$st Fréchet derivative is $L$-Lipschitz on $\mathcal{X}_P$: $\|D^{q-1}g(x) - D^{q-1}g(y)\| \leq L\|x - y\|_2$ for all $x, y \in \mathcal{X}_P$.*

**Definition 5.** [`synth_72`] (Faithful random-design law space $\mathcal{P}$) *We say $P \in \mathcal{P}$ when $P$ is a probability law of $(Y, X) \in \mathbb{R} \times \mathbb{R}^2$ with square-integrable outcome, whose carried objects $\mathcal{X}_P$, $f_P$, $\mu_P$, and $\sigma_P^2$ are tied to the law as follows: $\mathcal{X}_P$ is the topological support of the $X$-marginal, the $X$-marginal has Lebesgue density $f_P \mathbf{1}_{\mathcal{X}_P}$, $\mu_P$ is an $X$-marginal-almost-everywhere version of $\mathbb{E}_P[Y \mid X = x]$, and $\sigma_P^2$ is an $X$-marginal-almost-everywhere version of $\operatorname{Var}_P(Y \mid X = x)$.*

**Definition 6.** [`synth_138`] (Bivariate design support $\mathcal{X}$) *For a law $P$ of $(Y, X)$ on $\mathbb{R} \times \mathbb{R}^2$, we say $\mathcal{X}$ is the bivariate design support when $\mathcal{X} = \operatorname{supp}(\mathcal{L}_P(X))$, the topological support of the $X$-marginal under $P$.*

**Definition 7.** [`synth_29`] (Total variation distance $\|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}}$) *For the two probability laws $Q_{j,0}$ and $Q_{j,1}$ on the common measurable sample space of $S_j$, define their total variation distance by*

$$\|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} = \sup_A |Q_{j,0}(A) - Q_{j,1}(A)|,$$

*where the supremum ranges over all measurable sets $A$ in that sample space.*

These preliminary conventions give the basic probabilistic and geometric vocabulary. In particular, the law-space convention ties the support, density, regression, and conditional variance to the same underlying law, while the Hölder and ball notation provide the local smoothness and Euclidean localization language used in the packing arguments. Total variation enters later through the direct-product comparison of coordinatewise compressed experiments.

**Definition 8.** [`def:frontier-rates`] (Frontier rate $a_n$) *The frontier normalization is*

$$a_n = \left(\frac{\log n}{n}\right)^{1/4}.$$

The scale in Definition 8 is the common normalization for the unsigned boundary lower bounds. Its logarithmic factor reflects the uniform boundary supremum loss rather than a pointwise loss criterion.

The nonparametric class now fixes the bivariate random-design model. It combines compact support, a bounded continuous design density, Hölder regression smoothness, and uniformly nondegenerate conditional variance.

**Definition 9.** [`def:cty-nonparametric-class`] (Nonparametric law class $\mathcal{P}_{\mathrm{NP}}(L,q)$) *For $q \in \mathbb{N}$ and $L \in \mathbb{R}$, the class $\mathcal{P}_{\mathrm{NP}}(L,q)$ consists of all random-design regression laws $P$ satisfying the following conditions.*

- ***(Parameter regime.)*** $1 \le q$ *and* $4 \le L$.
- ***(Sampling.)*** *$P$ is a probability law for one observation $(Y,X) \in \mathbb{R} \times \mathbb{R}^2$, $Y$ is square-integrable under $P$, and for every $n$, $(Y_i,X_i)_{i=1}^n$ are sampled independently and identically from $P$.*
- ***(Design.)*** *The $X$-marginal has Lebesgue density $f_P \mathbf{1}_{\mathcal{X}_P}$, where $\mathcal{X}_P$ is the exact topological support of the $X$-marginal. Write $f_P$ for this density. It is continuous on $\mathcal{X}_P$, $\mathcal{X}_P$ is compact and satisfies $\mathcal{X}_P \subseteq [-L,L]^2$, and*

  $$L^{-1} \le f_P(x) \le L$$

  *for every $x \in \mathcal{X}_P$. The boundary $\mathrm{bd}(\mathcal{X}_P)$ admits a Lipschitz parametrization by $[0,1]$.*
- ***(Regression.)*** *The function $\mu_P$ is an $X$-marginal-almost-everywhere version of $\mathbb{E}_P[Y \mid X = x]$ and belongs to $\mathcal{H}^q(\mathcal{X}_P, L)$.*
- ***(Variance.)*** *The function $\sigma_P^2$ is an $X$-marginal-almost-everywhere version of $\mathrm{Var}_P(Y \mid X = x)$, is continuous on $\mathcal{X}_P$, and satisfies*

  $$L^{-1} \le \sigma_P^2(x) \le L$$

  *for every $x \in \mathcal{X}_P$.*

*Equivalently,*

$$\mathcal{P}_{\mathrm{NP}}(L,q) = \{P : P \text{ satisfies the conditions above}\}.$$

The class in Definition 9 is a compact-support version of a random-design nonparametric regression model. The density bounds keep local sample sizes comparable across admissible laws, the Lipschitz boundary condition gives a one-dimensional interface over which a supremum can be evaluated, and the Hölder condition calibrates how quickly regression values may change near separated boundary locations. These conditions are standard in nonparametric minimax analysis, with compactness and smoothness playing the same organizational role as in classical regression lower bounds [Stone, 1982, Tsybakov, 2009].

At a boundary query point, the unsigned experiment compresses each covariate to its distance from that point. The rule retains the outcome and the scalar distance, so all angular information around the query point is absorbed into this one-dimensional summary.

**Definition 10.** [`def:cty-distance-data`] (Distance data $V_{n,P}(x)$) *For $x \in \mathrm{bd}(\mathcal{X}_P)$, the distance data are*

$$V_{n,P}(x) = \big((Y_i, \|X_i - x\|_2) : 1 \le i \le n\big).$$

The observation in Definition 10 is indexed by the boundary point because the same sample is re-expressed relative to each query location. This formulation matches the distance-to-boundary reductions studied by Cattaneo et al. [2026], while the present risk criterion evaluates the resulting boundary-indexed regression rule uniformly over $\mathrm{bd}(\mathcal{X}_P)$.

The first decision class uses one law-independent map for all query points. This common-map restriction captures procedures that process the distance data by a single measurable rule once the sample size is fixed.

**Definition 11.** [`def:cty-distance-decision-class`] (Common-map class $\mathcal{T}_n^{\mathrm{NP}}$) *For a represented rule, write $T_n$ for its selected law-independent measurable map. Define the sample-size-indexed class by*

$$\mathcal{T}_n^{\mathrm{NP}} = \left\{ \widehat{\mu}_n : \begin{array}{l} \text{there exists a measurable map } T_n : (\mathbb{R} \times [0,\infty))^n \to \mathbb{R} \text{ selected independently of } P, \\ \widehat{\mu}_n(x) = T_n(V_{n,P}(x)) \text{ for every } x \in \mathrm{bd}(\mathcal{X}_P) \end{array} \right\}.$$

**Definition 12.** [`def:cty-distance-risk`] (Common-map risk $R_n^{\mathrm{NP}}$) *The common-map minimax risk is*

$$R_n^{\mathrm{NP}} = \inf_{\widehat{\mu}_n \in \mathcal{T}_n^{\mathrm{NP}}} \sup_{P \in \mathcal{P}_{\mathrm{NP}}(L,q)} \mathbb{E}_P \left[ \sup_{x \in \mathrm{bd}(\mathcal{X}_P)} |\widehat{\mu}_n(x) - \mu_P(x)| \right].$$

Together, Definitions 11 and 12 define the completed-expectation minimax problem for common-map rules. The expectation is the ordinary expectation under the product law generated by $P$, and the loss is the supremum error along the support boundary. The infimum therefore ranges over estimators that share the same measurable distance-data transformation across all admissible laws and all boundary points.

For comparison, the point-indexed decision class lets the measurable section depend on the query point while remaining law independent. This keeps the unsigned distance information set fixed and enlarges the admissible class by allowing a separate Borel rule at each $x$.

**Definition 13.** [`def:point-indexed-distance-decision-class`] (Point-indexed class $\mathcal{T}_n^{\mathrm{PI}}$) *The point-indexed decision class is*

$$\mathcal{T}_n^{\mathrm{PI}} = \left\{ \widehat{\mu}_n^{\mathrm{PI}} : \begin{array}{l} \text{there exists a family } \{T_{n,x} : x \in \mathbb{R}^2\} \text{ selected independently of } P, \\ T_{n,x} : (\mathbb{R} \times [0,\infty))^n \to \mathbb{R} \text{ is Borel measurable for every fixed } x, \\ \widehat{\mu}_{n,P}^{\mathrm{PI}}(x) = T_{n,x}(V_{n,P}(x)) \text{ for every } P \in \mathcal{P}_{\mathrm{NP}}(L,q) \\ \text{and every } x \in \mathrm{bd}(\mathcal{X}_P) \end{array} \right\}.$$

*This class consists exactly of such law-independent fixed-x Borel sections.*

**Definition 14.** [def:point-indexed-distance-risk] (Point-indexed risk $R_n^{\mathrm{PI}}$) *For an extended nonnegative function $Z$ on the sample space, define the outer expectation by*

$$\mathbb{E}_P^*[Z] = \inf\left\{\mathbb{E}_P[W] : W \text{ is extended nonnegative and measurable, and } W \geq Z \text{ pointwise}\right\}.$$

*The point-indexed distance risk is*

$$R_n^{\mathrm{PI}} = \inf_{\widehat{\mu}_n^{\mathrm{PI}} \in \mathcal{T}_n^{\mathrm{PI}}} \sup_{P \in \mathcal{P}_{\mathrm{NP}}(L,q)} \mathbb{E}_P^*\left[\sup_{x \in \mathrm{bd}(\mathcal{X}_P)} \left|\widehat{\mu}_{n,P}^{\mathrm{PI}}(x) - \mu_P(x)\right|\right].$$

The outer expectation in Definition 14 follows the standard empirical-process convention for suprema that may require measurable majorants [van der Vaart and Wellner, 1996]. With this convention, Definitions 13 and 14 define a sectionwise Borel problem that preserves the same distance-compressed observations as Definition 10. The main lower bounds compare these two unsigned risks after the common notation and signed-distance setup have been fixed.

## 4 Setup: Known-Geometry Signed-Distance Designs

The second experiment keeps the boundary geometry available to the rule and records which side of the interface each nearby observation occupies. This is the setting in which local-polynomial methods use distance as a one-dimensional coordinate while preserving the sign induced by treatment assignment, as in boundary regression designs developed by Cattaneo et al. [2026]. We first fix the assignment, interface, basis, and coordinate conventions that underlie the signed-distance reduction.

**Definition 15.** [synth_4] (Assignment regions $\mathcal{A}_{t,P}$) *We say $\mathcal{A}_{t,P}$ is the arm-$t$ assignment region of $P$, for $t \in \{0,1\}$, when $\mathcal{A}_{0,P}$ and $\mathcal{A}_{1,P}$ are Borel subsets of $\mathcal{X}_P$ satisfying $\mathcal{A}_{0,P} \cup \mathcal{A}_{1,P} = \mathcal{X}_P$ and $\mathcal{A}_{0,P} \cap \mathcal{A}_{1,P} = \varnothing$, with deterministic treatment assignment $T = \mathbf{1}\{X \in \mathcal{A}_{1,P}\}$.*

**Definition 16.** [synth_35] (Known treatment interface $\mathcal{B}$) *We say $\mathcal{B}$ is the known treatment interface for the fixed hypercube geometry when it is the boundary of the treated rectangle,*

$$\mathcal{B} = \partial([-1,1] \times [0,2]).$$

**Definition 17.** [synth_71] (Degree-$p$ Polynomial Basis $r_p$) *For an integer $p \geq 0$ and $u \in \mathbb{R}$, define $r_p(u) = (1, u, \ldots, u^p)^\top \in \mathbb{R}^{p+1}$.*

**Definition 18.** [synth_24] (Score-coordinate projection score) *We say* score *is the score-coordinate projection when, for every potential-outcome observation $w = (y_0, y_1, x) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$,* $\mathrm{score}(w) = x$.

The partition in Definition 15 makes treatment deterministic given location, so the side of the boundary is a feature of the design rather than an additional random assignment variable. The fixed interface in Definition 16 is the rectangular geometry used in the lower-bound construction, while Definition 17 supplies the signed-distance polynomial basis used by the local fits. The projection in Definition 18 records that the score coordinate is the Euclidean covariate.

The known-geometry object packages the support, assignment regions, interface, Euclidean metric, and smoothing weight. The canonical smoothing weight is the one-dimensional uniform kernel $K_\square$, which selects observations whose signed distance lies within the bandwidth window.

**Definition 19.** [def:cty-uniform-kernel] (Uniform kernel $K_\square$) *The fixed uniform kernel is*

$$K_\square(u) = \mathbf{1}\{|u| \le 1\}, \qquad u \in \mathbb{R}.$$

**Definition 20.** [def:cty-known-geometry] (Known geometry $G_P$) *Define $\mathcal{A}$ to be the class of admissible known geometries $G = (\mathcal{X}, \mathcal{A}_0, \mathcal{A}_1, \mathcal{B}, d, K_\square)$, where $\mathcal{A}_0, \mathcal{A}_1$ are a Borel partition of $\mathcal{X}$,*

$$\mathcal{B} = \mathrm{bd}(\mathcal{A}_0) \cap \mathrm{bd}(\mathcal{A}_1) \subset \mathrm{int}(\mathcal{X}), \qquad d(x,z) = \|x - z\|_2.$$

*For a member law $P$, write $\mathcal{A}_{0,P}, \mathcal{A}_{1,P}$ for its assignment regions and write $\mathcal{B}_P = \mathrm{bd}(\mathcal{A}_{0,P}) \cap \mathrm{bd}(\mathcal{A}_{1,P})$. Define its known design object by*

$$G_P = (\mathcal{X}_P, \mathcal{A}_{0,P}, \mathcal{A}_{1,P}, \mathcal{B}_P, d_P, K_\square) \in \mathcal{A}.$$

**Definition 21.** [def:cty-known-geometry-signed-distance-data] (Signed-distance data $U^{\pm}_{n,P}(x)$) *For $x \in \mathcal{B}_P$, define the signed distance*

$$D^{\pm}_{P,x} = \{\mathbf{1}(X \in \mathcal{A}_{1,P}) - \mathbf{1}(X \in \mathcal{A}_{0,P})\}\|X - x\|_2$$

*and the signed-distance sample*

$$U^{\pm}_{n,P}(x) = ((Y_i, D^{\pm}_{P,x,i}) : 1 \le i \le n).$$

The known design object $G_P$ in Definition 20 gives the rule the relevant boundary and side labels. For an interface point $x$, the signed distance $D^{\pm}_{P,x}$ in Definition 21 retains the Euclidean distance magnitude and assigns its sign according to the treatment region. Thus the signed-distance sample $U^{\pm}_{n,P}(x)$ provides the one-dimensional running coordinate used by local-polynomial RD estimators [Fan and Gijbels, 1996, Ruppert and Wand, 1994, Calonico et al., 2014, 2020, Imbens and Wager, 2019].

The law class below is the displayed Euclidean, uniform-kernel specialization of the CTY Assumptions 1–2 boundary design. Its clauses have four statistical roles. The support, assignment regions, interface, metric, and signed coordinate define the experiment and information available to a rule. The density, regression smoothness, conditional variance, and moment bounds provide uniform design and outcome envelopes. The population-Gram, arm-mass, distance-slice, and VC conditions keep local-polynomial fitting uniformly conditioned along the interface. The separate inputs $\mathsf{AI}_{p,\nu,L}$, introduced in Definition 33, supply identification, first-order approximation, and expected uniform stochastic control for the conditional upper bound.

**Definition 22.** [def:cty-a1-a2-class] (A1/A2 law class $\mathcal{P}_{12}(p,\nu,L)$) *For integers $p \ge 0$, real numbers $\nu \ge 2$, and $L \ge 4$, let $\mathcal{P}_{12}(p,\nu,L)$ be the set of laws $P$ satisfying the following conditions.*

*Write $Y_i(0), Y_i(1)$ for the sampled potential outcomes of observation $i$, and write $\mathcal{H}^1(\mathcal{B}_P)$ for the one-dimensional Hausdorff measure of the common interface.*

- ***(Sampling and consistency.)*** *For every $n$, the vectors $(Y_i(0), Y_i(1), X_i)$, $1 \le i \le n$, are independently and identically distributed,*

$$T = \mathbf{1}\{X \in \mathcal{A}_{1,P}\}, \qquad Y = TY(1) + (1-T)Y(0).$$

- ***(Rectangular support and density.)*** *The support has the form*

$$\mathcal{X}_P = [\underline{x}_P, \overline{x}_P]^2 \subseteq [-L, L]^2.$$

*The variable $X$ has a Lebesgue density $f_P$, continuous on $\mathcal{X}_P$, such that*

$$L^{-1} \leq f_P(x) \leq L, \qquad x \in \mathcal{X}_P.$$

- ***(Smooth potential-outcome regressions.)*** *For each $t \in \{0,1\}$, the regression*

$$\mu_{t,P}(x) = \mathbb{E}_P[Y(t) \mid X = x]$$

*has a $C^{p+1}$ extension to an open neighborhood of $\mathcal{X}_P$, and*

$$\max_{0\leq|\alpha|\leq p} \sup_{x\in\mathcal{X}_P} |\partial^\alpha \mu_{t,P}(x)| + \max_{0\leq|\alpha|\leq p} \sup_{x\neq z\in\mathcal{X}_P} \frac{|\partial^\alpha \mu_{t,P}(x) - \partial^\alpha \mu_{t,P}(z)|}{\|x-z\|_2} \leq L.$$

- ***(Conditional variances.)*** *For each $t \in \{0,1\}$, the conditional variance*

$$\sigma^2_{t,P}(x) = \mathrm{Var}_P(Y(t) \mid X = x)$$

*is continuous and satisfies*

$$L^{-1} \leq \sigma^2_{t,P}(x) \leq L, \qquad x \in \mathcal{X}_P.$$

- ***(Moment envelope.)*** *For each $t \in \{0,1\}$ and every $x \in \mathcal{X}_P$,*

$$\mathbb{E}_P[|Y(t)|^{2+\nu} \mid X = x] \leq L.$$

- ***(Assignment geometry.)*** *The sets $\mathcal{A}_{0,P}$ and $\mathcal{A}_{1,P}$ form a Borel partition of $\mathcal{X}_P$, and their known common interface $\mathcal{B}_P \subset \mathrm{int}(\mathcal{X}_P)$ is a compact rectifiable curve satisfying*

$$L^{-1} \leq \mathcal{H}^1(\mathcal{B}_P) \leq L.$$

- ***(Metric, kernel, and VC bound.)*** *The known metric and kernel are*

$$d_P(x,z) = \|x-z\|_2, \qquad K_\square.$$

*The collection*

$$\{\{z \in \mathcal{X}_P : \|z-x\|_2 \leq r\} : x \in \mathcal{X}_P,\ r \geq 0\}$$

*is a VC class with VC index at most four.*

- ***(Population Gram lower bound.)*** *Put*

$$r_p(u) = (1, u, \ldots, u^p)^\top, \qquad I_1 = [0,\infty), \qquad I_0 = (-\infty, 0),$$

*and*

$$\Psi_{t,P,x}(h) = \mathbb{E}_P\left[\frac{1}{h^2}\mathbf{1}\{D^\pm_{P,x} \in I_t\} K_\square(D^\pm_{P,x}/h) r_p(D^\pm_{P,x}/h) r_p(D^\pm_{P,x}/h)^\top\right].$$

*For every $0 < h \leq L^{-1}$, every $x \in \mathcal{B}_P$, and both $t \in \{0,1\}$,*

$$\lambda_{\min}(\Psi_{t,P,x}(h)) \geq L^{-1}.$$

- ***(Small-bandwidth arm mass.)*** *For every $0 < h \leq L^{-1}$, every $x \in \mathcal{B}_P$, and both $t \in \{0,1\}$,*

$$\frac{1}{h^2}\int_{\mathcal{A}_{t,P}} K_\square(\{(2t-1)\|z-x\|_2\}/h)\,dz \geq L^{-1}.$$

- ***(Distance-slice density.)*** *For every $x \in \mathcal{B}_P$, both $t \in \{0,1\}$, and every $0 < s \leq L^{-1}$, the one-dimensional Hausdorff integral of $f_P$ over*

$$\{z \in \mathcal{A}_{t,P} : \|z-x\|_2 = s\}$$

*is finite and strictly positive.*

*This class is the displayed L-uniformized Euclidean-distance, uniform-kernel specialization of the two baseline smoothness and design conditions, augmented by the displayed envelope restrictions and quantitative small-bandwidth thresholds.*

The assumptions in Definition 22 fall into three groups. First, the support, assignment partition, rectifiable interface, signed coordinate, and distance-slice conditions define the design and the information observed by the signed-distance procedure. Second, the density, moments, smoothness, variance, arm-mass, population-Gram, and VC conditions provide uniform outcome, conditioning, and complexity envelopes for local fitting. Third, $\mathsf{AI}_{p,\nu,L}$ in Definition 33 collects the identification, bias, and expected maximal inputs invoked specifically by the conditional upper bound. The unconditional signed lower bound uses the fixed-geometry subexperiment inside the first two groups; the stabilized estimator and conditional frontier use all three groups.

The potential outcomes $Y(t)$ and deterministic treatment $T$ in Definition 22 give the boundary target its causal RD interpretation. The arm-specific regressions $\mu_{t,P}$ support degree-$p$ signed-distance approximation, while the arm-specific variances $\sigma^2_{t,P}$ and $2+\nu$ moment envelope keep the outcome scale uniform. Geometrically, the class includes compact rectifiable interfaces of controlled length inside rectangular supports and requires two-sided local mass and stable signed-distance Gram matrices at every interface point. These quantitative conditions also cover the corners of the fixed treated rectangle: Lemma 6 verifies the Gram and arm-mass thresholds there rather than relying on a smooth-boundary shortcut.

The treatment-effect target is the jump between the two smooth potential-outcome regression surfaces at the interface. The population local-polynomial coefficient records the signed-distance normal equations that the estimator approximates.

**Definition 23.** [`synth_46`] (Interface treatment-effect curve $\tau_P(x)$) *For a law $P \in \mathcal{P}_{12}(p,\nu,L)$ and an interface point $x \in \mathcal{B}_P$, define the treatment-effect curve by $\tau_P(x) = \mu_{1,P}(x) - \mu_{0,P}(x)$, where $\mu_{t,P}(x) = \mathbb{E}_P[Y(t) \mid X = x]$ is the continuous conditional-mean version for arm $t \in \{0,1\}$.*

**Definition 24.** [`synth_128`] (Population local-polynomial coefficient $\beta^*_{t,P,x}(h)$) *We say $\beta^*_{t,P,x}(h)$ is the unwinsorized population degree-p signed-distance local-polynomial coefficient when $\beta^*_{t,P,x}(h) \in \mathbb{R}^{p+1}$ satisfies the normal equations*

$$\Psi_{t,P,x}(h)\,\beta^*_{t,P,x}(h) = \mathbb{E}_P\left[\frac{1}{h^2}\mathbf{1}\{D^\pm_{P,x} \in I_t\}K_\square(D^\pm_{P,x}/h)r_p(D^\pm_{P,x}/h)Y\right].$$

The curve $\tau_P$ in Definition 23 is evaluated uniformly over $\mathcal{B}_P$. The coefficient in Definition 24 is a population projection onto the signed-distance polynomial basis; its intercept difference is the population analogue of the treatment-effect estimator, and its approximation error is the first-order bias component isolated later.

Rules in the signed-distance experiment may depend on the known geometry and the query point through law-independent Borel sections. The risk then uses the outer expectation already introduced for point-indexed suprema in Definition 14.

**Definition 25.** [`def:cty-a1-a2-point-indexed-decision-class`] (Known-geometry rules $\mathcal{T}_n^{12,\pm}$) *Let $\mathcal{T}_n^{12,\pm}$ be the class of rules for which there is one collection*

$$\{T_{n,G,x} : G \textit{ is an admissible known geometry}, \; x \in \mathbb{R}^2\}$$

*selected independently of the unknown member law, such that every fixed $(G, x)$ section is a Borel map from $(\mathbb{R} \times \mathbb{R})^n$ to $\mathbb{R}$ and, for every $P \in \mathcal{P}_{12}(p, \nu, L)$, every $p \geq 0$, and every $x \in \mathcal{B}_P$,*

$$\widehat{\tau}_{n,P}(x) = T_{n,G_P,x}(U^{\pm}_{n,P}(x)).$$

*The imposed regularity is fixed-$(G, x)$ Borel measurability of each section.*

**Definition 26.** [`def:cty-a1-a2-outer-risk`] (Signed-distance minimax risk $R_n^{12,\pm}(p, \nu, L)$) *Using the outer expectation $\mathbb{E}_P^*$ defined in Definition 14, for $p \geq 0$ the signed-distance minimax risk is*

$$R_n^{12,\pm}(p, \nu, L) = \inf_{\widehat{\tau}_n \in \mathcal{T}_n^{12,\pm}} \sup_{P \in \mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^* \left[ \sup_{x \in \mathcal{B}_P} |\widehat{\tau}_{n,P}(x) - \tau_P(x)| \right].$$

**Definition 27.** [`synth_126`] (Fixed-geometry signed-distance outer risk $R_n^{12,\pm,\mathrm{fix}}(p, \nu, L)$) *We define $R_n^{12,\pm,\mathrm{fix}}(p, \nu, L)$ as the signed-distance outer minimax risk with the law supremum restricted to the fixed hard geometry:*

$$R_n^{12,\pm,\mathrm{fix}}(p, \nu, L) = \inf_{\widehat{\tau}_n \in \mathcal{T}_n^{12,\pm}} \sup_{P \in \mathcal{P}_{12}(p,\nu,L):\ \mathcal{X}_P = [-3,3]^2,\ \mathcal{A}_{1,P} = [-1,1] \times [0,2],\ \mathcal{A}_{0,P} = \mathcal{X}_P \setminus \mathcal{A}_{1,P},\ \mathcal{B}_P = \partial \mathcal{A}_{1,P},\ d_P(x,z) = \|x-z\|_2,\ \textit{and the}}$$

The decision class $\mathcal{T}_n^{12,\pm}$ in Definition 25 gives procedures access to the known geometry while keeping the section family fixed before the member law is chosen. The signed-distance risk $R_n^{12,\pm}$ in Definition 26 measures the expected worst interface error for treatment-effect estimation. The fixed-geometry risk in Definition 27 is the same criterion restricted to the rectangular hard family, which is the family used for the signed-distance converse in Theorem 3.

It remains to name the estimator used for the upper bound. The construction follows the standard local-polynomial template, with three stabilizing features: outcomes are winsorized at a bandwidth-dependent level, the empirical Gram matrix is inverted only when it has a uniform quadratic-form lower bound, and the final treatment-effect estimate is clipped to the natural envelope scale.

**Definition 28.** [`def:cty-stabilized-local-polynomial-estimator`] (Stabilized local-polynomial estimator $\widehat{\tau}_{n,h}^{12,\mathrm{LP}}$) *For $B \in \mathbb{R}$, define the winsorization map by*

$$\psi_B(y) = \operatorname{sign}(y) \min\{|y|, B\}.$$

*For each $t \in \{0,1\}$, query point $x$, and bandwidth $h$, define*

$$\widehat{\Psi}_{t,x}(h) = \frac{1}{nh^2} \sum_{i=1}^{n} \mathbf{1}\{D^{\pm}_{P,x,i} \in I_t\} K_{\square}(D^{\pm}_{P,x,i}/h) r_p(D^{\pm}_{P,x,i}/h) r_p(D^{\pm}_{P,x,i}/h)^{\top}.$$

*With $B(h) = h^{-1/3}$, define*

$$\widehat{m}^{B}_{t,x}(h) = \frac{1}{nh^2} \sum_{i=1}^{n} \mathbf{1}\{D^{\pm}_{P,x,i} \in I_t\} K_{\square}(D^{\pm}_{P,x,i}/h) r_p(D^{\pm}_{P,x,i}/h) \psi_{B(h)}(Y_i).$$

*The stabilized coefficient is*

$$\widehat{\beta}^{B}_{t,x}(h) = \begin{cases} \widehat{\Psi}_{t,x}(h)^{-1} \widehat{m}^{B}_{t,x}(h), & \text{if } \forall v \in \mathbb{R}^{p+1},\ (2L)^{-1} \sum_{j=0}^{p} v_j^2 \le v^{\top} \widehat{\Psi}_{t,x}(h) v, \\ 0, & \text{otherwise.} \end{cases}$$

*With $e_1 = (1, 0, \dots, 0)^{\top}$, the clipped signed-distance local-polynomial treatment-effect estimator is*

$$\widehat{\tau}^{12,\mathrm{LP}}_{n,h}(x) = \max\Big\{-2L, \min\Big\{2L, e_1^{\top}(\widehat{\beta}^{B}_{1,x}(h) - \widehat{\beta}^{B}_{0,x}(h))\Big\}\Big\}.$$

The winsorization map $\psi_B$ in Definition 28 converts the moment envelope into a bounded-score component at bandwidth $h$. The empirical Gram matrix estimates the population Gram $\Psi_{t,P,x}(h)$, and the fallback rule keeps the coefficient map stable on samples where the empirical design is ill conditioned. The resulting estimator $\widehat{\tau}^{12,\mathrm{LP}}_{n,h}$ is the concrete rule used in the signed-distance upper result.

The following three deviation objects separate approximation, design, and score variation. They are stated here because the main theorem collects analytic inputs in terms of these quantities before applying the estimator in Definition 28.

**Definition 29.** [`synth_39`] (Uniform first-order bias ratio $\mathrm{BiasRatio}_{p,\nu',L}(h_n)$) *For $h_n > 0$, define the uniform first-order bias ratio by*

$$\mathrm{BiasRatio}_{p,\nu',L}(h_n) = \sup_{P \in \mathcal{P}_{12}(p,\nu',L)} \sup_{x \in \mathcal{B}_P} \frac{\left| e_1^{\top}\{\beta^{*}_{1,P,x}(h_n) - \beta^{*}_{0,P,x}(h_n)\} - \tau_P(x) \right|}{h_n}.$$

**Definition 30.** [`synth_62`] (Uniform empirical Gram deviation $\mathrm{GramDev}_{n,p,P}(h_n)$) *For a law $P \in \mathcal{P}_{12}(p,\nu,L)$ and bandwidth $h_n > 0$, we write* $\mathrm{GramDev}_{n,p,P}(h_n)$ *for the extended nonnegative sample-space random variable*

$$\mathrm{GramDev}_{n,p,P}(h_n) = \sup_{t \in \{0,1\}} \sup_{x \in \mathcal{B}_P} \max_{0 \le j,k \le p} \left| \widehat{\Psi}_{t,x}(h_n)_{jk} - \Psi_{t,P,x}(h_n)_{jk} \right|.$$

**Definition 31.** [`synth_63`] (Uniform raw-score deviation $\mathrm{ScoreDev}_{n,p,P}(h_n)$) *Define* $\mathrm{ScoreDev}_{n,p,P}(h_n)$ *as the sample-dependent uniform centered local-polynomial score deviation*

$$\mathrm{ScoreDev}_{n,p,P}(h_n) = \sup_{t \in \{0,1\}} \sup_{x \in \mathcal{B}_P} \left\| \frac{1}{nh_n^2} \sum_{i=1}^{n} \mathbf{1}\{D^{\pm}_{P,x,i} \in I_t\} K_{\square}(D^{\pm}_{P,x,i}/h_n) r_p(D^{\pm}_{P,x,i}/h_n) Y_i - \mathbb{E}_P\Big[\frac{1}{h_n^2} \mathbf{1}\{D^{\pm}_{P,x} \in I_t\} K_{\square}$$

*with the norm taken in $\mathbb{R}^{p+1}$.*

**Definition 32.** [`synth_131`] (Winsorized score $\mathrm{score}^B_{t,x,h,j}$) *We say* $\mathrm{score}^B_{t,x,h,j}$ *is the winsorized local-polynomial residual score when, for* $P \in \mathcal{P}_{12}(p,\nu,L)$*, sample size* $n$*, arm* $t \in \{0,1\}$*, interface point* $x \in \mathcal{B}_P$*, bandwidth* $h > 0$*, winsorization level* $B > 0$*, and coordinate* $j \in \{0,\dots,p\}$*, it is the* $P^n$*-random variable*

$$\mathrm{score}^B_{t,x,h,j} = \frac{1}{nh^2}\sum_{i=1}^n \mathbf{1}\{D^{\pm}_{P,x,i} \in I_t\} K_{\square}\left(\frac{D^{\pm}_{P,x,i}}{h}\right) r_p\left(\frac{D^{\pm}_{P,x,i}}{h}\right)_j \left[\psi_B(Y_i) - r_p\left(\frac{D^{\pm}_{P,x,i}}{h}\right)^{\top} \beta^*_{t,P,x}(h)\right],$$

*with* $\beta^*_{t,P,x}(h)$ *the unwinsorized population local-polynomial coefficient at* $(t,x,h)$*.*

The bias ratio in Definition 29 measures the uniform first-order approximation error of the population signed-distance fit. The Gram deviation in Definition 30 tracks the uniform empirical conditioning of the local design, and the raw-score deviation in Definition 31 records the centered stochastic part before winsorization. The winsorized score in Definition 32 is the bounded-envelope score component used by the maximal inequality in the upper-bound appendix. These objects place the later risk result in the usual bias, design-stability, and empirical-process decomposition for local-polynomial RD estimators [Fan and Gijbels, 1996, Ruppert and Wand, 1994, Calonico et al., 2014, 2020, van der Vaart and Wellner, 1996].

## 5 Main Results

The results now put the two distance-compressed experiments on a common risk scale. The unsigned problem uses the nonparametric regression class and distance data introduced in Definitions 9 and 10; the signed problem uses the known-geometry law class, signed-distance data, and treatment-effect risk in Definitions 21, 22 and 26. Throughout the section, the normalizing sequence is the rate $a_n$ in Definition 8.

The first result is the same-class converse for the unsigned common-map experiment. It evaluates exactly the decision class in Definition 11 under the completed-expectation risk in Definition 12.

**Theorem 1.** [`thm:cty-same-class-log-converse`] (Same-Class Log Converse) *For every integer* $q \geq 1$ *and every* $L \geq 4$*, there exists a constant* $c_{\mathrm{CTY}} = c_{\mathrm{CTY}}(q,L) > 0$ *such that, for the common-map minimax risk* $R_n^{\mathrm{NP}}$ *in Definition 12 and* $a_n = (\log n/n)^{1/4}$*,*

$$\liminf_{n\to\infty}\left(\frac{n}{\log n}\right)^{1/4} R_n^{\mathrm{NP}} = \liminf_{n\to\infty}\frac{R_n^{\mathrm{NP}}}{a_n} \geq c_{\mathrm{CTY}}.$$

The theorem establishes that a single law-independent distance rule incurs expected boundary sup-loss at least of logarithmic order $a_n$ over $\mathcal{P}_{\mathrm{NP}}(L,q)$. The packing calculation behind the rate is easiest to read from the separated boundary construction used in the appendix. If the signal amplitude is $\Delta$, then the construction places $M \asymp \Delta^{-1/q}$ active locations along the boundary and uses cells with area scale $w^2 \asymp \Delta^{2/q}$. Unsigned distance compression leaves each coordinate with information of order $n\Delta^4$, so the many-coordinate testing barrier is calibrated by

$$n\Delta^4 \asymp \log M, \qquad \text{hence} \qquad \Delta \asymp \left(\frac{\log n}{n}\right)^{1/4}.$$

The smoothness order $q$ affects the packing constants and the geometric calibration of $w$, while the exponent is governed by the fourth-order distance-compression information calculation. In the signed-distance local-polynomial results below, the polynomial order $p$ likewise affects constants and regularity thresholds while the displayed rate normalization remains $a_n$.

The next result carries the same logarithmic converse to the point-indexed unsigned class. The point-indexed class in Definition 13 permits a separate Borel section at each query point, and the risk is evaluated with outer expectation as in Definition 14.

**Theorem 2.** [`thm:point-indexed-distance-log-converse`] (Logarithmic Distance Converse) *For every $q \in \mathbb{N}$ and $L \in \mathbb{R}$ satisfying $q \geq 1$ and $L \geq 4$, there exists a constant $c_{\mathrm{PI}} = c_{\mathrm{PI}}(q, L) > 0$ such that, with $R_n^{\mathrm{PI}}$ the point-indexed outer-expectation minimax risk in Definition 14 and $a_n = (\log n/n)^{1/4}$,*

$$\liminf_{n\to\infty} \left(\frac{n}{\log n}\right)^{1/4} R_n^{\mathrm{PI}} = \liminf_{n\to\infty} \frac{R_n^{\mathrm{PI}}}{a_n} \geq c_{\mathrm{PI}}.$$

Thus the lower-bound scale is a property of the unsigned distance information itself, even when the rule may choose its measurable section point by point. Outer expectation in this statement is the same measurable-majorant criterion introduced in Definition 14, which is the standard way to state uniform risks over potentially nonseparable boundary index sets [van der Vaart and Wellner, 1996].

The signed-distance lower bound uses the known-geometry experiment in Definitions 25 and 26. It is stated for all local-polynomial orders after an envelope threshold that depends on the order.

**Theorem 3.** [`thm:cty-a1-a2-point-indexed-log-converse-all-orders`] (Point-Indexed Log Converse) *For every nonnegative integer $p$, there exists a real number $L_0 = L_0(p)$ such that:*

- ***(Envelope threshold.)*** *$L_0 \geq 48$.*
- ***(Moment exponent.)*** *$\nu \geq 2$.*
- ***(Uniform envelope.)*** *$L \geq L_0$.*
- ***(Risk definition.)*** *$R_n^{12,\pm}(p, \nu, L)$ is the signed-distance minimax risk in Definition 26.*
- ***(Fixed-geometry risk.)*** *$R_n^{12,\pm,\mathrm{fix}}(p, \nu, L)$ is the infimum over the same known-geometry point-indexed decision rules of the same worst-case outer-expected interface sup-loss, with the law supremum further restricted to laws in $\mathcal{P}_{12}(p, \nu, L)$ satisfying the fixed hard-geometry condition.*

*For every $\nu$ and $L$ satisfying these conditions, there exists a constant $c = c(p, \nu, L) > 0$ such that, writing $a_n = (\log n/n)^{1/4}$,*

$$\liminf_{n\to\infty} a_n^{-1} R_n^{12,\pm}(p, \nu, L) = \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n},$$

*and*

$$c \leq \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n}, \qquad c \leq \liminf_{n\to\infty} \frac{R_n^{12,\pm,\mathrm{fix}}(p, \nu, L)}{a_n}.$$

Theorem 3 establishes the same logarithmic lower scale for the known-geometry signed-distance minimax risk. The second inequality records that the lower bound is already present inside the fixed rectangular geometry from Definition 27. Consequently, the converse is calibrated by a concrete Euclidean boundary experiment rather than by variation over arbitrary interfaces.

The upper result for the signed-distance experiment combines the estimator in Definition 28 with three analytic inputs. These inputs are distance identification, uniform first-order bias control, and expected maximal inequalities; every signed-distance upper or matched-rate statement below is conditioned on this collection.

**Definition 33.** [`synth_139`] (Analytic inputs for the signed-distance upper bound) *Define $\mathsf{AI}_{p,\nu,L}$ as the following three signed-distance analytic inputs used by the expected-risk upper bound.*

- ***(CTY distance identification.)*** *For every A1/A2 law $P$ and metric $d$, the support is rectangular, the density is continuous and positive on the support, both potential outcomes are integrable, both arm regressions have $C^{p+1}$ extensions, the boundary is a positive-length rectifiable curve, $d$ is a metric uniformly equivalent to Euclidean distance on the support, and every sufficiently small positive armwise distance slice at every boundary point has finite positive density mass. Under these conditions there is one source-coherent selected conditional law of the observed outcome given signed distance, with armwise mean versions $\theta_1, \theta_0$ whose one-sided limits at zero satisfy*

$$\tau_P(x) = \lim_{u \downarrow 0} \theta_1(u) - \lim_{u \uparrow 0} \theta_0(u).$$

- ***(Uniform first-order bias.)*** *There is $C_b > 0$, depending only on $p, L$, such that for every $\nu' \geq 2$ and every positive antitone deterministic sequence $h_n \to 0$ with $nh_n^2 \to \infty$,*

$$\limsup_{n \to \infty} \mathrm{BiasRatio}_{p,\nu',L}(h_n) \leq C_b.$$

- ***(Expected maximal bounds.)*** *There is $C_m > 0$ such that, along every positive deterministic $h_n \to 0$ satisfying*

$$\frac{n^{(1+\nu)/(2+\nu)} h_n^2}{\log(h_n^{-1})} \to \infty,$$

*all sufficiently large $n$ and every $P \in \mathcal{P}_{12}(p, \nu, L)$ satisfy*

$$\mathbb{E}_P^*[\mathrm{GramDev}_{n,p,P}(h_n)] \leq C_m \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}}$$

*and*

$$\mathbb{E}_P^*[\mathrm{ScoreDev}_{n,p,P}(h_n)] \leq C_m \left\{ \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}} + \frac{\log(h_n^{-1})}{n^{(1+\nu)/(2+\nu)} h_n^2} \right\}.$$

The three components of $\mathsf{AI}_{p,\nu,L}$ are statistical assumptions with distinct roles. *Identification* connects the one-sided signed-distance conditional-mean limits to the boundary treatment-effect curve [Cattaneo et al., 2026]. *Approximation* bounds the population local-polynomial intercept error uniformly at first order in the bandwidth [Fan and Gijbels, 1996, Ruppert and Wand, 1994]. *Stochastic control* bounds the expected uniform Gram and score deviations over arms, boundary points, and laws [Vapnik and Chervonenkis, 1971, Pollard, 1984, van der Vaart and Wellner, 1996]. The next proposition is conditional on all three displayed inputs and derives the expected-risk rate for the stabilized estimator from them.

**Proposition 1.** [`prop:cty-a1-a2-winsorized-expected-outer-upper`] (Expected Outer Upper Bound) *Let $p \geq 0$ be an integer, let $\nu \geq 2$, and let $L \geq 4$. Assume:*

- ***(Distance identification.)*** *For every A1/A2 law $P$, every distance map $d$, and every $x \in \mathcal{B}_P$ satisfying the CTY identification assumptions at order $p$, there are signed-distance conditional-mean versions $\theta_1, \theta_0 : \mathbb{R} \to \mathbb{R}$ and one-sided limits $\theta_1(0+)$ and $\theta_0(0-)$ such that*
$$\tau_P(x) = \theta_1(0+) - \theta_0(0-).$$
- ***(Uniform first-order bias.)*** *There is a constant $C_b > 0$, depending on $p$ and $L$, such that for every $\nu' \geq 2$ and every positive antitone deterministic bandwidth sequence $h_n$ with $h_n \to 0$ and $nh_n^2 \to \infty$,*
$$\limsup_{n\to\infty} \mathrm{BiasRatio}_{p,\nu',L}(h_n) \leq C_b.$$
- ***(Expected maximal bounds.)*** *There is a constant $C_m > 0$ such that, for every positive deterministic bandwidth sequence $h_n$ with $h_n \to 0$ and*
$$\frac{n^{(1+\nu)/(2+\nu)} h_n^2}{\log(h_n^{-1})} \to \infty,$$
*the following bounds hold for all sufficiently large $n$, uniformly over $P \in \mathcal{P}_{12}(p, \nu, L)$:*
$$\mathbb{E}_P^*[\mathrm{GramDev}_{n,p,P}(h_n)] \leq C_m \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}},$$
*and*
$$\mathbb{E}_P^*[\mathrm{ScoreDev}_{n,p,P}(h_n)] \leq C_m \left\{ \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}} + \frac{\log(h_n^{-1})}{n^{(1+\nu)/(2+\nu)} h_n^2} \right\}.$$

*With $a_n = (\log n/n)^{1/4}$ as in Definition 8, take the bandwidth $h_n = a_n$ and the winsorization level $B_n = a_n^{-1/3}$ in the winsorized, Gram-stabilized signed-distance local-polynomial estimator clipped to $[-2L, 2L]$. Then there are constants $C > 0$ and $N$ such that, for every $n \geq N$,*
$$\sup_{P \in \mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^* \left[ \sup_{x \in \mathcal{B}_P} \left| \widehat{\tau}_{n,h_n}^{12,\mathrm{LP}}(x) - \tau_P(x) \right| \right] \leq C a_n.$$

The proposition gives a concrete signed-distance procedure with expected outer risk bounded by a constant multiple of $a_n$. The bandwidth choice balances the first-order signed-distance bias, of order $h_n$, with the uniform stochastic scale $\{\log(h_n^{-1})/(nh_n^2)\}^{1/2}$. At $h_n = a_n$, both terms have order $a_n$, and the additional heavy-tail term in the assumed maximal bound is controlled under the displayed moment condition. The winsorized-score lemma used in the appendix supplies the bounded-envelope score component needed for this decomposition. The full proposition additionally uses the assumed uniform Gram control and the assumed raw-score, law-uniform, heavy-tail maximal bound with exponent $n^{(1+\nu)/(2+\nu)}$, so the displayed expected-risk conclusion is tied to the full collection $\mathsf{AI}_{p,\nu,L}$ in Definition 33.

**Conditional signed-distance frontier.** Under the analytic inputs in Definition 33, the signed-distance lower and upper statements combine to yield the matched outer-expected rate over $\mathcal{P}_{12}(p, \nu, L)$.

**Theorem 4.** [`thm:cty-a1-a2-winsorized-matched-frontier`] (Matched Frontier Rate) *For every integer $p \geq 0$, there exists a constant $L_0 \geq 48$ such that, for every $\nu \geq 2$ and every $L \geq L_0$, the following conditions imply the frontier conclusion below:*

- ***(Distance identification.)*** *For every A1/A2 law $P$, every map $d$ satisfying the CTY identification assumptions for $p$ and $P$, and every $x \in \mathcal{B}_P$, there are selected source-coherent signed-distance conditional-mean versions $\theta_1, \theta_0 : \mathbb{R} \to \mathbb{R}$ whose right and left limits at zero exist and identify*

$$\tau_P(x) = \lim_{u \downarrow 0} \theta_1(u) - \lim_{u \uparrow 0} \theta_0(u).$$

- ***(Uniform first-order bias.)*** *There is a constant depending only on $p$ and $L$ such that, uniformly over every $\nu' \geq 2$ and every positive antitone deterministic bandwidth sequence $h_n \to 0$ with $nh_n^2 \to \infty$, the finite limsup of the normalized uniform first-order bias ratio is bounded by that constant.*

- ***(Expected maximal bounds.)*** *For this $p, \nu, L$, there is a constant $C_{\max} > 0$ such that, along every positive bandwidth sequence $h_n \to 0$ with*

$$\frac{n^{(1+\nu)/(2+\nu)} h_n^2}{\log(h_n^{-1})} \to \infty,$$

*for all sufficiently large $n$ and every $P \in \mathcal{P}_{12}(p, \nu, L)$, the expected outer Gram-deviation supremum satisfies*

$$\mathbb{E}_P^* \left[ \sup_{t \in \{0,1\}} \sup_{x \in \mathcal{B}_P} \sup_{j,k} \left| \widehat{\Psi}_{t,x}(h_n)_{jk} - \Psi_{t,P,x}(h_n)_{jk} \right| \right] \leq C_{\max} \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}},$$

*and the expected outer centered raw-score supremum satisfies*

$$\mathbb{E}_P^* \left[ \sup_{t \in \{0,1\}} \sup_{x \in \mathcal{B}_P} \| \widehat{m}_{t,x}(h_n) - m_{t,P,x}(h_n) \| \right] \leq C_{\max} \left( \sqrt{\frac{\log(h_n^{-1})}{nh_n^2}} + \frac{\log(h_n^{-1})}{n^{(1+\nu)/(2+\nu)} h_n^2} \right).$$

*Then there exist constants $c$ and $C$ with $0 < c \leq C$ such that, for the signed-distance minimax risk $R_n^{12,\pm}(p, \nu, L)$ of Definition 26 and $a_n = (\log n / n)^{1/4}$,*

$$c \leq \liminf_{n \to \infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n} \leq \limsup_{n \to \infty} \frac{R_n^{12,\pm}(p, \nu, L)}{a_n} \leq C.$$

*Moreover, the explicit winsorized, Gram-stabilized degree-$p$ signed-distance local-polynomial estimator with bandwidth $h_n = a_n$ and winsorization level $B_n = a_n^{-1/3}$, clipped to $[-2L, 2L]$, has uniform outer risk over $P \in \mathcal{P}_{12}(p, \nu, L)$ satisfying*

$$\limsup_{n \to \infty} \frac{\sup_{P \in \mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^* \left[ \sup_{x \in \mathcal{B}_P} \left| \widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x) \right| \right]}{a_n} \leq C.$$

Under the analytic inputs in Definition 33, Theorem 4 characterizes the signed-distance known-geometry minimax risk up to constants under the same envelope threshold as the lower bound. The lower inequality is supplied by Theorem 3, while the upper inequality is attained by the explicit winsorized, Gram-stabilized local-polynomial estimator from Definition 28. The result places the signed-distance experiment on the logarithmic expected outer-risk scale $a_n$ for every fixed polynomial order $p$, moment exponent $\nu \geq 2$, and envelope $L$ above the stated order-dependent threshold, with the upper side governed by the three analytic inputs.

# 6 Discussion and Extensions

The preceding results separate two roles of distance in boundary regression designs. For the unsigned experiment, Theorems 1 and 2 establish logarithmic lower bounds for boundary sup-loss when the rule observes only outcomes and scalar Euclidean distances from the query point. The common-map statement applies to the decision class and completed-expectation risk in Definitions 11 and 12, while the point-indexed statement applies to the sectionwise Borel class and outer-expectation risk in Definitions 13 and 14. These conclusions identify an information barrier created by unsigned distance compression over the compact bivariate law class in Definition 9.

The signed-distance experiment has a different information set. There the rule observes signed distances generated by the known geometry in Definitions 20 and 21 and is evaluated over the A1/A2 law class and risk in Definitions 22 and 26. Assuming the analytic inputs collected in Definition 33, Proposition 1 and Theorem 4 combine the signed-distance lower bound with the stabilized local-polynomial upper argument to obtain a two-sided expected outer-risk rate. The role of Cattaneo et al. [2026] is therefore most direct in the signed design: their distance-identification framework supplies the population interpretation of the one-dimensional signed-distance coordinate, while the conditional result here packages the displayed inputs with the new lower bound and stabilized estimator.

The common normalization $a_n = (\log n/n)^{1/4}$ in Definition 8 indexes different delivered objects across the two experiments. The unsigned experiment has unconditional common-map and point-indexed lower bounds: many separated boundary perturbations remain difficult because scalar distance hides their angular locations and treatment sides. The signed experiment has an unconditional fixed-geometry lower bound and, under the three inputs in Definition 33, a conditional upper bound attained by a procedure that uses the treatment side, known boundary geometry, and local-polynomial structure. Accordingly, the shared normalization compares lower scales; a two-sided rate conclusion is available here for the signed experiment conditional on $\mathsf{AI}_{p,\nu,L}$.

The signed-distance result is stated for the known Euclidean geometry in Definition 20 and the one-dimensional uniform kernel in Definition 19. The law class in Definition 22 imposes rectangular support, bounded continuous density, smooth potential-outcome regressions, continuous variances, a $2 + \nu$ moment envelope, rectifiable interior interface, two-sided small-bandwidth mass, and a population Gram lower bound. These conditions describe the boundary designs for which the signed-distance local-polynomial estimator has uniformly controlled bias, stable local design matrices, and expected stochastic fluctuations. They also align with the geographic RD setting in which boundary location and treatment side are part of the design information [Keele and Titiunik, 2015, Keele et al., 2015, Keele and Titiunik, 2016, Kendall et al., 2026].

The Euclidean and envelope restrictions also clarify the intended empirical interpretation.

Geographic RD applications often use distance to a boundary as a practical running coordinate while retaining design knowledge about assignment regions and local geography [Papay et al., 2011, Reardon and Robinson, 2012, Wong et al., 2013, Choi and Lee, 2018, Díaz and Zubizarreta, 2023, Sawada et al., 2024, Merlano, 2025]. The signed-distance analysis formalizes that information set through the known-geometry class, whereas the unsigned analysis formalizes a more compressed experiment in which angular and side information are absent from the observation available at a query point. The results therefore organize comparisons by the data transformation and target risk, rather than by a universal ranking of distance-based empirical strategies.

**Open questions.** The unsigned results in Theorems 1 and 2 are lower bounds only and establish no matching unsigned upper bound here. The signed-distance matched rate in Theorem 4 is conditional on $\mathsf{AI}_{p,\nu,L}$, so a fully self-contained derivation of the signed-distance upper ingredients within the same displayed law class remains an open direction. Further extensions include non-Euclidean metrics, kernels beyond $K_{\square}$, alternative interface regularity classes, and risk criteria adapted to inference rather than uniform expected loss.

# Appendices

## A Direct-Product and Support-Boundary Lower Bound

This appendix records the finite-experiment ingredients used to derive the unsigned logarithmic lower bounds. The argument has two components. The first is an overlap product inequality for many binary coordinates after each coordinate has been compressed. The second is a support-boundary hypercube inside the nonparametric class of Definition 9, constructed so that the boundary regression values differ at many separated points while the corresponding unsigned distance experiments remain close coordinate by coordinate. These ingredients yield the common-map converse in Theorem 1; after the strict inclusion statement below, the same construction yields the point-indexed converse in Theorem 2.

The direct-product step is stated abstractly because the same form is also useful for the signed-distance lower experiment. It converts many small coordinatewise overlaps into a lower bound on the probability of at least one decoding error.

**Lemma 1.** [`lem:coordinatewise-overlap-direct-product`] (Coordinatewise overlap product) *Fix an integer $M \geq 1$. Assume the following conditions.*

- ***(Spaces.)*** *The spaces carrying $Z_0, Z_1, \ldots, Z_M, S_1, \ldots, S_M$ are standard Borel.*
- ***(Binary coordinates.)*** *The variables $\Omega_1, \ldots, \Omega_M$ are independent and uniformly distributed on $\{0, 1\}$.*
- ***(Conditional product structure.)*** *Conditionally on $(\Omega_1, \ldots, \Omega_M)$, the variables $Z_0, Z_1, \ldots, Z_M$ are independent, the law of $Z_j$ depends only on $\Omega_j$ for $1 \leq j \leq M$, and the law of $Z_0$ is independent of $(\Omega_1, \ldots, \Omega_M)$.*
- ***(Compressed coordinate decoding.)*** *For each $j$, $S_j = c_j(Z_j)$ for a measurable map $c_j$. The decoder $\widehat{\Omega}_j$ is a measurable $\{0, 1\}$-valued function of $(S_j, (Z_1, \ldots, Z_M), Z_0)$ and is*

*unchanged when its raw coordinate vector is altered only in coordinate $j$ while $(S_j, Z_{-j}, Z_0)$ is fixed. Define*

$$Q_{j,b} = \mathcal{L}(S_j \mid \Omega_j = b), \qquad \rho_j = 1 - \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}}.$$

*Then every such decentralized decoder satisfies*

$$\Pr\{\widehat{\Omega}_j = \Omega_j \text{ for every } j\} \le \frac{1}{2}\left\{1 + \prod_{j=1}^{M}(1-\rho_j)\right\}.$$

*Moreover, for every real $\kappa < 1$, if*

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \le \max\{\kappa \log M, 0\} \qquad \text{for all } j,$$

*then*

$$\Pr\{\text{at least one decoding error}\} \ge \frac{1}{2}\left[1 - \exp\{-M^{1-\kappa}/2\}\right].$$

*Proof of Lemma 1.* Write $\mathcal{S}_j$ for the measurable space carrying $S_j$, and let $g_j$ be the measurable map representing the $j$th decoder, so that

$$\widehat{\Omega}_j = g_j\big(S_j, (Z_1, \ldots, Z_M), Z_0\big),$$

and so that the invariance hypothesis reads: for every $s \in \mathcal{S}_j$, every $z_0$, and all raw vectors $z, z'$ with $z_k = z'_k$ for every $k \ne j$,

$$g_j(s, z, z_0) = g_j(s, z', z_0).$$

The two laws compared at coordinate $j$ are the compressed conditional laws $Q_{j,b} = \mathcal{L}(S_j \mid \Omega_j = b)$ of Definition 1, and $\rho_j = 1 - \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}}$ with the total variation distance of Definition 7. Since $0 \le \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} \le 1$, we have $0 \le \rho_j \le 1$ for every $j$.

*Step 1: a coordinatewise coupling whose compressions agree with probability $\rho_j$.* The raw bit-conditional law at coordinate $j$ is $\mathcal{L}(Z_j \mid \Omega_j = b)$, and $Q_{j,b}$ is its image under $c_j$. Let $\mathcal{K}_j$ be the common part of $Q_{j,0}$ and $Q_{j,1}$, that is, the measure on $\mathcal{S}_j$ with density

$$\frac{d\mathcal{K}_j}{d(Q_{j,0} + Q_{j,1})} = \min\left\{\frac{dQ_{j,0}}{d(Q_{j,0} + Q_{j,1})}, \frac{dQ_{j,1}}{d(Q_{j,0} + Q_{j,1})}\right\}$$

with respect to $Q_{j,0} + Q_{j,1}$. Its total mass is

$$\mathcal{K}_j(\mathcal{S}_j) = 1 - \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} = \rho_j,$$

because $\min\{a, b\} = \{a + b - |a - b|\}/2$, both densities integrate to one, and half the $L^1$ distance between them is the total variation distance.

Each residual $Q_{j,0} - \mathcal{K}_j$ and $Q_{j,1} - \mathcal{K}_j$ then has total mass $1 - \rho_j$. When $\rho_j < 1$, the maximal coupling of $Q_{j,0}$ and $Q_{j,1}$ is the law on $\mathcal{S}_j \times \mathcal{S}_j$ given by

$$\big(\text{image of } \mathcal{K}_j \text{ under } s \mapsto (s, s)\big) + (1-\rho_j)^{-1}\big\{(Q_{j,0} - \mathcal{K}_j) \times (Q_{j,1} - \mathcal{K}_j)\big\},$$

and when $\rho_j = 1$ it is the image of $Q_{j,0}$ under $s \mapsto (s, s)$. In both cases its two marginals are $Q_{j,0}$ and $Q_{j,1}$, and its diagonal carries mass at least $\mathcal{K}_j(\mathcal{S}_j) = \rho_j$.

Because all the spaces are standard Borel, the regular conditional distributions of $Z_j$ given $c_j(Z_j)$ exist under both raw bit-conditional laws; feeding the two coordinates of the maximal coupling into them produces a law $\mathcal{C}_j$ on pairs of raw coordinate blocks, whose coordinates we write $(Z_j^0, Z_j^1)$, such that

$$\mathcal{L}(Z_j^0) = \mathcal{L}(Z_j \mid \Omega_j = 0), \qquad \mathcal{L}(Z_j^1) = \mathcal{L}(Z_j \mid \Omega_j = 1), \qquad \mathcal{C}_j\big\{c_j(Z_j^0) = c_j(Z_j^1)\big\} \geq \rho_j.$$

The first two equalities hold because the lift preserves each marginal, and the third because the image of $\mathcal{C}_j$ under $(z, z') \mapsto (c_j(z), c_j(z'))$ is exactly the maximal coupling above, whose diagonal has mass at least $\rho_j$.

*Step 2: the coupled experiment reproduces the original one at every vertex.* Let $\mathcal{C}$ be the product of the coordinatewise couplings, that is, the law of the array $(Z_j^0, Z_j^1)_{1\leq j\leq M}$ determined by

$$\mathcal{C}\big\{(Z_1^0, Z_1^1) \in A_1, \ldots, (Z_M^0, Z_M^1) \in A_M\big\} = \prod_{j=1}^{M} \mathcal{C}_j(A_j),$$

and let $\mathcal{R} = \mathcal{L}(Z_0)$, which by the conditional product structure does not depend on $(\Omega_1, \ldots, \Omega_M)$. Write $\mathcal{R}\times\mathcal{C}$ for the law under which $Z_0$ has law $\mathcal{R}$ and is independent of the array. For a vertex $\omega \in \{0,1\}^M$ define the selected raw vector by

$$Z_j(\omega) = Z_j^{\omega_j}, \qquad 1 \leq j \leq M.$$

Since the two marginals of $\mathcal{C}_j$ are the two raw bit-conditional laws at coordinate $j$, the selection map pushes $\mathcal{C}$ forward to the product over $j$ of $\mathcal{L}(Z_j \mid \Omega_j = \omega_j)$, so that

$$\mathcal{L}\big(Z_0, (Z_1(\omega), \ldots, Z_M(\omega))\big) = \mathcal{L}\big(Z_0, (Z_1, \ldots, Z_M) \mid \Omega = \omega\big),$$

which is precisely the conditional product law assumed in the statement.

Consequently, writing

$$G(\omega) = \big\{g_j\big(c_j(Z_j(\omega)), (Z_1(\omega), \ldots, Z_M(\omega)), Z_0\big) = \omega_j \text{ for every } 1 \leq j \leq M\big\}$$

for the simultaneous-correctness event at the vertex $\omega$, the uniform hypercube prior gives

$$\Pr\{\widehat{\Omega}_j = \Omega_j \text{ for every } j\} = 2^{-M} \sum_{\omega\in\{0,1\}^M} (\mathcal{R}\times\mathcal{C})\big(G(\omega)\big).$$

*Step 3: the total-disagreement event.* Let

$$N = \big\{c_j(Z_j^0) \neq c_j(Z_j^1) \text{ for every } 1 \leq j \leq M\big\}$$

be the event that every coupled coordinate has unequal compressed values. It is a product event across coordinates, so Step 1 gives

$$\mathcal{C}(N) = \prod_{j=1}^{M} \mathcal{C}_j\big\{c_j(Z_j^0) \neq c_j(Z_j^1)\big\} \leq \prod_{j=1}^{M} \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} = \prod_{j=1}^{M} (1 - \rho_j),$$

each factor being the complement of an event of probability at least $\rho_j$.

*Step 4: a pointwise bound on the number of correctly decoded vertices.* Fix one realization of $Z_0$ and of the coupled array $(Z_j^0, Z_j^1)_{j \le M}$. We claim

$$\sum_{\omega \in \{0,1\}^M} \mathbf{1}\{G(\omega)\} \le \frac{2^M}{2}\big(1 + \mathbf{1}_N\big).$$

On $N$ this is immediate: each of the $2^M$ indicators is at most one, and the right-hand side equals $2^M$.

Off $N$ there is a coordinate $j$ with $c_j(Z_j^0) = c_j(Z_j^1)$. Let $\omega^{(j)}$ be $\omega$ with coordinate $j$ flipped. Selection at a coordinate $k \ne j$ reads the same bit under $\omega$ and $\omega^{(j)}$, while at coordinate $j$ the two selected blocks have equal compressions by the choice of $j$; hence

$$Z_k(\omega) = Z_k(\omega^{(j)}) \quad (k \ne j), \qquad c_j\big(Z_j(\omega)\big) = c_j\big(Z_j(\omega^{(j)})\big).$$

The invariance hypothesis applied to these two raw vectors therefore gives

$$g_j\big(c_j(Z_j(\omega)), (Z_k(\omega))_{k \le M}, Z_0\big) = g_j\big(c_j(Z_j(\omega^{(j)})), (Z_k(\omega^{(j)}))_{k \le M}, Z_0\big),$$

so this single value cannot equal both $\omega_j$ and $\omega_j^{(j)} = 1 - \omega_j$, and consequently

$$\mathbf{1}\{G(\omega)\} + \mathbf{1}\{G(\omega^{(j)})\} \le 1 \qquad \text{for every } \omega \in \{0,1\}^M.$$

Summing this over all $\omega$ and using that $\omega \mapsto \omega^{(j)}$ is an involution of $\{0,1\}^M$, hence a bijection under which the sum is unchanged,

$$2\sum_{\omega} \mathbf{1}\{G(\omega)\} = \sum_{\omega}\big(\mathbf{1}\{G(\omega)\} + \mathbf{1}\{G(\omega^{(j)})\}\big) \le \sum_{\omega} 1 = 2^M,$$

that is, $\sum_\omega \mathbf{1}\{G(\omega)\} \le 2^M/2$, which is the claim off $N$.

*Step 5: the simultaneous success bound.* Integrating the pointwise bound of Step 4 against $\mathcal{R} \times \mathcal{C}$ and using Step 3,

$$\sum_{\omega \in \{0,1\}^M} (\mathcal{R} \times \mathcal{C})\big(G(\omega)\big) = \int \sum_{\omega} \mathbf{1}\{G(\omega)\}\, d(\mathcal{R} \times \mathcal{C}) \le \frac{2^M}{2}\{1 + \mathcal{C}(N)\} \le \frac{2^M}{2}\left\{1 + \prod_{j=1}^{M}(1 - \rho_j)\right\}.$$

Dividing by $2^M$ and inserting the identity of Step 2 gives the first assertion,

$$\Pr\{\widehat{\Omega}_j = \Omega_j \text{ for every } j\} \le \frac{1}{2}\left\{1 + \prod_{j=1}^{M}(1 - \rho_j)\right\}.$$

*Step 6: calibrating the product by the Kullback–Leibler budget.* Fix a real $\kappa < 1$ and assume

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \le \max\{\kappa \log M, 0\} \qquad \text{for every } j.$$

Since $M \ge 1$ we have $\log M \ge 0$, and we split on the sign of $\kappa$.

If $\kappa \ge 0$, then $\kappa \log M \ge 0$, so $\max\{\kappa \log M, 0\} = \kappa \log M$ and the budget reads $\mathrm{KL}(Q_{j,0}, Q_{j,1}) \le \kappa \log M$. The Bretagnolle–Huber overlap bound applied to the pair $(Q_{j,0}, Q_{j,1})$ states

$$\frac{1}{2}\exp\{-\mathrm{KL}(Q_{j,0}, Q_{j,1})\} \le 1 - \|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} = \rho_j,$$

so monotonicity of $t \mapsto \exp(-t)$ together with the budget gives

$$\rho_j \geq \frac{1}{2}\exp\{-\kappa \log M\} \qquad \text{for every } 1 \leq j \leq M.$$

Every factor $1-\rho_j$ is nonnegative, so the elementary inequality $1-t \leq \exp(-t)$ applied coordinatewise and then the common floor just obtained yield

$$\prod_{j=1}^{M}(1-\rho_j) \leq \exp\left\{-\sum_{j=1}^{M}\rho_j\right\} \leq \exp\left\{-M \cdot \frac{1}{2}\exp(-\kappa \log M)\right\} = \exp\{-M^{1-\kappa}/2\},$$

the final equality because $M\exp(-\kappa \log M) = M^{1-\kappa}$.

If $\kappa < 0$, then $\kappa \log M \leq 0$, so $\max\{\kappa \log M, 0\} = 0$ and the budget forces $\mathrm{KL}(Q_{j,0}, Q_{j,1}) = 0$, hence $Q_{j,0} = Q_{j,1}$ and $\|Q_{j,0} - Q_{j,1}\|_{\mathrm{TV}} = 0$, that is, $\rho_j = 1$ for every $j$. Since $M \geq 1$, the product below has at least one factor, so

$$\prod_{j=1}^{M}(1-\rho_j) = 0 \leq \exp\{-M^{1-\kappa}/2\}.$$

*Step 7: passing to the error probability.* Write $p = \prod_{j=1}^{M}(1-\rho_j)$ and $e = \exp\{-M^{1-\kappa}/2\}$. Then $0 \leq p$, because every factor is a total variation distance; $e \leq 1$, because $M^{1-\kappa} \geq 0$; and $p \leq e$ by Step 6. Hence

$$\frac{1}{2}(1-e) + \frac{1}{2}(1+p) = 1 + \frac{p-e}{2} \leq 1,$$

so subtracting the success bound of Step 5 from one gives

$$\Pr\{\text{at least one decoding error}\} = 1 - \Pr\{\widehat{\Omega}_j = \Omega_j \text{ for every } j\} \geq \frac{1}{2}\left[1 - \exp\{-M^{1-\kappa}/2\}\right].$$

Finally, along any sequence $M_n \to \infty$ with $\kappa < 1$ fixed, $M_n^{1-\kappa} \to \infty$ and hence $\exp\{-M_n^{1-\kappa}/2\} \to 0$, so the displayed lower bound is $1/2 - o(1)$. ☐

The invariance requirement in Lemma 1 matches the distance-compressed experiment: when a rule estimates the value attached to coordinate $j$, the information about the $j$th perturbation that remains after compression is summarized by $S_j$, while the remaining raw coordinates and the background block may be held fixed. The total-variation overlap coefficient measures how often the compressed coordinate laws can be coupled to agree. The Kullback–Leibler consequence gives the logarithmic calibration used in the boundary packing, in the same testing spirit as standard minimax lower-bound arguments [Stone, 1982, Tsybakov, 2009].

The next statement supplies the geometric and probabilistic certificate for the unsigned law class. It places many separated locations on the boundary of a square support and assigns one binary regression perturbation to each local half-disc.

**Lemma 2.** [`lem:cty-support-boundary-angular-packing`] (Angular Packing Certificate) *For every integer $q \geq 1$ and every envelope $L \geq 4$, there exist constants*

$$c_0, c_1, c_w^-, c_w^+, c_{\mathrm{rad}}, \alpha > 0, \qquad \alpha < \frac{1}{8},$$

*such that, for all sufficiently large $n$, writing $a_n = (\log n/n)^{1/4}$, there are an integer $M_n$, a radius $w_n$, a mass $m_n$, boundary points $x_1, \ldots, x_{M_n}$, laws $P_\omega$ indexed by $\omega \in \{0,1\}^{M_n}$, and values $u_{j,0}, u_{j,1}$ such that:*

- ***(Packing size and scale.)***

$$c_0 a_n^{-1/q} \le M_n, \qquad c_w^- a_n^{1/q} \le w_n \le c_w^+ a_n^{1/q}.$$

- ***(Boundary packing.)*** *With $S = [-1/2, 1/2]^2$, each $x_j$ lies on $\mathrm{bd}(S)$, and for $i \ne j$,*

$$\|x_i - x_j\|_2 \ge c_w^- a_n^{1/q}.$$

  *The cells $C_j = B_2(x_j, w_n) \cap S$ are pairwise disjoint.*

- ***(Model class and support.)*** *For every $\omega \in \{0,1\}^{M_n}$, the law $P_\omega$ belongs to $\mathcal{P}_{\mathrm{NP}}(L, q)$ and has covariate support $S$.*

- ***(Cell masses.)*** *For every $\omega$ and every $j$,*

$$P_\omega(X \in C_j) = m_n.$$

- ***(Local dependence on one bit.)*** *If $\omega_j = \omega'_j$, then the restricted laws of $(Y, X)$ on the cell $C_j$ agree under $P_\omega$ and $P_{\omega'}$. The restricted laws of $(Y, X)$ on $S \setminus \bigcup_{j=1}^{M_n} C_j$ are the same for all $\omega$.*

- ***(Boundary signal.)*** *For every $\omega$ and $j$,*

$$\mu_{P_\omega}(x_j) = u_{j,\omega_j}, \qquad |u_{j,1} - u_{j,0}| \ge c_1 a_n.$$

- ***(Radial invariance.)*** *Let $\omega^{(j)}$ be $\omega$ with coordinate $j$ flipped. For every $\omega$ and $j$, the laws of $\|X - x_j\|_2$ under $P_\omega$ and $P_{\omega^{(j)}}$ agree. Moreover, the corresponding one-point distance laws agree after restriction to radii $z_2 \ge c_{\mathrm{rad}} a_n$.*

- ***(Compressed-distance KL bound.)*** *For every $\omega$ and $j$,*

$$\mathrm{KL}\Big(\mathcal{L}_{P_\omega}\{V_{n,P_\omega}(x_j)\}, \mathcal{L}_{P_{\omega^{(j)}}}\{V_{n,P_{\omega^{(j)}}}(x_j)\}\Big) \le \alpha \log M_n.$$

*Proof of Lemma 2.* Fix $q \ge 1$ and $L \ge 4$, and write $S = [-1/2, 1/2]^2$.

*Step 1: the fixed profiles.* Let $\varphi : \mathbb{R}^2 \to [0,1]$ be the normalized radial bump, that is, the smooth radially symmetric function with

$$\varphi(0) = 1, \qquad 0 \le \varphi \le 1, \qquad \varphi(z) = 0 \ \text{ whenever } \ \|z\|_2 \ge 1,$$

and let $\Psi : \mathbb{R} \to [0,1]$ be the smooth transition function, which vanishes on $(-\infty, 0]$ and equals one on $[1, \infty)$. For each integer $j \ge 0$, fix a nonnegative derivative envelope $b_j$ satisfying

$$\|D^j \varphi(z)\| \le b_j \qquad \text{for every } z \in \mathbb{R}^2.$$

These envelopes are finite fixed constants attached to the bump profile.

*Step 2: the constants.* Define the smoothness-dependent derivative scale and the one-observation divergence constant by

$$D_q = 1 + q + 1310720 + \sum_{j=0}^{q} b_j, \qquad C_{\mathrm{KL}} = 1310720,$$

and set

$$c_0 = \frac{1}{24}, \qquad c_1 = \frac{1}{1024D_q}, \qquad c_w^- = c_w^+ = 1, \qquad c_{\mathrm{rad}} = \frac{1}{8}, \qquad \alpha = \frac{1}{16}.$$

Since every $b_j \geq 0$, the scale $D_q$ is finite and satisfies

$$D_q \geq 1, \qquad D_q \geq q, \qquad D_q \geq C_{\mathrm{KL}}.$$

Hence all six constants are positive and $\alpha = 1/16 < 1/8$.

The three displayed inequalities give the divergence budget used below:

$$256\, q\, C_{\mathrm{KL}} \leq 256\, D_q^2 \leq (1024 D_q)^4,$$

the first step because $q \leq D_q$ and $C_{\mathrm{KL}} \leq D_q$, the second because $D_q \geq 1$.

*Step 3: the eventual regime.* With $a_n = (\log n/n)^{1/4}$ as in Definition 8, set

$$w_n = a_n^{1/q}, \qquad \delta_n = \frac{a_n}{1024 D_q}, \qquad M_n = \left\lfloor \frac{1}{12 w_n} \right\rfloor.$$

Because $a_n \to 0$, for all sufficiently large $n$ we have $n \geq 2$, $a_n \leq 1$, and

$$0 < w_n \leq \frac{1}{24}, \qquad 0 < \delta_n \leq \frac{1}{8}.$$

Moreover, since $n a_n^4 = \log n$,

$$n\, C_{\mathrm{KL}}\, \delta_n^4 = C_{\mathrm{KL}} \frac{\log n}{(1024 D_q)^4} \leq \frac{\log n}{256\, q} \qquad \text{by the budget of Step 2,}$$

while $M_n \geq 1/(24 w_n) = \frac{1}{24} a_n^{-1/q}$ gives $\log M_n \geq \frac{1}{4q} \log n - o(\log n)$; hence, enlarging $n$ once more,

$$n\, C_{\mathrm{KL}}\, \delta_n^4 \leq \alpha \log M_n.$$

Finally, for every $r \geq c_{\mathrm{rad}} a_n = a_n/8$,

$$2 \cdot (8\delta_n) \leq \frac{r}{8},$$

because $D_q \geq 1$ makes $16\delta_n = a_n/(64 D_q) \leq a_n/64 \leq r/8$.

Fix any $n$ in this eventual regime; all remaining claims are verified at that $n$.

*Step 4: the packing points and cells.* Take the equispaced grid on the middle half of the lower edge of $S$,

$$x_j = \left( -\frac{1}{4} + \frac{j}{2(M_n + 1)}, -\frac{1}{2} \right), \qquad j = 1, \ldots, M_n,$$

and let $C_j = B_2(x_j, w_n) \cap S$ as in the statement, with $B_2$ the closed Euclidean ball of Definition 3. Each $x_j$ has second coordinate $-1/2$ and first coordinate of modulus strictly below $1/4$, so $x_j \in \mathrm{bd}(S)$; and since $w_n \leq 1/24$, each $C_j$ is the closed upper half-disc $\{z \in B_2(x_j, w_n) : z^{(2)} \geq -1/2\}$, where $z = (z^{(1)}, z^{(2)})$.

From the definition of $M_n$, $M_n \leq 1/(12 w_n)$. The small-radius bound $w_n \leq 1/24$ also gives $1 \leq 1/(12 w_n)$, and therefore

$$M_n + 1 \leq \frac{1}{12 w_n} + \frac{1}{12 w_n} = \frac{1}{6 w_n}.$$

Thus the grid spacing obeys

$$3w_n \le \frac{1}{2(M_n+1)},$$

so that for $i \neq j$

$$\|x_i - x_j\|_2 \ge 3w_n \ge w_n = c_w^- a_n^{1/q},$$

and the closed cells $C_i, C_j$, of radius $w_n$ around centers at distance at least $3w_n$, are disjoint. The radius bounds $c_w^- a_n^{1/q} \le w_n \le c_w^+ a_n^{1/q}$ hold with equality since $c_w^\pm = 1$. For the packing size, $w_n \le 1/24$ forces $1/(12w_n) \ge 2$, and $\lfloor u \rfloor \ge u/2$ for $u \ge 2$, so

$$M_n \ge \frac{1}{24w_n} = \frac{1}{24}\, a_n^{-1/q} = c_0 a_n^{-1/q}.$$

*Step 5: the laws $P_\omega$.* Fix $\omega \in \{0,1\}^{M_n}$. Define the regression profile on $\mathbb{R}^2$ by

$$\mu_\omega(x) = \frac{1}{2} + \frac{x^{(1)}}{8} + \sum_{j=1}^{M_n} \omega_j\, \delta_n\, \varphi\!\left(\frac{x - x_j}{w_n}\right),$$

and the design density on $S$ by

$$f_\omega(x) = 1 + \sum_{j=1}^{M_n} \omega_j\, t_n\big(\|x - x_j\|_2\big)\, \kappa_j(x), \qquad \kappa_j(x) = \begin{cases} \dfrac{x^{(1)} - x_j^{(1)}}{\|x - x_j\|_2}, & x \neq x_j, \\ 0, & x = x_j, \end{cases}$$

where the radial tilt is

$$t_n(r) = -\frac{2\delta_n\, \varphi\!\left(\frac{r}{w_n} e_\circ\right)\, \chi_n(r)}{\max\{r/8,\ 8\delta_n\}}, \qquad \chi_n(r) = \Psi\!\left(\frac{r}{64\delta_n} - 1\right),$$

and $e_\circ$ is a unit vector of $\mathbb{R}^2$ (the bump is radial, so the choice is immaterial). Let $P_\omega$ be the joint law on observation coordinates obtained as follows. Its $X$-marginal is the measure with Lebesgue density $f_\omega \mathbf{1}_S$, and at score $x$ the outcome kernel is the probability measure

$$\mu_\omega(x)\, N(1,1) + \{1 - \mu_\omega(x)\}\, N(0,1).$$

Equivalently, $P_\omega$ is the image under $(x,y) \mapsto (y,x)$ of the product-kernel measure generated by this $X$-marginal and this outcome kernel. The declared regression and variance profiles of the resulting law are

$$\mu_{P_\omega}(x) = \mu_\omega(x), \qquad \sigma^2_{P_\omega}(x) = 1 + \mu_\omega(x)\{1 - \mu_\omega(x)\}.$$

They are $X$-marginal-almost-everywhere versions of the conditional mean and variance in the sense of Definition 9. Since $|x^{(1)}| \le 1/2$ on $S$, the bumps have disjoint supports $B_2(x_j, w_n)$, and $\delta_n \le 1/8$,

$$\frac{1}{4} \le \frac{1}{2} - \frac{1}{16} \le \mu_\omega(x) \le \frac{1}{2} + \frac{1}{16} + \delta_n \le \frac{3}{4} \qquad \text{for } x \in S,$$

so the mixture weight is legitimate and the clipping of $\mu_\omega$ to $[0,1]$ needed off $S$, and to $[1/4, 3/4]$ used in Step 10, are both silent on $S$.

*Step 6: model class, support, and cell masses.* The angular tilt obeys $|t_n(r)\kappa_j(x)| \le 1/4$ uniformly, and the bumps attached to distinct centers have disjoint supports, so

$$\frac{3}{4} \le f_\omega(x) \le \frac{5}{4} \qquad (x \in S),$$

whence $L^{-1} \le f_\omega \le L$ for $L \ge 4$; $f_\omega$ is continuous on $S$, and $S$ is compact with $S \subseteq [-L, L]^2$ and with $\mathrm{bd}(S)$ Lipschitz-parametrized by $[0, 1]$. The conditional variance $1 + \mu_\omega(1 - \mu_\omega)$ is continuous and lies in $[1, 5/4] \subseteq [L^{-1}, L]$. Finally, for $0 \le j \le q$,

$$\delta_n\, w_n^{-j}\, b_j = \big(a_n w_n^{-j}\big) \cdot \frac{b_j}{1024 D_q} = w_n^{q-j} \cdot \frac{b_j}{1024 D_q} \le 1,$$

using $w_n^q = a_n$, $w_n \le 1$ and $b_j \le D_q$; since the $j$th derivative of the localized bump $\delta_n\varphi((\cdot - x_j)/w_n)$ is bounded by $\delta_n w_n^{-j} b_j$, these bounds give $\|D^j \mu_\omega\| \le L$ on $S$ for $0 \le j \le q-1$ together with $L$-Lipschitz continuity of $D^{q-1}\mu_\omega$, that is

$$\mu_\omega \in \mathcal{H}^q(S, L)$$

in the sense of Definition 4. Consequently $P_\omega \in \mathcal{P}_{\mathrm{NP}}(L, q)$ as in Definition 9, with $\mathcal{X}_{P_\omega} = S$.

For the cell masses, the only $\omega$-dependent part of $f_\omega$ on $C_j$ is $\omega_j t_n(\|x - x_j\|_2)\kappa_j(x)$, and reflecting $C_j$ in the vertical line through $x_j$ changes the sign of $\kappa_j$ while fixing $\|x - x_j\|_2$; hence that term integrates to zero over $C_j$ and

$$P_\omega(X \in C_j) = \mathrm{Leb}(C_j) = m_n,$$

where $m_n = \mathrm{Leb}(C_1)$ is common to all $j$ because the cells are translates of each other along the lower edge.

*Step 7: locality in one bit.* The functions $\varphi((\cdot - x_i)/w_n)$ and $t_n(\|\cdot - x_i\|_2)\kappa_i(\cdot)$ vanish outside $B_2(x_i, w_n)$, and the cells are pairwise disjoint. Therefore, on $C_j$ both $f_\omega$ and $\mu_\omega$ depend on $\omega$ only through $\omega_j$, so

$$\omega_j = \omega'_j \quad \Longrightarrow \quad P_\omega\big|_{\{X \in C_j\}} = P_{\omega'}\big|_{\{X \in C_j\}},$$

and on $S \setminus \bigcup_j C_j$ we have $f_\omega \equiv 1$ and $\mu_\omega(x) \equiv \frac{1}{2} + \frac{x^{(1)}}{8}$, so

$$P_\omega\big|_{\{X \in S \setminus \bigcup_j C_j\}} = P_{\omega'}\big|_{\{X \in S \setminus \bigcup_j C_j\}} \qquad \text{for all } \omega, \omega'.$$

*Step 8: boundary signal.* Put

$$u_{j,0} = \frac{1}{2} + \frac{x_j^{(1)}}{8}, \qquad u_{j,1} = u_{j,0} + \delta_n.$$

At $x_j$ every bump attached to another center vanishes, because $\|x_j - x_i\|_2 \ge 3w_n > w_n$ for $i \ne j$, while $\varphi(0) = 1$; hence

$$\mu_{P_\omega}(x_j) = u_{j,\omega_j}, \qquad |u_{j,1} - u_{j,0}| = \delta_n = c_1 a_n.$$

*Step 9: radial invariance.* Let $\omega^{(j)}$ be $\omega$ with coordinate $j$ flipped. For every measurable $A \subseteq [0, \infty)$, the reflection of Step 6 applied to the annular slice $\{x : \|x - x_j\|_2 \in A\}$ again cancels the $\omega_j$-dependent density term, so

$$\mathcal{L}_{P_\omega}\big(\|X - x_j\|_2\big) = \mathcal{L}_{P_{\omega^{(j)}}}\big(\|X - x_j\|_2\big).$$

For the tail statement, fix $r \geq c_{\mathrm{rad}} a_n$. The last display of Step 3 gives $16\delta_n \leq r/8$, so $r/(64\delta_n) - 1 \geq 1$ and $r/8 \geq 8\delta_n$; therefore the cutoff is fully active,

$$\chi_n(r) = 1, \qquad t_n(r) = -\frac{16\,\delta_n\,\varphi\Big(\frac{r}{w_n} e_\circ\Big)}{r}.$$

Let $A \subseteq [c_{\mathrm{rad}} a_n, \infty)$ be measurable and let

$$D_{j,A} = \Big\{ x \in B_2(x_j, w_n) : x^{(2)} > -\tfrac{1}{2},\ \|x - x_j\|_2 \in A \Big\}$$

be the part of the cell $C_j$ on which the two vertices differ. Assume $\omega_j = 1$, so that $\omega_j^{(j)} = 0$; the reverse case is symmetric. On $D_{j,A}$ the difference of the two outcome-weighted densities is

$$\mu_\omega f_\omega - \mu_{\omega^{(j)}} f_{\omega^{(j)}} = \delta_n \varphi\Big(\tfrac{r}{w_n} e_\circ\Big) + \left\{ \frac{1}{2} + \frac{x^{(1)}}{8} + \delta_n \varphi\Big(\tfrac{r}{w_n} e_\circ\Big) \right\} t_n(r)\, \kappa_j(x), \qquad r = \|x - x_j\|_2,$$

because on the cell all other bumps and angular corrections vanish. Splitting the brace into the radius-only part $\frac{1}{2} + \frac{x_j^{(1)}}{8} + \delta_n \varphi(\frac{r}{w_n} e_\circ)$ and the remainder $\frac{1}{8}(x^{(1)} - x_j^{(1)})$, the first part contributes $\int_{D_{j,A}} g(r) \kappa_j = 0$ by the same horizontal reflection, and the second contributes

$$\int_{D_{j,A}} \frac{x^{(1)} - x_j^{(1)}}{8}\, t_n(r)\, \kappa_j(x)\, dx = \int_{D_{j,A}} \frac{r}{8}\, t_n(r)\, \kappa_j(x)^2\, dx = -\int_{D_{j,A}} 2\delta_n \varphi\Big(\tfrac{r}{w_n} e_\circ\Big)\, \kappa_j(x)^2\, dx.$$

Since the average of $\kappa_j^2$ over each half-circle $\{\|x - x_j\|_2 = r,\ x^{(2)} > -1/2\}$ equals $1/2$, the two remaining terms cancel:

$$\int_{D_{j,A}} \big(\mu_\omega f_\omega - \mu_{\omega^{(j)}} f_{\omega^{(j)}}\big)\, dx = 0.$$

Together with the equality of radial slice masses this says that, for every measurable $A \subseteq [c_{\mathrm{rad}} a_n, \infty)$, the two experiments assign the same mass and the same conditional-mean mass to the slice $\{\|X - x_j\|_2 \in A\}$; since the outcome mixture is determined by its conditional mean, the restricted one-point distance laws agree:

$$\mathcal{L}_{P_\omega}\big(Y, \|X - x_j\|_2\big)\Big|_{\{z_2 \geq c_{\mathrm{rad}} a_n\}} = \mathcal{L}_{P_{\omega^{(j)}}}\big(Y, \|X - x_j\|_2\big)\Big|_{\{z_2 \geq c_{\mathrm{rad}} a_n\}}.$$

*Step 10: the compressed-distance divergence.* Below the fully active radius the same cancellation is replaced by a localized setwise bound. For every measurable $A \subseteq [0, \infty)$,

$$\left| \int_{\{\|X - x_j\|_2 \in A\}} \mu_\omega\, dP_\omega^X - \int_{\{\|X - x_j\|_2 \in A\}} \mu_{\omega^{(j)}}\, dP_{\omega^{(j)}}^X \right| \leq \delta_n\, P_\omega\big(\|X - x_j\|_2 \in A \cap [0, 128\delta_n)\big),$$

where $P^X$ denotes the covariate marginal: outside the radii $r < 2 \cdot (8\delta_n)/(1/8) = 128\delta_n$ the cutoff is fully active and Step 9 gives exact cancellation, while on those radii the uncancelled increment is bounded by the bump amplitude per unit radial mass. Moreover, since $f_\omega \leq 5/4$ and a disc of radius $R$ has area $\pi R^2 \leq 4R^2$,

$$P_\omega\big(\|X - x_j\|_2 < R\big) \leq 5R^2 \qquad (R \geq 0).$$

The complete radial laws of Step 9 provide a common marginal law for $R = \|X - x_j\|_2$ under the two adjacent vertices. Disintegrating the one-observation laws with respect to this common radial marginal gives measurable radial success parameters, clipped to $[1/4, 3/4]$ and equal almost everywhere to the corresponding conditional Bernoulli weights, such that the preceding setwise bound implies

$$|g_\omega(r) - g_{\omega^{(j)}}(r)| \le 2\delta_n \, \mathbf{1}\{r < 128\delta_n\} \quad \text{for the common radial marginal-almost every } r.$$

Here the factor $2\delta_n$ is the envelope used for the almost-everywhere Radon–Nikodym comparison, and the middle-half clipping is silent by Step 5. For two Bernoulli-plus-Gaussian mixtures with success parameters in $[1/4, 3/4]$, the divergence at a fixed radius is at most four times the squared difference of the success parameters. Averaging over the common radial marginal therefore gives

$$\mathrm{KL}\Big(\mathcal{L}_{P_\omega}(Y, \|X - x_j\|_2), \mathcal{L}_{P_{\omega^{(j)}}}(Y, \|X - x_j\|_2)\Big) \le 4(2\delta_n)^2 \, P_\omega(\|X - x_j\|_2 < 128\delta_n) \le 4(2\delta_n)^2 \cdot 5(128\delta_n)^2 = C_{\mathrm{KL}} \, \delta_n^4,$$

using $4 \cdot 4 \cdot 5 \cdot 128^2 = 1310720 = C_{\mathrm{KL}}$.

The distance data $V_{n,P}(x_j)$ of Definition 10 are $n$ independent copies of the one-observation pair $(Y, \|X - x_j\|_2)$, so the divergence tensorizes; combining with the eventual budget of Step 3,

$$\mathrm{KL}\Big(\mathcal{L}_{P_\omega}\{V_{n,P_\omega}(x_j)\}, \mathcal{L}_{P_{\omega^{(j)}}}\{V_{n,P_{\omega^{(j)}}}(x_j)\}\Big) \le n \, C_{\mathrm{KL}} \, \delta_n^4 \le \alpha \log M_n.$$

Steps 4–10 supply every displayed clause with the constants of Step 2, for all sufficiently large $n$. □

Lemma 2 is the concrete bridge between smooth boundary regression and the abstract product experiment. The Hölder restriction determines the relation between the signal height $a_n$ and the cell radius $w_n$, while the boundary packing supplies $M_n$ separated query points. Local dependence on one bit makes each cell a coordinate of the hypercube, and the radial invariance clause expresses the angular information loss created by unsigned distance observations. The compressed-distance Kullback–Leibler bound then puts those coordinates in the range of Lemma 1. This is the finite-sample testing structure that underlies the rate calculation summarized in the main text.

For common-map rules, the decoder associated with a proposed estimator applies the same measurable map to each query point. If the estimator were uniformly accurate on all $x_j$'s, the signs of the boundary signals in Lemma 2 would be decoded correctly across all coordinates. Combining the signal separation with the overlap lower bound gives the positive lower scale in Theorem 1. The argument uses the law class and completed-expectation criterion exactly as defined in Definitions 9, 11 and 12.

The point-indexed class enlarges the set of admissible rules by allowing the Borel section to vary with the query point. The following inclusion statement records that relationship before applying the same hypercube construction to the outer-expectation risk.

**Lemma 3.** [`lem:common-map-strict-in-point-indexed`] (Strict Common-Map Inclusion) *For parameters satisfying*

- ***(Sample size.)*** $n \ge 1$.
- ***(Smoothness index.)*** $q \ge 1$.

- ***(Envelope.)*** $L \geq 4$.

*the common-map decision class and the point-indexed decision class of Definitions 11 and 13 satisfy*

$$\mathcal{T}_n^{\mathrm{NP}} \subsetneq \mathcal{T}_n^{\mathrm{PI}}.$$

*Proof of Lemma 3.* Throughout, a point of $\mathbb{R}^2$ is written $x = (x^{(1)}, x^{(2)})$, and a sample point is written $w = ((y_i, z_i) : 1 \leq i \leq n) \in (\mathbb{R} \times \mathbb{R}^2)^n$, so that the distance data of Definition 10 evaluated at $w$ and at a query point $x$ are

$$V_{n,P}(x)\big|_w = \big((y_i, \|z_i - x\|_2) : 1 \leq i \leq n\big).$$

*Inclusion.* Let $\widehat{\mu}_n \in \mathcal{T}_n^{\mathrm{NP}}$ and let $T_n : (\mathbb{R} \times [0,\infty))^n \to \mathbb{R}$ be a measurable map, selected independently of $P$, representing it as in Definition 11. Define a point-indexed family by

$$T_{n,x} = T_n \qquad \text{for every } x \in \mathbb{R}^2.$$

Each section is the single Borel map $T_n$, hence Borel measurable for every fixed $x$, and the family is selected independently of $P$ because $T_n$ is. Moreover, for every $P \in \mathcal{P}_{\mathrm{NP}}(L,q)$, every sample point, and every $x \in \mathrm{bd}(\mathcal{X}_P)$,

$$\widehat{\mu}_n(x) = T_n(V_{n,P}(x)) = T_{n,x}(V_{n,P}(x)),$$

so $\widehat{\mu}_n \in \mathcal{T}_n^{\mathrm{PI}}$ in the sense of Definition 13. Therefore

$$\mathcal{T}_n^{\mathrm{NP}} \subseteq \mathcal{T}_n^{\mathrm{PI}}.$$

*A point-indexed rule outside the common-map class.* Let the rule return the first coordinate of the query point, ignoring the data:

$$\widehat{\mu}_n^{\mathrm{PI}}(x) = x^{(1)}, \qquad \text{generated by the sections } \; T_{n,x}(v) = x^{(1)} \quad \text{for every } v \in (\mathbb{R} \times [0,\infty))^n.$$

For each fixed $x$ the section $T_{n,x}$ is constant in $v$, hence Borel, and the family $\{T_{n,x}\}$ does not depend on the member law. Thus $\widehat{\mu}_n^{\mathrm{PI}} \in \mathcal{T}_n^{\mathrm{PI}}$.

Suppose, for contradiction, that this same rule also belongs to $\mathcal{T}_n^{\mathrm{NP}}$, with common measurable map $T_n$. Apply Lemma 2 at the given $q \geq 1$ and $L \geq 4$, at any sample size beyond the threshold it provides, and take the vertex $\omega = (0, \ldots, 0)$; its model-class and support clause supplies a law

$$P \in \mathcal{P}_{\mathrm{NP}}(L,q), \qquad \mathcal{X}_P = S = [-1/2, 1/2]^2.$$

Set

$$x_- = (-1/4, -1/2), \qquad x_+ = (1/4, -1/2), \qquad z = (0, -1/2).$$

Each of $x_-$ and $x_+$ has second coordinate $-1/2$ and first coordinate of modulus $1/4 < 1/2$, so it is a non-corner point of the lower edge of $S$ and therefore

$$x_- \in \mathrm{bd}(\mathcal{X}_P), \qquad x_+ \in \mathrm{bd}(\mathcal{X}_P).$$

Take the deterministic sample point $w$ whose every observation equals $(0, z)$, that is $y_i = 0$ and $z_i = z$ for all $i \leq n$. Since $z$, $x_-$ and $x_+$ share their second coordinate, the Euclidean distances reduce to horizontal ones:

$$\|z - x_-\|_2 = |0 - (-1/4)| = \frac{1}{4} = |0 - 1/4| = \|z - x_+\|_2.$$

Consequently the compressed samples at the two query points coincide:

$$V_{n,P}(x_-)\big|_w = \big((0,1/4) : 1 \le i \le n\big) = V_{n,P}(x_+)\big|_w.$$

Both $x_-$ and $x_+$ lie on $\mathrm{bd}(\mathcal{X}_P)$, so the assumed common-map representation applies at each of them and, with the displayed equality of distance data, gives

$$x_-^{(1)} = T_n\big(V_{n,P}(x_-)\big|_w\big) = T_n\big(V_{n,P}(x_+)\big|_w\big) = x_+^{(1)}.$$

But $x_-^{(1)} = -1/4$ and $x_+^{(1)} = 1/4$, a contradiction.

Hence $\widehat{\mu}_n^{\mathrm{PI}} \in \mathcal{T}_n^{\mathrm{PI}} \setminus \mathcal{T}_n^{\mathrm{NP}}$, so the inclusion proved first is proper:

$$\mathcal{T}_n^{\mathrm{NP}} \subsetneq \mathcal{T}_n^{\mathrm{PI}}.$$

□

The strict inclusion in Lemma 3 places the two unsigned risks in the expected order of admissible decision classes. For the converse, the relevant fact is that each point-indexed section remains law independent and receives the same unsigned distance data $V_{n,P}(x)$ from Definition 10. Thus the coordinatewise testing reduction induced by Lemma 2 continues to apply after replacing a single common map by fixed-$x$ Borel sections. The outer expectation in Definition 14 supplies the measurable-majorant convention for the boundary supremum, and the resulting lower bound is the point-indexed logarithmic converse in Theorem 2.

# B Signed-Distance Geometry, Maximal Inequalities, and Upper Bound

This appendix collects the empirical-process ingredients used by the signed-distance upper bound in Proposition 1. The objects are those of the known-geometry design in Definitions 20 to 22 and 28: local neighborhoods are Euclidean balls generated by the uniform weight $K_\square$, the outcome moment envelope is handled through winsorization, and the stochastic term is measured by an outer-expected supremum over arms and interface points. The two statements below organize these roles before the estimator bound is assembled in Proposition 1.

The first ingredient is geometric. Because the local windows are closed Euclidean balls, the relevant index class has finite VC complexity, giving the covering control used by the maximal inequality [Vapnik and Chervonenkis, 1971, Pollard, 1984, van der Vaart and Wellner, 1996].

**Lemma 4.** [`lem:euclidean-balls-vc`] (Planar Balls VC Bound) *For every finite set $S$ of score points in $\mathbb{R}^2$, if $|S| \ge 4$, then $S$ is not shattered by the class of closed Euclidean balls*

$$\{B_2(x,r) : x \in \mathbb{R}^2,\ r \ge 0\}.$$

*Equivalently, no such $S$ has the property that every subset of $S$ can be written as $S \cap B_2(x,r)$ for some center $x \in \mathbb{R}^2$ and radius $r \ge 0$.*

*Proof of Lemma 4.* Suppose, for contradiction, that a finite set $S \subset \mathbb{R}^2$ with $|S| \ge 4$ is shattered by closed Euclidean balls, that is, for every subset $T \subseteq S$ there are a center $c \in \mathbb{R}^2$ and a radius $r \ge 0$ with

$$T = S \cap B_2(c,r) = \{z \in S : \|z - c\|_2 \le r\}.$$

Since $|S| \geq 4$ we may fix four pairwise distinct points $z_1, z_2, z_3, z_4 \in S$. Writing $z_i = (z_{i,1}, z_{i,2})$, put

$$\Phi(z_i) = \big(1, z_{i,1}, z_{i,2}, \|z_i\|_2^2\big) \in \mathbb{R}^4, \qquad i \in \{1,2,3,4\}.$$

*An algebraic identity.* Let $\ell = (\ell_1, \ell_2, \ell_3, \ell_4) \in \mathbb{R}^4$ be arbitrary coefficients and let $w \in \mathbb{R}^4$ be the corresponding combination,

$$w = \sum_{i=1}^{4} \ell_i \Phi(z_i).$$

For a center $c = (c_1, c_2) \in \mathbb{R}^2$ and a radius $r \geq 0$, expanding the squared distance gives, for each $i$,

$$\|z_i - c\|_2^2 - r^2 = \|z_i\|_2^2 - 2c_1 z_{i,1} - 2c_2 z_{i,2} + (\|c\|_2^2 - r^2) \cdot 1,$$

which is one and the same linear combination of the four entries of $\Phi(z_i)$, with coefficients not depending on $i$. Multiplying by $\ell_i$ and summing therefore replaces those entries by the entries of $w$:

$$\sum_{i=1}^{4} \ell_i \{\|z_i - c\|_2^2 - r^2\} = w_4 - 2c_1 w_2 - 2c_2 w_3 + (\|c\|_2^2 - r^2) w_1.$$

*A sign contradiction.* We next record the contradiction used in both cases below. Suppose the coefficients $\ell$ are not all zero, that their combination $w$ satisfies $w_1 = 0$, and that

$$\sum_{i=1}^{4} \ell_i \{\|z_i - c\|_2^2 - r^2\} \geq 0 \qquad \text{for every } c \in \mathbb{R}^2 \text{ and every } r \geq 0.$$

Reading off the first entry of $w$ and using $\Phi(z_i)_1 = 1$ for every $i$,

$$\sum_{i=1}^{4} \ell_i = w_1 = 0,$$

so, some $\ell_i$ being nonzero, at least one coefficient is strictly negative; fix an index $i_-$ with $\ell_{i_-} < 0$. Apply the shattering hypothesis to the subset

$$T = \{z_i : 1 \leq i \leq 4,\ \ell_i > 0\} \subseteq S,$$

obtaining a center $c$ and a radius $r \geq 0$ with $T = S \cap B_2(c, r)$. Because $z_1, \ldots, z_4$ are distinct, $z_i \in T$ holds exactly when $\ell_i > 0$. Hence every term of the weighted sum is nonpositive,

$$\ell_i \{\|z_i - c\|_2^2 - r^2\} \leq 0, \qquad i \in \{1,2,3,4\}:$$

if $\ell_i > 0$ then $z_i \in T \subseteq B_2(c, r)$, so $\|z_i - c\|_2 \leq r$, the bracket is nonpositive, and the product is nonpositive; if $\ell_i \leq 0$ then $z_i \notin T$, so $\|z_i - c\|_2 > r$, the bracket is strictly positive, and again the product is nonpositive. At $i_-$ the second alternative occurs with a strictly negative coefficient, so

$$\ell_{i_-} \{\|z_{i_-} - c\|_2^2 - r^2\} < 0,$$

and summing the four terms gives

$$\sum_{i=1}^{4} \ell_i \{\|z_i - c\|_2^2 - r^2\} < 0,$$

which contradicts the assumed nonnegativity at this $(c, r)$.

*Case 1: $\Phi(z_1), \ldots, \Phi(z_4)$ are linearly independent.* Four linearly independent vectors in $\mathbb{R}^4$ form a basis, so the fourth standard basis vector has an expansion

$$e_4 = (0, 0, 0, 1) = \sum_{i=1}^{4} \ell_i \Phi(z_i)$$

with coefficients $\ell$ that are not all zero, since $e_4 \neq 0$. Here $w = e_4$, so $w_1 = w_2 = w_3 = 0$ and $w_4 = 1$, and the algebraic identity gives, for every $c \in \mathbb{R}^2$ and every $r \geq 0$,

$$\sum_{i=1}^{4} \ell_i \{\|z_i - c\|_2^2 - r^2\} = 1 \geq 0.$$

The sign contradiction applies.

*Case 2: $\Phi(z_1), \ldots, \Phi(z_4)$ are linearly dependent.* Choose coefficients $\ell$, not all zero, with

$$\sum_{i=1}^{4} \ell_i \Phi(z_i) = 0.$$

Here $w = 0$, so in particular $w_1 = 0$, and the algebraic identity gives, for every $c \in \mathbb{R}^2$ and every $r \geq 0$,

$$\sum_{i=1}^{4} \ell_i \{\|z_i - c\|_2^2 - r^2\} = 0 \geq 0.$$

The sign contradiction applies again.

Both cases are impossible, so no finite $S$ with $|S| \geq 4$ is shattered by closed Euclidean balls. $\square$

The bound in Lemma 4 is used with the ball notation introduced in Definition 3. In the signed-distance design, the event selected by the uniform kernel is a side-restricted Euclidean neighborhood of the query point. Finite VC complexity therefore converts the continuum of possible interface centers into an empirical-process class with logarithmic entropy, matching the $\log(h^{-1})$ factors in the risk calculation.

The second auxiliary statement supplies the expected maximal bound for the centered winsorized local-polynomial score. It combines the VC geometry of Lemma 4 with the bounded envelope created by winsorizing the outcome at level $B$, using the outer-expectation convention already built into Definition 26.

**Lemma 5.** [`lem:cty-winsorized-score-maximal-bound`] (Winsorized Score Maximal Bound) *Fix a nonnegative integer $p$ and a real number $L$. There is a constant $C = C(p, L) > 0$ such that the following holds for every real $\nu$.*

- ***(Law class.)*** *The law $P$ belongs to the A1/A2 class $\mathcal{P}_{12}(p, \nu, L)$ of Definition 22.*
- ***(Sample size.)*** *The sample size satisfies $n \geq 1$.*
- ***(Bandwidth and clipping.)*** *The bandwidth and winsorization level satisfy $0 < h \leq L^{-1}$ and $B \geq 1$.*

*Let $P^n$ denote the $n$-fold i.i.d. law of the potential-outcome observations. For each $t \in \{0,1\}$ and $x \in \mathcal{B}_P$, let the winsorized centered local-polynomial score be the vector whose $j$th coordinate is*

$$\mathrm{score}^B_{t,x,h,j} - \mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h,j}\right],$$

*where the empirical score uses the uniform kernel, the polynomial basis $r_p$, bandwidth $h$, and the clipped outcome transform $\psi_B$, centered around the corresponding population score. Then*

$$\mathbb{E}^*_{P^n}\left[\sup_{t\in\{0,1\},\, x\in\mathcal{B}_P} \left\|\mathrm{score}^B_{t,x,h} - \mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h}\right]\right\|\right] \le C\left\{\sqrt{\frac{\log(B/h)}{nh^2}} + \frac{B\log(B/h)}{nh^2}\right\}.$$

*Proof of Lemma 5.* Throughout, $Z = (Y(0), Y(1), X)$ denotes one potential-outcome observation, $Y = TY(1) + (1-T)Y(0)$ its observed outcome, and for $u \in \mathbb{R}^{p+1}$ we write $\|u\| = \max_{0\le j\le p}|u_j|$ for the coordinate maximum, which is the norm appearing in the displayed conclusion. All constants below are fixed before the moment exponent $\nu$, the law $P$, the sample size $n$, the bandwidth $h$, and the winsorization level $B$.

**Step 1: a coefficient radius depending only on $p$ and $L$.** Set

$$R = 1 + |L(p+1)\cdot 16L(2+L)|,$$

which is positive. We claim that for every $P \in \mathcal{P}_{12}(p,\nu,L)$, every $t \in \{0,1\}$, every $x \in \mathcal{B}_P$, every $0 < h \le L^{-1}$ and every $0 \le j \le p$,

$$\left|\beta^*_{t,P,x}(h)_j\right| \le R.$$

Write

$$m_{t,P,x}(h) = \mathbb{E}_P\left[\frac{1}{h^2}\mathbf{1}\{D^{\pm}_{P,x} \in I_t\}K_{\square}(D^{\pm}_{P,x}/h)\, r_p(D^{\pm}_{P,x}/h)\, Y\right]$$

for the right-hand side of the population normal equations in Definition 24. Its coordinates obey

$$\max_{0\le j\le p}|m_{t,P,x}(h)_j| \le 16L(2+L).$$

Indeed, the uniform kernel forces $|D^{\pm}_{P,x}| \le h$, hence $X \in B_2(x,h)$ and $|r_p(D^{\pm}_{P,x}/h)_j| \le 1$, so the integrand is dominated pointwise by $h^{-2}\mathbf{1}\{X \in B_2(x,h)\}\{1 + 2Y(0)^2 + 2Y(1)^2\}$. The density envelope $f_P \le L$ of Definition 22 and the area $\pi h^2 \le 4h^2$ of a planar disk give

$$P\{X \in B_2(x,h)\} \le 4Lh^2.$$

The moment envelope in Definition 22 gives, since $\nu \ge 2$, the conditional second-moment estimate

$$\mathbb{E}_P\left[Y(t)^2 \mid X = x\right] \le 1 + L, \qquad x \in \mathcal{X}_P,$$

using $y^2 \le 1 + |y|^{2+\nu}$. Localizing this bound to $B_2(x,h)$ gives $\mathbb{E}_P[\mathbf{1}\{X \in B_2(x,h)\}Y(t)^2] \le (1+L)4Lh^2$ for each arm, so the displayed coordinate bound follows from

$$h^{-2}\left\{4Lh^2 + 4(1+L)4Lh^2\right\} = 20L + 16L^2 \le 16L(2+L).$$

The population Gram lower bound of Definition 22 and the normal equations of Definition 24, written with $\beta^* = \beta^*_{t,P,x}(h)$, then give

$$L^{-1}\|\beta^*\|^2 \le L^{-1}\sum_{k=0}^{p}(\beta^*_k)^2 \le (\beta^*)^\top \Psi_{t,P,x}(h)\beta^* = (\beta^*)^\top m_{t,P,x}(h) \le (p+1)\|\beta^*\|\cdot 16L(2+L),$$

the first step because $\|\beta^*\|^2$ is one of the summands. Hence $\|\beta^*\| \le L(p+1)\cdot 16L(2+L) \le R$, which is the claim.

**Step 2: the bounded score class.** For $b \in \mathbb{R}^{p+1}$ let $\Pi_R b$ be its componentwise clipping,

$$(\Pi_R b)_k = \max\{-R, \min\{R, b_k\}\}, \qquad 0 \le k \le p,$$

and let

$$\mathcal{I}_{P,h} = \left\{\iota = (q,t,x,b,j) : q \in [h,2h),\ t \in \{0,1\},\ x \in \mathcal{B}_P,\ b \in \mathbb{R}^{p+1},\ 0 \le j \le p\right\}.$$

For $\iota = (q,t,x,b,j) \in \mathcal{I}_{P,h}$ define the scalar function of one observation

$$g_\iota(Z) = \mathbf{1}\{D^{\pm}_{P,x} \in I_t\} K_\square\left(\frac{D^{\pm}_{P,x}}{q}\right) r_p\left(\frac{D^{\pm}_{P,x}}{h}\right)_j \left\{\psi_B(Y) - r_p\left(\frac{D^{\pm}_{P,x}}{h}\right)^\top \Pi_R b\right\}.$$

Only the kernel window uses the enlarged bandwidth $q$; the polynomial arguments keep $h$, so the target score is the member with $q = h$. The half-open enlargement $[h, 2h)$ supplies the support slack used in Step 6, and clipping the coefficient gives a constant-envelope class.

**Step 3: the score is a member of the class.** Fix $t \in \{0,1\}$, $x \in \mathcal{B}_P$ and $0 \le j \le p$, and put $\iota = (h, t, x, \beta^*_{t,P,x}(h), j) \in \mathcal{I}_{P,h}$. By Step 1 the clipping is inactive at this index, $\Pi_R \beta^*_{t,P,x}(h) = \beta^*_{t,P,x}(h)$, so comparing with the definition of the winsorized score in Definition 32,

$$\text{score}^B_{t,x,h,j} = \frac{1}{h^2}\cdot\frac{1}{n}\sum_{i=1}^n g_\iota(Z_i).$$

The envelope bound of Step 5 makes $g_\iota$ integrable, so averaging the previous display over the independent and identically distributed observations gives

$$\mathbb{E}_{P^n}\left[\text{score}^B_{t,x,h,j}\right] = \frac{1}{h^2}\int g_\iota \, dP.$$

**Step 4: reduction to one scalar empirical process.** For each fixed sample the coordinate maximum $\|\cdot\|$ is attained at some $j$, and every pair $(t,x)$ with $x \in \mathcal{B}_P$ together with that $j$ produces an index of $\mathcal{I}_{P,h}$ by Step 3. Hence, pointwise on the sample space,

$$\sup_{t\in\{0,1\},\, x\in\mathcal{B}_P} \left\|\text{score}^B_{t,x,h} - \mathbb{E}_{P^n}\left[\text{score}^B_{t,x,h}\right]\right\| \le \frac{1}{h^2}\sup_{\iota\in\mathcal{I}_{P,h}}\left|\frac{1}{n}\sum_{i=1}^n g_\iota(Z_i) - \int g_\iota\, dP\right|.$$

Outer expectation is monotone and pulls out the positive finite constant $h^{-2}$, so

$$\mathbb{E}^*_{P^n}\left[\sup_{t\in\{0,1\},\, x\in\mathcal{B}_P} \left\|\text{score}^B_{t,x,h} - \mathbb{E}_{P^n}\left[\text{score}^B_{t,x,h}\right]\right\|\right] \le \frac{1}{h^2}\,\mathbb{E}^*_{P^n}\left[\sup_{\iota\in\mathcal{I}_{P,h}}\left|\frac{1}{n}\sum_{i=1}^n g_\iota(Z_i) - \int g_\iota\, dP\right|\right].$$

**Step 5: envelope, radius and covering witnesses.** Put

$$\Lambda_{\text{env}} = (2\cdot 4^p + 2)\{1 + (p+1)R\}, \qquad \Lambda_{\text{rad}} = 1 + \left|3\cdot 4^p\left\{2(1+L) + \{(p+1)2^pR\}^2\right\}16L\right|,$$

$$C_0 = 1 + \max\{\Lambda_{\text{rad}}, \Lambda_{\text{env}}\} > 0.$$

There are constants $A \ge \exp(1)$ and $v \ge 1$, depending only on $p$ and $R$ and therefore only on $p$ and $L$, such that for every real $\nu$, every $P \in \mathcal{P}_{12}(p,\nu,L)$ and all $0 < h < B$ with $B \ge 1$, the class $\{g_\iota : \iota \in \mathcal{I}_{P,h}\}$ satisfies the following three properties.

- **(Envelope.)** For every $\iota \in \mathcal{I}_{P,h}$ and every observation $z$,

$$|g_\iota(z)| \le C_0 B.$$

On the kernel window $|D^{\pm}_{P,x}| \le q < 2h$, so $|r_p(D^{\pm}_{P,x}/h)_k| \le 2^p$ for every $k$. Together with $|\psi_B(Y)| \le B$ and $|(\Pi_R b)_k| \le R$, this gives

$$|g_\iota(z)| \le (2 \cdot 4^p + 2)\{B + (p+1)R\} \le \Lambda_{\mathrm{env}} B \le C_0 B,$$

where the middle inequality uses $B \ge 1$.

- **(Population radius.)** For every $\iota \in \mathcal{I}_{P,h}$,

$$\left( \int g_\iota^2 \, dP \right)^{1/2} \le C_0 h.$$

On the almost-sure support of $X$, the squared separable score is bounded pointwise by

$$3 \cdot 4^p \, \mathbf{1}\{X \in B_2(x, 2h)\} \left\{ Y(0)^2 + Y(1)^2 + \{(p+1)2^p R\}^2 \right\}.$$

The kernel window gives the radius $2h$ because $q < 2h$. The probability of this disk is at most $16Lh^2$, while the two localized arm-square integrals are each bounded by $(1 + L)16Lh^2$, using the conditional second-moment estimate from Step 1. Hence $\int g_\iota^2 \, dP$ is bounded by

$$3 \cdot 4^p \left\{ 2(1+L) + \{(p+1)2^p R\}^2 \right\} 16Lh^2.$$

Taking square roots and using $\sqrt{u} \le 1+u$ for $u \ge 0$ gives the displayed $\Lambda_{\mathrm{rad}} h \le C_0 h$ bound. The winsorization level is absorbed through $|\psi_B(Y)| \le |Y|$, so this radius is uniform in $B$.

- **(Polynomial covering.)** For every non-empty finite list $z_1, \ldots, z_m$ of observations and every $0 < \varepsilon \le 1$ there is a finite set $\mathcal{C} \subseteq \mathcal{I}_{P,h}$ with

$$\min_{\kappa \in \mathcal{C}} \left( \frac{1}{m} \sum_{i=1}^{m} \{g_\iota(z_i) - g_\kappa(z_i)\}^2 \right)^{1/2} < \varepsilon \, C_0 B \quad \text{for every } \iota \in \mathcal{I}_{P,h},$$

and with $|\mathcal{C}| \le (A/\varepsilon)^v$. The certificate is obtained as follows. First, the fixed-dimensional radial residual-polynomial class with bounded response $|\psi_B(Y)| \le B$, bounded arm weights, and coefficient box $[-R, R]^{p+1}$ has polynomial empirical $L^2$ covers with envelope $B+(p+1)R$; this is applied with radial ranges $[0, 2]$ for the positive and negative arm pieces and $[0, 0]$ for the zero-distance and outside-support pieces. Second, Lemma 4 bounds the shattering of the planar closed-ball class, so the classifier family $\mathbf{1}\{\|X - x\|_2 \le q\}$ indexed by the center-radius parameter has Vapnik–Chervonenkis dimension at most three; the standard passage from a finite Vapnik–Chervonenkis dimension to empirical $L^2$ covering numbers [Pollard, 1984, van der Vaart and Wellner, 1996] then supplies a fixed polynomial cover for that classifier family. Product closure combines each radial piece with the ball classifier for the positive and negative pieces, and finite-sum closure combines the four pieces

positive arm, negative arm, zero-distance arm, outside support.

On the negative arm the factors $(-1)^k$ are absorbed into the clipped coefficient vector, and the zero-distance and outside-support terms are degree-zero radial residual terms. A pullback then ties the ball center, radial center, arm, coefficient vector and coordinate to the same index $\iota$. These finitely many product, sum, envelope-enlargement and pullback operations preserve polynomial empirical $L^2$ entropy, changing only the constants; after the final envelope enlargement from $(2 \cdot 4^p + 2)\{B + (p+1)R\}$ to $C_0 B$, the witnesses are the displayed $A$ and $v$.

The three witnesses $A$, $v$, $C_0$ are chosen before $\nu$, $P$, $h$ and $B$, which is what makes the final constant uniform in the moment exponent.

The parameter-regime clause of $\mathcal{P}_{12}(p, \nu, L)$ gives $L \geq 4$, so $h \leq L^{-1} < 1 \leq B$ and $h < B$; the certificate therefore applies at the bandwidth and clipping level of the statement.

**Step 6: countable reduction.** Equip $\mathcal{I}_{P,h}$ with the product topology inherited from $[h, 2h) \subset \mathbb{R}$, the discrete arm and coordinate sets, the relative topology on $\mathcal{B}_P \subset \mathbb{R}^2$, and the Euclidean topology on $\mathbb{R}^{p+1}$. Choose a fixed countable dense sequence $(\kappa_m)_{m \geq 1}$ in this space. Let

$$S = \{z : X(z) \in \mathcal{X}_P,\ X(z) \notin \mathcal{B}_P\}.$$

The set $S$ has $P$-probability one: $X \in \mathcal{X}_P$ almost surely, and $P\{X \in \mathcal{B}_P\} = 0$ because the interface has finite one-dimensional Hausdorff measure by Definition 22 and hence vanishing planar Lebesgue measure under the displayed density.

Fix $\iota = (q, t, x, b, j) \in \mathcal{I}_{P,h}$. For $\ell \geq 1$ set

$$\delta_\ell = \frac{2h - q}{16\ell} > 0$$

and consider the target point $(q + 4\delta_\ell, t, x, b, j) \in \mathcal{I}_{P,h}$. By density of $(\kappa_m)$, choose $m_\ell$ so that $\kappa_{m_\ell} = (q_\ell, t_\ell, x_\ell, b_\ell, j_\ell)$ satisfies

$$\|x_\ell - x\|_2 < \delta_\ell, \qquad |q_\ell - (q + 4\delta_\ell)| < \delta_\ell, \qquad |b_\ell(k) - b(k)| < \delta_\ell \quad (0 \leq k \leq p),$$

with $t_\ell = t$ and $j_\ell = j$. Then $\kappa_{m_\ell} \to \iota$, and the one-sided construction gives

$$q + \|x_\ell - x\|_2 < q_\ell \qquad \text{for every } \ell.$$

Thus the approximating support bandwidth converges to the target bandwidth while staying above it by enough to cover the center perturbation. For every $z \in S$, this slack preserves the closed-kernel decision when $|D^{\pm}_{P,x}(z)| \leq q$, including the endpoint case $|D^{\pm}_{P,x}(z)| = q$; when $|D^{\pm}_{P,x}(z)| > q$, ordinary continuity of distance and of $q_\ell$ makes the approximating kernel decision eventually agree with the target one. Since $z \in \mathcal{X}_P \setminus \mathcal{B}_P$, the assignment partition in Definition 22 fixes the signed-arm indicator along the approximating boundary centers, and continuity handles the polynomial and clipped-coefficient factors. Therefore

$$g_{\kappa_{m_\ell}}(z) \longrightarrow g_\iota(z), \qquad z \in S.$$

The class is dominated by the integrable envelope

$$z \mapsto 2^p \left(|Y(0)| + |Y(1)| + (p+1)2^p R\right),$$

so dominated convergence also gives convergence of the corresponding population integrals. Consequently, for every $n$, almost surely under $P^n$,

$$\sup_{\iota\in\mathcal{I}_{P,h}}\left|\frac{1}{n}\sum_{i=1}^{n} g_\iota(Z_i)-\int g_\iota\,dP\right| = \sup_{m\ge 1}\left|\frac{1}{n}\sum_{i=1}^{n} g_{\kappa_m}(Z_i)-\int g_{\kappa_m}\,dP\right|.$$

**Step 7: the variance-adaptive maximal inequality.** We use the following countable-reduction form of the variance-adaptive VC maximal inequality. Let $\mu$ be a probability law and $\{G_\iota : \iota\in\mathcal{I}\}$ a real-valued class. Suppose that $0<\sigma<U$, $A\ge\exp(1)$, $v\ge 1$, every $G_\iota$ is measurable, $|G_\iota|\le U$, $(\int G_\iota^2\,d\mu)^{1/2}\le\sigma$, every countable subfamily has empirical $L^2$ covers of size at most $(A/\varepsilon)^v$ at radius $\varepsilon U$ for $0<\varepsilon\le 1$, and a countable subfamily realizes the centered empirical supremum almost surely under every finite product law. Then, with

$$\ell_\mu=\log\left(\max\left\{\exp(1),\frac{AU}{\sigma}\right\}\right),$$

one has, for every $n\ge 1$,

$$\mathbb{E}^*_{\mu^n}\left[\sup_{\iota\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^{n} G_\iota(W_i)-\int G_\iota\,d\mu\right|\right]\le c_{\mathrm{VC}}\left\{\sigma\sqrt{\frac{v\ell_\mu}{n}}+\frac{vU\ell_\mu}{n}\right\},\qquad c_{\mathrm{VC}}=16384.$$

Indeed, choose a reducing sequence $(G_m)_{m\ge 1}$. The envelope makes the countable supremum measurable, integrable, and bounded by $2U$, so the outer expectation of the continuum supremum is bounded by the ordinary expectation of the countable supremum. Countable-class symmetrization bounds that expectation by twice the Rademacher complexity. The Rademacher bound applies Dudley chaining to the empirical $L^2$ polynomial covers, uses the population $L^2$ radius $\sigma$, and controls the empirical radius by symmetrizing the squared bounded class; the resulting quadratic inequality yields the displayed leading $\sigma$-term and second-order $U$-term, with the numerical constant rounded to 16384.

Apply this inequality with $\mu=P$, $G_\iota=g_\iota$, $U=C_0B$, $\sigma=C_0h$, and with the entropy and countability witnesses from Steps 5 and 6. Writing

$$\ell=\log\left(\max\left\{\exp(1),\ \frac{A\,C_0B}{C_0h}\right\}\right),$$

we obtain, for every $n\ge 1$,

$$\mathbb{E}^*_{P^n}\left[\sup_{\iota\in\mathcal{I}_{P,h}}\left|\frac{1}{n}\sum_{i=1}^{n} g_\iota(Z_i)-\int g_\iota\,dP\right|\right]\le c_{\mathrm{VC}}\left\{C_0h\sqrt{\frac{v\,\ell}{n}}+\frac{v\,C_0B\,\ell}{n}\right\},\qquad c_{\mathrm{VC}}=16384.$$

**Step 8: calibrating the logarithm.** Since $L\ge 4$, $h\le L^{-1}$ and $B\ge 1$,

$$4h\le Lh\le 1\le B,\qquad\text{so}\qquad \frac{B}{h}\ge 4\qquad\text{and}\qquad \log(B/h)\ge 1.$$

Also $A\ge\exp(1)$ gives $\log A\ge 1$, so $A(B/h)\ge\exp(1)$ and the maximum in $\ell$ is attained at the second entry. Putting $k=1+\log A$,

$$\ell=\log A+\log(B/h)\le k\log(B/h),$$

because $\{\log A - 1\}\{\log(B/h) - 1\} \geq 0$.

**Step 9: conclusion.** Abbreviate

$$\Gamma_1 = \sqrt{\frac{\log(B/h)}{nh^2}}, \qquad \Gamma_2 = \frac{B\log(B/h)}{nh^2},$$

both nonnegative. Multiplying the two terms of Step 7 by $h^{-2}$ and using Step 8,

$$\frac{1}{h^2}\, C_0 h \sqrt{\frac{v\,\ell}{n}} \leq C_0\sqrt{vk}\,\frac{1}{h}\sqrt{\frac{\log(B/h)}{n}} = C_0\sqrt{vk}\,\Gamma_1, \qquad \frac{1}{h^2}\cdot\frac{v\,C_0 B\,\ell}{n} \leq C_0 vk\,\Gamma_2.$$

Therefore, chaining Steps 4, 6, 7 and the two displays above and setting

$$C = 1 + \left| c_{\mathrm{VC}} C_0 \left\{\sqrt{vk} + vk\right\}\right|,$$

we obtain

$$\mathbb{E}^*_{P^n}\left[\sup_{t\in\{0,1\},\, x\in\mathcal{B}_P} \left\|\mathrm{score}^B_{t,x,h} - \mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h}\right]\right\|\right] \leq c_{\mathrm{VC}} C_0\left\{\sqrt{vk} + vk\right\}(\Gamma_1 + \Gamma_2) \leq C\left\{\sqrt{\frac{\log(B/h)}{nh^2}} + \frac{B\log(B/h)}{nh^2}\right\}$$

The constant $C$ is positive and, since $R$, $C_0$, $A$ and $v$ were fixed from $p$ and $L$ alone, depends only on $p$ and $L$. □

The two terms in Lemma 5 have the usual variance and envelope forms for bounded empirical processes indexed by VC-type sets [Pollard, 1984, van der Vaart and Wellner, 1996]. The scaling $h^{-2}$ reflects the two-dimensional local window around an interface point, while the supremum over $t \in \{0, 1\}$ and $x \in \mathcal{B}_P$ matches the signed-distance risk criterion in Definition 26. Together with the Gram lower bound in Definition 22, the statement controls the stochastic perturbation of the intercepts in the stabilized estimator.

These ingredients enter the upper-bound calculation as follows. The population part is separated into the first-order local-polynomial approximation term and the winsorization term, the latter controlled by the $2 + \nu$ conditional moment envelope in Definition 22, which bounds the winsorization remainder by a multiple of $B^{-3}$. The sample part is decomposed into empirical Gram fluctuation and centered score fluctuation, with the score controlled by Lemma 5 and the Gram term handled by the expected maximal condition stated in Proposition 1. The stabilization rule in Definition 28 converts these uniform matrix and score bounds into a uniform intercept bound for each arm, and clipping keeps the treatment-effect estimate on the envelope scale.

At the bandwidth $h_n = a_n$ from Definition 8 and winsorization level $B_n = a_n^{-1/3}$, the first-order bias, winsorization bias, and stochastic terms in Proposition 1 are all bounded by a constant multiple of $a_n$ for sufficiently large $n$. Thus the auxiliary results in this appendix support the expected outer-risk upper bound for the explicit signed-distance local-polynomial estimator, completing the upper side of the matched-rate comparison stated in Theorem 4.

# C Signed-Distance Hypercube and Matched Rate

This appendix records the hard signed-distance subexperiment used for the lower side of the matched-rate statement. The construction fixes the known rectangular geometry from Definition 27 and places many separated local perturbations along the bottom edge of the treatment boundary. Each perturbation changes the treatment-effect target at one boundary point by a calibrated amount, while the corresponding one-observation signed-distance laws remain close enough for the overlap direct-product argument in Lemma 1 to apply.

**Lemma 6.** [lem:cty-a1-a2-rectangle-angular-hypercube-all-orders] (Rectangle Angular Hypercube) *For every integer $p \geq 0$, there is a constant $L_0 = L_0(p) \geq 48$ such that, for every $\nu \geq 2$ and every $L \geq L_0$, there are constants $c, A, C, C_0 > 0$, with $2C < A$ when $p = 0$, and a constant $\delta_0 > 0$ such that the following construction is available for every $0 < \Delta \leq \delta_0$. Writing $q = p + 1$, there exist an integer $M$, constants $w, \rho > 0$, points $x_1, \ldots, x_M$, laws $P_\omega$ indexed by $\omega \in \{0,1\}^M$, and probability laws $Q_{j,b}$ for $j = 1, \ldots, M$ and $b \in \{0,1\}$, satisfying:*

- ***(Size and calibration.)***

$$M \geq c\Delta^{-1/q}, \qquad w = A\Delta^{1/q}, \qquad \rho = \frac{\pi w^2}{36}.$$

- ***(Common geometry.)*** *The support is $\mathcal{X} = [-3,3]^2$, the treated region is $\mathcal{A}_1 = [-1,1] \times [0,2]$, the control region is $\mathcal{A}_0 = \mathcal{X} \setminus \mathcal{A}_1$, the boundary is $\mathcal{B} = \partial \mathcal{A}_1$, the metric is Euclidean distance, and the kernel is $K_\square$. The points $x_j$ lie on the middle of the bottom edge, and the disks $S_j = B_2(x_j, w)$ are contained in $\mathcal{X}$, pairwise separated by distance at least $2w$, and pairwise disjoint.*

- ***(Class membership.)*** *For every $\omega \in \{0,1\}^M$, the law $P_\omega$ belongs to $\mathcal{P}_{12}(p, \nu, L)$ with the common geometry above.*

- ***(Bit structure.)*** *For every $\omega$ and $j$,*

$$P_\omega\{z : \mathrm{score}(z) \in S_j\} = \rho.$$

*The restriction of $P_\omega$ to $S_j$ depends on $\omega$ only through $\omega_j$, and the restriction of $P_\omega$ outside $\bigcup_{j=1}^M S_j$ is common across all $\omega$.*

- ***(Signed-observation laws.)*** *Conditional on $\mathrm{score}(Z) \in S_j$, the one-observation signed-distance law at $x_j$ under $P_\omega$ is $Q_{j,\omega_j}$.*

- ***(Signal separation.)*** *If two vertices agree in coordinate $j$, then their treatment-effect targets at $x_j$ agree. Flipping coordinate $j$ changes the treatment-effect target at $x_j$ by exactly $\Delta$:*

$$\left|\tau_{P_\omega}(x_j) - \tau_{P_{\omega^{(j)}}}(x_j)\right| = \Delta.$$

- ***(Local comparison laws.)*** *For each $j$, the laws $Q_{j,0}$ and $Q_{j,1}$ have the same signed-distance marginal, agree outside the event $0 < D^{\pm}_{P,x_j} < 2C\Delta$, and obey the two Kullback–Leibler bounds*

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \leq C_0 \frac{\Delta^4}{w^2}, \qquad \mathrm{KL}(Q_{j,1}, Q_{j,0}) \leq C_0 \frac{\Delta^4}{w^2}.$$

*Proof of Lemma 6.* Throughout, a point of $\mathbb{R}^2$ is written $x = (x^{(1)}, x^{(2)})$, and we abbreviate $q = p+1$. The arm sign attached to the side intervals $I_1 = [0, \infty)$, $I_0 = (-\infty, 0)$ of Definition 22 is written $(2t-1)$, as in the arm-mass clause there.

*Two fixed smooth profiles.* Fix once and for all a radially symmetric function $\varphi : \mathbb{R}^2 \to [0,1]$ of class $C^\infty$ with

$$\varphi(0) = 1, \qquad \varphi(z) = 0 \;\text{ whenever }\; \|z\|_2 \geq 1,$$

let $\varphi_{\mathrm{r}} : [0, \infty) \to [0, 1]$ be its radial profile, so that $\varphi(z) = \varphi_{\mathrm{r}}(\|z\|_2)$, and put

$$b_k = \sup_{z \in \mathbb{R}^2} \|D^k \varphi(z)\| < \infty, \qquad k \geq 0.$$

Fix also a nondecreasing $C^\infty$ function $\chi : \mathbb{R} \to [0, 1]$ with $\chi(s) = 0$ for $s \leq 0$ and $\chi(s) = 1$ for $s \geq 1$.

*The constants.* Set

$$A = 257 + \big(1 + q + 1310720 + \sum_{k=0}^{q} b_k\big), \qquad c = \frac{1}{24A}, \qquad C = 128, \qquad C_0 = 262144.$$

All four are positive, $A > 256$ and $A \geq 1$ because the bracket is positive, and, since every $b_k$ is nonnegative and appears in the bracketed sum,

$$b_k \leq A \qquad \text{for every } 0 \leq k \leq q.$$

Because $2C = 256 < A$, the required order-zero separation $2C < A$ holds; it in fact holds for every $p$.

*The small-separation threshold.* The map $\Delta \mapsto A\Delta^{1/q}$ is continuous at 0 with value 0, so there is $\delta_0 > 0$ such that every $0 < \Delta \leq \delta_0$ satisfies

$$\Delta \leq \frac{1}{16}, \qquad \Delta A \leq 1, \qquad w := A\Delta^{1/q} \leq \frac{1}{24}.$$

*The envelope threshold.* The signed radial polynomial energy is coercive: there is $c_p > 0$ such that, for both $t \in \{0, 1\}$ and every $v = (v_0, \ldots, v_p)^\top \in \mathbb{R}^{p+1}$,

$$c_p \|v\|_2^2 \leq \int_0^1 \big\{v^\top r_p((2t-1)u)\big\}^2 u \, du,$$

with $r_p$ the basis of Definition 17. Put

$$\gamma_p = \frac{1}{48} \cdot \frac{\pi}{2} \cdot c_p > 0, \qquad L_0 = \max\{48, \gamma_p^{-1}\},$$

so that $L_0 \geq 48$. Fix $\nu \geq 2$ and $L \geq L_0$; in particular $L \geq 48$. Fix also $0 < \Delta \leq \delta_0$.

*Common geometry and grid.* Take the fixed known geometry of Definition 20,

$$\mathcal{X} = [-3, 3]^2, \quad \mathcal{A}_1 = [-1, 1] \times [0, 2], \quad \mathcal{A}_0 = \mathcal{X} \setminus \mathcal{A}_1, \quad \mathcal{B} = \partial \mathcal{A}_1, \quad d(x, z) = \|x - z\|_2,$$

with the uniform kernel $K_\square$ of Definition 19; the interface $\mathcal{B}$ is the one displayed in Definition 16. With $w = A\Delta^{1/q}$ put

$$M = \left\lfloor \frac{1}{12w} \right\rfloor, \qquad x_j = \Big(-\frac{1}{4} + \frac{j}{2(M+1)},\ 0\Big), \quad j = 1, \ldots, M, \qquad S_j = B_2(x_j, w),$$

the balls being those of Definition 3. Each $x_j$ has second coordinate 0 and first coordinate of modulus at most $1/4$, so it lies on the middle of the bottom edge $\{(s, 0) : |s| \leq 1/2\}$ of the treated rectangle, hence on $\mathcal{B}$. Since $w \leq 1/24$ gives $1/(12w) \geq 2$ and therefore $M \geq 2$ and $M \geq 1/(24w)$,

$$M \geq \frac{1}{24w} = \frac{1}{24A}\Delta^{-1/q} = c\,\Delta^{-1/q}.$$

The centers are equispaced with gap $1/\{2(M+1)\}$, so for $i \neq j$

$$\|x_i - x_j\|_2 = \frac{|i-j|}{2(M+1)} \geq \frac{1}{2(M+1)} \geq 4w \geq 2w,$$

using $2(M+1) \leq 3M \leq 1/(4w)$, which is where $M \geq 2$ and $M \leq 1/(12w)$ are used. Consequently the disks $S_j$ are pairwise disjoint, and each is contained in $\mathcal{X}$ because $|x_j^{(1)}| \leq 1/4$ and $w \leq 1/24$.

*The vertex laws.* For $\omega \in \{0,1\}^M$ let $P_\omega$ be the law with the following two ingredients. First, $X$ has Lebesgue density $f_{P_\omega}$ supported on $\mathcal{X}$, given there by

$$f_{P_\omega}(x) = \frac{1}{36}\left\{1 + \sum_{j=1}^{M} \omega_j\, t_\Delta\big(\|x - x_j\|_2\big)\, \eta_j(x)\right\},$$

where the radial tilt and the horizontal direction cosine at $x_j$ are

$$t_\Delta(r) = -\frac{2\Delta\, \varphi_{\mathrm{r}}(r/w)\, \chi\big(\frac{r}{128\Delta} - 1\big)}{\max\{r/16,\ 8\Delta\}}, \qquad \eta_j(x) = \frac{x^{(1)} - x_j^{(1)}}{\|x - x_j\|_2} \quad (x \neq x_j), \quad \eta_j(x_j) = 0.$$

Second, given $X = x$ the two potential outcomes are conditionally independent Bernoulli variables whose means are the arm regressions of Definition 22,

$$\mu_{0,P_\omega}(x) = \frac{1}{2}, \qquad \mu_{1,P_\omega}(x) = \frac{1}{2} + \frac{x^{(1)}}{16} + \sum_{j=1}^{M} \omega_j\, \Delta\, \varphi\left(\frac{x - x_j}{w}\right),$$

both truncated to $[0,1]$ off $\mathcal{X}$, with $T = \mathbf{1}\{X \in \mathcal{A}_1\}$ and $Y = TY(1) + (1-T)Y(0)$. Because the disks are $4w$-separated while $\varphi$ vanishes outside the unit ball, at most one summand is nonzero at any $x$; together with $|x^{(1)}|/16 \leq 3/16$ and $\Delta \leq 1/16$ this gives

$$\frac{1}{4} \leq \mu_{1,P_\omega}(x) \leq \frac{3}{4}, \qquad x \in \mathcal{X},$$

so the truncation is inactive on $\mathcal{X}$. For the same reason the angular sum has modulus at most $1/4$, whence

$$\frac{1}{48} \leq f_{P_\omega}(x) \leq \frac{5}{144}, \qquad x \in \mathcal{X}.$$

*Class membership.* We check the clauses of Definition 22 in turn, using $L \geq 48$ throughout.

The support is the square $\mathcal{X} = [-3,3]^2 \subseteq [-L,L]^2$, and $f_{P_\omega}$ is continuous on $\mathcal{X}$ with

$$L^{-1} \leq \frac{1}{48} \leq f_{P_\omega}(x) \leq \frac{5}{144} \leq L, \qquad x \in \mathcal{X}.$$

The control regression is constant, and the treatment regression is the restriction to $\mathcal{X}$ of the globally $C^\infty$ function

$$g_\omega(x) = \frac{1}{2} + \frac{x^{(1)}}{16} + \sum_{j=1}^{M} \omega_j\, \Delta\, \varphi\left(\frac{x - x_j}{w}\right),$$

which satisfies the displayed envelope

$$\max_{0 \leq |\alpha| \leq p} \sup_{x \in \mathcal{X}} |\partial^\alpha g_\omega(x)| + \max_{0 \leq |\alpha| \leq p} \sup_{x \neq z \in \mathcal{X}} \frac{|\partial^\alpha g_\omega(x) - \partial^\alpha g_\omega(z)|}{\|x - z\|_2} \leq 48 \leq L.$$

The bump terms are controlled because the bandwidth is calibrated to the amplitude: for every $0 \le k \le q$ the $k$th derivative of one bump is bounded by $\Delta w^{-k} b_k$, and

$$\Delta\, w^{-k} b_k = \Delta^{1-k/q} A^{-k} b_k \le 1,$$

since $\Delta \le 1$ and $k \le q$ give $\Delta^{1-k/q} \le 1$, while $b_k \le A \le A^k$ for $k \ge 1$; at $k = 0$ the bound is $\Delta b_0 \le \Delta A \le 1$. The affine part $1/2 + x^{(1)}/16$ contributes at most $1/2 + 3/16$ to the derivative maximum, because $|x^{(1)}|/16 \le 3/16$ on $\mathcal{X}$ and its only nonvanishing derivative is the constant $1/16$, and at most $1/16$ to the Lipschitz maximum; and at most one bump is active at a time.

The conditional variances are the Bernoulli variances of the two means, so

$$\sigma^2_{t,P_\omega}(x) = \mu_{t,P_\omega}(x)\{1 - \mu_{t,P_\omega}(x)\} \in \Big[\frac{3}{16}, \frac{1}{4}\Big] \subseteq [L^{-1}, L], \qquad x \in \mathcal{X},\ t \in \{0,1\},$$

and they are continuous on $\mathcal{X}$.

Each $Y(t)$ takes values in $\{0,1\}$, so the moment envelope holds with room to spare:

$$\mathbb{E}_{P_\omega}\big[|Y(t)|^{2+\nu} \mid X = x\big] \le 1 \le L, \qquad x \in \mathcal{X}.$$

The sets $\mathcal{A}_0, \mathcal{A}_1$ are a Borel partition of $\mathcal{X}$ with $\mathrm{bd}(\mathcal{A}_0) \cap \mathrm{bd}(\mathcal{A}_1) = \partial\mathcal{A}_1 = \mathcal{B} \subset \mathrm{int}(\mathcal{X})$, and $\mathcal{B}$ is a compact rectifiable curve, being the boundary of a $2 \times 2$ square. Its size obeys

$$2 \le \mathcal{H}^1(\mathcal{B}) \le 16, \qquad \text{hence} \qquad L^{-1} \le \mathcal{H}^1(\mathcal{B}) \le L,$$

the lower bound because $\mathcal{B}$ contains the segment from $(-1, 0)$ to $(1, 0)$ and the upper bound from a Lipschitz parametrization of the perimeter.

The metric is Euclidean and the kernel is $K_\square$; the traces of closed Euclidean balls on $\mathcal{X}$ form a VC class of index at most four by Lemma 4.

For the population Gram matrix, at every $x \in \mathcal{B}$, both $t \in \{0,1\}$, every $0 < h \le L^{-1}$ and every $v \in \mathbb{R}^{p+1}$,

$$\frac{1}{48} \cdot \frac{\pi}{2} \int_0^1 \big\{v^\top r_p((2t-1)u)\big\}^2 u\, du \ \le\ v^\top \Psi_{t,P_\omega,x}(h)\, v,$$

because the density is at least $1/48$ and, in polar coordinates around $x$, the arm-$t$ part of the window $B_2(x, h)$ contains a quarter-disk on which the uniform kernel equals one, namely a sector of opening $\pi/2$ with one sign choice per coordinate determined by where $x$ sits on $\mathcal{B}$. A quarter, not a half, is what is available at every interface point: at the four corners of $\mathcal{A}_1$ the arm-$t$ part of a small window is exactly a quarter-disk, and $h \le L^{-1} \le 1$ keeps that sector inside $\mathcal{X}$. Its angular measure $\pi/2$ is the second factor in the display. Combining this with the coercivity constant, $\gamma_p \|v\|_2^2 \le v^\top \Psi_{t,P_\omega,x}(h) v$, and $L \ge \gamma_p^{-1}$ gives the class threshold

$$\lambda_{\min}\big(\Psi_{t,P_\omega,x}(h)\big) \ge L^{-1}.$$

For the small-bandwidth arm mass, at every $x \in \mathcal{B}$, both $t$, and every $0 < h \le L^{-1}$,

$$\frac{1}{h^2} \int_{\mathcal{A}_t} K_\square\big(\{(2t-1)\|z - x\|_2\}/h\big)\, dz = \frac{1}{h^2} \int_{\mathcal{A}_t \cap B_2(x,h)} dz \ \ge\ \frac{1}{8}\ \ge\ L^{-1},$$

since every point of $\partial\mathcal{A}_1$, corners included, retains at least an eighth of the disk area in each arm.

Finally, for every $x \in \mathcal{B}$, both $t$, and every $0 < s \le L^{-1}$, the one-dimensional Hausdorff integral of $f_{P_\omega}$ over $\{z \in \mathcal{A}_t : \|z - x\|_2 = s\}$ is finite and strictly positive, because that slice is a nondegenerate circular arc and the density lies in $[1/48, 5/144]$. Hence $P_\omega \in \mathcal{P}_{12}(p, \nu, L)$, with the common geometry displayed above.

*Bit structure.* Put $\rho = \pi w^2/36 > 0$. Point reflection through $x_j$ maps the disk $S_j$ onto itself, preserves $\|x - x_j\|_2$ and negates $\eta_j$, so the angular correction integrates to zero over $S_j$; the remaining background density is the constant $1/36$. With the score projection of Definition 18,

$$P_\omega\{z : \mathrm{score}(z) \in S_j\} = \frac{1}{36} \int_{S_j} dz = \frac{\pi w^2}{36} = \rho.$$

Each bump and each angular correction indexed by $k$ vanishes off $S_k$, and the disks are disjoint; therefore the restriction of $P_\omega$ to $\{z : \mathrm{score}(z) \in S_j\}$ depends on $\omega$ only through $\omega_j$, and the restriction of $P_\omega$ to the complement of $\bigcup_j S_j$ is the same for every $\omega$.

*Conditional signed-observation laws.* Every $z \in S_j$ has $|z^{(1)}| \le 1/4 + w \le 1$ and $|z^{(2)}| \le w \le 2$, so $z \in \mathcal{A}_1$ exactly when $z^{(2)} \ge 0$. Since $x_j^{(2)} = 0$, the signed distance of Definition 21 therefore reduces on $S_j$ to the vertically signed radius,

$$D^{\pm}_{P_\omega, x_j} = \begin{cases} \|X - x_j\|_2, & X^{(2)} \ge 0, \\ -\|X - x_j\|_2, & X^{(2)} < 0. \end{cases}$$

Using the conditional-law notation of Definition 1, define

$$Q_{j,b} = \mathcal{L}_{P_\omega}\big((Y, D^{\pm}_{P_\omega, x_j}) \mid X \in S_j\big) \qquad \text{for any } \omega \text{ with } \omega_j = b,$$

which is unambiguous by the one-bit locality just proved and is a probability law. Equivalently, the image of $P_\omega$ restricted to $\{X \in S_j\}$ under $(Y, D^{\pm}_{P_\omega, x_j})$ equals $\rho\, Q_{j,\omega_j}$.

*Signal separation.* On $\mathcal{X}$ the treatment-effect curve of Definition 23 is

$$\tau_{P_\omega}(x) = \mu_{1,P_\omega}(x) - \mu_{0,P_\omega}(x) = \frac{x^{(1)}}{16} + \sum_{j=1}^{M} \omega_j \Delta\, \varphi\left(\frac{x - x_j}{w}\right),$$

and at $x = x_j$ only the $j$th bump is active, with $\varphi(0) = 1$, so

$$\tau_{P_\omega}(x_j) = \frac{x_j^{(1)}}{16} + \omega_j \Delta.$$

Hence two vertices agreeing in coordinate $j$ have the same target at $x_j$, and flipping coordinate $j$ moves that target by exactly the amplitude,

$$\big|\tau_{P_\omega}(x_j) - \tau_{P_{\omega(j)}}(x_j)\big| = \Delta.$$

*Common signed-distance marginal.* The same reflection argument applies slice by slice: for every Borel $A \subseteq \mathbb{R}$, the set $\{z \in S_j :\ D^{\pm}_{P_\omega, x_j}(z) \in A\}$ is invariant under reflection through $x_j$, so the angular correction contributes nothing to its mass and

$$Q_{j,0}(\mathbb{R} \times A) = \frac{1}{36\rho} \int_{\{z \in S_j :\ D^{\pm}_{P_\omega, x_j}(z) \in A\}} dz = Q_{j,1}(\mathbb{R} \times A),$$

which is the asserted equality of signed-distance marginals.

*Localized comparison of the two bit laws.* Under $Q_{j,b}$ the outcome is Bernoulli given the signed distance, with success parameter in $[1/4, 3/4]$ by the middle-half bound above. The two success parameters differ only at short positive radii: for every Borel $B \subseteq \mathbb{R}$,

$$\left| \int_{\{u \in B\}} y \, dQ_{j,1}(y,u) - \int_{\{u \in B\}} y \, dQ_{j,0}(y,u) \right| \leq \Delta \; Q_{j,0}\big(\mathbb{R} \times (B \cap (0, 2C\Delta))\big).$$

Indeed, enabling the $j$th bit adds $\Delta\varphi_{\mathrm{r}}(r/w)$ to the conditional mean on the upper half of the radial slice at radius $r$ and simultaneously tilts the design density by $t_\Delta(r)\eta_j$. For $r \geq 2C\Delta = 256\Delta$ the cutoff is fully active, $\chi(r/(128\Delta) - 1) = 1$ and $\max\{r/16, 8\Delta\} = r/16$, so $t_\Delta(r) = -32\Delta\varphi_{\mathrm{r}}(r/w)/r$; since the affine part of the treatment regression contributes the factor $(x^{(1)} - x_j^{(1)})/16 = r\eta_j(x)/16$, the tilt contributes $-2\Delta\varphi_{\mathrm{r}}(r/w)\eta_j(x)^2$ per unit slice mass, and the mean of $\eta_j^2$ over the upper half circle of radius $r$ is $1/2$, which cancels the added bump exactly. For $0 < r < 256\Delta$ the two conditional means differ by at most $\Delta$, and on the control side, where the signed distance is negative, both laws carry the common mean $1/2$. Consequently the restrictions of the two laws to the complement of the short positive window agree: writing $E = \{(y,u) : 0 < u < 2C\Delta\}$,

$$Q_{j,0}(\,\cdot\, \cap E^c) = Q_{j,1}(\,\cdot\, \cap E^c),$$

which is the asserted agreement outside $0 < D^{\pm}_{P,x_j} < 2C\Delta$.

*The two divergence bounds.* Two Bernoulli outcome laws over a common statistic, with success parameters in $[1/4, 3/4]$ and setwise mean-mass discrepancy at most $\Delta$ localized to $(0, 2C\Delta)$, satisfy

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \leq 4\Delta^2 \, Q_{j,0}\big(\mathbb{R} \times (0, 2C\Delta)\big),$$

and the same bound with the two laws interchanged, the marginals being common. Because $\{z \in S_j : \; 0 < D^{\pm}_{P_\omega,x_j}(z) < R\} \subseteq B_2(x_j, R)$, the localized mass obeys

$$Q_{j,0}\big(\mathbb{R} \times (0,R)\big) = \frac{1}{36\rho} \int_{\{z \in S_j : \; 0 < D^{\pm}_{P_\omega,x_j}(z) < R\}} dz \leq \frac{\pi R^2}{36\rho}.$$

Taking $R = 2C\Delta = 256\Delta$ and substituting $\rho = \pi w^2/36$,

$$4\Delta^2 \cdot \frac{\pi(256\Delta)^2}{36\rho} = \frac{4 \cdot 65536 \, \Delta^4}{w^2} = C_0 \frac{\Delta^4}{w^2},$$

so both directed divergences are at most $C_0\Delta^4/w^2$.

Collecting the displayed size and calibration identities, the common geometry, class membership, the bit structure, the conditional signed-observation laws, the exact signal separation, and the three local comparison facts gives every clause of the statement for the constants $c, A, C = 128, C_0 = 262144$ and the threshold $\delta_0$. □

The constants in Lemma 6 separate geometry from signal size. The radius $w = A\Delta^{1/q}$ gives room for a $C^{p+1}$ perturbation of height $\Delta$, and the packing size $M \gtrsim \Delta^{-1/q}$ supplies enough boundary coordinates to generate the logarithmic uniform-risk scale. The common rectangular geometry places every $P_\omega$ inside the same fixed-known-geometry subexperiment, while the bit structure makes each disk a coordinate of the hypercube. The local comparison clause supplies the testing input: after $n$ observations, the effective information in a coordinate is calibrated by the cell probability $\rho$ and the Kullback–Leibler bounds.

To obtain Theorem 3, set the hypercube signal to the risk scale $\Delta \asymp (\log n/n)^{1/4}$. Since $q = p + 1$, the calibration in Lemma 6 gives $M \gtrsim \Delta^{-1/q}$, $w = A\Delta^{1/q}$, and $\rho = \pi w^2/36$. The comparison laws then satisfy

$$n\rho \ \mathrm{KL}(Q_{j,0}, Q_{j,1}) \lesssim nw^2 \frac{\Delta^4}{w^2} = n\Delta^4 \asymp \log n,$$

with the same order for the reverse divergence. This calibration keeps the coordinatewise overlap controlled while the maximum over the packed boundary points contributes the logarithmic factor. Applying Lemma 1 to the signed-distance observations $U^{\pm}_{n,P_\omega}(x_j)$ converts the overlap into a lower bound for recovering the hypercube coordinates, and the signal-separation clause converts that testing loss into the interface sup-loss in Definition 26. Because all vertices lie in the fixed rectangular subexperiment, the conclusion applies both to $R_n^{12,\pm,\mathrm{fix}}(p, \nu, L)$ from Definition 27 and to the full risk $R_n^{12,\pm}(p, \nu, L)$.

The matched statement follows by combining the lower and upper components already stated in the main text. Theorem 3 gives the positive lower constant for the signed-distance risk after the envelope threshold $L_0(p)$ is met. Proposition 1 gives a uniform outer-expected upper bound of order $a_n$ for the explicit winsorized, Gram-stabilized local-polynomial estimator with bandwidth $h_n = a_n$ and winsorization level $B_n = a_n^{-1/3}$. Therefore Theorem 4 characterizes the displayed known-geometry signed-distance minimax risk on the normalization $a_n = (\log n/n)^{1/4}$, with the estimator in Definition 28 attaining the upper side of the rate over $\mathcal{P}_{12}(p, \nu, L)$.

# D Reproducibility Note

The displayed definitions and theorem implications have been machine checked, including the two unsigned converses, the signed fixed-geometry lower bound, and the assembly of the conditional signed upper and frontier results. In Proposition 1 and Theorem 4, distance identification, uniform first-order bias, and expected maximal control enter as hypotheses.

# E Proofs of the main results

*Proof of Proposition 1.* Throughout, $Z = (Y(0), Y(1), X)$ denotes one potential-outcome observation and $Y = TY(1) + (1 - T)Y(0)$ its observed outcome, as in Definition 22; $Z_1, \ldots, Z_n$ is the sample and $P^n$ its law. For $u \in \mathbb{R}^{p+1}$ we write

$$\|u\| = \max_{0 \le j \le p} |u_j|, \qquad \|u\|_2 = \left( \sum_{j=0}^{p} u_j^2 \right)^{1/2},$$

so that $\|u\|_2 \le (p+1)\|u\|$, the norm $\|\cdot\|$ being the one appearing in Lemma 5 and in Definition 30. Fix $p \ge 0$, $\nu \ge 2$ and $L \ge 4$, and abbreviate the bandwidth and clipping schedules of the statement by

$$h_n = a_n = \left( \frac{\log n}{n} \right)^{1/4}, \qquad B_n = a_n^{-1/3},$$

as in Definition 8 and Definition 28. We write $\widehat{\Psi}_{t,x}(h)$, $\widehat{m}^{B}_{t,x}(h)$, $\widehat{\beta}^{B}_{t,x}(h)$ and $\widehat{\tau}^{12,\mathrm{LP}}_{n,h}(x)$ for the objects of Definition 28, $\Psi_{t,P,x}(h)$ for the population Gram of Definition 22, $\beta^{*}_{t,P,x}(h)$ for the population coefficient of Definition 24, and $\tau_P$ for the target of Definition 23.

The first input of Definition 33 fixes what is being estimated: it is what makes $\tau_P(x)$ the boundary contrast approached by the two one-sided signed-distance conditional-mean limits, and the bias input of the same collection is stated relative to that same target through Definition 29. Everything below is quantitative and uses the bias and expected-maximal inputs, Lemma 5, and the envelopes of Definition 22.

**Step 0: the constants.** Let $C_b > 0$ be the constant supplied by the uniform first-order bias input of the statement and put

$$\bar{C}_b = C_b + 1 > 0.$$

Let $C_m > 0$ be the constant supplied by the expected maximal bounds of the statement, and let $C_{\mathrm{sc}} = C_{\mathrm{sc}}(p, L) > 0$ be the constant of Lemma 5. Define

$$C = \bar{C}_b + 32L^3(p+1) + 16L(p+1)C_{\mathrm{sc}} + 8L^2(p+1)C_m > 0.$$

All four summands are positive because $L \geq 4 > 0$. This is the constant of the conclusion; the threshold $N$ is fixed at the end of Step 2.

**Step 1: the four bandwidth calibrations.** Let $n \geq 3$ and write $h = h_n$, $B = B_n$. We claim

$$B^{-3} \leq h, \qquad \sqrt{\frac{\log(B/h)}{nh^2}} \leq 2h, \qquad \frac{B\log(B/h)}{nh^2} \leq 2h, \qquad \sqrt{\frac{\log(h^{-1})}{nh^2}} \leq h.$$

First, $\log n < n$ for every $n \geq 1$, so $\log n/n < 1$ and hence $0 < h < 1$ and $B = h^{-1/3} \geq 1$. Raising the definition of $a_n$ to the fourth power gives the identity that drives every calibration,

$$h^4 = \frac{\log n}{n}, \qquad \text{equivalently} \qquad \frac{\log n}{nh^2} = h^2.$$

The first claim is the exponent computation $B^{-3} = (h^{-1/3})^{-3} = h$. For the remaining three, take logarithms in the definition of $h$:

$$\log(h^{-1}) = \frac{\log n - \log\log n}{4}, \qquad \log(B/h) = \log\bigl(h^{-4/3}\bigr) = \frac{4}{3}\log(h^{-1}),$$

and, since $n \geq 3$ forces $\log\log n > 0$, both are nonnegative and satisfy

$$\log(h^{-1}) \leq \log n, \qquad \log(B/h) \leq \frac{\log n}{3} \leq 2\log n.$$

Dividing by $nh^2$ and using the displayed identity,

$$\frac{\log(B/h)}{nh^2} \leq \frac{2\log n}{nh^2} = 2h^2, \qquad \frac{\log(h^{-1})}{nh^2} \leq \frac{\log n}{nh^2} = h^2.$$

Taking square roots gives the second claim, because $2h^2 \leq (2h)^2$, and the fourth claim directly. For the third, factor out $B$ and use the previous display together with

$$Bh = h^{-1/3}h = h^{2/3} \leq 1,$$

which holds because $0 < h < 1$; thus $B\log(B/h)/(nh^2) \le B \cdot 2h^2 = 2h(Bh) \le 2h$.

**Step 2: activating the two assumed inputs along the frontier bandwidth.** The bias input of the statement is asserted along *antitone* bandwidth sequences, and $n \mapsto a_n$ is antitone only from $n = 3$ onward. Set therefore

$$h^\sharp_n = a_{\max\{n,3\}}, \qquad n \in \mathbb{N}.$$

Then $h^\sharp_n > 0$ for every $n$, the sequence $h^\sharp$ is antitone, $h^\sharp_n \to 0$, and

$$n(h^\sharp_n)^2 = \sqrt{n \log n} \to \infty \qquad (n \ge 3),$$

so the bias input applies to $h^\sharp$ with $\nu' = \nu$ and yields

$$\limsup_{n\to\infty} \mathrm{BiasRatio}_{p,\nu,L}(h^\sharp_n) \le C_b < \bar{C}_b.$$

Since $h^\sharp_n = a_n$ for every $n \ge 3$, there is $N_1$ such that

$$\mathrm{BiasRatio}_{p,\nu,L}(a_n) \le \bar{C}_b, \qquad n \ge N_1.$$

The expected maximal bounds of the statement are asserted along positive deterministic sequences tending to zero whose heavy-tail regime diverges. At $h_n = a_n$, using $\log(a_n^{-1}) = (\log n - \log\log n)/4$ and $a_n^2 = \sqrt{\log n/n}$,

$$\frac{n^{(1+\nu)/(2+\nu)} a_n^2}{\log(a_n^{-1})} = \frac{4\, n^{(1+\nu)/(2+\nu)-1/2} (\log n)^{1/2}}{\log n - \log\log n} \longrightarrow \infty,$$

because $(1+\nu)/(2+\nu) - 1/2 = \nu/\{2(2+\nu)\} > 0$ for $\nu \ge 2$, so the numerator's polynomial factor dominates the logarithm. Hence there is $N_2$ such that, for every $n \ge N_2$ and every $P \in \mathcal{P}_{12}(p,\nu,L)$,

$$\mathbb{E}^*_P[\mathrm{GramDev}_{n,p,P}(a_n)] \le C_m \sqrt{\frac{\log(a_n^{-1})}{n a_n^2}},$$

with GramDev as in Definition 30. Only this first half of the expected maximal bounds is used; the stochastic control of the score is supplied instead by Lemma 5, whose bounded envelope is created by the winsorization built into Definition 28. Finally $a_n \to 0$, so there is $N_3$ with $a_n \le L^{-1}$ for $n \ge N_3$, which is the small-bandwidth regime in which the population Gram floor, the arm-mass threshold and Lemma 5 are all available. Put

$$N = \max\{3, N_1, N_2, N_3\}$$

and fix $n \ge N$ and $P \in \mathcal{P}_{12}(p,\nu,L)$ for the rest of the proof. Write $h = a_n$ and $B = B_n = h^{-1/3}$; by Step 1, $1 \le B$ and $0 < h \le L^{-1}$.

**Step 3: a small Gram deviation activates the stabilization guard.** Put

$$\delta = \frac{1}{2L(p+1)} > 0.$$

We claim that on the event $\{\mathrm{GramDev}_{n,p,P}(h) \le \delta\}$ the guard of Definition 28 holds simultaneously at every arm $t \in \{0,1\}$ and every interface point $x \in \mathcal{B}_P$; that is,

$$(2L)^{-1} \sum_{j=0}^{p} v_j^2 \le v^\top \widehat{\Psi}_{t,x}(h) v, \qquad v \in \mathbb{R}^{p+1}.$$

Indeed, $\mathrm{GramDev}_{n,p,P}(h) \le \delta$ means exactly that every entry satisfies $|\widehat{\Psi}_{t,x}(h)_{jk} - \Psi_{t,P,x}(h)_{jk}| \le \delta$, and for any symmetric perturbation with entries bounded by $\delta$,

$$\left|v^\top\{\widehat{\Psi}_{t,x}(h) - \Psi_{t,P,x}(h)\}v\right| \le \sum_{j,k} |v_j|\,\delta\,|v_k| \le \delta \sum_{j,k} \frac{v_j^2 + v_k^2}{2} = (p+1)\,\delta \sum_{j=0}^{p} v_j^2 = (2L)^{-1} \sum_{j=0}^{p} v_j^2,$$

the middle step by $2|v_j||v_k| \le v_j^2 + v_k^2$. The population Gram lower bound of Definition 22, valid because $0 < h \le L^{-1}$ and $x \in \mathcal{B}_P$, gives $v^\top \Psi_{t,P,x}(h) v \ge L^{-1} \sum_j v_j^2 = 2(2L)^{-1} \sum_j v_j^2$. Subtracting the previous display leaves $v^\top \widehat{\Psi}_{t,x}(h) v \ge (2L)^{-1} \sum_j v_j^2$, which is the guard. This is where the factor 2 in the guard threshold $(2L)^{-1}$ is spent: it is exactly the room left by the population floor $L^{-1}$ after the empirical perturbation is absorbed.

**Step 4: the guarded coefficient error is the inverted score.** Suppose the guard holds at $(t,x)$. Then $\widehat{\Psi}_{t,x}(h)$ is invertible: if $\widehat{\Psi}_{t,x}(h)v = 0$ then the guard forces $\sum_j v_j^2 \le 0$, so $v = 0$. Moreover, for every $\zeta \in \mathbb{R}^{p+1}$,

$$\left\|\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\|_2 \le 2L\|\zeta\|_2.$$

Indeed, applying the guard to the vector $\widehat{\Psi}_{t,x}(h)^{-1}\zeta$ and then the Cauchy–Schwarz inequality,

$$(2L)^{-1}\left\|\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\|_2^2 \le \left\{\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\}^\top \widehat{\Psi}_{t,x}(h)\left\{\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\} = \left\{\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\}^\top \zeta \le \left\|\widehat{\Psi}_{t,x}(h)^{-1}\zeta\right\|_2 \|\zeta\|_2,$$

and dividing by $\|\widehat{\Psi}_{t,x}(h)^{-1}\zeta\|_2$ when it is nonzero gives the claim; when it is zero the claim is trivial. On the guarded branch $\widehat{\beta}^B_{t,x}(h) = \widehat{\Psi}_{t,x}(h)^{-1}\widehat{m}^B_{t,x}(h)$, so subtracting $\beta^*_{t,P,x}(h)$ and using the definition of the winsorized score in Definition 32,

$$\widehat{\beta}^B_{t,x}(h) - \beta^*_{t,P,x}(h) = \widehat{\Psi}_{t,x}(h)^{-1}\left\{\widehat{m}^B_{t,x}(h) - \widehat{\Psi}_{t,x}(h)\beta^*_{t,P,x}(h)\right\} = \widehat{\Psi}_{t,x}(h)^{-1}\,\mathrm{score}^B_{t,x,h}.$$

Selecting the intercept, bounding a coordinate by the Euclidean norm, and converting norms through $\|u\|_2 \le (p+1)\|u\|$,

$$\left|e_1^\top\left\{\widehat{\beta}^B_{t,x}(h) - \beta^*_{t,P,x}(h)\right\}\right| \le \left\|\widehat{\Psi}_{t,x}(h)^{-1}\,\mathrm{score}^B_{t,x,h}\right\|_2 \le 2L\left\|\mathrm{score}^B_{t,x,h}\right\|_2 \le 2L(p+1)\left\|\mathrm{score}^B_{t,x,h}\right\|.$$

**Step 5: the mean of the winsorized score is of order $B^{-3}$.** Write

$$\mathrm{WScoreDev}_{n,p,P}(h,B) = \sup_{t\in\{0,1\}} \sup_{x\in\mathcal{B}_P} \left\|\mathrm{score}^B_{t,x,h} - \mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h}\right]\right\|$$

for the centered supremum appearing in Lemma 5, so that for every $t$ and every $x \in \mathcal{B}_P$,

$$\left\|\mathrm{score}^B_{t,x,h}\right\| \le \mathrm{WScoreDev}_{n,p,P}(h,B) + \left\|\mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h}\right]\right\|.$$

We claim the deterministic bound

$$\left\|\mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h}\right]\right\| \le 8L^2B^{-3}, \qquad t\in\{0,1\},\ x\in\mathcal{B}_P.$$

Fix $t$, $x \in \mathcal{B}_P$, and a coordinate $0 \le j \le p$, and write the single-observation weight

$$\xi_j(Z) = \mathbf{1}\{D^\pm_{P,x} \in I_t\} K_\square(D^\pm_{P,x}/h)\, r_p(D^\pm_{P,x}/h)_j,$$

with $D^{\pm}_{P,x}$ the signed distance of Definition 21, $K_{\square}$ the uniform kernel of Definition 19, and $r_p$ the basis of Definition 17. Since $K_{\square}$ vanishes off $\{|D^{\pm}_{P,x}| \leq h\}$ and $|r_p(u)_j| \leq 1$ for $|u| \leq 1$,

$$|\xi_j(Z)| \leq \mathbf{1}\{X \in B_2(x,h)\},$$

with $B_2(x,h)$ the ball of Definition 3; the equivalence $|D^{\pm}_{P,x}| \leq h \iff X \in B_2(x,h)$ is immediate from the definition of the signed distance, whose magnitude is $\|X - x\|_2$. By Definition 32 the $j$th coordinate of the winsorized score is the sample average $h^{-2}n^{-1}\sum_{i\leq n}\xi_j(Z_i)\{\psi_B(Y_i) - r_p(D^{\pm}_{P,x,i}/h)^\top \beta^*_{t,P,x}(h)\}$, so the observations being independent and identically distributed and the summand integrable,

$$\mathbb{E}_{P^n}\left[\mathrm{score}^B_{t,x,h,j}\right] = \frac{1}{h^2}\int \xi_j(Z)\left\{\psi_B(Y) - r_p(D^{\pm}_{P,x}/h)^\top \beta^*_{t,P,x}(h)\right\} dP.$$

Split the integrand by inserting and removing the unwinsorized outcome,

$$\xi_j(Z)\left\{\psi_B(Y) - r_p^\top \beta^*\right\} = \underbrace{\xi_j(Z)\left\{Y - r_p^\top \beta^*\right\}}_{\text{population residual}} + \underbrace{\xi_j(Z)\left\{\psi_B(Y) - Y\right\}}_{\text{winsorization remainder}}.$$

The first term integrates to zero: by the normal equations of Definition 24,

$$\frac{1}{h^2}\int \xi_j(Z)\left\{Y - r_p(D^{\pm}_{P,x}/h)^\top \beta^*_{t,P,x}(h)\right\} dP = \mathbb{E}_P\left[\frac{1}{h^2}\mathbf{1}\{D^{\pm}_{P,x} \in I_t\}K_{\square}(D^{\pm}_{P,x}/h) r_p(D^{\pm}_{P,x}/h) Y\right]_j - \{\Psi_{t,P,x}(h)\beta^*_{t,P,x}($$

For the second term, the moment envelope of Definition 22 is converted into a deterministic $B^{-3}$ envelope pointwise in the outcome. For $B \geq 1$ and $\nu \geq 2$, every real $y$ satisfies

$$|y - \psi_B(y)| \leq |y|^{2+\nu}B^{-3}.$$

If $|y| \leq B$ the left side vanishes; otherwise $|y| > B \geq 1$, so $|y - \psi_B(y)| = |y| - B \leq |y| \leq |y|^4 B^{-3} \leq |y|^{2+\nu}B^{-3}$, using $|y|^3 \geq B^3$ and then $|y| \geq 1$ with $2 + \nu \geq 4$. Integrating this against the selected conditional law of $Y(t')$ given $X = z$ and applying the moment envelope $\mathbb{E}_P[|Y(t')|^{2+\nu} \mid X = z] \leq L$ of Definition 22 gives, for each arm $t' \in \{0,1\}$ and each $z \in \mathcal{X}_P$,

$$\mathbb{E}_P\big[\,|Y(t') - \psi_B(Y(t'))| \;\big|\; X = z\big] \leq LB^{-3}.$$

Note that it is this absolute conditional remainder, and not merely the difference of the two conditional means, that the argument requires, because the remainder is multiplied by the random weight $\xi_j$ before it is integrated. Combining the weight bound, the consistency relation $Y = TY(1) + (1-T)Y(0)$ of Definition 22, which makes $|\psi_B(Y) - Y|$ at most the sum of the two armwise remainders, and the density envelope $f_P \leq L$ together with the planar area bound $\pi h^2 \leq 4h^2$, which give

$$P\{X \in B_2(x,h)\} \leq 4Lh^2,$$

we obtain

$$\int |\xi_j(Z)\{\psi_B(Y) - Y\}|\, dP \leq \sum_{t' \in \{0,1\}} \mathbb{E}_P\big[\mathbf{1}\{X \in B_2(x,h)\}\,\big|Y(t') - \psi_B(Y(t'))\big|\big] \leq 2\cdot LB^{-3}\cdot 4Lh^2 = 8L^2B^{-3}h^2.$$

Multiplying by $h^{-2}$ and taking the maximum over $j$ proves the claim, since the bound is free of $j$, $t$ and $x$.

**Step 6: the pointwise interface loss.** We claim that for every sample,

$$\sup_{x\in\mathcal{B}_P}\left|\widehat{\tau}^{\,12,\mathrm{LP}}_{n,h}(x)-\tau_P(x)\right| \le \left\{\bar{C}_b h + 4L(p+1)\cdot 8L^2B^{-3}\right\} + 4L(p+1)\,\mathrm{WScoreDev}_{n,p,P}(h,B) + 8L^2(p+1)\,\mathrm{GramDev}_{n,p}$$

Two preliminary facts hold at every $x\in\mathcal{B}_P$. First, $\mathcal{B}_P\subset\mathrm{int}(\mathcal{X}_P)$, and the derivative envelope of Definition 22 bounds each arm regression by $|\mu_{t,P}(x)|\le L$, so

$$|\tau_P(x)|\le 2L, \qquad \text{that is} \qquad \tau_P(x)\in[-2L,2L].$$

Consequently, since projecting onto an interval containing $\tau_P(x)$ cannot increase the distance to $\tau_P(x)$, the clipping in Definition 28 is harmless:

$$\left|\widehat{\tau}^{\,12,\mathrm{LP}}_{n,h}(x)-\tau_P(x)\right| \le \left|e_1^\top\left\{\widehat{\beta}^B_{1,x}(h)-\widehat{\beta}^B_{0,x}(h)\right\}-\tau_P(x)\right|.$$

Second, $n\ge N_1$ and the definition of the bias ratio in Definition 29 give the deterministic first-order bias bound

$$\left|e_1^\top\left\{\beta^*_{1,P,x}(h)-\beta^*_{0,P,x}(h)\right\}-\tau_P(x)\right|\le \bar{C}_b h, \qquad x\in\mathcal{B}_P.$$

Now split on the design.

*Case 1:* $\mathrm{GramDev}_{n,p,P}(h)\le\delta$. By Step 3 the guard holds at both arms and every $x\in\mathcal{B}_P$, so Step 4 applies at both arms. Inserting the two population coefficients and using the triangle inequality,

$$\left|e_1^\top\left\{\widehat{\beta}^B_{1,x}(h)-\widehat{\beta}^B_{0,x}(h)\right\}-\tau_P(x)\right| \le \sum_{t\in\{0,1\}}\left|e_1^\top\left\{\widehat{\beta}^B_{t,x}(h)-\beta^*_{t,P,x}(h)\right\}\right| + \bar{C}_b h.$$

Bounding each summand by Step 4 and then each score by Step 5,

$$\sum_{t\in\{0,1\}}\left|e_1^\top\left\{\widehat{\beta}^B_{t,x}(h)-\beta^*_{t,P,x}(h)\right\}\right| \le 2\cdot 2L(p+1)\left\{\mathrm{WScoreDev}_{n,p,P}(h,B)+8L^2B^{-3}\right\},$$

which is

$$4L(p+1)\cdot 8L^2B^{-3} + 4L(p+1)\,\mathrm{WScoreDev}_{n,p,P}(h,B).$$

The right-hand side no longer depends on $x$, so taking the supremum over $x\in\mathcal{B}_P$ and adding the nonnegative Gram term gives the claim in this case.

*Case 2:* $\mathrm{GramDev}_{n,p,P}(h)>\delta$. Here the clipping alone controls the loss: $|\widehat{\tau}^{\,12,\mathrm{LP}}_{n,h}(x)|\le 2L$ by construction and $|\tau_P(x)|\le 2L$, so

$$\sup_{x\in\mathcal{B}_P}\left|\widehat{\tau}^{\,12,\mathrm{LP}}_{n,h}(x)-\tau_P(x)\right|\le 4L.$$

On the other hand the case hypothesis and the value of $\delta$ give

$$8L^2(p+1)\,\mathrm{GramDev}_{n,p,P}(h)\ge 8L^2(p+1)\cdot\frac{1}{2L(p+1)}=4L,$$

so the Gram term alone already dominates the loss, and the claim holds because the two remaining terms are nonnegative. This is the calibration that fixes the coefficient $8L^2(p+1)$: it is the smallest multiple of $\delta^{-1}$ that pays for the trivial envelope $4L$ on the ill-conditioned event.

**Step 7: integration.** Take outer expectations under $P^n$ in Step 6. The outer expectation of Definition 14 is monotone, is subadditive on sums, returns a constant on a constant because $P^n$ is a probability measure, and satisfies $\mathbb{E}_P^*[cW] \le c\,\mathbb{E}_P^*[W]$ for every finite constant $c > 0$. Applying these four properties in turn to the three summands of Step 6,

$$\mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P}\left|\widehat{\tau}_{n,h}^{12,\mathrm{LP}}(x)-\tau_P(x)\right|\right] \le \bar{C}_b h+32L^3(p+1)B^{-3}+4L(p+1)\,\mathbb{E}_P^*[\mathrm{WScoreDev}_{n,p,P}(h,B)]+8L^2(p+1)\,\mathbb{E}_P^*[\text{Gram}$$

where $4L(p+1)\cdot 8L^2 = 32L^3(p+1)$. The measurability, integrability and extended-arithmetic bookkeeping needed to pass from the pointwise inequality to this display is routine and is suppressed.

**Step 8: substituting the two maximal bounds.** The three terms are now evaluated at $h = a_n$ and $B = B_n = a_n^{-1/3}$ using Step 1.

The deterministic term. By the first calibration $B_n^{-3} = a_n$, so

$$\bar{C}_b a_n + 32L^3(p+1)B_n^{-3} = \left\{\bar{C}_b + 32L^3(p+1)\right\} a_n.$$

The score term. Since $n \ge 1$, $0 < a_n \le L^{-1}$ and $B_n \ge 1$, Lemma 5 applies to $P \in \mathcal{P}_{12}(p,\nu,L)$ at this bandwidth and clipping level and supplies

$$\mathbb{E}_P^*[\mathrm{WScoreDev}_{n,p,P}(a_n,B_n)] \le C_{\mathrm{sc}}\left\{\sqrt{\frac{\log(B_n/a_n)}{na_n^2}}+\frac{B_n\log(B_n/a_n)}{na_n^2}\right\} \le C_{\mathrm{sc}}\{2a_n+2a_n\} = 4C_{\mathrm{sc}}a_n,$$

the last inequality by the second and third calibrations of Step 1. Hence

$$4L(p+1)\,\mathbb{E}_P^*[\mathrm{WScoreDev}_{n,p,P}(a_n,B_n)] \le 16L(p+1)C_{\mathrm{sc}}\,a_n.$$

This is the step at which the winsorization level is paid for twice over, and in opposite directions: raising $B_n$ shrinks the deterministic term $32L^3(p+1)B_n^{-3}$ but inflates the envelope term $B_n\log(B_n/a_n)/(na_n^2)$. The choice $B_n = a_n^{-1/3}$ is exactly the one that places both on the scale $a_n$.

The Gram term. By Step 2 and the fourth calibration of Step 1,

$$8L^2(p+1)\,\mathbb{E}_P^*[\mathrm{GramDev}_{n,p,P}(a_n)] \le 8L^2(p+1)C_m\sqrt{\frac{\log(a_n^{-1})}{na_n^2}} \le 8L^2(p+1)C_m\,a_n.$$

Adding the three displays and recalling the definition of $C$ in Step 0,

$$\mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P}\left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x)-\tau_P(x)\right|\right] \le \left\{\bar{C}_b+32L^3(p+1)+16L(p+1)C_{\mathrm{sc}}+8L^2(p+1)C_m\right\}a_n = Ca_n.$$

Every quantity in $C$ was fixed in Step 0, before $P$ and $n$, so the bound is uniform over the law class, and taking the supremum over $P \in \mathcal{P}_{12}(p,\nu,L)$ gives

$$\sup_{P\in\mathcal{P}_{12}(p,\nu,L)}\mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P}\left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x)-\tau_P(x)\right|\right] \le Ca_n, \qquad n \ge N,$$

which is the assertion, with $C > 0$ and $N$ as constructed. □

*Proof of Theorem 3. Step 1 (constants).* Fix $p \geq 0$ and put $q = p+1$. Let $L_0 = L_0(p)$ be the envelope threshold supplied by Lemma 6, so $L_0 \geq 48$. Fix $\nu \geq 2$ and $L \geq L_0$, and let $c_\square, A, C, C_0, \delta_0$ be the positive constants that lemma attaches to $(p, \nu, L)$. Set

$$z = 128\, q\, C_0, \qquad \gamma = \min\{1, z^{-1}\}, \qquad d = \frac{1}{2}\{1 - \exp(-1/2)\}, \qquad c = \frac{\gamma d}{8}.$$

Then $\gamma > 0$, $d > 0$ and $c > 0$. Since $0 < \gamma \leq 1$ we have $\gamma^4 \leq \gamma$, and $\gamma \leq z^{-1}$ therefore gives the budget inequality

$$128\, q\, C_0\, \gamma^4 \leq 128\, q\, C_0\, \gamma = z\gamma \leq 1.$$

*Step 2 (the logarithmic packing budget).* We first record, as an eventual statement in $n$, the calibration that will put the hypercube of Lemma 6 inside the range of Lemma 1. For $n \geq 2$ the rate $a_n = (\log n/n)^{1/4}$ of Definition 8 satisfies $n\, a_n^4 = \log n$, so the budget inequality above gives

$$C_0\, n\, (\gamma a_n)^4 = C_0 \gamma^4\, n\, a_n^4 = C_0 \gamma^4 \log n \leq \frac{\log n}{128\, q}.$$

Also $\log a_n = \frac{1}{4}(\log\log n - \log n)$, so every integer $M$ with $M \geq c_\square (\gamma a_n)^{-1/q}$ obeys

$$\log M \geq \log c_\square - \frac{1}{q}\log\gamma - \frac{1}{4q}\bigl(\log\log n - \log n\bigr).$$

For all sufficiently large $n$ we have $\log\log n \leq \frac{1}{2}\log n$, whence $-\frac{1}{4q}(\log\log n - \log n) \geq \frac{1}{8q}\log n$; and, since $\log n \to \infty$, also $\log c_\square - \frac{1}{q}\log\gamma \geq -\frac{1}{32q}\log n$. For such $n$ the last display yields

$$\log M \geq -\frac{1}{32q}\log n + \frac{1}{8q}\log n \geq \frac{1}{32q}\log n.$$

Combining the two estimates, for all sufficiently large $n$ and every integer $M \geq c_\square(\gamma a_n)^{-1/q}$,

$$C_0\, n\, (\gamma a_n)^4 \leq \frac{\log n}{128\, q} = \frac{1}{4} \cdot \frac{\log n}{32\, q} \leq \frac{1}{4}\log M.$$

The de-Poissonization remainder $\exp\{-n(1 - \log 2)\}$ decays exponentially in $n$, while $a_n$ decays only polynomially, so $\exp\{-n(1-\log 2)\} = o(a_n)$. Since $a_n \to 0$, we may therefore fix $N \geq 2$ such that, for every $n \geq N$,

$$\gamma a_n \leq \delta_0, \qquad 2L\exp\{-n(1-\log 2)\} \leq c\, a_n,$$

and the displayed logarithmic packing budget holds.

*Step 3 (the hard family at scale $\Delta$).* Fix $n \geq N$ and a known-geometry point-indexed rule $\widehat{\tau}_n \in \mathcal{T}_n^{12,\pm}$, and choose a family of Borel sections $\{T_{n,G,x}\}$ representing it as in Definition 25. Put $\Delta = \gamma a_n$, so $0 < \Delta \leq \delta_0$. Applying Lemma 6 at separation $\Delta$ produces an integer $M$, a radius $w = A\Delta^{1/q}$, the disk mass $\rho = \pi w^2/36 > 0$, centers $x_1, \ldots, x_M$ on the middle of the bottom edge with pairwise disjoint disks $S_j = B_2(x_j, w) \subseteq [-3,3]^2$, laws $P_\omega \in \mathcal{P}_{12}(p, \nu, L)$ indexed by $\omega \in \{0,1\}^M$ all carrying the same fixed rectangular geometry

$$\mathcal{X} = [-3,3]^2, \quad \mathcal{A}_1 = [-1,1] \times [0,2], \quad \mathcal{A}_0 = \mathcal{X} \setminus \mathcal{A}_1, \quad \mathcal{B} = \partial\mathcal{A}_1,$$

and conditional signed-distance laws $Q_{j,b}$ with $M \geq c_\square \Delta^{-1/q}$. In particular $M \geq 1$, because $c_\square \Delta^{-1/q} > 0$.

*Step 4 (per-coordinate divergence budget).* The Kullback–Leibler clause of Lemma 6, the identity $\rho = \pi w^2/36$, and $\pi \le 4$ give, for every $j$,

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \cdot 2n\rho \le \frac{C_0 \Delta^4}{w^2} \cdot 2n\, \frac{\pi w^2}{36} = C_0\, n\, \Delta^4 \cdot \frac{\pi}{18} \le C_0\, n\, \Delta^4 \le \frac{1}{4} \log M,$$

where the last inequality uses the logarithmic packing budget above and $M \ge c_{\square}(\gamma a_n)^{-1/q}$.

*Step 5 (target values at the centers).* For $j \le M$ and $b \in \{0,1\}$ let $\epsilon(j,b) \in \{0,1\}^M$ be the vertex with $j$th coordinate $b$ and all other coordinates 0, and put

$$v_{j,b} = \tau_{P_{\epsilon(j,b)}}(x_j).$$

Because the treatment-effect target at $x_j$ depends on the vertex only through its $j$th coordinate, and because flipping that coordinate moves the target at $x_j$ by exactly $\Delta$ (Lemma 6),

$$\tau_{P_\omega}(x_j) = v_{j,\omega_j} \qquad \text{for every } \omega, \qquad \text{and} \qquad |v_{j,1} - v_{j,0}| = \Delta.$$

*Step 6 (the Poissonized experiment).* For a probability law $Q$ on a standard Borel space and $\lambda \ge 0$, write $\Pi_\lambda(Q)$ for the law of the finite marked configuration obtained by drawing $N \sim \mathrm{Pois}(\lambda)$ points independently from $Q$, attaching to each point an independent mark uniform on $(0,1]$, and listing the points in increasing order of their marks; the marks are independent of the points and almost surely distinct, so they only fix a canonical ordering. For a configuration $s$ write $|s|$ for its number of points and, when $|s| \ge n$, write $(Y_1, X_1), \dots, (Y_n, X_n)$ for its first $n$ points. Let $G_0$ be the common known geometry of the hypercube, which is the same tuple for every vertex because all $P_\omega$ share $\mathcal{X}, \mathcal{A}_0, \mathcal{A}_1, \mathcal{B}$, the Euclidean metric and $K_\square$. Define the Poissonized value of the rule at a point $x$ by

$$\widehat{\tau}_n^{\Pi}(x; s) = \begin{cases} T_{n,G_0,x}\big((Y_i, D^{\pm}_{x,i}) : 1 \le i \le n\big), & |s| \ge n, \\ 0, & |s| < n, \end{cases} \qquad D^{\pm}_{x,i} = \big\{\mathbf{1}(X_i \in \mathcal{A}_1) - \mathbf{1}(X_i \in \mathcal{A}_0)\big\} \|X_i - x\|_2,$$

and the Poissonized maximum loss at a vertex $\omega$ by

$$\Lambda_\omega(s) = \max_{1 \le j \le M} \big|\widehat{\tau}_n^{\Pi}(x_j; s) - v_{j,\omega_j}\big|.$$

Both are measurable, each section $T_{n,G_0,x_j}$ being Borel.

*Step 7 (independent blocks and their compressions).* Split a configuration drawn from $\Pi_{2n}(P_\omega)$ according to the partition of $[-3,3]^2$ into the pairwise disjoint disks $S_1, \dots, S_M$ and their complement. Because thinning a marked Poisson configuration by a measurable partition produces independent marked Poisson configurations, the blocks

$$Z_j = \{\text{points with score} \in S_j\}\ (1 \le j \le M), \qquad Z_0 = \Big\{\text{points with score} \notin \bigcup_{k=1}^{M} S_k\Big\}$$

are independent, $Z_j \sim \Pi_{2n\rho}\big(P_\omega(\,\cdot \mid \text{score} \in S_j)\big)$, and $Z_0$ is the corresponding marked Poisson configuration built from the normalized law of $P_\omega$ off $\bigcup_k S_k$ at the complementary intensity. By the bit-structure clause of Lemma 6 the law of $Z_j$ depends on $\omega$ only through $\omega_j$, each disk having the same mass $\rho$, while the law of $Z_0$ is common across all vertices.

Compress the $j$th block by replacing each of its points by its signed observation at $x_j$ under the common geometry, keeping its mark; the compressed blocks $\widetilde{S}_1, \ldots, \widetilde{S}_M$ are the compressed coordinates required by Lemma 1:

$$\widetilde{S}_j = c_j(Z_j), \qquad c_j : (Y, X, \text{mark}) \longmapsto \big(Y,\ \{\mathbf{1}(X \in \mathcal{A}_1) - \mathbf{1}(X \in \mathcal{A}_0)\}\|X - x_j\|_2,\ \text{mark}\big).$$

By the signed-observation clause of Lemma 6, the point law of the compressed $j$th block at bit $b$ is exactly $Q_{j,b}$, so

$$\mathcal{L}(\widetilde{S}_j \mid \Omega_j = b) = \Pi_{2n\rho}(Q_{j,b}), \qquad b \in \{0, 1\}.$$

Poisson tensorization gives KL equal to $2n\rho\,\mathrm{KL}(Q_{j,0}, Q_{j,1})$ for the two unordered marked configurations, and mark-ordering is a measurable map, which cannot increase divergence; hence, with the per-coordinate divergence bound above,

$$\mathrm{KL}\big(\mathcal{L}(\widetilde{S}_j \mid \Omega_j = 0), \mathcal{L}(\widetilde{S}_j \mid \Omega_j = 1)\big) \le 2n\rho\ \mathrm{KL}(Q_{j,0}, Q_{j,1}) \le \frac{1}{4}\log M.$$

*Step 8 (direct-product testing bound).* Let $\Omega_1, \ldots, \Omega_M$ be independent uniform bits, so that the vertex $\Omega = (\Omega_1, \ldots, \Omega_M)$ is uniform on $\{0, 1\}^M$. The block product experiment first draws the common block and the cell blocks conditionally independently according to the bit-conditional laws above, and then superposes those blocks; at vertex $\omega$, the superposed configuration has the canonical marked-Poisson law $\Pi_{2n}(P_\omega)$. Attach to coordinate $j$ the midpoint decoder

$$\widehat{\Omega}_j = \mathbf{1}\big\{|\,\widehat{\tau}_n^{\Pi}(x_j; \cdot) - v_{j,1}| \le |\,\widehat{\tau}_n^{\Pi}(x_j; \cdot) - v_{j,0}|\big\}.$$

The configuration feeding $\widehat{\tau}_n^{\Pi}(x_j; \cdot)$ is obtained by superposing the compressed blocks $c_j(Z_k)$, $k \neq j$, the compressed ancillary block $c_j(Z_0)$, and the already compressed block $\widetilde{S}_j$; the raw $j$th block therefore enters only through $\widetilde{S}_j$. Hence $\widehat{\Omega}_j$ is a measurable function of $(\widetilde{S}_j, (Z_1, \ldots, Z_M), Z_0)$ that is unchanged when the raw coordinate vector is altered only in coordinate $j$ while $(\widetilde{S}_j, Z_{-j}, Z_0)$ is held fixed. All hypotheses of Lemma 1 are thus met, and its Kullback–Leibler consequence with $\kappa = 1/4$, fed by the compressed-coordinate divergence bound, gives for the block experiment

$$\Pr\{\text{at least one decoding error}\} \ge \frac{1}{2}\big\{1 - \exp(-M^{3/4}/2)\big\} \ge \frac{1}{2}\big\{1 - \exp(-1/2)\big\} = d,$$

the middle inequality because $M \ge 1$ forces $M^{3/4} \ge 1$.

*Step 9 (from decoding error to loss, and averaging).* If the midpoint decoder is wrong at coordinate $j$ under the block product experiment, then the estimate is at least as close to the wrong value as to the true one, so the separation $|v_{j,1} - v_{j,0}| = \Delta$ forces

$$\widehat{\Omega}_j \neq \omega_j \quad \Longrightarrow \quad \big|\widehat{\tau}_n^{\Pi}(x_j; s) - v_{j,\omega_j}\big| \ge \frac{\Delta}{2}.$$

After superposition, this is a statement under the canonical marked-Poisson law $\Pi_{2n}(P_\omega)$. Consequently, for every vertex $\omega$,

$$\frac{\Delta}{2} \Pr_{\Pi_{2n}(P_\omega)}\big\{\exists\, j : \widehat{\Omega}_j \neq \omega_j\big\} \le \mathbb{E}_{\Pi_{2n}(P_\omega)}\big[\Lambda_\omega\big].$$

Averaging this inequality over the $2^M$ vertices and bounding the average of the right-hand side by its maximum selects a vertex $\omega$ with

$$\frac{\Delta}{2}\cdot 2^{-M}\sum_{\omega'\in\{0,1\}^M}\Pr_{\Pi_{2n}(P_{\omega'})}\big\{\exists j:\widehat{\Omega}_j\neq\omega'_j\big\}\le\mathbb{E}_{\Pi_{2n}(P_\omega)}\big[\Lambda_\omega\big].$$

The left factor is the average block-product decoding-error probability bounded below in the direct-product display, transported through the superposition map; and $(\Delta/2)d=(\gamma a_n/2)d=4c\,a_n$ by the definition of $c$. Therefore

$$4c\,a_n\le\mathbb{E}_{\Pi_{2n}(P_\omega)}\big[\Lambda_\omega\big].$$

*Step 10 (de-Poissonization).* Write $P=P_\omega$. The hypercube construction gives $P\in\mathcal{P}_{12}(p,\nu,L)$ and the fixed rectangular geometry, so $G_P=G_0$ and each center lies on $\mathcal{B}_P=\partial([-1,1]\times[0,2])$. Define the finite packing loss of the rule on a sample $W$ of size $n$ by

$$\ell^{\mathrm{pack}}_{n,P}(W)=\max_{1\le j\le M}\big|\widehat{\tau}_{n,P}(x_j)-\tau_P(x_j)\big|,$$

which is measurable in $W$ because each section $T_{n,G_P,x_j}$ is Borel and $W\mapsto U^{\pm}_{n,P}(x_j)$ is measurable. On the event $\{|s|\ge n\}$ the Poissonized value at $x_j$ is the value of the rule on the retained prefix, because $G_0=G_P$ makes the two signed observations at $x_j$ coincide, and $\tau_P(x_j)=v_{j,\omega_j}$ by the target identity above; hence

$$\Lambda_\omega(s)=\ell^{\mathrm{pack}}_{n,P}\big((Y_i,X_i)_{i\le n}\big)\qquad\text{whenever }|s|\ge n.$$

On the complementary event the Poissonized value is 0, so $\Lambda_\omega(s)\le\max_j|v_{j,\omega_j}|$; and since the two potential-outcome regressions of Definition 22 satisfy $\sup_{x\in\mathcal{X}_P}|\mu_{t,P}(x)|\le L$, the target $\tau_P=\mu_{1,P}-\mu_{0,P}$ of Definition 23 obeys

$$|v_{j,\omega_j}|=|\tau_P(x_j)|\le 2L.$$

Conditionally on $\{|s|\ge n\}$ the first $n$ points of the configuration are i.i.d. from $P$, and $\Pr\{\mathrm{Pois}(2n)<n\}\le\exp\{-n(1-\log 2)\}$, so splitting the Poissonized expectation over the two events gives

$$\mathbb{E}_{\Pi_{2n}(P)}\big[\Lambda_\omega\big]\le\int\ell^{\mathrm{pack}}_{n,P}\,dP^n+2L\exp\{-n(1-\log 2)\}.$$

By the choice of $N$ the second term is at most $c\,a_n$. Combining this upper bound with the preceding Poissonized lower bound and weakening $4c\,a_n$ to $2c\,a_n$,

$$2c\,a_n\le\int\ell^{\mathrm{pack}}_{n,P}\,dP^n+c\,a_n.$$

Cancelling the finite common term $c\,a_n$ yields

$$c\,a_n\le\int\ell^{\mathrm{pack}}_{n,P}\,dP^n.$$

*Step 11 (passage to the two risks).* Each $x_j$ lies on $\mathcal{B}_P$, so pointwise in the sample

$$\ell^{\mathrm{pack}}_{n,P}\le\sup_{x\in\mathcal{B}_P}\big|\widehat{\tau}_{n,P}(x)-\tau_P(x)\big|.$$

The finite packing loss is measurable, so its integral is at most the outer expectation of the boundary supremum. Hence

$$c\,a_n \le \mathbb{E}_P^*\Big[\sup_{x\in\mathcal{B}_P}\big|\widehat{\tau}_{n,P}(x)-\tau_P(x)\big|\Big].$$

Since $P \in \mathcal{P}_{12}(p,\nu,L)$, taking the supremum over the law class and then the infimum over $\widehat{\tau}_n \in \mathcal{T}_n^{12,\pm}$ gives, by Definition 26,

$$c\,a_n \le R_n^{12,\pm}(p,\nu,L) \qquad \text{for every } n \ge N.$$

Since $P$ also satisfies the fixed hard-geometry condition, the same two steps taken inside the restricted supremum of Definition 27 give

$$c\,a_n \le R_n^{12,\pm,\mathrm{fix}}(p,\nu,L) \qquad \text{for every } n \ge N.$$

*Step 12 (normalization and liminf).* For $n \ge 2$ we have $a_n = (\log n/n)^{1/4} > 0$, hence

$$a_n^{-1} R_n^{12,\pm}(p,\nu,L) = \frac{R_n^{12,\pm}(p,\nu,L)}{a_n}.$$

The two sequences are eventually equal, so their lower limits agree, which is the first asserted identity. Dividing the two eventual risk inequalities by $a_n > 0$ gives, for every $n \ge \max\{N,2\}$,

$$c \le \frac{R_n^{12,\pm}(p,\nu,L)}{a_n}, \qquad c \le \frac{R_n^{12,\pm,\mathrm{fix}}(p,\nu,L)}{a_n},$$

and the order property of $\liminf$ under an eventual lower bound yields

$$c \le \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p,\nu,L)}{a_n}, \qquad c \le \liminf_{n\to\infty} \frac{R_n^{12,\pm,\mathrm{fix}}(p,\nu,L)}{a_n}.$$

This proves the theorem. □

*Proof of Theorem 1.* Write $a_n = (\log n/n)^{1/4}$.

1. By Theorem 2, applied with the same $q$ and $L$, choose $c_{\mathrm{PI}} > 0$ such that

   $$c_{\mathrm{PI}} \le \liminf_{n\to\infty} \frac{R_n^{\mathrm{PI}}}{a_n}.$$

   Set $c_{\mathrm{CTY}} = c_{\mathrm{PI}}$.

2. The two normalizations of the common-map risk agree eventually. Indeed, for all sufficiently large $n$, $a_n > 0$ and

   $$\left(\frac{n}{\log n}\right)^{1/4} R_n^{\mathrm{NP}} = \frac{R_n^{\mathrm{NP}}}{a_n},$$

   because $a_n = (\log n/n)^{1/4}$ is the reciprocal of $(n/\log n)^{1/4}$. Taking liminfs of these eventually equal sequences gives

   $$\liminf_{n\to\infty} \left(\frac{n}{\log n}\right)^{1/4} R_n^{\mathrm{NP}} = \liminf_{n\to\infty} \frac{R_n^{\mathrm{NP}}}{a_n}.$$

3. *For each $n$, the common-map risk dominates the point-indexed risk:*

$$R_n^{\mathrm{PI}} \le R_n^{\mathrm{NP}}.$$

Two facts combine to give this.

First, a common map in Definition 11 is a point-indexed rule in Definition 13: take the same Borel section $T_{n,x} = T_n$ at every boundary point, as recorded by Lemma 3. Hence the infimum defining $R_n^{\mathrm{PI}}$ runs over a superset of the rules defining $R_n^{\mathrm{NP}}$.

Second, for a common-map rule the two expectation conventions coincide term by term. Fix $P \in \mathcal{P}_{\mathrm{NP}}(L, q)$ and a common map $T_n$, and write the boundary loss at the sample point $w$ as

$$\Lambda_P(w) = \sup_{x \in \mathrm{bd}(\mathcal{X}_P)} \left| T_n\big(V_{n,P}(x)\big|_w\big) - \mu_P(x) \right|.$$

The map

$$(x, w) \longmapsto \left| T_n\big(V_{n,P}(x)\big|_w\big) - \mu_P(x) \right|, \qquad (x, w) \in \mathrm{bd}(\mathcal{X}_P) \times (\mathbb{R} \times \mathbb{R}^2)^n,$$

is jointly Borel: $T_n$ is measurable, $(x, w) \mapsto V_{n,P}(x)|_w$ is continuous in $x$ and Borel in $w$, and $\mu_P$ is continuous on the compact set $\mathcal{X}_P \supseteq \mathrm{bd}(\mathcal{X}_P)$ because it lies in the Hölder ball of Definition 4. Consequently, for every level $a$,

$$\{w : \Lambda_P(w) > a\} = \Big\{ w : \exists\, x \in \mathrm{bd}(\mathcal{X}_P),\ \left| T_n\big(V_{n,P}(x)\big|_w\big) - \mu_P(x) \right| > a \Big\}$$

is the projection onto the sample coordinate of a Borel subset of the product of the Polish space $\mathrm{bd}(\mathcal{X}_P)$ with the sample space, hence an analytic set. Analytic sets are universally measurable, so every superlevel set of $\Lambda_P$ is measurable for the completion of the product law generated by $P$; therefore $\Lambda_P$ is measurable on that completed space and its completed integral coincides with its outer expectation,

$$\int \Lambda_P \, d\overline{P^n} = \mathbb{E}_P^* \left[ \sup_{x \in \mathrm{bd}(\mathcal{X}_P)} \left| \widehat{\mu}_n(x) - \mu_P(x) \right| \right].$$

Taking the supremum over $P \in \mathcal{P}_{\mathrm{NP}}(L, q)$ in the last display and then the infimum over the larger point-indexed class in Definition 14 bounds $R_n^{\mathrm{PI}}$ above by the common-map risk in Definition 12, which is the displayed inequality. Dividing by the same nonnegative normalizing factor preserves this order:

$$\frac{R_n^{\mathrm{PI}}}{a_n} \le \frac{R_n^{\mathrm{NP}}}{a_n}.$$

4. Applying monotonicity of liminf to the preceding pointwise normalized-risk inequality gives

$$\liminf_{n \to \infty} \frac{R_n^{\mathrm{PI}}}{a_n} \le \liminf_{n \to \infty} \frac{R_n^{\mathrm{NP}}}{a_n}.$$

Combining this with the choice of $c_{\mathrm{CTY}}$ in the first step yields

$$c_{\mathrm{CTY}} \le \liminf_{n \to \infty} \frac{R_n^{\mathrm{NP}}}{a_n},$$

and the equality of normalizations from the second step gives the displayed conclusion.

$\square$

*Proof of Theorem 2.* Fix $q \geq 1$ and $L \geq 4$, and write $a_n = (\log n/n)^{1/4}$.

1. *A finite packing lower bound, uniform over point-indexed rules.* Let $c_0, c_1, c_w^-, c_w^+, c_{\mathrm{rad}}, \alpha > 0$ with $\alpha < 1/8$ be the constants of Lemma 2, and put

$$d = \frac{1}{2}\left(1 - e^{-1/2}\right) > 0, \qquad c = \frac{c_1 d}{4} > 0.$$

Take $n$ beyond the threshold of that lemma, and let $M_n$, $w_n$, $m_n$, the boundary points $x_1, \ldots, x_{M_n}$, the laws $P_\omega$, and the values $u_{j,0}, u_{j,1}$ be the objects it supplies, with $S = [-1/2, 1/2]^2$ the common covariate support. Two of its clauses are used at once: $M_n \geq c_0 a_n^{-1/q} > 0$, so $M_n \geq 1$; and $m_n > 0$, because $C_j$ contains a relatively open neighbourhood of the support point $x_j$ inside $S$, which therefore has positive covariate mass, and that mass is $m_n$.

Fix a point-indexed rule $\widehat{\mu}_{n,P}^{\mathrm{PI}} \in \mathcal{T}_n^{\mathrm{PI}}$ and law-independent Borel sections $\{T_{n,x}\}$ representing it as in Definition 13.

*The Poissonized product experiment.* Put independent uniform binary coordinates $\Omega_1, \ldots, \Omega_{M_n}$ on $\{0,1\}$. Conditionally on $\Omega = \omega$, let $N$ be Poisson with mean $2n$, independent of an i.i.d. stream of observations from $P_\omega$, and attach to each observation an independent mark uniform on $(0,1]$; list the resulting $N$ marked observations in increasing mark order. Split this configuration along the partition of $S$ into the cells $C_1, \ldots, C_{M_n}$ and their complement, and write

$Z_j =$ the sub-configuration with covariate in $C_j$ $(1 \leq j \leq M_n)$, $\qquad Z_0 =$ the sub-configuration with covari

Splitting a Poisson configuration along a measurable partition makes the blocks independent conditionally on $\Omega$. For $j \geq 1$, $Z_j$ is the marked Poisson configuration with intensity $2n$ times the restriction of $P_\omega$ to $C_j$, equivalently with mean $2nm_n$ and the conditional observation law on $C_j$. The complement block $Z_0$ is the marked Poisson configuration with intensity $2n$ times the restriction of $P_\omega$ to $S \setminus \bigcup_k C_k$; when this complement has positive mass this is the corresponding conditional law with that mean, and when the complement has zero mass it is the zero-intensity Poisson configuration. By the equal-cell-mass, one-bit-locality and common-outside clauses of Lemma 2, the conditional law of $Z_j$ depends on $\Omega$ only through $\Omega_j$, and the conditional law of $Z_0$ is common for all values of $\Omega$, including the zero-intensity case.

*The compressions and the decoders.* The compressed blocks required by Lemma 1 are written $\widetilde{S}_1, \ldots, \widetilde{S}_{M_n}$ here, the plain letter $S$ being reserved for the covariate support; let

$$\widetilde{S}_j = c_j(Z_j), \qquad c_j : \ \big((y,z), \mathrm{mark}\big) \longmapsto \big((y, \|z - x_j\|_2), \mathrm{mark}\big),$$

so that $\widetilde{S}_j$ records, for each atom of the $j$th cell block, its outcome, its distance from $x_j$, and its mark. For coordinate $j$, superpose in mark order the compressed own block $\widetilde{S}_j$, the images under $c_j$ of the other cell blocks $\{Z_k\}_{k \neq j}$, and the image under $c_j$ of the complement block $Z_0$; call the result $s_j$, and set

$$\widehat{u}_j = \begin{cases} T_{n,x_j}\big(\text{first } n \text{ atoms of } s_j \text{ in mark order}\big), & \text{if } s_j \text{ has at least } n \text{ atoms,} \\ 0, & \text{otherwise,} \end{cases} \qquad \widehat{\Omega}_j = \mathbf{1}\big\{|\widehat{u}_j - u_{j,1}| \leq |\widehat{u}_j - u$$

Each $\widehat{\Omega}_j$ is a measurable $\{0,1\}$-valued function of $(\widetilde{S}_j, (Z_1, \ldots, Z_{M_n}), Z_0)$, and it is unchanged when the raw block $Z_j$ is altered while $(\widetilde{S}_j, Z_{-j}, Z_0)$ is held fixed, because $Z_j$ enters $s_j$ only through $\widetilde{S}_j$.

*The coordinatewise divergence budget.* Because $\Pr\{\Omega_j = b\} = 1/2$ for $b \in \{0,1\}$, the conditional law in Definition 1 is defined; write

$$Q_{j,b} = \mathcal{L}(\widetilde{S}_j \mid \Omega_j = b).$$

Then

$$\mathrm{KL}(Q_{j,0}, Q_{j,1}) \le 2\alpha \log M_n \qquad \text{for every } j.$$

Indeed, under $\Omega_j = b$, $\widetilde{S}_j$ is exactly the block of radii at most $w_n$ of the marked Poisson experiment generated by the one-observation outcome-distance law at $x_j$ for the single-bit vertex with $j$th bit $b$. Restricting to that block is a measurable map, so it cannot increase divergence, and a marked Poisson experiment of mean $2n$ has divergence at most twice that of the corresponding $n$-fold product experiment, which is the compressed-distance divergence bounded by $\alpha \log M_n$ in Lemma 2.

*The direct-product step.* The uniform bits $\Omega_1, \ldots, \Omega_{M_n}$, the conditionally independent blocks $(Z_0, Z_1, \ldots, Z_{M_n})$, the compressions $c_j$, and the decoders $\widehat{\Omega}_j$ satisfy all the hypotheses of Lemma 1. Applying that lemma with $\kappa = 2\alpha < 1/4$ and using the previous display gives

$$\Pr\{\text{at least one decoding error}\} \ge \frac{1}{2}\left[1 - \exp\{-M_n^{1-2\alpha}/2\}\right] \ge d,$$

the last inequality because $M_n \ge 1$ and $1 - 2\alpha \ge 0$ give $M_n^{1-2\alpha} \ge 1$.

*From decoding errors to loss, and choice of a vertex.* If $\widehat{\Omega}_j \ne \Omega_j$, then the nearest-value decoder has picked the wrong endpoint, so the triangle inequality and the signal separation $|u_{j,1} - u_{j,0}| \ge c_1 a_n$ give

$$\left|\widehat{u}_j - u_{j,\Omega_j}\right| \;\ge\; \frac{c_1 a_n}{2}.$$

Multiplying the previous two displays and averaging over the uniform prior on $\{0,1\}^{M_n}$ selects a vertex $\omega$ such that, writing $\mathbb{E}_{2n}^{\omega}$ for expectation in the mean-$2n$ marked Poisson experiment under $P_\omega$,

$$\mathbb{E}_{2n}^{\omega}\left[\max_{1 \le j \le M_n} \left|\widehat{u}_j - u_{j,\omega_j}\right|\right] \;\ge\; \frac{c_1 a_n}{2} \cdot d = 2c\, a_n.$$

*De-poissonization.* Put $P = P_\omega$, so $P \in \mathcal{P}_{\mathrm{NP}}(L, q)$, $\mathcal{X}_P = S$, and every $x_j \in \mathrm{bd}(\mathcal{X}_P)$. Define the finite packing loss of the rule at a sample point $w$ of size $n$ by

$$\ell_{n,P}^{\mathrm{pack}}(w) = \max_{1 \le j \le M_n} \left|\widehat{\mu}_{n,P}^{\mathrm{PI}}(x_j) - \mu_P(x_j)\right| \quad \text{evaluated at } w,$$

a measurable function of $w$, being a finite maximum of compositions of the Borel sections $T_{n,x_j}$ with the measurable distance compressions of Definition 10. On the event $\{N \ge n\}$ the first $n$ atoms in mark order are an exact i.i.d. sample from $P$, and by construction $\widehat{u}_j$ is then the value of the rule at $x_j$ on that retained sample and $u_{j,\omega_j} = \mu_P(x_j)$; hence the Poissonized maximum equals $\ell_{n,P}^{\mathrm{pack}}$ of the retained sample. On the complementary event

$\{N < n\}$ every $\widehat{u}_j$ equals the fallback value 0, so the maximum is at most $\max_j |u_{j,\omega_j}| = \max_j |\mu_P(x_j)| \le L$, the envelope coming from the zeroth-order derivative bound in $\mu_P \in \mathcal{H}^q(\mathcal{X}_P, L)$ of Definition 4. Since $N$ is Poisson with mean $2n$, a Chernoff bound for its lower tail gives $\Pr\{N < n\} \le \exp\{-n(1-\log 2)\}$, and therefore

$$\mathbb{E}^{\omega}_{2n}\left[\max_{1\le j\le M_n}\left|\widehat{u}_j - u_{j,\omega_j}\right|\right] \le \mathbb{E}_P\left[\ell^{\mathrm{pack}}_{n,P}\right] + L\exp\{-n(1-\log 2)\}.$$

Because $\exp\{-n(1-\log 2)\} = o(a_n)$, we may enlarge the threshold on $n$ so that $L\exp\{-n(1-\log 2)\} \le c\,a_n$. Combining the last three displays and cancelling one copy of the finite quantity $c\,a_n$ yields: for every point-indexed rule $\widehat{\mu}^{\mathrm{PI}}_{n,P} \in \mathcal{T}^{\mathrm{PI}}_n$ there are $P \in \mathcal{P}_{\mathrm{NP}}(L,q)$, an integer $M_n$, and boundary points $x_1, \ldots, x_{M_n} \in \mathrm{bd}(\mathcal{X}_P)$ with

$$\mathbb{E}_{P^n}\left[\max_{1\le j\le M_n}\left|\widehat{\mu}^{\mathrm{PI}}_{n,P}(x_j) - \mu_P(x_j)\right|\right] \ge c\,a_n.$$

2. For such $P$ and $x_1, \ldots, x_{M_n}$, the finite maximum is bounded pointwise by the full boundary loss because each $x_j$ lies on $\mathrm{bd}(\mathcal{X}_P)$:

$$\max_{1\le j\le M_n}\left|\widehat{\mu}^{\mathrm{PI}}_{n,P}(x_j) - \mu_P(x_j)\right| \le \sup_{x\in\mathrm{bd}(\mathcal{X}_P)}\left|\widehat{\mu}^{\mathrm{PI}}_{n,P}(x) - \mu_P(x)\right|.$$

The left-hand side is measurable, so its ordinary expectation is bounded above by the outer expectation of the right-hand side, that is, by the quantity appearing in Definition 14. Taking the supremum over $P \in \mathcal{P}_{\mathrm{NP}}(L,q)$, and then the infimum over all rules in $\mathcal{T}^{\mathrm{PI}}_n$, gives a sample-size threshold $n_0$, distinct from the Poisson count $N$ introduced in the Poissonized experiment, such that

$$R^{\mathrm{PI}}_n \ge c\,a_n \qquad \text{for every } n \ge n_0.$$

3. For every $n \ge 2$, $a_n > 0$ and

$$\left(\frac{n}{\log n}\right)^{1/4} R^{\mathrm{PI}}_n = \frac{R^{\mathrm{PI}}_n}{a_n},$$

because $(n/\log n)^{1/4}$ is the reciprocal of $a_n$. Thus the two normalized sequences agree eventually, so monotonicity of $\liminf$ in both directions gives

$$\liminf_{n\to\infty}\left(\frac{n}{\log n}\right)^{1/4} R^{\mathrm{PI}}_n = \liminf_{n\to\infty}\frac{R^{\mathrm{PI}}_n}{a_n}.$$

4. For every $n \ge \max\{n_0, 2\}$, the positivity of $a_n$ allows division in the preceding eventual lower bound:

$$c \le \frac{R^{\mathrm{PI}}_n}{a_n}.$$

Taking $\liminf$ preserves this eventual lower bound. With $c_{\mathrm{PI}} = c > 0$, the asserted equality of normalizations and the lower bound

$$\liminf_{n\to\infty}\frac{R^{\mathrm{PI}}_n}{a_n} \ge c_{\mathrm{PI}}$$

follow.

$\square$

*Proof of Theorem 4.* 1. Apply Theorem 3 for the chosen order $p$, and let $L_{\mathrm{conv}} \geq 48$ be the envelope threshold it provides. Set

$$L_0 = \max\{L_{\mathrm{conv}}, 4\}.$$

Then $L_0 \geq 48$. Fix $\nu \geq 2$, $L \geq L_0$, and the three displayed hypotheses of the theorem. The choice of $L_0$ gives both $L \geq L_{\mathrm{conv}}$ and $L \geq 4$.

2. The converse theorem at $L_{\mathrm{conv}}$ supplies a constant $c > 0$ such that

$$c \leq \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p,\nu,L)}{a_n}.$$

That theorem also records the parallel lower bound for the fixed-geometry risk $R_n^{12,\pm,\mathrm{fix}}(p,\nu,L)$ of Definition 27; only the full-risk statement just displayed is used below.

3. By Proposition 1, applied with $L \geq 4$ and the distance-identification, first-order-bias, and expected-maximal-bound hypotheses, there are $C_0 > 0$ and $N$ such that, for every $n \geq N$,

$$\sup_{P\in\mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P} \left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x)\right|\right] \leq C_0 a_n.$$

Since $a_n > 0$ eventually, division by $a_n$ and passage to the limsup give

$$\limsup_{n\to\infty} \frac{\sup_{P\in\mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P} \left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x)\right|\right]}{a_n} \leq C_0.$$

4. *The explicit rule is admissible.* For every $n \geq 2$ the frontier bandwidth $a_n$ is positive, so the winsorization level $B(a_n) = a_n^{-1/3}$ of Definition 28 is well defined, and the estimator is generated by the law-independent family of sections

$$T_{n,G,x}(u) = \max\Big\{-2L, \min\Big\{2L,\ e_1^\top\big(\widehat{\beta}_1^B(a_n;u) - \widehat{\beta}_0^B(a_n;u)\big)\Big\}\Big\}, \qquad u \in (\mathbb{R}\times\mathbb{R})^n,$$

where $\widehat{\beta}_t^B(a_n;u)$ is the stabilized coefficient of Definition 28 computed from the signed-distance sample $u$. This family does not depend on the unknown member law, and each fixed-$(G,x)$ section is Borel in $u$, because the empirical Gram $\widehat{\Psi}_{t,x}(a_n)$, the uniform conditioning event $\{(2L)^{-1}\sum_j v_j^2 \leq v^\top\widehat{\Psi}_{t,x}(a_n)v \text{ for all } v\}$, the matrix inverse on that event, the winsorized empirical moment $\widehat{m}_{t,x}^B(a_n)$, and the clipping map are all Borel functions of $u$. Moreover, for every $P \in \mathcal{P}_{12}(p,\nu,L)$ and every $x \in \mathcal{B}_P$, the signed-distance sample built from the known geometry $G_P$ of Definition 20 is exactly $U_{n,P}^{\pm}(x)$ of Definition 21, since the treatment indicator is the deterministic function $\mathbf{1}\{X \in \mathcal{A}_{1,P}\}$ of the covariate recorded in $G_P$. Hence

$$\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}} \in \mathcal{T}_n^{12,\pm}.$$

Therefore the minimax risk of Definition 26 is bounded above by the risk of this explicit rule:

$$R_n^{12,\pm}(p,\nu,L) \leq \sup_{P\in\mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P} \left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x)\right|\right].$$

After normalizing by $a_n$, this inequality holds eventually in $n$, and monotonicity of limsup yields

$$\limsup_{n\to\infty} \frac{R_n^{12,\pm}(p,\nu,L)}{a_n} \le \limsup_{n\to\infty} \frac{\sup_{P\in\mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P} \left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x)\right|\right]}{a_n} \le C_0.$$

5. Finally set

$$C = \max\{C_0, c\}.$$

Then $0 < c \le C$. The lower bound from Step 2, the general inequality $\liminf \le \limsup$, and the two upper bounds from Steps 3 and 4 give

$$c \le \liminf_{n\to\infty} \frac{R_n^{12,\pm}(p,\nu,L)}{a_n} \le \limsup_{n\to\infty} \frac{R_n^{12,\pm}(p,\nu,L)}{a_n} \le C,$$

and

$$\limsup_{n\to\infty} \frac{\sup_{P\in\mathcal{P}_{12}(p,\nu,L)} \mathbb{E}_P^*\left[\sup_{x\in\mathcal{B}_P} \left|\widehat{\tau}_{n,a_n}^{12,\mathrm{LP}}(x) - \tau_P(x)\right|\right]}{a_n} \le C.$$

These are exactly the asserted frontier bounds. □

# References


Timothy B. Armstrong and Michal Kolesar. Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, 11(1):1–39, 2020. doi: 10.3982/QE1199.

Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014. doi: 10.3982/ECTA11757.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210, 2020. doi: 10.1093/ectj/utz022.

Matias D. Cattaneo and Rocio Titiunik. Regression discontinuity designs. *Annual Review of Economics*, 14:821–851, 2022. doi: 10.1146/annurev-economics-051520-021409.

Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik. *A Practical Introduction to Regression Discontinuity Designs*. Cambridge University Press, 2019. doi: 10.1017/9781108684606.

Matias D. Cattaneo, Nicolas Idrobo, and Rocio Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press, 2023.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. Boundary discontinuity designs: Theory and practice, 2025.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. Estimation and inference in boundary discontinuity designs: Distance-based methods. *Journal of Econometrics*, 256:106266, 2026. doi: 10.1016/j.jeconom.2026.106266. URL `https://mdcattaneo.github.io/papers/Cattaneo-Titiunik-Yu_2026_JOE--Supplement.pdf`. Supplemental Appendix, May 25, 2026.

Jin-young Choi and Myoung-jae Lee. Minimum distance estimator for sharp regression discontinuity with multiple running variables. *Economics Letters*, 162:10–14, 2018. doi: 10.1016/j.econlet.2017.10.002.

Antonio Cuevas and Alberto Rodriguez-Casal. On boundary estimation. *Advances in Applied Probability*, 36(2):340–354, 2004. doi: 10.1239/aap/1086957575.

Juan D. Díaz and José R. Zubizarreta. Complex discontinuity designs using covariates: Impact of school grade retention on later life outcomes in chile. *The Annals of Applied Statistics*, 17 (1):67–88, 2023. doi: 10.1214/22-AOAS1616.

Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1996. ISBN 9780412983214.

Jinyong Hahn, Petra Todd, and Wilbert van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001. doi: 10.1111/1468-0262.00183.

Guido W. Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959, 2012. doi: 10.1093/restud/rdr043.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008. doi: 10.1016/j.jeconom.2007.05.001.

Guido W. Imbens and Stefan Wager. Optimized regression discontinuity designs. *The Review of Economics and Statistics*, 101(2):264–278, 2019. doi: 10.1162/rest_a_00793.

Luke J. Keele and Rocio Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015. doi: 10.1093/pan/mpu014.

Luke J. Keele and Rocio Titiunik. Natural experiments based on geography. *Political Science Research and Methods*, 4(1):65–95, 2016. doi: 10.1017/psrm.2015.4.

Luke J. Keele, Rocio Titiunik, and Jose R. Zubizarreta. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1): 223–239, 2015. doi: 10.1111/rssa.12056.

Emmett B. Kendall, Brenden Beck, and Joseph Antonelli. Robust inference for geographic regression discontinuity designs: Assessing the impact of police precincts. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 189(3):1521–1545, 2026. doi: 10.1093/jrsssa/qnaf128.

Michal Kolesár and Christoph Rothe. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8):2277–2304, 2018. doi: 10.1257/aer.20160945.

David S. Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010. doi: 10.1257/jel.48.2.281.

Enno Mammen and Alexandre B. Tsybakov. Asymptotical minimax recovery of sets with smooth boundaries. *The Annals of Statistics*, 23(2):502–524, 1995. doi: 10.1214/aos/1176324533.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2008. doi: 10.1016/j.jeconom.2007.05.005.

Eugenio Felipe Merlano. Boundary estimation in the regression-discontinuity design: Evidence for a merit- and need-based financial aid program, 2025.

John P. Papay, John B. Willett, and Richard J. Murnane. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161(2):203–207, 2011. doi: 10.1016/j.jeconom.2010.12.008.

David Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer, New York, 1984. ISBN 9780387909905. doi: 10.1007/978-1-4612-5254-2.

Sean F. Reardon and Joseph P. Robinson. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104, 2012. doi: 10.1080/19345747.2011.609583.

David Ruppert and Matthew P. Wand. Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3):1346–1370, 1994. doi: 10.1214/aos/1176325632.

Masayuki Sawada, Takuya Ishihara, Daisuke Kurisu, and Yasumasa Matsuda. Local-polynomial estimation for multivariate regression discontinuity designs, 2024.

Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982. doi: 10.1214/aos/1176345969.

Donald L. Thistlethwaite and Donald T. Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317, 1960. doi: 10.1037/h0044319.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 9780387790510. doi: 10.1007/b13794.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York, 1996. ISBN 9780387946405. doi: 10.1007/978-1-4757-2545-2.

Vladimir N. Vapnik and Alexey Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971. doi: 10.1137/1116025.

Vivian C. Wong, Peter M. Steiner, and Thomas D. Cook. Analyzing regression-discontinuity designs with multiple assignment variables. *Journal of Educational and Behavioral Statistics*, 38(2):107–141, 2013. doi: 10.3102/1076998611432172.