Critical-Grid Debiasing and Transport Certificates for Supercritical Minimax Wasserstein Distance Estimation

April 2026

Abstract

Let $P, Q \in \mathcal{P}([0,1]^d)$ be observed through two independent samples of size N, and consider minimax estimation of $W_p(P,Q)$ in the supercritical regime $d > 2p$. The empirical plug-in estimator gives the scale $N^{-1/d}$, while the Niles–Weed–Rigollet lower-bound mechanism gives the smaller candidate scale

$$\eta_N = (N \log N)^{-1/d}.$$

The manuscript develops critical-grid debiasing, multiscale polynomial transport estimation, transport certificates, and finite-LP curvature methods for the sharp law. The unrestricted upper bound is reduced, with constants and no loss of scale, to an adaptive finite Kantorovich linear-program value on a Euclidean grid with $\asymp N \log N$ atoms. The target law is proved on a large family of critical subclasses retaining the same $N \log N$-alphabet difficulty as the lower-bound construction.

The first positive engine is an exact rooted total-variation skeleton principle. It converts the Wasserstein value into finite weighted sums of large-alphabet L_1 distances, where the effective $N \log N$ gain from functional estimation is available. This yields exact minimax laws on paired Euclidean grids, finite-band and packed direct sums, dyadic pair-isolation models, critical laminar hierarchies, hierarchical tree classes, sparse-shortcut graph classes, continuum blob lifts, partition lifts, full-support paired cores, contiguous split shells, martingale/Haar shells, smooth and real-analytic lower cores, and local block models.

A new complete multiscale theorem solves the full additive p-cost dyadic tree transport functional for all laws on the leaf set:

$$\mathfrak{T}_{p,J}(P,Q) = \sum_{j \le J} 2^{-pj} \|P_j - Q_j\|_1, \qquad 2^{dJ} \asymp N \log N,$$

with powered risk $O(2^{-pJ})$ and root risk $O(2^{-J})$, and proves the matching lower bound. The proof uses polynomial large-alphabet L_1 debiasing on every scale. A complementary theorem shows that raw four-sample diagonal centering misses the logarithmic improvement already on a single tree level with $2^{dj} \asymp N$. Thus the missing Euclidean proof must contain genuine polynomial or regularized LP debiasing, not merely diagonal cancellation.

The second positive engine is transport certification. A high-mass tight-graph theorem gives the critical rate whenever an optimal plan is mostly supported on a low-entropy family of nearly tight graphs. Its continuum form yields arbitrary-source Monge dictionary results, including translations, affine Brenier maps, finite piecewise cyclically monotone dictionaries, and smooth uniformly convex Brenier certificate classes. These theorems solve regimes in which the source law may have full ambient complexity while the transport rule has small description complexity.

The third engine is localized dual compression for the finite Kantorovich LP. Gap-shell, contact-defect, mass-contact, smooth-phase, and low-dimensional-phase theorems are proved. A diagonal entropy theorem shows that a purely uncentered localized-contact proof cannot hold on the full Euclidean grid, because at $r = s$ the near-active dual face contains an exponentially large h_N^p-separated sign cube. Four-sample centering has exact diagonal cancellation and $N^{-1/2}$ variance, but the new tree obstruction shows that centering must be coupled to large-alphabet debiasing.

The remaining unrestricted finite theorem is therefore formulated as adaptive polynomialized debiasing of the critical Euclidean Kantorovich LP. The manuscript proves scale-adaptive powered-to-root reduction, finite von Mises trace identities, active-face and Richardson criteria, and a theorem showing that any estimator satisfying the adaptive finite-grid powered risk yields the full unrestricted minimax upper bound. For $p = 2, d > 4$, the weighted tangent theory is retained: the bounded-density critical cone is solved at scale η_N, the empirical estimator solves the far annulus, and a cross-fitted weighted H_ρ^{-1} spectral estimator closes the intermediate strip for smooth non-flat small Brenier deformations. A counterexample shows that the unweighted H^{-1} strip method is false on non-uniform backgrounds.

Contents

1 Problem, scales, and objective

Let $d \geq 1$, $p \geq 1$, and $\mathcal{P}_d = \mathcal{P}([0,1]^d)$. For $P, Q \in \mathcal{P}_d$,

$$W_p(P,Q) := \left(\inf_{\pi \in \Pi(P,Q)} \int \|x-y\|_2^p \, d\pi(x,y) \right)^{1/p}.$$

We observe

$$X_1, \dots, X_N \overset{\text{i.i.d.}}{\sim} P, \qquad Y_1, \dots, Y_N \overset{\text{i.i.d.}}{\sim} Q$$

and estimate the scalar functional $W_p(P,Q)$. The absolute and quadratic minimax risks are

$$M^{\text{abs}}_{N,d,p} := \inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}_d} \mathbb{E}|\widehat{W} - W_p(P,Q)|,$$

$$M^{\text{sq}}_{N,d,p} := \inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}_d} \mathbb{E}(\widehat{W} - W_p(P,Q))^2.$$

Throughout the paper

$$\eta_N = (N \log N)^{-1/d}.$$

When $p = 2$ we also write

$$\gamma_N := N^{-1/d} (\log N)^{1/d} = \frac{N^{-2/d}}{\eta_N}.$$

The known supercritical picture is this. If $d > 2p$, empirical-Wasserstein theory gives

$$\sup_{P,Q \in \mathcal{P}_d} \mathbb{E}|W_p(P_N, Q_N) - W_p(P,Q)| \lesssim_{d,p} N^{-1/d},$$

where P_N, Q_N are the empirical measures; see, for example, [3, 13]. Niles-Weed and Rigollet's spiked construction gives

$$M^{\text{abs}}_{N,d,p} \gtrsim_{d,p} \eta_N, \qquad M^{\text{sq}}_{N,d,p} \gtrsim_{d,p} \eta_N^2$$

in the high-dimensional regime [10]. The open minimax question is whether the logarithmic lower scale is sharp for the unrestricted class.

A natural global convex-function covering route does not provide a proof of the unrestricted theorem. Such a route would require the convex-body exponent $(d-1)/2$, whereas the Bronshtein exponent for uniformly bounded Lipschitz convex functions on $[0,1]^d$ is $d/2$. Consequently that route yields only the smooth-cost scale $N^{-2/d}$ for cost estimation and cannot close the unrestricted distance problem. The analysis below uses this obstruction as a guide: global uncentered entropy is replaced by critical-grid reduction, exact skeletons, transport certificates, localized dual geometry, and centered curvature.

1.1 What is proved here

The paper has seven proved components.

(i) A target-scale continuum problem is reduced exactly to a critical discrete grid with $|G_N| \asymp N \log N$. This is the central finite obstruction.

(ii) The local diagonal law is sharp at scale η_N. Away from the diagonal, smooth-cost plug-in theory gives a target upper bound once $W_p(P,Q)$ is sufficiently large. For $p = 2$ this threshold is exactly γ_N.

(iii) A total-variation skeleton theorem solves all exact rooted L_1-transport models at scale η_N. This preserves the useful part of the discrete construction without presenting an unsupported global solution.

(iv) Transport-certificate theorems solve a complementary regime in which an optimal plan is certified by a low-complexity tight graph or Monge dictionary. These include arbitrary-source translation laws, affine and smooth Brenier dictionaries, finite piecewise cyclically monotone dictionaries, and high-mass near-graph plans.

(v) Finite dual catalogs, active-face margins, and low-dimensional dual manifolds give target-scale estimators whenever the active Kantorovich face is compressible.

(vi) A localized gap-shell theorem replaces global entropy by near-active entropy. The contact-defect identity shows that near-active duals are geometrically pinned along optimal plans. The mass-contact net theorem gives a proved target-scale class beyond exact skeletons. Smooth and low-dimensional semidual phase theorems give another route beyond L_1-exactness.

(vii) For W_2, $d > 4$, bounded densities have a solved critical cone $W_2 \lesssim \eta_N$ and a solved far annulus $W_2 \gtrsim \gamma_N$. The weighted tangent theorem closes the intermediate strip on smooth non-flat small-deformation classes by replacing the flat H^{-1} model with the correct weighted metric.

The unrestricted overlapping critical grid remains the decisive object. The reduction in Section 2 gives a precise finite-dimensional problem whose solution would immediately imply the full minimax law, and Sections 18 to 21 and 23 identify concrete mechanisms for that finite problem.

2 Critical-grid reduction

Let $h \in (0,1)$ be dyadic and let $G_h \subset [0,1]^d$ denote the grid of dyadic cell centers. Let

$$\kappa_h : [0,1]^d \to G_h$$

map each point to the center of its dyadic cell, and write

$$P^h := (\kappa_h)_\# P, \qquad Q^h := (\kappa_h)_\# Q.$$

Lemma 2.1 (Quantization error)**.** *For every $p \geq 1$ and every $P, Q \in \mathcal{P}_d$,*

$$W_p(P, P^h) \leq \frac{\sqrt{d}}{2} h, \qquad W_p(Q, Q^h) \leq \frac{\sqrt{d}}{2} h,$$

and therefore

$$|W_p(P,Q) - W_p(P^h, Q^h)| \leq \sqrt{d}\, h.$$

Proof. Couple $X \sim P$ with $\kappa_h(X)$. The displacement is at most $\sqrt{d}\,h/2$ pointwise. This proves the first bound; the second is identical. The final display follows from the triangle inequality for W_p. □

Choose h_N dyadic with

$$c_d \eta_N \leq h_N \leq C_d \eta_N, \qquad |G_{h_N}| \asymp h_N^{-d} \asymp N \log N.$$

Define the finite-grid minimax risk

$$M^{\mathrm{grid}}_{N,h,p} := \inf_{\widehat{W}} \sup_{\mu,\nu \in \mathcal{P}(G_h)} \mathbb{E}_{\mu,\nu} |\widehat{W} - W_p(\mu,\nu)|,$$

where the samples are now cell labels in G_h.

Theorem 2.2 (Equivalence with the critical grid)**.** *For $h_N \asymp \eta_N$,*

$$M^{\mathrm{abs}}_{N,d,p} \leq M^{\mathrm{grid}}_{N,h_N,p} + C_{d,p} h_N.$$

Consequently an estimator on the critical grid with risk $O(h_N)$ implies the unrestricted continuum upper bound

$$M^{\mathrm{abs}}_{N,d,p} \lesssim_{d,p} \eta_N.$$

The same assertion holds for quadratic risk, with h_N^2 in place of h_N.

Proof. Given continuum samples, replace them by their cell centers. This produces exactly N samples from P^{h_N} and N samples from Q^{h_N}. Apply an optimal grid estimator $\widehat{W}_{\mathrm{grid}}$ to these labels. By Theorem 2.1,

$$|\widehat{W}_{\mathrm{grid}} - W_p(P,Q)| \leq |\widehat{W}_{\mathrm{grid}} - W_p(P^{h_N}, Q^{h_N})| + C_d h_N.$$

Taking expectations and then the supremum proves the absolute-risk bound. Squaring the same inequality gives the quadratic-risk version. □

Thus the complete minimax upper bound is equivalent to the following finite problem.

Problem 2.3 (Critical finite-grid transport)**.** *Let G_N be a d-dimensional Euclidean grid with $|G_N| \asymp N \log N$ and mesh $h_N \asymp (N \log N)^{-1/d}$. Decide whether*

$$M^{\mathrm{grid}}_{N,h_N,p} \lesssim_{d,p} h_N.$$

The Niles–Weed–Rigollet lower bound is supported on such critical grids. Hence a positive answer to Problem 2.3 would close the unrestricted problem; a negative answer would have to improve the known lower bound.

3 Local and annular laws

The following elementary conversion is used repeatedly.

Lemma 3.1 (Cost-to-distance conversion)**.** *Let $p > 1$. If $x \geq 0$, $y \geq r > 0$, then*

$$|x - y| \leq C_p r^{1-p} |x^p - y^p|.$$

Proof. If $x \geq r/2$, the mean-value theorem gives

$$|x^p - y^p| = p\xi^{p-1}|x - y|$$

with $\xi \geq r/2$. If $x < r/2$, then $y \geq r$ and

$$\frac{y - x}{y^p - x^p} \leq \frac{y}{y^p - (r/2)^p} \leq \frac{r^{1-p}}{1 - 2^{-p}}.$$

Combining the two cases proves the claim. □

3.1 Exact local diagonal law

Theorem 3.2 (Local diagonal law)**.** *Assume $d > 2p$. There exists $A_0 = A_0(d,p) > 0$ such that*

$$\inf_{\widehat{W}} \sup_{W_p(P,Q)\le A_0\eta_N} \mathbb{E}|\widehat{W} - W_p(P,Q)| \asymp_{d,p} \eta_N,$$

and

$$\inf_{\widehat{W}} \sup_{W_p(P,Q)\le A_0\eta_N} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \asymp_{d,p} \eta_N^2.$$

Proof. The upper bound is attained by the estimator $\widehat{W} \equiv 0$. For the lower bound, use the spiked transport construction of Niles-Weed and Rigollet [10]. Their hard alternatives live on a dyadic grid with $2^{Jd} \asymp N \log N$, and each active displacement has length $2^{-J} \asymp \eta_N$. Thus all hard pairs satisfy $W_p(P,Q) \le A_0\eta_N$ for a fixed A_0, while Le Cam's method gives absolute separation of order η_N. The quadratic bound follows from the same two-point or mixture testing argument with squared loss. □

The local theorem explains why no estimator can beat η_N uniformly, even arbitrarily close to the diagonal. It also explains why the empirical plug-in is not locally minimax.

Proposition 3.3 (Plug-in is locally suboptimal)**.** *For every fixed $A > 0$,*

$$\sup_{W_p(P,Q)\le A\eta_N} \mathbb{E}\big(W_p(P_N,Q_N) - W_p(P,Q)\big)^2 \gtrsim_{d,p} N^{-2/d}.$$

Proof. Let $z_1,\ldots,z_N \in [0,1]^d$ be $c_d N^{-1/d}$-separated and let

$$Q_0 = \frac{1}{N}\sum_{j=1}^N \delta_{z_j}.$$

Take $P = Q = Q_0$. If K_j, K_j' are the two independent multinomial count vectors, then

$$\mathrm{TV}(Q_{0,N}, Q'_{0,N}) = \frac{1}{2N}\sum_{j=1}^N |K_j - K_j'|.$$

By symmetry,

$$\mathbb{E}\,\mathrm{TV}(Q_{0,N}, Q'_{0,N}) = \frac{1}{2}\mathbb{E}|K_1 - K_1'| \ge \mathbb{P}(K_1 = 0, K_1' = 1)$$

which is bounded below by a positive constant. Any unmatched mass must move at least $c_d N^{-1/d}$, hence

$$\mathbb{E}W_p(Q_{0,N}, Q'_{0,N}) \gtrsim_d N^{-1/d}.$$

Jensen's inequality gives the squared lower bound. □

3.2 Macroscopic plug-in law

The sharp cost-level empirical theorem of Manole and Niles-Weed implies

$$\sup_{P,Q\in\mathcal{P}_d} \mathbb{E}\Big|W_p(P_N,Q_N)^p - W_p(P,Q)^p\Big| \lesssim_{d,p} N^{-(p\wedge 2)/d}.$$

Set

$$a_{N,p} := N^{-(p\wedge 2)/d}.$$

For $p > 1$, define

$$r_{N,p} := \left(\frac{a_{N,p}}{\eta_N}\right)^{1/(p-1)}.$$

For $p = 2$, $r_{N,2} = \gamma_N$.

Theorem 3.4 (Macroscopic plug-in control)**.** *Assume $p > 1$ and $d > 2p$. Let*

$$\mathcal{A}_t := \{(P,Q) : t/2 \le W_p(P,Q) \le 2t\}.$$

Then

$$\sup_{(P,Q)\in\mathcal{A}_t} \mathbb{E}|W_p(P_N,Q_N) - W_p(P,Q)| \lesssim_{d,p} a_{N,p}t^{1-p}.$$

In particular, for every fixed $B > 0$,

$$\sup_{W_p(P,Q)\ge Br_{N,p}} \mathbb{E}|W_p(P_N,Q_N) - W_p(P,Q)| \lesssim_{B,d,p} \eta_N.$$

The corresponding quadratic risk is bounded by $C_{B,d,p}\eta_N^2$.

Proof. Let

$$Z = \Big|W_p(P_N,Q_N)^p - W_p(P,Q)^p\Big|.$$

On $\mathcal{A}_t$, Theorem 3.1 gives

$$|W_p(P_N,Q_N) - W_p(P,Q)| \le C_p t^{1-p} Z.$$

Taking expectations and using the cost-level theorem proves the first assertion. If $t \ge Br_{N,p}$, then $a_{N,p}t^{1-p} \le B^{1-p}\eta_N$.

For the quadratic statement, changing one observation changes $W_p(P_N,Q_N)^p$ by at most $C_{d,p}/N$. Efron–Stein therefore gives variance $O(N^{-1})$, which is $O(a_{N,p}^2)$ in the supercritical regime $d > 2p$. Thus $\mathbb{E}Z^2 \lesssim a_{N,p}^2$, and the same conversion yields the claim. □

4 Annular linearization and the finite critical core

4.1 Critical linearization

For a dyadic level $J \geq 1$, let $\mathcal{D}_J$ be the partition of $[0,1]^d$ into cubes of side length $h_J := 2^{-J}$, and let c_R be the center of $R \in \mathcal{D}_J$. For $p > 1$, define the piecewise-affine linearized cost

$$c^{\mathrm{lin}}_{J,p}(x,y) := \begin{cases} 0, & R = S, \\ \|c_R - c_S\|_2^p + p\|c_R - c_S\|_2^{p-2}(c_R - c_S) \cdot \big[(x - c_R) - (y - c_S)\big], & R \neq S, \end{cases}$$

for $x \in R$, $y \in S$, and the associated transport value

$$T^{\mathrm{lin}}_{J,p}(P,Q) := \inf_{\pi \in \Pi(P,Q)} \int c^{\mathrm{lin}}_{J,p}(x,y)\, d\pi(x,y).$$

Lemma 4.1 (Sup-norm Lipschitz property). *For bounded measurable costs g, h on $[0,1]^d \times [0,1]^d$,*

$$|T_g(P,Q) - T_h(P,Q)| \leq \|g - h\|_\infty \qquad \forall\, P, Q \in \mathcal{P}_d.$$

Proof. For any coupling π,

$$\int g\, d\pi \leq \int h\, d\pi + \|g - h\|_\infty.$$

Taking infima over π and interchanging g, h gives the result. ☐

Proposition 4.2 (Critical linearization). *Let $p > 1$ and set $\beta_p := p \wedge 2$. Then for every dyadic level J,*

$$\sup_{P,Q \in \mathcal{P}_d} \left| W_p(P,Q)^p - T^{\mathrm{lin}}_{J,p}(P,Q) \right| \lesssim_{d,p} h_J^{\beta_p}.$$

Proof. Fix cells $R \neq S$ and write $z_0 := c_R - c_S$ and

$$\Delta := (x - c_R) - (y - c_S), \qquad x \in R,\ y \in S.$$

Then $\|\Delta\|_2 \leq 2\sqrt{d}\, h_J$. For $h_p(z) := \|z\|_2^p$ we have

$$h_p(z_0 + \Delta) = h_p(z_0) + \nabla h_p(z_0) \cdot \Delta + \mathcal{R}.$$

If $1 < p \leq 2$, then $\nabla h_p(z) = p\|z\|_2^{p-2} z$ is $(p-1)$-Hölder, hence

$$|\mathcal{R}| \leq C_p \|\Delta\|_2^p \lesssim_{d,p} h_J^p.$$

If $p \geq 2$, the Hessian satisfies

$$\|\nabla^2 h_p(z)\|_{\mathrm{op}} \leq p(p-1)\|z\|_2^{p-2} \leq C_{d,p}$$

on $[0,1]^d - [0,1]^d$, so Taylor's theorem gives

$$|\mathcal{R}| \lesssim_{d,p} h_J^2.$$

Same-cell pairs contribute at most $C_d h_J^p$, which is also $O(h_J^{\beta_p})$. Thus

$$\|\, \|x - y\|_2^p - c^{\mathrm{lin}}_{J,p}(x,y) \|_\infty \lesssim_{d,p} h_J^{\beta_p}.$$

Applying Theorem 4.1 proves the claim. ☐

4.2 Annular reduction

Let $t \in [\eta_N, 1]$ and choose $J(t)$ so that

$$h_t^{\beta_p} \asymp \eta_N t^{p-1}, \qquad h_t := 2^{-J(t)}.$$

Write

$$T^{\mathrm{lin}}_{t,p} := T^{\mathrm{lin}}_{J(t),p}.$$

Corollary 4.3 (Annular linearization). *For every $p > 1$ and every $t \in [\eta_N, 1]$,*

$$\sup_{P,Q \in \mathcal{P}_d} \left| W_p(P,Q)^p - T^{\mathrm{lin}}_{t,p}(P,Q) \right| \lesssim_{d,p} \eta_N t^{p-1}.$$

Proof. Apply Theorem 4.2 with $h_J = h_t$. ☐

Theorem 4.4 (Fixed-annulus completion criterion). *Assume $p > 1$, $d > 2p$, and let $t \in [\eta_N, 1]$. Suppose an estimator $\hat{T}_{t,p}$ satisfies*

$$\sup_{(P,Q) \in \mathcal{A}_t} \mathbb{E}\left| \hat{T}_{t,p} - T^{\mathrm{lin}}_{t,p}(P,Q) \right| \leq C_1 \eta_N t^{p-1}$$

and

$$\sup_{(P,Q) \in \mathcal{A}_t} \mathbb{E}\big(\hat{T}_{t,p} - T^{\mathrm{lin}}_{t,p}(P,Q) \big)^2 \leq C_2 \eta_N^2 t^{2p-2}.$$

Then the distance estimator $\hat{W}_{t,p} := (\hat{T}_{t,p})_+^{1/p}$ satisfies

$$\sup_{(P,Q) \in \mathcal{A}_t} \mathbb{E}|\hat{W}_{t,p} - W_p(P,Q)| \lesssim_{d,p,C_1} \eta_N$$

and

$$\sup_{(P,Q) \in \mathcal{A}_t} \mathbb{E}(\hat{W}_{t,p} - W_p(P,Q))^2 \lesssim_{d,p,C_1,C_2} \eta_N^2.$$

Proof. On $\mathcal{A}_t$, we have $W_p(P,Q) \geq t/2$. Therefore Theorem 3.1 with $r = t/2$ and Theorem 4.3 give

$$|\hat{W}_{t,p} - W_p(P,Q)| \lesssim_{d,p} t^{1-p}\Big(|\hat{T}_{t,p} - T^{\mathrm{lin}}_{t,p}(P,Q)| + \eta_N t^{p-1}\Big).$$

Taking expectations yields the absolute bound.

The same inequality squared and Jensen's inequality give

$$\mathbb{E}(\hat{W}_{t,p} - W_p(P,Q))^2 \lesssim_{d,p} t^{2-2p}\left(\mathbb{E}(\hat{T}_{t,p} - T^{\mathrm{lin}}_{t,p}(P,Q))^2 + \eta_N^2 t^{2p-2}\right),$$

which is $O(\eta_N^2)$ under the assumptions. □

4.3 Critical-grid reduction

For a dyadic mesh $h = 2^{-J}$, let $\mathcal{D}_h$ be the partition of $[0,1]^d$ into dyadic cubes of side length h, and let G_h be the set of cube centers. Define the quantizer $q_h : [0,1]^d \to G_h$ by sending each point to the center of the cube that contains it. For $P \in \mathcal{P}_d$ write

$$P^{(h)} := (q_h)_\# P \in \mathcal{P}(G_h).$$

Let

$$M^{\mathrm{abs,grid}}_{n,m,h,p} := \inf_{\widehat{W}} \sup_{\mu,\nu \in \mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W} - W_p(\mu,\nu)\Big|$$

and similarly

$$M^{\mathrm{sq,grid}}_{n,m,h,p} := \inf_{\widehat{W}} \sup_{\mu,\nu \in \mathcal{P}(G_h)} \mathbb{E}\Big(\widehat{W} - W_p(\mu,\nu)\Big)^2.$$

Theorem 4.5 (Critical-grid reduction)**.** *Let $p \geq 1$, $d \geq 1$, and let $h = 2^{-J}$. Then*

$$M^{\mathrm{abs,grid}}_{n,m,h,p} \leq M^{\mathrm{abs}}_{n,m,d,p} \leq M^{\mathrm{abs,grid}}_{n,m,h,p} + \sqrt{d}\,h. \tag{4.1}$$

Moreover,

$$M^{\mathrm{sq,grid}}_{n,m,h,p} \leq M^{\mathrm{sq}}_{n,m,d,p} \leq 2M^{\mathrm{sq,grid}}_{n,m,h,p} + 2d\,h^2. \tag{4.2}$$

Proof. The lower bounds are immediate because $\mathcal{P}(G_h) \subset \mathcal{P}_d$.

For the upper bounds, note first that every $x \in [0,1]^d$ satisfies

$$\|x - q_h(x)\|_2 \leq \frac{\sqrt{d}}{2} h.$$

Hence coupling P with $P^{(h)}$ by the map $x \mapsto q_h(x)$ gives

$$W_p\Big(P, P^{(h)}\Big) \leq \frac{\sqrt{d}}{2} h, \qquad W_p\Big(Q, Q^{(h)}\Big) \leq \frac{\sqrt{d}}{2} h.$$

By the triangle inequality,

$$\Big|W_p(P,Q) - W_p(P^{(h)}, Q^{(h)})\Big| \leq \sqrt{d}\,h. \tag{4.3}$$

Now fix an arbitrary estimator $\widehat{W}_h$ for the grid model. Given samples $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$, quantize them and apply the grid estimator to $q_h(X_1), \ldots, q_h(X_n)$ and $q_h(Y_1), \ldots, q_h(Y_m)$. These quantized observations are i.i.d. from $P^{(h)}$ and $Q^{(h)}$, so

$$\mathbb{E}\Big|\widehat{W}_h - W_p(P^{(h)}, Q^{(h)})\Big| \leq \sup_{\mu,\nu \in \mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W}_h - W_p(\mu,\nu)\Big|.$$

Combining with (4.3),

$$\mathbb{E}\Big|\widehat{W}_h - W_p(P,Q)\Big| \leq \sup_{\mu,\nu \in \mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W}_h - W_p(\mu,\nu)\Big| + \sqrt{d}\,h.$$

Taking the supremum over P, Q and then the infimum over $\widehat{W}_h$ proves (4.1).

For squared loss,

$$(\widehat{W}_h - W_p(P,Q))^2 \leq 2(\widehat{W}_h - W_p(P^{(h)}, Q^{(h)}))^2 + 2\Big(W_p(P^{(h)}, Q^{(h)}) - W_p(P,Q)\Big)^2$$

and the second term is bounded by $2dh^2$ by (4.3). Taking expectations, suprema, and infima yields (4.2). □

Corollary 4.6 (Critical-scale grid equivalence)**.** *Assume $d > 2p$ and choose a dyadic mesh $h_N = 2^{-J_N}$ such that $h_N \asymp \eta_N$. Then*

$$M^{\mathrm{abs}}_{n,m,d,p} \asymp \eta_N \quad \Longleftrightarrow \quad M^{\mathrm{abs,grid}}_{n,m,h_N,p} \asymp h_N,$$

and similarly

$$M^{\mathrm{sq}}_{n,m,d,p} \asymp \eta_N^2 \quad \Longleftrightarrow \quad M^{\mathrm{sq,grid}}_{n,m,h_N,p} \asymp h_N^2.$$

In particular, the unrestricted supercritical problem is equivalent, up to the target scale, to a finite-support OT estimation problem on a grid with

$$|G_{h_N}| = h_N^{-d} \asymp N \log N$$

support points.

Proof. Apply Theorem 4.5 with $h = h_N$ and use $h_N \asymp \eta_N$. □

4.4 One-sided known-support reduction

The two-sided grid reduction can be sharpened further. Only one marginal needs to be quantized.

Define the one-sided known-support risks

$$M^{\mathrm{abs,ks}}_{n,m,h,p} := \inf_{\widehat{W}} \sup_{P \in \mathcal{P}_d,\ R \in \mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W} - W_p(P,R)\Big|$$

and

$$M^{\mathrm{sq,ks}}_{n,m,h,p} := \inf_{\widehat{W}} \sup_{P \in \mathcal{P}_d,\ R \in \mathcal{P}(G_h)} \mathbb{E}\Big(\widehat{W} - W_p(P,R)\Big)^2.$$

Theorem 4.7 (One-sided known-support reduction). *Let $p \geq 1$, $d \geq 1$, and let $h = 2^{-J}$. Then*

$$M^{\mathrm{abs,ks}}_{n,m,h,p} \leq M^{\mathrm{abs}}_{n,m,d,p} \leq M^{\mathrm{abs,ks}}_{n,m,h,p} + \frac{\sqrt{d}}{2}h.$$

Moreover,

$$M^{\mathrm{sq,ks}}_{n,m,h,p} \leq M^{\mathrm{sq}}_{n,m,d,p} \leq 2M^{\mathrm{sq,ks}}_{n,m,h,p} + \frac{d}{2}\,h^2.$$

Proof. The lower bounds are immediate because the one-sided class

$$\{(P,R) : P \in \mathcal{P}_d,\ R \in \mathcal{P}(G_h)\}$$

is a subclass of the unrestricted class $\mathcal{P}_d^2$.

For the upper bounds, fix $P, Q \in \mathcal{P}_d$ and quantize only the second marginal:

$$Q^{(h)} := (q_h)_\# Q \in \mathcal{P}(G_h).$$

As above,

$$W_p(Q, Q^{(h)}) \leq \frac{\sqrt{d}}{2}h.$$

Therefore

$$\Big|W_p(P,Q) - W_p(P,Q^{(h)})\Big| \leq \frac{\sqrt{d}}{2}h \tag{4.4}$$

by the triangle inequality.

Now fix an arbitrary estimator $\widehat{W}^{\mathrm{ks}}_h$ for the one-sided known-support model. Given samples $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$, quantize only the second sample and apply the estimator to

$$X_1, \ldots, X_n, \qquad q_h(Y_1), \ldots, q_h(Y_m).$$

These observations are distributed exactly as a sample from $(P, Q^{(h)})$ in the one-sided model, so

$$\mathbb{E}\Big|\widehat{W}^{\mathrm{ks}}_h - W_p(P,Q^{(h)})\Big| \leq \sup_{\mu\in\mathcal{P}_d,\ \nu\in\mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W}^{\mathrm{ks}}_h - W_p(\mu,\nu)\Big|.$$

Combining with (4.4),

$$\mathbb{E}\Big|\widehat{W}^{\mathrm{ks}}_h - W_p(P,Q)\Big| \leq \sup_{\mu\in\mathcal{P}_d,\ \nu\in\mathcal{P}(G_h)} \mathbb{E}\Big|\widehat{W}^{\mathrm{ks}}_h - W_p(\mu,\nu)\Big| + \frac{\sqrt{d}}{2}h.$$

Taking the supremum over P, Q and the infimum over $\widehat{W}^{\mathrm{ks}}_h$ proves the absolute-loss statement.

For squared loss,

$$(\widehat{W}^{\mathrm{ks}}_h - W_p(P,Q))^2 \leq 2(\widehat{W}^{\mathrm{ks}}_h - W_p(P,Q^{(h)}))^2 + 2\Big(W_p(P,Q^{(h)}) - W_p(P,Q)\Big)^2,$$

and the second term is bounded by $2(\sqrt{d}\,h/2)^2 = dh^2/2$ by (4.4). Again taking expectations, suprema, and infima yields the result. □

Corollary 4.8 (Critical-scale known-support equivalence). *Assume $d > 2p$ and choose a dyadic mesh $h_N = 2^{-J_N}$ such that $h_N \asymp \eta_N$. Then*

$$M^{\mathrm{abs}}_{n,m,d,p} \asymp \eta_N \iff M^{\mathrm{abs,ks}}_{n,m,h_N,p} \asymp h_N,$$

and similarly

$$M^{\mathrm{sq}}_{n,m,d,p} \asymp \eta_N^2 \iff M^{\mathrm{sq,ks}}_{n,m,h_N,p} \asymp h_N^2.$$

Proof. Apply Theorem 4.7 with $h = h_N$ and use $h_N \asymp \eta_N$. □

Proposition 4.9 (Critical known-support hardness). *Assume $d > 2p$ and $n = N$. Choose a dyadic mesh $h_N = 2^{-J_N}$ with $h_N \asymp \eta_N$, and let U_{h_N} denote the uniform distribution on G_{h_N}. Then there exist constants $A_*, c_* > 0$, depending only on (d,p), such that*

$$\inf_{\widehat{W}} \sup_{P:\ W_p(P,U_{h_N})\leq A_*\eta_N} \mathbb{E}\Big|\widehat{W} - W_p(P,U_{h_N})\Big| \geq c_*\eta_N.$$

Consequently,

$$\inf_{\widehat{W}} \sup_{P:\ W_p(P,U_{h_N})\leq A_*\eta_N} \mathbb{E}\Big(\widehat{W} - W_p(P,U_{h_N})\Big)^2 \gtrsim_{d,p} \eta_N^2.$$

Proof. Let

$$S_N := |G_{h_N}|.$$

Since $h_N \asymp (N \log N)^{-1/d}$, we have

$$S_N \asymp h_N^{-d} \asymp N \log N.$$

Choose a bijection

$$F : [S_N] \to G_{h_N},$$

and let u denote the uniform distribution on $[S_N]$, so that

$$F_\# u = U_{h_N}.$$

Because $G_{h_N} \subset [0,1]^d$ has diameter $O_d(1)$ and pairwise separation $\gtrsim_d h_N \asymp S_N^{-1/d}$, Proposition 9 of Niles-Weed–Rigollet [10] applies to this metric set. Combined with Proposition 10 and the proof of their Theorem 11, it yields a family of distributions q on $[S_N]$ with the following properties:

(i) the induced alternatives

$$P_q := F_\# q$$

satisfy

$$W_p(P_q, U_{h_N}) \leq A_* h_N$$

for some constant $A_* = A_*(d,p)$;

(ii) for every estimator $\widehat{W}$ based on N samples from the unknown law,

$$\sup_q \mathbb{E}\Big|\widehat{W} - W_p(P_q, U_{h_N})\Big| \geq c_* h_N$$

for some constant $c_* = c_*(d,p) > 0$.
The only point to note is that the Niles-Weed–Rigollet argument averages over the random bijection F from Proposition 9; therefore some deterministic realization of F must satisfy the same lower bound up to absolute constants, and we fix such an F once and for all.

Since $h_N \asymp \eta_N$, the family $\{P_q\}$ lies inside

$$\{P: \ W_p(P, U_{h_N}) \le A_* \eta_N\},$$

and hence

$$\inf_{\widehat{W}} \sup_{P:\ W_p(P,U_{h_N}) \le A_*\eta_N} \mathbb{E}\left|\widehat{W} - W_p(P, U_{h_N})\right| \ge c_* \eta_N.$$

This proves the absolute-risk statement.

The squared-risk bound follows from Jensen's inequality: for every estimator and every P,

$$\mathbb{E}\left(\widehat{W} - W_p(P, U_{h_N})\right)^2 \ge \left(\mathbb{E}\left|\widehat{W} - W_p(P, U_{h_N})\right|\right)^2.$$

□

5 Large-alphabet skeletons and exact Euclidean critical classes

5.1 A general exact TV-skeleton principle

Many of the exact positive classes below share a common hidden mechanism: after a suitable geometric reduction, the transport cost becomes an exact weighted sum of finitely many discrete total-variation distances. The next theorem isolates this mechanism once and for all. It is the common abstract engine behind the paired, multiband, laminar, tree, and semi-discrete models developed later in the paper.

Theorem 5.1 (Exact TV-skeleton principle). *Fix $\alpha \in (0,1)$ and $C_0 \ge 1$. Let $\mathcal{C} \subseteq \mathcal{P}([0,1]^d)^2$, and assume that for each $N = n \wedge m$ the following data are given: an integer $L_N \ge 1$; for each $1 \le \ell \le L_N$, an alphabet size $M_{\ell,N} \ge 2$, measurable maps*

$$\pi_{\ell,N}, \sigma_{\ell,N} : [0,1]^d \to \{1, \dots, M_{\ell,N}\},$$

and a weight $\lambda_{\ell,N} > 0$. Assume that

$$M_{\ell,N} \le C_0\, N \log N \qquad (1 \le \ell \le L_N),$$

and that for every $(P,Q) \in \mathcal{C}$,

$$T_N(P,Q) = \sum_{\ell=1}^{L_N} \lambda_{\ell,N}\ \mathrm{TV}\big((\pi_{\ell,N})_\# P, (\sigma_{\ell,N})_\# Q\big).$$

Define

$$\varepsilon_{\ell,N} := \begin{cases} \sqrt{\dfrac{M_{\ell,N}}{N \log N}}, & M_{\ell,N} \ge N^\alpha, \\ \sqrt{\dfrac{M_{\ell,N}}{N}}, & M_{\ell,N} < N^\alpha, \end{cases} \qquad \Delta_N := \sum_{\ell=1}^{L_N} \lambda_{\ell,N}\, \varepsilon_{\ell,N}.$$

Then there exists an estimator $\widehat{T}_N$ such that

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\left|\widehat{T}_N - T_N(P,Q)\right| \le C_{\alpha,C_0}\, \Delta_N,$$

and

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\left(\widehat{T}_N - T_N(P,Q)\right)^2 \le C_{\alpha,C_0}\, \Delta_N^2.$$

Proof. Fix ℓ. Discarding surplus observations if necessary, we may work with the first $N = n \wedge m$ samples from each side. For $X_1, \dots, X_N \sim P$ and $Y_1, \dots, Y_N \sim Q$, the pushed samples

$$\pi_{\ell,N}(X_1), \dots, \pi_{\ell,N}(X_N), \qquad \sigma_{\ell,N}(Y_1), \dots, \sigma_{\ell,N}(Y_N)$$

are i.i.d. from the discrete laws

$$R_{\ell,N}(P) := (\pi_{\ell,N})_\# P, \qquad S_{\ell,N}(Q) := (\sigma_{\ell,N})_\# Q$$

on the known alphabet $\{1, \dots, M_{\ell,N}\}$.

If $M_{\ell,N} \ge N^\alpha$, then

$$\log N \le \alpha^{-1} \log M_{\ell,N},$$

while by assumption

$$M_{\ell,N} \le C_0 N \log N.$$

Hence the large-alphabet regime of Jiao–Han–Weissman [5, Theorem 5] applies to the two-sample L_1-distance problem on this known alphabet (after the standard Poissonization/de-Poissonization transfer recorded there), and yields an estimator $\widehat{L}_{\ell,N}$ satisfying

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\left(\widehat{L}_{\ell,N} - \|R_{\ell,N}(P) - S_{\ell,N}(Q)\|_1\right)^2 \le C_{\alpha,C_0} \frac{M_{\ell,N}}{N \log N}.$$

If $M_{\ell,N} < N^\alpha$, we use the empirical plug-in estimator

$$\widehat{L}_{\ell,N}^{\mathrm{emp}} := \left\|R_{\ell,N}^{\mathrm{emp}} - S_{\ell,N}^{\mathrm{emp}}\right\|_1,$$

where $R_{\ell,N}^{\mathrm{emp}}$ and $S_{\ell,N}^{\mathrm{emp}}$ are the empirical laws of the pushed samples. Since

$$\left|\widehat{L}_{\ell,N}^{\mathrm{emp}} - \|R_{\ell,N}(P) - S_{\ell,N}(Q)\|_1\right| \le \|R_{\ell,N}^{\mathrm{emp}} - R_{\ell,N}(P)\|_1 + \|S_{\ell,N}^{\mathrm{emp}} - S_{\ell,N}(Q)\|_1,$$

we obtain

$$\mathbb{E}\left(\widehat{L}_{\ell,N}^{\mathrm{emp}} - \|R_{\ell,N}(P) - S_{\ell,N}(Q)\|_1\right)^2 \le 2\mathbb{E}\|R_{\ell,N}^{\mathrm{emp}} - R_{\ell,N}(P)\|_1^2 + 2\mathbb{E}\|S_{\ell,N}^{\mathrm{emp}} - S_{\ell,N}(Q)\|_1^2.$$

For any empirical law $\widehat{r}$ on an alphabet of size M based on N samples from r,

$$\|\widehat{r}-r\|_1^2 \le M\sum_{i=1}^{M}(\widehat{r}_i-r_i)^2,$$

hence

$$\mathbb{E}\|\widehat{r}-r\|_1^2 \le \frac{M}{N}\sum_{i=1}^{M} r_i(1-r_i) \le \frac{M}{N}.$$

Therefore

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\Big(\widehat{L}^{\mathrm{emp}}_{\ell,N} - \|R_{\ell,N}(P)-S_{\ell,N}(Q)\|_1\Big)^2 \le 4\,\frac{M_{\ell,N}}{N}.$$

Combining the two regimes, there exists an estimator $\widehat{L}^\star_{\ell,N}$ such that

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\Big(\widehat{L}^\star_{\ell,N} - \|R_{\ell,N}(P)-S_{\ell,N}(Q)\|_1\Big)^2 \le C_{\alpha,C_0}\,\varepsilon^2_{\ell,N}.$$

Clip to the natural range by

$$\widetilde{L}_{\ell,N} := 0\vee \widehat{L}^\star_{\ell,N}\wedge 2.$$

Clipping cannot increase squared error. Set

$$\widehat{D}_{\ell,N} := \frac{\widetilde{L}_{\ell,N}}{2}, \qquad D_{\ell,N}(P,Q) := \mathrm{TV}\big(R_{\ell,N}(P), S_{\ell,N}(Q)\big)\in[0,1].$$

Then

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\Big(\widehat{D}_{\ell,N} - D_{\ell,N}(P,Q)\Big)^2 \le C_{\alpha,C_0}\,\varepsilon^2_{\ell,N}.$$

Define

$$\widehat{T}_N := \sum_{\ell=1}^{L_N}\lambda_{\ell,N}\,\widehat{D}_{\ell,N}.$$

Since

$$T_N(P,Q) = \sum_{\ell=1}^{L_N}\lambda_{\ell,N}\,D_{\ell,N}(P,Q),$$

Minkowski's inequality in L^2 yields

$$\Big(\mathbb{E}\big(\widehat{T}_N - T_N(P,Q)\big)^2\Big)^{1/2} \le \sum_{\ell=1}^{L_N}\lambda_{\ell,N}\Big(\mathbb{E}\big(\widehat{D}_{\ell,N}-D_{\ell,N}(P,Q)\big)^2\Big)^{1/2} \le C_{\alpha,C_0}\,\Delta_N.$$

This proves the squared-risk bound. The absolute-risk bound follows from Cauchy–Schwarz. □

Corollary 5.2 (Rooted exact TV-skeleton principle)**.** *Fix $p\ge 1$, $\alpha\in(0,1)$, and constants $B, C_0\ge 1$. Assume that, in the setting of Theorem 5.1, there exists a scale $h_N\in(0,1]$ such that*

$$W_p(P,Q)^p = h_N^p\,T_N(P,Q), \qquad 0\le T_N(P,Q)\le B \qquad \forall\,(P,Q)\in\mathcal{C}.$$

Then there exists an estimator $\widehat{W}_N$ such that

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\big|\widehat{W}_N - W_p(P,Q)\big| \le C_{\alpha,C_0,B,p}\,h_N\,\Delta_N^{1/p},$$

and

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\big(\widehat{W}_N - W_p(P,Q)\big)^2 \le C_{\alpha,C_0,B,p}\,h_N^2\,\Delta_N^{2/(p\vee 2)}.$$

Proof. Let $\widehat{T}_N$ be the estimator from Theorem 5.1, and clip once more by

$$\widetilde{T}_N := 0\vee\widehat{T}_N\wedge B.$$

Clipping cannot increase squared error because $T_N(P,Q)\in[0,B]$. Define

$$\widehat{W}_N := h_N\,\widetilde{T}_N^{1/p}.$$

Since $x\mapsto x^{1/p}$ is $1/p$-Hölder on $[0,\infty)$,

$$|\widetilde{T}_N^{1/p} - T_N(P,Q)^{1/p}| \le |\widetilde{T}_N - T_N(P,Q)|^{1/p}.$$

Therefore

$$\mathbb{E}\big|\widehat{W}_N - W_p(P,Q)\big| \le h_N\,\mathbb{E}|\widetilde{T}_N - T_N(P,Q)|^{1/p} \le h_N\Big(\mathbb{E}|\widetilde{T}_N - T_N(P,Q)|^2\Big)^{1/(2p)} \lesssim h_N\,\Delta_N^{1/p}.$$

If $p\ge 2$, then $2/p\le 1$, so

$$\mathbb{E}\big(\widehat{W}_N - W_p(P,Q)\big)^2 \le h_N^2\,\mathbb{E}|\widetilde{T}_N - T_N(P,Q)|^{2/p} \le h_N^2\Big(\mathbb{E}|\widetilde{T}_N - T_N(P,Q)|^2\Big)^{1/p} \lesssim h_N^2\,\Delta_N^{2/p}.$$

If $1\le p<2$, write

$$u := \frac{\widetilde{T}_N}{B}, \qquad v := \frac{T_N(P,Q)}{B},$$

so that $u,v\in[0,1]$. Because $2/p>1$ and $|u-v|\le 1$,

$$|u^{1/p} - v^{1/p}|^2 \le |u-v|^{2/p} \le |u-v|.$$

Hence

$$|\widetilde{T}_N^{1/p} - T_N(P,Q)^{1/p}|^2 = B^{2/p}|u^{1/p} - v^{1/p}|^2 \le B^{2/p-1}|\widetilde{T}_N - T_N(P,Q)|.$$

Therefore

$$\mathbb{E}\Big(\widehat{W}_N - W_p(P,Q)\Big)^2 \le h_N^2 B^{2/p-1}\,\mathbb{E}|\widetilde{T}_N - T_N(P,Q)| \lesssim h_N^2\,\Delta_N.$$

Combining the two cases gives the claim. □

Corollary 5.3 (Annular rooted exact TV-skeleton principle)**.** *Fix $p \ge 1$, $\alpha \in (0,1)$, and constants $0 < \tau_- < B$, $C_0 \ge 1$. In the setting of Theorem 5.2, assume moreover that*

$$\tau_- \le T_N(P,Q) \le B \qquad \forall\,(P,Q) \in \mathcal{C}.$$

Then there exists an estimator $\widehat{W}_N^{\mathrm{ann}}$ such that

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\Big|\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\Big| \le C_{\alpha,C_0,\tau_-,B,p}\,h_N\,\Delta_N,$$

and

$$\sup_{(P,Q)\in\mathcal{C}} \mathbb{E}\Big(\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\Big)^2 \le C_{\alpha,C_0,\tau_-,B,p}\,h_N^2\,\Delta_N^2.$$

Proof. Let $\widehat{T}_N$ be the estimator from Theorem 5.1, and clip it to the interval

$$\widetilde{T}_N := \frac{\tau_-}{2} \vee \widehat{T}_N \wedge B.$$

Because $T_N(P,Q) \in [\tau_-, B]$, this clipping cannot increase either the absolute or the squared loss. Define

$$\widehat{W}_N^{\mathrm{ann}} := h_N\,\widetilde{T}_N^{1/p}.$$

On the interval $[\tau_-/2, B]$ the map $x \mapsto x^{1/p}$ is Lipschitz with constant

$$L_{\tau_-,p} := \begin{cases} 1, & p = 1,\\ \dfrac{1}{p}\Big(\dfrac{\tau_-}{2}\Big)^{1/p-1}, & p > 1. \end{cases}$$

Hence

$$\Big|\widetilde{T}_N^{1/p} - T_N(P,Q)^{1/p}\Big| \le L_{\tau_-,p}\,|\widetilde{T}_N - T_N(P,Q)|.$$

Therefore

$$\mathbb{E}\Big|\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\Big| \le h_N L_{\tau_-,p}\,\mathbb{E}|\widetilde{T}_N - T_N(P,Q)| \lesssim_{\tau_-,p} h_N\,\Delta_N,$$

and likewise

$$\mathbb{E}\Big(\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\Big)^2 \le h_N^2 L_{\tau_-,p}^2\,\mathbb{E}(\widetilde{T}_N - T_N(P,Q))^2 \lesssim_{\tau_-,p} h_N^2\,\Delta_N^2.$$

The conclusion follows from Theorem 5.1. □

Corollary 5.4 (Critical exact-skeleton envelope)**.** *Fix $p \ge 1$, $\alpha \in (0,1)$, and constants $B, C_0, c_0, C_1 > 0$. Let $\mathcal{C}_N \subseteq \mathcal{P}([0,1]^d)^2$ be a sequence of classes with scales $h_N \downarrow 0$. Assume that for every N the class $\mathcal{C}_N$ satisfies the hypotheses of Theorem 5.2, that*

$$\Delta_N \le C_1,$$

and that there exists a distinguished level $\ell_\star(N)$ such that

$$M_{\ell_\star(N),N} \asymp N\log N, \qquad \lambda_{\ell_\star(N),N} \asymp 1.$$

Assume moreover that $\mathcal{C}_N$ contains an embedded one-level subclass $\mathcal{C}_N^\star$ on which all levels except $\ell_\star(N)$ are frozen and

$$T_N(P,Q) = \lambda_{\ell_\star(N),N}\ \mathrm{TV}(r,s)$$

after identifying the distinguished level with an arbitrary pair (r,s) of distributions on the alphabet

$$\{1,\dots,M_{\ell_\star(N),N}\}.$$

Then

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_N} \mathbb{E}\Big|\widehat{W} - W_p(P,Q)\Big| \asymp h_N,$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_N} \mathbb{E}\Big(\widehat{W} - W_p(P,Q)\Big)^2 \asymp h_N^2.$$

Proof. The upper bounds follow immediately from Theorem 5.2.

For the lower bounds, restrict to the subclass $\mathcal{C}_N^\star$. By the minimax lower bound of Jiao–Han–Weissman [5, Theorem 3 and Theorem 5] for two-sample L_1-distance estimation on an alphabet of size $\asymp N\log N$,

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_N^\star} \mathbb{E}\Big(\widehat{T} - T_N(P,Q)\Big)^2 \gtrsim 1.$$

Since $0 \le T_N(P,Q) \le B$ on $\mathcal{C}_N^\star$, clipping any estimator to $[0,B]$ cannot increase squared loss, and then

$$\Big|\widehat{T} - T_N(P,Q)\Big|^2 \le B\,\Big|\widehat{T} - T_N(P,Q)\Big|.$$

Hence

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_N^\star} \mathbb{E}\Big|\widehat{T} - T_N(P,Q)\Big| \gtrsim_B 1.$$

Now let $\widehat{W}$ be an arbitrary estimator of $W_p(P,Q)$ on $\mathcal{C}_N^\star$, and clip it to the range

$$0 \le \widehat{W} \le h_N B^{1/p},$$

which again cannot increase squared error. Define the induced estimator of T_N by

$$\widehat{T} := \left(\frac{\widehat{W}}{h_N}\right)^p.$$

Because $x \mapsto x^p$ is Lipschitz on $[0, B^{1/p}]$ with constant $pB^{(p-1)/p}$,

$$|\widehat{T} - T_N(P,Q)| \le pB^{(p-1)/p}\,\frac{|\widehat{W} - W_p(P,Q)|}{h_N}.$$

Taking expectations and the supremum over $\mathcal{C}_N^\star$ therefore gives

$$\sup_{(P,Q)\in\mathcal{C}_N^\star} \mathbb{E}\left|\widehat{W} - W_p(P,Q)\right| \gtrsim_{B,p} h_N.$$

This proves the absolute-risk lower bound. The squared-risk lower bound follows from Jensen's inequality. □

Remark 5.5 (Scope of the skeleton principle). The geometric burden in the explicit models below is to prove an exact identity of the form required by Theorem 5.1 and then to verify the critical envelope $\Delta_N \lesssim 1$. Once this is done, the target-scale upper law is automatic. The paired class, finite-band class, packed direct sums, dyadic pair-isolation class, nested and critical laminar hierarchies, hierarchical trees, and the semi-discrete support-box model all fit into this scheme. What differs from model to model is only the geometry of the exact reduction, not the statistics of the reduced functional.

5.2 Exact one-level classes beyond the critical point

The critical envelope from Theorem 5.4 is only the first consequence of the exact reduction. On one-level exact classes the reduced functional is itself a large-alphabet discrete TV problem, and this makes it possible to solve the whole noncritical large-alphabet scale exactly for the powered functional W_p^p. For W_p itself one still pays for the singular root at the origin, but away from zero the map $t \mapsto t^{1/p}$ is bi-Lipschitz and the same linear large-alphabet law reappears.

Lemma 5.6 (Large-alphabet lower bound for discrete TV)**.** *Fix $\alpha \in (0,1)$ and $C_0 \ge 1$, and write $[M] := \{1,\dots,M\}$. Whenever*

$$N^\alpha \le M \le C_0 N \log N,$$

one has

$$\inf_{\widehat{D}} \sup_{r,s\in\mathcal{P}([M])} \mathbb{E}\left|\widehat{D} - \mathrm{TV}(r,s)\right| \gtrsim_{\alpha,C_0} \sqrt{\frac{M}{N\log N}},$$

and

$$\inf_{\widehat{D}} \sup_{r,s\in\mathcal{P}([M])} \mathbb{E}\left(\widehat{D} - \mathrm{TV}(r,s)\right)^2 \gtrsim_{\alpha,C_0} \frac{M}{N\log N},$$

where $\widehat{D}$ ranges over all estimators based on N samples from r and N samples from s.

Proof. Let u_M be the uniform law on $[M]$. Since the full two-sample problem contains the known-baseline subclass

$$\{(r,u_M) : r \in \mathcal{P}([M])\},$$

it suffices to prove the lower bounds there.

The regime assumptions imply

$$\alpha\log N \le \log M \le \log N + \log(C_0\log N),$$

hence

$$\log M \asymp_{\alpha,C_0} \log N \qquad \text{and} \qquad N \gtrsim_{\alpha,C_0} \frac{M}{\log M}.$$

Therefore the large-alphabet regime in Jiao–Han–Weissman [5, Theorem 4] applies to estimation of $\|r-u_M\|_1$ from N samples of r, and yields

$$\inf_{\widehat{L}} \sup_{r\in\mathcal{P}([M])} \mathbb{E}\left(\widehat{L} - \|r-u_M\|_1\right)^2 \gtrsim_{\alpha,C_0} \frac{M}{N\log N}.$$

Since $\mathrm{TV}(r,u_M) = \frac{1}{2}\|r-u_M\|_1$, this gives the displayed squared lower bound.

For the absolute lower bound we use the known-Q lower-bound construction in the proof of [5, Theorem 4], again with $Q = u_M$. That argument constructs two priors μ_0,μ_1 on $\mathcal{P}([M])$ and a quantity

$$\chi \asymp_{\alpha,C_0} \sqrt{\frac{M}{N\log N}}$$

with the following properties. Writing

$$D(r) := \mathrm{TV}(r,u_M), \qquad D_i := \mathbb{E}_{\mu_i} D(r),$$

one has

$$D_1 - D_0 = \frac{\chi}{2}.$$

Moreover, after conditioning μ_i onto events on which

$$|D(r) - D_i| \le \frac{\chi}{8},$$

the corresponding mixture laws F_0, F_1 of the N-sample experiment satisfy

$$\mathrm{TV}(F_0,F_1) = o(1).$$

Let

$$\zeta := D_0 + \frac{\chi}{4}$$

and consider the threshold test

$$\psi := \mathbf{1}_{\{\widehat{D} > \zeta\}}.$$

Under the conditioned prior F_0 the parameter satisfies $D(r) \le \zeta - \chi/8$, hence

$$\left|\widehat{D} - D(r)\right| \ge \frac{\chi}{8}\,\psi.$$

Under F_1 one has $D(r) \ge \zeta + \chi/8$, hence

$$\left|\widehat{D} - D(r)\right| \ge \frac{\chi}{8}\,(1-\psi).$$

Averaging over the two conditioned priors gives

$$\sup_{r\in\mathcal{P}([M])} \mathbb{E}\left|\widehat{D} - D(r)\right| \ge \frac{\chi}{16}\Big(F_0(\psi=1) + F_1(\psi=0)\Big).$$

By the standard testing inequality,

$$\inf_{\psi}\Big(F_0(\psi=1) + F_1(\psi=0)\Big) \ge 1 - \mathrm{TV}(F_0, F_1).$$

Therefore, for all sufficiently large N,

$$\sup_{r\in\mathcal{P}([M])} \mathbb{E}\left|\widehat{D} - D(r)\right| \gtrsim_{\alpha,C_0} \chi \asymp_{\alpha,C_0} \sqrt{\frac{M}{N\log N}}.$$

This proves the absolute lower bound on the known-baseline subclass and therefore on the full two-sample problem. □

Lemma 5.7 (Large-alphabet annular lower bound for discrete TV)**.** *Fix $\alpha \in (0,1)$, $C_0 \ge 1$, and numbers $0 < \tau_- < \tau_+ < 1$. Whenever*

$$N^\alpha \le M \le C_0 N \log N,$$

one has

$$\inf_{\widehat{D}} \sup_{\substack{r,s\in\mathcal{P}([M]) \\ \tau_- \le \mathrm{TV}(r,s) \le \tau_+}} \mathbb{E}\left|\widehat{D} - \mathrm{TV}(r,s)\right| \gtrsim_{\alpha,C_0,\tau_-,\tau_+} \sqrt{\frac{M}{N\log N}},$$

and

$$\inf_{\widehat{D}} \sup_{\substack{r,s\in\mathcal{P}([M]) \\ \tau_- \le \mathrm{TV}(r,s) \le \tau_+}} \mathbb{E}\Big(\widehat{D} - \mathrm{TV}(r,s)\Big)^2 \gtrsim_{\alpha,C_0,\tau_-,\tau_+} \frac{M}{N\log N}.$$

Proof. Write

$$\tau_0 := \frac{\tau_- + \tau_+}{2}, \qquad \theta := \min\left\{\frac{\tau_+ - \tau_-}{4},\ \frac{1-\tau_+}{2}\right\} > 0.$$

Then

$$[\tau_0, \tau_0 + \theta] \subset [\tau_-, \tau_+]$$

and

$$a := \frac{\tau_0}{1-\theta} \in (0,1).$$

Choose two probability vectors

$$\rho := \left(\frac{1+a}{2}, \frac{1-a}{2}\right), \qquad \sigma := \left(\frac{1-a}{2}, \frac{1+a}{2}\right)$$

on the two-point alphabet $\{0,1\}$, so that $\mathrm{TV}(\rho,\sigma) = a$.

Let $M' := M - 2$. For all sufficiently large N, one has $M' \asymp M$, $M' \ge \frac{1}{2}N^\alpha$, and $M' \le C_0 N \log N$. Fix arbitrary

$$r, s \in \mathcal{P}([M']) = \mathcal{P}(\{1,\dots,M'\}),$$

and define lifted laws on the disjoint union

$$\{0,1\} \sqcup [M'] \cong [M]$$

by

$$\bar{r} := (1-\theta)\rho \oplus \theta r, \qquad \bar{s} := (1-\theta)\sigma \oplus \theta s.$$

Because the background and hard parts live on disjoint supports,

$$\mathrm{TV}(\bar{r},\bar{s}) = (1-\theta)\,\mathrm{TV}(\rho,\sigma) + \theta\,\mathrm{TV}(r,s) = \tau_0 + \theta\,\mathrm{TV}(r,s) \in [\tau_0, \tau_0+\theta] \subset [\tau_-, \tau_+].$$

Given N i.i.d. samples from (r,s), one can generate N i.i.d. samples from $(\bar{r},\bar{s})$ by independent Bernoulli thinning: with probability $1-\theta$ one draws from the fixed background laws (ρ,σ), and with probability θ one uses the next raw sample from (r,s). Hence any estimator $\widehat{A}$ for the annular problem on the M-point alphabet yields an estimator

$$\widehat{D} := \frac{\left(\widehat{A} \vee \tau_0\right) \wedge (\tau_0 + \theta) - \tau_0}{\theta}$$

of $\mathrm{TV}(r,s)$ on the M'-point alphabet. Since $\mathrm{TV}(\bar{r},\bar{s}) = \tau_0 + \theta\,\mathrm{TV}(r,s)$ and clipping can only reduce the distance to this interval,

$$\left|\widehat{D} - \mathrm{TV}(r,s)\right| \le \theta^{-1}\left|\widehat{A} - \mathrm{TV}(\bar{r},\bar{s})\right|.$$

Therefore

$$\sup_{r,s\in\mathcal{P}([M'])} \mathbb{E}\left|\widehat{D} - \mathrm{TV}(r,s)\right| \le \theta^{-1} \sup_{\substack{\mu,\nu\in\mathcal{P}([M]) \\ \tau_- \le \mathrm{TV}(\mu,\nu) \le \tau_+}} \mathbb{E}\left|\widehat{A} - \mathrm{TV}(\mu,\nu)\right|.$$

Taking the infimum over $\widehat{A}$ and applying Theorem 5.6 on the M'-point alphabet gives

$$\inf_{\widehat{A}} \sup_{\substack{\mu,\nu\in\mathcal{P}([M]) \\ \tau_- \le \mathrm{TV}(\mu,\nu) \le \tau_+}} \mathbb{E}\left|\widehat{A} - \mathrm{TV}(\mu,\nu)\right| \gtrsim_{\alpha,C_0,\tau_-,\tau_+} \sqrt{\frac{M'}{N\log N}} \asymp \sqrt{\frac{M}{N\log N}}.$$

The squared lower bound follows from Jensen's inequality. □

Proposition 5.8 (Exact large-alphabet laws for one-level exact classes)**.** *Fix $p \ge 1$, $\alpha \in (0,1)$, $C_0 \ge 1$, constants $0 < c_- \le c_+ < \infty$, and numbers $0 < \tau_- < \tau_+ < 1$. For each N, let $h_N \in (0,1]$, let $M_N \ge 2$, let $\lambda_N \in [c_-, c_+]$, and let $\mathcal{E}_N \subseteq \mathcal{P}([0,1]^d)^2$. Assume that*

$$N^\alpha \le M_N \le C_0 N \log N,$$

and that there exist measurable maps

$$\pi_N, \sigma_N : [0,1]^d \to [M_N]$$

such that

$$W_p(P,Q)^p = h_N^p \lambda_N \ \mathrm{TV}\big((\pi_N)_\# P, (\sigma_N)_\# Q\big) \qquad \forall (P,Q) \in \mathcal{E}_N,$$

and, conversely, for every pair

$$r, s \in \mathcal{P}([M_N])$$

there exists $(P_{r,s}, Q_{r,s}) \in \mathcal{E}_N$ with

$$(\pi_N)_\# P_{r,s} = r, \qquad (\sigma_N)_\# Q_{r,s} = s.$$

Then

$$\inf_{\widehat{U}} \sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big|\widehat{U} - W_p(P,Q)^p\big| \asymp_{p,\alpha,C_0,c_-,c_+} h_N^p \sqrt{\frac{M_N}{N \log N}},$$

and

$$\inf_{\widehat{U}} \sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big(\widehat{U} - W_p(P,Q)^p\big)^2 \asymp_{p,\alpha,C_0,c_-,c_+} h_N^{2p} \frac{M_N}{N \log N}.$$

If

$$\mathcal{E}_N^{\mathrm{ann}}(\tau_-, \tau_+) := \Big\{(P,Q) \in \mathcal{E}_N : \tau_- \le \mathrm{TV}\big((\pi_N)_\# P, (\sigma_N)_\# Q\big) \le \tau_+\Big\},$$

then also

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{E}_N^{\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \asymp_{p,\alpha,C_0,c_-,c_+,\tau_-,\tau_+} h_N \sqrt{\frac{M_N}{N \log N}},$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{E}_N^{\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\big(\widehat{W} - W_p(P,Q)\big)^2 \asymp_{p,\alpha,C_0,c_-,c_+,\tau_-,\tau_+} h_N^2 \frac{M_N}{N \log N}.$$

Proof. Set

$$T_N(P,Q) := \lambda_N \ \mathrm{TV}\big((\pi_N)_\# P, (\sigma_N)_\# Q\big).$$

Since $L_N = 1$ and

$$\Delta_N = \lambda_N \sqrt{\frac{M_N}{N \log N}},$$

Theorem 5.1 yields an estimator $\widehat{T}_N$ with

$$\sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big|\widehat{T}_N - T_N(P,Q)\big| \lesssim_{\alpha,C_0,c_+} \sqrt{\frac{M_N}{N \log N}},$$

and

$$\sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big(\widehat{T}_N - T_N(P,Q)\big)^2 \lesssim_{\alpha,C_0,c_+} \frac{M_N}{N \log N}.$$

Therefore

$$\widehat{U}_N := h_N^p \widehat{T}_N$$

satisfies

$$\sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big|\widehat{U}_N - W_p(P,Q)^p\big| \lesssim h_N^p \sqrt{\frac{M_N}{N \log N}},$$

and

$$\sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\big(\widehat{U}_N - W_p(P,Q)^p\big)^2 \lesssim h_N^{2p} \frac{M_N}{N \log N}.$$

On the annular subclass one has

$$\lambda_N \tau_- \le T_N(P,Q) \le \lambda_N \tau_+,$$

so Theorem 5.3 gives an estimator $\widehat{W}_N^{\mathrm{ann}}$ with

$$\sup_{(P,Q)\in\mathcal{E}_N^{\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\big|\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\big| \lesssim h_N \sqrt{\frac{M_N}{N \log N}},$$

and

$$\sup_{(P,Q)\in\mathcal{E}_N^{\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\big(\widehat{W}_N^{\mathrm{ann}} - W_p(P,Q)\big)^2 \lesssim h_N^2 \frac{M_N}{N \log N}.$$

For the lower bound for W_p^p, let $\widehat{U}$ be any estimator on $\mathcal{E}_N$. Restrict to the realized subclass

$$\{(P_{r,s}, Q_{r,s}) : r, s \in \mathcal{P}([M_N])\},$$

and define

$$\widehat{D} := \frac{\widehat{U}}{h_N^p \lambda_N}.$$

Then

$$\big|\widehat{D} - \mathrm{TV}(r,s)\big| = \frac{1}{h_N^p \lambda_N} \big|\widehat{U} - W_p(P_{r,s}, Q_{r,s})^p\big|.$$

Taking expectations and suprema over r, s and invoking Theorem 5.6 yields

$$\sup_{(P,Q)\in\mathcal{E}_N} \mathbb{E}\left|\widehat{U} - W_p(P,Q)^p\right| \gtrsim_{p,\alpha,C_0,c_-,c_+} h_N^p \sqrt{\frac{M_N}{N\log N}}.$$

The squared lower bound follows from Jensen's inequality.

For the annular lower bound, define

$$g_N(t) := h_N(\lambda_N t)^{1/p}, \qquad t \in [\tau_-, \tau_+].$$

Since $\lambda_N \in [c_-, c_+]$ and $t \in [\tau_-, \tau_+]$, both g_N and g_N^{-1} are Lipschitz on the relevant intervals, with

$$|g_N^{-1}(u) - g_N^{-1}(v)| \le C_{p,c_-,c_+,\tau_-,\tau_+} h_N^{-1}|u - v|.$$

Let $\widehat{W}$ be any estimator on $\mathcal{E}_N^{\mathrm{ann}}(\tau_-, \tau_+)$, clip it to the range

$$h_N(\lambda_N\tau_-)^{1/p} \le \widetilde{W} \le h_N(\lambda_N\tau_+)^{1/p},$$

and define

$$\widehat{D}^{\mathrm{ann}} := g_N^{-1}(\widetilde{W}).$$

Restricting again to the realized annular subclass

$$\{(P_{r,s}, Q_{r,s}) : \tau_- \le \mathrm{TV}(r,s) \le \tau_+\} \subseteq \mathcal{E}_N^{\mathrm{ann}}(\tau_-, \tau_+),$$

one obtains

$$\left|\widehat{D}^{\mathrm{ann}} - \mathrm{TV}(r,s)\right| \le C_{p,c_-,c_+,\tau_-,\tau_+} h_N^{-1}\left|\widehat{W} - W_p(P_{r,s}, Q_{r,s})\right|.$$

Applying Theorem 5.7 gives

$$\sup_{(P,Q)\in\mathcal{E}_N^{\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\left|\widehat{W} - W_p(P,Q)\right| \gtrsim_{p,\alpha,C_0,c_-,c_+,\tau_-,\tau_+} h_N\sqrt{\frac{M_N}{N\log N}}.$$

The squared lower bound again follows from Jensen's inequality. □

Remark 5.9 (The remaining noncritical issue on one-level exact classes). Theorem 5.8 solves the large-alphabet statistical problem completely for the powered functional W_p^p on every one-level exact model, and also for W_p itself on every fixed annulus away from zero. Thus, on such exact classes, the only remaining noncritical loss for W_p is the singular behavior of the map $t \mapsto t^{1/p}$ at $t = 0$. In particular, when $p = 1$ the annular restriction disappears and the full minimax law is

$$h_N\sqrt{\frac{M_N}{N\log N}}.$$

5.3 A paired-grid support

Lemma 5.10 (Separated paired-grid support)**.** *There exists a constant $c_d > 0$ such that for every $h \in (0, 1/12]$ one can find*

$$M_h \ge c_d h^{-d}$$

pairs of points

$$(a_i, b_i)_{1\le i\le M_h} \subset [0,1]^d$$

with the following properties:

(i) $\|a_i - b_i\|_2 = h$ *for every* i*;*

(ii) *for every* $i \ne j$ *and every* $u \in \{a_i, b_i\}$*,* $v \in \{a_j, b_j\}$ *one has* $\|u - v\|_2 \ge 5h$*.*

Proof. Let

$$L_h := \left\lfloor \frac{1}{6h} \right\rfloor.$$

For each multi-index $k = (k_1, \ldots, k_d) \in \{0, \ldots, L_h - 1\}^d$, define

$$a_k := 6h\,k, \qquad b_k := a_k + he_1.$$

Because $6hL_h \le 1$, all these points lie in $[0,1]^d$. Moreover $\|a_k - b_k\|_2 = h$.

If $k \ne \ell$, then in some coordinate r we have $|k_r - \ell_r| \ge 1$. If $r \ne 1$, then every point in the k-pair differs from every point in the ℓ-pair by at least $6h$ in the r-th coordinate, hence their Euclidean distance is at least $6h$. If $r = 1$, then the first coordinates belong to the sets $\{6hk, 6hk + h\}$ and $\{6h\ell, 6h\ell + h\}$; the minimal possible distance between these two sets is $5h$. Thus every inter-pair distance is at least $5h$. Finally $M_h = L_h^d \ge c_d h^{-d}$ for a constant $c_d > 0$ depending only on d. □

Fix such a support once and for all and write

$$\mathcal{S}_h := \{a_i, b_i :\ 1 \le i \le M_h\}.$$

Definition 5.11 (Paired-grid class)**.** Let $\mathcal{C}_h^{\mathrm{pair}}$ denote the class of all pairs $(P,Q) \in \mathcal{P}(\mathcal{S}_h)^2$ for which there exist numbers

$$r_i \ge 0, \qquad \sum_{i=1}^{M_h} r_i = 1,$$

and imbalances $\alpha_i, \beta_i \in [-r_i, r_i]$ such that

$$P(a_i) = \frac{r_i + \alpha_i}{2}, \qquad P(b_i) = \frac{r_i - \alpha_i}{2},$$

$$Q(a_i) = \frac{r_i + \beta_i}{2}, \qquad Q(b_i) = \frac{r_i - \beta_i}{2} \qquad (1 \le i \le M_h).$$

Equivalently, the two measures have the same total mass on each pair $\{a_i, b_i\}$, but may have different within-pair splits.

5.4 Exact reduction to a discrete L_1 functional

Proposition 5.12 (Exact pairwise transport identity). *Let $(P,Q) \in \mathcal{C}_h^{\mathrm{pair}}$, and write*

$$\Delta_i := \alpha_i - \beta_i \qquad (1 \le i \le M_h).$$

Then

$$W_p(P,Q)^p = \frac{h^p}{2} \sum_{i=1}^{M_h} |\Delta_i| = h^p \ \mathrm{TV}(P,Q). \tag{5.1}$$

In particular, on $\mathcal{C}_h^{\mathrm{pair}}$ the transport cost is exactly a scaled discrete L_1 functional.

Proof. For each pair i, the source and target have the same total mass r_i on $\{a_i, b_i\}$. Hence the only discrepancy inside that pair is the signed imbalance

$$P(a_i) - Q(a_i) = \frac{\Delta_i}{2}, \qquad P(b_i) - Q(b_i) = -\frac{\Delta_i}{2}.$$

If $\Delta_i > 0$, transport the mass $\Delta_i/2$ from a_i to b_i; if $\Delta_i < 0$, transport $|\Delta_i|/2$ from b_i to a_i. Doing this independently for all pairs produces a coupling between P and Q with total cost

$$\sum_{i=1}^{M_h} h^p \frac{|\Delta_i|}{2} = \frac{h^p}{2} \sum_{i=1}^{M_h} |\Delta_i|.$$

This proves the upper bound.

For the matching lower bound, fix any coupling π between P and Q. If $\Delta_i > 0$, then the point b_i has target deficit $\Delta_i/2$, so under π at least $\Delta_i/2$ mass must arrive at b_i from points other than b_i itself. Every such source point is at Euclidean distance at least h from b_i: the paired point a_i is at distance exactly h, while every point from a different pair is at distance at least $5h$ by Theorem 5.10. Thus the contribution of the mass entering b_i to the transport cost is at least $h^p \Delta_i/2$. The same argument with a_i and b_i interchanged applies when $\Delta_i < 0$. Summing over i yields

$$\int \|x-y\|_2^p \, d\pi(x,y) \ge \frac{h^p}{2} \sum_{i=1}^{M_h} |\Delta_i|.$$

Taking the infimum over π proves the first equality in (5.1).

For the second equality, note that

$$\mathrm{TV}(P,Q) = \frac{1}{2} \sum_{i=1}^{M_h} \Big(|P(a_i) - Q(a_i)| + |P(b_i) - Q(b_i)| \Big) = \frac{1}{2} \sum_{i=1}^{M_h} |\Delta_i|.$$

Substituting into the first equality proves (5.1). □

5.5 Sharp estimation on the critical paired class

Theorem 5.13 (Paired-grid estimator via large-alphabet L_1 theory). *Let $p \ge 1$, let $h \in (0, 1/12]$, and let $\mathcal{C}_h^{\mathrm{pair}}$ be as above. Assume that*

$$N := n \wedge m \ge c \, \frac{M_h}{\log M_h} \qquad \textit{and} \qquad \log N \le C \log M_h$$

for fixed constants $c, C > 0$. Then there exists an estimator $\widehat{W}_h^{\mathrm{pair}}$, based on the two samples from (P,Q) and the known support $\mathcal{S}_h$, such that

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big|\widehat{W}_h^{\mathrm{pair}} - W_p(P,Q)\Big| \le C_{d,p,c,C} \, h \Big(\frac{M_h}{N \log N} \Big)^{1/(2p)},$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big(\widehat{W}_h^{\mathrm{pair}} - W_p(P,Q)\Big)^2 \le C_{d,p,c,C} \, h^2 \Big(\frac{M_h}{N \log N} \Big)^{1/(p\vee 2)}.$$

Consequently, if $h = h_N$ is chosen so that $M_{h_N} \asymp N \log N$, then

$$\sup_{(P,Q)\in\mathcal{C}_{h_N}^{\mathrm{pair}}} \mathbb{E}\Big|\widehat{W}_{h_N}^{\mathrm{pair}} - W_p(P,Q)\Big| \lesssim_{d,p} h_N,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_{h_N}^{\mathrm{pair}}} \mathbb{E}\Big(\widehat{W}_{h_N}^{\mathrm{pair}} - W_p(P,Q)\Big)^2 \lesssim_{d,p} h_N^2.$$

Proof. By Theorem 5.12,

$$W_p(P,Q) = h \ \mathrm{TV}(P,Q)^{1/p} = h \Big(\frac{\|P-Q\|_1}{2} \Big)^{1/p} \qquad \text{for every } (P,Q) \in \mathcal{C}_h^{\mathrm{pair}}.$$

The support $\mathcal{S}_h$ has size $S_h := 2M_h$. Since the support is known, we may identify P and Q with probability vectors on a known alphabet of size S_h and invoke the large-alphabet L_1-distance estimator of Jiao–Han–Weissman [5, Theorem 6]. That theorem is stated in the Poisson model, and the standard Poissonization equivalence recorded in [5] transfers the same bound to fixed sample sizes up to absolute constants. Under the stated regime assumptions, there exists an estimator $\widehat{L}_h$ such that

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big(\widehat{L}_h - \|P-Q\|_1\Big)^2 \le C_{c,C} \, \frac{S_h}{N \log N} \le C'_{c,C} \, \frac{M_h}{N \log N}.$$

Clip $\widehat{L}_h$ to the natural parameter range by setting

$$\widetilde{L}_h := 0 \vee \widehat{L}_h \wedge 2.$$

Because $\|P-Q\|_1 \in [0,2]$, this clipping cannot increase squared error. Define

$$\widehat{W}_h^{\mathrm{pair}} := h\Big(\frac{\widetilde{L}_h}{2}\Big)^{1/p}, \qquad V(P,Q) := \frac{\|P-Q\|_1}{2} = \mathrm{TV}(P,Q) \in [0,1].$$

Then

$$\widehat{W}_h^{\mathrm{pair}} = h\,\widehat{V}_h^{\,1/p}, \qquad W_p(P,Q) = h\,V(P,Q)^{1/p},$$

where $\widehat{V}_h := \widetilde{L}_h/2 \in [0,1]$.

Since $x \mapsto x^{1/p}$ is $1/p$-Hölder on $[0,\infty)$,

$$|\widehat{V}_h^{\,1/p} - V^{1/p}| \le |\widehat{V}_h - V|^{1/p}.$$

Therefore

$$\mathbb{E}\Big|\widehat{W}_h^{\mathrm{pair}} - W_p(P,Q)\Big| \le h\,\mathbb{E}|\widehat{V}_h - V|^{1/p} \le h\left(\mathbb{E}|\widehat{V}_h - V|^2\right)^{1/(2p)} \lesssim h\Big(\frac{M_h}{N\log N}\Big)^{1/(2p)}.$$

For the squared risk,

$$\Big|\widehat{W}_h^{\mathrm{pair}} - W_p(P,Q)\Big|^2 \le h^2|\widehat{V}_h - V|^{2/p}.$$

If $p \ge 2$, then $2/p \le 1$, so by Jensen's inequality

$$\mathbb{E}|\widehat{V}_h - V|^{2/p} \le \left(\mathbb{E}|\widehat{V}_h - V|^2\right)^{1/p} \lesssim \Big(\frac{M_h}{N\log N}\Big)^{1/p}.$$

If $1 \le p < 2$, then $2/p > 1$, but $|\widehat{V}_h - V| \le 1$ because both $\widehat{V}_h$ and V are nonnegative and bounded by 1 after clipping. Hence $|\widehat{V}_h - V|^{2/p} \le |\widehat{V}_h - V|$, and therefore

$$\mathbb{E}|\widehat{V}_h - V|^{2/p} \le \mathbb{E}|\widehat{V}_h - V| \le \left(\mathbb{E}|\widehat{V}_h - V|^2\right)^{1/2} \lesssim \Big(\frac{M_h}{N\log N}\Big)^{1/2}.$$

Combining the two cases gives

$$\mathbb{E}\Big(\widehat{W}_h^{\mathrm{pair}} - W_p(P,Q)\Big)^2 \lesssim h^2\Big(\frac{M_h}{N\log N}\Big)^{1/(p\vee 2)}.$$

Finally, when $M_{h_N} \asymp N\log N$, the displayed bounds reduce to $O(h_N)$ and $O(h_N^2)$. $\square$

Corollary 5.14 (Large-alphabet exact laws for the paired class)**.** *Fix $p \ge 1$, $\alpha \in (0,1)$, $C_0 \ge 1$, and numbers $0 < \tau_- < \tau_+ < 1$. Assume that*

$$N^\alpha \le M_h \le C_0 N\log N.$$

Then

$$\inf_{\widehat{U}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big|\widehat{U} - W_p(P,Q)^p\Big| \asymp_{d,p,\alpha,C_0} h^p\sqrt{\frac{M_h}{N\log N}},$$

and

$$\inf_{\widehat{U}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big(\widehat{U} - W_p(P,Q)^p\Big)^2 \asymp_{d,p,\alpha,C_0} h^{2p}\frac{M_h}{N\log N}.$$

If

$$\mathcal{C}_{h,\mathrm{ann}}^{\mathrm{pair}}(\tau_-,\tau_+) := \Big\{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}} : \tau_- \le \mathrm{TV}(P,Q) \le \tau_+\Big\},$$

then also

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_{h,\mathrm{ann}}^{\mathrm{pair}}(\tau_-,\tau_+)} \mathbb{E}\Big|\widehat{W} - W_p(P,Q)\Big| \asymp_{d,p,\alpha,C_0,\tau_-,\tau_+} h\sqrt{\frac{M_h}{N\log N}},$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_{h,\mathrm{ann}}^{\mathrm{pair}}(\tau_-,\tau_+)} \mathbb{E}\Big(\widehat{W} - W_p(P,Q)\Big)^2 \asymp_{d,p,\alpha,C_0,\tau_-,\tau_+} h^2\frac{M_h}{N\log N}.$$

Proof. Let

$$\ell_h : [0,1]^d \to \{1,\dots,2M_h\}$$

be any measurable labeling map that sends the support points

$$a_1, b_1, \dots, a_{M_h}, b_{M_h}$$

to distinct labels. Because every $(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}$ is supported on $\mathcal{S}_h$, Theorem 5.12 gives

$$W_p(P,Q)^p = h^p\ \mathrm{TV}(P,Q) = h^p\ \mathrm{TV}\Big((\ell_h)_\#P, (\ell_h)_\#Q\Big).$$

Hence Theorem 5.1 with one level yields an estimator $\widehat{T}_h$ satisfying

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big|\widehat{T}_h - \mathrm{TV}(P,Q)\Big| \lesssim_{d,p,\alpha,C_0} \sqrt{\frac{M_h}{N\log N}},$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\Big(\widehat{T}_h - \mathrm{TV}(P,Q)\Big)^2 \lesssim_{d,p,\alpha,C_0} \frac{M_h}{N\log N}.$$

Therefore

$$\widehat{U}_h := h^p\widehat{T}_h$$

gives the displayed upper bounds for W_p^p. On the annular subclass, Theorem 5.3 applied with $T_N(P,Q) = \mathrm{TV}(P,Q)$ gives

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\left|\widehat{W}_h^{\mathrm{ann}} - W_p(P,Q)\right| \lesssim h\sqrt{\frac{M_h}{N\log N}},$$

and

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\left(\widehat{W}_h^{\mathrm{ann}} - W_p(P,Q)\right)^2 \lesssim h^2\frac{M_h}{N\log N}.$$

For the lower bounds, define for every pair

$$r, s \in \mathcal{P}(\{1,\dots,M_h\})$$

the measures

$$P_{r,s}(a_i) = \frac{r_i}{2}, \qquad P_{r,s}(b_i) = \frac{s_i}{2},$$

$$Q_{r,s}(a_i) = \frac{s_i}{2}, \qquad Q_{r,s}(b_i) = \frac{r_i}{2} \qquad (1 \le i \le M_h).$$

Then $(P_{r,s}, Q_{r,s}) \in \mathcal{C}_h^{\mathrm{pair}}$, and a direct check shows

$$\mathrm{TV}(P_{r,s}, Q_{r,s}) = \mathrm{TV}(r,s).$$

Combining this with Theorem 5.12 yields

$$W_p(P_{r,s}, Q_{r,s})^p = h^p\ \mathrm{TV}(r,s).$$

Let $\widehat{U}$ be any estimator of W_p^p on $\mathcal{C}_h^{\mathrm{pair}}$, and restrict it to the embedded subclass

$$\{(P_{r,s}, Q_{r,s}) : r, s \in \mathcal{P}(\{1,\dots,M_h\})\}.$$

Then

$$\widehat{D} := \frac{\widehat{U}}{h^p}$$

is an estimator of $\mathrm{TV}(r,s)$ on an alphabet of size M_h. Therefore Theorem 5.6 gives

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pair}}} \mathbb{E}\left|\widehat{U} - W_p(P,Q)^p\right| \gtrsim_{d,p,\alpha,C_0} h^p\sqrt{\frac{M_h}{N\log N}}.$$

The squared lower bound follows from Jensen's inequality.

For the annular lower bound, restrict further to

$$\tau_- \le \mathrm{TV}(r,s) \le \tau_+.$$

Then $(P_{r,s}, Q_{r,s}) \in \mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)$, and on this subclass

$$W_p(P_{r,s}, Q_{r,s}) = h\ \mathrm{TV}(r,s)^{1/p}.$$

Let $\widehat{W}$ be any estimator on $\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)$, clip it to

$$h\tau_-^{1/p} \le \widetilde{W} \le h\tau_+^{1/p},$$

and define

$$\widehat{D}^{\mathrm{ann}} := \left(\frac{\widetilde{W}}{h}\right)^p.$$

Because the inverse map $w \mapsto (w/h)^p$ is Lipschitz on $[h\tau_-^{1/p}, h\tau_+^{1/p}]$,

$$\left|\widehat{D}^{\mathrm{ann}} - \mathrm{TV}(r,s)\right| \le C_{p,\tau_-,\tau_+}\, h^{-1}\left|\widehat{W} - W_p(P_{r,s}, Q_{r,s})\right|.$$

Applying Theorem 5.7 gives

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)} \mathbb{E}\left|\widehat{W} - W_p(P,Q)\right| \gtrsim_{d,p,\alpha,C_0,\tau_-,\tau_+} h\sqrt{\frac{M_h}{N\log N}}.$$

The squared lower bound again follows from Jensen's inequality. □

Corollary 5.15 (Critical paired-grid Euclidean core)**.** *Let h_N satisfy $M_{h_N} \asymp N\log N$. Then the class $\mathcal{C}^{\mathrm{pair}}_{h_N}$ is a genuinely Euclidean critical-support class of cardinality $\asymp N\log N$ on which the minimax target scales are attained:*

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h_N}} \mathbb{E}\left|\widehat{W}^{\mathrm{pair}}_{h_N} - W_p(P,Q)\right| \lesssim_{d,p} \eta_N, \qquad \sup_{(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h_N}} \mathbb{E}\left(\widehat{W}^{\mathrm{pair}}_{h_N} - W_p(P,Q)\right)^2 \lesssim_{d,p} \eta_N^2.$$

5.6 An exact finite multiscale band

The one-scale paired theorem shows that a single critical block is fully solvable. The next result pushes this further in a genuinely multiscale direction. We show that any *fixed finite band* of adjacent dyadic scales can be realized inside $[0,1]^d$ by disjoint Euclidean paired supports, and on the resulting class the transport cost splits exactly into a finite sum of scale-wise discrete L_1 functionals. Thus the minimax target is attained not only for one critical block, but for any fixed finite family of neighboring critical/mesoscopic blocks.

Lemma 5.16 (Finite-band paired support)**.** *Fix an integer $L \ge 1$. There exist constants $s_{d,L} \in (0,1)$, $c_{d,L} > 0$, and $h_{d,L} > 0$ such that for every*

$$h \in (0, h_{d,L}]$$

one can find disjoint axis-parallel cubes

$$U_0,\dots,U_{L-1} \subset [0,1]^d$$

each of side length $s_{d,L}$ with the following property.

For every $\ell = 0,\dots,L-1$, writing

$$h_\ell := 2^\ell h,$$

there exist integers $M_{\ell,h}$ satisfying

$$c_{d,L} h_\ell^{-d} \le M_{\ell,h} \le C_{d,L} h_\ell^{-d}$$

for a constant $C_{d,L} < \infty$, and corresponding pairs of points

$$(a_{\ell,i}, b_{\ell,i})_{1\le i\le M_{\ell,h}} \subset U_\ell$$

such that

(i) $\|a_{\ell,i} - b_{\ell,i}\|_2 = h_\ell$ *for every i;*
(ii) *if $i \neq j$, then every point of the i-th pair is at distance at least $5h_\ell$ from every point of the j-th pair;*
(iii) *if $\ell \neq r$, then every point of U_ℓ is at distance at least $10\max(h_\ell, h_r)$ from every point of U_r.*

Proof. Choose L disjoint axis-parallel cubes $U_0, \dots, U_{L-1}$ of the same side length $s_{d,L} \in (0,1)$; for example, one may partition a corner of $[0,1]^d$ into L disjoint subcubes of side length $s_{d,L} := (3L)^{-1/d}$. Let

$$h_{d,L} := \frac{s_{d,L}}{120 \cdot 2^{L-1}}.$$

Then for every $h \le h_{d,L}$ and every $\ell \le L-1$ one has

$$h_\ell = 2^\ell h \le \frac{s_{d,L}}{120}.$$

Translate and scale the construction of Theorem 5.10 inside each U_ℓ. Because U_ℓ has side length $s_{d,L}$, this yields at least

$$M_{\ell,h} \ge c_d \left(\frac{s_{d,L}}{h_\ell}\right)^d =: c_{d,L} h_\ell^{-d}$$

pairs in U_ℓ satisfying (i) and (ii). The matching upper bound $M_{\ell,h} \le C_{d,L} h_\ell^{-d}$ follows from the obvious packing estimate inside a fixed cube of side length $s_{d,L}$.

Since the cubes U_ℓ are disjoint and their mutual separation is a positive constant depending only on (d, L), shrinking $h_{d,L}$ further if necessary ensures that the distance between distinct cubes is at least

$$10 \max_{\ell \le L-1} h_\ell$$

for every admissible h. This gives (iii). □

Fix such a support and write

$$\mathcal{S}_{h,L} := \{a_{\ell,i}, b_{\ell,i} : \ 0 \le \ell \le L-1, \ 1 \le i \le M_{\ell,h}\}, \qquad \mathcal{S}_{h,L}^{(\ell)} := \{a_{\ell,i}, b_{\ell,i} : \ 1 \le i \le M_{\ell,h}\}.$$

Definition 5.17 (Finite-band multiscale paired class)**.** Let $\mathcal{C}_{h,L}^{\text{band}}$ denote the class of all pairs $(P, Q) \in \mathcal{P}(\mathcal{S}_{h,L})^2$ for which there exist numbers

$$r_{\ell,i} \ge 0, \qquad \sum_{\ell=0}^{L-1} \sum_{i=1}^{M_{\ell,h}} r_{\ell,i} = 1,$$

and imbalances $\alpha_{\ell,i}, \beta_{\ell,i} \in [-r_{\ell,i}, r_{\ell,i}]$ such that

$$P(a_{\ell,i}) = \frac{r_{\ell,i} + \alpha_{\ell,i}}{2}, \qquad P(b_{\ell,i}) = \frac{r_{\ell,i} - \alpha_{\ell,i}}{2},$$
$$Q(a_{\ell,i}) = \frac{r_{\ell,i} + \beta_{\ell,i}}{2}, \qquad Q(b_{\ell,i}) = \frac{r_{\ell,i} - \beta_{\ell,i}}{2}$$

for every ℓ and i. Equivalently, for every individual pair $\{a_{\ell,i}, b_{\ell,i}\}$, the two measures have the same total mass on that pair, while the within-pair split may differ.

Proposition 5.18 (Exact finite-band decomposition)**.** *Let $(P, Q) \in \mathcal{C}_{h,L}^{\text{band}}$, and write*

$$\Delta_{\ell,i} := \alpha_{\ell,i} - \beta_{\ell,i}.$$

Then

$$W_p(P,Q)^p = \sum_{\ell=0}^{L-1} \frac{h_\ell^p}{2} \sum_{i=1}^{M_{\ell,h}} |\Delta_{\ell,i}|. \tag{5.2}$$

Equivalently, if

$$L_\ell(P,Q) := \sum_{i=1}^{M_{\ell,h}} \Big(|P(a_{\ell,i}) - Q(a_{\ell,i})| + |P(b_{\ell,i}) - Q(b_{\ell,i})| \Big),$$

then

$$W_p(P,Q)^p = \sum_{\ell=0}^{L-1} \frac{h_\ell^p}{2} L_\ell(P,Q). \tag{5.3}$$

Proof. For the upper bound, transport each pair independently exactly as in Theorem 5.12. If $\Delta_{\ell,i} > 0$, move the mass $\Delta_{\ell,i}/2$ from $a_{\ell,i}$ to $b_{\ell,i}$; if $\Delta_{\ell,i} < 0$, move $|\Delta_{\ell,i}|/2$ from $b_{\ell,i}$ to $a_{\ell,i}$. Because the total mass on every pair is the same under P and Q, this produces a valid coupling between P and Q with total cost

$$\sum_{\ell=0}^{L-1} \sum_{i=1}^{M_{\ell,h}} h_\ell^p \frac{|\Delta_{\ell,i}|}{2}.$$

For the lower bound, fix an arbitrary coupling π between P and Q. Suppose $\Delta_{\ell,i} > 0$. Then the target point $b_{\ell,i}$ has deficit $\Delta_{\ell,i}/2$, so under π at least $\Delta_{\ell,i}/2$ mass must enter $b_{\ell,i}$ from points other than $b_{\ell,i}$ itself. Every such source point is at Euclidean distance at least h_ℓ from $b_{\ell,i}$: its partner $a_{\ell,i}$ is at distance exactly h_ℓ, every other point of the same scale is at distance at least $5h_\ell$ by Theorem 5.16, and every point of a different scale is at distance at least $10\max(h_\ell, h_r) \ge h_\ell$. Thus the contribution of the mass entering $b_{\ell,i}$ to the transport cost is at least $h_\ell^p\Delta_{\ell,i}/2$. The same argument with $a_{\ell,i}$ and $b_{\ell,i}$ interchanged applies when $\Delta_{\ell,i}<0$. Summing over all pairs gives

$$\int \|x-y\|_2^p\,d\pi(x,y) \ge \sum_{\ell=0}^{L-1}\sum_{i=1}^{M_{\ell,h}} h_\ell^p\frac{|\Delta_{\ell,i}|}{2}.$$

Taking the infimum over π proves (5.2).

Finally,

$$L_\ell(P,Q)=\sum_{i=1}^{M_{\ell,h}}\left(\left|\frac{\Delta_{\ell,i}}{2}\right|+\left|\frac{-\Delta_{\ell,i}}{2}\right|\right)=\sum_{i=1}^{M_{\ell,h}}|\Delta_{\ell,i}|,$$

and substituting this into (5.2) yields (5.3). □

Theorem 5.19 (Exact finite-band multiscale critical law)**.** *Fix $L\ge 1$, $p\ge 1$, and assume $d>2p$. Let $h\in(0,h_{d,L}]$, let $h_\ell=2^\ell h$, and let $\mathcal{C}_{h,L}^{\mathrm{band}}$ be as in Theorem 5.17. Assume that*

$$N:=n\wedge m\ge c\,\frac{h^{-d}}{\log(1/h)}$$

for a sufficiently large constant $c=c(d,p,L)$. Then there exists an estimator $\hat W_{h,L}^{\mathrm{band}}$ such that

$$\sup_{(P,Q)\in\mathcal{C}_{h,L}^{\mathrm{band}}}\mathbb{E}\left|\hat W_{h,L}^{\mathrm{band}}-W_p(P,Q)\right|\le C_{d,p,L}\,h,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_{h,L}^{\mathrm{band}}}\mathbb{E}\left(\hat W_{h,L}^{\mathrm{band}}-W_p(P,Q)\right)^2\le C_{d,p,L}\,h^2.$$

Moreover, if in addition

$$h^{-d}\asymp N\log N,$$

then the minimax risks on $\mathcal{C}_{h,L}^{\mathrm{band}}$ satisfy the exact law

$$\inf_{\hat W}\sup_{(P,Q)\in\mathcal{C}_{h,L}^{\mathrm{band}}}\mathbb{E}\left|\hat W-W_p(P,Q)\right|\asymp_{d,p,L} h,$$

and

$$\inf_{\hat W}\sup_{(P,Q)\in\mathcal{C}_{h,L}^{\mathrm{band}}}\mathbb{E}\left(\hat W-W_p(P,Q)\right)^2\asymp_{d,p,L} h^2.$$

Proof. For each scale ℓ, collapse all points outside $\mathcal{S}_{h,L}^{(\ell)}$ to a cemetery symbol $\star_\ell$. This defines probability measures $\overline{P}_\ell,\overline{Q}_\ell$ on the known alphabet

$$\mathcal{A}_\ell:=\mathcal{S}_{h,L}^{(\ell)}\cup\{\star_\ell\},\qquad |\mathcal{A}_\ell|=2M_{\ell,h}+1\asymp_{d,L}h_\ell^{-d}.$$

Because P and Q have the same total mass on every pair, they also have the same total mass on the whole ℓ-th scale:

$$P(\mathcal{S}_{h,L}^{(\ell)})=\sum_{i=1}^{M_{\ell,h}}r_{\ell,i}=Q(\mathcal{S}_{h,L}^{(\ell)}).$$

Hence the cemetery masses coincide,

$$\overline{P}_\ell(\star_\ell)=\overline{Q}_\ell(\star_\ell),$$

and therefore

$$\|\overline{P}_\ell-\overline{Q}_\ell\|_1=L_\ell(P,Q). \tag{5.4}$$

Now map each original sample point to its image in $\mathcal{A}_\ell$. For fixed ℓ, this yields i.i.d. samples from $\overline{P}_\ell$ and $\overline{Q}_\ell$. Since L is fixed and $h_\ell=2^\ell h$, the alphabet size satisfies

$$|\mathcal{A}_\ell|\le C_{d,L}h_\ell^{-d}\le C'_{d,L}h^{-d},$$

and by the assumption $N\gtrsim h^{-d}/\log(1/h)$ the large-alphabet regime of [5, Theorems 2 and 4] applies uniformly in ℓ. Thus there exists an estimator $\widehat{L}_\ell$ of $\|\overline{P}_\ell-\overline{Q}_\ell\|_1$ such that

$$\sup_{(P,Q)\in\mathcal{C}_{h,L}^{\mathrm{band}}}\mathbb{E}\left(\widehat{L}_\ell-\|\overline{P}_\ell-\overline{Q}_\ell\|_1\right)^2\le C_{d,p,L}\,\frac{h_\ell^{-d}}{N\log N}. \tag{5.5}$$

Clip each $\widehat{L}_\ell$ to $[0,2]$; this does not increase squared error because the parameter itself belongs to $[0,2]$. Using (5.4) and Theorem 5.18, define

$$\widehat{T}_{h,L}:=\sum_{\ell=0}^{L-1}\frac{h_\ell^p}{2}\,\widehat{L}_\ell,\qquad \hat W_{h,L}^{\mathrm{band}}:=\widehat{T}_{h,L}^{1/p}.$$

Then

$$W_p(P,Q)^p=\sum_{\ell=0}^{L-1}\frac{h_\ell^p}{2}\,\|\overline{P}_\ell-\overline{Q}_\ell\|_1.$$

We first bound the cost-level error. By Cauchy–Schwarz and (5.5),

$$\mathbb{E}|\widehat{T}_{h,L}-W_p(P,Q)^p|\le\sum_{\ell=0}^{L-1}\frac{h_\ell^p}{2}\left(\mathbb{E}(\widehat{L}_\ell-\|\overline{P}_\ell-\overline{Q}_\ell\|_1)^2\right)^{1/2}$$

$$\le C_{d,p,L}\sum_{\ell=0}^{L-1} h_\ell^p\Big(\frac{h_\ell^{-d}}{N\log N}\Big)^{1/2} = C_{d,p,L}\sum_{\ell=0}^{L-1}\frac{h_\ell^{p-d/2}}{\sqrt{N\log N}}.$$

Because $h^{-d}\lesssim N\log N$ and $h_\ell = 2^\ell h$, this becomes

$$\mathbb{E}|\widehat{T}_{h,L} - W_p(P,Q)^p| \le C_{d,p,L}\, h^p \sum_{\ell=0}^{L-1} 2^{\ell(p-d/2)} \le C'_{d,p,L}\, h^p,$$

where the final sum is finite because L is fixed.

Similarly, using $(\sum_{\ell=0}^{L-1} x_\ell)^2 \le L\sum_{\ell=0}^{L-1} x_\ell^2$ and (5.5),

$$\mathbb{E}|\widehat{T}_{h,L} - W_p(P,Q)^p|^2 \le C_L \sum_{\ell=0}^{L-1} h_\ell^{2p}\mathbb{E}(\widehat{L}_\ell - \|\overline{P}_\ell - \overline{Q}_\ell\|_1)^2$$

$$\le C_{d,p,L}\sum_{\ell=0}^{L-1} h_\ell^{2p}\frac{h_\ell^{-d}}{N\log N} \le C_{d,p,L}\, h^{2p}\sum_{\ell=0}^{L-1} 2^{\ell(2p-d)} \le C'_{d,p,L}\, h^{2p},$$

since $d > 2p$ in the supercritical regime and L is fixed.

We now convert from the cost level to the distance level. For the absolute risk, Theorem 5.18 gives $W_p(P,Q)^p \in [0, C_{p,L}h^p]$, and therefore

$$\mathbb{E}|\hat{W}^{\mathrm{band}}_{h,L} - W_p(P,Q)| \le \Big(\mathbb{E}|\widehat{T}_{h,L} - W_p(P,Q)^p|\Big)^{1/p} \le C_{d,p,L}\, h.$$

For the squared risk, if $p\ge 2$, then $2/p \le 1$ and Jensen yields

$$\mathbb{E}\Big(\hat{W}^{\mathrm{band}}_{h,L} - W_p(P,Q)\Big)^2 \le \Big(\mathbb{E}|\widehat{T}_{h,L} - W_p(P,Q)^p|^2\Big)^{1/p} \le C_{d,p,L}\, h^2.$$

If $1\le p<2$, then $2/p>1$, but the bounded cost range implies

$$|\widehat{T}_{h,L} - W_p(P,Q)^p|^{2/p} \le (C_{p,L}h^p)^{2/p-1}|\widehat{T}_{h,L} - W_p(P,Q)^p| \le C'_{p,L}h^{2-p}|\widehat{T}_{h,L} - W_p(P,Q)^p|.$$

Hence the absolute cost bound gives

$$\mathbb{E}\Big(\hat{W}^{\mathrm{band}}_{h,L} - W_p(P,Q)\Big)^2 \le C'_{p,L}h^{2-p}\mathbb{E}|\widehat{T}_{h,L} - W_p(P,Q)^p| \le C_{d,p,L}\, h^2.$$

This proves the upper bounds.

For the lower bounds at critical resolution $h^{-d}\asymp N\log N$, note that $\mathcal{C}^{\mathrm{band}}_{h,L}$ contains the subclass obtained by placing all mass on the finest scale $\ell = 0$ and zero mass on every coarser scale. On this subclass, Theorem 5.18 reduces to the one-scale identity

$$W_p(P,Q) = h\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p}$$

over a known alphabet of cardinality $\asymp h^{-d}\asymp N\log N$. By [5, Theorem 3], the minimax squared error for estimating $\|P-Q\|_1$ on that alphabet is bounded below by a positive constant depending only on (d,L). Since the parameter is bounded by 2, this also implies a positive lower bound on its absolute risk. If $\hat{W}$ is any estimator of $W_p(P,Q)$ on the band class, define

$$\widehat{L} := 2\Big(\frac{\hat{W}\wedge(2^{1/p}h)}{h}\Big)^p.$$

The map $w\mapsto 2(w/h)^p$ is $C_p h^{-1}$-Lipschitz on $[0, 2^{1/p}h]$, so on the one-scale subclass

$$\mathbb{E}|\widehat{L} - \|P-Q\|_1| \le C_p h^{-1}\mathbb{E}|\hat{W} - W_p(P,Q)|.$$

Therefore the constant lower bound for estimating $\|P-Q\|_1$ forces

$$\inf_{\hat{W}}\sup_{(P,Q)\in\mathcal{C}^{\mathrm{band}}_{h,L}} \mathbb{E}|\hat{W} - W_p(P,Q)| \gtrsim_{d,p,L} h.$$

The squared-risk lower bound then follows from Jensen:

$$\sup_{(P,Q)}\mathbb{E}(\hat{W} - W_p(P,Q))^2 \ge \Big(\sup_{(P,Q)}\mathbb{E}|\hat{W} - W_p(P,Q)|\Big)^2.$$

This completes the proof. □

5.7 A maximal packed direct-sum theorem

The finite-band theorem freezes only finitely many neighboring scales. We now push the exact Euclidean decomposition much further. The next result solves *arbitrary finite or infinite families of dyadic scales* as long as the active blocks satisfy the natural Euclidean packing constraint. This is the maximal positive reach of the disjoint-block strategy.

For a base mesh $h\in(0,1/24]$, write

$$h_\ell := 2^\ell h, \qquad 0\le \ell\le L_h := \max\{\ell\ge 0:\ h_\ell\le 1/24\}.$$

Lemma 5.20 (Dyadic host-cube packing)**.** *There exists a constant $\kappa_d^{\mathrm{pack}}\in(0,1)$ with the following property. Let $(\sigma_j)_{1\le j\le J}$ be dyadic side lengths, $\sigma_j = 2^{-k_j}$, such that*

$$\sum_{j=1}^J \sigma_j^d \le \kappa_d^{\mathrm{pack}}.$$

Then there exist pairwise disjoint dyadic cubes

$$V_j\subset[0,1]^d, \qquad \mathrm{side}(V_j) = \sigma_j, \qquad 1\le j\le J.$$

Proof. Choose $\kappa_d^{\mathrm{pack}} := 1/4$. For each $k \geq 0$, let

$$n_k := \#\{j : \ \sigma_j = 2^{-k}\}.$$

Then

$$\sum_{k\geq 0} n_k 2^{-kd} \leq \frac{1}{4}.$$

We place the cubes scale by scale, in increasing k (that is, from large cubes to small cubes).

Fix $k \geq 0$ and suppose all cubes of side strictly larger than 2^{-k} have already been placed. Every previously placed cube of side 2^{-j} with $j < k$ occupies exactly $2^{d(k-j)}$ level-k dyadic cells. Hence the number of level-k cells already occupied is

$$\sum_{j<k} n_j 2^{d(k-j)} = 2^{kd} \sum_{j<k} n_j 2^{-jd} \leq \frac{1}{4}\, 2^{kd}.$$

But the total number of level-k dyadic cells in $[0,1]^d$ is 2^{kd}, so at least

$$\frac{3}{4}\, 2^{kd}$$

level-k cells remain free. On the other hand,

$$n_k 2^{-kd} \leq \sum_{j\geq 0} n_j 2^{-jd} \leq \frac{1}{4},$$

hence

$$n_k \leq \frac{1}{4}\, 2^{kd}.$$

Therefore there are enough free level-k cells to place all n_k cubes of side 2^{-k}. Proceeding inductively over k gives the required disjoint family. □

Lemma 5.21 (Packed direct-sum support geometry)**.** *There exist constants $\kappa_d \in (0,1)$ and $C_d < \infty$ with the following property. Let $h \in (0, 1/C_d]$ and let integers*

$$M_{\ell,h} \in \mathbb{N} \cup \{0\}, \qquad 0 \leq \ell \leq L_h,$$

satisfy the packing constraint

$$\sum_{\ell=0}^{L_h} M_{\ell,h} h_\ell^d \leq \kappa_d. \tag{5.6}$$

Then one can construct a finite set

$$\mathcal{S}_h^{\mathrm{pack}} = \{a_{\ell,i}, b_{\ell,i} : \ 0 \leq \ell \leq L_h,\ 1 \leq i \leq M_{\ell,h}\} \subset [0,1]^d$$

such that for every active scale ℓ and every $1 \leq i \leq M_{\ell,h}$:
- (i) *$\|a_{\ell,i} - b_{\ell,i}\|_2 = h_\ell$;*
- (ii) *if $i \neq j$, then every point of the i-th pair is at distance at least $5h_\ell$ from every point of the j-th pair;*
- (iii) *there exists a host cube $V_{\ell,h} \subset [0,1]^d$ such that all points of the ℓ-th scale lie in $V_{\ell,h}$, the host cubes are pairwise disjoint across ℓ, and every support point at scale ℓ is at distance at least $3h_\ell$ from $[0,1]^d \setminus V_{\ell,h}$.*

Proof. Let

$$\Lambda_h := \{\ell \in \{0, \ldots, L_h\} : \ M_{\ell,h} \geq 1\}.$$

For every $\ell \in \Lambda_h$, define

$$s_{\ell,h} := 12\, M_{\ell,h}^{1/d} h_\ell.$$

Choose a dyadic number $\sigma_{\ell,h}$ such that

$$s_{\ell,h} \leq \sigma_{\ell,h} < 2s_{\ell,h}.$$

Then

$$\sum_{\ell\in\Lambda_h} \sigma_{\ell,h}^d \leq 2^d 12^d \sum_{\ell\in\Lambda_h} M_{\ell,h} h_\ell^d.$$

Choose

$$\kappa_d := 2^{-d} 12^{-d} \kappa_d^{\mathrm{pack}},$$

where κ_d^{pack} is the constant from Theorem 5.20. Then (5.6) implies

$$\sum_{\ell\in\Lambda_h} \sigma_{\ell,h}^d \leq \kappa_d^{\mathrm{pack}},$$

so Theorem 5.20 yields pairwise disjoint dyadic cubes

$$V_{\ell,h} \subset [0,1]^d, \qquad \mathrm{side}(V_{\ell,h}) = \sigma_{\ell,h}, \qquad \ell \in \Lambda_h.$$

For each active ℓ, let $U_{\ell,h}$ be the concentric subcube of $V_{\ell,h}$ with side length $\sigma_{\ell,h}/2$. Then

$$\mathrm{side}(U_{\ell,h}) = \frac{\sigma_{\ell,h}}{2} \geq 6M_{\ell,h}^{1/d} h_\ell.$$

Inside $U_{\ell,h}$, place a translated and rescaled copy of the separated paired-grid construction from Theorem 5.10 with mesh h_ℓ. Because the side length is at least $6M_{\ell,h}^{1/d} h_\ell$, this yields at least $M_{\ell,h}$ pairs satisfying (i) and (ii); we keep exactly $M_{\ell,h}$ of them.

Finally, since $U_{\ell,h}$ is concentric in $V_{\ell,h}$,

$$\mathrm{dist}\big(U_{\ell,h}, [0,1]^d \setminus V_{\ell,h}\big) = \frac{\sigma_{\ell,h}}{4} \geq 3M_{\ell,h}^{1/d} h_\ell \geq 3h_\ell.$$

Thus every support point of scale ℓ lies at distance at least $3h_\ell$ from $[0,1]^d \setminus V_{\ell,h}$. This proves (iii). □

Fix such a support once and for all.

Definition 5.22 (Packed direct-sum multiscale class)**.** Let $\mathcal{C}_h^{\text{pack}}$ denote the class of all pairs $(P,Q) \in \mathcal{P}(\mathcal{S}_h^{\text{pack}})^2$ such that for every active scale ℓ and every $1 \le i \le M_{\ell,h}$ there exist numbers

$$r_{\ell,i} \ge 0, \qquad \alpha_{\ell,i}, \beta_{\ell,i} \in [-r_{\ell,i}, r_{\ell,i}],$$

with

$$P(a_{\ell,i}) = \frac{r_{\ell,i} + \alpha_{\ell,i}}{2}, \qquad P(b_{\ell,i}) = \frac{r_{\ell,i} - \alpha_{\ell,i}}{2},$$
$$Q(a_{\ell,i}) = \frac{r_{\ell,i} + \beta_{\ell,i}}{2}, \qquad Q(b_{\ell,i}) = \frac{r_{\ell,i} - \beta_{\ell,i}}{2},$$

and

$$\sum_{\ell=0}^{L_h} \sum_{i=1}^{M_{\ell,h}} r_{\ell,i} = 1.$$

Equivalently, on every individual pair $\{a_{\ell,i}, b_{\ell,i}\}$ the two measures have the same total mass, while the within-pair split may differ arbitrarily.

Proposition 5.23 (Exact packed direct-sum decomposition)**.** *For every $(P,Q) \in \mathcal{C}_h^{\text{pack}}$, writing*

$$\Delta_{\ell,i} := \alpha_{\ell,i} - \beta_{\ell,i},$$

one has

$$W_p(P,Q)^p = \sum_{\ell=0}^{L_h} \frac{h_\ell^p}{2} \sum_{i=1}^{M_{\ell,h}} |\Delta_{\ell,i}|. \tag{5.7}$$

Equivalently, if

$$L_\ell(P,Q) := \sum_{i=1}^{M_{\ell,h}} \Big(|P(a_{\ell,i}) - Q(a_{\ell,i})| + |P(b_{\ell,i}) - Q(b_{\ell,i})| \Big),$$

then

$$W_p(P,Q)^p = \sum_{\ell=0}^{L_h} \frac{h_\ell^p}{2} L_\ell(P,Q). \tag{5.8}$$

Proof. As in Theorems 5.12 and 5.18, transport each pair independently: if $\Delta_{\ell,i} > 0$, move the mass $\Delta_{\ell,i}/2$ from $a_{\ell,i}$ to $b_{\ell,i}$; if $\Delta_{\ell,i} < 0$, move $|\Delta_{\ell,i}|/2$ from $b_{\ell,i}$ to $a_{\ell,i}$. Because the total mass on every pair is the same under P and Q, this gives a feasible coupling with cost equal to the right-hand side of (5.7).

For the matching lower bound, fix an arbitrary coupling π. Suppose $\Delta_{\ell,i} > 0$. Then the target point $b_{\ell,i}$ has deficit $\Delta_{\ell,i}/2$, so at least $\Delta_{\ell,i}/2$ mass must arrive at $b_{\ell,i}$ from points other than $b_{\ell,i}$ itself. Any such source point is at Euclidean distance at least h_ℓ from $b_{\ell,i}$: its partner $a_{\ell,i}$ is at distance exactly h_ℓ; any other point at the same scale is at distance at least $5h_\ell$ by Theorem 5.21; and any point at a different scale lies outside the host cube $V_{\ell,h}$ and is therefore at distance at least $3h_\ell$ from $b_{\ell,i}$ by Theorem 5.21(iii). Thus the contribution of the mass entering $b_{\ell,i}$ is at least $h_\ell^p \Delta_{\ell,i}/2$. The same argument with $a_{\ell,i}$ and $b_{\ell,i}$ interchanged applies when $\Delta_{\ell,i} < 0$. Summing over all pairs and taking the infimum over π proves (5.7).

The identity (5.8) follows exactly as in Theorem 5.18:

$$L_\ell(P,Q) = \sum_{i=1}^{M_{\ell,h}} |\Delta_{\ell,i}|.$$

□

Theorem 5.24 (Packed direct-sum critical law)**.** *Assume $d > 2p$. There exist constants $\kappa_d \in (0,1)$ and $C_{d,p} < \infty$ such that the following holds.*

Let $h \in (0, 1/C_{d,p}]$, let $(M_{\ell,h})_{0 \le \ell \le L_h}$ satisfy the packing condition

$$\sum_{\ell=0}^{L_h} M_{\ell,h} h_\ell^d \le \kappa_d,$$

and assume moreover that

$$M_{0,h} \ge c_d h^{-d} \tag{5.9}$$

for some constant $c_d \in (0, \kappa_d]$. Then:

(i) *the support size obeys*

$$\#\mathcal{S}_h^{\text{pack}} = 2 \sum_{\ell=0}^{L_h} M_{\ell,h} \asymp_d h^{-d};$$

(ii) *if*

$$N \log N \ge C_{d,p} h^{-d},$$

there exists an estimator $\hat{W}_h^{\text{pack}}$ such that

$$\sup_{(P,Q) \in \mathcal{C}_h^{\text{pack}}} \mathbb{E} \Big| \hat{W}_h^{\text{pack}} - W_p(P,Q) \Big| \le C_{d,p}\, h,$$

and

$$\sup_{(P,Q) \in \mathcal{C}_h^{\text{pack}}} \mathbb{E} \big(\hat{W}_h^{\text{pack}} - W_p(P,Q) \big)^2 \le C_{d,p}\, h^2;$$

(iii) *if in addition*

$$h^{-d} \asymp N \log N,$$

then the minimax risks on $\mathcal{C}_h^{\mathrm{pack}}$ *satisfy*

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pack}}} \mathbb{E}\big|\hat{W} - W_p(P,Q)\big| \asymp_{d,p} h,$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{pack}}} \mathbb{E}\big(\hat{W} - W_p(P,Q)\big)^2 \asymp_{d,p} h^2.$$

Proof. Set

$$S_h := \frac{1}{2}\,\#\mathcal{S}_h^{\mathrm{pack}} = \sum_{\ell=0}^{L_h} M_{\ell,h}.$$

By (5.9),

$$S_h \geq M_{0,h} \gtrsim h^{-d}.$$

On the other hand, the packing bound implies

$$M_{\ell,h} \leq \kappa_d h_\ell^{-d} = \kappa_d 2^{-\ell d} h^{-d},$$

hence

$$S_h \leq \kappa_d h^{-d} \sum_{\ell=0}^{L_h} 2^{-\ell d} \lesssim_{d,p} h^{-d}.$$

This proves part (i).

For part (ii), fix

$$\alpha \in \left(0, 1 - \frac{2p}{d}\right).$$

For each scale ℓ, collapse all support points outside the ℓ-th scale to a cemetery symbol $\star_\ell$. This yields probability measures $\overline{P}_\ell, \overline{Q}_\ell$ on a known alphabet of size

$$S_{\ell,h} := 2M_{\ell,h} + 1.$$

Because P and Q have the same total mass on every pair, they also have the same total mass on the whole ℓ-th scale, so the cemetery masses coincide and

$$\|\overline{P}_\ell - \overline{Q}_\ell\|_1 = L_\ell(P,Q).$$

We estimate each $L_\ell(P,Q)$ separately.

Large-alphabet scales. If $M_{\ell,h} \geq N^\alpha$, then $S_{\ell,h} \asymp M_{\ell,h}$ and

$$\log N \leq \alpha^{-1} \log S_{\ell,h}$$

for all large N. The sharp estimator of Jiao–Han–Weissman for unknown (P,Q) on a known alphabet then gives

$$\mathbb{E}\Big(\widehat{L}_\ell - L_\ell(P,Q)\Big)^2 \lesssim_\alpha \frac{M_{\ell,h}}{N \log N}.$$

Small-alphabet scales. If $M_{\ell,h} < N^\alpha$, then $S_{\ell,h} \lesssim N^\alpha$. Since $\alpha < 1$, the empirical plug-in/MLE estimator on the alphabet $\mathcal{A}_\ell$ satisfies

$$\mathbb{E}\Big(\widehat{L}_\ell - L_\ell(P,Q)\Big)^2 \lesssim \frac{M_{\ell,h}}{N}$$

for all large N, by the classical L_1 plug-in bound and the MLE risk bound of Jiao–Han–Weissman.

Define

$$\widehat{T}_h^{\mathrm{pack}} := \sum_{\ell=0}^{L_h} \frac{h_\ell^p}{2}\,\widehat{L}_\ell, \qquad \hat{W}_h^{\mathrm{pack}} := \big(\widehat{T}_h^{\mathrm{pack}}\big)^{1/p}.$$

By Theorem 5.23,

$$W_p(P,Q)^p = \sum_{\ell=0}^{L_h} \frac{h_\ell^p}{2}\, L_\ell(P,Q).$$

We first bound the cost-level error. By Cauchy–Schwarz,

$$\mathbb{E}\Big|\widehat{T}_h^{\mathrm{pack}} - W_p(P,Q)^p\Big| \leq \sum_{\ell=0}^{L_h} \frac{h_\ell^p}{2} \sqrt{\mathbb{E}\Big(\widehat{L}_\ell - L_\ell(P,Q)\Big)^2}.$$

Hence

$$\mathbb{E}\Big|\widehat{T}_h^{\mathrm{pack}} - W_p(P,Q)^p\Big| \lesssim \frac{1}{\sqrt{N\log N}} \sum_{\ell:\, M_{\ell,h}\geq N^\alpha} h_\ell^p \sqrt{M_{\ell,h}} + \frac{1}{\sqrt{N}} \sum_{\ell:\, M_{\ell,h}<N^\alpha} h_\ell^p \sqrt{M_{\ell,h}}.$$

For the large-alphabet part, use the packing condition:

$$\sum_{\ell:\, M_{\ell,h}\geq N^\alpha} h_\ell^p \sqrt{M_{\ell,h}} = \sum_{\ell:\, M_{\ell,h}\geq N^\alpha} \sqrt{M_{\ell,h} h_\ell^d}\, h_\ell^{p-d/2} \leq \kappa_d^{1/2} \Bigg(\sum_{\ell=0}^{L_h} h_\ell^{2p-d}\Bigg)^{1/2}.$$

Since $d > 2p$ and $h_\ell = 2^\ell h$,

$$\sum_{\ell=0}^{L_h} h_\ell^{2p-d} = h^{2p-d} \sum_{\ell=0}^{L_h} 2^{\ell(2p-d)} \lesssim_{d,p} h^{2p-d}.$$

Therefore

$$\frac{1}{\sqrt{N\log N}}\sum_{\ell:\,M_{\ell,h}\ge N^\alpha} h_\ell^p\sqrt{M_{\ell,h}} \lesssim_{d,p} \frac{h^{p-d/2}}{\sqrt{N\log N}} \lesssim_{d,p} h^p$$

because $N\log N \ge C_{d,p}h^{-d}$.

For the small-alphabet part, use $\sqrt{M_{\ell,h}} \le N^{\alpha/2}$:

$$\frac{1}{\sqrt{N}}\sum_{\ell:\,M_{\ell,h}<N^\alpha} h_\ell^p\sqrt{M_{\ell,h}} \le N^{(\alpha-1)/2}\sum_{\ell=0}^{L_h} h_\ell^p \lesssim_{d,p} N^{(\alpha-1)/2}.$$

Because $\alpha < 1-2p/d$, we have

$$\frac{1-\alpha}{2} > \frac{p}{d},$$

hence

$$N^{(\alpha-1)/2} = o(h^p)$$

under the critical relation $h^{-d} \lesssim N\log N$. Combining the two contributions yields

$$\sup_{(P,Q)\in\mathcal{C}_h^{\text{pack}}} \mathbb{E}\Big|\widehat{T}_h^{\text{pack}} - W_p(P,Q)^p\Big| \lesssim_{d,p} h^p.$$

The same estimates imply

$$\sup_{(P,Q)\in\mathcal{C}_h^{\text{pack}}} \mathbb{E}\Big(\widehat{T}_h^{\text{pack}} - W_p(P,Q)^p\Big)^2 \lesssim_{d,p} h^{2p}.$$

Now convert to the distance level. Since for every $p \ge 1$,

$$|a-b|^p \le |a^p - b^p|, \qquad a,b\ge 0,$$

we obtain

$$\mathbb{E}\Big|\hat{W}_h^{\text{pack}} - W_p(P,Q)\Big| \le \left(\mathbb{E}\Big|\widehat{T}_h^{\text{pack}} - W_p(P,Q)^p\Big|\right)^{1/p} \lesssim_{d,p} h,$$

and

$$\mathbb{E}\Big(\hat{W}_h^{\text{pack}} - W_p(P,Q)\Big)^2 \le \left(\mathbb{E}\Big(\widehat{T}_h^{\text{pack}} - W_p(P,Q)^p\Big)^2\right)^{1/p} \lesssim_{d,p} h^2.$$

This proves part (ii).

For the lower bound in part (iii), restrict $\mathcal{C}_h^{\text{pack}}$ to the subclass obtained by setting all masses on scales $\ell \ge 1$ equal to zero. Then only the scale-0 pairs remain. By (5.9), this one-scale restriction contains

$$M_{0,h} \asymp h^{-d} \asymp N\log N$$

paired sites, and it is precisely the finite-band class with $L = 1$. Therefore the lower-bound part of Theorem 5.19 applies and yields

$$\inf_{\hat{W}}\sup \mathbb{E}|\hat{W} - W_p(P,Q)| \gtrsim_{d,p} h, \qquad \inf_{\hat{W}}\sup \mathbb{E}(\hat{W} - W_p(P,Q))^2 \gtrsim_{d,p} h^2.$$

Together with part (ii), this proves part (iii). □

5.8 The exact dyadic pair-isolation law

The proof of Theorem 5.24 uses the host cubes only to produce an exact dyadic pair decomposition and the count bound $M_{\ell,h} \lesssim h_\ell^{-d}$. Once these two inputs are available abstractly, the same statistical argument goes through unchanged.

Fix $h \in (0,1/2]$, write

$$h_\ell := 2^\ell h, \qquad 0 \le \ell \le L_h := \max\{\ell \ge 0:\ h_\ell \le 1/2\},$$

and let

$$\mathcal{S}_h^{\text{iso}} = \{a_{\ell,i}, b_{\ell,i}:\ 0\le \ell \le L_h,\ 1 \le i \le M_{\ell,h}\} \subset [0,1]^d$$

satisfy

$$\|a_{\ell,i} - b_{\ell,i}\|_2 = h_\ell$$

and

$$\min\{\|z-a_{\ell,i}\|_2, \|z-b_{\ell,i}\|_2\} \ge h_\ell \qquad \forall z \in \mathcal{S}_h^{\text{iso}} \setminus \{a_{\ell,i}, b_{\ell,i}\}.$$

Let $\mathcal{C}_h^{\text{iso}}$ be defined exactly as in Theorem 5.22 on this support.

Proposition 5.25 (Exact dyadic pair-isolation decomposition)**.** *For every $(P,Q) \in \mathcal{C}_h^{\text{iso}}$, writing*

$$\Delta_{\ell,i} := \alpha_{\ell,i} - \beta_{\ell,i}, \qquad L_\ell(P,Q) := \sum_{i=1}^{M_{\ell,h}} |\Delta_{\ell,i}|,$$

one has

$$W_p(P,Q)^p = \sum_{\ell=0}^{L_h}\frac{h_\ell^p}{2}\sum_{i=1}^{M_{\ell,h}}|\Delta_{\ell,i}| = \sum_{\ell=0}^{L_h}\frac{h_\ell^p}{2}L_\ell(P,Q).$$

Proof. The proof is identical to that of Theorem 5.23: the upper bound transports each pair independently, and the lower bound uses only the pointwise isolation property above to show that every deficit at $a_{\ell,i}$ or $b_{\ell,i}$ must be supplied from Euclidean distance at least h_ℓ. □

Theorem 5.26 (Exact dyadic pair-isolation critical law)**.** *Assume $d > 2p$, and fix constants $K < \infty$ and $c_0 > 0$. Then there exists $C_{K,c_0,d,p} < \infty$ such that the following holds.*

If

$$\sum_{\ell=0}^{L_h} M_{\ell,h} h_\ell^d \le K \qquad \text{and} \qquad M_{0,h} \ge c_0 h^{-d},$$

then

$$\#\mathcal{S}_h^{\mathrm{iso}} \asymp_{K,c_0,d} h^{-d}.$$

If moreover

$$N \log N \ge C_{K,c_0,d,p}\, h^{-d},$$

there exists an estimator $\hat{W}_h^{\mathrm{iso}}$ *such that*

$$\sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{iso}}} \mathbb{E}\big|\hat{W}_h^{\mathrm{iso}} - W_p(P,Q)\big| \le C_{K,c_0,d,p}\, h, \qquad \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{iso}}} \mathbb{E}\big(\hat{W}_h^{\mathrm{iso}} - W_p(P,Q)\big)^2 \le C_{K,c_0,d,p}\, h^2.$$

If in addition

$$h^{-d} \asymp N \log N,$$

then the minimax risks on $\mathcal{C}_h^{\mathrm{iso}}$ *satisfy*

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{iso}}} \mathbb{E}\big|\hat{W} - W_p(P,Q)\big| \asymp_{K,c_0,d,p} h, \qquad \inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\mathrm{iso}}} \mathbb{E}\big(\hat{W} - W_p(P,Q)\big)^2 \asymp_{K,c_0,d,p} h^2.$$

Proof. The count assumptions imply

$$M_{\ell,h} \le K h_\ell^{-d} = K\, 2^{-\ell d} h^{-d},$$

hence

$$c_0 h^{-d} \le \sum_{\ell=0}^{L_h} M_{\ell,h} \le K h^{-d} \sum_{\ell=0}^{L_h} 2^{-\ell d} \lesssim_{K,d} h^{-d},$$

which proves the support-size claim.

For the upper bounds, the proof of Theorem 5.24 applies verbatim once one replaces Theorem 5.23 by Theorem 5.25. Indeed, that argument uses only the exact identity above and the count bound $M_{\ell,h} \le K h_\ell^{-d}$. Repeating the same large-/small-alphabet split therefore produces an estimator $\hat{W}_h^{\mathrm{iso}}$ with the stated risks.

For the lower bounds under $h^{-d} \asymp N \log N$, restrict to the subclass with all mass on scale 0. Then

$$W_p(P,Q) = h\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p}$$

over a known alphabet of size

$$2M_{0,h} \asymp_{K,c_0} h^{-d} \asymp N \log N.$$

The lower-bound argument from the end of the proof of Theorem 5.19 therefore applies verbatim and yields the claimed absolute lower bound, and the squared lower bound follows by Jensen's inequality. □

5.9 A genuinely nested near-critical class

The packed direct-sum theorem still freezes each active scale into disjoint host cubes. We now show that deep laminar overlap by itself is not the remaining obstruction. For every sufficiently large integer B, there is an explicit B-adic Euclidean class with overlap depth $L+1$, support exponent

$$s_B = d + \log_B\Big(1 - \frac{2}{B}\Big) \uparrow d,$$

and exact minimax law at the natural critical scale $h^{-s_B} \asymp N \log N$.

Fix an integer $B \ge 8$. Set

$$\mathcal{D}_B := \{2, 3, \ldots, B-1\} \times \{0, 1, \ldots, B-1\}^{d-1}, \qquad A_B := |\mathcal{D}_B| = (B-2)B^{d-1},$$

and

$$s_B := \log_B A_B = d + \log_B\Big(1 - \frac{2}{B}\Big) \in (d-1, d).$$

For $\ell \ge 0$, let $\Sigma_{B,\ell} := \mathcal{D}_B^\ell$, with $\Sigma_{B,0} := \{\varnothing\}$. For $\sigma = (\sigma_1, \ldots, \sigma_\ell) \in \Sigma_{B,\ell}$ define

$$u_\ell := B^{-\ell}, \qquad x_\sigma := \sum_{k=1}^{\ell} B^{-k}\sigma_k \in [0,1)^d, \qquad Q_\sigma := x_\sigma + [0, u_\ell)^d.$$

Thus $(Q_\sigma)_\sigma$ is a prefix-nested family of B-adic cells and

$$|\Sigma_{B,\ell}| = A_B^\ell = B^{s_B \ell}.$$

For every $\sigma \in \Sigma_{B,\ell}$ set

$$a_\sigma := x_\sigma + \frac{1}{4} u_{\ell+1}\mathbf{1}, \qquad b_\sigma := a_\sigma + \frac{1}{4} u_{\ell+1} e_1, \qquad r_\ell := \|a_\sigma - b_\sigma\|_2 = \frac{1}{4} u_{\ell+1}.$$

Given $L \ge 0$, define the nested support

$$\mathcal{S}_{B,L}^{\mathrm{nest}} := \{a_\sigma, b_\sigma :\ 0 \le |\sigma| \le L\}.$$

Proposition 5.27 (Laminar separation)**.** *Fix* $B \ge 8$*,* $L \ge 0$*, and* $\sigma \in \Sigma_{B,\ell}$ *with* $\ell \le L$*. Then:*

(i) $a_\sigma, b_\sigma \in Q_\sigma$ *and*

$$\mathrm{dist}(a_\sigma, \partial Q_\sigma) \ge r_\ell, \qquad \mathrm{dist}(b_\sigma, \partial Q_\sigma) \ge r_\ell;$$

(ii) *every point of* $\mathcal{S}^{\mathrm{nest}}_{B,L} \cap Q_\sigma$ *other than* a_σ, b_σ *belongs to a proper descendant of* Q_σ *and hence satisfies*

$$x^{(1)} \ge x^{(1)}_\sigma + 2u_{\ell+1};$$

in particular it lies at Euclidean distance at least $6r_\ell$ *from both* a_σ *and* b_σ*;*

(iii) *consequently every* $z \in \mathcal{S}^{\mathrm{nest}}_{B,L} \setminus \{a_\sigma, b_\sigma\}$ *satisfies*

$$\|z - a_\sigma\|_2 \ge r_\ell, \qquad \|z - b_\sigma\|_2 \ge r_\ell.$$

Proof. Part (i) is immediate from the definitions: the first coordinate of a_σ is $x^{(1)}_\sigma + \frac{1}{4}u_{\ell+1}$, the first coordinate of b_σ is $x^{(1)}_\sigma + \frac{1}{2}u_{\ell+1}$, and every other coordinate equals $x^{(j)}_\sigma + \frac{1}{4}u_{\ell+1}$. Thus both points belong to Q_σ and each coordinate stays at distance at least $\frac{1}{4}u_{\ell+1} = r_\ell$ from the boundary of Q_σ.

For part (ii), note that every proper descendant of Q_σ must first choose a child digit in $\mathcal{D}_B$, whose first coordinate is at least 2. Hence the first coordinate of every support point in a proper descendant is at least $x^{(1)}_\sigma + 2u_{\ell+1}$. Since the first coordinates of a_σ and b_σ are at most $x^{(1)}_\sigma + \frac{1}{2}u_{\ell+1}$, the separation along the first axis is at least

$$2u_{\ell+1} - \frac{1}{2}u_{\ell+1} = \frac{3}{2}u_{\ell+1} = 6r_\ell.$$

For part (iii), if $z \in Q_\sigma$, then z is covered by part (ii). If $z \notin Q_\sigma$, then part (i) shows that both a_σ and b_σ lie at Euclidean distance at least r_ℓ from the complement of Q_σ. Thus $\|z - a_\sigma\|_2 \ge r_\ell$ and $\|z - b_\sigma\|_2 \ge r_\ell$ in all cases. ☐

Definition 5.28 (The nested laminar class)**.** Fix $B \ge 8$ and $L \ge 0$. Let $\mathcal{C}^{\mathrm{nest}}_{B,L}$ be the family of pairs (P, Q) of the form

$$P = \sum_{|\sigma| \le L} \Big(\alpha_\sigma \delta_{a_\sigma} + (w_\sigma - \alpha_\sigma)\delta_{b_\sigma}\Big), \qquad Q = \sum_{|\sigma| \le L} \Big(\beta_\sigma \delta_{a_\sigma} + (w_\sigma - \beta_\sigma)\delta_{b_\sigma}\Big),$$

where

$$w_\sigma \ge 0, \qquad \sum_{|\sigma| \le L} w_\sigma = 1, \qquad 0 \le \alpha_\sigma, \beta_\sigma \le w_\sigma.$$

Proposition 5.29 (Exact L_1 representation on the nested class)**.** *Fix $B \ge 8$ and $L \ge 0$. For $(P, Q) \in \mathcal{C}^{\mathrm{nest}}_{B,L}$, define*

$$L_\ell(P, Q) := \sum_{|\sigma| = \ell} |\alpha_\sigma - \beta_\sigma|.$$

Then

$$W_p(P, Q)^p = \sum_{\ell=0}^{L} \frac{r_\ell^p}{2} L_\ell(P, Q) = \sum_{|\sigma| \le L} \frac{r_{|\sigma|}^p}{2} |\alpha_\sigma - \beta_\sigma|.$$

Proof. Write $\Delta_\sigma := \alpha_\sigma - \beta_\sigma$. Transporting $|\Delta_\sigma|/2$ units of mass directly along the segment (a_σ, b_σ) for every σ gives the upper bound

$$W_p(P, Q)^p \le \sum_{|\sigma| \le L} \frac{r_{|\sigma|}^p}{2} |\Delta_\sigma|.$$

For the reverse inequality, let π be any coupling of (P, Q). Fix σ of length ℓ. If $\Delta_\sigma > 0$, then the point b_σ has deficit $\Delta_\sigma/2$ under Q relative to P, so at least $\Delta_\sigma/2$ units of mass must enter b_σ from support points different from b_σ. By Theorem 5.27, every such source lies at distance at least r_ℓ from b_σ. Hence the contribution to the transport cost is at least $r_\ell^p \Delta_\sigma/2$. If $\Delta_\sigma < 0$, the same argument with a_σ and b_σ interchanged gives the lower bound $r_\ell^p |\Delta_\sigma|/2$. Summing over all σ and taking the infimum over π proves the claimed identity. ☐

Theorem 5.30 (Genuinely nested near-critical law)**.** *Assume $d > 2p$. Let $B \ge 8$ be so large that*

$$s_B > 2p.$$

Then there exists a constant $C_{B,d,p} < \infty$ such that for every $L \ge 1$, with

$$h_{B,L} := r_L = \frac{1}{4} B^{-(L+1)},$$

the following hold:

(i) *the support size obeys*

$$\#\mathcal{S}^{\mathrm{nest}}_{B,L} = 2 \sum_{\ell=0}^{L} A_B^\ell \asymp_B A_B^L \asymp_B h_{B,L}^{-s_B};$$

(ii) *if*

$$N \log N \ge C_{B,d,p}\, h_{B,L}^{-s_B},$$

then there exists an estimator $\hat{W}^{\mathrm{nest}}_{B,L}$ such that

$$\sup_{(P,Q) \in \mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big|\hat{W}^{\mathrm{nest}}_{B,L} - W_p(P, Q)\Big| \le C_{B,d,p}\, h_{B,L},$$

and

$$\sup_{(P,Q) \in \mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big(\hat{W}^{\mathrm{nest}}_{B,L} - W_p(P, Q)\Big)^2 \le C_{B,d,p}\, h_{B,L}^2;$$

(iii) *if in addition*

$$h_{B,L}^{-s_B} \asymp N\log N,$$

then the minimax risks on $\mathcal{C}_{B,L}^{\mathrm{nest}}$ *satisfy*

$$\inf_{\hat W}\sup_{(P,Q)\in\mathcal{C}_{B,L}^{\mathrm{nest}}} \mathbb{E}\left|\hat W - W_p(P,Q)\right| \asymp_{B,d,p} h_{B,L},$$

and

$$\inf_{\hat W}\sup_{(P,Q)\in\mathcal{C}_{B,L}^{\mathrm{nest}}} \mathbb{E}\left(\hat W - W_p(P,Q)\right)^2 \asymp_{B,d,p} h_{B,L}^2.$$

Proof. Part (i) follows from the geometric-series identity

$$\sum_{\ell=0}^{L} A_B^\ell \asymp_B A_B^L$$

and from

$$A_B^L = B^{s_B L} \asymp_B h_{B,L}^{-s_B}.$$

For part (ii), fix

$$\alpha \in \left(0, 1-\frac{2p}{s_B}\right).$$

For each scale ℓ, collapse all support points outside the level-ℓ set

$$\mathcal{S}_{B,L}^{(\ell)} := \{a_\sigma, b_\sigma :\ |\sigma| = \ell\}$$

to a cemetery symbol $\star_\ell$. Because P and Q assign the same total mass w_σ to every pair, they also assign the same total mass to the whole level ℓ. Thus the collapsed measures $\overline{P}_\ell, \overline{Q}_\ell$ satisfy

$$\|\overline{P}_\ell - \overline{Q}_\ell\|_1 = L_\ell(P,Q)$$

on a known alphabet of size

$$S_{\ell,B} := 2A_B^\ell + 1 \asymp_B A_B^\ell \asymp_B r_\ell^{-s_B}.$$

If $A_B^\ell \geq N^\alpha$, then $S_{\ell,B} \asymp_B A_B^\ell$ and therefore

$$\log N \leq \alpha^{-1}\log S_{\ell,B} + O_{B,\alpha}(1)$$

for all large N. The large-alphabet estimator of [5, Theorem 6] then gives

$$\mathbb{E}\left(\widehat{L}_\ell - L_\ell(P,Q)\right)^2 \lesssim_{B,\alpha} \frac{A_B^\ell}{N\log N}.$$

If $A_B^\ell < N^\alpha$, then the empirical plug-in/MLE estimator on the alphabet of size $S_{\ell,B} \lesssim_B N^\alpha$ gives

$$\mathbb{E}\left(\widehat{L}_\ell - L_\ell(P,Q)\right)^2 \lesssim_B \frac{A_B^\ell}{N}.$$

Clip each estimator to the natural parameter range by setting

$$\widetilde{L}_\ell := 0 \vee \widehat{L}_\ell \wedge 2.$$

Clipping cannot increase squared error.

Define

$$\widehat{T}_{B,L} := \sum_{\ell=0}^{L} \frac{r_\ell^p}{2}\widetilde{L}_\ell, \qquad \hat W_{B,L}^{\mathrm{nest}} := \widehat{T}_{B,L}^{1/p}.$$

By Theorem 5.29,

$$W_p(P,Q)^p = \sum_{\ell=0}^{L}\frac{r_\ell^p}{2} L_\ell(P,Q).$$

Therefore

$$\mathbb{E}\left|\widehat{T}_{B,L} - W_p(P,Q)^p\right| \leq \sum_{\ell=0}^{L}\frac{r_\ell^p}{2}\sqrt{\mathbb{E}\left(\widetilde{L}_\ell - L_\ell(P,Q)\right)^2}.$$

Split the sum into large and small alphabets. For the large-alphabet part,

$$\frac{1}{\sqrt{N\log N}}\sum_{\ell:\, A_B^\ell \geq N^\alpha} r_\ell^p\sqrt{A_B^\ell} \lesssim_B \frac{1}{\sqrt{N\log N}}\sum_{\ell=0}^{L} B^{\ell(s_B/2-p)} \lesssim_{B,d,p} \frac{h_{B,L}^{p-s_B/2}}{\sqrt{N\log N}} \lesssim_{B,d,p} h_{B,L}^p,$$

because $s_B > 2p$ and $N\log N \geq C_{B,d,p} h_{B,L}^{-s_B}$.

For the small-alphabet part, $\sqrt{A_B^\ell} \leq N^{\alpha/2}$, so

$$\frac{1}{\sqrt{N}}\sum_{\ell:\, A_B^\ell < N^\alpha} r_\ell^p\sqrt{A_B^\ell} \leq N^{(\alpha-1)/2}\sum_{\ell=0}^{L} r_\ell^p \lesssim_{B,p} N^{(\alpha-1)/2}.$$

Since $(1-\alpha)s_B/2 > p$, the sampling relation $h_{B,L}^{-s_B} \lesssim N\log N$ implies

$$N^{(\alpha-1)/2} = O(h_{B,L}^p).$$

Hence

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big|\widehat{T}_{B,L} - W_p(P,Q)^p\Big| \lesssim_{B,d,p} h^p_{B,L}.$$

The same bound with the square outside follows from

$$\mathbb{E}\Big(\widehat{T}_{B,L} - W_p(P,Q)^p\Big)^2 \le \Big(\sum_{\ell=0}^{L} \frac{r^p_\ell}{2}\sqrt{\mathbb{E}\big(\widetilde{L}_\ell - L_\ell(P,Q)\big)^2}\Big)^2 \lesssim_{B,d,p} h^{2p}_{B,L}.$$

Now convert to the distance level. Since for every $p \ge 1$,

$$|u-v|^p \le |u^p - v^p|, \qquad u,v \ge 0,$$

we obtain

$$\mathbb{E}\Big|\hat{W}^{\mathrm{nest}}_{B,L} - W_p(P,Q)\Big| \le \Big(\mathbb{E}\Big|\widehat{T}_{B,L} - W_p(P,Q)^p\Big|\Big)^{1/p} \lesssim_{B,d,p} h_{B,L},$$

and likewise

$$\mathbb{E}\Big(\hat{W}^{\mathrm{nest}}_{B,L} - W_p(P,Q)\Big)^2 \le \Big(\mathbb{E}\Big(\widehat{T}_{B,L} - W_p(P,Q)^p\Big)^2\Big)^{1/p} \lesssim_{B,d,p} h^2_{B,L}.$$

For part (iii), restrict to the subclass in which all masses w_σ vanish off the finest scale $|\sigma| = L$. Then Theorem 5.29 reduces to

$$W_p(P,Q) = h_{B,L}\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p}$$

over a known alphabet of cardinality

$$2A^L_B \asymp_B h^{-s_B}_{B,L} \asymp N\log N.$$

By [5, Theorem 3], the minimax squared error for estimating $\|P-Q\|_1$ on this alphabet is bounded below by a positive constant depending only on (B,d). Since the parameter is bounded by 2, the minimax absolute risk is also bounded below by a positive constant. If $\hat{W}$ is any estimator of $W_p(P,Q)$ on $\mathcal{C}^{\mathrm{nest}}_{B,L}$, define

$$\widehat{L} := 2\Big(\frac{\hat{W}\wedge(2^{1/p}h_{B,L})}{h_{B,L}}\Big)^p.$$

Exactly as in the finite-band lower bound, the map $w \mapsto 2(w/h_{B,L})^p$ is $C_p h^{-1}_{B,L}$-Lipschitz on $[0, 2^{1/p}h_{B,L}]$. Therefore a constant lower bound for estimating $\|P-Q\|_1$ forces

$$\inf_{\hat{W}}\sup_{(P,Q)\in\mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \gtrsim_{B,d,p} h_{B,L},$$

and

$$\inf_{\hat{W}}\sup_{(P,Q)\in\mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big(\hat{W} - W_p(P,Q)\Big)^2 \gtrsim_{B,d,p} h^2_{B,L}.$$

Together with part (ii), this proves part (iii). ☐

Corollary 5.31 (Nested classes arbitrarily close to the critical exponent)**.** *Assume $d > 2p$ and fix $\varepsilon \in (0, d-2p)$. Then there exists an integer $B = B(d,p,\varepsilon)$ and explicit genuinely nested Euclidean classes $\mathcal{C}^{\mathrm{nest}}_{B,L}$ such that*

$$s_B > d - \varepsilon, \qquad \#\mathcal{S}^{\mathrm{nest}}_{B,L} \asymp_B h^{-s_B}_{B,L},$$

and whenever

$$h^{-s_B}_{B,L} \asymp N\log N,$$

their minimax risks satisfy

$$\inf_{\hat{W}}\sup_{(P,Q)\in\mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \asymp_{B,d,p} h_{B,L},$$

and

$$\inf_{\hat{W}}\sup_{(P,Q)\in\mathcal{C}^{\mathrm{nest}}_{B,L}} \mathbb{E}\Big(\hat{W} - W_p(P,Q)\Big)^2 \asymp_{B,d,p} h^2_{B,L}.$$

Proof. Because $s_B = d + \log_B(1-2/B) \uparrow d$ as $B \to \infty$, one can choose B so that $s_B > d - \varepsilon$. The claim then follows immediately from Theorem 5.30. ☐

5.10 Critical laminar universality

The fixed-B theorem reaches every exponent $s < d$, but it still pays the same corridor factor $(1-2/B)$ at every generation. The explicit accelerating-branch construction shows that this loss can be made summable. The next theorem upgrades that example to a full universality statement: *every* Euclidean laminar hierarchy whose corridor losses are summable and whose generation-wise occupancy remains asymptotically critical already satisfies the sharp minimax law at the full scale $h \asymp (N\log N)^{-1/d}$.

Let $(B_\ell)_{\ell\ge0}$ be integers with

$$B_\ell \ge 8 \qquad (\ell \ge 0),$$

and let $(\varepsilon_\ell)_{\ell\ge0}$ satisfy

$$\frac{2}{B_\ell} \le \varepsilon_\ell \le \frac{1}{2}, \qquad \sum_{\ell=0}^{\infty}\varepsilon_\ell < \infty.$$

Set

$$c_\star := \prod_{\ell=0}^{\infty}(1-\varepsilon_\ell) \in (0,1].$$

Fix a depth $L \geq 0$ and define

$$u_0 := 1, \qquad u_{\ell+1} := \frac{u_\ell}{B_\ell} \qquad (0 \leq \ell \leq L-1).$$

We recursively build a laminar family of cubes. Let

$$\Sigma_0^{\mathrm{lam,crit}} := \{\varnothing\}, \qquad x_\varnothing := 0, \qquad Q_\varnothing := [0,1)^d.$$

Given a node $\sigma \in \Sigma_\ell^{\mathrm{lam,crit}}$ with $\ell < L$, choose a set of child digits

$$\mathcal{D}_\sigma \subseteq \{2,3,\dots,B_\ell - 1\} \times \{0,1,\dots,B_\ell - 1\}^{d-1}$$

with cardinality

$$(1-\varepsilon_\ell)B_\ell^d \leq A_\sigma := |\mathcal{D}_\sigma| \leq (B_\ell - 2)B_\ell^{d-1} = \Big(1 - \frac{2}{B_\ell}\Big)B_\ell^d. \tag{5.10}$$

Define

$$\Sigma_{\ell+1}^{\mathrm{lam,crit}} := \{(\sigma, v) : \ \sigma \in \Sigma_\ell^{\mathrm{lam,crit}}, \ v \in \mathcal{D}_\sigma\}.$$

For a child $\tau = (\sigma, v) \in \Sigma_{\ell+1}^{\mathrm{lam,crit}}$, set

$$x_\tau := x_\sigma + u_{\ell+1} v, \qquad Q_\tau := x_\tau + [0, u_{\ell+1})^d.$$

For every node $\sigma \in \Sigma_\ell^{\mathrm{lam,crit}}$, define support points

$$a_\sigma, b_\sigma \in Q_\sigma$$

as follows. If $\ell < L$, set

$$a_\sigma := x_\sigma + \frac{1}{4}u_{\ell+1}\mathbf{1}, \qquad b_\sigma := a_\sigma + \frac{1}{4}u_{\ell+1}e_1, \qquad r_\ell := \frac{1}{4}u_{\ell+1}.$$

At the deepest level, set

$$a_\sigma := x_\sigma + \frac{1}{32}u_L\mathbf{1}, \qquad b_\sigma := a_\sigma + \frac{1}{32}u_L e_1, \qquad r_L := \frac{1}{32}u_L.$$

Finally define

$$\mathcal{S}_L^{\mathrm{lam,crit}} := \{a_\sigma, b_\sigma : \ \sigma \in \Sigma_\ell^{\mathrm{lam,crit}} \text{ for some } 0 \leq \ell \leq L\}, \qquad h_L^{\mathrm{lam,crit}} := r_L = \frac{1}{32}u_L.$$

Proposition 5.32 (General critical laminar separation)**.** *Fix $L \geq 0$ and $\sigma \in \Sigma_\ell^{\mathrm{lam,crit}}$ with $\ell \leq L$. Then:*

(i) *$a_\sigma, b_\sigma \in Q_\sigma$ and*

$$\mathrm{dist}(a_\sigma, \partial Q_\sigma) \geq r_\ell, \qquad \mathrm{dist}(b_\sigma, \partial Q_\sigma) \geq r_\ell;$$

(ii) *if $\ell < L$, then every point of $\mathcal{S}_L^{\mathrm{lam,crit}} \cap Q_\sigma$ other than a_σ, b_σ belongs to a proper descendant of Q_σ and hence satisfies*

$$x^{(1)} \geq x_\sigma^{(1)} + 2u_{\ell+1};$$

in particular it lies at Euclidean distance at least $6r_\ell$ from both a_σ and b_σ;

(iii) *consequently every $z \in \mathcal{S}_L^{\mathrm{lam,crit}} \setminus \{a_\sigma, b_\sigma\}$ satisfies*

$$\|z - a_\sigma\|_2 \geq r_\ell, \qquad \|z - b_\sigma\|_2 \geq r_\ell.$$

Proof. For $\ell < L$, part (i) is identical to Theorem 5.27: the first coordinates of a_σ, b_σ equal $x_\sigma^{(1)} + \frac{1}{4}u_{\ell+1}$ and $x_\sigma^{(1)} + \frac{1}{2}u_{\ell+1}$, while all other coordinates equal $x_\sigma^{(j)} + \frac{1}{4}u_{\ell+1}$. Hence both points belong to Q_σ and stay at distance at least $r_\ell = \frac{1}{4}u_{\ell+1}$ from the boundary.

For $\ell = L$, the first coordinates of a_σ, b_σ equal $x_\sigma^{(1)} + \frac{1}{32}u_L$ and $x_\sigma^{(1)} + \frac{1}{16}u_L$, while every other coordinate equals $x_\sigma^{(j)} + \frac{1}{32}u_L$. Thus both points again belong to Q_σ and lie at distance at least $r_L = \frac{1}{32}u_L$ from ∂Q_σ. This proves part (i).

For part (ii), fix $\ell < L$. Every proper descendant of Q_σ is contained in some child cube

$$Q_{(\sigma,v)} \subset Q_\sigma \qquad \text{with } v \in \mathcal{D}_\sigma.$$

Since the first coordinate of every $v \in \mathcal{D}_\sigma$ is at least 2, every support point in a proper descendant has first coordinate at least

$$x_\sigma^{(1)} + 2u_{\ell+1}.$$

Since the first coordinates of a_σ, b_σ are at most $x_\sigma^{(1)} + \frac{1}{2}u_{\ell+1}$, the separation along the first axis is at least

$$2u_{\ell+1} - \frac{1}{2}u_{\ell+1} = \frac{3}{2}u_{\ell+1} = 6r_\ell.$$

For part (iii), if $z \in Q_\sigma$, then either $z = a_\sigma$, $z = b_\sigma$, or z is covered by part (ii). If $z \notin Q_\sigma$, then part (i) implies that both a_σ and b_σ lie at Euclidean distance at least r_ℓ from the complement of Q_σ. Therefore $\|z - a_\sigma\|_2 \geq r_\ell$ and $\|z - b_\sigma\|_2 \geq r_\ell$ in all cases. ☐

Definition 5.33 (The general critical laminar class)**.** Fix $L \geq 0$ and a laminar hierarchy as above. Let $\mathcal{C}_L^{\mathrm{lam,crit}}$ be the family of pairs (P, Q) of the form

$$P = \sum_\sigma \Big(\alpha_\sigma \delta_{a_\sigma} + (w_\sigma - \alpha_\sigma)\delta_{b_\sigma}\Big), \qquad Q = \sum_\sigma \Big(\beta_\sigma \delta_{a_\sigma} + (w_\sigma - \beta_\sigma)\delta_{b_\sigma}\Big),$$

where the sum runs over all nodes σ with $0 \leq |\sigma| \leq L$ and

$$w_\sigma \geq 0, \qquad \sum_\sigma w_\sigma = 1, \qquad 0 \leq \alpha_\sigma, \beta_\sigma \leq w_\sigma.$$

Proposition 5.34 (Exact L_1 representation on the critical laminar class)**.** *Fix $L \geq 0$ and a laminar hierarchy as above. For $(P,Q) \in \mathcal{C}_L^{\mathrm{lam,crit}}$, define*

$$L_\ell^{\mathrm{lam,crit}}(P,Q) := \sum_{\sigma \in \Sigma_\ell^{\mathrm{lam,crit}}} |\alpha_\sigma - \beta_\sigma|.$$

Then

$$W_p(P,Q)^p = \sum_{\ell=0}^{L} \frac{r_\ell^p}{2} L_\ell^{\mathrm{lam,crit}}(P,Q) = \sum_{\sigma} \frac{r_{|\sigma|}^p}{2} |\alpha_\sigma - \beta_\sigma|,$$

where the last sum runs over all nodes σ with $0 \leq |\sigma| \leq L$.

Proof. Write $\Delta_\sigma := \alpha_\sigma - \beta_\sigma$. Transporting $|\Delta_\sigma|/2$ units of mass directly along the segment (a_σ, b_σ) for every node σ yields

$$W_p(P,Q)^p \leq \sum_{\sigma} \frac{r_{|\sigma|}^p}{2} |\Delta_\sigma|.$$

For the reverse inequality, let π be any coupling of (P,Q) and fix a node σ at level ℓ. If $\Delta_\sigma > 0$, then the point b_σ has deficit $\Delta_\sigma/2$ under Q relative to P, so at least $\Delta_\sigma/2$ units of mass must enter b_σ from support points different from b_σ. By Theorem 5.32, every such source lies at distance at least r_ℓ from b_σ. Hence the contribution to the transport cost is at least $r_\ell^p \Delta_\sigma/2$. If $\Delta_\sigma < 0$, the same argument with a_σ and b_σ interchanged yields the lower bound $r_\ell^p |\Delta_\sigma|/2$. Summing over all nodes σ and taking the infimum over π proves the claim. □

Theorem 5.35 (General critical laminar law)**.** *Assume $d > 2p$. There exists a constant $C_{d,p,c_\star} < \infty$ such that for every depth $L \geq 1$ and every critical laminar hierarchy above the following hold.*

(i) *The support size obeys*

$$\#\mathcal{S}_L^{\mathrm{lam,crit}} \asymp_{d,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^{-d}.$$

(ii) *If*

$$N \log N \geq C_{d,p,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^{-d},$$

then there exists an estimator $\hat{W}_L^{\mathrm{lam,crit}}$ such that

$$\sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\left|\hat{W}_L^{\mathrm{lam,crit}} - W_p(P,Q)\right| \leq C_{d,p,c_\star}\, h_L^{\mathrm{lam,crit}},$$

and

$$\sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\left(\hat{W}_L^{\mathrm{lam,crit}} - W_p(P,Q)\right)^2 \leq C_{d,p,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^2.$$

(iii) *If in addition*

$$\left(h_L^{\mathrm{lam,crit}}\right)^{-d} \asymp N \log N,$$

then the minimax risks on $\mathcal{C}_L^{\mathrm{lam,crit}}$ satisfy

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\left|\hat{W} - W_p(P,Q)\right| \asymp_{d,p,c_\star} h_L^{\mathrm{lam,crit}},$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\left(\hat{W} - W_p(P,Q)\right)^2 \asymp_{d,p,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^2.$$

Proof. For each level ℓ, let

$$M_\ell := |\Sigma_\ell^{\mathrm{lam,crit}}| \qquad (0 \leq \ell \leq L).$$

For part (i), the hierarchy construction gives the recursion

$$M_{\ell+1} = \sum_{\sigma\in\Sigma_\ell^{\mathrm{lam,crit}}} A_\sigma \qquad (0 \leq \ell \leq L-1).$$

Since every chosen child cube is one of the B_ℓ^d B_ℓ-adic subcubes of its parent,

$$M_{\ell+1} \leq B_\ell^d M_\ell.$$

Starting from $M_0 = 1$, induction yields

$$M_\ell \leq u_\ell^{-d} \qquad (0 \leq \ell \leq L).$$

On the other hand, (5.10) gives

$$M_{\ell+1} \geq (1-\varepsilon_\ell) B_\ell^d M_\ell,$$

hence

$$M_\ell \geq u_\ell^{-d} \prod_{j=0}^{\ell-1} (1-\varepsilon_j) \geq c_\star\, u_\ell^{-d} \qquad (0 \leq \ell \leq L).$$

Moreover, again by (5.10),

$$A_\sigma \geq (1-\varepsilon_\ell) B_\ell^d \geq \frac{1}{2} \cdot 8^d \geq 2 \qquad (\sigma \in \Sigma_\ell^{\mathrm{lam,crit}},\ \ell < L),$$

so $M_{\ell+1} \geq 2M_\ell$ and therefore

$$M_L \leq \sum_{\ell=0}^{L} M_\ell \leq 2M_L.$$

Consequently

$$2c_\star\, u_L^{-d} \le \#\mathcal{S}_L^{\mathrm{lam,crit}} = 2\sum_{\ell=0}^{L} M_\ell \le 4u_L^{-d}.$$

Because $h_L^{\mathrm{lam,crit}} = u_L/32$, this is equivalent to

$$\#\mathcal{S}_L^{\mathrm{lam,crit}} \asymp_{d,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^{-d}.$$

For part (ii), choose

$$\alpha \in \left(0, 1-\frac{2p}{d}\right).$$

For each level ℓ, collapse all support points outside

$$\mathcal{S}_{L,\ell}^{\mathrm{lam,crit}} := \{a_\sigma, b_\sigma :\ \sigma \in \Sigma_\ell^{\mathrm{lam,crit}}\}$$

to a cemetery symbol $\star_\ell$. Because P and Q assign the same total mass w_σ to every pair, the collapsed measures $\overline{P}_\ell, \overline{Q}_\ell$ satisfy

$$\|\overline{P}_\ell - \overline{Q}_\ell\|_1 = L_\ell^{\mathrm{lam,crit}}(P,Q)$$

on a known alphabet of size

$$S_\ell := 2M_\ell + 1.$$

If $M_\ell \ge N^\alpha$, then $S_\ell \asymp M_\ell$ and therefore

$$\log N \le \alpha^{-1}\log S_\ell + O_\alpha(1).$$

The large-alphabet estimator of [5, Theorem 6] then gives

$$\mathbb{E}\Big(\widehat{L}_\ell - L_\ell^{\mathrm{lam,crit}}(P,Q)\Big)^2 \lesssim_{d,p} \frac{M_\ell}{N\log N}.$$

If $M_\ell < N^\alpha$, then the empirical plug-in/MLE estimator on the alphabet of size $S_\ell \lesssim N^\alpha$ gives

$$\mathbb{E}\Big(\widehat{L}_\ell - L_\ell^{\mathrm{lam,crit}}(P,Q)\Big)^2 \lesssim_{d,p} \frac{M_\ell}{N}.$$

Clip each estimator to the natural range by setting

$$\widetilde{L}_\ell := 0 \vee \widehat{L}_\ell \wedge 2.$$

Clipping cannot increase squared error.

Define

$$\widehat{T}_L^{\mathrm{lam,crit}} := \sum_{\ell=0}^{L} \frac{r_\ell^p}{2}\,\widetilde{L}_\ell, \qquad \hat{W}_L^{\mathrm{lam,crit}} := \left(\widehat{T}_L^{\mathrm{lam,crit}}\right)^{1/p}.$$

By Theorem 5.34,

$$W_p(P,Q)^p = \sum_{\ell=0}^{L} \frac{r_\ell^p}{2}\, L_\ell^{\mathrm{lam,crit}}(P,Q).$$

For every ℓ one has

$$M_\ell \le u_\ell^{-d}, \qquad r_\ell \le \frac{u_\ell}{32}.$$

Hence

$$r_\ell^p\sqrt{M_\ell} \le 32^{-p} u_\ell^{p-d/2}.$$

Set $q := d/2 - p > 0$. Because $u_{\ell+1} \le u_\ell/8$ for $0 \le \ell \le L-1$,

$$u_\ell^{-q} \le 8^{-q(L-\ell)} u_L^{-q}.$$

Therefore

$$\sum_{\ell=0}^{L} r_\ell^p\sqrt{M_\ell} \lesssim_{d,p} u_L^{p-d/2} \asymp \left(h_L^{\mathrm{lam,crit}}\right)^{p-d/2}.$$

Now split the error into large and small alphabets. For the large-alphabet part,

$$\frac{1}{\sqrt{N\log N}} \sum_{\ell:\, M_\ell \ge N^\alpha} r_\ell^p\sqrt{M_\ell} \lesssim_{d,p} \frac{\left(h_L^{\mathrm{lam,crit}}\right)^{p-d/2}}{\sqrt{N\log N}} \lesssim_{d,p,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^p,$$

because $N\log N \ge C_{d,p,c_\star}(h_L^{\mathrm{lam,crit}})^{-d}$.

For the small-alphabet part,

$$\frac{1}{\sqrt{N}} \sum_{\ell:\, M_\ell < N^\alpha} r_\ell^p\sqrt{M_\ell} \le N^{(\alpha-1)/2}\sum_{\ell=0}^{L} r_\ell^p.$$

Since $r_{\ell+1} \le r_\ell/8$ for $0 \le \ell \le L-2$ and $r_L = r_{L-1}/8$, the geometric series $\sum_{\ell=0}^{L} r_\ell^p$ is bounded by a constant depending only on p. Moreover, $(1-\alpha)d/2 > p$, so

$$N^{(\alpha-1)/2} = O\Big((h_L^{\mathrm{lam,crit}})^p\Big)$$

under the relation $(h_L^{\mathrm{lam,crit}})^{-d} \lesssim N\log N$. Consequently

$$\sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\left|\widehat{T}_L^{\mathrm{lam,crit}} - W_p(P,Q)^p\right| \lesssim_{d,p,c_\star} \left(h_L^{\mathrm{lam,crit}}\right)^p.$$

Exactly as before,

$$\mathbb{E}\Big(\widehat{T}_L^{\mathrm{lam,crit}} - W_p(P,Q)^p\Big)^2 \le \Big(\sum_{\ell=0}^{L} \frac{r_\ell^p}{2}\sqrt{\mathbb{E}\Big(\widetilde{L}_\ell - L_\ell^{\mathrm{lam,crit}}(P,Q)\Big)^2}\Big)^2 \lesssim_{d,p,c_\star} \big(h_L^{\mathrm{lam,crit}}\big)^{2p}.$$

Since $|u-v|^p \le |u^p - v^p|$ for every $u, v \ge 0$,

$$\mathbb{E}\Big|\hat{W}_L^{\mathrm{lam,crit}} - W_p(P,Q)\Big| \le \Big(\mathbb{E}\Big|\widehat{T}_L^{\mathrm{lam,crit}} - W_p(P,Q)^p\Big|\Big)^{1/p} \lesssim_{d,p,c_\star} h_L^{\mathrm{lam,crit}},$$

and similarly

$$\mathbb{E}\Big(\hat{W}_L^{\mathrm{lam,crit}} - W_p(P,Q)\Big)^2 \le \Big(\mathbb{E}\Big(\widehat{T}_L^{\mathrm{lam,crit}} - W_p(P,Q)^p\Big)^2\Big)^{1/p} \lesssim_{d,p,c_\star} \big(h_L^{\mathrm{lam,crit}}\big)^2.$$

This proves part (ii).

For part (iii), restrict to the subclass in which all masses w_σ vanish off the deepest level $\Sigma_L^{\mathrm{lam,crit}}$. Then Theorem 5.34 reduces to

$$W_p(P,Q) = h_L^{\mathrm{lam,crit}}\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p}$$

over a known alphabet of cardinality

$$2M_L \asymp_{d,c_\star} \big(h_L^{\mathrm{lam,crit}}\big)^{-d} \asymp N\log N.$$

By [5, Theorem 3], the minimax squared error for estimating $\|P-Q\|_1$ on this alphabet is bounded below by a positive constant depending only on d and $c_\star$. Since the parameter is bounded by 2, the minimax absolute risk is also bounded below by a positive constant. If $\hat{W}$ is any estimator of $W_p(P,Q)$ on $\mathcal{C}_L^{\mathrm{lam,crit}}$, define

$$\widehat{L} := 2\Big(\frac{\hat{W} \wedge (2^{1/p}h_L^{\mathrm{lam,crit}})}{h_L^{\mathrm{lam,crit}}}\Big)^p.$$

The map $w \mapsto 2(w/h_L^{\mathrm{lam,crit}})^p$ is $C_p(h_L^{\mathrm{lam,crit}})^{-1}$-Lipschitz on $[0, 2^{1/p}h_L^{\mathrm{lam,crit}}]$. Therefore a constant lower bound for estimating $\|P-Q\|_1$ forces

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \gtrsim_{d,p,c_\star} h_L^{\mathrm{lam,crit}},$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_L^{\mathrm{lam,crit}}} \mathbb{E}\Big(\hat{W} - W_p(P,Q)\Big)^2 \gtrsim_{d,p,c_\star} \big(h_L^{\mathrm{lam,crit}}\big)^2.$$

Together with part (ii), this proves part (iii). □

Corollary 5.36 (The accelerating-branch model is a special case)**.** *Take*

$$B_\ell := 2^{\ell+4}, \qquad \varepsilon_\ell := \frac{2}{B_\ell} = 2^{-\ell-3},$$

and at every node choose all admissible children

$$\mathcal{D}_\sigma = \{2,3,\dots,B_\ell - 1\} \times \{0,1,\dots,B_\ell-1\}^{d-1}.$$

Then the resulting hierarchy satisfies the assumptions above, with

$$c_\star = \prod_{\ell=0}^{\infty}\Big(1 - \frac{2}{B_\ell}\Big) > 0.$$

Hence Theorem 5.35 recovers the explicit accelerating-branch critical-density laminar class.

6 Tree exactness, catalogs, and stochastic envelopes

The packed direct-sum theorem and the laminar theorems show that large genuinely Euclidean subclasses of the critical core are already solvable exactly once the transport can be frozen either into disjoint multiscale blocks or into explicit laminar overlap, including an exact critical-density nested class. The next results identify the complementary tree and sparse-cycle geometry of the discrete W_1 core. Every *individual* Euclidean finite-support instance is already exact on some spanning tree. More positively, every fixed critical hierarchical tree metric is itself estimable at the sharp $O(h)$ scale, every finite catalog of dominating such trees satisfies an oracle inequality against its best member, and one may even add a shortcut system of total tree footprint $O(h\sqrt{N})$ without losing the target law. At the same time, the full Euclidean grid transportation norm is globally captured by dominating random trees only up to a sharp logarithmic factor. Thus the unresolved difficulty is neither the absence of tree structure nor the impossibility of sparse graph overlap; it is the genuinely macroscopic cycle space of the unrestricted Euclidean core.

6.1 Pointwise Euclidean tree exactness

Let $X \subset \mathbb{R}^d$ be finite. For a function $f : X \to \mathbb{R}$ with total sum zero, define the Euclidean transportation-cost norm

$$\|f\|_{\mathrm{TC}(X)} := \inf\Big\{\sum_{x,y\in X} P(x,y)\|x-y\|_2 : P \ge 0,\ f = \sum_{x,y\in X} P(x,y)(\delta_x - \delta_y)\Big\}.$$

If $P, Q \in \mathcal{P}(X)$ then

$$W_1(P,Q) = \|P-Q\|_{\mathrm{TC}(X)}.$$

Proposition 6.1 (Pointwise Euclidean tree exactness)**.** *Let $X \subset \mathbb{R}^d$ be finite and let $f : X \to \mathbb{R}$ satisfy $\sum_{x\in X} f(x) = 0$. Then there exists a spanning tree T on the vertex set X, with edge lengths*

$$\ell(uv) := \|u - v\|_2 \qquad (uv \in E(T)),$$

such that

$$\|f\|_{\mathrm{TC}(X)} = \|f\|_{\mathrm{TC}(T)}.$$

In particular, every discrete Euclidean W_1 instance is exact on some spanning tree.

Proof. Write

$$X_+ := \{x \in X : \ f(x) > 0\}, \qquad X_- := \{y \in X : \ f(y) < 0\}.$$

By splitting each transport from a positive site to another positive site (and similarly on the negative side), one may restrict without loss of generality to transportation plans

$$P : X_+ \times X_- \to [0, \infty)$$

satisfying

$$\sum_{y\in X_-} P(x, y) = f(x) \quad (x \in X_+), \qquad \sum_{x\in X_+} P(x, y) = -f(y) \quad (y \in X_-).$$

The cost

$$\sum_{x\in X_+} \sum_{y\in X_-} P(x, y)\|x - y\|_2$$

is the objective of a classical transportation linear program. Choose an optimal basic feasible solution $P^\star$. Its support graph on the bipartite vertex set $X_+ \sqcup X_-$ is acyclic: if the support contained a cycle, one could perturb $P^\star$ by the usual alternating-sign cycle flow and preserve all marginals, contradicting basicness. Hence the support of $P^\star$ is a forest F.

Extend F by adding arbitrary edges on X until it becomes a spanning tree T. Because $F \subseteq T$, the plan $P^\star$ is also feasible as a transportation plan on T and has exactly the same cost there, since each used edge $xy \in F$ has length $\|x - y\|_2$ both in X and in T. Therefore

$$\|f\|_{\mathrm{TC}(T)} \le \|f\|_{\mathrm{TC}(X)}.$$

Conversely, for any two vertices $u, v \in X$, the unique path distance $d_T(u, v)$ in the tree T satisfies

$$d_T(u, v) \ge \|u - v\|_2$$

by the Euclidean triangle inequality. Hence every transport plan on T has cost at least its cost in the complete Euclidean metric on X, so

$$\|f\|_{\mathrm{TC}(T)} \ge \|f\|_{\mathrm{TC}(X)}.$$

The two inequalities imply equality. □

6.2 Critical hierarchical tree metrics

The pointwise exact-tree proposition becomes statistically useful once the exact tree carries a critical multiscale hierarchy. We record the abstract tree formula first.

Let X be a finite set and let T be a rooted weighted tree whose leaves are exactly the points of X. Assume that every nonleaf vertex lies on one of the levels

$$V_0 = \{\rho\}, \ V_1, \dots , V_L,$$

that every edge joining a vertex in $V_{\ell-1}$ to one in V_ℓ has the same length $r_\ell > 0$, and that every leaf lies in V_L. For $v \in V_\ell$ let $A_v \subseteq X$ be the set of descendant leaves below v. For $P, Q \in \mathcal{P}(X)$ define

$$L^T_\ell(P, Q) := \sum_{v\in V_\ell} \big|P(A_v) - Q(A_v)\big| \qquad (1 \le \ell \le L),$$

and write $W^T_1(P, Q)$ for the transportation cost induced by the tree metric on X.

Proposition 6.2 (Tree cut formula on a leveled hierarchy)**.** *For every $P, Q \in \mathcal{P}(X)$,*

$$W^T_1(P, Q) = \sum_{\ell=1}^{L} r_\ell \, L^T_\ell(P, Q) = \sum_{\ell=1}^{L} r_\ell \sum_{v\in V_\ell} \big|P(A_v) - Q(A_v)\big|.$$

Proof. Let

$$f := P - Q, \qquad s_v := f(A_v) = P(A_v) - Q(A_v) \qquad (v \in V_\ell, \ 1 \le \ell \le L).$$

For the edge $e(v)$ joining v to its parent, every transport plan on the tree must send net amount s_v across $e(v)$, because A_v can exchange mass with its complement only through this edge. Hence every feasible transport on T has cost at least

$$\sum_{\ell=1}^{L} r_\ell \sum_{v\in V_\ell} |s_v|.$$

Conversely, one can balance the tree recursively from the leaves to the root. If $s_v > 0$, move s_v units of excess mass from the subtree A_v to the parent side across $e(v)$; if $s_v < 0$, move $|s_v|$ units in the opposite direction. After performing this on all edges from level L upward to level 1, every subtree is balanced and the total cost is exactly

$$\sum_{\ell=1}^{L} r_\ell \sum_{v\in V_\ell} |s_v|.$$

This proves equality. □

Theorem 6.3 (Critical hierarchical-tree upper law)**.** *Assume $d > 2$ and fix $K < \infty$. There exists a constant $C_{K,d} < \infty$ such that the following holds.*

Let $h \in (0, 1/8]$ and let T_h be a rooted weighted tree whose leaves form a known set $X_h \subset [0,1]^d$. Assume that T_h has depth L_h, level sets $V_{\ell,h}$, and level lengths $r_{\ell,h}$ satisfying

$$r_{1,h} \le 1, \qquad h \le r_{L_h,h} \le 8h, \qquad r_{\ell+1,h} \le \frac{r_{\ell,h}}{8} \quad (1 \le \ell < L_h),$$

and

$$M_{\ell,h} := |V_{\ell,h}| \le K\, r_{\ell,h}^{-d} \qquad (1 \le \ell \le L_h).$$

If

$$N \log N \ge C_{K,d}\, h^{-d},$$

then there exists an estimator $\hat{W}_h^T$ such that

$$\sup_{P,Q\in\mathcal{P}(X_h)} \mathbb{E}\left|\hat{W}_h^T - W_1^{T_h}(P,Q)\right| \le C_{K,d}\, h,$$

and

$$\sup_{P,Q\in\mathcal{P}(X_h)} \mathbb{E}\left(\hat{W}_h^T - W_1^{T_h}(P,Q)\right)^2 \le C_{K,d}\, h^2.$$

Proof. For each level ℓ, let

$$\Pi_{\ell,h} : X_h \to V_{\ell,h}$$

send each leaf to its unique ancestor at level ℓ, and write

$$P_{\ell,h} := (\Pi_{\ell,h})_\# P, \qquad Q_{\ell,h} := (\Pi_{\ell,h})_\# Q.$$

Because the sets $(A_v)_{v\in V_{\ell,h}}$ partition X_h,

$$\|P_{\ell,h} - Q_{\ell,h}\|_1 = \sum_{v\in V_{\ell,h}} |P(A_v) - Q(A_v)| = L_\ell^{T_h}(P,Q).$$

Therefore Theorem 6.2 gives

$$W_1^{T_h}(P,Q) = \sum_{\ell=1}^{L_h} r_{\ell,h}\, \|P_{\ell,h} - Q_{\ell,h}\|_1.$$

Choose

$$\alpha \in \left(0, 1 - \frac{2}{d}\right).$$

For each level ℓ, estimate $\|P_{\ell,h} - Q_{\ell,h}\|_1$ as follows. If $M_{\ell,h} \ge N^\alpha$, apply the large-alphabet estimator of [5, Theorem 6] on the known alphabet $V_{\ell,h}$ to obtain an estimator $\widehat{L}_{\ell,h}$ satisfying

$$\sup_{P,Q\in\mathcal{P}(X_h)} \mathbb{E}\left(\widehat{L}_{\ell,h} - \|P_{\ell,h} - Q_{\ell,h}\|_1\right)^2 \lesssim_d \frac{M_{\ell,h}}{N \log N}.$$

If $M_{\ell,h} < N^\alpha$, use the empirical plug-in estimator on the same alphabet; the standard multinomial bound gives

$$\sup_{P,Q\in\mathcal{P}(X_h)} \mathbb{E}\left(\widehat{L}_{\ell,h} - \|P_{\ell,h} - Q_{\ell,h}\|_1\right)^2 \lesssim_d \frac{M_{\ell,h}}{N}.$$

Clip each estimator by

$$\widetilde{L}_{\ell,h} := 0 \vee \widehat{L}_{\ell,h} \wedge 2.$$

Clipping cannot increase squared error.

Define

$$\hat{W}_h^T := \sum_{\ell=1}^{L_h} r_{\ell,h}\, \widetilde{L}_{\ell,h}.$$

Then

$$\hat{W}_h^T - W_1^{T_h}(P,Q) = \sum_{\ell=1}^{L_h} r_{\ell,h} \left(\widetilde{L}_{\ell,h} - \|P_{\ell,h} - Q_{\ell,h}\|_1\right).$$

Hence, by Cauchy–Schwarz level by level,

$$\mathbb{E}\left|\hat{W}_h^T - W_1^{T_h}(P,Q)\right| \le \sum_{\ell=1}^{L_h} r_{\ell,h} \sqrt{\mathbb{E}\left(\widetilde{L}_{\ell,h} - \|P_{\ell,h} - Q_{\ell,h}\|_1\right)^2}.$$

For the large-alphabet part,

$$\frac{1}{\sqrt{N\log N}} \sum_{\ell:\, M_{\ell,h}\ge N^\alpha} r_{\ell,h} \sqrt{M_{\ell,h}} \le \frac{\sqrt{K}}{\sqrt{N \log N}} \sum_{\ell=1}^{L_h} r_{\ell,h}^{1-d/2}.$$

Set

$$q := \frac{d}{2} - 1 > 0.$$

Since $r_{\ell+1,h} \le r_{\ell,h}/8$, one has

$$r_{\ell,h}^{-q} \le 8^{-q(L_h-\ell)} r_{L_h,h}^{-q},$$

and therefore

$$\sum_{\ell=1}^{L_h} r_{\ell,h}^{1-d/2} = \sum_{\ell=1}^{L_h} r_{\ell,h}^{-q} \lesssim_d r_{L_h,h}^{-q} \lesssim_d h^{1-d/2}.$$

Thus

$$\frac{1}{\sqrt{N\log N}}\sum_{\ell:\,M_{\ell,h}\geq N^\alpha} r_{\ell,h}\sqrt{M_{\ell,h}} \lesssim_{K,d} \frac{h^{1-d/2}}{\sqrt{N\log N}} \lesssim_{K,d} h$$

under the hypothesis $N\log N \geq C_{K,d}h^{-d}$.

For the small-alphabet part,

$$\frac{1}{\sqrt{N}}\sum_{\ell:\,M_{\ell,h}<N^\alpha} r_{\ell,h}\sqrt{M_{\ell,h}} \leq N^{(\alpha-1)/2}\sum_{\ell=1}^{L_h} r_{\ell,h}.$$

Because $r_{1,h}\leq 1$ and $r_{\ell+1,h}\leq r_{\ell,h}/8$,

$$\sum_{\ell=1}^{L_h} r_{\ell,h} \lesssim 1.$$

Moreover, since $(1-\alpha)d/2>1$ and $h^{-d}\lesssim N\log N$, the logarithmic factor is harmless and

$$N^{(\alpha-1)/2}\lesssim_{d,\alpha} h.$$

Consequently

$$\sup_{P,Q\in\mathcal{P}(X_h)} \mathbb{E}\Big|\hat W_h^T - W_1^{T_h}(P,Q)\Big| \lesssim_{K,d} h.$$

Exactly the same decomposition gives the squared bound:

$$\mathbb{E}\Big(\hat W_h^T - W_1^{T_h}(P,Q)\Big)^2 \leq \Bigg(\sum_{\ell=1}^{L_h} r_{\ell,h}\sqrt{\mathbb{E}\Big(\widetilde L_{\ell,h}-\|P_{\ell,h}-Q_{\ell,h}\|_1\Big)^2}\Bigg)^2 \lesssim_{K,d} h^2.$$

This proves the theorem. □

Corollary 6.4 (Euclidean exact hierarchical-tree classes)**.** *Assume $d>2$ and let T_h satisfy the hypotheses of Theorem 6.3. Assume moreover that its leaf set is $X_h\subset[0,1]^d$ and that the tree metric dominates Euclidean distance:*

$$d_{T_h}(x,y)\geq \|x-y\|_2 \qquad (x,y\in X_h).$$

Define the Euclidean exact-tree class

$$\mathcal{C}_h(T_h) := \Big\{(P,Q)\in\mathcal{P}(X_h)^2:\ W_1(P,Q)=W_1^{T_h}(P,Q)\Big\}.$$

If $N\log N\geq C_{K,d}h^{-d}$, then the estimator from Theorem 6.3 satisfies

$$\sup_{(P,Q)\in\mathcal{C}_h(T_h)} \mathbb{E}\Big|\hat W_h^T - W_1(P,Q)\Big| \lesssim_{K,d} h,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h(T_h)} \mathbb{E}\Big(\hat W_h^T - W_1(P,Q)\Big)^2 \lesssim_{K,d} h^2.$$

Proof. On $\mathcal{C}_h(T_h)$ one has $W_1(P,Q)=W_1^{T_h}(P,Q)$ by definition, so Theorem 6.3 applies verbatim. □

6.3 Finite dominating hierarchical tree catalogs

The stochastic-tree barrier rules out uniform convex averages of tree norms. A different mechanism remains viable: a nonlinear minimum over a finite catalog of sharp tree estimators. The next result is an oracle inequality of exactly this type.

Theorem 6.5 (Finite hierarchical-tree catalog oracle inequality)**.** *Assume $d>2$ and fix an integer $K\geq 1$. For $1\leq k\leq K$, let $T_h^{(k)}$ be a rooted weighted tree on the same leaf set $X_h\subset[0,1]^d$, and assume that each $T_h^{(k)}$ satisfies the hypotheses of Theorem 6.3 with constants bounded uniformly in k. Assume also that each tree metric dominates Euclidean distance on X_h:*

$$d_{T_h^{(k)}}(x,y)\geq\|x-y\|_2 \qquad (x,y\in X_h,\ 1\leq k\leq K).$$

Let $\hat W_h^{(k)}$ be the estimator from Theorem 6.3 for the metric $T_h^{(k)}$, and define

$$\hat W_h^{\mathrm{cat}} := \min_{1\leq k\leq K}\hat W_h^{(k)}.$$

For $(P,Q)\in\mathcal{P}(X_h)^2$, define the catalog approximation error

$$\Delta_h^{\mathrm{cat}}(P,Q) := \min_{1\leq k\leq K}\Big\{W_1^{T_h^{(k)}}(P,Q)-W_1(P,Q)\Big\}\geq 0.$$

Then

$$\mathbb{E}\Big|\hat W_h^{\mathrm{cat}}-W_1(P,Q)\Big| \leq \Delta_h^{\mathrm{cat}}(P,Q)+C_{K,d}\,h,$$

and

$$\mathbb{E}\Big(\hat W_h^{\mathrm{cat}}-W_1(P,Q)\Big)^2 \leq 2\Delta_h^{\mathrm{cat}}(P,Q)^2+C_{K,d}\,h^2.$$

Proof. Choose $k_\star\in\{1,\dots,K\}$ attaining the minimum in the definition of $\Delta_h^{\mathrm{cat}}(P,Q)$.

For the positive deviation,

$$\hat W_h^{\mathrm{cat}}-W_1(P,Q)\leq \hat W_h^{(k_\star)}-W_1(P,Q) = \Big(\hat W_h^{(k_\star)}-W_1^{T_h^{(k_\star)}}(P,Q)\Big)+\Delta_h^{\mathrm{cat}}(P,Q).$$

Hence

$$\mathbb{E}\Big(\hat W_h^{\mathrm{cat}}-W_1(P,Q)\Big)_+ \leq \Delta_h^{\mathrm{cat}}(P,Q)+C_d h$$

and

$$\mathbb{E}\Big(\hat{W}_h^{\mathrm{cat}} - W_1(P,Q)\Big)_+^2 \le 2\Delta_h^{\mathrm{cat}}(P,Q)^2 + C_d h^2$$

by Theorem 6.3.

For the negative deviation, the domination $W_1(P,Q) \le W_1^{T_h^{(k)}}(P,Q)$ gives

$$\Big(W_1(P,Q) - \hat{W}_h^{\mathrm{cat}}\Big)_+ \le \max_{1\le k\le K} \Big(W_1^{T_h^{(k)}}(P,Q) - \hat{W}_h^{(k)}\Big)_+.$$

Therefore

$$\mathbb{E}\Big(W_1(P,Q) - \hat{W}_h^{\mathrm{cat}}\Big)_+ \le \sum_{k=1}^{K} \mathbb{E}\Big|W_1^{T_h^{(k)}}(P,Q) - \hat{W}_h^{(k)}\Big| \le C_{K,d} h,$$

and, using $\max a_k^2 \le \sum a_k^2$,

$$\mathbb{E}\Big(W_1(P,Q) - \hat{W}_h^{\mathrm{cat}}\Big)_+^2 \le \sum_{k=1}^{K} \mathbb{E}\Big(W_1^{T_h^{(k)}}(P,Q) - \hat{W}_h^{(k)}\Big)^2 \le C_{K,d} h^2.$$

Combining the positive and negative parts proves both claims. □

Corollary 6.6 (Approximate catalog criterion)**.** *Under the assumptions of Theorem 6.5, let $\mathcal{C}_h \subseteq \mathcal{P}(X_h)^2$ satisfy*

$$\sup_{(P,Q)\in\mathcal{C}_h} \Delta_h^{\mathrm{cat}}(P,Q) \le Ah$$

for some $A \ge 0$. Then

$$\sup_{(P,Q)\in\mathcal{C}_h} \mathbb{E}\Big|\hat{W}_h^{\mathrm{cat}} - W_1(P,Q)\Big| \le C_{A,K,d}\, h,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h} \mathbb{E}\Big(\hat{W}_h^{\mathrm{cat}} - W_1(P,Q)\Big)^2 \le C_{A,K,d}\, h^2.$$

In particular, the same conclusion holds on the exact union class

$$\bigcup_{k=1}^{K} \mathcal{C}_h(T_h^{(k)}),$$

for which $A = 0$.

Proof. Insert the approximation bound into Theorem 6.5. □

6.4 Sparse shortcut graphs beyond trees

The exact tree classes solve the zero-cycle regime. The next theorem shows that one may add a quantitatively controlled family of shortcut edges and still retain the sharp critical law. The reason is structural as well as statistical. The tree part still requires the full large-alphabet machinery of Theorem 6.3, but the cycle correction depends only on the subtree imbalances along the union of the shortcut paths, and those are ordinary one-dimensional empirical means. Thus a genuinely overlapping graph geometry becomes tractable as long as the total tree footprint of the shortcuts remains below the critical noise budget.

Fix a rooted weighted tree T with finite leaf set X. For every non-root edge $e \in E(T)$, let $A_e \subseteq X$ denote the descendant leaf set below e, and let $\ell(e) > 0$ denote the length of e. Write

$$\mathscr{F}(X) := \Big\{f : X \to \mathbb{R} : \sum_{x\in X} f(x) = 0\Big\}, \qquad s_e(f) := f(A_e) \qquad (f \in \mathscr{F}(X)).$$

By the tree formula of Theorem 6.2,

$$\|f\|_{\mathrm{TC}(X,d_T)} = \sum_{e\in E(T)} \ell(e)\, |s_e(f)| \qquad (f \in \mathscr{F}(X)).$$

Lemma 6.7 (Edge-flow representation of graph transportation)**.** *Let $H = (V, E, \ell)$ be a finite connected weighted graph, endowed with its shortest-path metric d_H on V. Choose an arbitrary orientation $e = (e^+, e^-)$ of each edge. Then for every $f \in \mathscr{F}(V)$,*

$$\|f\|_{\mathrm{TC}(V,d_H)} = \inf\Big\{\sum_{e\in E} \ell(e)\, |u_e| : \ f = \sum_{e\in E} u_e(\delta_{e^+} - \delta_{e^-})\Big\}.$$

Proof. This is standard minimum-cost-flow geometry for graph metrics. For completeness we give a direct proof.

Let π be any transport plan for f on (V, d_H). For each ordered pair $(x, y) \in V \times V$, choose a shortest path γ_{xy} in H joining x to y. Sending the mass $\pi(x, y)$ along γ_{xy} produces edge coefficients (u_e) such that

$$f = \sum_{e\in E} u_e(\delta_{e^+} - \delta_{e^-}),$$

and whose total edge cost is at most

$$\sum_{x,y\in V} \pi(x,y)\, d_H(x,y).$$

Taking the infimum over π yields

$$\inf_{(u_e)} \sum_{e\in E} \ell(e)|u_e| \le \|f\|_{\mathrm{TC}(V,d_H)}.$$

Conversely, fix edge coefficients (u_e) satisfying

$$f = \sum_{e \in E} u_e(\delta_{e^+} - \delta_{e^-}).$$

Replace every edge $e = (e^+, e^-)$ by the directed arc $e^+ \to e^-$ carrying mass $(u_e)_+$ and the reverse directed arc $e^- \to e^+$ carrying mass $(u_e)_-$. This yields a nonnegative directed flow on the bidirected graph whose divergence is exactly f.

Repeatedly extract a directed path from some vertex of positive divergence to some vertex of negative divergence, subtracting along that path the largest amount compatible with nonnegativity. After finitely many steps all divergences are exhausted, and the remaining directed flow is a nonnegative circulation. Decompose that circulation into directed cycles and discard those cycles. We are left with directed path flows

$$\gamma_r : \; x_r \to y_r, \qquad t_r > 0,$$

such that

$$\sum_r t_r(\delta_{x_r} - \delta_{y_r}) = f$$

and

$$\sum_r t_r \operatorname{length}(\gamma_r) \le \sum_{e \in E} \ell(e)|u_e|.$$

Define a transport plan on $V \times V$ by

$$\pi(x, y) := \sum_{r:\,(x_r, y_r) = (x, y)} t_r.$$

Then π transports the positive part of f to its negative part, and because $d_H(x_r, y_r) \le \operatorname{length}(\gamma_r)$ for every r,

$$\sum_{x,y \in V} \pi(x, y)\, d_H(x, y) \le \sum_r t_r \operatorname{length}(\gamma_r) \le \sum_{e \in E} \ell(e)|u_e|.$$

Taking the infimum over all feasible edge coefficients (u_e) gives

$$\|f\|_{\mathrm{TC}(V, d_H)} \le \inf_{(u_e)} \sum_{e \in E} \ell(e)|u_e|,$$

which proves the claim. □

Proposition 6.8 (Exact shortcut reduction relative to a rooted tree)**.** *Let T be a rooted weighted tree with leaf set X, and let H be obtained from T by adjoining shortcut edges*

$$g_j = \{a_j, b_j\} \qquad (1 \le j \le c)$$

of lengths $\lambda_j > 0$. Choose an orientation $a_j \to b_j$ for each shortcut. For every tree edge e, define

$$\varepsilon_{e,j} := \mathbf{1}_{\{b_j \in A_e,\ a_j \notin A_e\}} - \mathbf{1}_{\{a_j \in A_e,\ b_j \notin A_e\}} \in \{-1, 0, 1\}.$$

Then for every $f \in \mathscr{F}(X)$,

$$\|f\|_{\mathrm{TC}(X, d_H)} = \min_{z \in \mathbb{R}^c} \left\{ \sum_{e \in E(T)} \ell(e) \left| s_e(f) + \sum_{j=1}^{c} \varepsilon_{e,j} z_j \right| + \sum_{j=1}^{c} \lambda_j |z_j| \right\}.$$

Let

$$U := \Big\{ e \in E(T) : \; \exists j \text{ with } \varepsilon_{e,j} \ne 0 \Big\}, \qquad \operatorname{Len}^{\mathrm{sh}}(H; T) := \sum_{e \in U} \ell(e).$$

Define

$$\Psi_H(u) := \min_{z \in \mathbb{R}^c} \left\{ \sum_{e \in U} \ell(e) \left(\left| u_e + \sum_{j=1}^{c} \varepsilon_{e,j} z_j \right| - |u_e| \right) + \sum_{j=1}^{c} \lambda_j |z_j| \right\}, \qquad u = (u_e)_{e \in U} \in \mathbb{R}^U.$$

Then

$$\|f\|_{\mathrm{TC}(X, d_H)} = \|f\|_{\mathrm{TC}(X, d_T)} + \Psi_H\big((s_e(f))_{e \in U}\big),$$

and for all $u, v \in \mathbb{R}^U$,

$$|\Psi_H(u) - \Psi_H(v)| \le 2 \sum_{e \in U} \ell(e)\, |u_e - v_e|.$$

Proof. Orient every tree edge toward the root. By Theorem 6.7, the graph transportation norm on (X, d_H) is the minimum of

$$\sum_{e \in E(T)} \ell(e)\, |u_e| + \sum_{j=1}^{c} \lambda_j |z_j|$$

over all coefficients $(u_e)_{e \in E(T)}$ and $z \in \mathbb{R}^c$ satisfying

$$f = \sum_{e \in E(T)} u_e(\delta_{e^+} - \delta_{e^-}) + \sum_{j=1}^{c} z_j(\delta_{a_j} - \delta_{b_j}).$$

For fixed z, the tree coefficients must therefore realize the signed leaf function

$$f - \sum_{j=1}^{c} z_j(\delta_{a_j} - \delta_{b_j}).$$

On a rooted tree the coefficient on an edge e is uniquely determined by the mass on the descendant side:

$$u_e = \Big(f - \sum_{j=1}^{c} z_j(\delta_{a_j} - \delta_{b_j})\Big)(A_e) = s_e(f) + \sum_{j=1}^{c} \varepsilon_{e,j} z_j.$$

Substituting this unique choice proves the first formula.

Now split the sum over $E(T)$ into the edges in U and in $E(T) \setminus U$. On the latter set every coefficient $\varepsilon_{e,j}$ vanishes, so

$$\|f\|_{\mathrm{TC}(X,d_H)} = \sum_{e \notin U} \ell(e)|s_e(f)| + \min_{z \in \mathbb{R}^c} \left\{ \sum_{e \in U} \ell(e) \Big| s_e(f) + \sum_{j=1}^{c} \varepsilon_{e,j} z_j \Big| + \sum_{j=1}^{c} \lambda_j |z_j| \right\}.$$

Since

$$\|f\|_{\mathrm{TC}(X,d_T)} = \sum_{e \notin U} \ell(e)|s_e(f)| + \sum_{e \in U} \ell(e)|s_e(f)|,$$

the claimed decomposition with Ψ_H follows.

For the Lipschitz bound, define

$$\Phi_H(u) := \min_{z \in \mathbb{R}^c} \left\{ \sum_{e \in U} \ell(e) \Big| u_e + \sum_{j=1}^{c} \varepsilon_{e,j} z_j \Big| + \sum_{j=1}^{c} \lambda_j |z_j| \right\}.$$

For fixed z, the displayed function is $\sum_{e \in U} \ell(e)|u_e - v_e|$-Lipschitz in u. Taking the infimum over z preserves the same Lipschitz constant, hence

$$|\Phi_H(u) - \Phi_H(v)| \le \sum_{e \in U} \ell(e)\,|u_e - v_e|.$$

Because

$$\Psi_H(u) = \Phi_H(u) - \sum_{e \in U} \ell(e)|u_e|,$$

the triangle inequality gives

$$|\Psi_H(u) - \Psi_H(v)| \le |\Phi_H(u) - \Phi_H(v)| + \sum_{e \in U} \ell(e)\,\big||u_e| - |v_e|\big| \le 2 \sum_{e \in U} \ell(e)|u_e - v_e|.$$

□

Theorem 6.9 (Critical law on exact sparse-shortcut graph classes)**.** *Assume $d > 2$ and fix $K < \infty$. There exists a constant $C_{K,d} < \infty$ such that the following holds.*

Let $h \in (0, 1/8]$ and let T_h satisfy the hypotheses of Theorem 6.3, with leaf set

$$X_h \subset [0,1]^d.$$

Let H_h be obtained from T_h by adjoining shortcut edges

$$g_{j,h} = \{a_{j,h}, b_{j,h}\} \qquad (1 \le j \le c_h)$$

of lengths $\lambda_{j,h} > 0$, and let d_{H_h} be the resulting shortest-path metric on X_h. Assume that

$$d_{H_h}(x,y) \ge \|x - y\|_2 \qquad (x, y \in X_h).$$

Let $U_h \subseteq E(T_h)$ be the union of the tree edges crossed by the shortcut system, and define the shortcut footprint

$$\mathrm{Len}_h^{\mathrm{sh}} := \sum_{e \in U_h} \ell_h(e).$$

If

$$N \log N \ge C_{K,d}\, h^{-d},$$

then there exists an estimator $\hat{W}_h^H$ such that

$$\sup_{P,Q \in \mathcal{P}(X_h)} \mathbb{E}\Big|\hat{W}_h^H - W_1^{H_h}(P,Q)\Big| \le C_{K,d}\, h + C\, \frac{\mathrm{Len}_h^{\mathrm{sh}}}{\sqrt{N}},$$

and

$$\sup_{P,Q \in \mathcal{P}(X_h)} \mathbb{E}\Big(\hat{W}_h^H - W_1^{H_h}(P,Q)\Big)^2 \le C_{K,d}\, h^2 + C\, \frac{(\mathrm{Len}_h^{\mathrm{sh}})^2}{N},$$

for a universal constant $C < \infty$.

Consequently, on the Euclidean exact graph class

$$\mathcal{C}_h(H_h) := \Big\{(P,Q) \in \mathcal{P}(X_h)^2 :\ W_1(P,Q) = W_1^{H_h}(P,Q)\Big\},$$

the same estimator satisfies

$$\sup_{(P,Q) \in \mathcal{C}_h(H_h)} \mathbb{E}\Big|\hat{W}_h^H - W_1(P,Q)\Big| \le C_{K,d}\, h + C\, \frac{\mathrm{Len}_h^{\mathrm{sh}}}{\sqrt{N}},$$

and

$$\sup_{(P,Q) \in \mathcal{C}_h(H_h)} \mathbb{E}\Big(\hat{W}_h^H - W_1(P,Q)\Big)^2 \le C_{K,d}\, h^2 + C\, \frac{(\mathrm{Len}_h^{\mathrm{sh}})^2}{N}.$$

In particular, if

$$\mathrm{Len}_h^{\mathrm{sh}} \le A\,h\sqrt{N},$$

then

$$\sup_{(P,Q)\in\mathcal{C}_h(H_h)} \mathbb{E}\big|\hat W_h^H - W_1(P,Q)\big| \le C_{A,K,d}\,h,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h(H_h)} \mathbb{E}\big(\hat W_h^H - W_1(P,Q)\big)^2 \le C_{A,K,d}\,h^2.$$

Proof. Let $\hat W_h^T$ be the estimator from Theorem 6.3. For every tree edge $e\in U_h$, let $A_e \subseteq X_h$ denote the descendant leaf set below e and define

$$s_{e,h}(P,Q) := P(A_e) - Q(A_e), \qquad \hat s_{e,h} := P_n(A_e) - Q_m(A_e).$$

By Theorem 6.8,

$$W_1^{H_h}(P,Q) = W_1^{T_h}(P,Q) + \Psi_h\Big((s_{e,h}(P,Q))_{e\in U_h}\Big),$$

where Ψ_h is the shortcut correction functional associated with (H_h,T_h). Define the estimator

$$\hat W_h^H := \hat W_h^T + \Psi_h\Big((\hat s_{e,h})_{e\in U_h}\Big).$$

By the triangle inequality and the Lipschitz bound from Theorem 6.8,

$$\Big|\hat W_h^H - W_1^{H_h}(P,Q)\Big| \le \Big|\hat W_h^T - W_1^{T_h}(P,Q)\Big| + 2\sum_{e\in U_h} \ell_h(e)\,|\hat s_{e,h} - s_{e,h}(P,Q)|.$$

For every $e\in U_h$,

$$\mathbb{E}|\hat s_{e,h} - s_{e,h}(P,Q)| \le \sqrt{\mathbb{E}(\hat s_{e,h} - s_{e,h}(P,Q))^2} \le N^{-1/2},$$

because $P_n(A_e)$ and $Q_m(A_e)$ are Bernoulli averages with variances at most $1/(4n)$ and $1/(4m)$. Therefore

$$\mathbb{E}\Big|\hat W_h^H - W_1^{H_h}(P,Q)\Big| \le \mathbb{E}\Big|\hat W_h^T - W_1^{T_h}(P,Q)\Big| + 2\,\frac{\mathrm{Len}_h^{\mathrm{sh}}}{\sqrt{N}}.$$

Applying Theorem 6.3 yields the absolute-risk bound.

For the squared risk, use

$$\Big(\hat W_h^H - W_1^{H_h}(P,Q)\Big)^2 \le 2\Big(\hat W_h^T - W_1^{T_h}(P,Q)\Big)^2 + 8\Big(\sum_{e\in U_h}\ell_h(e)\,|\hat s_{e,h} - s_{e,h}(P,Q)|\Big)^2.$$

By Cauchy–Schwarz,

$$\Big(\sum_{e\in U_h}\ell_h(e)\,|\hat s_{e,h} - s_{e,h}(P,Q)|\Big)^2 \le \mathrm{Len}_h^{\mathrm{sh}} \sum_{e\in U_h}\ell_h(e)\,(\hat s_{e,h} - s_{e,h}(P,Q))^2.$$

Taking expectations and using

$$\mathbb{E}(\hat s_{e,h} - s_{e,h}(P,Q))^2 \le N^{-1}$$

for every $e\in U_h$, we obtain

$$\mathbb{E}\Big(\sum_{e\in U_h}\ell_h(e)\,|\hat s_{e,h} - s_{e,h}(P,Q)|\Big)^2 \le \frac{(\mathrm{Len}_h^{\mathrm{sh}})^2}{N}.$$

Combining this with Theorem 6.3 proves

$$\mathbb{E}\Big(\hat W_h^H - W_1^{H_h}(P,Q)\Big)^2 \le C_{K,d}h^2 + 8\,\frac{(\mathrm{Len}_h^{\mathrm{sh}})^2}{N},$$

and the stated bound follows after absorbing constants.

On $\mathcal{C}_h(H_h)$ one has $W_1(P,Q) = W_1^{H_h}(P,Q)$ by definition, so the same estimates apply verbatim. If $\mathrm{Len}_h^{\mathrm{sh}} \le Ah\sqrt{N}$, the shortcut term is at most a constant multiple of h in absolute loss and of h^2 in squared loss. ☐

Corollary 6.10 (Growing shortcut counts are still critical)**.** *Under the assumptions of Theorem 6.9, suppose that every shortcut joins two leaves of T_h. Then*

$$\mathrm{Len}_h^{\mathrm{sh}} \le \sum_{j=1}^{c_h} d_{T_h}(a_{j,h}, b_{j,h}) \le \frac{16}{7}\,c_h.$$

Consequently, if

$$c_h \le A\,h\sqrt{N},$$

then

$$\sup_{(P,Q)\in\mathcal{C}_h(H_h)} \mathbb{E}\big|\hat W_h^H - W_1(P,Q)\big| \le C_{A,K,d}\,h,$$

and

$$\sup_{(P,Q)\in\mathcal{C}_h(H_h)} \mathbb{E}\big(\hat W_h^H - W_1(P,Q)\big)^2 \le C_{A,K,d}\,h^2.$$

Proof. For any leaf x of T_h,

$$d_{T_h}(x,\mathrm{root}) = \sum_{\ell=1}^{L_h} r_{\ell,h} \le \sum_{\ell=1}^{\infty} 8^{-(\ell-1)} = \frac{8}{7},$$

because $r_{1,h}\le 1$ and $r_{\ell+1,h} \le r_{\ell,h}/8$. Hence any two leaves satisfy

$$d_{T_h}(x,y) \le \frac{16}{7}.$$

Summing over the shortcuts gives the bound on $\mathrm{Len}_h^{\mathrm{sh}}$, and the risk statement follows from Theorem 6.9. ☐

Remark 6.11 (What remains beyond sparse shortcut overlap). Within the present tree-plus-shortcut framework, Theorem 6.9 solves every exact Euclidean class whose total shortcut footprint is $O(h\sqrt{N})$. Since $h \asymp (N \log N)^{-1/d}$, this already allows

$$h\sqrt{N} = N^{1/2-1/d}(\log N)^{-1/d}$$

shortcuts of bounded tree length. Thus the unresolved core is not merely the presence of cycles. It is the genuinely macroscopic cycle space in which overlapping reuse occurs on a footprint much larger than the empirical $N^{-1/2}$ fluctuation scale.

6.5 A sharp stochastic-tree envelope law

We now freeze the support to the regular Euclidean grid

$$G_m := m^{-1}\{1, \dots, m\}^d \subset [0,1]^d, \qquad |G_m| = m^d.$$

Write $\mathscr{F}(G_m)$ for the vector space of all real functions on G_m with total sum zero. For any metric ρ on G_m, let

$$\|f\|_{\mathrm{TC}(G_m,\rho)}$$

denote the transportation-cost norm induced by ρ. When $\rho(x,y) = \|x-y\|_2$, we abbreviate this as

$$\|f\|_{\mathrm{TC}(G_m)}.$$

Lemma 6.12 (Transfer of the grid L_1-distortion barrier to the Euclidean metric)**.** *Assume $d \geq 2$. Then there exists $c_d > 0$ such that every embedding of the normed space*

$$(\mathscr{F}(G_m), \|\cdot\|_{\mathrm{TC}(G_m)})$$

into L_1 has distortion at least

$$c_d \log |G_m|.$$

Proof. Let ρ_1 and ρ_2 denote the ℓ_1 and Euclidean metrics on G_m. For all $x, y \in G_m$,

$$\rho_2(x,y) \leq \rho_1(x,y) \leq \sqrt{d}\,\rho_2(x,y).$$

By monotonicity of transportation-cost norms under pointwise comparison of ground metrics,

$$\|f\|_{\mathrm{TC}(G_m,\rho_2)} \leq \|f\|_{\mathrm{TC}(G_m,\rho_1)} \leq \sqrt{d}\,\|f\|_{\mathrm{TC}(G_m,\rho_2)} \qquad \forall f \in \mathscr{F}(G_m).$$

Therefore an embedding of the Euclidean transportation-cost space into L_1 with distortion D would induce an embedding of the ℓ_1-grid transportation-cost space into L_1 with distortion at most $\sqrt{d}\,D$. The lower bound of Gartland–Ostrovskii–Rabani–Young for d-dimensional grids then gives

$$\sqrt{d}\,D \geq c \log |G_m|$$

for a universal constant $c > 0$. Absorbing $\sqrt{d}$ into the constant yields the claim. ☐

If τ is a tree metric on G_m, then the transportation-cost norm on (G_m, τ) is L_1-embeddable: if T_τ is a weighted tree realizing τ and $A_e \subseteq G_m$ denotes one side of the leaf partition induced by deleting an edge $e \in E(T_\tau)$, then

$$\|f\|_{\mathrm{TC}(G_m,\tau)} = \sum_{e \in E(T_\tau)} \ell_\tau(e)\,|f(A_e)|.$$

Consequently, if μ is a probability distribution on tree metrics on G_m, the averaged envelope

$$N_\mu(f) := \int \|f\|_{\mathrm{TC}(G_m,\tau)}\, d\mu(\tau)$$

is again an L_1-embeddable norm, as an L_1-direct integral of tree norms.

Theorem 6.13 (Sharp stochastic-tree envelope law)**.** *Assume $d \geq 2$. Then there exist constants $c_d, C_d > 0$ such that the following holds for every $m \geq 2$.*

(i) *There exists a probability distribution μ_m on tree metrics τ on G_m such that*

$$\|x-y\|_2 \leq \tau(x,y) \qquad \forall x, y \in G_m,$$

and

$$\int \tau(x,y)\, d\mu_m(\tau) \leq C_d \log |G_m|\, \|x-y\|_2 \qquad \forall x, y \in G_m.$$

Hence

$$\|f\|_{\mathrm{TC}(G_m)} \leq N_{\mu_m}(f) \leq C_d \log |G_m|\, \|f\|_{\mathrm{TC}(G_m)} \qquad \forall f \in \mathscr{F}(G_m).$$

(ii) *Conversely, let μ be any probability distribution on tree metrics τ on G_m satisfying*

$$\tau(x,y) \geq \|x-y\|_2 \qquad \forall x, y \in G_m$$

for μ-almost every τ. If for some $D \geq 1$ one has

$$N_\mu(f) \leq D\,\|f\|_{\mathrm{TC}(G_m)} \qquad \forall f \in \mathscr{F}(G_m),$$

then

$$D \geq c_d \log |G_m|.$$

Therefore the optimal stochastic-tree/L_1 distortion of the Euclidean grid transportation-cost space is

$$\Theta_d(\log |G_m|).$$

Proof. For part (i), apply the theorem of Fakcharoenphol–Rao–Talwar [4] to the finite metric space $(G_m, \|\cdot\|_2)$. It gives a probability distribution μ_m on dominating tree metrics τ such that

$$\|x-y\|_2 \le \tau(x,y) \qquad \text{and} \qquad \int \tau(x,y)\, d\mu_m(\tau) \le C_d \log|G_m|\, \|x-y\|_2$$

for all $x, y \in G_m$.

Fix $f \in \mathscr{F}(G_m)$ and choose an optimal transport plan π_f for $\|f\|_{\mathrm{TC}(G_m)}$. For every tree metric τ,

$$\|f\|_{\mathrm{TC}(G_m,\tau)} \le \sum_{x,y\in G_m} \pi_f(x,y)\tau(x,y).$$

Integrating over μ_m and using Fubini's theorem yields

$$\begin{aligned} N_{\mu_m}(f) &\le \sum_{x,y\in G_m} \pi_f(x,y) \int \tau(x,y)\, d\mu_m(\tau) \\ &\le C_d \log|G_m| \sum_{x,y\in G_m} \pi_f(x,y)\|x-y\|_2 \\ &= C_d \log|G_m|\, \|f\|_{\mathrm{TC}(G_m)}. \end{aligned}$$

The lower bound

$$\|f\|_{\mathrm{TC}(G_m)} \le N_{\mu_m}(f)$$

is immediate from the pointwise domination $\tau \ge \|\cdot\|_2$.

For part (ii), the same domination implies

$$\|f\|_{\mathrm{TC}(G_m)} \le N_\mu(f) \qquad \forall f \in \mathscr{F}(G_m).$$

By the discussion preceding the theorem, N_μ is L_1-embeddable. Therefore

$$\|f\|_{\mathrm{TC}(G_m)} \le N_\mu(f) \le D\,\|f\|_{\mathrm{TC}(G_m)}$$

gives an embedding of $(\mathscr{F}(G_m), \|\cdot\|_{\mathrm{TC}(G_m)})$ into L_1 with distortion at most D. Applying Theorem 6.12 yields

$$D \ge c_d \log|G_m|.$$

□

Corollary 6.14 (Convex-combination barrier for L_1 surrogates)**.** *Assume $d \ge 2$. Let $N_1, \dots, N_K$ be norms on $\mathscr{F}(G_m)$ such that*

$$\|f\|_{\mathrm{TC}(G_m)} \le N_k(f) \qquad \forall f \in \mathscr{F}(G_m),\ \forall k,$$

and assume each N_k embeds isometrically into L_1. If weights $w_1, \dots, w_K \ge 0$ with $\sum_{k=1}^K w_k = 1$ satisfy

$$\sum_{k=1}^K w_k N_k(f) \le D\,\|f\|_{\mathrm{TC}(G_m)} \qquad \forall f \in \mathscr{F}(G_m),$$

then

$$D \ge c_d \log|G_m|.$$

In particular, no convex combination of finitely many tree norms or packed direct-sum L_1 models can approximate the full Euclidean grid transportation norm with distortion $o(\log|G_m|)$.

Proof. The norm

$$N(f) := \sum_{k=1}^K w_k N_k(f)$$

is L_1-embeddable and dominates $\|f\|_{\mathrm{TC}(G_m)}$. Hence

$$\|f\|_{\mathrm{TC}(G_m)} \le N(f) \le D\,\|f\|_{\mathrm{TC}(G_m)}$$

for all f, and Theorem 6.12 yields the conclusion. □

Theorem 6.15 (No exact or bounded-distortion difference-linear L_1 skeleton on the full Euclidean grid)**.** *Assume $d \ge 2$. Let $L \ge 1$, let $\lambda_1, \dots, \lambda_L > 0$, and let*

$$A_\ell : \mathscr{F}(G_m) \to \mathbb{R}^{M_\ell} \qquad (1 \le \ell \le L)$$

be linear maps. If for some $D \ge 1$ one has

$$\|f\|_{\mathrm{TC}(G_m)} \le \sum_{\ell=1}^L \lambda_\ell \|A_\ell f\|_1 \le D\,\|f\|_{\mathrm{TC}(G_m)} \qquad \forall f \in \mathscr{F}(G_m),$$

then

$$D \ge c_d \log|G_m|.$$

In particular, no exact representation of the form

$$\|f\|_{\mathrm{TC}(G_m)} = \sum_{\ell=1}^L \lambda_\ell \|A_\ell f\|_1$$

can hold along the full asymptotic grid sequence $m \to \infty$.

Proof. Define

$$\Phi : \mathscr{F}(G_m) \to \mathbb{R}^{M_1+\cdots+M_L}, \qquad \Phi(f) := (\lambda_1 A_1 f, \dots, \lambda_L A_L f).$$

Then

$$\|\Phi(f)\|_1 = \sum_{\ell=1}^{L} \lambda_\ell \|A_\ell f\|_1.$$

Therefore the assumed two-sided bound yields an embedding of the normed space

$$(\mathscr{F}(G_m), \|\cdot\|_{\mathrm{TC}(G_m)})$$

into $\ell_1^{M_1+\cdots+M_L} \subset L_1$ with distortion at most D. Applying Theorem 6.12 gives

$$D \geq c_d \log |G_m|.$$

The final claim follows because $c_d \log |G_m| \to \infty$ as $m \to \infty$. ☐

Corollary 6.16 (No exact same-map TV skeleton on the full Euclidean grid)**.** *Assume $d \geq 2$. Let $\pi_\ell : G_m \to \{1, \dots, M_\ell\}$ and $\lambda_\ell > 0$ for $1 \leq \ell \leq L$. If for some $D \geq 1$ one has*

$$W_1(P,Q) \leq \sum_{\ell=1}^{L} \lambda_\ell \,\mathrm{TV}\big((\pi_\ell)_\# P, (\pi_\ell)_\# Q\big) \leq D\, W_1(P,Q) \qquad \forall P,Q \in \mathcal{P}(G_m),$$

then

$$D \geq c_d \log |G_m|.$$

In particular, there is no exact representation

$$W_1(P,Q) = \sum_{\ell=1}^{L} \lambda_\ell \,\mathrm{TV}\big((\pi_\ell)_\# P, (\pi_\ell)_\# Q\big) \qquad \forall P,Q \in \mathcal{P}(G_m)$$

valid for all sufficiently large m.

Proof. For each ℓ, let

$$A_\ell : \mathscr{F}(G_m) \to \mathbb{R}^{M_\ell}$$

be the linear aggregation map defined by

$$(A_\ell f)(r) := \sum_{x \in G_m :\ \pi_\ell(x) = r} f(x), \qquad 1 \leq r \leq M_\ell.$$

Then for every $P, Q \in \mathcal{P}(G_m)$,

$$\mathrm{TV}\big((\pi_\ell)_\# P, (\pi_\ell)_\# Q\big) = \frac{1}{2} \|A_\ell (P - Q)\|_1.$$

Hence the assumed representation is exactly of the form covered by Theorem 6.15, after absorbing the factor $1/2$ into the weights λ_ℓ. ☐

7 Multiscale polynomial debiasing on transport trees

The preceding tree results were stated for exact Euclidean subclasses. The next theorem is stronger in a different direction: it solves, for every pair of laws on the full leaf set, the complete multiscale tree transportation functional with the same $N \log N$ critical gain. This is the cleanest model in which the lower-bound scale, the diagonal empirical width, and the large-alphabet bias correction all coexist. It also clarifies a key point for the Euclidean problem: diagonal centering alone is not the whole mechanism. One must combine centering with polynomial large-alphabet debiasing at every active scale.

Let X be a finite set equipped with a nested sequence of partitions

$$\mathcal{A}_0 = \{X\}, \mathcal{A}_1, \dots, \mathcal{A}_J,$$

where every atom of $\mathcal{A}_{j+1}$ is contained in an atom of $\mathcal{A}_j$. For $P \in \mathcal{P}(X)$, write P_j for the push-forward of P to the finite alphabet $\mathcal{A}_j$,

$$P_j(A) = P(A), \qquad A \in \mathcal{A}_j.$$

For $p \geq 1$ define the additive p-cost tree functional

$$\mathfrak{T}_{p,J}(P,Q) := \sum_{j=1}^{J} 2^{-pj} \|P_j - Q_j\|_1$$

and the associated root functional

$$\mathfrak{W}_{p,J}(P,Q) := \mathfrak{T}_{p,J}(P,Q)^{1/p}.$$

For $p = 1$, $\mathfrak{T}_{1,J}$ is exactly the transportation distance on the rooted tree whose edge from level $j-1$ to level j has length 2^{-j}. For $p > 1$, it is the same cut-additive tree transport with edge cost 2^{-pj}, followed by the natural p-th root.

Theorem 7.1 (Complete multiscale polynomial tree law)**.** *Let $d > 2p$. Fix $A < \infty$. Assume that*

$$|\mathcal{A}_j| \leq A\, 2^{dj}, \qquad 1 \leq j \leq J,$$

and let $h_J = 2^{-J}$. If

$$N \log N \geq A\, h_J^{-d},$$

then there exists an estimator $\widehat{\mathfrak{T}}_{p,J}$, based on two independent samples of size N, such that

$$\sup_{P,Q \in \mathcal{P}(X)} \mathbb{E} \left| \widehat{\mathfrak{T}}_{p,J} - \mathfrak{T}_{p,J}(P,Q) \right| \leq C_{A,d,p}\, h_J^p$$

and

$$\sup_{P,Q\in\mathcal{P}(X)} \mathbb{E}\left|\left(\widehat{\mathfrak{T}}_{p,J}\right)_+^{1/p} - \mathfrak{W}_{p,J}(P,Q)\right| \le C_{A,d,p}\, h_J.$$

Proof. For each level j, the samples induce two empirical samples from P_j, Q_j on the alphabet $\mathcal{A}_j$, whose size is

$$M_j := |\mathcal{A}_j| \le A2^{dj}.$$

Choose

$$\alpha \in \left(0,\, 1 - \frac{2p}{d}\right).$$

If $M_j \ge N^\alpha$, use the large-alphabet L_1-distance estimator of Jiao–Han–Weissman on $\mathcal{A}_j$; after clipping to $[0,2]$, denote it by $\widehat{L}_j$. We use the consequence

$$\sup_{P,Q} \mathbb{E}\left(\widehat{L}_j - \|P_j - Q_j\|_1\right)^2 \lesssim \frac{M_j}{N\log N}. \tag{6.1}$$

If $M_j < N^\alpha$, use the empirical plug-in estimator, again clipped to $[0,2]$; then the standard multinomial inequality gives

$$\sup_{P,Q} \mathbb{E}\left(\widehat{L}_j - \|P_j - Q_j\|_1\right)^2 \lesssim \frac{M_j}{N} \le N^{\alpha-1}. \tag{6.2}$$

Set

$$\widehat{\mathfrak{T}}_{p,J} := \sum_{j=1}^{J} 2^{-pj}\widehat{L}_j.$$

Then

$$\mathbb{E}\left|\widehat{\mathfrak{T}}_{p,J} - \mathfrak{T}_{p,J}(P,Q)\right| \le \sum_{j=1}^{J} 2^{-pj}\left[\mathbb{E}\left(\widehat{L}_j - \|P_j - Q_j\|_1\right)^2\right]^{1/2}.$$

For the large-alphabet levels,

$$\sum_{j:\, M_j \ge N^\alpha} 2^{-pj}\sqrt{\frac{M_j}{N\log N}} \le C_A (N\log N)^{-1/2}\sum_{j=1}^{J} 2^{(d/2-p)j}.$$

Since $d/2 - p > 0$, the last sum is dominated by its terminal scale:

$$\sum_{j=1}^{J} 2^{(d/2-p)j} \lesssim_{d,p} 2^{(d/2-p)J}.$$

Using $2^J = h_J^{-1}$ and $N\log N \ge Ah_J^{-d}$,

$$(N\log N)^{-1/2}2^{(d/2-p)J} \lesssim_{A,d,p} h_J^p.$$

For the small-alphabet levels, (6.2) gives

$$\sum_{j:\, M_j < N^\alpha} 2^{-pj} N^{(\alpha-1)/2} \le C_p\, N^{(\alpha-1)/2}.$$

The choice of α implies $(1-\alpha)/2 > p/d$. Since $h_J^{-d} \lesssim_A N\log N$, the logarithmic factor is harmless and

$$N^{(\alpha-1)/2} \lesssim_{A,d,p} h_J^p.$$

Combining the two estimates proves the powered bound.

Finally,

$$|a^{1/p} - b^{1/p}| \le |a-b|^{1/p}, \qquad a,b \ge 0.$$

Hence, by Jensen's inequality,

$$\begin{aligned}\mathbb{E}\left|\left(\widehat{\mathfrak{T}}_{p,J}\right)_+^{1/p} - \mathfrak{W}_{p,J}(P,Q)\right| &\le \mathbb{E}\left|\left(\widehat{\mathfrak{T}}_{p,J}\right)_+ - \mathfrak{T}_{p,J}(P,Q)\right|^{1/p} \\ &\le \left[\mathbb{E}\left|\widehat{\mathfrak{T}}_{p,J} - \mathfrak{T}_{p,J}(P,Q)\right|\right]^{1/p} \lesssim h_J.\end{aligned}$$

□

Corollary 7.2 (Sharpness on the complete dyadic tree)**.** *Let $\mathcal{A}_j$ be the dyadic partition of $[0,1]^d$ up to level J, restricted to its 2^{dJ} terminal cells, and assume*

$$2^{dJ} \asymp N\log N.$$

Then

$$\inf_{\widehat{\mathfrak{W}}}\sup_{P,Q} \mathbb{E}\left|\widehat{\mathfrak{W}} - \mathfrak{W}_{p,J}(P,Q)\right| \asymp_{d,p} 2^{-J},$$

and, equivalently,

$$\inf_{\widehat{\mathfrak{T}}}\sup_{P,Q} \mathbb{E}\left|\widehat{\mathfrak{T}} - \mathfrak{T}_{p,J}(P,Q)\right| \asymp_{d,p} 2^{-pJ}.$$

Proof. The upper bound is Theorem 7.1. For the lower bound, choose inside every dyadic parent at level $J-1$ two children and perturb mass only between those two children. All coarser partition marginals agree, while the level J term satisfies

$$\mathfrak{T}_{p,J}(P,Q) = 2^{-pJ}\,\|P_J - Q_J\|_1.$$

The family contains a large-alphabet L_1-distance problem over $M \asymp 2^{dJ} \asymp N\log N$ symbols. The Jiao–Han–Weissman lower bound gives a constant lower bound for estimating $\|P_J - Q_J\|_1$ in this regime. Multiplication by 2^{-pJ} yields the powered lower bound. The root lower bound follows by restricting to alternatives for which $\|P_J - Q_J\|_1$ is bounded above and below by fixed positive constants, so that the map $t \mapsto t^{1/p}$ has derivative comparable to one on the relevant interval after factoring out 2^{-pJ}. □

Theorem 7.3 (Raw centering misses the logarithm on a tree level)**.** *Let $d > 2p$, and set*

$$h_N = (N \log N)^{-1/d}.$$

For the complete dyadic tree functional above, let $\widehat{\mathfrak{T}}^{\mathrm{raw}}$ be the levelwise four-sample centered plug-in estimator obtained by replacing every $\|P_j - Q_j\|_1$ by

$$\|\widehat{P}_j - \widehat{Q}_j\|_1 - \frac{1}{2}\|\widehat{P}_j - \widehat{P}'_j\|_1 - \frac{1}{2}\|\widehat{Q}_j - \widehat{Q}'_j\|_1.$$

Then there are dyadic levels j_N, with $2^{dj_N} \asymp N$, and laws P_N, Q_N such that

$$\mathfrak{W}_{p,J}(P_N, Q_N) \asymp N^{-1/d},$$

but

$$\left|\mathbb{E}\widehat{\mathfrak{T}}^{\mathrm{raw}} - \mathfrak{T}_{p,J}(P_N, Q_N)\right| \gtrsim_{d,p} N^{-p/d}.$$

Consequently

$$\frac{\left|\mathbb{E}\widehat{\mathfrak{T}}^{\mathrm{raw}} - \mathfrak{T}_{p,J}(P_N, Q_N)\right|}{h_N\left(\mathfrak{W}_{p,J}(P_N, Q_N) + h_N\right)^{p-1}} \to \infty.$$

Thus exact diagonal cancellation without polynomial large-alphabet debiasing does not recover the $N \log N$ improvement, even on a completely solved multiscale tree model.

Proof. Choose j_N so that

$$k_N := |\mathcal{A}_{j_N}| \asymp N,$$

and take P_N, Q_N to differ only inside level-j_N atoms, with all coarser masses equal. It is enough to work on the alphabet of level j_N. Choose signs $\sigma_i \in \{-1, 1\}$ with $\sum_i \sigma_i = 0$, and fix a small constant $a \in (0, 1/10)$. Put

$$p_i = \frac{1 + a\sigma_i}{k_N}, \qquad q_i = \frac{1 - a\sigma_i}{k_N}.$$

Then

$$\|p - q\|_1 = 2a.$$

Consider the scalar Poissonized version with $N/k_N \to \lambda \in (0, \infty)$. For one coordinate with $\sigma_i = 1$, the centered absolute-value contribution converges after multiplication by N to

$$B_\lambda(a) := \mathbb{E}|X_+ - Y_-| - \frac{1}{2}\mathbb{E}|X_+ - X'_+| - \frac{1}{2}\mathbb{E}|Y_- - Y'_-|,$$

where

$$X_+, X'_+ \sim \mathrm{Poi}(\lambda(1 + a)), \qquad Y_-, Y'_- \sim \mathrm{Poi}(\lambda(1 - a))$$

are independent. The same limit holds for $\sigma_i = -1$ by symmetry. The function $B_\lambda(a)$ is analytic near 0, satisfies $B_\lambda(0) = 0$, and is even to first order: $B'_\lambda(0) = 0$. Indeed, interchanging a and $-a$ swaps the two independent coordinates in the first absolute value and swaps the two diagonal terms. Hence

$$B_\lambda(a) = O_\lambda(a^2) \qquad (a \downarrow 0).$$

Taking $a > 0$ sufficiently small gives

$$|2a - B_\lambda(a)| \geq a.$$

The usual de-Poissonization changes the expectation by $o(1)$ after summing over $k_N \asymp N$ coordinates. Therefore the level-j_N centered plug-in estimator of $\|p - q\|_1$ has an expectation bias bounded below by a positive constant depending only on a and the bounded ratio N/k_N.

All other levels are arranged to have identical coarser masses, so the displayed single-level bias contributes to the tree functional with weight 2^{-pj_N}. Since $2^{dj_N} \asymp N$,

$$2^{-pj_N} \asymp N^{-p/d}.$$

Moreover

$$\mathfrak{T}_{p,J}(P_N, Q_N) \asymp 2^{-pj_N}, \qquad \mathfrak{W}_{p,J}(P_N, Q_N) \asymp 2^{-j_N} \asymp N^{-1/d}.$$

Finally,

$$h_N\left(\mathfrak{W}_{p,J}(P_N, Q_N) + h_N\right)^{p-1} \asymp (N \log N)^{-1/d} N^{-(p-1)/d} = N^{-p/d} (\log N)^{-1/d},$$

which is smaller than $N^{-p/d}$ by a diverging logarithmic factor. □

Remark 7.4 (Consequence for the Euclidean finite LP)**.** The obstruction in Theorem 7.3 does not disprove a special Euclidean centered-curvature identity. It does show that such an identity, if true, cannot be a generic consequence of diagonal cancellation for high-dimensional transportation norms. A robust proof of the unrestricted Euclidean theorem must contain an $N \log N$-type polynomial debiasing mechanism comparable to Theorem 7.1, either explicitly at multiscale cuts or implicitly inside a debiased Kantorovich LP regularization.

8 Dual compression and annular completion

The tree construction of Section 6 compresses the Euclidean transport norm through a special family of hierarchical metrics. We now pass to the semi-discrete dual itself. This produces a broader positive mechanism: for a bounded cost on a finite known support, a maximum over any feasible dual catalog is statistically cheap, and the only price of allowing many candidate phases is a $\sqrt{\log K/N}$ factor. Low-dimensional continuous dual manifolds are equally admissible after discretization. Applied to the annular linearizations from Section 4, this turns the unresolved supercritical core into a concrete *dual compression problem.*

8.1 Finite dual catalogs

Let $\mathcal{X} := [0, 1]^d$, let $Y \subset \mathcal{X}$ be finite, and let

$$c : \mathcal{X} \times Y \to \mathbb{R}$$

be a bounded measurable cost. For $P \in \mathcal{P}(\mathcal{X})$ and $R \in \mathcal{P}(Y)$, define

$$T_c(P,R) := \inf_{\pi \in \Pi(P,R)} \int c(x,y)\, d\pi(x,y).$$

By Kantorovich duality,

$$T_c(P,R) = \sup\Big\{ Pf + Rg : \ f : \mathcal{X} \to \mathbb{R},\ g : Y \to \mathbb{R},\ f(x) + g(y) \le c(x,y)\ \forall (x,y) \Big\}.$$

A *finite dual catalog* for c is a finite family

$$\mathscr{A} = \{(f_k, g_k)\}_{k=1}^K$$

of dual-feasible pairs,

$$f_k(x) + g_k(y) \le c(x,y) \qquad (x \in \mathcal{X},\ y \in Y,\ 1 \le k \le K).$$

Given samples

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P, \qquad Y_1, \dots, Y_m \overset{\text{i.i.d.}}{\sim} R,$$

write

$$P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \qquad R_m := \frac{1}{m} \sum_{j=1}^m \delta_{Y_j}, \qquad N := n \wedge m.$$

For every catalog element define the empirical dual score

$$\hat{D}_k := P_n f_k + R_m g_k,$$

and the catalog estimator

$$\hat{T}_{\mathscr{A}} := \max_{1 \le k \le K} \hat{D}_k.$$

The corresponding population score is

$$D_k(P,R) := Pf_k + Rg_k,$$

and the catalog approximation gap is

$$\Delta_{\mathscr{A}}(P,R) := T_c(P,R) - \max_{1 \le k \le K} D_k(P,R) \ge 0.$$

Lemma 8.1 (Finite dual-catalog maximal inequality)**.** *Assume*

$$\|f_k\|_\infty \le B, \qquad \|g_k\|_\infty \le B \qquad (1 \le k \le K).$$

Then

$$\mathbb{E} \max_{1 \le k \le K} \big|(P_n - P) f_k + (R_m - R) g_k\big| \le C\, B \sqrt{\frac{\log(2K)}{N}},$$

and

$$\mathbb{E} \max_{1 \le k \le K} \big|(P_n - P) f_k + (R_m - R) g_k\big|^2 \le C\, B^2\, \frac{\log(2K)}{N},$$

for a universal constant $C < \infty$.

Proof. For each k define

$$Z_k := (P_n - P) f_k + (R_m - R) g_k.$$

Since

$$\frac{f_k(X_i) - Pf_k}{n} \in \Big[-\frac{2B}{n}, \frac{2B}{n}\Big], \qquad \frac{g_k(Y_j) - Rg_k}{m} \in \Big[-\frac{2B}{m}, \frac{2B}{m}\Big],$$

Hoeffding's inequality gives, for every $u > 0$,

$$\mathbb{P}(|Z_k| \ge u) \le 2 \exp\Big(-\frac{u^2}{8B^2(n^{-1} + m^{-1})}\Big).$$

Taking the union bound over $1 \le k \le K$ yields

$$\mathbb{P}\Big(\max_{1 \le k \le K} |Z_k| \ge u\Big) \le 2K \exp\Big(-\frac{u^2}{8B^2(n^{-1} + m^{-1})}\Big).$$

Since $n^{-1} + m^{-1} \le 2/N$, there is a universal constant $c_0 > 0$ such that

$$\mathbb{P}\Big(\max_{1 \le k \le K} |Z_k| \ge u\Big) \le 2K \exp\Big(-\frac{c_0 N u^2}{B^2}\Big).$$

Set

$$M_K := \max_{1 \le k \le K} |Z_k|, \qquad u_0 := \frac{2B}{\sqrt{c_0 N}} \sqrt{\log(2K)}.$$

Then

$$\mathbb{E} M_K = \int_0^\infty \mathbb{P}(M_K \ge u)\, du \le u_0 + \int_{u_0}^\infty 2K e^{-c_0 N u^2 / B^2}\, du \le C\, B \sqrt{\frac{\log(2K)}{N}}$$

for a universal constant C. Likewise,

$$\mathbb{E} M_K^2 = 2 \int_0^\infty u\, \mathbb{P}(M_K \ge u)\, du \le u_0^2 + 4K \int_{u_0}^\infty u e^{-c_0 N u^2 / B^2}\, du \le C\, B^2\, \frac{\log(2K)}{N}.$$

□

Theorem 8.2 (Dual-catalog oracle inequality)**.** *Let $\mathscr{A} = \{(f_k, g_k)\}_{k=1}^K$ be a finite dual catalog for c satisfying*

$$\|f_k\|_\infty \le B, \qquad \|g_k\|_\infty \le B \qquad (1 \le k \le K).$$

Then for every $P \in \mathcal{P}(\mathcal{X})$ and every $R \in \mathcal{P}(Y)$,

$$\mathbb{E}\big|\hat{T}_{\mathscr{A}} - T_c(P,R)\big| \le \Delta_{\mathscr{A}}(P,R) + C\,B\sqrt{\frac{\log(2K)}{N}},$$

and

$$\mathbb{E}\big(\hat{T}_{\mathscr{A}} - T_c(P,R)\big)^2 \le 2\Delta_{\mathscr{A}}(P,R)^2 + C\,B^2\,\frac{\log(2K)}{N},$$

for a universal constant $C < \infty$.

Proof. For each k write

$$D_k := D_k(P,R) = Pf_k + Rg_k, \qquad \hat{D}_k := P_n f_k + R_m g_k, \qquad Z_k := \hat{D}_k - D_k.$$

By dual feasibility,

$$D_k \le T_c(P,R) \qquad (1 \le k \le K).$$

Hence

$$\hat{T}_{\mathscr{A}} - T_c(P,R) = \max_{1\le k\le K} \hat{D}_k - T_c(P,R) \le \max_{1\le k\le K}(\hat{D}_k - D_k) \le \max_{1\le k\le K}|Z_k|.$$

Choose $k_\star$ attaining the maximum in the definition of $\Delta_{\mathscr{A}}(P,R)$. Then

$$T_c(P,R) - \hat{T}_{\mathscr{A}} \le T_c(P,R) - \hat{D}_{k_\star} = \Delta_{\mathscr{A}}(P,R) + (D_{k_\star} - \hat{D}_{k_\star}) \le \Delta_{\mathscr{A}}(P,R) + \max_{1\le k\le K}|Z_k|.$$

Therefore

$$\big|\hat{T}_{\mathscr{A}} - T_c(P,R)\big| \le \Delta_{\mathscr{A}}(P,R) + \max_{1\le k\le K}|Z_k|,$$

and also

$$\big(\hat{T}_{\mathscr{A}} - T_c(P,R)\big)^2 \le 2\Delta_{\mathscr{A}}(P,R)^2 + 2\max_{1\le k\le K}|Z_k|^2.$$

Now apply Theorem 8.1. □

Corollary 8.3 (Approximate K-phase families)**.** *Under the assumptions of Theorem 8.2, let $\mathcal{C} \subseteq \mathcal{P}(\mathcal{X}) \times \mathcal{P}(Y)$ satisfy*

$$\sup_{(P,R)\in\mathcal{C}} \Delta_{\mathscr{A}}(P,R) \le A$$

for some $A \ge 0$. Then

$$\sup_{(P,R)\in\mathcal{C}} \mathbb{E}\big|\hat{T}_{\mathscr{A}} - T_c(P,R)\big| \le A + C\,B\sqrt{\frac{\log(2K)}{N}},$$

and

$$\sup_{(P,R)\in\mathcal{C}} \mathbb{E}\big(\hat{T}_{\mathscr{A}} - T_c(P,R)\big)^2 \le 2A^2 + C\,B^2\,\frac{\log(2K)}{N}.$$

In particular, if $A = 0$, then every exact K-phase family is estimable at rate

$$\sqrt{\frac{\log K}{N}}$$

up to constants depending only on the envelope bound.

Proof. Insert the uniform approximation bound into Theorem 8.2. □

8.2 Low-dimensional dual manifolds

The previous theorem is finite. A continuous dual family can be reduced to this situation by a net argument.

Theorem 8.4 (Parametric dual compression)**.** *Let $\Theta \subset [-R,R]^q$ and let*

$$\{(f_\theta, g_\theta) : \theta \in \Theta\}$$

be a family of dual-feasible pairs for c. Assume

$$\|f_\theta\|_\infty \le B, \qquad \|g_\theta\|_\infty \le B \qquad (\theta \in \Theta),$$

and suppose that for some $L \ge 1$,

$$\|f_\theta - f_{\theta'}\|_\infty + \|g_\theta - g_{\theta'}\|_\infty \le L\|\theta - \theta'\|_\infty \qquad (\theta, \theta' \in \Theta).$$

Define

$$\Delta_\Theta(P,R) := T_c(P,R) - \sup_{\theta\in\Theta}\big\{Pf_\theta + Rg_\theta\big\} \ge 0.$$

Then for every $\varepsilon \in (0,1]$ there exists an estimator $\hat{T}_{\Theta,\varepsilon}$ such that

$$\mathbb{E}\big|\hat{T}_{\Theta,\varepsilon} - T_c(P,R)\big| \le \Delta_\Theta(P,R) + \varepsilon + C\,B\sqrt{\frac{q\log\big(1 + 2RL/\varepsilon\big)}{N}},$$

and

$$\mathbb{E}\big(\hat{T}_{\Theta,\varepsilon} - T_c(P,R)\big)^2 \le 2\big(\Delta_\Theta(P,R) + \varepsilon\big)^2 + C\, B^2\, \frac{q \log\big(1 + 2RL/\varepsilon\big)}{N},$$

where $C < \infty$ is universal.

Proof. Choose an ℓ_∞-net

$$\mathcal{N}_\varepsilon \subset \Theta$$

of mesh ε/L. Since $\Theta \subset [-R,R]^q$, one may choose it so that

$$|\mathcal{N}_\varepsilon| \le \Big(1 + \frac{2RL}{\varepsilon}\Big)^q.$$

Apply Theorem 8.2 to the finite catalog

$$\mathscr{A}_\varepsilon := \{(f_\theta, g_\theta) : \theta \in \mathcal{N}_\varepsilon\}.$$

For any (P,R) and any $\theta \in \Theta$, choose $\theta_\varepsilon \in \mathcal{N}_\varepsilon$ with

$$\|\theta - \theta_\varepsilon\|_\infty \le \varepsilon/L.$$

Then

$$|Pf_\theta - Pf_{\theta_\varepsilon}| \le \|f_\theta - f_{\theta_\varepsilon}\|_\infty, \qquad |Rg_\theta - Rg_{\theta_\varepsilon}| \le \|g_\theta - g_{\theta_\varepsilon}\|_\infty,$$

hence

$$\Big|Pf_\theta + Rg_\theta - Pf_{\theta_\varepsilon} - Rg_{\theta_\varepsilon}\Big| \le \varepsilon.$$

Therefore

$$\Delta_{\mathscr{A}_\varepsilon}(P,R) \le \Delta_\Theta(P,R) + \varepsilon.$$

Since

$$\log\Big(2|\mathscr{A}_\varepsilon|\Big) \le q \log\Big(1 + \frac{2RL}{\varepsilon}\Big) + \log 2,$$

Theorem 8.2 gives the stated bounds after absorbing $\log 2$ into the constant. □

Corollary 8.5 (Feature-compressed semi-dual weights)**.** *Let $\psi_1, \dots, \psi_q : Y \to [-1,1]$ and fix $R_0 \ge 1$. For $\theta \in [-R_0, R_0]^q$, define*

$$g_\theta(y) := \sum_{b=1}^{q} \theta_b \psi_b(y), \qquad f_\theta(x) := \min_{y \in Y}\big(c(x,y) - g_\theta(y)\big).$$

Then each (f_θ, g_θ) is dual-feasible for c. If $|c(x,y)| \le C_c$ on $\mathcal{X} \times Y$, then

$$\|g_\theta\|_\infty \le qR_0, \qquad \|f_\theta\|_\infty \le C_c + qR_0,$$

and

$$\|f_\theta - f_{\theta'}\|_\infty + \|g_\theta - g_{\theta'}\|_\infty \le 2q\|\theta - \theta'\|_\infty.$$

Consequently, if a class $\mathcal{C} \subseteq \mathcal{P}(\mathcal{X}) \times \mathcal{P}(Y)$ satisfies

$$\sup_{(P,R)\in\mathcal{C}} \Delta_\Theta(P,R) \le A,$$

then for every $\varepsilon \in (0,1]$ there exists an estimator with

$$\sup_{(P,R)\in\mathcal{C}} \mathbb{E}\big|\hat{T} - T_c(P,R)\big| \le A + \varepsilon + C\,(C_c + qR_0)\sqrt{\frac{q \log\big(1 + 4qR_0/\varepsilon\big)}{N}},$$

and an analogous squared-error bound.

Proof. Dual feasibility is immediate from the definition of f_θ. The envelope bounds follow from $|g_\theta(y)| \le qR_0$ and

$$-C_c - qR_0 \le f_\theta(x) \le C_c + qR_0.$$

Finally,

$$\|g_\theta - g_{\theta'}\|_\infty \le q\|\theta - \theta'\|_\infty,$$

and since minima are 1-Lipschitz with respect to the sup-norm perturbation of the family being minimized,

$$\|f_\theta - f_{\theta'}\|_\infty \le \|g_\theta - g_{\theta'}\|_\infty \le q\|\theta - \theta'\|_\infty.$$

Apply Theorem 8.4. □

8.3 The full semi-discrete box: removing the discretization logarithm

The net argument in Theorem 8.4 is sharp for a generic q-parameter family, but it is not sharp for the full semi-discrete family itself. If the support Y is fixed and finite, the exact support-weight box has entropy dimension $M := |Y|$ in the sup norm, and this removes the additional $\log(1/\varepsilon)$ loss coming from a crude ε-net of the parameter space.

Fix

$$Y = \{y_1, \dots, y_M\} \subset \mathcal{X},$$

and define the oscillation scale

$$D_Y(c) := \max_{1 \le a,b \le M} \sup_{x \in \mathcal{X}} \big|c(x,y_a) - c(x,y_b)\big|.$$

Also write

$$B_Y(c) := \|c\|_\infty + D_Y(c).$$

For a weight function $g : Y \to \mathbb{R}$, let

$$f_g(x) := \min_{y \in Y}\big(c(x,y) - g(y)\big).$$

Then (f_g, g) is dual-feasible for c.

Lemma 8.6 (Exact oscillation-bounded normalization)**.** *For every $P \in \mathcal{P}(\mathcal{X})$ and every $R \in \mathcal{P}(Y)$,*

$$T_c(P,R) = \sup_{g \in \mathcal{G}_Y} \big\{ Pf_g + Rg \big\},$$

where

$$\mathcal{G}_Y := \Big\{ g : Y \to \mathbb{R} : \ g(y_M) = 0, \ \|g\|_\infty \le D_Y(c) \Big\}.$$

Proof. Fix any $g : Y \to \mathbb{R}$ and set $f := f_g$. Define

$$\tilde g(y) := \inf_{x \in \mathcal{X}} \big(c(x,y) - f(x) \big), \qquad y \in Y.$$

Because $f(x) \le c(x,y) - g(y)$ for every (x,y), we have $g(y) \le \tilde g(y)$ for every $y \in Y$. Moreover, for every $x \in \mathcal{X}$ and every $y \in Y$,

$$\tilde g(y) \le c(x,y) - f(x),$$

hence

$$f(x) \le c(x,y) - \tilde g(y).$$

Taking the minimum over y gives

$$f(x) \le f_{\tilde g}(x).$$

On the other hand, since $\tilde g \ge g$ pointwise,

$$f_{\tilde g}(x) \le f_g(x) = f(x).$$

Therefore

$$f_{\tilde g} = f_g.$$

Consequently

$$Pf_{\tilde g} + R\tilde g = Pf_g + R\tilde g \ge Pf_g + Rg.$$

Thus the full dual supremum may be restricted to pairs of the form (f_g, g) with

$$g(y) = \inf_{x \in \mathcal{X}} \big(c(x,y) - f_g(x) \big).$$

For such a normalized g, fix $y, z \in Y$. Since

$$f_g(x) \le c(x,z) - g(z) \qquad \forall x \in \mathcal{X},$$

we obtain

$$c(x,y) - f_g(x) \ge g(z) + c(x,y) - c(x,z) \qquad \forall x \in \mathcal{X}.$$

Taking the infimum over x yields

$$g(y) - g(z) \ge \inf_{x \in \mathcal{X}} \big(c(x,y) - c(x,z) \big).$$

Interchanging y and z gives

$$g(y) - g(z) \le \sup_{x \in \mathcal{X}} \big(c(x,y) - c(x,z) \big),$$

hence

$$|g(y) - g(z)| \le D_Y(c).$$

Now subtract the constant $g(y_M)$ from all coordinates. The dual value is unchanged because

$$f_{g-a}(x) = f_g(x) + a \qquad \text{and} \qquad R(g-a) = Rg - a \qquad (a \in \mathbb{R}),$$

so

$$Pf_{g-a} + R(g-a) = Pf_g + Rg.$$

After this centering we have $g(y_M) = 0$ and therefore

$$\|g\|_\infty \le D_Y(c).$$

This proves the claim. □

Lemma 8.7 (Entropy of the full semi-discrete box)**.** *Let*

$$\mathcal{F}_Y := \{ f_g : \ g \in \mathcal{G}_Y \}.$$

Then for every $\varepsilon \in (0, D_Y(c)]$,

$$\log N\big(\varepsilon, \mathcal{G}_Y, \|\cdot\|_\infty\big) \le (M-1) \log\Big(1 + \frac{2D_Y(c)}{\varepsilon}\Big),$$

and

$$\log N\big(\varepsilon, \mathcal{F}_Y, \|\cdot\|_\infty\big) \le (M-1) \log\Big(1 + \frac{2D_Y(c)}{\varepsilon}\Big).$$

Moreover,

$$\|f_g\|_\infty \le B_Y(c) \qquad (g \in \mathcal{G}_Y),$$

and

$$\|f_g - f_{g'}\|_\infty \le \|g - g'\|_\infty \qquad (g, g' \in \mathcal{G}_Y).$$

Proof. The set $\mathcal{G}_Y$ is an $(M-1)$-dimensional ℓ_∞-box of radius $D_Y(c)$ after fixing the coordinate $g(y_M) = 0$. Hence it admits an ε-net in $\|\cdot\|_\infty$ of cardinality at most

$$\Big(1 + \frac{2D_Y(c)}{\varepsilon}\Big)^{M-1}.$$

This proves the first entropy bound.

For the envelope bound, since $|c(x,y)| \le \|c\|_\infty$ and $|g(y)| \le D_Y(c)$,

$$-B_Y(c) \le c(x,y) - g(y) \le B_Y(c) \qquad \forall (x,y) \in \mathcal{X} \times Y,$$

and therefore

$$\|f_g\|_\infty \le B_Y(c).$$

Finally, for every $x \in \mathcal{X}$,

$$|f_g(x) - f_{g'}(x)| = \Big| \min_{y \in Y} \big(c(x,y) - g(y) \big) - \min_{y \in Y} \big(c(x,y) - g'(y) \big) \Big| \le \|g - g'\|_\infty,$$

because the minimum is 1-Lipschitz with respect to uniform perturbations of the family being minimized. Hence

$$\|f_g - f_{g'}\|_\infty \le \|g - g'\|_\infty.$$

The second entropy bound follows by transporting any ε-net of $\mathcal{G}_Y$ through the map $g \mapsto f_g$. □

Proposition 8.8 (Empirical process control on the full support box)**.** *Define*

$$Z_X := \sup_{g\in\mathcal{G}_Y} \left|(P_n - P)f_g\right|, \qquad Z_Y := \sup_{g\in\mathcal{G}_Y} \left|(R_m - R)g\right|.$$

Then

$$\mathbb{E}Z_X \le C\, B_Y(c)\sqrt{\frac{M}{n}}, \qquad \mathbb{E}Z_X^2 \le C\, B_Y(c)^2\,\frac{M}{n},$$

and

$$\mathbb{E}Z_Y \le D_Y(c)\sqrt{\frac{M}{m}}, \qquad \mathbb{E}Z_Y^2 \le D_Y(c)^2\,\frac{M}{m},$$

for a universal constant $C < \infty$.

Proof. For Z_X, symmetrization gives

$$\mathbb{E}Z_X \le 2\,\mathbb{E}\sup_{g\in\mathcal{G}_Y}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i f_g(X_i)\right|,$$

where $(\sigma_i)_{1\le i\le n}$ are i.i.d. Rademacher signs independent of the data. Conditionally on $X_1,\dots,X_n$, Dudley's entropy integral bound (see, e.g., [12, Chapter 2]) and Theorem 8.7 yield

$$\begin{aligned}\mathbb{E}_\sigma \sup_{g\in\mathcal{G}_Y}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i f_g(X_i)\right| &\le \frac{C}{\sqrt{n}}\int_0^{2B_Y(c)} \sqrt{\log N\big(u, \mathcal{F}_Y, L_2(P_n)\big)}\,du \\ &\le \frac{C\sqrt{M}}{\sqrt{n}}\int_0^{2B_Y(c)} \sqrt{\log\Big(1 + \frac{2D_Y(c)}{u}\Big)}\,du.\end{aligned}$$

Since $D_Y(c) \le B_Y(c)$, the change of variables $u = 2B_Y(c)s$ gives

$$\int_0^{2B_Y(c)} \sqrt{\log\Big(1 + \frac{2D_Y(c)}{u}\Big)}\,du \le 2B_Y(c)\int_0^1 \sqrt{\log(1+s^{-1})}\,ds \le C\,B_Y(c).$$

Therefore

$$\mathbb{E}Z_X \le C\,B_Y(c)\sqrt{\frac{M}{n}}.$$

For the second moment, changing one observation X_i alters every empirical average $P_n f_g$ by at most

$$\frac{2\|f_g\|_\infty}{n} \le \frac{2B_Y(c)}{n},$$

hence changes Z_X by at most $2B_Y(c)/n$. By the bounded-differences variance bound,

$$\mathrm{Var}(Z_X) \le C\,\frac{B_Y(c)^2}{n}.$$

Combining this with the first-moment estimate yields

$$\mathbb{E}Z_X^2 = \mathrm{Var}(Z_X) + (\mathbb{E}Z_X)^2 \le C\,B_Y(c)^2\,\frac{M}{n}.$$

For Z_Y, write

$$\Delta_a := R_m(\{y_a\}) - R(\{y_a\}) \qquad (1 \le a \le M).$$

Because every $g \in \mathcal{G}_Y$ satisfies $g(y_M) = 0$ and $|g(y_a)| \le D_Y(c)$,

$$Z_Y = \sup_{g\in\mathcal{G}_Y}\left|\sum_{a=1}^{M-1} g(y_a)\Delta_a\right| = D_Y(c)\sum_{a=1}^{M-1}|\Delta_a|.$$

Hence, by Cauchy–Schwarz,

$$\mathbb{E}Z_Y \le D_Y(c)\sqrt{M-1}\left(\sum_{a=1}^{M-1}\mathbb{E}\Delta_a^2\right)^{1/2}.$$

Since $(R_m(\{y_a\}))_{a=1}^M$ is multinomial,

$$\sum_{a=1}^M \mathbb{E}\Delta_a^2 = \frac{1 - \sum_{a=1}^M R(\{y_a\})^2}{m} \le \frac{1}{m}.$$

Therefore

$$\mathbb{E}Z_Y \le D_Y(c)\sqrt{\frac{M}{m}}.$$

Similarly,

$$\mathbb{E}Z_Y^2 = D_Y(c)^2\,\mathbb{E}\Big(\sum_{a=1}^{M-1}|\Delta_a|\Big)^2 \le D_Y(c)^2\,(M-1)\sum_{a=1}^{M-1}\mathbb{E}\Delta_a^2 \le D_Y(c)^2\,\frac{M}{m}.$$

□

Theorem 8.9 (Exact support-box compression for the full semi-discrete dual)**.** *Define*

$$\hat{T}_Y := \sup_{g\in\mathcal{G}_Y} \big\{P_n f_g + R_m g\big\}.$$

Then for every $P \in \mathcal{P}(\mathcal{X})$ and every $R \in \mathcal{P}(Y)$,

$$\mathbb{E}\Big|\hat{T}_Y - T_c(P,R)\Big| \le C\, B_Y(c)\sqrt{\frac{M}{N}},$$

and

$$\mathbb{E}\Big(\hat{T}_Y - T_c(P,R)\Big)^2 \le C\, B_Y(c)^2\,\frac{M}{N},$$

where $C < \infty$ is universal.

Proof. By Theorem 8.6, the population value satisfies

$$T_c(P,R) = \sup_{g\in\mathcal{G}_Y} \big\{P f_g + R g\big\}.$$

The map

$$g \mapsto P f_g + R g$$

is continuous on the compact box $\mathcal{G}_Y$, so let $g_\star \in \mathcal{G}_Y$ attain the supremum. Then

$$\hat{T}_Y - T_c(P,R) \le \sup_{g\in\mathcal{G}_Y} \big\{(P_n - P) f_g + (R_m - R) g\big\} \le Z_X + Z_Y,$$

and also

$$T_c(P,R) - \hat{T}_Y \le (P - P_n) f_{g_\star} + (R - R_m) g_\star \le Z_X + Z_Y.$$

Therefore

$$\Big|\hat{T}_Y - T_c(P,R)\Big| \le Z_X + Z_Y,$$

and

$$\Big(\hat{T}_Y - T_c(P,R)\Big)^2 \le 2Z_X^2 + 2Z_Y^2.$$

Apply Theorem 8.8 and recall that $N = n \wedge m$. ☐

Corollary 8.10 (Geometry-free known-support law for W_p)**.** *Let $p \ge 1$ and let $Y \subset [0,1]^d$ be fixed with $|Y| = M$. Then there exists an estimator $\hat{W}_Y$ such that*

$$\sup_{P\in\mathcal{P}([0,1]^d),\, R\in\mathcal{P}(Y)} \mathbb{E}\Big|\hat{W}_Y - W_p(P,R)\Big| \le C_{d,p}\Big(\frac{M}{N}\Big)^{1/(2p)},$$

and

$$\sup_{P\in\mathcal{P}([0,1]^d),\, R\in\mathcal{P}(Y)} \mathbb{E}\Big(\hat{W}_Y - W_p(P,R)\Big)^2 \le C_{d,p}\Big(\frac{M}{N}\Big)^{1/p}.$$

In particular, if $h \in (0,1]$ and

$$M \le c_{d,p}\, N h^{2p},$$

then

$$\sup_{P\in\mathcal{P}([0,1]^d),\, R\in\mathcal{P}(Y)} \mathbb{E}\Big|\hat{W}_Y - W_p(P,R)\Big| \le C_{d,p}\, h,$$

and

$$\sup_{P\in\mathcal{P}([0,1]^d),\, R\in\mathcal{P}(Y)} \mathbb{E}\Big(\hat{W}_Y - W_p(P,R)\Big)^2 \le C_{d,p}\, h^2.$$

Proof. Apply Theorem 8.9 to the Euclidean cost

$$c(x,y) = \|x-y\|_2^p.$$

On $[0,1]^d \times [0,1]^d$,

$$\|c\|_\infty \le d^{p/2} \qquad \text{and} \qquad D_Y(c) \le d^{p/2},$$

so

$$B_Y(c) \le 2d^{p/2}.$$

Thus there exists $\hat{T}_Y$ with

$$\sup_{P,R} \mathbb{E}\Big|\hat{T}_Y - W_p(P,R)^p\Big| \le C_{d,p}\sqrt{\frac{M}{N}},$$

and

$$\sup_{P,R} \mathbb{E}\Big(\hat{T}_Y - W_p(P,R)^p\Big)^2 \le C_{d,p}\,\frac{M}{N}.$$

Set

$$\hat{W}_Y := (\hat{T}_Y)_+^{1/p}.$$

Since the map $u \mapsto u^{1/p}$ is $1/p$-Hölder on $[0,\infty)$,

$$|\hat{W}_Y - W_p(P,R)| \le \Big|\hat{T}_Y - W_p(P,R)^p\Big|^{1/p}.$$

Hence

$$\mathbb{E}|\hat{W}_Y - W_p(P,R)| \le \Big(\mathbb{E}\Big|\hat{T}_Y - W_p(P,R)^p\Big|\Big)^{1/p} \le C_{d,p}\Big(\frac{M}{N}\Big)^{1/(2p)},$$

and

$$\mathbb{E}(\hat{W}_Y - W_p(P,R))^2 \le \Big(\mathbb{E}\Big(\hat{T}_Y - W_p(P,R)^p\Big)^2\Big)^{1/p} \le C_{d,p}\Big(\frac{M}{N}\Big)^{1/p}.$$

The final claims follow immediately when $M \le c_{d,p} N h^{2p}$. ☐

Corollary 8.11 (Annular support-size law)**.** *Assume $p > 1$. Let $Y \subset [0,1]^d$ be fixed with $|Y| = M$, and let*

$$\mathcal{C}_t \subseteq \Big\{(P,R) \in \mathcal{P}([0,1]^d) \times \mathcal{P}(Y): \ t/2 \le W_p(P,R) \le 2t\Big\}$$

for some $t \in (0,1]$. Then there exists an estimator $\hat{W}_t$ such that

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{W}_t - W_p(P,R)\Big| \le C_{d,p}\, t^{1-p}\sqrt{\frac{M}{N}},$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{W}_t - W_p(P,R)\Big)^2 \le C_{d,p}\, t^{2-2p}\frac{M}{N}.$$

In particular, in the supercritical notation of this paper, if

$$M \le c_{d,p}\, N\eta_N^2 t^{2p-2},$$

then

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{W}_t - W_p(P,R)\Big| \le C_{d,p}\,\eta_N,$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{W}_t - W_p(P,R)\Big)^2 \le C_{d,p}\,\eta_N^2.$$

Proof. Let $\hat{T}_Y$ be the cost estimator from Theorem 8.9 for

$$c(x,y) = \|x-y\|_2^p,$$

and set $\hat{W}_t := (\hat{T}_Y)_+^{1/p}$. By Theorem 3.1, on the annulus $\mathcal{C}_t$,

$$|\hat{W}_t - W_p(P,R)| \le C_p\, t^{1-p}\Big|\hat{T}_Y - W_p(P,R)^p\Big|.$$

Taking expectations and using the cost estimate from Theorem 8.9 gives

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{W}_t - W_p(P,R)\Big| \le C_{d,p}\, t^{1-p}\sqrt{\frac{M}{N}}.$$

Squaring the same pointwise inequality yields

$$\mathbb{E}\Big(\hat{W}_t - W_p(P,R)\Big)^2 \le C_p\, t^{2-2p}\,\mathbb{E}\Big(\hat{T}_Y - W_p(P,R)^p\Big)^2,$$

so the squared bound follows from Theorem 8.9. The last display is immediate when $M \le c_{d,p} N\eta_N^2 t^{2p-2}$. ☐

8.4 Completion of the annular core

We now return to the supercritical problem. Fix $p > 1$, $d > 2p$, let

$$h_N \asymp \eta_N = (N \log N)^{-1/d},$$

and let G_{h_N} be the critical grid from Section 4. For $t \in [\eta_N, r_{N,p}]$, recall the linearized transport value

$$T_{t,p}^{\mathrm{lin}} = T_{c_{t,p}^{\mathrm{lin}}}$$

from Theorem 4.3, where $c_{t,p}^{\mathrm{lin}}$ is a bounded cost on $[0,1]^d \times [0,1]^d$. Restrict its second argument to the critical grid:

$$c_{t,p}^{\mathrm{lin}}: \ [0,1]^d \times G_{h_N} \to \mathbb{R}.$$

Theorem 8.12 (Annular completion via dual catalogs)**.** *Assume $p > 1$ and $d > 2p$. Let*

$$\mathcal{C}_t \subseteq \Big\{(P,R) \in \mathcal{P}([0,1]^d) \times \mathcal{P}(G_{h_N}): \ t/2 \le W_p(P,R) \le 2t\Big\}$$

for some $t \in [\eta_N, r_{N,p}]$. Suppose there exist constants $A_0, A_1, B_0 < \infty$ and a finite dual catalog

$$\mathscr{A}_t = \{(f_k^{(t)}, g_k^{(t)})\}_{k=1}^{K_t}$$

for the linearized cost $c_{t,p}^{\mathrm{lin}}$ such that

$$\|f_k^{(t)}\|_\infty \le B_0, \qquad \|g_k^{(t)}\|_\infty \le B_0 \qquad (1 \le k \le K_t),$$

$$\sup_{(P,R)\in\mathcal{C}_t} \Delta_{\mathscr{A}_t}^{\mathrm{lin}}(P,R) \le A_0\,\eta_N\, t^{p-1},$$

and

$$\log(2K_t) \le A_1\, N\eta_N^2 t^{2p-2}.$$

Then there exists an estimator $\hat{W}_t$ such that

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{W}_t - W_p(P,R)\Big| \le C_{d,p,A_0,A_1,B_0}\,\eta_N,$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{W}_t - W_p(P,R)\Big)^2 \le C_{d,p,A_0,A_1,B_0}\, \eta_N^2.$$

Proof. Apply Theorem 8.2 to the bounded cost $c^{\mathrm{lin}}_{t,p}$ and the finite catalog $\mathscr{A}_t$. This yields an estimator $\hat{T}_t$ satisfying

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{T}_t - T^{\mathrm{lin}}_{t,p}(P,R)\Big| \le A_0\, \eta_N t^{p-1} + CB_0\sqrt{\frac{\log(2K_t)}{N}},$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{T}_t - T^{\mathrm{lin}}_{t,p}(P,R)\Big)^2 \le 2A_0^2\, \eta_N^2 t^{2p-2} + CB_0^2\frac{\log(2K_t)}{N}.$$

Under the entropy budget on K_t,

$$\sqrt{\frac{\log(2K_t)}{N}} \le \sqrt{A_1}\, \eta_N\, t^{p-1}, \qquad \frac{\log(2K_t)}{N} \le A_1\, \eta_N^2 t^{2p-2}.$$

Therefore

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{T}_t - T^{\mathrm{lin}}_{t,p}(P,R)\Big| \le C_{A_0,A_1,B_0}\, \eta_N\, t^{p-1},$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{T}_t - T^{\mathrm{lin}}_{t,p}(P,R)\Big)^2 \le C_{A_0,A_1,B_0}\, \eta_N^2\, t^{2p-2}.$$

Now apply the fixed-annulus completion theorem Theorem 4.4. □

Corollary 8.13 (Annular completion via low-dimensional dual manifolds)**.** *Assume the setup of Theorem 8.12. Suppose that for the linearized cost $c^{\mathrm{lin}}_{t,p}$ there exists a parameter set $\Theta_t \subset [-R,R]^q$ and a dual-feasible family*

$$\{(f^{(t)}_\theta, g^{(t)}_\theta) : \theta \in \Theta_t\}$$

such that

$$\|f^{(t)}_\theta\|_\infty \le B_0, \qquad \|g^{(t)}_\theta\|_\infty \le B_0 \qquad (\theta \in \Theta_t),$$
$$\|f^{(t)}_\theta - f^{(t)}_{\theta'}\|_\infty + \|g^{(t)}_\theta - g^{(t)}_{\theta'}\|_\infty \le L\|\theta-\theta'\|_\infty \qquad (\theta,\theta' \in \Theta_t),$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \Delta^{\mathrm{lin}}_{\Theta_t}(P,R) \le A_0\, \eta_N\, t^{p-1}.$$

If moreover

$$q\log\Big(1 + \frac{2RL}{\eta_N t^{p-1}}\Big) \le A_1\, N\eta_N^2 t^{2p-2},$$

then there exists an estimator $\hat{W}_t$ satisfying

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big|\hat{W}_t - W_p(P,R)\Big| \le C_{d,p,A_0,A_1,B_0,R,L}\, \eta_N,$$

and

$$\sup_{(P,R)\in\mathcal{C}_t} \mathbb{E}\Big(\hat{W}_t - W_p(P,R)\Big)^2 \le C_{d,p,A_0,A_1,B_0,R,L}\, \eta_N^2.$$

Proof. Apply Theorem 8.4 with

$$\varepsilon := \eta_N t^{p-1}.$$

The complexity assumption ensures that the resulting cost estimator satisfies the hypotheses of Theorem 4.4, so the same annular completion argument as in Theorem 8.12 applies. □

Corollary 8.14 (Critical W_1 phase families on the one-sided grid)**.** *Assume $d > 2$ and let*

$$h_N \asymp (N\log N)^{-1/d}.$$

Let

$$\mathcal{C}_N \subseteq \mathcal{P}([0,1]^d) \times \mathcal{P}(G_{h_N}).$$

Suppose there exist constants $A_0, A_1, B_0 < \infty$ and a finite dual catalog

$$\mathscr{A}_N = \{(f_k, g_k)\}_{k=1}^{K_N}$$

for the Euclidean cost

$$c(x,y) = \|x-y\|_2 \qquad (x \in [0,1]^d,\ y \in G_{h_N})$$

such that

$$\|f_k\|_\infty \le B_0, \qquad \|g_k\|_\infty \le B_0 \qquad (1 \le k \le K_N),$$
$$\sup_{(P,R)\in\mathcal{C}_N} \Delta_{\mathscr{A}_N}(P,R) \le A_0 h_N,$$

and

$$\log(2K_N) \le A_1 N h_N^2.$$

Then

$$\sup_{(P,R)\in\mathcal{C}_N} \mathbb{E}\Big|\hat{W} - W_1(P,R)\Big| \le C_{A_0,A_1,B_0,d}\, h_N,$$

and

$$\sup_{(P,R)\in\mathcal{C}_N} \mathbb{E}\big(\hat{W} - W_1(P,R)\big)^2 \le C_{A_0,A_1,B_0,d}\, h_N^2.$$

In particular, every exact K_N-phase family is sharp whenever

$$K_N \le \exp(cNh_N^2) = \exp\!\big(cN^{1-2/d}(\log N)^{-2/d}\big)$$

for a sufficiently small constant $c = c(d, B_0) > 0$.

Proof. Apply Theorem 8.2 directly to the W_1 cost. Under the entropy budget,

$$B_0\sqrt{\frac{\log(2K_N)}{N}} \lesssim h_N, \qquad B_0^2\frac{\log(2K_N)}{N} \lesssim h_N^2,$$

which gives the claim. □

8.5 The full one-sided semidual class already has full critical entropy

Corollary 8.14 shows that an $O(h_N)$ phase approximation of entropy $O(Nh_N^2)$ would be statistically sufficient on the one-sided critical grid. The next theorem shows that such a route cannot proceed by globally covering the full semidual potential class itself. After anchoring at a single grid point, the full one-sided semidual class is already $O(h)$-dense in the entire normalized Lipschitz ball and has metric entropy of exact order h^{-d} at scale h.

Fix a dyadic mesh $h = 2^{-J}$ and a distinguished grid point $z_\star \in G_h$. Define the anchored one-sided semidual class

$$\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star) := \big\{\tilde{f}_g := f_g - f_g(z_\star) :\ g : G_h \to \mathbb{R}\big\}, \qquad f_g(x) := \min_{y\in G_h}\big(\|x-y\|_2 - g(y)\big),$$

and the anchored Lipschitz ball

$$\mathrm{Lip}_1(z_\star) := \big\{f : [0,1]^d \to \mathbb{R} :\ \mathrm{Lip}(f) \le 1,\ f(z_\star) = 0\big\}.$$

Proposition 8.15 (McShane saturation of the full one-sided semidual class)**.** *Every element of $\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ belongs to $\mathrm{Lip}_1(z_\star)$. Conversely, for every $f \in \mathrm{Lip}_1(z_\star)$, if*

$$g_f(y) := -f(y) \qquad (y \in G_h),$$

then

$$f_{g_f}(z_\star) = 0$$

and

$$0 \le f_{g_f}(x) - f(x) \le \sqrt{d}\, h \qquad (x \in [0,1]^d).$$

In particular,

$$\sup_{f\in\mathrm{Lip}_1(z_\star)}\ \inf_{u\in\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)} \|f-u\|_\infty \le \sqrt{d}\, h.$$

Proof. For every $y \in G_h$, the map

$$x \longmapsto \|x-y\|_2 - g(y)$$

is 1-Lipschitz on $[0,1]^d$. Hence f_g, being the pointwise minimum of such functions, is also 1-Lipschitz, and the same is true for $\tilde{f}_g = f_g - f_g(z_\star)$. Since $\tilde{f}_g(z_\star) = 0$, we obtain

$$\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star) \subseteq \mathrm{Lip}_1(z_\star).$$

Now fix $f \in \mathrm{Lip}_1(z_\star)$ and set $g_f(y) := -f(y)$. Then

$$f_{g_f}(x) = \min_{y\in G_h}\big(\|x-y\|_2 + f(y)\big).$$

Because f is 1-Lipschitz,

$$f(y) + \|x-y\|_2 \ge f(x) \qquad (x \in [0,1]^d,\ y \in G_h),$$

hence

$$f_{g_f}(x) \ge f(x).$$

On the other hand, taking $y = q_h(x)$ gives

$$f_{g_f}(x) \le \|x-q_h(x)\|_2 + f(q_h(x)) \le \|x-q_h(x)\|_2 + f(x) + \|x-q_h(x)\|_2.$$

Since

$$\|x-q_h(x)\|_2 \le \frac{\sqrt{d}}{2}\, h,$$

we obtain

$$f_{g_f}(x) \le f(x) + \sqrt{d}\, h.$$

Finally, at $x = z_\star$ we may choose $y = z_\star$ and use $f(z_\star) = 0$ to get

$$0 \le f_{g_f}(z_\star) \le \|z_\star - z_\star\|_2 + f(z_\star) = 0.$$

Thus $f_{g_f}(z_\star) = 0$, so $f_{g_f} \in \widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ and the uniform approximation bound follows. □

Lemma 8.16 (Large packings inside the anchored Lipschitz ball)**.** *There exist constants $r_d, c_d \in (0,1)$ depending only on d such that the following holds. For every $z_\star \in [0,1]^d$ and every $r \in (0, r_d]$, the class $\mathrm{Lip}_1(z_\star)$ contains functions*

$$f_1, \dots, f_M$$

satisfying

$$M \ge \exp(c_d r^{-d})$$

and

$$\|f_a - f_b\|_\infty \ge r \qquad (1 \le a < b \le M).$$

Proof. For each coordinate $1 \le i \le d$, define

$$I_i := \begin{cases} [1/16,\ 3/16], & (z_\star)_i \ge 1/2, \\ [13/16,\ 15/16], & (z_\star)_i < 1/2. \end{cases}$$

Let

$$Q_\star := \prod_{i=1}^{d} I_i.$$

Then $Q_\star \subset [0,1]^d$, has side length 1/8, and

$$\operatorname{dist}(z_\star, Q_\star) \ge \frac{5}{16}.$$

Assume $r \le 1/128$. Inside each interval I_i choose the arithmetic progression

$$\Lambda_i = \Big\{ a_i + 4rk :\ 0 \le k \le m_r - 1 \Big\}, \qquad m_r := \Big\lfloor \frac{1}{32r} \Big\rfloor,$$

where $a_i = 1/16$ or $13/16$ according to the definition of I_i. Then every point of

$$\Lambda_\star := \Lambda_1 \times \cdots \times \Lambda_d \subseteq Q_\star$$

is at distance at least $4r$ from every other point of $\Lambda_\star$, and

$$|\Lambda_\star| = m_r^d \ge c_d r^{-d}$$

for a constant $c_d > 0$ and all sufficiently small r. Enumerate $\Lambda_\star = \{z_1, \ldots, z_{M_0}\}$.

For each $1 \le j \le M_0$, define the tent function

$$\phi_j(x) := \Big(r - \|x - z_j\|_2 \Big)_+ \qquad (x \in [0,1]^d).$$

Each ϕ_j is 1-Lipschitz, satisfies

$$\phi_j(z_\star) = 0$$

because $\operatorname{dist}(z_\star, Q_\star) \ge 5/16 > r$, and the supports are pairwise disjoint because the centers are $4r$-separated while each support has radius r.

For every binary vector $\sigma = (\sigma_1, \ldots, \sigma_{M_0}) \in \{0,1\}^{M_0}$ define

$$f_\sigma := \sum_{j=1}^{M_0} \sigma_j \phi_j.$$

Since the supports of the ϕ_j are disjoint and the tents are nonnegative,

$$f_\sigma = \max\Big(0, \sigma_1\phi_1, \ldots, \sigma_{M_0}\phi_{M_0}\Big),$$

hence f_σ is 1-Lipschitz as a maximum of 1-Lipschitz functions. Also $f_\sigma(z_\star) = 0$, so

$$f_\sigma \in \mathrm{Lip}_1(z_\star).$$

If $\sigma \ne \tau$, choose j with $\sigma_j \ne \tau_j$. At the point z_j, all tents except ϕ_j vanish, so

$$|f_\sigma(z_j) - f_\tau(z_j)| = |\sigma_j - \tau_j|\, \phi_j(z_j) = r.$$

Therefore

$$\|f_\sigma - f_\tau\|_\infty \ge r.$$

Taking all 2^{M_0} choices of σ gives

$$M := 2^{M_0} \ge \exp(c_d r^{-d}),$$

after changing c_d if necessary. This proves the claim for $r \le r_d := 1/128$. ☐

Theorem 8.17 (Exact h^{-d} entropy law for the full one-sided semidual class)**.** *Fix*

$$A > \frac{\sqrt{d}}{2}.$$

Then there exist constants $c_{d,A}, C_{d,A}, h_{d,A} \in (0,\infty)$ *such that for every dyadic*

$$h \in (0, h_{d,A}]$$

and every $z_\star \in G_h$*,*

$$c_{d,A}\, h^{-d} \le \log \mathcal{N}\Big(\widetilde{\mathcal{F}}_h^{\mathrm{sd}}(z_\star), Ah, \|\cdot\|_\infty\Big) \le C_{d,A}\, h^{-d}.$$

Proof. For the lower bound, apply Lemma 8.16 with

$$r := 2(A + \sqrt{d})\, h.$$

For $h \le h_{d,A}$ this satisfies $r \le r_d$, so there exist

$$f_1, \ldots, f_M \in \mathrm{Lip}_1(z_\star)$$

with

$$M \ge \exp\Big(c_{d,A} h^{-d}\Big)$$

and

$$\|f_a - f_b\|_\infty \ge 2(A + \sqrt{d})\, h \qquad (a \ne b).$$

By Proposition 8.15, for each a there exists

$$u_a \in \widetilde{\mathcal{F}}_h^{\mathrm{sd}}(z_\star)$$

such that

$$\|u_a - f_a\|_\infty \le \sqrt{d}\, h.$$

Hence, for $a \ne b$,

$$\|u_a - u_b\|_\infty \ge \|f_a - f_b\|_\infty - \|u_a - f_a\|_\infty - \|u_b - f_b\|_\infty \ge 2Ah.$$

Therefore $\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ contains a $2Ah$-packing of cardinality at least $\exp(c_{d,A}h^{-d})$, which implies

$$\log\mathcal{N}\Big(\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star), Ah, \|\cdot\|_\infty\Big) \ge c_{d,A}\, h^{-d}.$$

For the upper bound, let

$$\Gamma_h$$

be the nearest-neighbor graph on G_h and fix a spanning tree T_h of Γ_h rooted at $z_\star$. Since $|G_h| = h^{-d}$, the tree T_h has exactly $h^{-d} - 1$ edges.

Set

$$\rho := 2A - \sqrt{d} > 0.$$

For each $u \in \widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ and each $y \in G_h$, let $a_u(y)$ be the nearest multiple of ρh to $u(y)$. Then

$$|a_u(y) - u(y)| \le \frac{\rho h}{2} \qquad (y \in G_h),$$

and since $u(z_\star) = 0$, we may take

$$a_u(z_\star) = 0.$$

If $y, z \in G_h$ are neighbors in Γ_h, then $\|y - z\|_2 = h$ and u is 1-Lipschitz, so

$$|u(y) - u(z)| \le h.$$

Therefore

$$|a_u(y) - a_u(z)| \le |u(y) - u(z)| + |a_u(y) - u(y)| + |a_u(z) - u(z)| \le (1 + \rho)h.$$

It follows that for each oriented edge $e = (y, z)$ of T_h,

$$\frac{a_u(y) - a_u(z)}{\rho h}$$

is an integer whose absolute value is bounded by

$$1 + \rho^{-1}.$$

Hence each edge increment admits at most

$$L_{d,A} := 2\Big\lceil 1 + \rho^{-1} \Big\rceil + 1$$

possible values. Because $a_u(z_\star) = 0$, the whole rounded array $a_u : G_h \to \mathbb{R}$ is determined by these edge increments, so the number of possible arrays satisfies

$$|\mathscr{A}_h| \le L_{d,A}^{h^{-d}-1} \le \exp\big(C_{d,A} h^{-d}\big)$$

for a constant $C_{d,A} < \infty$.

For each admissible array $a \in \mathscr{A}_h$, define the piecewise-constant function

$$v_a(x) := a(q_h(x)) \qquad (x \in [0, 1]^d).$$

If $u \in \widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ and $a = a_u$, then

$$|u(x) - v_a(x)| \le |u(x) - u(q_h(x))| + |u(q_h(x)) - a_u(q_h(x))| \le \frac{\sqrt{d}}{2} h + \frac{\rho h}{2} = Ah.$$

Thus the family $\{v_a :\ a \in \mathscr{A}_h\}$ is an Ah-cover of $\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star)$ in $\|\cdot\|_\infty$, and therefore

$$\log\mathcal{N}\Big(\widetilde{\mathcal{F}}_h^{\rm sd}(z_\star), Ah, \|\cdot\|_\infty\Big) \le C_{d,A}\, h^{-d}.$$

This proves the theorem. □

Corollary 8.18 (No low-entropy global phase cover on the full critical one-sided grid)**.** *Assume $d > 2$ and let*

$$h_N \asymp (N \log N)^{-1/d}.$$

Fix

$$A > \frac{\sqrt{d}}{2}$$

and choose any $z_{\star,N} \in G_{h_N}$. Then

$$\log\mathcal{N}\Big(\widetilde{\mathcal{F}}_{h_N}^{\rm sd}(z_{\star,N}), Ah_N, \|\cdot\|_\infty\Big) \asymp_{d,A} N \log N.$$

Consequently, any global $O(h_N)$ approximation of the full one-sided semidual class by a finite family of candidate phases requires

$$K_N \ge \exp(c_{d,A} N \log N)$$

for some $c_{d,A} > 0$. Moreover, if

$$\{u_\theta :\ \theta \in [-R, R]^q\}$$

is an L-Lipschitz family in $\|\cdot\|_\infty$ and its $(Ah_N/2)$-neighborhood covers

$$\widetilde{\mathcal{F}}_{h_N}^{\rm sd}(z_{\star,N}),$$

then

$$q \log\Big(1 + \frac{4RL}{Ah_N}\Big) \ge c_{d,A} N \log N$$

for a constant $c_{d,A} > 0$. In particular, the low-entropy regime of Corollary 8.14 cannot be reached by globally covering the full critical one-sided semidual class itself.

Proof. The entropy statement is Theorem 8.17 together with

$$h_N^{-d} \asymp N \log N.$$

The lower bound on K_N is immediate.

For the parameterized claim, let

$$r_N := \frac{Ah_N}{2L}.$$

An r_N-net of $[-R,R]^q$ in $\|\cdot\|_\infty$ has cardinality at most

$$\left(1+\frac{2R}{r_N}\right)^q = \left(1+\frac{4RL}{Ah_N}\right)^q.$$

Because the map $\theta \mapsto u_\theta$ is L-Lipschitz, the image of this parameter net is an $(Ah_N/2)$-net of the family $\{u_\theta\}$ in $\|\cdot\|_\infty$. By the covering hypothesis, it follows that

$$\widetilde{\mathcal{F}}^{\mathrm{sd}}_{h_N}(z_{\star,N})$$

admits an Ah_N-cover of cardinality at most

$$\left(1+\frac{4RL}{Ah_N}\right)^q.$$

Applying Theorem 8.17 yields

$$q\log\!\left(1+\frac{4RL}{Ah_N}\right) \ge c_{d,A}h_N^{-d} \asymp c_{d,A}N\log N.$$

Finally,

$$Nh_N^2 = N^{1-2/d}(\log N)^{-2/d} = o(N\log N),$$

so the entropy budget required by Theorem 8.14 is exponentially smaller than the entropy of a global $O(h_N)$ cover of the full one-sided semidual class. □

Proposition 8.19 (Empirical-process saturation on the full one-sided semidual class)**.** *Fix a dyadic mesh $h = 2^{-J}$ and $z_\star \in G_h$. For $P \in \mathcal{P}([0,1]^d)$ and i.i.d. $X_1,\dots,X_N \sim P$, write*

$$P_N := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$$

and

$$\Xi_h(P) := \sup_{u\in\widetilde{\mathcal{F}}^{\mathrm{sd}}_h(z_\star)} \left|(P_N-P)u\right|.$$

Then almost surely

$$\Xi_h(P) \ge W_1(P_N,P) - 2\sqrt{d}\,h.$$

Consequently, there exists a constant $c_d > 0$ such that

$$\sup_{P\in\mathcal{P}([0,1]^d)} \mathbb{E}\Xi_h(P) \ge c_d\,N^{-1/d}$$

whenever

$$h \le c_d\,N^{-1/d}.$$

In particular, at the critical mesh $h_N \asymp (N\log N)^{-1/d}$, the raw empirical supremum over the full one-sided semidual class still fluctuates at the plug-in scale $N^{-1/d}$.

Proof. By Kantorovich–Rubinstein duality,

$$W_1(P_N,P) = \sup_{f\in\mathrm{Lip}_1(z_\star)}(P_N-P)f = \sup_{f\in\mathrm{Lip}_1(z_\star)}\left|(P_N-P)f\right|,$$

because $f\in\mathrm{Lip}_1(z_\star)$ implies $-f\in\mathrm{Lip}_1(z_\star)$. Fix $\varepsilon>0$ and choose $f_\varepsilon\in\mathrm{Lip}_1(z_\star)$ such that

$$(P_N-P)f_\varepsilon \ge W_1(P_N,P)-\varepsilon.$$

By Proposition 8.15, there exists

$$u_\varepsilon \in \widetilde{\mathcal{F}}^{\mathrm{sd}}_h(z_\star)$$

with

$$\|u_\varepsilon - f_\varepsilon\|_\infty \le \sqrt{d}\,h.$$

Since the signed measure $P_N - P$ has total variation norm at most 2,

$$\left|(P_N-P)(u_\varepsilon-f_\varepsilon)\right| \le 2\sqrt{d}\,h.$$

Therefore

$$\Xi_h(P) \ge (P_N-P)u_\varepsilon \ge W_1(P_N,P)-\varepsilon-2\sqrt{d}\,h.$$

Letting $\varepsilon\downarrow 0$ gives

$$\Xi_h(P) \ge W_1(P_N,P)-2\sqrt{d}\,h.$$

For the lower bound on the expectation, choose N points

$$z_1,\dots,z_N \in [0,1]^d$$

with pairwise separation at least $c_dN^{-1/d}$ and define

$$P_0 := \frac{1}{N}\sum_{j=1}^N \delta_{z_j}.$$

If K_j denotes the empirical count of z_j, then

$$\mathrm{TV}(P_{0,N},P_0) = \frac{1}{2N}\sum_{j=1}^N |K_j-1|,$$

and every unmatched unit of mass must move by at least $c_d N^{-1/d}$. Hence

$$W_1(P_{0,N}, P_0) \geq c_d N^{-1/d}\, \mathrm{TV}(P_{0,N}, P_0).$$

By symmetry,

$$\mathbb{E}\,\mathrm{TV}(P_{0,N}, P_0) = \frac{1}{2}\,\mathbb{E}|K_1 - 1| \geq \frac{1}{2}\,\mathbb{P}(K_1 = 0) = \frac{1}{2}\Big(1 - \frac{1}{N}\Big)^N \geq c$$

for a universal constant $c > 0$ and all $N \geq 2$. Thus

$$\mathbb{E}W_1(P_{0,N}, P_0) \geq c'_d N^{-1/d}.$$

Combining this with the deterministic inequality above yields

$$\mathbb{E}\Xi_h(P_0) \geq c'_d N^{-1/d} - 2\sqrt{d}\, h.$$

If $h \leq c_d N^{-1/d}$ with c_d sufficiently small, the right-hand side is bounded below by a positive constant multiple of $N^{-1/d}$. Taking the supremum over P proves the claim. □

Remark 8.20. Theorem 8.17 and Proposition 8.19 do not solve the one-sided critical problem by themselves. What they do show is that two of the most natural global routes are impossible on the full class: a global $O(h_N)$ semidual phase cover already has entropy $\exp(\Theta(N \log N))$, and the raw empirical supremum over the full semidual still lives at the plug-in scale $N^{-1/d}$. Any eventual critical-rate estimator on the full one-sided grid must therefore exploit a subtler value-estimation mechanism than either global phase covering or raw empirical optimization.

9 Continuum lifts and coarse quantization

We now pass from solved discrete models to continuum classes. The exact critical classes of Section 5 were originally formulated on finite Euclidean supports. We show that the same statistical content survives under arbitrary unresolved microstructure inside known $O(h)$ blobs. This produces genuinely continuum, and even absolutely continuous, critical classes with the same minimax laws.

For any statistical class $\mathfrak{C} \subset \mathcal{P}([0,1]^d)^2$, write

$$M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) := \inf_{\hat{W}} \sup_{(P,Q)\in\mathfrak{C}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big|,$$

$$M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) := \inf_{\hat{W}} \sup_{(P,Q)\in\mathfrak{C}} \mathbb{E}\Big(\hat{W} - W_p(P,Q)\Big)^2.$$

Definition 9.1 (Blob lift of an atomic class)**.** Fix $h > 0$, $\kappa > 0$, a finite anchor set

$$S = \{z_1, \ldots, z_M\} \subset [0,1]^d,$$

and pairwise disjoint measurable blobs

$$B_1, \ldots, B_M \subset [0,1]^d$$

such that

$$z_j \in B_j, \qquad \sup_{x\in B_j} \|x - z_j\|_2 \leq \kappa h \qquad \text{for every } 1 \leq j \leq M.$$

Let

$$U := \bigcup_{j=1}^{M} B_j.$$

For a probability measure P supported on U, define its coarse quantization

$$\Gamma_{\mathcal{B}} P := \sum_{j=1}^{M} P(B_j)\, \delta_{z_j}.$$

For every class $\mathfrak{C} \subset \mathcal{P}(S)^2$, define its blob lift

$$\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) := \Big\{(P,Q) \in \mathcal{P}(U)^2 :\ (\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q) \in \mathfrak{C}\Big\},$$

and its absolutely continuous blob lift

$$\mathcal{L}^{\mathrm{ac}}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) := \Big\{(P,Q) \in \mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) :\ P, Q \ll dx\Big\}.$$

Lemma 9.2 (Coarse quantization costs only $O(h)$)**.** *In the setting of Theorem 9.1, every $P \in \mathcal{P}(U)$ satisfies*

$$W_p(P, \Gamma_{\mathcal{B}}P) \leq \kappa h.$$

Consequently, every $(P,Q) \in \mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})$ obeys

$$\Big|W_p(P,Q) - W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q)\Big| \leq 2\kappa h.$$

Proof. For each j, couple every point $x \in B_j$ to the anchor z_j. This produces a coupling between P and $\Gamma_{\mathcal{B}}P$ whose transport cost is at most

$$\int \|x - z_j\|_2^p \, dP(x) \leq (\kappa h)^p.$$

Hence $W_p(P, \Gamma_{\mathcal{B}}P) \leq \kappa h$. Applying the triangle inequality twice gives

$$\Big|W_p(P,Q) - W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q)\Big| \leq W_p(P, \Gamma_{\mathcal{B}}P) + W_p(Q, \Gamma_{\mathcal{B}}Q) \leq 2\kappa h.$$

□

Theorem 9.3 (Exact coarse-quantization equivalence)**.** *In the setting of Theorem 9.1, for every atomic class $\mathfrak{C} \subset \mathcal{P}(S)^2$ one has*

$$M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) - 2\kappa h \leq M_{n,m}^{\mathrm{abs}}\big(\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\big) \leq M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) + 2\kappa h,$$

and

$$\frac{1}{2}\, M_{n,m}^{\mathrm{sq}}(\mathfrak{C}) - 4\kappa^2 h^2 \leq M_{n,m}^{\mathrm{sq}}\big(\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\big) \leq 2\, M_{n,m}^{\mathrm{sq}}(\mathfrak{C}) + 8\kappa^2 h^2.$$

Proof. We begin with the upper bounds. Fix $\varepsilon > 0$ and choose an estimator $\hat{W}_S$ on the atomic class $\mathfrak{C}$ such that

$$\sup_{(\mu,\nu)\in\mathfrak{C}} \mathbb{E}\Big|\hat{W}_S - W_p(\mu,\nu)\Big| \leq M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) + \varepsilon,$$

and

$$\sup_{(\mu,\nu)\in\mathfrak{C}} \mathbb{E}\Big(\hat{W}_S - W_p(\mu,\nu)\Big)^2 \leq M_{n,m}^{\mathrm{sq}}(\mathfrak{C}) + \varepsilon.$$

Now let $(P,Q) \in \mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})$. Because the blobs are pairwise disjoint and known, every raw sample point determines its blob index. Replacing each sample in B_j by the anchor z_j therefore yields a measurable pair of anchor samples distributed exactly as i.i.d. samples from

$$\Gamma_{\mathcal{B}}P \qquad \text{and} \qquad \Gamma_{\mathcal{B}}Q.$$

Apply $\hat{W}_S$ to these anchor samples, and denote the resulting statistic by $\hat{W}_{\mathcal{B}}$. Then

$$\mathbb{E}\Big|\hat{W}_{\mathcal{B}} - W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q)\Big| \leq M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) + \varepsilon,$$

and similarly for the squared risk.

By Theorem 9.2,

$$\Big|W_p(P,Q) - W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q)\Big| \leq 2\kappa h.$$

Hence

$$\mathbb{E}\Big|\hat{W}_{\mathcal{B}} - W_p(P,Q)\Big| \leq M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) + \varepsilon + 2\kappa h.$$

For the squared risk, write

$$\hat{W}_{\mathcal{B}} - W_p(P,Q) = \Big(\hat{W}_{\mathcal{B}} - W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q)\Big) + \Big(W_p(\Gamma_{\mathcal{B}}P, \Gamma_{\mathcal{B}}Q) - W_p(P,Q)\Big)$$

and use $(a+b)^2 \leq 2a^2 + 2b^2$ together with Theorem 9.2 to obtain

$$\mathbb{E}\Big(\hat{W}_{\mathcal{B}} - W_p(P,Q)\Big)^2 \leq 2\Big(M_{n,m}^{\mathrm{sq}}(\mathfrak{C}) + \varepsilon\Big) + 8\kappa^2 h^2.$$

Taking the supremum over (P,Q) and then letting $\varepsilon \downarrow 0$ proves the upper bounds.

For the lower bounds, fix any estimator $\widetilde{W}$ on the lifted class $\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})$. Choose arbitrary probability measures

$$\lambda_j \in \mathcal{P}(B_j), \qquad 1 \leq j \leq M,$$

and define the thickening operator

$$T(\mu) := \sum_{j=1}^{M} \mu(\{z_j\})\, \lambda_j, \qquad \mu \in \mathcal{P}(S).$$

By construction,

$$\Gamma_{\mathcal{B}}T(\mu) = \mu,$$

hence

$$(T(\mu), T(\nu)) \in \mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) \qquad \text{for every } (\mu,\nu) \in \mathfrak{C}.$$

Now suppose we observe atomic samples

$$Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} \mu, \qquad T_1, \ldots, T_m \overset{\text{i.i.d.}}{\sim} \nu$$

for some $(\mu,\nu) \in \mathfrak{C}$. Conditionally on these samples, replace each occurrence of z_j by an independent draw from λ_j. The resulting randomized samples

$$\widetilde{Z}_1, \ldots, \widetilde{Z}_n, \qquad \widetilde{T}_1, \ldots, \widetilde{T}_m$$

are i.i.d. from $T(\mu)$ and $T(\nu)$. Define the induced atomic estimator

$$\widetilde{W}_S := \widetilde{W}(\widetilde{Z}_1, \ldots, \widetilde{Z}_n; \widetilde{T}_1, \ldots, \widetilde{T}_m).$$

Then

$$\mathbb{E}\Big|\widetilde{W}_S - W_p(T(\mu), T(\nu))\Big| \leq \sup_{(P,Q)\in\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})} \mathbb{E}\Big|\widetilde{W} - W_p(P,Q)\Big|.$$

Applying Theorem 9.2 to $T(\mu)$ and $T(\nu)$ gives

$$\Big|W_p(T(\mu), T(\nu)) - W_p(\mu,\nu)\Big| \leq 2\kappa h.$$

Therefore

$$\mathbb{E}\Big|\widetilde{W}_S - W_p(\mu,\nu)\Big| \leq \sup_{(P,Q)\in\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})} \mathbb{E}\Big|\widetilde{W} - W_p(P,Q)\Big| + 2\kappa h.$$

Taking the supremum over $(\mu,\nu) \in \mathfrak{C}$ and then the infimum over $\widetilde{W}$ yields

$$M_{n,m}^{\mathrm{abs}}(\mathfrak{C}) \leq M_{n,m}^{\mathrm{abs}}\Big(\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\Big) + 2\kappa h,$$

which is the desired absolute-risk lower bound.

For the squared risk, the same construction gives

$$\mathbb{E}\Big(\widetilde{W}_S - W_p(\mu,\nu)\Big)^2 \leq 2 \sup_{(P,Q)\in\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})} \mathbb{E}\Big(\widetilde{W} - W_p(P,Q)\Big)^2 + 8\kappa^2 h^2.$$

Taking the supremum over (μ,ν) and then the infimum over $\widetilde{W}$ proves

$$M_{n,m}^{\mathrm{sq}}(\mathfrak{C}) \leq 2\, M_{n,m}^{\mathrm{sq}}\Big(\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\Big) + 8\kappa^2 h^2,$$

equivalently

$$M^{\mathrm{sq}}_{n,m}\big(\mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\big) \ge \frac{1}{2}\, M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) - 4\kappa^2 h^2.$$

□

Corollary 9.4 (Absolutely continuous blob lifts)**.** *Assume in addition that every blob B_j has positive Lebesgue measure. Then*

$$M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) - 2\kappa h \le M^{\mathrm{abs}}_{n,m}\big(\mathcal{L}^{\mathrm{ac}}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\big) \le M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) + 2\kappa h,$$

and

$$\frac{1}{2}\, M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) - 4\kappa^2 h^2 \le M^{\mathrm{sq}}_{n,m}\big(\mathcal{L}^{\mathrm{ac}}_{h,\kappa,\mathcal{B}}(\mathfrak{C})\big) \le 2\, M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) + 8\kappa^2 h^2.$$

Proof. The upper bounds follow immediately from

$$\mathcal{L}^{\mathrm{ac}}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) \subset \mathcal{L}_{h,\kappa,\mathcal{B}}(\mathfrak{C})$$

and the upper half of Theorem 9.3. For the lower bounds, choose

$$\lambda_j := \frac{\mathbf{1}_{B_j}}{|B_j|}\, dx.$$

Then every thickened measure $T(\mu) = \sum_j \mu(\{z_j\})\lambda_j$ is absolutely continuous, so

$$(T(\mu), T(\nu)) \in \mathcal{L}^{\mathrm{ac}}_{h,\kappa,\mathcal{B}}(\mathfrak{C}) \qquad \text{for every } (\mu,\nu) \in \mathfrak{C}.$$

Repeating verbatim the lower-bound reduction from the proof of Theorem 9.3 gives the claim. □

Corollary 9.5 (The exact critical laws survive on genuinely continuum classes)**.** *Assume $d > 2p$. For each Euclidean realization constructed in Section 5, choose $\kappa > 0$ smaller than the corresponding bottom-scale separation constant and select pairwise disjoint positive-measure blobs around the support points, each contained in the Euclidean ball of radius κh around its anchor at the finest scale h of the class. Then, under the same sample-size hypotheses as in the parent theorems, the following absolutely continuous blob lifts satisfy the same exact minimax laws as their atomic ancestors:*

(i) *the finite-band lift of $\mathcal{C}^{\mathrm{band}}_{h,L}$ from Theorem 5.19;*
(ii) *the packed direct-sum lift of $\mathcal{C}^{\mathrm{pack}}_{h}$ from Theorem 5.24;*
(iii) *the dyadic pair-isolation lift of $\mathcal{C}^{\mathrm{iso}}_{h}$ from Theorem 5.26;*
(iv) *the nested near-critical lift of $\mathcal{C}^{\mathrm{nest}}_{B,L}$ from Theorem 5.30, with class scale $h_{B,L}$;*
(v) *the critical laminar lift of $\mathcal{C}^{\mathrm{lam,crit}}_{L}$ from Theorem 5.35, with class scale $h^{\mathrm{lam,crit}}_{L}$.*

In each case,

$$M^{\mathrm{abs}}_{n,m} \asymp h, \qquad M^{\mathrm{sq}}_{n,m} \asymp h^2,$$

with h understood as the intrinsic finest scale of the parent theorem. Thus every exact atomic critical class solved in Section 5 has a genuinely continuum, indeed absolutely continuous, counterpart with the same sharp law.

Proof. Each parent class listed above already satisfies an exact atomic law of order h and h^2 by the cited theorems. The geometric constructions in Section 5 place the support points at mutual separation comparable to the finest active scale, so for sufficiently small fixed κ the required disjoint positive-measure blobs exist. Applying Theorem 9.4 to each parent class yields the same upper and lower rates on the corresponding absolutely continuous lift. □

10 Partition lifts and full-support critical classes

The blob-lift theorem of Theorem 9.3 still localizes the variable mass to a sparse union of microregions around the atomic anchors. At the critical scale there is a more global continuum mechanism. If the anchor set genuinely tiles the cube at mesh h, then one may forget the exact point locations and retain only the coarse cell masses on a full partition of $[0,1]^d$. The next results make this precise and turn the one-scale paired core into a full-support continuum model.

Lemma 10.1 (Critical paired support with covering partition)**.** *There exist constants $0 < c_d \le C_d < \infty$ such that for every $h \in (0, 1/C_d]$ one can find*

$$M_h \asymp_d h^{-d}$$

pairs of points

$$(a_i, b_i)_{1\le i\le M_h} \subset [0,1]^d$$

with the following properties:

(i) *$\|a_i - b_i\|_2 = h$ for every i;*
(ii) *for every $i \ne j$ and every $u \in \{a_i, b_i\}$, $v \in \{a_j, b_j\}$ one has*

$$\|u - v\|_2 \ge 7h;$$

(iii) *every $x \in [0,1]^d$ satisfies*

$$\mathrm{dist}\big(x, \{a_1, b_1, \dots, a_{M_h}, b_{M_h}\}\big) \le C_d h.$$

Consequently, if $\Pi^{\mathrm{pair}}_h = \{C_u : u \in \mathcal{S}_h\}$ denotes the Voronoi partition of $[0,1]^d$ induced by the anchor set

$$\mathcal{S}_h := \{a_1, b_1, \dots, a_{M_h}, b_{M_h}\},$$

then every cell C_u satisfies

$$c_d h^d \le |C_u| \le C_d h^d, \qquad \mathrm{diam}(C_u) \le C_d h.$$

Proof. Choose an integer

$$m_h := \left\lfloor \frac{1}{8h} \right\rfloor \asymp h^{-1},$$

and for each $k = (k_1, \dots, k_d) \in \{0, \dots, m_h - 1\}^d$ define the center

$$c_k := \big(8hk_1 + 4h, \dots, 8hk_d + 4h\big).$$

Set

$$a_k := c_k - \frac{h}{2}e_1, \qquad b_k := c_k + \frac{h}{2}e_1.$$

Writing $M_h := m_h^d$, we have $M_h \asymp_d h^{-d}$. Property (i) is immediate.

If $k \neq \ell$, then c_k and c_ℓ differ by at least $8h$ in some coordinate. Hence every two anchors belonging to distinct indices remain at Euclidean distance at least $7h$, which proves (ii).

For (iii), let $x \in [0,1]^d$. Choose k so that each coordinate of c_k is one of the grid centers nearest to the corresponding coordinate of x. Then $|x_r - c_{k,r}| \le 8h$ for every r, hence

$$\|x - c_k\|_2 \le 8\sqrt{d}\,h.$$

Therefore

$$\operatorname{dist}(x, \mathcal{S}_h) \le \|x - c_k\|_2 + \frac{h}{2} \le \Big(8\sqrt{d} + \frac{1}{2}\Big)h,$$

which proves (iii) after enlarging C_d.

Now let $u \in \mathcal{S}_h$ and let C_u be its Voronoi cell. Because every point of C_u is no farther from u than from the nearest anchor, (iii) yields

$$C_u \subseteq B(u, C_d h),$$

hence $\operatorname{diam}(C_u) \le 2C_d h$ and

$$|C_u| \le |B(u, 2C_d h)| \lesssim_d h^d.$$

On the other hand, (ii) implies that the open ball $B(u, 7h/2)$ contains no other anchor. Therefore C_u contains $B(u, 7h/2) \cap [0,1]^d$, whose volume is bounded below by $c_d h^d$ after decreasing c_d if necessary. This proves the cell-volume lower bound and completes the proof. □

Corollary 10.2 (Exact critical law for the one-scale paired core)**.** *Assume $d > 2p$ and choose the paired support from Theorem 10.1. If*

$$h^{-d} \asymp N \log N,$$

then

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\text{pair}}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \asymp_{d,p} h,$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\text{pair}}} \mathbb{E}\big(\hat{W} - W_p(P,Q)\big)^2 \asymp_{d,p} h^2.$$

Proof. The upper bounds follow from Theorem 5.13, because

$$M_h \asymp_d h^{-d} \asymp N \log N.$$

For the lower bounds, Theorem 5.12 gives

$$W_p(P,Q) = h\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p} \qquad \big((P,Q) \in \mathcal{C}_h^{\text{pair}}\big)$$

on a known alphabet of size

$$S_h := 2M_h \asymp_d h^{-d} \asymp N \log N.$$

By [5, Theorem 3], the minimax squared error for estimating $\|P-Q\|_1$ on that alphabet is bounded below by a positive constant depending only on d. Since the parameter range is contained in $[0,2]$, the absolute minimax risk for estimating $\|P-Q\|_1$ is also bounded below by a positive constant.

If $\hat{W}$ is any estimator of $W_p(P,Q)$ on $\mathcal{C}_h^{\text{pair}}$, define

$$\widehat{L} := 2\Big(\frac{\hat{W} \wedge (2^{1/p}h)}{h}\Big)^p.$$

The map $w \mapsto 2(w/h)^p$ is $C_p h^{-1}$-Lipschitz on $[0, 2^{1/p}h]$, so on the paired class

$$\mathbb{E}\Big|\widehat{L} - \|P-Q\|_1\Big| \le C_p h^{-1}\mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big|.$$

Therefore the constant lower bound for estimating $\|P-Q\|_1$ forces

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}_h^{\text{pair}}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \gtrsim_{d,p} h.$$

The squared lower bound follows from Jensen's inequality. □

Definition 10.3 (Partition lift)**.** Let $S = \{z_1, \dots, z_M\} \subset [0,1]^d$, let

$$\Pi = \{C_1, \dots, C_M\}$$

be a measurable partition of $[0,1]^d$ into positive-measure cells with $z_i \in C_i$ for every i, and write

$$P^\Pi := \sum_{i=1}^M P(C_i)\,\delta_{z_i} \qquad (P \in \mathcal{P}([0,1]^d)).$$

For an atomic class $\mathfrak{C} \subset \mathcal{P}(S)^2$, define its partition lift by

$$\mathcal{L}_\Pi^{\text{part}}(\mathfrak{C}) := \Big\{(P,Q) \in \mathcal{P}([0,1]^d)^2 :\ (P^\Pi, Q^\Pi) \in \mathfrak{C}\Big\}.$$

If $|C_i| > 0$, we write U_{C_i} for the normalized Lebesgue measure on C_i.

Theorem 10.4 (Exact partition-lift equivalence)**.** *Let $\mathfrak{C} \subset \mathcal{P}(S)^2$ be an atomic class as in Theorem 10.3, and assume that*

$$\operatorname{diam}(C_i) \le \kappa h \qquad (1 \le i \le M)$$

for some $\kappa \ge 1$. Then

$$M_{n,m}^{\text{abs}}(\mathfrak{C}) - 2\kappa h \le M_{n,m}^{\text{abs}}\big(\mathcal{L}_\Pi^{\text{part}}(\mathfrak{C})\big) \le M_{n,m}^{\text{abs}}(\mathfrak{C}) + 2\kappa h,$$

and
$$\frac{1}{2}\,M^{\mathrm{sq}}_{n,m}(\mathfrak{C})-4\kappa^2h^2\le M^{\mathrm{sq}}_{n,m}\Big(\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})\Big)\le 2\,M^{\mathrm{sq}}_{n,m}(\mathfrak{C})+8\kappa^2h^2.$$

Proof. Let $(P,Q)\in\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})$. Coupling each point to the anchor of its cell yields
$$W_p(P,P^{\Pi})\le\kappa h,\qquad W_p(Q,Q^{\Pi})\le\kappa h.$$
Hence
$$\Big|W_p(P,Q)-W_p(P^{\Pi},Q^{\Pi})\Big|\le 2\kappa h.$$
Now fix $\varepsilon>0$ and let $\hat{W}_{\mathrm{at}}$ be an estimator on the atomic class $\mathfrak{C}$ satisfying
$$\sup_{(\mu,\nu)\in\mathfrak{C}}\mathbb{E}\Big|\hat{W}_{\mathrm{at}}-W_p(\mu,\nu)\Big|\le M^{\mathrm{abs}}_{n,m}(\mathfrak{C})+\varepsilon.$$
Given raw samples from $(P,Q)\in\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})$, map each observation to its cell index. The resulting label samples are i.i.d. from $(P^{\Pi},Q^{\Pi})\in\mathfrak{C}$. Running $\hat{W}_{\mathrm{at}}$ on these labels therefore gives
$$\sup_{(P,Q)\in\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})}\mathbb{E}\Big|\hat{W}_{\mathrm{at}}-W_p(P,Q)\Big|\le M^{\mathrm{abs}}_{n,m}(\mathfrak{C})+\varepsilon+2\kappa h.$$
Letting $\varepsilon\downarrow 0$ proves the upper absolute bound.

For the lower absolute bound, fix any estimator $\hat{W}_{\mathrm{part}}$ on $\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})$. For $(\mu,\nu)\in\mathfrak{C}$, define the canonical partition lifts
$$\widetilde{\mu}:=\sum_{i=1}^{M}\mu(z_i)\,U_{C_i},\qquad \widetilde{\nu}:=\sum_{i=1}^{M}\nu(z_i)\,U_{C_i}.$$
Then $(\widetilde{\mu},\widetilde{\nu})\in\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})$ and
$$\Big|W_p(\widetilde{\mu},\widetilde{\nu})-W_p(\mu,\nu)\Big|\le 2\kappa h.$$
Given atomic samples from (μ,ν), we may simulate raw samples from $(\widetilde{\mu},\widetilde{\nu})$ by replacing every observation z_i with an independent draw from U_{C_i}. Applying $\hat{W}_{\mathrm{part}}$ to the simulated raw samples yields an atomic estimator $\hat{W}^{\star}_{\mathrm{at}}$ with
$$\mathbb{E}_{\mu,\nu}\Big|\hat{W}^{\star}_{\mathrm{at}}-W_p(\mu,\nu)\Big|\le\mathbb{E}_{\widetilde{\mu},\widetilde{\nu}}\Big|\hat{W}_{\mathrm{part}}-W_p(\widetilde{\mu},\widetilde{\nu})\Big|+2\kappa h.$$
Taking the supremum over $(\mu,\nu)\in\mathfrak{C}$ and then the infimum over $\hat{W}_{\mathrm{part}}$ gives
$$M^{\mathrm{abs}}_{n,m}(\mathfrak{C})\le M^{\mathrm{abs}}_{n,m}\Big(\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})\Big)+2\kappa h,$$
which is the desired lower absolute bound.

For the squared upper bound, use
$$(a+b)^2\le 2a^2+2b^2$$
with
$$a=\hat{W}_{\mathrm{at}}-W_p(P^{\Pi},Q^{\Pi}),\qquad b=W_p(P^{\Pi},Q^{\Pi})-W_p(P,Q),$$
and then optimize over atomic estimators. For the squared lower bound, the same argument with the canonical lifts gives
$$M^{\mathrm{sq}}_{n,m}(\mathfrak{C})\le 2\,M^{\mathrm{sq}}_{n,m}\Big(\mathcal{L}^{\mathrm{part}}_{\Pi}(\mathfrak{C})\Big)+8\kappa^2h^2,$$
which rearranges to the claimed lower bound. □

Proposition 10.5 (Background-stable paired core)**.** *Let $u\in\mathcal{P}(\mathcal{S}_h)$ and $\lambda\in[0,1)$. Define*
$$\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}:=\Big\{\Big(\lambda u+(1-\lambda)P,\ \lambda u+(1-\lambda)Q\Big):(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h}\Big\}.$$
Then
$$\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}\subseteq\mathcal{C}^{\mathrm{pair}}_{h},$$
and for every $(P,Q)\in\mathcal{C}^{\mathrm{pair}}_{h}$ one has
$$W_p\Big(\lambda u+(1-\lambda)P,\ \lambda u+(1-\lambda)Q\Big)=(1-\lambda)^{1/p}\,W_p(P,Q).$$
If moreover $h^{-d}\asymp N\log N$, then
$$\inf_{\hat{W}}\sup_{(P',Q')\in\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}}\mathbb{E}\Big|\hat{W}-W_p(P',Q')\Big|\asymp_{d,p,\lambda}h,$$
and
$$\inf_{\hat{W}}\sup_{(P',Q')\in\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}}\mathbb{E}\Big(\hat{W}-W_p(P',Q')\Big)^2\asymp_{d,p,\lambda}h^2.$$

Proof. Let
$$P':=\lambda u+(1-\lambda)P,\qquad Q':=\lambda u+(1-\lambda)Q.$$
Write (r_i,α_i,β_i) for the pair parameters of (P,Q). Then
$$P'(a_i)+P'(b_i)=Q'(a_i)+Q'(b_i)=\lambda\Big(u(a_i)+u(b_i)\Big)+(1-\lambda)r_i,$$
so $(P',Q')\in\mathcal{C}^{\mathrm{pair}}_{h}$. Moreover
$$P'(a_i)-Q'(a_i)=(1-\lambda)\Big(P(a_i)-Q(a_i)\Big),\qquad P'(b_i)-Q'(b_i)=(1-\lambda)\Big(P(b_i)-Q(b_i)\Big),$$
hence
$$\|P'-Q'\|_1=(1-\lambda)\|P-Q\|_1.$$
By Theorem 5.12,
$$W_p(P',Q')^p=\frac{h^p}{2}\|P'-Q'\|_1=(1-\lambda)\frac{h^p}{2}\|P-Q\|_1=(1-\lambda)W_p(P,Q)^p,$$

which proves the exact scaling identity.

For the upper bounds, simply note that

$$\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u} \subseteq \mathcal{C}^{\mathrm{pair}}_{h},$$

so Theorem 10.2 implies

$$M^{\mathrm{abs}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}\right) \lesssim_{d,p} h, \qquad M^{\mathrm{sq}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}\right) \lesssim_{d,p} h^2.$$

For the lower bounds, let $\hat{W}$ be any estimator on $\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}$. Given samples from $(P,Q) \in \mathcal{C}^{\mathrm{pair}}_{h}$, independently replace each observation by an independent sample from u with probability λ. The resulting samples have the same law as samples from $(P',Q') \in \mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}$. Therefore

$$\hat{W}^\star := (1-\lambda)^{-1/p}\hat{W}$$

is an estimator on $\mathcal{C}^{\mathrm{pair}}_{h}$ with

$$\mathbb{E}_{P,Q}\left|\hat{W}^\star - W_p(P,Q)\right| = (1-\lambda)^{-1/p}\mathbb{E}_{P',Q'}\left|\hat{W} - W_p(P',Q')\right|.$$

Taking the supremum over $\mathcal{C}^{\mathrm{pair}}_{h}$ and then the infimum over $\hat{W}$ yields

$$M^{\mathrm{abs}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}\right) \geq (1-\lambda)^{1/p} M^{\mathrm{abs}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h}\right),$$

and similarly

$$M^{\mathrm{sq}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u}\right) \geq (1-\lambda)^{2/p} M^{\mathrm{sq}}_{n,m}\left(\mathcal{C}^{\mathrm{pair}}_{h}\right).$$

Now apply Theorem 10.2. □

Corollary 10.6 (Background-stable annular phase law on the paired class). *Fix $p \geq 1$, $\alpha \in (0,1)$, $C_0 \geq 1$, $\lambda \in [0,1)$, and numbers $0 < \tau_- < \tau_+ < 1$. Assume that*

$$N^\alpha \leq M_h \leq C_0 N \log N.$$

Let $u \in \mathcal{P}(\mathcal{S}_h)$, and define

$$\mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+) := \Big\{\big(\lambda u + (1-\lambda)P,\ \lambda u + (1-\lambda)Q\big) : (P,Q) \in \mathcal{C}^{\mathrm{pair}}_{h},\ \tau_- \leq \mathrm{TV}(P,Q) \leq \tau_+\Big\}.$$

Then

$$\inf_{\hat{W}} \sup_{(P',Q') \in \mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+)} \mathbb{E}\left|\hat{W} - W_p(P',Q')\right| \asymp_{d,p,\alpha,C_0,\lambda,\tau_-,\tau_+} h\sqrt{\frac{M_h}{N\log N}},$$

and

$$\inf_{\hat{W}} \sup_{(P',Q') \in \mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+)} \mathbb{E}\big(\hat{W} - W_p(P',Q')\big)^2 \asymp_{d,p,\alpha,C_0,\lambda,\tau_-,\tau_+} h^2\frac{M_h}{N\log N}.$$

Proof. Set

$$\tau'_- := (1-\lambda)\tau_-, \qquad \tau'_+ := (1-\lambda)\tau_+.$$

If

$$P' := \lambda u + (1-\lambda)P, \qquad Q' := \lambda u + (1-\lambda)Q,$$

then Theorem 10.5 gives

$$W_p(P',Q') = (1-\lambda)^{1/p}\, W_p(P,Q)$$

and

$$\mathrm{TV}(P',Q') = (1-\lambda)\,\mathrm{TV}(P,Q).$$

Hence

$$\mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+) \subseteq \mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau'_-,\tau'_+).$$

Applying Theorem 5.14 to the annulus $\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau'_-,\tau'_+)$ therefore yields the desired upper bounds.

For the lower bounds, let $\hat{W}$ be any estimator on $\mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+)$. Given samples from $(P,Q) \in \mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)$, independently replace each observation by an independent sample from u with probability λ. Exactly as in the proof of Theorem 10.5, the resulting samples have the same law as samples from

$$(P',Q') := \big(\lambda u + (1-\lambda)P,\ \lambda u + (1-\lambda)Q\big) \in \mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+),$$

and

$$W_p(P',Q') = (1-\lambda)^{1/p}\, W_p(P,Q).$$

Therefore

$$\hat{W}^\star := (1-\lambda)^{-1/p}\hat{W}$$

is an estimator on $\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)$ satisfying

$$\mathbb{E}_{P,Q}\left|\hat{W}^\star - W_p(P,Q)\right| = (1-\lambda)^{-1/p}\mathbb{E}_{P',Q'}\left|\hat{W} - W_p(P',Q')\right|.$$

Taking the supremum over $\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)$ and then the infimum over $\hat{W}$ gives

$$M^{\mathrm{abs}}_{n,m}\Big(\mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+)\Big) \geq (1-\lambda)^{1/p} M^{\mathrm{abs}}_{n,m}\Big(\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)\Big),$$

and similarly

$$M^{\mathrm{sq}}_{n,m}\Big(\mathcal{C}^{\mathrm{pair,ann}}_{h,\lambda,u}(\tau_-,\tau_+)\Big) \geq (1-\lambda)^{2/p} M^{\mathrm{sq}}_{n,m}\Big(\mathcal{C}^{\mathrm{pair}}_{h,\mathrm{ann}}(\tau_-,\tau_+)\Big).$$

Now apply Theorem 5.14. □

Corollary 10.7 (Full-support paired continuum core). *Assume $d > 2p$, fix $\lambda \in (0,1)$, and let $\Pi^{\mathrm{pair}}_h = \{C_u : u \in \mathcal{S}_h\}$ be the partition from Theorem 10.1. Set*

$$u_h(u) := |C_u|, \qquad u \in \mathcal{S}_h,$$

so that $u_h \in \mathcal{P}(\mathcal{S}_h)$. Define the full partition lift

$$\widetilde{\mathcal{C}}^{\mathrm{pair,full}}_{h,\lambda} := \mathcal{L}^{\mathrm{part}}_{\Pi^{\mathrm{pair}}_h}\big(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u_h}\big).$$

If

$$h^{-d} \asymp N \log N,$$

then

$$\inf_{\hat W} \sup_{(P,Q)\in \widetilde{\mathcal{C}}^{\mathrm{pair,full}}_{h,\lambda}} \mathbb{E}\big|\hat W - W_p(P,Q)\big| \asymp_{d,p,\lambda} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in \widetilde{\mathcal{C}}^{\mathrm{pair,full}}_{h,\lambda}} \mathbb{E}\big(\hat W - W_p(P,Q)\big)^2 \asymp_{d,p,\lambda} h^2.$$

Moreover the canonical cell-uniform class

$$\widetilde{\mathcal{U}}^{\mathrm{pair,full}}_{h,\lambda} := \Big\{ \big(\sum_{u\in\mathcal{S}_h} \big(\lambda u_h(u) + (1-\lambda)P(u)\big) U_{C_u},\ \sum_{u\in\mathcal{S}_h} \big(\lambda u_h(u) + (1-\lambda)Q(u)\big) U_{C_u} \big) : (P,Q)\in \mathcal{C}^{\mathrm{pair}}_h \Big\}$$

consists of piecewise-constant absolutely continuous measures with full support on $[0,1]^d$*, satisfies the pointwise lower bound*

$$f, g \geq \lambda,$$

and obeys the same exact minimax law.

Proof. Because $\mathrm{diam}(C_u) \leq C_d h$ by Theorem 10.1, Theorem 10.4 yields

$$M^{\mathrm{abs}}_{n,m}\big(\widetilde{\mathcal{C}}^{\mathrm{pair,full}}_{h,\lambda}\big) = M^{\mathrm{abs}}_{n,m}\big(\mathcal{C}^{\mathrm{pair}}_{h,\lambda,u_h}\big) + O_d(h),$$

and similarly for the squared risk with error $O_d(h^2)$. Now apply Theorem 10.5.

For the canonical class, the same upper bound is obtained by running the coarse-cell estimator on the cell labels. For the lower bound, the proof of Theorem 10.4 uses exactly these canonical lifts, so the same comparison applies. Finally, on each cell C_u the density of a canonical lift equals

$$\lambda + (1-\lambda)\frac{P(u)}{|C_u|} \qquad \text{or} \qquad \lambda + (1-\lambda)\frac{Q(u)}{|C_u|},$$

hence is at least λ. Because Π^{pair}_h partitions the whole cube, every such density has full support. □

11 Contiguous full-support critical shells

The exact continuum classes obtained so far arise from partition or blob lifts of atomic configurations that still retain some explicit geometric separation. We now show that the critical $(N\log N)^{-1/d}$ law already survives on a genuinely contiguous full-support continuum model whose fine cells fill the whole cube and touch along faces everywhere. The first result gives an exact critical class for every $p \geq 1$; the second broadens this to a much larger touching-cell shell class in the special case W_1.

11.1 A contiguous split-shell core for every $p \geq 1$

Fix a dyadic level $J \geq 1$ and write

$$h := 2^{-J}.$$

For each parent cube

$$R = \prod_{r=1}^d [a_r, a_r + 2h] \in \mathcal{D}_{J-1},$$

define its left and right slabs

$$R^- := R \cap \{x_1 < a_1 + h\}, \qquad R^+ := R \cap \{x_1 \geq a_1 + h\}.$$

Both slabs have volume

$$|R^\pm| = 2^{d-1}h^d.$$

Definition 11.1 (Contiguous split-shell class)**.** Fix $\lambda \in (0,1)$. Let $\mathcal{C}^{\mathrm{split}}_{J,\lambda}$ be the class of all pairs (P,Q) of the form

$$P = \sum_{R\in\mathcal{D}_{J-1}} \Big(2^d h^d \,\frac{1+\alpha_R}{2}\, U_{R^-} + 2^d h^d \,\frac{1-\alpha_R}{2}\, U_{R^+} \Big),$$

$$Q = \sum_{R\in\mathcal{D}_{J-1}} \Big(2^d h^d \,\frac{1+\beta_R}{2}\, U_{R^-} + 2^d h^d \,\frac{1-\beta_R}{2}\, U_{R^+} \Big),$$

where

$$\alpha_R, \beta_R \in [-1+\lambda,\ 1-\lambda] \qquad (R\in\mathcal{D}_{J-1}).$$

Equivalently, on every parent cube R both measures are uniform on the two adjacent half-slabs $R^\pm$, have the same total parent mass

$$P(R) = Q(R) = 2^d h^d,$$

and the slab densities satisfy

$$\lambda \leq f_P, f_Q \leq 2-\lambda$$

pointwise on $[0,1]^d$.

Proposition 11.2 (Exact transport identity on the contiguous split-shell core)**.** *Let* $(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}$ *and let* $p \geq 1$*. Then*

$$W_p(P,Q)^p = h^p\ \mathrm{TV}(P,Q) = \frac{h^p}{2}\,\|P-Q\|_1.$$

Equivalently,

$$W_p(P,Q) = h\Big(\frac{\|P-Q\|_1}{2}\Big)^{1/p}.$$

Proof. Write

$$m_R := 2^d h^d, \qquad \Delta_R := \alpha_R - \beta_R \qquad (R \in \mathcal{D}_{J-1}).$$

Then the discrepancy on the two slabs inside R is

$$P(R^-) - Q(R^-) = \frac{m_R \Delta_R}{2}, \qquad P(R^+) - Q(R^+) = -\frac{m_R \Delta_R}{2}.$$

For the upper bound, if $\Delta_R > 0$ move the mass $m_R\Delta_R/2$ from R^- to R^+ by the translation $x \mapsto x + he_1$; if $\Delta_R < 0$, use the inverse translation from R^+ to R^-. This preserves the transverse coordinates and moves every transported point by Euclidean distance exactly h. Doing this independently on every parent cube produces a coupling with total cost

$$\sum_{R\in\mathcal{D}_{J-1}} h^p \, \frac{m_R|\Delta_R|}{2} = h^p \ \mathrm{TV}(P, Q).$$

Hence

$$W_p(P, Q)^p \le h^p \ \mathrm{TV}(P, Q).$$

For the reverse inequality, let $\pi_1(x) = x_1$ be the first-coordinate projection. Since π_1 is 1-Lipschitz,

$$W_p(P, Q) \ge W_p\big((\pi_1)_\# P, (\pi_1)_\# Q\big).$$

Now $(\pi_1)_\# P$ and $(\pi_1)_\# Q$ are one-dimensional piecewise-uniform measures on the dyadic intervals

$$I_R^- := [a_1, a_1 + h), \qquad I_R^+ := [a_1 + h, a_1 + 2h)$$

with the same paired masses $m_R(1 \pm \alpha_R)/2$ and $m_R(1 \pm \beta_R)/2$. Because the total mass on every parent interval $I_R^- \cup I_R^+$ agrees for the two projected measures, the monotone one-dimensional optimal transport acts independently on each parent interval, moving exactly $m_R|\Delta_R|/2$ units of mass across a distance h. Therefore

$$W_p\big((\pi_1)_\# P, (\pi_1)_\# Q\big)^p = \sum_{R\in\mathcal{D}_{J-1}} h^p \, \frac{m_R|\Delta_R|}{2} = h^p \ \mathrm{TV}(P, Q).$$

Combining this with the projection lower bound gives

$$W_p(P, Q)^p \ge h^p \ \mathrm{TV}(P, Q).$$

Together with the upper bound, this proves equality. □

Corollary 11.3 (Exact critical law on the contiguous split-shell core)**.** *Assume $d > 2p$ and fix $\lambda \in (0, 1)$. Let $h = 2^{-J}$ satisfy*

$$h^{-d} \asymp N \log N.$$

Then

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\big|\hat W - W_p(P, Q)\big| \asymp_{d,p,\lambda} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\big(\hat W - W_p(P, Q)\big)^2 \asymp_{d,p,\lambda} h^2.$$

Proof. Map each observation to its slab label $(R, \pm)$. This produces samples from a known alphabet of size

$$S_J := 2\,|\mathcal{D}_{J-1}| = 2^{1+(J-1)d} \asymp_d h^{-d} \asymp N \log N.$$

By Theorem 11.2,

$$W_p(P, Q) = h\Big(\frac{\|P - Q\|_1}{2}\Big)^{1/p} \qquad \big((P, Q) \in \mathcal{C}^{\mathrm{split}}_{J,\lambda}\big),$$

where $\|P - Q\|_1$ is computed on that known finite alphabet. The sharp large-alphabet L_1 theory of [5, Theorem 6 and Theorem 3], together with the same clipping and transformation argument used in the proof of Theorems 5.13 and 10.2, therefore gives the upper and lower bounds

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\big|\hat W - W_p(P, Q)\big| \asymp_{d,p,\lambda} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\big(\hat W - W_p(P, Q)\big)^2 \asymp_{d,p,\lambda} h^2.$$

□

Corollary 11.4 (A contiguous full-support lower-bound core for the unrestricted problem)**.** *Assume $d > 2p$ and fix $\lambda \in (0, 1)$. Let $h \asymp (N \log N)^{-1/d}$. Then even on the subclass of absolutely continuous piecewise-constant measures on $[0, 1]^d$ with full support and pointwise density bounds*

$$\lambda \le f, g \le 2 - \lambda,$$

the minimax risks for estimating $W_p(P, Q)$ are bounded below by

$$\inf_{\hat W} \sup \mathbb{E}\big|\hat W - W_p(P, Q)\big| \gtrsim_{d,p,\lambda} h,$$

and

$$\inf_{\hat W} \sup \mathbb{E}\big(\hat W - W_p(P, Q)\big)^2 \gtrsim_{d,p,\lambda} h^2.$$

Proof. This is immediate from Theorem 11.3, since $\mathcal{C}^{\mathrm{split}}_{J,\lambda}$ is such a full-support bounded-density subclass. □

11.2 The full one-step martingale shell for W_1

The split-shell class keeps only one left/right mode inside each parent cube. We now allow the *entire* child configuration inside every parent cube, while retaining the same parent masses and the same global full-support positivity.

For each fine dyadic cube $Q \in \mathcal{D}_J$, let U_Q denote the normalized Lebesgue measure on Q.

Definition 11.5 (One-step dyadic martingale shell)**.** Fix $\lambda \in (0,1)$ and write $h = 2^{-J}$. Let $\mathcal{C}^{\mathrm{shell}}_{J,\lambda}$ consist of all pairs

$$P = \sum_{Q \in \mathcal{D}_J} p_Q \, U_Q, \qquad Q = \sum_{Q \in \mathcal{D}_J} q_Q \, U_Q,$$

such that

$$\lambda h^d \le p_Q, q_Q \le (2-\lambda) h^d \qquad (Q \in \mathcal{D}_J),$$

and

$$\sum_{Q \subset R} p_Q = \sum_{Q \subset R} q_Q = 2^d h^d \qquad (R \in \mathcal{D}_{J-1}).$$

Equivalently, both measures are positive piecewise-constant densities on the whole cube, every parent cube carries the same total mass for the two measures, and the average density on each parent cube is exactly one.

Lemma 11.6 (Haar-shell localization and coefficient comparability)**.** *Let $(P,Q) \in \mathcal{C}^{\mathrm{shell}}_{J,\lambda}$, write*

$$P = f\,dx, \qquad Q = g\,dx,$$

and let

$$u_Q := p_Q - q_Q \qquad (Q \in \mathcal{D}_J).$$

Then $f-g$ belongs to the single tensor-Haar shell at level $J-1$. More precisely, for every parent cube $R \in \mathcal{D}_{J-1}$ and every nontrivial Walsh sign

$$\varepsilon \in \{0,1\}^d \setminus \{0\},$$

let $\psi_{R,\varepsilon}$ be the standard L^2-normalized tensor-Haar wavelet on R. Then

$$f - g = \sum_{R \in \mathcal{D}_{J-1}} \sum_{\varepsilon \neq 0} \beta_{R,\varepsilon} \, \psi_{R,\varepsilon},$$

and there exist constants $0 < c_d \le C_d < \infty$ such that

$$c_d \, (2h)^{-d/2} \sum_{R \in \mathcal{D}_{J-1}} \sum_{Q \subset R} |u_Q| \le \sum_{R \in \mathcal{D}_{J-1}} \sum_{\varepsilon \neq 0} |\beta_{R,\varepsilon}| \le C_d \, (2h)^{-d/2} \sum_{R \in \mathcal{D}_{J-1}} \sum_{Q \subset R} |u_Q|.$$

Equivalently,

$$\sum_{R,\varepsilon} |\beta_{R,\varepsilon}| \asymp_d (2h)^{-d/2} \, \|P - Q\|_1.$$

Proof. Because P and Q assign the same total mass to every parent cube $R \in \mathcal{D}_{J-1}$, the piecewise-constant function $f-g$ has zero average on every such parent cube. Hence it lies in the direct sum of the nonconstant tensor-Haar modes on those parent cubes, which gives the stated expansion.

Fix one parent cube R and label its children by

$$Q_\sigma(R), \qquad \sigma \in \{0,1\}^d.$$

Write

$$u_{R,\sigma} := u_{Q_\sigma(R)}, \qquad u_R := (u_{R,\sigma})_{\sigma \in \{0,1\}^d} \in \mathbb{R}^{2^d}.$$

Since the parent masses agree,

$$\sum_{\sigma \in \{0,1\}^d} u_{R,\sigma} = 0.$$

For the standard tensor-Haar wavelets on R, one has

$$\beta_{R,\varepsilon} = |R|^{-1/2} \sum_{\sigma \in \{0,1\}^d} (-1)^{\varepsilon \cdot \sigma} \, u_{R,\sigma} \qquad (\varepsilon \neq 0).$$

Thus the vector $(\beta_{R,\varepsilon})_{\varepsilon \neq 0}$ is obtained from u_R by applying the Walsh–Hadamard transform and restricting to the nonconstant coordinates, then multiplying by $|R|^{-1/2} = (2h)^{-d/2}$. On the zero-sum subspace

$$E_d := \Big\{ v \in \mathbb{R}^{2^d} : \sum_\sigma v_\sigma = 0 \Big\},$$

this linear map is invertible. Since E_d has fixed finite dimension depending only on d, all norms on E_d are equivalent; in particular there exist constants $c_d, C_d > 0$ such that

$$c_d \, \|u_R\|_{\ell_1} \le (2h)^{d/2} \sum_{\varepsilon \neq 0} |\beta_{R,\varepsilon}| \le C_d \, \|u_R\|_{\ell_1}.$$

Summing this over all parent cubes R proves the claim. □

Proposition 11.7 (The one-step shell is equivalent to fine-grid L_1 at scale h)**.** *Fix $\lambda \in (0,1)$. There exist constants $0 < c_{d,\lambda} \le C_{d,\lambda} < \infty$ such that for every $(P,Q) \in \mathcal{C}^{\mathrm{shell}}_{J,\lambda}$,*

$$c_{d,\lambda} \, h \, \|P - Q\|_1 \le W_1(P,Q) \le C_{d,\lambda} \, h \, \|P - Q\|_1.$$

Equivalently,

$$W_1(P,Q) \asymp_{d,\lambda} h \, \|P - Q\|_1 \qquad \big((P,Q) \in \mathcal{C}^{\mathrm{shell}}_{J,\lambda}\big).$$

Proof. Write $P = f\,dx$ and $Q = g\,dx$. By construction,

$$\lambda \le f, g \le 2 - \lambda.$$

Since $f-g$ occupies only the single tensor-Haar shell at level $J-1$, [9, Theorem 3] gives

$$c'_{d,\lambda}\, 2^{-(J-1)} 2^{-(J-1)d/2} \sum_{R,\varepsilon} |\beta_{R,\varepsilon}| \le W_1(P,Q) \le C'_{d,\lambda}\, 2^{-(J-1)} 2^{-(J-1)d/2} \sum_{R,\varepsilon} |\beta_{R,\varepsilon}|.$$

Now apply Theorem 11.6. Because

$$2^{-(J-1)} 2^{-(J-1)d/2} (2h)^{-d/2} = 2h,$$

the previous display becomes

$$c_{d,\lambda}\, h\, \|P-Q\|_1 \le W_1(P,Q) \le C_{d,\lambda}\, h\, \|P-Q\|_1$$

after adjusting constants. □

Theorem 11.8 (Exact critical law on the contiguous one-step W_1 shell)**.** *Assume $d>2$ and fix $\lambda \in (0,1)$. Let $h = 2^{-J}$ satisfy*

$$h^{-d} \asymp N \log N.$$

Then the minimax risks on $\mathcal{C}^{\mathrm{shell}}_{J,\lambda}$ obey

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big|\hat W - W_1(P,Q)\big| \asymp_{d,\lambda} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big(\hat W - W_1(P,Q)\big)^2 \asymp_{d,\lambda} h^2.$$

Proof. For the upper bound, map every observation to its fine dyadic cell label in $\mathcal{D}_J$. This yields samples from a known alphabet of size

$$M_J := |\mathcal{D}_J| = 2^{Jd} \asymp h^{-d} \asymp N \log N.$$

By [5, Theorem 6] there exists an estimator $\widehat{L}_J$ of $\|P-Q\|_1$ on that alphabet with

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big(\widehat{L}_J - \|P-Q\|_1\big)^2 \le C_{d,\lambda},$$

hence

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big|\widehat{L}_J - \|P-Q\|_1\big| \le C_{d,\lambda}.$$

Set

$$\widehat{W}^{\mathrm{shell}}_J := h\, \widehat{L}_J.$$

Then Theorem 11.7 yields

$$\big|\widehat{W}^{\mathrm{shell}}_J - W_1(P,Q)\big| \le h\, \big|\widehat{L}_J - \|P-Q\|_1\big| + C_{d,\lambda}\, h,$$

and therefore

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big|\widehat{W}^{\mathrm{shell}}_J - W_1(P,Q)\big| \le C_{d,\lambda}\, h.$$

Squaring the same pointwise bound gives

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big(\widehat{W}^{\mathrm{shell}}_J - W_1(P,Q)\big)^2 \le C_{d,\lambda}\, h^2.$$

For the lower bounds, observe that the split-shell subclass

$$\mathcal{C}^{\mathrm{split}}_{J,\lambda} \subseteq \mathcal{C}^{\mathrm{shell}}_{J,\lambda}.$$

Therefore Theorem 11.3 with $p=1$ implies

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big|\hat W - W_1(P,Q)\big| \ge \inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\big|\hat W - W_1(P,Q)\big| \gtrsim_{d,\lambda} h,$$

and similarly for the squared risk. □

12 Overlapping local smoothing and smooth critical lower cores

The partition lifts of Section 10 preserve exact critical laws because the underlying cell label remains directly observable from the data. The next theorem shows that lower bounds are much more robust than upper bounds: any *known* Markov smoothing whose transport radius is $O(h)$ can change W_p only by $O(h)$, so every critical lower-bound core survives even under heavily overlapping local regularization.

For a statistical class $\mathfrak{C} \subset \mathcal{P}([0,1]^d)^2$, write

$$M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) := \inf_{\hat W} \sup_{(P,Q)\in\mathfrak{C}} \mathbb{E}\big|\hat W - W_p(P,Q)\big|,$$

$$M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) := \inf_{\hat W} \sup_{(P,Q)\in\mathfrak{C}} \mathbb{E}\big(\hat W - W_p(P,Q)\big)^2.$$

Theorem 12.1 (Critical lower bounds survive overlapping local smoothing)**.** *Fix $p \ge 1$, $h>0$, and $\kappa>0$. Let K be a Markov kernel on $[0,1]^d$ such that*

$$\sup_{x\in[0,1]^d} \left(\int \|y-x\|_2^p\, K(x,dy) \right)^{1/p} \le \kappa h.$$

For a class $\mathfrak{C} \subset \mathcal{P}([0,1]^d)^2$, define its smoothing image

$$K\mathfrak{C} := \Big\{ (K_\# P, K_\# Q) : \ (P,Q) \in \mathfrak{C} \Big\}.$$

Then every $(P,Q)\in\mathcal{P}([0,1]^d)^2$ *satisfies*

$$\left|W_p(K_\# P, K_\# Q) - W_p(P,Q)\right| \le 2\kappa h.$$

Consequently,

$$M^{\mathrm{abs}}_{n,m}(K\mathfrak{C}) \ge M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) - 2\kappa h,$$

and

$$M^{\mathrm{sq}}_{n,m}(K\mathfrak{C}) \ge \frac{1}{2}\, M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) - 4\kappa^2 h^2.$$

Proof. Fix $P\in\mathcal{P}([0,1]^d)$. The measure

$$\pi_P(dx,dy) := P(dx)\,K(x,dy)$$

is a coupling between P and $K_\# P$. Therefore

$$W_p(P,K_\# P)^p \le \int \|y-x\|_2^p\,\pi_P(dx,dy) = \int\left(\int \|y-x\|_2^p\,K(x,dy)\right)P(dx) \le \kappa^p h^p,$$

hence

$$W_p(P,K_\# P)\le \kappa h.$$

The same argument gives

$$W_p(Q,K_\# Q)\le \kappa h.$$

By the triangle inequality,

$$W_p(K_\# P, K_\# Q) \le W_p(K_\# P, P) + W_p(P,Q) + W_p(Q, K_\# Q) \le W_p(P,Q) + 2\kappa h,$$

and symmetrically

$$W_p(P,Q)\le W_p(K_\# P, K_\# Q) + 2\kappa h.$$

This proves the pointwise perturbation bound.

Now let $\hat W$ be any estimator on $K\mathfrak{C}$. Given raw samples

$$X_1,\dots,X_n\sim P, \qquad Y_1,\dots,Y_m\sim Q,$$

generate conditionally independent smoothed samples

$$\widetilde X_i \sim K(X_i,\cdot), \qquad \widetilde Y_j\sim K(Y_j,\cdot).$$

Then

$$\widetilde X_1,\dots,\widetilde X_n \overset{\text{i.i.d.}}{\sim} K_\# P, \qquad \widetilde Y_1,\dots,\widetilde Y_m \overset{\text{i.i.d.}}{\sim} K_\# Q.$$

Hence

$$\hat W^\sharp := \hat W(\widetilde X_1,\dots,\widetilde X_n,\widetilde Y_1,\dots,\widetilde Y_m)$$

is an estimator on $\mathfrak{C}$. For every $(P,Q)\in\mathfrak{C}$,

$$\left|\hat W^\sharp - W_p(P,Q)\right| \le \left|\hat W^\sharp - W_p(K_\# P, K_\# Q)\right| + 2\kappa h.$$

Taking expectations gives

$$\mathbb{E}\left|\hat W^\sharp - W_p(P,Q)\right| \le \mathbb{E}\left|\hat W - W_p(K_\# P, K_\# Q)\right| + 2\kappa h.$$

After taking the supremum over $(P,Q)\in\mathfrak{C}$ and then the infimum over $\hat W$,

$$M^{\mathrm{abs}}_{n,m}(\mathfrak{C}) \le M^{\mathrm{abs}}_{n,m}(K\mathfrak{C}) + 2\kappa h,$$

which rearranges to the stated lower bound.

For squared loss,

$$\left(\hat W^\sharp - W_p(P,Q)\right)^2 \le 2\left(\hat W^\sharp - W_p(K_\# P, K_\# Q)\right)^2 + 8\kappa^2 h^2.$$

Taking expectations, suprema, and infima as above yields

$$M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) \le 2\, M^{\mathrm{sq}}_{n,m}(K\mathfrak{C}) + 8\kappa^2 h^2,$$

hence

$$M^{\mathrm{sq}}_{n,m}(K\mathfrak{C}) \ge \frac{1}{2}\, M^{\mathrm{sq}}_{n,m}(\mathfrak{C}) - 4\kappa^2 h^2.$$ □

We now turn this abstract principle into an explicit smooth lower-bound core. Fix once and for all an even bump

$$\psi\in C_c^\infty((-1,1)), \qquad \psi\ge 0, \qquad \int_{\mathbb{R}}\psi(u)\,du = 1.$$

For $\varepsilon>0$ and $x,y\in[0,1]$, define the one-dimensional reflected kernel

$$k^{(1)}_\varepsilon(x,y) := \sum_{m\in\mathbb{Z}}\frac{1}{\varepsilon}\,\psi\!\left(\frac{y+2m-x}{\varepsilon}\right) + \sum_{m\in\mathbb{Z}}\frac{1}{\varepsilon}\,\psi\!\left(\frac{2m-y-x}{\varepsilon}\right),$$

and in dimension d set

$$k_\varepsilon(x,y) := \prod_{j=1}^d k^{(1)}_\varepsilon(x_j,y_j), \qquad x,y\in[0,1]^d.$$

Let K_ε denote the associated Markov operator,

$$(K_\varepsilon f)(y) := \int_{[0,1]^d} k_\varepsilon(x,y)\,f(x)\,dx.$$

Lemma 12.2 (Reflected smooth local kernel)**.** *For every* $\varepsilon>0$*, the kernel* K_ε *satisfies:*

(i) k_ε *is symmetric and doubly stochastic:*

$$\int_{[0,1]^d} k_\varepsilon(x,y)\,dy = 1 \qquad \text{and} \qquad \int_{[0,1]^d} k_\varepsilon(x,y)\,dx = 1 \qquad (x,y\in[0,1]^d);$$

(ii) *for every* $x \in [0,1]^d$,

$$\operatorname{supp} K_\varepsilon(x,\cdot) \subseteq B_\infty(x,\varepsilon),$$

hence

$$\sup_{x\in[0,1]^d} \left(\int \|y-x\|_2^p \, K_\varepsilon(x,dy) \right)^{1/p} \le \sqrt{d}\,\varepsilon \qquad (p\ge 1);$$

(iii) *if* $f \in L^\infty([0,1]^d)$, *then* $K_\varepsilon f \in C^\infty([0,1]^d)$ *and for every multi-index* α,

$$\|\partial^\alpha K_\varepsilon f\|_\infty \le C_{\alpha,d,\psi}\,\varepsilon^{-|\alpha|}\,\|f\|_\infty;$$

(iv) *if* $m \le f \le M$ *almost everywhere, then*

$$m \le K_\varepsilon f \le M \qquad \text{everywhere on } [0,1]^d.$$

Proof. Let $r:\mathbb{R}\to[0,1]$ be the standard reflection map

$$r(u) := \begin{cases} u-2k, & u\in[2k,2k+1], \\ 2k+2-u, & u\in[2k+1,2k+2], \end{cases} \qquad k\in\mathbb{Z},$$

and write

$$R(u_1,\dots,u_d) := (r(u_1),\dots,r(u_d)).$$

If $Z=(Z_1,\dots,Z_d)$ has product density

$$\Phi(z) := \prod_{j=1}^d \psi(z_j),$$

then $K_\varepsilon(x,\cdot)$ is exactly the law of

$$R(x+\varepsilon Z).$$

Indeed, in one dimension the preimages of $y\in[0,1]$ under r are precisely

$$\{\,y+2m,\ 2m-y:\ m\in\mathbb{Z}\,\},$$

which gives the stated formula for $k_\varepsilon^{(1)}$; taking products yields the d-dimensional density formula.

Because $\operatorname{supp}\Phi \subset [-1,1]^d$ and R is coordinatewise 1-Lipschitz with $R(x)=x$ for $x\in[0,1]^d$, one has

$$\|R(x+\varepsilon Z)-x\|_\infty \le \varepsilon$$

almost surely. This proves the support statement and therefore

$$\|R(x+\varepsilon Z)-x\|_2 \le \sqrt{d}\,\varepsilon$$

almost surely, giving (ii).

For (i), fix $z\in[-1,1]^d$. For each coordinate, the map

$$x_j \mapsto r(x_j+\varepsilon z_j)$$

is an interval exchange with slopes ±1 and hence preserves Lebesgue measure on $[0,1]$. Therefore the product map

$$x \mapsto R(x+\varepsilon z)$$

preserves Lebesgue measure on $[0,1]^d$. Averaging over z shows that K_ε preserves Lebesgue measure, i.e.

$$\int_{[0,1]^d} k_\varepsilon(x,y)\,dx = 1 \qquad (y\in[0,1]^d).$$

Similarly, $\int k_\varepsilon(x,y)\,dy = 1$ because k_ε is a Markov kernel by construction. The symmetry

$$k_\varepsilon(x,y) = k_\varepsilon(y,x)$$

is immediate from the explicit formula and the evenness of ψ.

For (iii), the density formula shows that for fixed x only finitely many image terms contribute to $k_\varepsilon(x,y)$, uniformly in (x,y). Differentiating under the integral sign,

$$\partial^\alpha K_\varepsilon f(y) = \int_{[0,1]^d} \partial_y^\alpha k_\varepsilon(x,y)\,f(x)\,dx.$$

Each derivative of order $|\alpha|$ brings a factor $\varepsilon^{-|\alpha|}$, while the summed image kernel remains uniformly integrable. Hence

$$\sup_{y\in[0,1]^d} \int_{[0,1]^d} |\partial_y^\alpha k_\varepsilon(x,y)|\,dx \le C_{\alpha,d,\psi}\,\varepsilon^{-|\alpha|},$$

which yields the claimed bound.

Finally, (iv) follows from positivity and doubly stochasticity: if $m\le f\le M$, then

$$K_\varepsilon f(y) = \int k_\varepsilon(x,y)\,f(x)\,dx$$

is a convex combination of values in $[m,M]$. ☐

Corollary 12.3 (A smooth full-support lower-bound core at the critical scale)**.** *Assume $d > 2p$ and fix $\lambda \in (0,1)$. Let*

$$h = 2^{-J} \qquad \textit{satisfy} \qquad h^{-d} \asymp N \log N.$$

For $\tau > 0$, define the smoothed split-shell class

$$\mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau} := \Big\{ \big((K_{\tau h})_{\#} P, (K_{\tau h})_{\#} Q \big) : \ (P,Q) \in \mathcal{C}^{\mathrm{split}}_{J,\lambda} \Big\}.$$

Then:

(i) every member of $\mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau}$ consists of C^∞ probability densities f, g on $[0,1]^d$ with full support and pointwise bounds

$$\lambda \le f, g \le 2 - \lambda;$$

(ii) for every multi-index α,

$$\|\partial^\alpha f\|_\infty + \|\partial^\alpha g\|_\infty \le C_{\alpha,d,\psi,\tau}\, h^{-|\alpha|};$$

(iii) there exists $\tau_0 = \tau_0(d,p,\lambda) > 0$ such that for every $\tau \in (0, \tau_0]$,

$$\inf_{\hat W} \sup_{(P,Q) \in \mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau}} \mathbb{E}\big|\hat W - W_p(P,Q)\big| \gtrsim_{d,p,\lambda,\tau} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q) \in \mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau}} \mathbb{E}\big(\hat W - W_p(P,Q)\big)^2 \gtrsim_{d,p,\lambda,\tau} h^2.$$

Proof. Let $(P,Q) \in \mathcal{C}^{\mathrm{split}}_{J,\lambda}$, say

$$P = f\,dx, \qquad Q = g\,dx.$$

By definition of the split-shell class,

$$\lambda \le f, g \le 2 - \lambda \qquad \text{on } [0,1]^d.$$

Applying Theorem 12.2 with $\varepsilon = \tau h$ yields the smoothness, full-support, pointwise bounds, and derivative bounds in (i)–(ii).

For the minimax lower bounds, Theorem 12.2 gives

$$\sup_x \left(\int \|y - x\|_2^p \, K_{\tau h}(x, dy) \right)^{1/p} \le \sqrt{d}\,\tau h.$$

Therefore Theorem 12.1 applies with

$$\kappa = \sqrt{d}\,\tau \qquad \text{and} \qquad \mathfrak{C} = \mathcal{C}^{\mathrm{split}}_{J,\lambda},$$

yielding

$$M^{\mathrm{abs}}_{n,m}\big(\mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau}\big) \ge M^{\mathrm{abs}}_{n,m}\big(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\big) - 2\sqrt{d}\,\tau h,$$

and

$$M^{\mathrm{sq}}_{n,m}\big(\mathcal{C}^{\mathrm{split},\infty}_{J,\lambda,\tau}\big) \ge \frac{1}{2} M^{\mathrm{sq}}_{n,m}\big(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\big) - 4d\,\tau^2 h^2.$$

By Theorem 11.3,

$$M^{\mathrm{abs}}_{n,m}\big(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\big) \asymp_{d,p,\lambda} h, \qquad M^{\mathrm{sq}}_{n,m}\big(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\big) \asymp_{d,p,\lambda} h^2.$$

Choosing $\tau_0 = \tau_0(d,p,\lambda) > 0$ sufficiently small makes the subtractive terms harmless, proving (iii). ☐

We can sharpen the same conclusion from C^∞ to real analyticity by replacing the compactly supported bump kernel with the reflected Gaussian heat kernel.

For $t > 0$, let

$$\phi_t(u) := \frac{1}{\sqrt{4\pi t}}\, e^{-u^2/(4t)} \qquad (u \in \mathbb{R}),$$

and define the one-dimensional reflected heat kernel on $[0,1]$ by

$$h^{(1)}_t(x,y) := \sum_{m \in \mathbb{Z}} \phi_t(y - x - 2m) + \sum_{m \in \mathbb{Z}} \phi_t(y + x - 2m).$$

In dimension d, set

$$h_t(x,y) := \prod_{j=1}^d h^{(1)}_t(x_j, y_j), \qquad x, y \in [0,1]^d,$$

and let H_t denote the associated Markov operator,

$$(\mathsf{H}_t f)(y) := \int_{[0,1]^d} h_t(x,y)\, f(x)\, dx.$$

Lemma 12.4 (Reflected Gaussian heat kernel)**.** *For every $t > 0$, the kernel H_t satisfies:*

(i) h_t is symmetric, strictly positive, doubly stochastic, and real analytic on $[0,1]^d \times [0,1]^d$;

(ii) if $f \in L^\infty([0,1]^d)$, then $\mathsf{H}_t f$ is real analytic on $[0,1]^d$ and, for every multi-index α,

$$\|\partial^\alpha \mathsf{H}_t f\|_\infty \le C_{\alpha,d}\, t^{-|\alpha|/2}\, \|f\|_\infty;$$

(iii) for every $p \geq 1$,

$$\sup_{x\in[0,1]^d} \left(\int \|y-x\|_2^p \, \mathsf{H}_t(x,dy) \right)^{1/p} \leq C_{d,p}\sqrt{t};$$

(iv) if $m \leq f \leq M$ almost everywhere, then

$$m \leq \mathsf{H}_t f \leq M \qquad \textit{everywhere on } [0,1]^d.$$

Proof. Let $r : \mathbb{R} \to [0,1]$ be the reflection map from the proof of Theorem 12.2, and let

$$R(u_1,\dots,u_d) := (r(u_1),\dots,r(u_d)).$$

If $G = (G_1,\dots,G_d) \sim N(0, 2t\, I_d)$, then $\mathsf{H}_t(x,\cdot)$ is exactly the law of

$$R(x+G).$$

Indeed, in one dimension the preimages of $y \in [0,1]$ under r are precisely

$$\{\, y + 2m,\ 2m - y :\ m \in \mathbb{Z} \,\},$$

and summing the Gaussian density over those preimages gives the formula for $h_t^{(1)}$.

Because r is 1-Lipschitz and $r(x) = x$ on $[0,1]$, we have

$$\|R(x+G) - x\|_2 \leq \|G\|_2$$

almost surely. Therefore

$$\sup_x \left(\int \|y-x\|_2^p \, \mathsf{H}_t(x,dy) \right)^{1/p} \leq \left(\mathbb{E}\|G\|_2^p \right)^{1/p} \leq C_{d,p}\sqrt{t},$$

which proves (iii).

For (i), fix $g \in \mathbb{R}^d$. For each coordinate, the map

$$x_j \mapsto r(x_j + g_j)$$

preserves Lebesgue measure on $[0,1]$; hence so does

$$x \mapsto R(x+g)$$

on $[0,1]^d$. Averaging over G shows that H_t preserves Lebesgue measure, i.e.

$$\int_{[0,1]^d} h_t(x,y)\, dx = 1 \qquad (y \in [0,1]^d).$$

Since h_t is a Markov kernel, also

$$\int_{[0,1]^d} h_t(x,y)\, dy = 1 \qquad (x \in [0,1]^d).$$

Symmetry is immediate from the explicit series. Strict positivity follows because every Gaussian term is positive. Moreover, for every pair of multi-indices α, β, the series defining $\partial_x^\alpha \partial_y^\beta h_t(x,y)$ converges absolutely and uniformly on $[0,1]^d \times [0,1]^d$. Indeed, Gaussian derivatives satisfy bounds of the form

$$|\phi_t^{(k)}(u)| \leq A\, B^k\, k!\, t^{-k/2}\, e^{-u^2/(8t)} \qquad (u \in \mathbb{R},\ k \geq 0),$$

for universal constants $A, B > 0$. Summing the exponentially decaying lattice tails over $m \in \mathbb{Z}$ therefore gives

$$\sup_{x,y\in[0,1]^d} |\partial_x^\alpha \partial_y^\beta h_t(x,y)| \leq A_d\, B_d^{|\alpha|+|\beta|}\, \alpha!\, \beta!\, t^{-(|\alpha|+|\beta|)/2}.$$

By the standard derivative criterion, h_t is real analytic on $[0,1]^d \times [0,1]^d$.

For (ii), each derivative of order k of ϕ_t has the form

$$\phi_t^{(k)}(u) = t^{-k/2} P_k\!\left(\frac{u}{\sqrt{t}}\right) \phi_t(u)$$

for a polynomial P_k. Hence, for each $k \geq 0$,

$$\sup_{x,y\in[0,1]} \sum_{m\in\mathbb{Z}} \left| \partial_y^k \phi_t(y \pm x - 2m) \right| \leq C_k\, t^{-k/2}.$$

Taking products over coordinates gives

$$\sup_{y\in[0,1]^d} \int_{[0,1]^d} |\partial_y^\alpha h_t(x,y)|\, dx \leq C_{\alpha,d}\, t^{-|\alpha|/2}.$$

Therefore

$$\partial^\alpha \mathsf{H}_t f(y) = \int_{[0,1]^d} \partial_y^\alpha h_t(x,y)\, f(x)\, dx$$

and

$$\|\partial^\alpha \mathsf{H}_t f\|_\infty \leq C_{\alpha,d}\, t^{-|\alpha|/2}\, \|f\|_\infty.$$

The sharper factorial derivative bound proved above shows that $\mathsf{H}_t f$ is real analytic as well.

Finally, (iv) follows exactly as in Theorem 12.2: positivity and double stochasticity imply that $\mathsf{H}_t f(y)$ is a convex combination of values of f. ☐

Corollary 12.5 (A real-analytic full-support lower-bound core at the critical scale)**.** *Assume $d > 2p$ and fix $\lambda \in (0,1)$. Let*

$$h = 2^{-J} \qquad \textit{satisfy} \qquad h^{-d} \asymp N \log N.$$

For $\tau > 0$, define

$$\mathcal{C}_{J,\lambda,\tau}^{\mathrm{split},\omega} := \left\{ \left((\mathsf{H}_{(\tau h)^2})_{\#} P, (\mathsf{H}_{(\tau h)^2})_{\#} Q \right) :\ (P,Q) \in \mathcal{C}_{J,\lambda}^{\mathrm{split}} \right\}.$$

Then:

(i) every member of $\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau}$ *consists of real-analytic probability densities* f, g *on* $[0,1]^d$ *with full support and pointwise bounds*

$$\lambda \le f, g \le 2 - \lambda;$$

(ii) for every multi-index α,

$$\|\partial^\alpha f\|_\infty + \|\partial^\alpha g\|_\infty \le C_{\alpha,d,\tau}\, h^{-|\alpha|};$$

(iii) there exists $\tau_0 = \tau_0(d,p,\lambda) > 0$ *such that for every* $\tau \in (0, \tau_0]$,

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau}} \mathbb{E}\left|\hat W - W_p(P,Q)\right| \gtrsim_{d,p,\lambda,\tau} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau}} \mathbb{E}\big(\hat W - W_p(P,Q)\big)^2 \gtrsim_{d,p,\lambda,\tau} h^2.$$

Proof. Let $(P,Q) \in \mathcal{C}^{\mathrm{split}}_{J,\lambda}$, say

$$P = f\,dx, \qquad Q = g\,dx.$$

By definition,

$$\lambda \le f, g \le 2-\lambda \qquad \text{on } [0,1]^d.$$

Applying Theorem 12.4 with $t = (\tau h)^2$ gives the analyticity, full-support, pointwise bounds, and derivative estimates in (i)–(ii).
By Theorem 12.4,

$$\sup_x \left(\int \|y-x\|_2^p\, \mathsf{H}_{(\tau h)^2}(x,dy)\right)^{1/p} \le C_{d,p}\,\tau h.$$

Hence Theorem 12.1 applies with

$$\kappa = C_{d,p}\tau \qquad \text{and} \qquad \mathfrak{C} = \mathcal{C}^{\mathrm{split}}_{J,\lambda},$$

so

$$M^{\mathrm{abs}}_{n,m}\left(\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau}\right) \ge M^{\mathrm{abs}}_{n,m}\left(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\right) - 2C_{d,p}\tau h,$$

and

$$M^{\mathrm{sq}}_{n,m}\left(\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau}\right) \ge \frac{1}{2} M^{\mathrm{sq}}_{n,m}\left(\mathcal{C}^{\mathrm{split}}_{J,\lambda}\right) - 4C_{d,p}^2\tau^2 h^2.$$

Now use Theorem 11.3 and choose $\tau_0 = \tau_0(d,p,\lambda) > 0$ sufficiently small. ☐

Corollary 12.6 (Exact target-scale law on a real-analytic full-support critical neighborhood)**.** *Assume* $d > 4$ *and* $p = 2$*. Fix* $\lambda \in (0,1)$*,* $\tau > 0$*, and let* $C_{d,2}$ *denote the constant from Theorem 12.4*(iii)*. Choose*

$$L \ge 1 + 2C_{d,2}\tau.$$

Let

$$h = 2^{-J} \qquad \textit{satisfy} \qquad h^{-d} \asymp N \log N,$$

and define the real-analytic critical neighborhood

$$\mathcal{A}^\omega_{J,\lambda,\tau}(L) := \left\{(P,Q) : \begin{array}{l} P = f\,dx,\ Q = g\,dx,\ \lambda \le f,g \le 2-\lambda,\ W_2(P,Q) \le Lh, \\ \|\partial^\alpha f\|_\infty + \|\partial^\alpha g\|_\infty \le C_{\alpha,d,\tau}\, h^{-|\alpha|} \quad \forall \alpha \end{array}\right\},$$

where $C_{\alpha,d,\tau}$ *is the derivative envelope from Theorem 12.5*(ii)*. Then*

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{A}^\omega_{J,\lambda,\tau}(L)} \mathbb{E}\left|\hat W - W_2(P,Q)\right| \asymp_{d,\lambda,\tau,L} h,$$

and

$$\inf_{\hat W} \sup_{(P,Q)\in\mathcal{A}^\omega_{J,\lambda,\tau}(L)} \mathbb{E}\big(\hat W - W_2(P,Q)\big)^2 \asymp_{d,\lambda,\tau,L} h^2.$$

Proof. Because

$$\mathcal{A}^\omega_{J,\lambda,\tau}(L) \subset \mathcal{U}^{\mathrm{crit}}_{N,\lambda,2-\lambda}(L),$$

the upper bounds follow immediately from Theorem 14.2.
For the lower bounds, let

$$(P',Q') \in \mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau},$$

say

$$P' = (\mathsf{H}_{(\tau h)^2})_\# P, \qquad Q' = (\mathsf{H}_{(\tau h)^2})_\# Q$$

for some $(P,Q) \in \mathcal{C}^{\mathrm{split}}_{J,\lambda}$. By Theorem 12.5(i)–(ii), the pair (P',Q') satisfies the pointwise and derivative bounds defining $\mathcal{A}^\omega_{J,\lambda,\tau}(L)$. Moreover, Theorem 12.4(iii) gives

$$W_2(P,P') \le C_{d,2}\tau h, \qquad W_2(Q,Q') \le C_{d,2}\tau h,$$

while Theorem 11.2 with $p = 2$ yields

$$W_2(P,Q) \le h.$$

Hence

$$W_2(P',Q') \le W_2(P',P) + W_2(P,Q) + W_2(Q,Q') \le (1 + 2C_{d,2}\tau)h \le Lh.$$

Therefore

$$\mathcal{C}^{\mathrm{split},\omega}_{J,\lambda,\tau} \subset \mathcal{A}^\omega_{J,\lambda,\tau}(L).$$

Applying Theorem 12.5 now gives

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{A}^{\omega}_{J,\lambda,\tau}(L)} \mathbb{E}\left|\hat{W} - W_2(P,Q)\right| \gtrsim_{d,\lambda,\tau} h,$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{A}^{\omega}_{J,\lambda,\tau}(L)} \mathbb{E}\left(\hat{W} - W_2(P,Q)\right)^2 \gtrsim_{d,\lambda,\tau} h^2.$$

Combining with the upper bounds proves the claim. □

Remark 12.7 (Regularization alone cannot remove the critical hardness). The class of Theorem 12.3 is genuinely smooth and has full support, but its derivative bounds naturally scale like $h^{-|\alpha|}$ with the critical mesh size. Thus this result does not contradict the faster minimax laws known on *fixed* smoothness classes; rather, it shows that the critical lower-bound mechanism survives every known local regularization whose displacement radius is only $O(h)$. In particular, pure smoothing of the hard full-support core cannot by itself close the unrestricted problem.

13 Scale-local blocks and fixed-depth towers

The one-step shell theorem of Section 11 already shows that the critical $(N \log N)^{-1/d}$ law survives on a broad touching-cell full-support continuum model. We now isolate the underlying mechanism. What matters for the universal upper law is not a special left/right split, but the fact that all discrepancies are confined inside blocks of Euclidean diameter $O(h)$ on the critical mesh. Exact additive decomposition is subtler: it does hold under a genuine inter-block cost gap, but it fails in general without such a gap. This section proves the universal upper result on arbitrary local block partitions, the exact critical law on broad active-block families, and an explicit exactness criterion for genuinely separated local decompositions.

13.1 Atomic block-local models on the critical grid

Fix a dyadic level $J \geq 1$ and write

$$h := 2^{-J}.$$

For every fine dyadic cube $C \in \mathcal{D}_J$, let z_C denote its center, and set

$$Z_J := \{z_C : C \in \mathcal{D}_J\}, \qquad |Z_J| = 2^{Jd} = h^{-d}.$$

Distinct points of Z_J satisfy

$$\|z_C - z_{C'}\|_2 \geq h \qquad (C \neq C').$$

Definition 13.1 (A-local block partition and atomic block-local class). Fix $A \geq 1$. An *A-local block partition* of Z_J is a partition

$$\mathfrak{B}_J = \{B_1, \ldots, B_M\}$$

such that

$$\operatorname{diam}(B_r) \leq Ah \qquad (1 \leq r \leq M).$$

For such a partition, let $\mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)$ be the class of all pairs

$$(\mu, \nu) \in \mathcal{P}(Z_J)^2$$

satisfying

$$\mu(B_r) = \nu(B_r) \qquad (1 \leq r \leq M).$$

Equivalently, all transport discrepancy is confined inside blocks of diameter at most Ah.

Proposition 13.2 (Block-local atomic transport is sandwiched by fine-grid L_1). *Let $\mathfrak{B}_J$ be an A-local block partition of Z_J. Then for every $(\mu,\nu) \in \mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)$ and every $p \geq 1$,*

$$h^p \ \mathrm{TV}(\mu,\nu) \leq W_p(\mu,\nu)^p \leq (Ah)^p \ \mathrm{TV}(\mu,\nu).$$

Equivalently,

$$h\left(\frac{\|\mu-\nu\|_1}{2}\right)^{1/p} \leq W_p(\mu,\nu) \leq Ah\left(\frac{\|\mu-\nu\|_1}{2}\right)^{1/p}.$$

Proof. For the lower bound, first match the common mass

$$\mu \wedge \nu$$

at each atom at zero cost. The remaining unmatched mass has total amount

$$\mathrm{TV}(\mu,\nu) = \frac{\|\mu-\nu\|_1}{2},$$

and every unit of it must move from one grid point of Z_J to a *different* grid point. Since distinct points of Z_J are at Euclidean distance at least h, every such unit pays cost at least h^p. Hence

$$W_p(\mu,\nu)^p \geq h^p \ \mathrm{TV}(\mu,\nu).$$

For the upper bound, fix one block B_r. Because $\mu(B_r) = \nu(B_r)$, the positive and negative parts of $(\mu-\nu)|_{B_r}$ have the same total mass. Thus one may couple the excess of μ inside B_r to the deficit of ν inside B_r using only pairs of points belonging to B_r. Every such pair has Euclidean distance at most $\operatorname{diam}(B_r) \leq Ah$, so the cost incurred inside B_r is at most

$$(Ah)^p \ \mathrm{TV}(\mu|_{B_r}, \nu|_{B_r}).$$

Summing over all blocks gives

$$W_p(\mu,\nu)^p \leq (Ah)^p \sum_{r=1}^{M} \mathrm{TV}(\mu|_{B_r}, \nu|_{B_r}) = (Ah)^p \ \mathrm{TV}(\mu,\nu),$$

which proves the claim. □

Theorem 13.3 (Critical upper law on arbitrary local block partitions). *Fix $A \geq 1$ and $p \geq 1$. Assume that*

$$h^{-d} \asymp N \log N.$$

Then for every A-local block partition $\mathfrak{B}_J$ of Z_J, there exists an estimator

$$\widehat{W}_{J,A}^{\mathrm{bloc}}$$

such that

$$\sup_{(\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\left|\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right| \le C_{A,d,p}\, h,$$

and

$$\sup_{(\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\left(\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right)^2 \le C_{A,d,p}\, h^2.$$

Proof. Since the support Z_J is known and has size

$$|Z_J| = h^{-d} \asymp N\log N,$$

the large-alphabet two-sample L_1 theory of Jiao–Han–Weissman [5, Theorem 6] yields an estimator $\widehat{L}_J$ of $\|\mu-\nu\|_1$ such that

$$\sup_{(\mu,\nu)\in\mathcal{P}(Z_J)^2} \mathbb{E}\left(\widehat{L}_J - \|\mu-\nu\|_1\right)^2 \le C_d.$$

Clip to the natural parameter range:

$$\widetilde{L}_J := 0\vee \widehat{L}_J \wedge 2, \qquad \widehat{V}_J := \frac{\widetilde{L}_J}{2} \in [0,1], \qquad V(\mu,\nu) := \mathrm{TV}(\mu,\nu)\in[0,1].$$

Define

$$\widehat{W}_{J,A}^{\mathrm{bloc}} := h\,\widehat{V}_J^{\,1/p}.$$

By Theorem 13.2,

$$h\,V(\mu,\nu)^{1/p} \le W_p(\mu,\nu) \le Ah\,V(\mu,\nu)^{1/p} \qquad \left((\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)\right).$$

Therefore

$$\left|\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right| \le h\left|\widehat{V}_J^{\,1/p} - V(\mu,\nu)^{1/p}\right| + (A-1)h\,V(\mu,\nu)^{1/p}.$$

Since $V(\mu,\nu)\in[0,1]$ and $x\mapsto x^{1/p}$ is $1/p$-Hölder on $[0,\infty)$,

$$\left|\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right| \le h\,|\widehat{V}_J - V(\mu,\nu)|^{1/p} + (A-1)h.$$

Taking expectations and using Jensen's inequality gives

$$\mathbb{E}\left|\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right| \le h\left(\mathbb{E}|\widehat{V}_J - V(\mu,\nu)|^2\right)^{1/(2p)} + (A-1)h \le C_{A,d,p}h.$$

For the squared risk,

$$\left|\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right|^2 \le 2h^2|\widehat{V}_J - V(\mu,\nu)|^{2/p} + 2(A-1)^2h^2.$$

If $p\ge 2$, then $2/p\le 1$, so Jensen gives

$$\mathbb{E}|\widehat{V}_J - V(\mu,\nu)|^{2/p} \le \left(\mathbb{E}|\widehat{V}_J - V(\mu,\nu)|^2\right)^{1/p} \le C_{d,p}.$$

If $1\le p<2$, then $2/p>1$, but $|\widehat{V}_J - V(\mu,\nu)|\le 1$, hence

$$|\widehat{V}_J - V(\mu,\nu)|^{2/p} \le |\widehat{V}_J - V(\mu,\nu)|$$

and so

$$\mathbb{E}|\widehat{V}_J - V(\mu,\nu)|^{2/p} \le \left(\mathbb{E}|\widehat{V}_J - V(\mu,\nu)|^2\right)^{1/2} \le C_d.$$

Thus in all cases

$$\mathbb{E}\left(\widehat{W}_{J,A}^{\mathrm{bloc}} - W_p(\mu,\nu)\right)^2 \le C_{A,d,p}h^2.$$

□

13.2 Continuum block-local lifts

For every fine cube $C\in\mathcal{D}_J$, let U_C be the normalized Lebesgue measure on C.

Definition 13.4 (Continuum block-local lift). Fix $\lambda\in(0,1)$ and let $\mathfrak{B}_J$ be an A-local block partition of Z_J. Define $\mathcal{L}_{J,\lambda}^{\mathrm{bloc}}(\mathfrak{B}_J)$ to be the class of all pairs

$$P = \sum_{C\in\mathcal{D}_J} p_C\,U_C, \qquad Q = \sum_{C\in\mathcal{D}_J} q_C\,U_C,$$

such that

$$\lambda h^d \le p_C, q_C \le (2-\lambda)h^d \qquad (C\in\mathcal{D}_J),$$

and

$$\sum_{z_C\in B_r} p_C = \sum_{z_C\in B_r} q_C \qquad (1\le r\le M).$$

Equivalently, the measures are positive piecewise-constant full-support densities on the whole cube, and all discrepancy is confined inside blocks of diameter $O(h)$.

Corollary 13.5 (Continuum block-local upper law). *Fix $A\ge 1$, $\lambda\in(0,1)$, and $p\ge 1$. Assume that*

$$h^{-d} \asymp N\log N.$$

Then for every A-local block partition $\mathfrak{B}_J$ of Z_J, there exists an estimator

$$\widehat{W}_{J,A,\lambda}^{\mathrm{bloc,cont}}$$

such that

$$\sup_{(P,Q)\in\mathcal{L}^{\mathrm{bloc}}_{J,\lambda}(\mathfrak{B}_J)} \mathbb{E}\left|\widehat{W}^{\mathrm{bloc,cont}}_{J,A,\lambda} - W_p(P,Q)\right| \le C_{A,d,p}\, h,$$

and

$$\sup_{(P,Q)\in\mathcal{L}^{\mathrm{bloc}}_{J,\lambda}(\mathfrak{B}_J)} \mathbb{E}\left(\widehat{W}^{\mathrm{bloc,cont}}_{J,A,\lambda} - W_p(P,Q)\right)^2 \le C_{A,d,p}\, h^2.$$

Proof. Let

$$\Pi_J := \mathcal{D}_J$$

be the fine dyadic partition and let

$$\mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J) \subset \mathcal{P}(Z_J)^2$$

be the atomic block-local class from Theorem 13.1. By construction,

$$\mathcal{L}^{\mathrm{bloc}}_{J,\lambda}(\mathfrak{B}_J) \subseteq \mathcal{L}^{\mathrm{part}}_{\Pi_J}\left(\mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)\right),$$

where the anchors are the cell centers z_C and every cell of Π_J has diameter at most $\sqrt{d}\,h$. Therefore Theorem 10.4 and Theorem 13.3 imply

$$M^{\mathrm{abs}}_{n,m}\left(\mathcal{L}^{\mathrm{bloc}}_{J,\lambda}(\mathfrak{B}_J)\right) \le M^{\mathrm{abs}}_{n,m}\left(\mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)\right) + 2\sqrt{d}\,h \le C_{A,d,p}h,$$

and similarly

$$M^{\mathrm{sq}}_{n,m}\left(\mathcal{L}^{\mathrm{bloc}}_{J,\lambda}(\mathfrak{B}_J)\right) \le 2\, M^{\mathrm{sq}}_{n,m}\left(\mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)\right) + 8d\,h^2 \le C_{A,d,p}h^2.$$

□

13.3 Many active local blocks force the exact critical law

The universal $O(h)$ upper law above is sharp on a much wider class of local partitions than the fixed-depth tower examples alone. The only ingredient needed for the lower bound is a positive-density supply of genuinely movable microscopic blocks.

Definition 13.6 (Nontrivial block count)**.** For an A-local block partition

$$\mathfrak{B}_J = \{B_1,\dots,B_M\}$$

of Z_J, define

$$K_{\ge 2}(\mathfrak{B}_J) := \#\{1\le r\le M:\ |B_r|\ge 2\}.$$

Thus $K_{\ge 2}(\mathfrak{B}_J)$ counts the blocks that contain at least two distinct grid points and hence allow nonzero within-block transport.

Proposition 13.7 (Separated repeated microscopic pairs)**.** *Fix $A\ge 1$. There exist constants $\tau_{A,d}>0$ and a finite set*

$$\mathcal{V}_{A,d}\subset \mathbb{Z}^d\setminus\{0\}$$

such that the following holds. Let $\mathfrak{B}_J=\{B_1,\dots,B_M\}$ be an A-local block partition of Z_J, and let

$$K := K_{\ge 2}(\mathfrak{B}_J).$$

Then there exist a vector $v_\star\in\mathcal{V}_{A,d}$ and a subcollection

$$\mathcal{I}_\star \subseteq \{r:\ |B_r|\ge 2\}$$

with

$$|\mathcal{I}_\star| \ge \tau_{A,d}\, K$$

such that for every $r\in\mathcal{I}_\star$ one can choose points

$$a_r, b_r \in B_r$$

satisfying

$$b_r - a_r = h\, v_\star,$$

and the selected blocks are pairwise separated in the sense that

$$\mathrm{dist}(B_r,B_s) > 4Ah \qquad (r\ne s,\ r,s\in\mathcal{I}_\star).$$

Proof. Let

$$\mathcal{V}_{A,d} := \{v\in\mathbb{Z}^d\setminus\{0\}:\ \|v\|_2\le A\},$$

which is finite. For every nontrivial block B_r, choose two distinct points $x_r,y_r\in B_r$. Since $x_r,y_r\in Z_J$, one has

$$y_r - x_r = h\,v_r \qquad \text{for some } v_r\in\mathcal{V}_{A,d}.$$

By the pigeonhole principle there exists $v_\star\in\mathcal{V}_{A,d}$ and a subcollection

$$\mathcal{J}_\star \subseteq \{r:\ |B_r|\ge 2\}$$

such that

$$|\mathcal{J}_\star| \ge \frac{K}{|\mathcal{V}_{A,d}|}$$

and every $r\in\mathcal{J}_\star$ contains a pair (a_r,b_r) with

$$b_r - a_r = h\,v_\star.$$

Now build a graph on the vertex set $\mathcal{J}_\star$ by joining $r\ne s$ whenever

$$\mathrm{dist}(B_r,B_s)\le 4Ah.$$

We claim that this graph has degree bounded by a constant depending only on A and d. Indeed, fix $r\in\mathcal{J}_\star$ and choose any $z_r\in B_r$. If s is adjacent to r, there exist $x\in B_r$ and $y\in B_s$ with

$$\|x-y\|_2\le 4Ah.$$

Since $\mathrm{diam}(B_r),\mathrm{diam}(B_s)\le Ah$, every point of B_s lies within distance at most $6Ah$ of z_r. Because distinct blocks are disjoint subsets of the critical grid Z_J, each adjacent block contributes at least one distinct grid point inside the Euclidean ball

$$B(z_r, 6Ah).$$

The number of grid points in that ball is $O_{A,d}(1)$, so the graph degree is bounded by some constant $D_{A,d} < \infty$.

Every finite graph of maximum degree $D_{A,d}$ contains an independent set of size at least

$$\frac{|\mathcal{J}_\star|}{D_{A,d}+1}.$$

Let $\mathcal{I}_\star$ be such an independent set. Then for distinct $r, s \in \mathcal{I}_\star$,

$$\operatorname{dist}(B_r, B_s) > 4Ah,$$

and

$$|\mathcal{I}_\star| \ge \frac{1}{(D_{A,d}+1)|\mathcal{V}_{A,d}|} K.$$

This proves the claim with

$$\tau_{A,d} := \frac{1}{(D_{A,d}+1)|\mathcal{V}_{A,d}|}.$$

□

Theorem 13.8 (Exact critical law for arbitrary local partitions with many active blocks)**.** *Assume $d > 2p$, fix $A \ge 1$, and let*

$$h^{-d} \asymp N \log N.$$

Then for every $c_0 > 0$ there exists $C_{A,d,p,c_0} < \infty$ such that the following holds. If $\mathfrak{B}_J$ is an A-local block partition of Z_J satisfying

$$K_{\ge 2}(\mathfrak{B}_J) \ge c_0\, h^{-d},$$

then

$$\inf_{\hat W} \sup_{(\mu,\nu)\in \mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\left|\hat W - W_p(\mu,\nu)\right| \asymp_{A,d,p,c_0} h,$$

and

$$\inf_{\hat W} \sup_{(\mu,\nu)\in \mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\left(\hat W - W_p(\mu,\nu)\right)^2 \asymp_{A,d,p,c_0} h^2.$$

Proof. The upper bounds are exactly Theorem 13.3.

For the lower bounds, let

$$K := K_{\ge 2}(\mathfrak{B}_J).$$

By Theorem 13.7, there exist a vector $v_\star \in \mathbb{Z}^d \setminus \{0\}$ with

$$1 \le \|v_\star\|_2 \le A,$$

and a subcollection

$$\mathcal{I}_\star = \{1, \dots, M_\star\}$$

of nontrivial blocks such that

$$M_\star \ge \tau_{A,d} K \gtrsim_{A,d,c_0} h^{-d} \asymp N \log N,$$

each block B_i contains points a_i, b_i with

$$b_i - a_i = h\, v_\star,$$

and the selected blocks satisfy

$$\operatorname{dist}(B_i, B_j) > 4Ah \qquad (i \ne j).$$

For distributions $r = (r_i)_{i=1}^{M_\star}$ and $s = (s_i)_{i=1}^{M_\star}$ on the alphabet $\{1, \dots, M_\star\}$, define atomic measures on Z_J by

$$P_{r,s}(a_i) := \frac{r_i}{2}, \qquad P_{r,s}(b_i) := \frac{s_i}{2},$$

$$Q_{r,s}(a_i) := \frac{s_i}{2}, \qquad Q_{r,s}(b_i) := \frac{r_i}{2} \qquad (1 \le i \le M_\star),$$

and set all remaining masses to zero. Then $P_{r,s}$ and $Q_{r,s}$ are probability measures, and for every selected block B_i,

$$P_{r,s}(B_i) = Q_{r,s}(B_i) = \frac{r_i + s_i}{2}.$$

Hence

$$(P_{r,s}, Q_{r,s}) \in \mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J).$$

We next compute the exact transport cost on this embedded subclass. Let

$$d_\star := \|v_\star\|_2\, h.$$

Inside block B_i the only discrepancy is

$$P_{r,s}(a_i) - Q_{r,s}(a_i) = \frac{r_i - s_i}{2}, \qquad P_{r,s}(b_i) - Q_{r,s}(b_i) = -\frac{r_i - s_i}{2}.$$

Transporting this imbalance directly between a_i and b_i for each i yields the upper bound

$$W_p(P_{r,s}, Q_{r,s})^p \le d_\star^p \ \mathrm{TV}(r, s).$$

For the matching lower bound, fix any coupling π between $P_{r,s}$ and $Q_{r,s}$. If $r_i > s_i$, then the target atom b_i has deficit $(r_i - s_i)/2$. Under π, that much mass must therefore arrive at b_i from points other than b_i itself. The source atom a_i lies at distance exactly $d_\star$ from b_i, whereas every atom belonging to a different selected block lies at distance strictly larger than

$$4Ah \ge d_\star,$$

because the selected blocks are pairwise $4Ah$-separated and each block has diameter at most Ah. Therefore at least $(r_i - s_i)/2$ mass pays cost at least $d_\star^p$ in order to supply b_i. The same argument with a_i and b_i interchanged applies when $s_i > r_i$. Summing over i gives

$$\int \|x - y\|_2^p \, d\pi(x, y) \ge d_\star^p \ \mathrm{TV}(r, s).$$

Taking the infimum over π yields the exact identity

$$W_p(P_{r,s}, Q_{r,s})^p = d_\star^p \ \mathrm{TV}(r, s) = d_\star^p \frac{\|r - s\|_1}{2}.$$

Thus estimating W_p on the block-local class is at least as hard as estimating

$$\mathrm{TV}(r,s) = \frac{\|r-s\|_1}{2}$$

for two unknown distributions on an alphabet of size

$$M_\star \asymp N \log N.$$

By the large-alphabet minimax lower bounds of Jiao–Han–Weissman [5, Theorem 3 and Theorem 5],

$$\inf_{\widehat{T}} \sup_{r,s} \mathbb{E}\Big(\widehat{T} - \mathrm{TV}(r,s)\Big)^2 \gtrsim_{A,d,c_0} 1.$$

Since $\mathrm{TV}(r,s) \in [0,1]$, clipping implies the corresponding absolute minimax risk is also bounded below by a positive constant.

Now let $\hat{W}$ be any estimator on $\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)$, and define on the embedded subclass

$$\widehat{T} := \left(\frac{0 \vee \hat{W} \wedge d_\star}{d_\star}\right)^p.$$

The map $w \mapsto (w/d_\star)^p$ is $(p/d_\star)$-Lipschitz on $[0, d_\star]$, so on the embedded subclass,

$$\mathbb{E}\Big|\widehat{T} - \mathrm{TV}(r,s)\Big| \le \frac{p}{d_\star}\,\mathbb{E}\Big|\hat{W} - W_p(P_{r,s}, Q_{r,s})\Big| \lesssim_{A,p} h^{-1}\mathbb{E}\Big|\hat{W} - W_p(P_{r,s}, Q_{r,s})\Big|.$$

Since the left-hand side has minimax lower bound bounded below by a positive constant, we obtain

$$\inf_{\hat{W}} \sup_{(\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\Big|\hat{W} - W_p(\mu,\nu)\Big| \gtrsim_{A,d,p,c_0} h.$$

Finally,

$$\left(\inf_{\hat{W}} \sup_{(\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\Big|\hat{W} - W_p(\mu,\nu)\Big|\right)^2 \le \inf_{\hat{W}} \sup_{(\mu,\nu)\in\mathfrak{C}_J^{\mathrm{bloc}}(\mathfrak{B}_J)} \mathbb{E}\Big(\hat{W} - W_p(\mu,\nu)\Big)^2,$$

so the squared-risk lower bound follows as well. □

Remark 13.9 (Scope of the active-block criterion). Theorem 13.8 shows that the exact critical law does not rely on any dyadic hierarchy, laminar organization, tree structure, or sparse-shortcut description. At the critical mesh, it is enough that a positive fraction of the $h^{-d} \asymp N \log N$ microscopic blocks contain *some* movable pair of grid sites. In that sense the exact target-scale theory now covers arbitrary bounded-diameter local partitions with macroscopic active-block count, far beyond the previously isolated shell and fixed-depth tower subclasses.

13.4 Fixed-depth contiguous dyadic towers

The most concrete consequence is a genuinely broad contiguous full-support family that strictly extends the one-step shell world.

Definition 13.10 (Fixed-depth dyadic tower). Fix an integer $L \ge 1$, a dyadic level $J \ge L$, and $\lambda \in (0,1)$. Let $\mathcal{T}_{J,L,\lambda}$ consist of all pairs

$$P = \sum_{C\in\mathcal{D}_J} p_C\, U_C, \qquad Q = \sum_{C\in\mathcal{D}_J} q_C\, U_C,$$

such that

$$\lambda h^d \le p_C, q_C \le (2-\lambda) h^d \qquad (C \in \mathcal{D}_J),$$

and

$$\sum_{C\subset R} p_C = \sum_{C\subset R} q_C = 2^{Ld} h^d \qquad (R \in \mathcal{D}_{J-L}).$$

Equivalently, P and Q are positive piecewise-constant densities on the critical fine grid, the average density on every level-$(J-L)$ coarse cube is exactly one for both measures, and all discrepancy is confined inside the descendants of each such coarse cube.

Corollary 13.11 (Exact critical law on fixed-depth contiguous towers). *Assume $d > 2p$, fix $L \ge 1$ and $\lambda \in (0,1)$, and let*

$$h = 2^{-J} \qquad \textit{satisfy} \qquad h^{-d} \asymp N \log N.$$

Then

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{T}_{J,L,\lambda}} \mathbb{E}\Big|\hat{W} - W_p(P,Q)\Big| \asymp_{d,p,L,\lambda} h,$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{T}_{J,L,\lambda}} \mathbb{E}\Big(\hat{W} - W_p(P,Q)\Big)^2 \asymp_{d,p,L,\lambda} h^2.$$

Proof. For every coarse cube $R \in \mathcal{D}_{J-L}$ let

$$B_R := \{z_C :\ C \subset R,\ C \in \mathcal{D}_J\} \subset Z_J.$$

The family

$$\mathfrak{B}_{J,L} := \{B_R :\ R \in \mathcal{D}_{J-L}\}$$

is a block partition of Z_J. If $z_C, z_{C'} \in B_R$, then both points lie in R, so

$$\|z_C - z_{C'}\|_2 \le \mathrm{diam}(R) = \sqrt{d}\, 2^L h.$$

Hence $\mathfrak{B}_{J,L}$ is $A_{d,L}$-local with

$$A_{d,L} := \sqrt{d}\, 2^L.$$

By definition,

$$\mathcal{T}_{J,L,\lambda} \subseteq \mathcal{L}_{J,\lambda}^{\mathrm{bloc}}(\mathfrak{B}_{J,L}),$$

so the upper bounds follow from Theorem 13.5.

For the lower bounds, note that the split-shell class from Theorem 11.1 is contained in $\mathcal{T}_{J,L,\lambda}$. Indeed, every member of $\mathcal{C}^{\mathrm{split}}_{J,\lambda}$ is piecewise constant on $\mathcal{D}_J$, satisfies the same pointwise density bounds, and has equal mass on every parent cube in $\mathcal{D}_{J-1}$; summing those parent equalities over the descendants of any $R \in \mathcal{D}_{J-L}$ yields the level-$(J-L)$ block equalities required in Theorem 13.10. Therefore

$$\mathcal{C}^{\mathrm{split}}_{J,\lambda} \subseteq \mathcal{T}_{J,L,\lambda}.$$

Applying Theorem 11.3 gives

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{T}_{J,L,\lambda}} \mathbb{E}\left|\hat{W} - W_p(P,Q)\right| \geq \inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{split}}_{J,\lambda}} \mathbb{E}\left|\hat{W} - W_p(P,Q)\right| \gtrsim_{d,p,\lambda} h,$$

and similarly for the squared risk. Combining this with the upper bound proves the claim. □

Corollary 13.12 (The one-step contiguous shell is critical for every $p \geq 1$)**.** *Assume $d > 2p$, fix $\lambda \in (0,1)$, and let*

$$h = 2^{-J} \qquad \textit{satisfy} \qquad h^{-d} \asymp N \log N.$$

Then

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\left|\hat{W} - W_p(P,Q)\right| \asymp_{d,p,\lambda} h,$$

and

$$\inf_{\hat{W}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{shell}}_{J,\lambda}} \mathbb{E}\big(\hat{W} - W_p(P,Q)\big)^2 \asymp_{d,p,\lambda} h^2.$$

Proof. When $L = 1$, the block condition in Theorem 13.10 reads

$$\sum_{C\subset R} p_C = \sum_{C\subset R} q_C = 2^d h^d \qquad (R \in \mathcal{D}_{J-1}),$$

which is exactly the defining parent-mass condition in Theorem 11.5. Hence

$$\mathcal{T}_{J,1,\lambda} = \mathcal{C}^{\mathrm{shell}}_{J,\lambda}.$$

Apply Theorem 13.11 with $L = 1$. □

13.5 A sharp exactness criterion for separated local blocks

The universal upper law of Theorem 13.3 and the exact critical law of Theorem 13.8 do *not* rely on any exact additive block decomposition. Nevertheless there is a natural sharp sufficient condition under which such a decomposition does hold: a strict inter-block cost gap. The next proposition isolates this mechanism on the critical grid.

Proposition 13.13 (Exact block decomposition under a strict separation gap)**.** *Let $\mathfrak{B}_J = \{B_1, \ldots, B_M\}$ be an A-local block partition of Z_J such that*

$$\operatorname{dist}(B_r, B_s) > Ah \qquad (r \neq s).$$

Then for every $(\mu,\nu) \in \mathfrak{C}^{\mathrm{bloc}}_J(\mathfrak{B}_J)$,

$$W_p(\mu,\nu)^p = \sum_{r=1}^{M} \mathsf{T}_r(\mu,\nu),$$

where

$$\mathsf{T}_r(\mu,\nu) := \inf_{\pi_r\in\Pi(\mu|_{B_r},\nu|_{B_r})} \int_{B_r\times B_r} \|x-y\|_2^p \, d\pi_r(x,y)$$

is the blockwise optimal transport cost on B_r.

Proof. For the upper bound, choose for each r an optimal blockwise coupling

$$\pi_r \in \Pi(\mu|_{B_r}, \nu|_{B_r})$$

attaining $\mathsf{T}_r(\mu,\nu)$. Since the blocks are disjoint and

$$\mu(B_r) = \nu(B_r) \qquad (1 \leq r \leq M),$$

the sum

$$\pi := \sum_{r=1}^{M} \pi_r$$

is a coupling between μ and ν. Hence

$$W_p(\mu,\nu)^p \leq \sum_{r=1}^{M} \mathsf{T}_r(\mu,\nu).$$

For the reverse inequality, fix any coupling $\pi \in \Pi(\mu,\nu)$. For $r \neq s$, let

$$m_{rs} := \pi(B_r \times B_s).$$

Because $\mu(B_r) = \nu(B_r)$, one has for every r

$$\sum_{s\neq r} m_{rs} = \mu(B_r) - m_{rr} = \nu(B_r) - m_{rr} = \sum_{s\neq r} m_{sr}.$$

Thus the off-diagonal block-flow matrix $(m_{rs})_{r\neq s}$ is a nonnegative circulation.
Assume first that some off-diagonal mass is present. Choose r_1 with

$$\sum_{s\neq r_1} m_{r_1 s} > 0.$$

Since the off-diagonal flow is divergence-free, every visited block that receives positive off-diagonal mass also emits positive off-diagonal mass. Following positive outgoing edges and using finiteness of the block set, we obtain a directed cycle

$$r_1 \to r_2 \to \cdots \to r_k \to r_1$$

such that

$$m_{r_i r_{i+1}} > 0 \qquad (1 \le i \le k),$$

with the convention $r_{k+1} := r_1$. Set

$$\lambda := \min_{1 \le i \le k} m_{r_i r_{i+1}} > 0.$$

For each i, choose a subcoupling

$$\sigma_i \le \pi|_{B_{r_i} \times B_{r_{i+1}}}$$

of total mass λ. Define finite measures on B_{r_i} by

$$\alpha_i := (\mathrm{pr}_1)_\# \sigma_i, \qquad \beta_i := (\mathrm{pr}_2)_\# \sigma_{i-1},$$

where $\sigma_0 := \sigma_k$. Both α_i and β_i have total mass λ, so there exists a coupling

$$\tau_i \in \Pi(\alpha_i, \beta_i)$$

supported on $B_{r_i} \times B_{r_i}$.

Now define

$$\pi' := \pi - \sum_{i=1}^k \sigma_i + \sum_{i=1}^k \tau_i.$$

Because the incoming and outgoing cycle marginals match blockwise, π' is again a coupling between μ and ν. Moreover, at least one off-diagonal block mass vanishes in π'. The cost removed from the cycle is at least

$$\sum_{i=1}^k \lambda \ \mathrm{dist}(B_{r_i}, B_{r_{i+1}})^p > k\lambda (Ah)^p,$$

whereas the cost added inside the blocks is at most

$$\sum_{i=1}^k \lambda \ \mathrm{diam}(B_{r_i})^p \le k\lambda (Ah)^p.$$

Hence

$$\int \|x-y\|_2^p \, d\pi'(x,y) \le \int \|x-y\|_2^p \, d\pi(x,y).$$

Repeating this cycle-elimination step finitely many times, we arrive at a coupling

$$\widetilde{\pi} \in \Pi(\mu, \nu)$$

with

$$\widetilde{\pi}(B_r \times B_s) = 0 \qquad (r \ne s)$$

and

$$\int \|x-y\|_2^p \, d\widetilde{\pi}(x,y) \le \int \|x-y\|_2^p \, d\pi(x,y).$$

Therefore every coupling π has cost at least that of some within-block coupling. But any within-block coupling decomposes as

$$\widetilde{\pi} = \sum_{r=1}^M \widetilde{\pi}_r, \qquad \widetilde{\pi}_r \in \Pi(\mu|_{B_r}, \nu|_{B_r}),$$

so

$$\int \|x-y\|_2^p \, d\widetilde{\pi}(x,y) = \sum_{r=1}^M \int_{B_r \times B_r} \|x-y\|_2^p \, d\widetilde{\pi}_r(x,y) \ge \sum_{r=1}^M \mathsf{T}_r(\mu, \nu).$$

Taking the infimum over π yields the reverse inequality

$$W_p(\mu,\nu)^p \ge \sum_{r=1}^M \mathsf{T}_r(\mu,\nu).$$

Combining both bounds proves the claim. □

14 Quadratic bounded-density theory

We now specialize to $p = 2$ and $d > 4$. Let

$$\eta_N = (N \log N)^{-1/d}, \qquad \gamma_N = N^{-1/d} (\log N)^{1/d}.$$

14.1 Weighted negative Sobolev geometry

For a density a satisfying $0 < m \le a \le M < \infty$, define the weighted negative Sobolev norm of a zero-mean distribution u by

$$\|u\|_{-1,a}^2 := \sup_{\varphi \in H^1([0,1]^d)/\mathbb{R}} \left\{ 2 \int \varphi \, u - \int a |\nabla \varphi|^2 \right\}.$$

The unweighted norm is denoted $\|u\|_{-1}$.

Proposition 14.1 (Bounded-density comparison)**.** *Let $P = f\,dx$ and $Q = g\,dx$ satisfy*

$$0 < m \le f, g \le M < \infty.$$

Then

$$\sqrt{m}\,W_2(P,Q) \le \|f-g\|_{-1} \le \sqrt{M}\,W_2(P,Q).$$

Proof. For the upper bound on W_2, use the linear interpolation $\rho_t = (1-t)f + tg$. Since $\rho_t \ge m$, the Benamou–Brenier formula and the time-independent solution of

$$\partial_t \rho_t + \nabla \cdot j = 0$$

give

$$W_2(P,Q) \le m^{-1/2}\|f-g\|_{-1}.$$

For the reverse inequality, take the displacement interpolation (ρ_t, v_t) between P and Q. The standard L^∞ displacement-convexity bound gives $\rho_t \le M$. Since

$$f - g = -\int_0^1 \partial_t \rho_t \, dt = \int_0^1 \nabla \cdot (\rho_t v_t)\, dt,$$

duality and Cauchy–Schwarz imply

$$\|f-g\|_{-1} \le \int_0^1 \|\partial_t \rho_t\|_{-1}\, dt \le \sqrt{M} \int_0^1 \left(\int |v_t|^2 \rho_t \right)^{1/2} dt = \sqrt{M}\, W_2(P,Q).$$

□

The comparison is only multiplicative. It is sufficient inside a ball of radius $O(\eta_N)$, because an $O(1)$ relative error is still an $O(\eta_N)$ absolute error.

14.2 Critical cone

Let J_N be dyadic with $h_N = 2^{-J_N} \asymp \eta_N$, and let Π_{J_N} be conditional expectation on dyadic cells. For a cell-constant density difference u, let $\widehat{Q}_{J_N}$ be the usual cross-fitted unbiased quadratic estimator of

$$\|\Pi_{J_N}(f-g)\|_{-1}^2$$

in the Neumann spectral basis up to frequency h_N^{-1}. Explicitly, if (e_k, λ_k) are Neumann eigenpairs and $0 < \lambda_k \le h_N^{-2}$,

$$u_k = \int e_k(f-g), \qquad \|\Pi_{J_N}(f-g)\|_{-1,h_N}^2 := \sum_{\lambda_k \le h_N^{-2}} \frac{u_k^2}{\lambda_k},$$

and $\widehat{Q}_{J_N}$ replaces u_k^2 by the product of two independent unbiased coefficient estimates.

The variance calculation is standard:

$$\mathbb{E}(\widehat{Q}_{J_N} - \|\Pi_{J_N}(f-g)\|_{-1,h_N}^2)^2 \lesssim \frac{\|\Pi_{J_N}(f-g)\|_{-1,h_N}^2}{N} + \frac{h_N^{4-d}}{N^2}.$$

Because $h_N^{-d} \asymp N \log N$ and $d > 4$, the second term is $o(h_N^4)$.

Theorem 14.2 (Target upper bound on the bounded-density critical cone)**.** *Fix $0 < m < 1 < M < \infty$, $L < \infty$, and $d > 4$. On*

$$\mathcal{U}_N^{\mathrm{crit}}(L) := \{(P,Q) : P = f dx,\ Q = g dx,\ m \le f, g \le M,\ W_2(P,Q) \le L\eta_N\},$$

there is an estimator $\widehat{W}_N^{\mathrm{crit}}$ satisfying

$$\sup_{\mathcal{U}_N^{\mathrm{crit}}(L)} \mathbb{E}|\widehat{W}_N^{\mathrm{crit}} - W_2(P,Q)| \lesssim_{d,m,M,L} \eta_N$$

and

$$\sup_{\mathcal{U}_N^{\mathrm{crit}}(L)} \mathbb{E}(\widehat{W}_N^{\mathrm{crit}} - W_2(P,Q))^2 \lesssim_{d,m,M,L} \eta_N^2.$$

The rate η_N is minimax sharp on this class when $L \ge 1$.

Proof. Project P, Q to dyadic histograms P^{h_N}, Q^{h_N}. By Theorem 2.1,

$$W_2(P^{h_N}, Q^{h_N}) \le (L + C_d) h_N.$$

Their densities remain in $[m, M]$. By Theorem 14.1,

$$\|\Pi_{J_N}(f-g)\|_{-1} \lesssim_{m,M,L} h_N.$$

The variance display above gives

$$\mathbb{E}|\widehat{Q}_{J_N} - \|\Pi_{J_N}(f-g)\|_{-1,h_N}^2| \lesssim h_N^2.$$

Thus $\sqrt{(\widehat{Q}_{J_N})_+}$ estimates the truncated H^{-1} norm to $O(h_N)$, and the multiplicative comparison plus quantization error converts this into $O(h_N)$ error for $W_2(P,Q)$.

For the lower bound, embed the Niles–Weed–Rigollet split-cell alternatives in the bounded-density cone: on each dyadic cell pair, put densities λ and $2-\lambda$, with $\lambda \in (m,1)$ and $2 - \lambda < M$. Their W_2-distance is $O(h_N)$, and the induced L_1-alphabet has size $N \log N$. Theorem 5.1 gives the lower bound. □

14.3 Far annulus

Let $\widehat{T}_N = W_2(P_N, Q_N)^2$. The quadratic cost is smooth, and the Manole–Niles-Weed smooth-cost theorem gives [8]

$$\sup_{P,Q \in \mathcal{P}_d} \mathbb{E}|\widehat{T}_N - W_2(P,Q)^2| \lesssim_d N^{-2/d}.$$

Changing one observation changes $\widehat{T}_N$ by at most d/N, hence

$$\mathrm{Var}(\widehat{T}_N) \lesssim_d N^{-1}.$$

Since $d > 4$, $N^{-1} = o(N^{-4/d})$, and therefore

$$\sup_{P,Q} \mathbb{E}(\widehat{T}_N - W_2(P,Q)^2)^2 \lesssim_d N^{-4/d}.$$

Theorem 14.3 (Far-annulus plug-in law)**.** *Assume $d > 4$. If $\tau_N > 0$, then*

$$\sup_{W_2(P,Q)\geq \tau_N} \mathbb{E}|W_2(P_N,Q_N) - W_2(P,Q)| \lesssim_d \frac{N^{-2/d}}{\tau_N}.$$

In particular, for every fixed $c > 0$,

$$\sup_{W_2(P,Q)\geq c\gamma_N} \mathbb{E}|W_2(P_N,Q_N) - W_2(P,Q)| \lesssim_{c,d} \eta_N,$$

and the squared risk is $O(\eta_N^2)$.

Proof. Write $T = W_2(P,Q)^2$. On $W_2(P,Q) \geq \tau_N$,

$$|W_2(P_N,Q_N) - W_2(P,Q)| \leq \frac{|\widehat{T}_N - T|}{\tau_N}.$$

Taking expectations gives the first bound. The special case follows from

$$N^{-2/d}/\gamma_N = \eta_N.$$

The squared-risk assertion is identical using the second moment estimate for $\widehat{T}_N - T$. □

Corollary 14.4 (Two-zone bounded-density theorem)**.** *For $d > 4$, fixed $0 < m < 1 < M < \infty$, $L < \infty$, and $c > 0$, the target upper bound $O(\eta_N)$ holds uniformly over all bounded-density pairs satisfying either*

$$W_2(P,Q) \leq L\eta_N$$

or

$$W_2(P,Q) \geq c\gamma_N.$$

The remaining bounded-density gap is the logarithmic strip

$$\eta_N \ll W_2(P,Q) \ll \gamma_N.$$

15 Weighted quadratic strip theorem

The flat H^{-1} norm is the wrong tangent metric away from the uniform density. The correct tangent norm at a background density ρ is generated by

$$L_\rho = -\nabla \cdot (\rho\nabla).$$

This section proves that the weighted geometry does close a genuinely non-flat part of the quadratic strip.

15.1 Smooth small-deformation class

Fix $d > 4$, $0 < m < 1 < M < \infty$, and $A < \infty$. Let $h_N \asymp \eta_N$. Define $\mathfrak{S}_N(A,m,M)$ to be the set of pairs (P,Q) such that:

(i) $P = \rho\, dx$ with $m \leq \rho \leq M$ and $\|\rho\|_{C^2} \leq A$;

(ii) the Brenier map from P to Q has the form

$$T = \mathrm{Id} + v, \qquad v = \nabla\psi,$$

with $T([0,1]^d) \subset [0,1]^d$, $I + Dv \succeq \frac{1}{2}I$, and

$$\|v\|_{C^2} \leq A\,t, \qquad t := W_2(P,Q);$$

(iii) the distance lies in the intermediate strip

$$h_N \leq t \leq \gamma_N.$$

The condition $\|v\|_{C^2} \lesssim t$ excludes oscillatory microscopic rearrangements; those are already handled by the critical-cone theorem at scale h_N, and they are the source of the hard unrestricted critical grid.

15.2 Weighted tangent expansion

Lemma 15.1 (Weighted tangent expansion)**.** *For $(P,Q) \in \mathfrak{S}_N(A,m,M)$, let $Q = q\, dx$ and $u = \rho - q$. Then*

$$\left| \|u\|_{-1,\rho}^2 - W_2(P,Q)^2 \right| \leq C_{A,m,M,d}\, W_2(P,Q)^3.$$

Proof. Put $t = W_2(P,Q)$ and $v = \nabla\psi$. Since $T = \mathrm{Id} + v$ is the Brenier map,

$$t^2 = \int |v|^2 \rho\, dx.$$

For $s \in [0,1]$, let

$$T_s = \mathrm{Id} + sv, \qquad \rho_s = (T_s)_\#(\rho dx),$$

and let $w_s(T_s x) = v(x)$. The curve (ρ_s, w_s) satisfies the continuity equation

$$\partial_s \rho_s + \nabla \cdot (\rho_s w_s) = 0.$$

Consequently,

$$u = \rho - q = \nabla \cdot \int_0^1 \rho_s w_s\, ds.$$

Let $J_s = \rho_s w_s$. In Eulerian variables,

$$J_s(y) = \frac{\rho(x_s(y))v(x_s(y))}{\det(I + sDv(x_s(y)))}, \qquad x_s = T_s^{-1}(y).$$

The assumptions $\|v\|_{C^2} \le At$, $m \le \rho \le M$, and $\|\rho\|_{C^2} \le A$ imply, for all sufficiently large N,

$$\|J_s - \rho v\|_{L^2} \le C_{A,m,M,d} t^2 \qquad (0 \le s \le 1).$$

Indeed $x_s(y) - y = O(t)$, $v(x_s(y)) - v(y) = O(t^2)$, $\rho(x_s(y)) - \rho(y) = O(t)$, and $\det(I + sDv) = 1 + O(t)$, while $\|v\|_{L^2} \le M^{-1/2}t$. Hence

$$u = \nabla \cdot (\rho v) + r = -L_\rho \psi + r, \qquad \|r\|_{-1,\rho} \le C_{A,m,M,d} t^2.$$

The last bound follows from the representation $r = \nabla \cdot F$ with $\|F\|_{L^2} \le Ct^2$ and the dual definition of $\|\cdot\|_{-1,\rho}$.

Since

$$\|-L_\rho \psi\|_{-1,\rho}^2 = \int \rho |\nabla \psi|^2 = t^2,$$

we obtain

$$\left| \|u\|_{-1,\rho}^2 - t^2 \right| \le 2t\, \|r\|_{-1,\rho} + \|r\|_{-1,\rho}^2 \le Ct^3 + Ct^4.$$

The diameter bound $t \le \sqrt{d}$ absorbs the last term. ☐

15.3 Spectral weighted estimator

Let $(e_k, \lambda_k)_{k \ge 0}$ be the Neumann eigenbasis on $[0,1]^d$, with $e_0 \equiv 1$, $\lambda_0 = 0$, and $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots$. Put

$$\Lambda_N := h_N^{-1}, \qquad I_N := \{k \ge 1 : \lambda_k \le \Lambda_N^2\}, \qquad D_N := |I_N| \asymp h_N^{-d} \asymp N \log N.$$

For a weight a satisfying $m/2 \le a \le 2M$, define

$$A_N(a)_{k\ell} := \int a \,\nabla e_k \cdot \nabla e_\ell \, dx, \qquad k, \ell \in I_N.$$

For $u = \rho - q$, set $u_N = (\int u e_k)_{k \in I_N}$ and

$$E_N(u, a) := u_N^\top A_N(a)^{-1} u_N.$$

The estimator uses three independent data blocks. A pilot block from the P-sample constructs a clipped density estimate $\widehat{\rho}$ satisfying

$$m/2 \le \widehat{\rho} \le 2M.$$

Two further independent blocks estimate u_N:

$$\widehat{u}_N^{(r)}(k) = \frac{1}{n_r} \sum_{i \in B_r^X} e_k(X_i) - \frac{1}{n_r} \sum_{i \in B_r^Y} e_k(Y_i), \qquad r = 1, 2.$$

Then define

$$\widehat{E}_N := \big(\widehat{u}_N^{(1)}\big)^\top A_N(\widehat{\rho})^{-1} \widehat{u}_N^{(2)}, \qquad \widehat{W}_N^{\text{wgt}} := \sqrt{(\widehat{E}_N)_+}.$$

Lemma 15.2 (Weighted quadratic variance)**.** *Conditionally on the pilot $\widehat{\rho}$,*

$$\mathbb{E}\left[\Big(\widehat{E}_N - E_N(u, \widehat{\rho}) \Big)^2 \,\middle|\, \widehat{\rho} \right] \le C_{d,m,M} \left(\frac{E_N(u,\widehat{\rho})}{N} + \frac{h_N^{4-d}}{N^2} \right).$$

Proof. Write $B = A_N(\widehat{\rho})^{-1}$. Conditional on the pilot,

$$\widehat{u}_N^{(r)} = u_N + \xi_r, \qquad r = 1, 2,$$

where ξ_1, ξ_2 are independent mean-zero vectors. For every vector $z = (z_k)$,

$$\operatorname{Var}(z^\top \xi_r) \le \frac{C_M}{N} \left\| \sum_{k \in I_N} z_k e_k \right\|_{L^2}^2 = \frac{C_M}{N} \|z\|_2^2.$$

Since $m/2 \le \widehat{\rho} \le 2M$,

$$A_N(\widehat{\rho}) \succeq \frac{m}{2} \operatorname{diag}(\lambda_k), \qquad \|B\|_{\text{op}} \le C_{m,d},$$

and

$$\operatorname{Tr}(B^2) \le C_{m,d} \sum_{\lambda_k \le h_N^{-2}} \lambda_k^{-2} \lesssim_{d,m} h_N^{4-d}$$

by Weyl's law, because $d > 4$.

Expanding

$$\widehat{E}_N - E_N(u, \widehat{\rho}) = u_N^\top B \xi_2 + \xi_1^\top B u_N + \xi_1^\top B \xi_2$$

and using independence gives

$$\mathbb{E}[(\widehat{E}_N - E_N)^2 | \widehat{\rho}] \le \frac{C}{N} u_N^\top B^2 u_N + \frac{C}{N^2} \operatorname{Tr}(B^2).$$

Since $\|B\|_{\text{op}} \le C$,

$$u_N^\top B^2 u_N \le C\, u_N^\top B u_N = C E_N(u, \widehat{\rho}).$$

The claim follows. ☐

Lemma 15.3 (Pilot stability)**.** *If $a, b \in [m/2, 2M]$, then*

$$|E_N(u,a) - E_N(u,b)| \le C_{m,M} \|a - b\|_\infty \, E_N(u,a).$$

Proof. The quadratic forms satisfy

$$|A_N(a) - A_N(b)| \le \|a-b\|_\infty \int |\nabla\phi|^2$$

on every vector $\phi \in \operatorname{span}\{e_k : k \in I_N\}$. Since $A_N(a) \succeq (m/2)\int |\nabla\phi|^2$, the forms $A_N(a)$ and $A_N(b)$ are relatively within $C_m\|a-b\|_\infty$. The same is true for their inverses by the order-reversing property of positive matrices, giving the asserted bound. ☐

Lemma 15.4 (Galerkin truncation)**.** *For* $(P,Q) \in \mathfrak{S}_N(A,m,M)$*,*

$$\left|E_N(u,\rho) - \|u\|_{-1,\rho}^2\right| \le C_{A,m,M,d}\, W_2(P,Q)^2 h_N^2.$$

Proof. Let $\phi \in H^1/\mathbb{R}$ solve

$$L_\rho\phi = u, \qquad \int \phi = 0.$$

Then

$$\|u\|_{-1,\rho}^2 = \int \rho|\nabla\phi|^2.$$

The Galerkin value $E_N(u,\rho)$ is the same variational problem restricted to

$$V_N = \operatorname{span}\{e_k : \lambda_k \le h_N^{-2},\ k \ge 1\}.$$

By Galerkin orthogonality,

$$0 \le \|u\|_{-1,\rho}^2 - E_N(u,\rho) \le C_{m,M} \inf_{\psi\in V_N} \|\phi-\psi\|_{H^1}^2.$$

Neumann elliptic regularity for the uniformly elliptic operator L_ρ, with $\rho \in C^2$ and $m \le \rho \le M$, gives

$$\|\phi\|_{H^2} \le C_{A,m,M,d}\|u\|_{L^2}.$$

Taking ψ to be the Neumann spectral projection of ϕ onto V_N,

$$\inf_{\psi\in V_N} \|\phi-\psi\|_{H^1}^2 \le Ch_N^2\|\phi\|_{H^2}^2 \le Ch_N^2\|u\|_{L^2}^2.$$

Finally the smooth-deformation expansion in the proof of Theorem 15.1 yields $\|u\|_{L^2} \le C_{A,m,M,d}W_2(P,Q)$. The claim follows. ☐

Lemma 15.5 (Pilot availability)**.** *If ρ ranges over a fixed C^2-ball and $m \le \rho \le M$, then there is a clipped pilot estimator $\widehat{\rho}$, constructed from an independent block of P-samples, such that*

$$\sup_\rho \mathbb{P}\left(\|\widehat{\rho}-\rho\|_\infty > c_d \frac{h_N}{\gamma_N}\right) \le N^{-10}$$

for all sufficiently large N.

Proof. Use a boundary-corrected kernel or wavelet projection estimator with bandwidth $b_N = N^{-1/(4+d)}$. The standard sup-norm bound over a fixed C^2-ball is

$$\|\widehat{\rho}-\rho\|_\infty = O_{\mathbb{P}}\left(b_N^2 + \sqrt{\frac{\log N}{Nb_N^d}}\right) = O_{\mathbb{P}}\left((\log N)^{1/2}N^{-2/(4+d)}\right).$$

This is $o((\log N)^{-2/d})$. Since

$$\frac{h_N}{\gamma_N} = (\log N)^{-2/d},$$

the displayed high-probability statement follows after increasing the constant and clipping to $[m/2, 2M]$. ☐

Theorem 15.6 (Weighted strip upper theorem)**.** *Assume $d > 4$. On the smooth small-deformation strip $\mathfrak{S}_N(A,m,M)$,*

$$\sup_{(P,Q)\in\mathfrak{S}_N(A,m,M)} \mathbb{E}|\widehat{W}_N^{\mathrm{wgt}} - W_2(P,Q)| \le C_{A,m,M,d}\, h_N,$$

and

$$\sup_{(P,Q)\in\mathfrak{S}_N(A,m,M)} \mathbb{E}(\widehat{W}_N^{\mathrm{wgt}} - W_2(P,Q))^2 \le C_{A,m,M,d}\, h_N^2.$$

Proof. Let $t = W_2(P,Q)$, so $h_N \le t \le \gamma_N$. Work on the pilot event

$$\|\widehat{\rho}-\rho\|_\infty \le Ch_N/\gamma_N.$$

The complement has probability $O(N^{-10})$ and contributes negligibly because all distances are bounded by $\sqrt{d}$.

By Theorems 15.1, 15.3 and 15.4,

$$\begin{aligned}|E_N(u,\widehat{\rho}) - t^2| &\le |E_N(u,\widehat{\rho}) - E_N(u,\rho)| + |E_N(u,\rho) - \|u\|_{-1,\rho}^2| + |\|u\|_{-1,\rho}^2 - t^2| \\ &\le C\frac{h_N}{\gamma_N}t^2 + Ct^2h_N^2 + Ct^3.\end{aligned}$$

Since $t \le \gamma_N$ and $\gamma_N^2 = o(h_N)$,

$$\frac{h_N}{\gamma_N}t^2 \le h_N t, \qquad t^2h_N^2 \le h_N t, \qquad t^3 \le \gamma_N^2 t = o(h_N t).$$

Therefore

$$|E_N(u,\widehat{\rho}) - t^2| \le Ch_N t.$$

Next, Theorem 15.2 and $E_N(u,\widehat{\rho}) \lesssim t^2$ yield

$$\mathbb{E}\left[|\widehat{E}_N - E_N(u,\widehat{\rho})| \,\middle|\, \widehat{\rho}\right] \le C\left(\frac{t}{\sqrt{N}} + \frac{h_N^{2-d/2}}{N}\right).$$

Because $h_N\sqrt{N}\to\infty$ for $d>2$,

$$\frac{t}{\sqrt{N}}\le Cth_N.$$

Also $h_N^{-d}\asymp N\log N$, so

$$\frac{h_N^{2-d/2}}{N}=h_N^2\sqrt{\frac{\log N}{N}}=o(h_N^2)\le o(h_N t),$$

as $t\ge h_N$. Thus

$$\mathbb{E}|\widehat{E}_N-t^2|\le Ch_N t.$$

Finally,

$$|\sqrt{(\widehat{E}_N)_+}-t|\le\frac{|\widehat{E}_N-t^2|}{t},$$

so the absolute risk is $O(h_N)$. For squared risk we use the same denominator and the second-moment version of the preceding estimate. The deterministic part contributes $Ch_N^2t^2$. The stochastic part is bounded by

$$C\left(\frac{t^2}{N}+\frac{h_N^{4-d}}{N^2}\right).$$

Dividing by t^2, and using $t\ge h_N$, gives

$$\mathbb{E}(\sqrt{(\widehat{E}_N)_+}-t)^2\le C\left(h_N^2+\frac{1}{N}+\frac{h_N^{4-d}}{N^2h_N^2}\right).$$

Because $h_N^{-d}\asymp N\log N$ and $d>4$,

$$\frac{1}{N}=o(h_N^2),\qquad\frac{h_N^{4-d}}{N^2h_N^2}=h_N^2\,\frac{\log N}{N}=o(h_N^2).$$

Thus the squared risk is $O(h_N^2)$. ☐

Corollary 15.7 (Smooth non-flat quadratic three-zone law)**.** *Let $d>4$. On bounded-density pairs satisfying either*

$$W_2(P,Q)\le L\eta_N,$$

or

$$(P,Q)\in\mathfrak{S}_N(A,m,M),$$

or

$$W_2(P,Q)\ge c\gamma_N,$$

there is an estimator with absolute risk $O(\eta_N)$ and squared risk $O(\eta_N^2)$.

Proof. Use the critical-cone estimator in the first region, the weighted estimator in the smooth strip, and the empirical plug-in estimator in the far annulus. A standard sample-splitting threshold using $\widehat{T}_N=W_2(P_N,Q_N)^2$ separates $W_2\lesssim\eta_N$ from $W_2\gtrsim\gamma_N$ with negligible error; in the overlap one may take the minimum of the three risks. The bounds are those of Theorems 14.2, 14.3 and 15.6. ☐

16 Why the unweighted strip method is false

The new weighted theorem is not cosmetic. The unweighted critical H^{-1} energy has the wrong tangent metric at non-uniform densities.

Proposition 16.1 (Failure of unweighted H^{-1} in the strip)**.** *Assume $d>4$. Fix $0<m<1<M<\infty$, and let $h_N\asymp\eta_N$. There are smooth densities f_N,g_N with*

$$m\le f_N,g_N\le M,\qquad\eta_N\le W_2(f_Ndx,g_Ndx)\le\gamma_N,$$

such that

$$\frac{\left|\|\Pi_{h_N}(f_N-g_N)\|_{-1}^2-W_2(f_Ndx,g_Ndx)^2\right|}{W_2(f_Ndx,g_Ndx)\,h_N}\longrightarrow\infty.$$

Proof. Take

$$\rho_a(x)=1+a\cos(\pi x_1),\qquad\varphi(x)=\cos(\pi x_1),\qquad v=\nabla\varphi,$$

with $a>0$ small enough that $\rho_a\in[m,M]$. Let

$$T_\varepsilon=\mathrm{Id}+\varepsilon v,\qquad Q_\varepsilon=(T_\varepsilon)_\#(\rho_a dx).$$

For small ε, T_ε is the Brenier map and

$$W_2(\rho_a dx,Q_\varepsilon)^2=\varepsilon^2\int|v|^2\rho_a dx=\varepsilon^2\frac{\pi^2}{2}.$$

The density q_ε of Q_ε satisfies

$$\rho_a-q_\varepsilon=\varepsilon\big[-\nabla\cdot(\rho_a\nabla\varphi)\big]+O(\varepsilon^2)$$

in C^1, and

$$-\nabla\cdot(\rho_a\nabla\varphi)=\pi^2\cos(\pi x_1)+a\pi^2\cos(2\pi x_1).$$

Therefore the unweighted H^{-1} energy is

$$\|\rho_a-q_\varepsilon\|_{-1}^2=\varepsilon^2\left(\frac{\pi^2}{2}+\frac{a^2\pi^2}{8}\right)+o(\varepsilon^2).$$

The excess $a^2\pi^2\varepsilon^2/8$ is precisely the metric mismatch. Choose

$$\varepsilon_N\asymp\gamma_N.$$

Then $W_2\asymp\gamma_N$, while the deterministic discrepancy divided by W_2h_N is of order

$$\frac{\gamma_N}{h_N}=(\log N)^{2/d}\to\infty.$$

Dyadic projection at mesh h_N changes the smooth H^{-1} energy by $o(\varepsilon_N^2)$, so the same conclusion holds for $\Pi_{h_N}(f_N - g_N)$. □

17 Finite critical-grid LP: active faces and debiasing criteria

The exact skeleton results solve many critical Euclidean subclasses because their LP value is already a weighted L_1 functional. The full grid problem is harder because the active Kantorovich face may change across a large and highly degenerate family. This section records three finite principles which are independent of continuum regularity. Each is proved directly at the level of the critical grid.

17.1 The normalized dual polytope

Let $G \subset [0,1]^d$ be finite and let $c(x,y) = \|x-y\|_2^p$. For $r, s \in \mathcal{P}(G)$,

$$T(r,s) := W_p(r,s)^p = \min_{\pi \in \Pi(r,s)} \sum_{x,y \in G} c(x,y)\pi_{xy}.$$

The dual form is

$$T(r,s) = \sup_{(\phi,\psi)\in\mathcal{D}_G} \left\{ \sum_{x\in G} \phi_x r_x + \sum_{y \in G} \psi_y s_y \right\},$$

where

$$\mathcal{D}_G := \{(\phi,\psi) \in \mathbb{R}^G \times \mathbb{R}^G : \ \phi_x + \psi_y \le c(x,y)\ \forall x,y \in G\}.$$

Since $(\phi + a\mathbf{1}, \psi - a\mathbf{1})$ gives the same objective, we fix a root $x_0 \in G$ and normalize by $\phi_{x_0} = 0$. Let

$$\mathcal{D}_G^0 := \{(\phi,\psi) \in \mathcal{D}_G : \phi_{x_0} = 0,\ \|\phi\|_\infty + \|\psi\|_\infty \le 4d^{p/2}\}.$$

Lemma 17.1 (Bounded normalized dual potentials)**.** *For every $r, s \in \mathcal{P}(G)$,*

$$T(r,s) = \sup_{(\phi,\psi)\in\mathcal{D}_G^0} \{\langle \phi, r\rangle + \langle \psi, s\rangle\}.$$

Proof. Let (ϕ,ψ) be any feasible dual pair. Replacing (ϕ,ψ) by $(\phi - \phi_{x_0}\mathbf{1}, \psi + \phi_{x_0}\mathbf{1})$, we may suppose $\phi_{x_0} = 0$. Feasibility gives

$$\psi_y \le c(x_0,y) \le d^{p/2}.$$

For any x, y,

$$\phi_x \le c(x,y) - \psi_y.$$

Taking $y = x_0$ and using $\psi_{x_0} \le d^{p/2}$ bounds ϕ_x above by $2d^{p/2}$ after replacing ψ by the largest feasible c-transform associated to ϕ. The same c-transform normalization gives

$$\psi_y = \inf_x \{c(x,y) - \phi_x\}.$$

Since $c \in [0, d^{p/2}]$, the oscillations of ϕ and ψ are bounded by $d^{p/2}$. With $\phi_{x_0} = 0$, both sup norms are bounded by a universal multiple of $d^{p/2}$. Truncating to the displayed compact set therefore does not change the dual value. □

17.2 Finite dual catalogs

Suppose $X_1, \dots, X_N \sim r$ and $Y_1, \dots, Y_N \sim s$. For a dual pair $a = (\phi,\psi) \in \mathcal{D}_G^0$, write

$$L_a(r,s) := \langle \phi, r\rangle + \langle \psi, s\rangle, \qquad \widehat{L}_a := \frac{1}{N}\sum_{i=1}^N \phi(X_i) + \frac{1}{N}\sum_{i=1}^N \psi(Y_i).$$

Theorem 17.2 (Finite dual-catalog theorem)**.** *Let $G_N \subset [0,1]^d$ be a grid of mesh $h_N \asymp (N\log N)^{-1/d}$. Let $\mathcal{A}_N \subset \mathcal{D}_{G_N}^0$ be a finite catalog and let $\varepsilon_N > 0$. Assume that for every (r,s) in a class $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$,*

$$0 \le T(r,s) - \max_{a\in\mathcal{A}_N} L_a(r,s) \le \varepsilon_N.$$

Then the estimator

$$\widehat{T}_N^{\mathcal{A}} := \max_{a \in \mathcal{A}_N} \widehat{L}_a$$

satisfies

$$\sup_{(r,s)\in\mathcal{C}_N} \mathbb{E}|\widehat{T}_N^{\mathcal{A}} - T(r,s)| \le \varepsilon_N + C_{d,p}\sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}},$$

and

$$\sup_{(r,s)\in\mathcal{C}_N} \mathbb{E}(\widehat{T}_N^{\mathcal{A}} - T(r,s))^2 \le C_{d,p}\left[\varepsilon_N^2 + \frac{\log(2|\mathcal{A}_N|)}{N}\right].$$

Consequently, if

$$\varepsilon_N \lesssim h_N^p, \qquad \log|\mathcal{A}_N| \lesssim N h_N^{2p},$$

then W_p^p is estimable on $\mathcal{C}_N$ to powered accuracy $O(h_N^p)$. On any annulus $W_p(r,s) \ge \tau h_N$, the distance W_p itself is estimable to accuracy $O_{\tau,p}(h_N)$.

Proof. For every (r,s),

$$\left| \max_{a\in\mathcal{A}_N} \widehat{L}_a - \max_{a\in\mathcal{A}_N} L_a(r,s) \right| \le \sup_{a\in\mathcal{A}_N} |\widehat{L}_a - L_a(r,s)|.$$

By Theorem 17.1, the summands in $\widehat{L}_a - L_a$ are uniformly bounded by $C_{d,p}$. Symmetrization and Massart's finite-class inequality give

$$\mathbb{E} \sup_{a\in\mathcal{A}_N} |\widehat{L}_a - L_a(r,s)| \le C_{d,p}\sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}}.$$

The same inequality with bounded differences, or integration of the corresponding sub-Gaussian tail, gives

$$\mathbb{E} \sup_{a\in\mathcal{A}_N} |\widehat{L}_a - L_a(r,s)|^2 \le C_{d,p}\frac{\log(2|\mathcal{A}_N|)}{N}.$$

Combining these bounds with the deterministic catalog approximation gives the absolute and squared-risk estimates for T. If $T = W_p^p$ and $W_p(r,s) \ge \tau h_N$, then

$$|u^{1/p} - v^{1/p}| \le C_{\tau,p} h_N^{1-p}|u-v|$$

whenever u, v lie in a fixed bounded interval and $v \ge (\tau h_N)^p$. Thus powered error $O(h_N^p)$ gives distance error $O(h_N)$. ☐

Corollary 17.3 (Active-face margin regime)**.** *In the setting of Theorem 17.2, suppose the catalog is exact on $\mathcal{C}_N$,*

$$T(r,s) = \max_{a\in\mathcal{A}_N} L_a(r,s),$$

and for every $(r,s) \in \mathcal{C}_N$ there exists an active $a_\star(r,s) \in \mathcal{A}_N$ such that

$$L_{a_\star}(r,s) - \max_{a\ne a_\star} L_a(r,s) \ge 4C_{d,p}\sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}}.$$

Then the empirical maximizer over $\mathcal{A}_N$ selects an active face with probability at least $1 - \exp[-c\log(2|\mathcal{A}_N|)]$, and the resulting value estimator has absolute risk

$$O_{d,p}\left(\sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}}\right).$$

In particular, it attains the critical powered target whenever $\log|\mathcal{A}_N| \lesssim Nh_N^{2p}$.

Proof. The same finite-class concentration used in Theorem 17.2 gives

$$\sup_{a\in\mathcal{A}_N} |\widehat{L}_a - L_a(r,s)| \le 2C_{d,p}\sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}}$$

with the stated probability after increasing the constants. On this event the displayed margin prevents any inactive catalog element from overtaking $a_\star$. The value bound is then the fixed-potential empirical mean bound, and the expectation follows by adding the negligible failure contribution. ☐

17.3 A Richardson debiasing criterion for the full grid

The full grid cannot be solved by a small catalog unless the optimal dual geometry is compressible. A different route is to regularize the LP value and cancel the deterministic regularization bias. The next theorem is purely finite-dimensional and identifies the exact quantitative assumptions needed for a critical-grid proof.

Let $T_\lambda(r,s)$ be a family of computable regularized LP values, indexed by $\lambda > 0$, and let $\widehat{T}_\lambda$ denote its plug-in or debiased empirical version based on the two samples.

Theorem 17.4 (Richardson criterion for critical-grid debiasing)**.** *Assume $d > 2p$, $h_N = (N\log N)^{-1/d}$, and G_N is the critical grid. Suppose that for all $r, s \in \mathcal{P}(G_N)$ and all sufficiently small λ the following two estimates hold with constants independent of N, r, s:*

$$T_\lambda(r,s) = T(r,s) + b_1(r,s)\lambda + R_\lambda(r,s), \qquad |R_\lambda(r,s)| \le A\lambda^2,$$

and

$$\mathbb{E}|\widehat{T}_\lambda - T_\lambda(r,s)| \le B\left(N^{-1/2} + \frac{1}{N\lambda}\right).$$

Define

$$\widehat{T}_N^{\mathrm{Rich}} := 2\widehat{T}_{\lambda_N/2} - \widehat{T}_{\lambda_N}, \qquad \lambda_N := h_N^{p/2}.$$

Then

$$\sup_{r,s\in\mathcal{P}(G_N)} \mathbb{E}|\widehat{T}_N^{\mathrm{Rich}} - T(r,s)| \lesssim_{A,B,d,p} h_N^p.$$

Consequently the critical-grid LP powered value is estimable at the target scale. On annuli $W_p(r,s) \ge \tau h_N$, this gives W_p-risk $O_{\tau,d,p}(h_N)$.

Proof. By the assumed expansion,

$$2T_{\lambda/2} - T_\lambda - T = 2R_{\lambda/2} - R_\lambda,$$

hence

$$|2T_{\lambda/2} - T_\lambda - T| \le \frac{A}{2}\lambda^2 + A\lambda^2 \lesssim_A \lambda^2.$$

The stochastic part obeys

$$\begin{aligned}\mathbb{E}|\widehat{T}_N^{\mathrm{Rich}} - (2T_{\lambda_N/2} - T_{\lambda_N})| &\le 2\mathbb{E}|\widehat{T}_{\lambda_N/2} - T_{\lambda_N/2}| + \mathbb{E}|\widehat{T}_{\lambda_N} - T_{\lambda_N}| \\ &\le C_B\left(N^{-1/2} + \frac{1}{N\lambda_N}\right).\end{aligned}$$

With $\lambda_N = h_N^{p/2}$, the deterministic term is $\lambda_N^2 = h_N^p$. Since $d > 2p$,

$$N^{-1/2} = o(h_N^p),$$

and

$$\frac{1}{N\lambda_N} = \frac{1}{Nh_N^{p/2}} = o(h_N^p),$$

because $Nh_N^{3p/2} \to \infty$, which follows from $d > 2p$. Therefore the total powered risk is $O(h_N^p)$. The annular conversion from powered risk to distance risk is the same mean-value argument as in Theorem 17.2. □

Remark 17.5 (Connection with recent debiased LP theory). Recent fixed-alphabet work on debiased Gaussian estimators for discrete optimal transport and general linear programs constructs two-tuning-parameter cancellations of regularization bias and proves centered Gaussian limits for random LP solutions [7]. Theorem 17.4 is not a restatement of those fixed-dimensional results; it is the growing-alphabet target theorem that their method would have to satisfy uniformly when $|G_N| \asymp N \log N$. The theorem shows that a first-order regularization expansion plus the displayed sampling bound is already enough to close the supercritical minimax upper bound.

18 Localized dual geometry and contact certificates

The preceding finite-catalog theorem uses the size of a single global catalog. This is too crude for the full Euclidean grid: the global semidual class at mesh h has entropy of order h^{-d}, whereas the critical statistical budget is only

$$Nh^{2p} \asymp N^{1-2p/d}(\log N)^{-2p/d}.$$

A genuinely new route must use the fact that most dual potentials are not competitive for a fixed pair (r, s). This section makes that principle quantitative. It gives a proved localized replacement for the global catalog criterion and exact contact-defect identities which turn near-optimality of a dual potential into a geometric restriction along an optimal plan.

18.1 Gap shells for support-function estimators

Let $G \subset [0, 1]^d$ be finite and let $\mathcal{A} \subset \mathcal{D}_G^0$ be a finite dual catalog. For $a = (\phi, \psi) \in \mathcal{A}$, write

$$L_a(r, s) = \langle \phi, r\rangle + \langle \psi, s\rangle, \qquad \widehat{L}_a = \frac{1}{N}\sum_{i=1}^N \phi(X_i) + \frac{1}{N}\sum_{i=1}^N \psi(Y_i).$$

Assume throughout this subsection that the catalog is exact on a class $\mathcal{C}$:

$$T(r, s) := W_p(r, s)^p = \max_{a\in\mathcal{A}} L_a(r, s) \qquad ((r, s) \in \mathcal{C}).$$

For $(r, s) \in \mathcal{C}$ define the deterministic gap

$$\Delta_a(r, s) := T(r, s) - L_a(r, s) \geq 0.$$

For a target powered accuracy $\varepsilon > 0$, set

$$\mathcal{A}_{-1}(r, s; \varepsilon) := \{a \in \mathcal{A} : \Delta_a(r, s) \leq \varepsilon\},$$

and, for $j \geq 0$,

$$\mathcal{A}_j(r, s; \varepsilon) := \{a \in \mathcal{A} : 2^j\varepsilon < \Delta_a(r, s) \leq 2^{j+1}\varepsilon\}.$$

The next theorem says that the empirical maximum needs only controlled entropy of the near-optimal gap shells, not controlled entropy of the entire catalog.

Theorem 18.1 (Gap-peeling dual theorem). *Let $B < \infty$ and suppose*

$$\|\phi\|_\infty + \|\psi\|_\infty \leq B \qquad (a = (\phi, \psi) \in \mathcal{A}).$$

There exist constants $c_0, C_0 \in (0, \infty)$, depending only on B, with the following property. If for every $(r, s) \in \mathcal{C}$

$$\log\bigl(2|\mathcal{A}_{-1}(r, s; \varepsilon)|\bigr) \leq c_0 N\varepsilon^2$$

and, for every $j \geq 0$ with $\mathcal{A}_j(r, s; \varepsilon) \neq \varnothing$,

$$\log\bigl(2|\mathcal{A}_j(r, s; \varepsilon)|\bigr) \leq c_0\, 2^{2j} N\varepsilon^2,$$

then the estimator

$$\widehat{T}_{\mathcal{A}} := \max_{a\in\mathcal{A}} \widehat{L}_a$$

satisfies

$$\sup_{(r,s)\in\mathcal{C}} \mathbb{E}|\widehat{T}_{\mathcal{A}} - T(r, s)| \leq C_0\varepsilon.$$

The same assumptions give

$$\sup_{(r,s)\in\mathcal{C}} \mathbb{E}(\widehat{T}_{\mathcal{A}} - T(r, s))^2 \leq C_0\varepsilon^2.$$

Proof. Fix $(r, s) \in \mathcal{C}$ and abbreviate $\Delta_a = \Delta_a(r, s)$. Let

$$Z_a := \widehat{L}_a - L_a(r, s).$$

For every a, Z_a is a centered average of two bounded empirical means, and Hoeffding's inequality gives

$$\mathbb{P}\{|Z_a| > u\} \leq 2\exp\left(-\frac{Nu^2}{C_B}\right)$$

with $C_B < \infty$ depending only on B.

The lower deviation is localized on the active shell. Indeed, choose $a_\star \in \mathcal{A}$ with $\Delta_{a_\star} = 0$. Then

$$T - \widehat{T}_{\mathcal{A}} \le T - \widehat{L}_{a_\star} = -Z_{a_\star} \le \sup_{a \in \mathcal{A}_{-1}} |Z_a|.$$

By the finite-class maximal inequality and the first entropy assumption,

$$\mathbb{E}(T - \widehat{T}_{\mathcal{A}})_+ \le C_B \sqrt{\frac{\log(2|\mathcal{A}_{-1}|)}{N}} \le C\varepsilon.$$

The second moment is bounded in the same way.

For the upper deviation,

$$\widehat{T}_{\mathcal{A}} - T = \max_{a \in \mathcal{A}} \{Z_a - \Delta_a\}.$$

On $\mathcal{A}_{-1}$ this is at most

$$\varepsilon + \sup_{a \in \mathcal{A}_{-1}} |Z_a|.$$

On $\mathcal{A}_j$, $j \ge 0$, it is at most

$$\sup_{a \in \mathcal{A}_j} Z_a - 2^j \varepsilon.$$

By the entropy assumption and the choice of $c_0(B)$, the expected supremum on $\mathcal{A}_j$ is at most $2^{j-2}\varepsilon$. The sub-Gaussian tail of the supremum gives

$$\mathbb{E}\left(\sup_{a \in \mathcal{A}_j} Z_a - 2^j \varepsilon\right)_+ \le C_B\, 2^j \varepsilon \exp(-c_B\, 2^{2j} N\varepsilon^2).$$

The number of non-empty shells is at most $O(\log(B/\varepsilon))$, because all dual values are bounded. In the supercritical applications $N\varepsilon^2 = Nh_N^{2p} \to \infty$. Enlarging C_0 to cover the finitely many small N, summing over $j \ge 0$ yields

$$\mathbb{E}(\widehat{T}_{\mathcal{A}} - T)_+ \le C\varepsilon.$$

The squared bound follows by integrating the same tail estimates and using that all values are uniformly bounded by B. Combining upper and lower deviations proves the theorem. □

Remark 18.2 (Why this is stronger than a global catalog). The global theorem $\log|\mathcal{A}| \lesssim N\varepsilon^2$ is the special case in which every shell is bounded by the critical budget. Theorem 18.1 allows exponentially more potentials at large deterministic gap, because a potential which loses by $2^j\varepsilon$ needs a $2^j\varepsilon$-sized statistical fluctuation to matter. This is exactly the statistical mechanism missing from a global entropy attack on the full Lipschitz or semiconcave dual class.

18.2 The contact-defect identity

The preceding theorem is useful only if one can control the entropy of near-optimal dual shells. The next identity converts this question into transport geometry. Let

$$\mathrm{slack}_a(x, y) := c(x, y) - \phi_x - \psi_y \qquad (a = (\phi, \psi)).$$

For $a \in \mathcal{D}^0_G$ this slack is nonnegative on $G \times G$.

Lemma 18.3 (Contact-defect identity)**.** *Let $r, s \in \mathcal{P}(G)$, let $\pi_\star$ be any optimal coupling for $T(r, s)$, and let $a = (\phi, \psi) \in \mathcal{D}^0_G$. Then*

$$T(r, s) - L_a(r, s) = \sum_{x,y \in G} \big(c(x, y) - \phi_x - \psi_y\big)\pi_\star(x, y).$$

Consequently, if a is δ-optimal, then for every $\tau > 0$

$$\pi_\star\{(x, y) : \mathrm{slack}_a(x, y) > \tau\} \le \frac{\delta}{\tau}.$$

Proof. Because $\pi_\star$ has marginals r and s,

$$L_a(r, s) = \sum_{x,y} (\phi_x + \psi_y)\pi_\star(x, y).$$

Since $\pi_\star$ is optimal,

$$T(r, s) = \sum_{x,y} c(x, y)\pi_\star(x, y).$$

Subtracting gives the identity. The Markov estimate follows from nonnegativity of the slack. □

The identity says that a near-active potential must almost satisfy complementary slackness on most of the optimal transport graph. Thus a proof of the full critical-grid theorem can be sought through the following finite geometric statement: for every optimal plan on the Euclidean grid, the class of c-concave potentials with small average contact defect has gap-shell entropy at most $O(2^{2j} Nh_N^{2p})$ at gap $2^j h_N^p$.

18.3 Finite contact certificates

We isolate a certificate version which is fully proved and directly usable. For $K \in \mathbb{N}$, let $\mathfrak{Z} = \{(x_\ell, y_\ell)\}_{\ell=1}^K \subset G \times G$. Given $\tau > 0$, say that $\mathfrak{Z}$ (τ, ε)-*determines* a dual family $\mathcal{U} \subset \mathcal{D}^0_G$ at (r, s) if whenever $a, b \in \mathcal{U}$ satisfy

$$|\phi_a(x_\ell) - \phi_b(x_\ell)| + |\psi_a(y_\ell) - \psi_b(y_\ell)| \le \tau \qquad (1 \le \ell \le K),$$

then

$$|L_a(r, s) - L_b(r, s)| \le \varepsilon.$$

This definition is deliberately operational: it asks that the objective value of every near-active dual potential be controlled by its values on finitely many contact points.

Proposition 18.4 (Contact-certificate catalog). *Let $\mathcal{U}(r,s) \subset \mathcal{D}_G^0$ be a family of dual potentials which contains at least one optimizer. Suppose that there is a data-independent collection of possible certificate sets*

$$\mathfrak{Z}_\alpha = \{(x_{\alpha,\ell}, y_{\alpha,\ell})\}_{\ell=1}^{K_N}, \qquad \alpha \in I_N,$$

such that for every $(r,s) \in \mathcal{C}$ at least one $\mathfrak{Z}_\alpha$ (τ_N, ε_N)-determines $\mathcal{U}(r,s)$ at (r,s). Assume also that all potentials in $\mathcal{U}(r,s)$ are bounded by B. If

$$\log |I_N| + K_N \log\Big(1 + \frac{B}{\tau_N}\Big) \lesssim N\varepsilon_N^2,$$

then there is a finite catalog estimator with risk $O(\varepsilon_N)$ on $\mathcal{C}$, provided each $(r,s) \in \mathcal{C}$ admits an ε_N-optimal potential in $\mathcal{U}(r,s)$.

Proof. Fix a certificate index α. Quantize the $2K_N$ values

$$\{\phi(x_{\alpha,\ell}), \psi(y_{\alpha,\ell}) : 1 \le \ell \le K_N\}$$

on a grid of mesh τ_N inside $[-B, B]$. For this α, the number of possible quantized traces is at most

$$\Big(1 + \frac{2B}{\tau_N}\Big)^{2K_N}.$$

For every trace cell and every α, choose one feasible representative dual potential if the cell contains a member of

$$\bigcup_{(r,s)\in\mathcal{C}} \mathcal{U}(r,s),$$

and choose nothing otherwise. This produces a data-independent catalog with logarithmic size bounded by

$$\log |I_N| + 2K_N \log\Big(1 + \frac{2B}{\tau_N}\Big).$$

Now fix $(r,s) \in \mathcal{C}$ and choose an index α whose certificate set determines $\mathcal{U}(r,s)$. Let $a_\star \in \mathcal{U}(r,s)$ be ε_N-optimal. The quantized trace of $a_\star$ has a representative $\tilde{a}$ in the catalog whose trace differs from that of $a_\star$ by at most τ_N at every certified coordinate. The determining property gives

$$|L_{\tilde{a}}(r,s) - L_{a_\star}(r,s)| \le \varepsilon_N.$$

Therefore the catalog has deterministic dual loss at most $2\varepsilon_N$ on (r,s). The entropy assumption and Theorem 17.2 give the statistical risk $O(\varepsilon_N)$. ☐

18.4 Mass-contact net theorem

The previous proposition is useful only when a small contact certificate exists. The next theorem proves one such certificate from a geometric property of an optimal plan. It is intentionally finite and non-asymptotic.

Let $L_{d,p} := pd^{(p-1)/2}$. Every c-concave potential for $c(x,y) = \|x-y\|_2^p$ on $[0,1]^d$ is $L_{d,p}$-Lipschitz in each variable after taking a c-transform normalization, because

$$|c(x,y) - c(x',y)| \le L_{d,p}\|x - x'\|_2$$

and similarly in y.

Theorem 18.5 (Mass-contact net certificate). *Let $G \subset [0,1]^d$ be finite, let $\mathcal{U}(r,s)$ be a family of normalized c-concave dual potentials bounded by B, and let $\pi_\star$ be an optimal plan between r and s. Suppose that for some pairs*

$$\mathfrak{Z} = \{(x_\ell, y_\ell)\}_{\ell=1}^K$$

there is a set $A \subset G \times G$ with

$$\pi_\star(A) \ge 1 - \zeta$$

such that every $(x,y) \in A$ is within product distance ρ of some certified pair:

$$\min_{1\le \ell \le K} \big(\|x - x_\ell\|_2 + \|y - y_\ell\|_2\big) \le \rho.$$

Then $\mathfrak{Z}$ (τ, ε)-determines $\mathcal{U}(r,s)$ at (r,s) with

$$\varepsilon = 2\tau + 2L_{d,p}\rho + 4B\zeta.$$

Proof. Let $a = (\phi_a, \psi_a)$ and $b = (\phi_b, \psi_b)$ belong to $\mathcal{U}(r,s)$, and assume their certified traces differ by at most τ in the sense of the definition. Since $\pi_\star$ has marginals r, s,

$$L_a(r,s) - L_b(r,s) = \sum_{x,y} \{(\phi_a - \phi_b)(x) + (\psi_a - \psi_b)(y)\}\pi_\star(x,y).$$

For $(x,y) \in A$, choose ℓ with

$$\|x - x_\ell\|_2 + \|y - y_\ell\|_2 \le \rho.$$

The Lipschitz bounds give

$$|(\phi_a - \phi_b)(x)| \le |\phi_a(x_\ell) - \phi_b(x_\ell)| + 2L_{d,p}\|x - x_\ell\|_2,$$
$$|(\psi_a - \psi_b)(y)| \le |\psi_a(y_\ell) - \psi_b(y_\ell)| + 2L_{d,p}\|y - y_\ell\|_2.$$

Hence the integrand on A is bounded by $2\tau + 2L_{d,p}\rho$. On A^c it is bounded by $4B$. Integrating with respect to $\pi_\star$ gives

$$|L_a(r,s) - L_b(r,s)| \le 2\tau + 2L_{d,p}\rho + 4B\zeta.$$

☐

Corollary 18.6 (Critical law for sparse mass-contact plans). *Let G_N be the critical grid, $M_N = |G_N| \asymp N\log N$, $h_N \asymp (N\log N)^{-1/d}$, and $\varepsilon_N = h_N^p$. Let $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$ be a class such that for every $(r,s) \in \mathcal{C}_N$ there is an optimal plan admitting a mass-contact net with parameters*

$$K_N, \qquad \rho_N \lesssim \varepsilon_N, \qquad \zeta_N \lesssim \varepsilon_N,$$

and such that

$$K_N \log\Big(\frac{M_N}{\varepsilon_N}\Big) \lesssim N\varepsilon_N^2.$$

Then W_p^p is estimable on $\mathcal{C}_N$ with powered risk $O(\varepsilon_N)$, and W_p is estimable with risk $O(h_N)$.

Proof. Let $\mathcal{U}(r,s)$ be the full normalized c-concave dual class with a fixed anchor. Since $0 \le c \le d^{p/2}$, this class is uniformly bounded by a constant $B_{d,p}$ and contains a Kantorovich optimizer. Use Theorem 18.5 with $\tau_N \asymp \varepsilon_N$. The data-independent collection I_N consists of all K_N-tuples in $G_N \times G_N$, so

$$\log|I_N| \le 2K_N \log M_N.$$

The entropy condition in Theorem 18.4 is therefore implied by the displayed assumption. The proposition gives the powered risk bound, and the passage from powered risk to W_p-risk uses

$$|u^{1/p} - v^{1/p}| \le |u-v|^{1/p}, \qquad u,v \ge 0.$$

□

Remark 18.7 (What this proves and what it does not). The theorem solves critical-grid classes whose optimal transport contact mass is concentrated near K_N edges with K_N below the statistical budget Nh_N^{2p}, up to logarithms. It does not solve the full Euclidean grid, because a generic optimal plan may have a contact set of size comparable to $|G_N|$. Its value is that the obstruction is now geometric and measurable: the only remaining hard cases are those whose near-contact set has genuinely high metric complexity at powered resolution h_N^p.

19 Graph certificates and low-complexity Monge laws

The localized contact estimates control near-active dual potentials by recording their values on finite contact sets. A sharper mechanism is available when the contact set itself has a low-complexity description. For value estimation one does not need to enumerate every optimal dual potential. It is enough to exhibit one feasible dual certificate whose slack is small on most of an optimal primal plan.

This section converts that observation into finite and continuum theorems. The resulting distribution classes may have arbitrary source law and full ambient support; the low-complexity object is the transport graph, the Monge map, or the dual certificate, not the density.

19.1 Finite high-mass tight graph certificates

Let $G \subset [0,1]^d$ be finite and $c(x,y) = \|x-y\|_2^p$. For a normalized feasible dual pair $a = (\phi,\psi) \in \mathcal{D}_G^0$, define

$$\mathrm{slack}_a(x,y) = c(x,y) - \phi_x - \psi_y \ge 0.$$

If $H \subset G \times G$ and $\alpha \ge 0$, say that a is α-tight on H when

$$\mathrm{slack}_a(x,y) \le \alpha \qquad ((x,y) \in H).$$

Theorem 19.1 (High-mass tight graph catalog)**.** *Let $G_N \subset [0,1]^d$ be a finite Euclidean grid and let $\mathfrak{H}_N$ be a finite family of subsets of $G_N \times G_N$. Suppose that for every $H \in \mathfrak{H}_N$ one has chosen a normalized feasible dual pair $a_H = (\phi_H,\psi_H) \in \mathcal{D}_{G_N}^0$. Let $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$. Assume that there are numbers $\alpha_N, \zeta_N \ge 0$ such that for every $(r,s) \in \mathcal{C}_N$ there exist $H = H(r,s) \in \mathfrak{H}_N$ and an optimal coupling $\pi_\star \in \Pi(r,s)$ satisfying*

$$\pi_\star(H) \ge 1 - \zeta_N, \qquad a_H \text{ is } \alpha_N\text{-tight on } H.$$

Then the finite-catalog estimator

$$\widehat{T}_N^{\mathfrak{H}} := \max_{H \in \mathfrak{H}_N} \left\{ \frac{1}{N}\sum_{i=1}^N \phi_H(X_i) + \frac{1}{N}\sum_{i=1}^N \psi_H(Y_i) \right\}$$

satisfies

$$\sup_{(r,s)\in\mathcal{C}_N} \mathbb{E}|\widehat{T}_N^{\mathfrak{H}} - T(r,s)| \le C_{d,p}\left(\alpha_N + \zeta_N + \sqrt{\frac{\log(2|\mathfrak{H}_N|)}{N}}\right).$$

The analogous squared-risk bound is

$$\sup_{(r,s)\in\mathcal{C}_N} \mathbb{E}(\widehat{T}_N^{\mathfrak{H}} - T(r,s))^2 \le C_{d,p}\left((\alpha_N+\zeta_N)^2 + \frac{\log(2|\mathfrak{H}_N|)}{N}\right).$$

Consequently, on the critical grid $h_N \asymp (N\log N)^{-1/d}$, if

$$\alpha_N + \zeta_N \lesssim h_N^p, \qquad \log|\mathfrak{H}_N| \lesssim Nh_N^{2p},$$

then W_p^p is estimable on $\mathcal{C}_N$ with powered risk $O(h_N^p)$, and W_p is estimable with risk $O(h_N)$.

Proof. Fix $(r,s) \in \mathcal{C}_N$ and choose $H, \pi_\star$ as in the assumption. Since a_H is feasible,

$$L_{a_H}(r,s) \le T(r,s).$$

Conversely, using the optimal plan $\pi_\star$,

$$\begin{aligned} T(r,s) - L_{a_H}(r,s) &= \sum_{x,y} \mathrm{slack}_{a_H}(x,y)\pi_\star(x,y) \\ &\le \alpha_N \pi_\star(H) + \sup_{x,y} \mathrm{slack}_{a_H}(x,y)\,\pi_\star(H^c). \end{aligned}$$

For $a_H \in \mathcal{D}_{G_N}^0$, the potentials are bounded by $4d^{p/2}$ and $0 \le c \le d^{p/2}$, hence the slack is bounded by a constant $C_{d,p}$. Thus

$$0 \le T(r,s) - L_{a_H}(r,s) \le \alpha_N + C_{d,p}\zeta_N.$$

The catalog $\mathcal{A}_N = \{a_H : H \in \mathfrak{H}_N\}$ therefore has deterministic approximation error at most $C_{d,p}(\alpha_N + \zeta_N)$. The same symmetrization and finite-class bound as in Theorem 17.2 gives the two powered-risk estimates. The passage from powered risk $O(h_N^p)$ to distance risk $O(h_N)$ follows by clipping the powered estimator to $[0, d^{p/2}]$ and using

$$|u^{1/p} - v^{1/p}| \le |u - v|^{1/p}, \qquad u, v \ge 0.$$

□

Remark 19.2 (Diagonal sign cubes versus graph certificates). At the diagonal $r = s$, the optimal dual face is enormous, as proved in Theorem 25.1. Nevertheless $T(r, r) = 0$ has the single graph certificate $H = \{(x, x) : x \in G_N\}$ and the single dual pair $(0, 0)$, which is exactly tight on H. The obstruction is therefore not the existence of many optimal potentials; it is the absence, in the unrestricted problem, of a uniformly low-entropy family of tight graphs or an analytic centering that makes such a family unnecessary.

19.2 Continuum dual catalogs

Let

$$\mathcal{D} = \{(\phi, \psi) \in L^\infty([0,1]^d)^2 : \phi(x) + \psi(y) \le \|x - y\|_2^p \text{ for all } x, y\}.$$

For $a = (\phi, \psi) \in \mathcal{D}$, write

$$L_a(P, Q) = \int \phi \, dP + \int \psi \, dQ, \qquad \widehat{L}_a = \frac{1}{N}\sum_{i=1}^N \phi(X_i) + \frac{1}{N}\sum_{i=1}^N \psi(Y_i).$$

Theorem 19.3 (Continuum dual-catalog theorem)**.** *Let $\mathcal{A}_N \subset \mathcal{D}$ be finite and assume*

$$\|\phi\|_\infty + \|\psi\|_\infty \le B \qquad ((\phi, \psi) \in \mathcal{A}_N).$$

Let $\mathcal{C} \subset \mathcal{P}([0,1]^d)^2$. If for every $(P, Q) \in \mathcal{C}$

$$0 \le W_p(P, Q)^p - \max_{a \in \mathcal{A}_N} L_a(P, Q) \le \varepsilon_N,$$

then

$$\sup_{(P,Q) \in \mathcal{C}} \mathbb{E}|\max_{a \in \mathcal{A}_N} \widehat{L}_a - W_p(P, Q)^p| \le \varepsilon_N + C_B \sqrt{\frac{\log(2|\mathcal{A}_N|)}{N}}.$$

The corresponding squared-risk bound is

$$\sup_{(P,Q) \in \mathcal{C}} \mathbb{E}(\max_{a \in \mathcal{A}_N} \widehat{L}_a - W_p(P, Q)^p)^2 \le C_B \left(\varepsilon_N^2 + \frac{\log(2|\mathcal{A}_N|)}{N}\right).$$

Proof. The deterministic error is the displayed approximation error. For the stochastic part,

$$\left|\max_{a \in \mathcal{A}_N} \widehat{L}_a - \max_{a \in \mathcal{A}_N} L_a(P, Q)\right| \le \sup_{a \in \mathcal{A}_N} |\widehat{L}_a - L_a(P, Q)|.$$

The summands are uniformly bounded by B. Symmetrization and Massart's finite-class inequality give the first-moment bound, and integration of the same bounded sub-Gaussian tail gives the second-moment bound. □

An infinite certificate class is handled by a sup-norm net. If $\mathcal{A} \subset \mathcal{D}$ is B-bounded and $\mathcal{A}_N$ is a δ_N-net in

$$d_\infty((\phi, \psi), (\phi', \psi')) = \|\phi - \phi'\|_\infty + \|\psi - \psi'\|_\infty,$$

then replacing $\mathcal{A}$ by $\mathcal{A}_N$ costs at most δ_N in deterministic dual value. Hence the theorem applies whenever

$$\delta_N \lesssim h_N^p, \qquad \log N(\delta_N, \mathcal{A}, d_\infty) \lesssim N h_N^{2p}.$$

19.3 Cyclically monotone Monge dictionaries

Let $c(x, y) = \|x - y\|_2^p$. A Borel map $S : [0,1]^d \to [0,1]^d$ is c-cyclically monotone if for every finite set $x_1, \ldots, x_m$ and every permutation σ,

$$\sum_{i=1}^m c(x_i, Sx_i) \le \sum_{i=1}^m c(x_i, Sx_{\sigma(i)}).$$

Equivalently, the graph of S is contained in the c-superdifferential of a c-concave potential. Thus there are dual feasible functions (ϕ_S, ψ_S) such that

$$\phi_S(x) + \psi_S(Sx) = c(x, Sx) \qquad \text{for all } x.$$

Theorem 19.4 (Arbitrary-source Monge dictionary)**.** *Let $\mathcal{S}_N = \{S_\theta : \theta \in \Theta_N\}$ be a finite family of c-cyclically monotone maps from $[0,1]^d$ to itself. For each θ, choose a bounded dual certificate $(\phi_\theta, \psi_\theta)$ satisfying*

$$\phi_\theta(x) + \psi_\theta(S_\theta x) = \|x - S_\theta x\|_2^p \qquad (x \in [0,1]^d),$$

and suppose

$$\|\phi_\theta\|_\infty + \|\psi_\theta\|_\infty \le B \qquad (\theta \in \Theta_N).$$

Let

$$\mathcal{C}(\mathcal{S}_N) = \{(P, Q) : Q = (S_\theta)_\# P \text{ for some } \theta \in \Theta_N\}.$$

Then

$$\widehat{T}_N^{\mathcal{S}} = \max_{\theta \in \Theta_N} \left\{ \frac{1}{N} \sum_{i=1}^N \phi_\theta(X_i) + \frac{1}{N} \sum_{i=1}^N \psi_\theta(Y_i) \right\}$$

satisfies

$$\sup_{(P,Q)\in \mathcal{C}(\mathcal{S}_N)} \mathbb{E}|\widehat{T}_N^{\mathcal{S}} - W_p(P,Q)^p| \le C_B \sqrt{\frac{\log(2|\Theta_N|)}{N}}.$$

In particular, if $d > 2p$, $h_N = (N \log N)^{-1/d}$, and

$$\log |\Theta_N| \lesssim N h_N^{2p},$$

then $W_p(P,Q)$ is estimable on $\mathcal{C}(\mathcal{S}_N)$ with risk $O(h_N)$, uniformly over all source laws P.

Proof. If $Q = (S_\theta)_\# P$, the coupling $(\mathrm{Id}, S_\theta)_\# P$ is optimal by c-cyclical monotonicity. The chosen certificate is tight on its graph, hence

$$W_p(P,Q)^p = \int \|x - S_\theta x\|_2^p \, dP(x) = \int \phi_\theta \, dP + \int \psi_\theta \, dQ = L_{(\phi_\theta,\psi_\theta)}(P,Q).$$

Thus the deterministic approximation error in Theorem 19.3 is zero for the catalog indexed by Θ_N. The displayed risk bound follows. The final statement uses $N^{-1/2} \le C h_N^p$ in the supercritical regime after the allowed entropy factor, and then uses $u \mapsto u^{1/p}$. □

Corollary 19.5 (Translation laws with arbitrary source). *Assume $p > 1$. Let $\mathcal{T} \subset [-1,1]^d$ be compact and consider all pairs*

$$Q = (x \mapsto x + t)_\# P, \qquad t \in \mathcal{T},$$

for which the translated support remains in $[0,1]^d$. Then the critical rate $O(h_N)$ holds uniformly over this class for every $d > 2p$.

Proof. For $h(z) = \|z\|_2^p$, Fenchel's inequality gives

$$h(x-y) \ge u \cdot x - u \cdot y - h^*(u).$$

For a translation $S_t x = x + t$, choose $u_t = \nabla h(-t)$. Then equality holds whenever $y = x + t$. Hence

$$\phi_t(x) = u_t \cdot x, \qquad \psi_t(y) = -u_t \cdot y - h^*(u_t)$$

is a bounded certificate on the unit cube. Since $p > 1$, $t \mapsto u_t$ is continuous on the compact set $\mathcal{T}$, and the certificate class has metric entropy

$$\log N(\delta, \{(\phi_t, \psi_t) : t \in \mathcal{T}\}, d_\infty) \lesssim_{d,p,\mathcal{T}} \log(1/\delta).$$

Taking $\delta_N \asymp h_N^p$, this entropy is $o(N h_N^{2p})$ because $d > 2p$. The net version of Theorem 19.3 proves powered risk $O(h_N^p)$, hence root risk $O(h_N)$. □

Corollary 19.6 (Affine Brenier dictionaries for W_2). *Let $p = 2$. Fix $0 < \kappa < K < \infty$ and let $\mathcal{A}$ be a compact parameter set of affine maps*

$$S_{A,b}(x) = Ax + b, \qquad A = A^\top, \quad \kappa I \preceq A \preceq K I,$$

mapping the relevant source supports into $[0,1]^d$. Then the critical rate $O(h_N)$, $h_N = (N \log N)^{-1/d}$, holds uniformly over all pairs

$$Q = (S_{A,b})_\# P, \qquad (A,b) \in \mathcal{A},$$

with arbitrary source law P, whenever $d > 4$.

Proof. Let

$$u_{A,b}(x) = \frac{1}{2} x^\top A x + b \cdot x.$$

The map $S_{A,b} = \nabla u_{A,b}$ is cyclically monotone because $u_{A,b}$ is convex. For the cost $c(x,y) = \|x-y\|_2^2$,

$$\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2x \cdot y \ge \|x\|_2^2 + \|y\|_2^2 - 2u_{A,b}(x) - 2u_{A,b}^*(y),$$

with equality when $y = \nabla u_{A,b}(x) = S_{A,b} x$. Thus

$$\phi_{A,b}(x) = \|x\|_2^2 - 2u_{A,b}(x), \qquad \psi_{A,b}(y) = \|y\|_2^2 - 2u_{A,b}^*(y)$$

is a dual certificate. On $[0,1]^d$ and under the spectral bounds on A, these certificates are uniformly bounded and Lipschitz in the finite-dimensional parameter (A,b). Their metric entropy at radius δ is $O_{d,\kappa,K}(\log(1/\delta))$. The net-catalog theorem with $\delta_N \asymp h_N^2$ gives powered risk $O(h_N^2)$, hence W_2-risk $O(h_N)$. □

Theorem 19.7 (Smooth Brenier certificate classes). *Let $p = 2$, $d > 4$, and $h_N = (N \log N)^{-1/d}$. Let $\mathfrak{U}$ be a class of convex functions u on a neighbourhood of $[0,1]^d$ such that $S_u = \nabla u$ maps the relevant source supports into $[0,1]^d$. Assume that the associated certificates*

$$\phi_u(x) = \|x\|_2^2 - 2u(x), \qquad \psi_u(y) = \|y\|_2^2 - 2u^*(y)$$

are uniformly bounded and satisfy

$$\log N(\delta, \{(\phi_u, \psi_u) : u \in \mathfrak{U}\}, d_\infty) \le C \delta^{-q/s} \qquad (0 < \delta < 1)$$

for some $q > 0$ and

$$s > \frac{2q}{d-4}.$$

Then

$$\sup_{u \in \mathfrak{U}} \sup_{P} \mathbb{E} \left| \widehat{W}_N - W_2(P, (\nabla u)_\# P) \right| \lesssim h_N$$

for the dual-catalog estimator built from a d_∞-net at radius $\delta_N \asymp h_N^2$. In particular, a uniformly convex C^{s+2}-bounded Brenier class with $q = d$ satisfies the target rate whenever

$$s > \frac{2d}{d-4}.$$

Proof. For every u, the graph of ∇u is cyclically monotone and the displayed Fenchel certificate is tight on that graph. The metric-entropy hypothesis at $\delta_N \asymp h_N^2$ gives

$$\log N(\delta_N, \mathfrak{U}, d_\infty) \lesssim h_N^{-2q/s} = (N\log N)^{2q/(sd)}.$$

The condition $s > 2q/(d-4)$ is equivalent to

$$\frac{2q}{sd} < 1 - \frac{4}{d},$$

and therefore

$$h_N^{-2q/s} = o(Nh_N^4).$$

The continuum dual-catalog theorem gives powered risk $O(h_N^2)$. Taking square roots gives W_2-risk $O(h_N)$. The last assertion follows from the classical Kolmogorov–Tikhomirov entropy bound for C^s-balls and the stability of the Legendre transform under uniform convexity and bounded C^{s+2} norms. □

19.4 Finite mixtures of graph certificates

The graph-certificate theorem also handles finite unions of low-complexity Monge graphs. This is useful for piecewise transport where no single global formula describes the whole plan.

Corollary 19.8 (Piecewise cyclically monotone graph dictionaries)**.** *Let G_N be the critical grid and let $\mathfrak{H}_N^{(1)}$ be a finite family of subsets of $G_N \times G_N$, each equipped with a feasible dual certificate which is exactly tight on that subset. Let $\mathfrak{H}_N^{(K)}$ be the family of all unions of at most K_N members of $\mathfrak{H}_N^{(1)}$ for which there exists a single feasible dual certificate exactly tight on the union. Consider the class $\mathcal{C}_N^{(K)}$ of grid pairs admitting an optimal plan supported on some $H \in \mathfrak{H}_N^{(K)}$. If*

$$K_N \log |\mathfrak{H}_N^{(1)}| \lesssim N h_N^{2p},$$

then W_p^p is estimable on $\mathcal{C}_N^{(K)}$ with powered risk $O(h_N^p)$, and W_p is estimable with risk $O(h_N)$.

Proof. For every admissible union H, choose one feasible certificate exactly tight on H. Since every pair in $\mathcal{C}_N^{(K)}$ has an optimal plan supported on such an H, Theorem 19.1 applies with $\alpha_N = \zeta_N = 0$. The number of admissible unions is at most

$$\sum_{k=1}^{K_N} |\mathfrak{H}_N^{(1)}|^k \le (K_N + 1)|\mathfrak{H}_N^{(1)}|^{K_N},$$

and the displayed entropy condition absorbs the harmless $\log(K_N + 1)$ term. □

Remark 19.9 (Interpretation). The exact L_1-skeleton results solve classes where the value decomposes into many independent one-dimensional nonsmooth functionals. The graph-certificate results solve a different regime: the value is certified by a small-description optimal graph, even when the source law has full $N\log N$-dimensional complexity. The unrestricted Euclidean grid may contain transport graphs of much larger description complexity; those cases are precisely where centered curvature or growing-alphabet LP debiasing is still needed.

20 Smooth and low-dimensional semidual phases

The previous section gives a local entropy principle. We now prove a concrete positive theorem from it: the target rate holds whenever the relevant semidual potentials belong, up to $O(h_N^p)$ dual loss, to a smooth or low-dimensional phase class whose metric entropy fits inside the critical budget. This is a genuine extension of the exact L_1-skeleton results. It does not assume that the transport value is itself a total-variation norm; it uses the Kantorovich dual directly.

20.1 Metric entropy of smooth phase classes

Let $s > 0$ and $R \ge 1$. For a compact set $U \subset \mathbb{R}^q$, write $\mathcal{H}_R^s(U)$ for the ball of real-valued functions on U with Hölder norm at most R, with the usual interpretation $s = k + \alpha$, $k \in \mathbb{N}_0$, $\alpha \in (0,1]$. The following elementary entropy bound is sufficient for the critical-grid estimates.

Lemma 20.1 (Hölder phase entropy)**.** *Let $U \subset [0,1]^q$. For every $\delta \in (0,1)$,*

$$\log N\big(\delta, \mathcal{H}_R^s(U), \|\cdot\|_\infty\big) \le C_{q,s,R}\, \delta^{-q/s} \log\Big(\frac{2}{\delta}\Big).$$

The same bound holds for restrictions of $\mathcal{H}_R^s([0,1]^q)$ to any finite subset of U.

Proof. Choose a mesh $\ell \asymp (\delta/R)^{1/s}$ on $[0,1]^q$. On each mesh cube, approximate a function by its Taylor polynomial of degree $\lfloor s \rfloor$ at the cube center. The Hölder remainder is $O(R\ell^s) \le \delta/4$. The number of coefficients per cube depends only on q, s, each coefficient is bounded by R, and quantization with mesh a fixed multiple of δ after the natural derivative scaling gives at most

$$\exp\{C_{q,s,R}\ell^{-q} \log(2/\delta)\}$$

possibilities. Since $\ell^{-q} \asymp \delta^{-q/s}$, the asserted bound follows. Restricting to a finite subset cannot increase the covering number. □

20.2 Critical theorem for smooth semidual phases

Let $G_N \subset [0,1]^d$ be the critical grid,

$$h_N \asymp (N\log N)^{-1/d}, \qquad |G_N| \asymp N\log N,$$

and set

$$\varepsilon_N := h_N^p.$$

For $g : G_N \to \mathbb{R}$, define the c-transform

$$f_g(x) := \min_{y\in G_N} \{\|x - y\|_2^p - g(y)\}, \qquad x \in G_N.$$

The pair (f_g, g) is dual-feasible.

Definition 20.2 (Smooth semidual phase)**.** Let $s, R > 0$ and $A < \infty$. A class $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$ is called (s, R, A)-smooth-semiphase if for every $(r, s_0) \in \mathcal{C}_N$ there exists $g_\star : G_N \to \mathbb{R}$ such that:

(i) $g_\star$ is the restriction to G_N of a function in $\mathcal{H}^s_R([0, 1]^d)$;

(ii) $g_\star(y_0) = 0$ for a fixed anchor $y_0 \in G_N$;

(iii) the semidual pair $(f_{g_\star}, g_\star)$ is $A\varepsilon_N$-optimal:

$$0 \le T(r, s_0) - \{\langle f_{g_\star}, r\rangle + \langle g_\star, s_0\rangle\} \le A\varepsilon_N.$$

Theorem 20.3 (Critical law for smooth semidual phases)**.** *Assume $d > 2p$. Let $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$ be (s, R, A)-smooth-semiphase. If*

$$s > \frac{pd}{d - 2p},$$

then there exists an estimator $\widehat{T}_N$ such that

$$\sup_{(r,s_0)\in\mathcal{C}_N} \mathbb{E}|\widehat{T}_N - W_p(r, s_0)^p| \le C_{d,p,s,R,A}\, h_N^p,$$

and

$$\sup_{(r,s_0)\in\mathcal{C}_N} \mathbb{E}(\widehat{T}_N - W_p(r, s_0)^p)^2 \le C_{d,p,s,R,A}\, h_N^{2p}.$$

Consequently $\widehat{W}_N := (\widehat{T}_N)_+^{1/p}$ satisfies

$$\sup_{(r,s_0)\in\mathcal{C}_N} \mathbb{E}|\widehat{W}_N - W_p(r, s_0)| \le C_{d,p,s,R,A}\, h_N,$$

and the analogous squared-risk bound $O(h_N^2)$.

Proof. Let

$$\delta_N := ch_N^p$$

with $c > 0$ sufficiently small. By Theorem 20.1, the anchored smooth phase class admits a δ_N-net $\mathcal{G}_N$ in the sup norm satisfying

$$\log|\mathcal{G}_N| \le C_{d,s,R}\, \delta_N^{-d/s} \log(2/\delta_N) \le Ch_N^{-pd/s} \log(1/h_N).$$

The condition $s > pd/(d - 2p)$ is equivalent to

$$\frac{pd}{s} < d - 2p.$$

Since $h_N^{-d} \asymp N \log N$,

$$Nh_N^{2p} \asymp \frac{h_N^{-d+2p}}{\log N}.$$

Therefore

$$\log|\mathcal{G}_N| = o(Nh_N^{2p}).$$

For each $g \in \mathcal{G}_N$, include the feasible dual pair (f_g, g) in the catalog. The map $g \mapsto f_g$ is 1-Lipschitz in the sup norm:

$$\|f_g - f_{\tilde g}\|_\infty \le \|g - \tilde g\|_\infty.$$

Hence if $g_\star$ is the $A\varepsilon_N$-optimal phase from Theorem 20.2 and $\tilde g \in \mathcal{G}_N$ satisfies $\|g_\star - \tilde g\|_\infty \le \delta_N$, then

$$\begin{aligned}&|\{\langle f_{g_\star}, r\rangle + \langle g_\star, s_0\rangle\} - \{\langle f_{\tilde g}, r\rangle + \langle \tilde g, s_0\rangle\}|\\ &\le \|f_{g_\star} - f_{\tilde g}\|_\infty + \|g_\star - \tilde g\|_\infty \le 2\delta_N.\end{aligned}$$

Thus the catalog approximates the true semidual value to error $(A+2c)\varepsilon_N$. Since the catalog entropy is $o(Nh_N^{2p}) = o(N\varepsilon_N^2)$, Theorem 17.2 gives the powered absolute and squared-risk bounds. The passage from powered risk to distance risk uses

$$|u^{1/p} - v^{1/p}| \le |u - v|^{1/p} \qquad (u, v \ge 0).$$

For squared distance risk, apply Lyapunov's inequality:

$$\mathbb{E}|\widehat{W}_N - W_p|^2 \le \mathbb{E}|\widehat{T}_N - W_p^p|^{2/p} \le \left(\mathbb{E}|\widehat{T}_N - W_p^p|^2\right)^{1/p} \lesssim h_N^2.$$

□

20.3 Low-dimensional phase families

The smoothness threshold in Theorem 20.3 concerns the intrinsic dimension of the semidual phase, not necessarily the ambient dimension. This matters because the known lower-bound mechanism is already low-dimensional inside a high-dimensional cube, and because many transport maps have potentials depending on a small number of features.

Let $\Phi_N : G_N \to [0, 1]^q$ be any fixed feature map. A semidual phase has feature dimension q if

$$g(y) = \bar g(\Phi_N(y))$$

for some $\bar g : [0, 1]^q \to \mathbb{R}$.

Theorem 20.4 (Critical law for low-dimensional smooth phases)**.** *Assume $d > 2p$. Let $q \le d$, $s, R, A > 0$, and suppose $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$ has the following property: for every $(r, s_0) \in \mathcal{C}_N$, there exists*

$$g_\star(y) = \bar g_\star(\Phi_N(y)), \qquad \bar g_\star \in \mathcal{H}^s_R([0, 1]^q),$$

anchored at y_0, such that $(f_{g_\star}, g_\star)$ is Ah_N^p-optimal. If

$$s > \frac{pq}{d-2p},$$

then W_p^p and W_p are estimable on $\mathcal{C}_N$ at powered accuracy $O(h_N^p)$ and distance accuracy $O(h_N)$, respectively.

Proof. Repeat the proof of Theorem 20.3, but cover $\mathcal{H}_R^s([0,1]^q)$ rather than $\mathcal{H}_R^s([0,1]^d)$. The entropy is

$$O\Big(h_N^{-pq/s}\log(1/h_N)\Big).$$

The condition $s > pq/(d-2p)$ is exactly the condition that this entropy be $o(Nh_N^{2p})$. The c-transform remains 1-Lipschitz with respect to $\|g-\tilde{g}\|_\infty$, so the same catalog proof applies. □

Corollary 20.5 (Smooth spiked semidual classes)**.** *If the optimal semiduals of a critical-grid class depend, up to $O(h_N^p)$ dual loss, on $q < d$ linear coordinates and have C^s norm uniformly bounded with*

$$s > \frac{pq}{d-2p},$$

then the class satisfies the sharp target upper bound $O(h_N)$ for W_p.

Proof. Take Φ_N to be the projection onto the q-dimensional coordinate subspace and apply Theorem 20.4. □

Remark 20.6 (Relation to smooth-cost plug-in theory). Smooth-cost empirical OT results control the stochastic error of the empirical optimizer when a fixed smooth dual phase is already present. Theorems 20.3 and 20.4 use a different mechanism: a smooth phase class is compressed into a finite semidual catalog of entropy below the critical statistical budget. This is why the theorem remains meaningful at the critical grid scale, where the full unsmoothed semidual class has entropy of order h_N^{-d} and is much too large.

21 Higher-order debiased regularization criteria

The finite LP may also be attacked without a small dual catalog. The idea is to replace the nonsmooth Kantorovich value T by a smooth regularized value T_λ, estimate T_λ, and cancel the regularization bias. Theorem 17.4 gave the first-order version. This section records the higher-order form needed for serious growing-alphabet work. The statements are deterministic once the displayed analytic estimates are verified.

21.1 Regularization-bias Richardson cancellation

Let $T_\lambda(r,s)$ be a family of regularized values. Fix an integer $K \ge 1$ and distinct positive numbers $q_0, \dots, q_K$. Let $w_0, \dots, w_K$ be the unique Richardson weights satisfying

$$\sum_{\ell=0}^{K} w_\ell = 1, \qquad \sum_{\ell=0}^{K} w_\ell q_\ell^j = 0 \quad (1 \le j \le K).$$

The weights depend only on K and the chosen q_ℓ's.

Theorem 21.1 (Higher-order Richardson criterion)**.** *Assume that for all $r, s \in \mathcal{P}(G_N)$ and all sufficiently small λ,*

$$T_\lambda(r,s) = T(r,s) + \sum_{j=1}^{K} b_j(r,s)\lambda^j + R_{K+1,\lambda}(r,s),$$

with

$$|R_{K+1,\lambda}(r,s)| \le A\lambda^{K+1}.$$

Assume also that empirical estimators $\widehat{T}_\lambda$ satisfy

$$\sup_{r,s} \mathbb{E}|\widehat{T}_\lambda - T_\lambda(r,s)| \le S_N(\lambda).$$

Then

$$\widehat{T}_\lambda^{(K)} := \sum_{\ell=0}^{K} w_\ell \widehat{T}_{q_\ell \lambda}$$

satisfies

$$\sup_{r,s} \mathbb{E}|\widehat{T}_\lambda^{(K)} - T(r,s)| \le C_{K,q}\left[A\lambda^{K+1} + \max_{0\le\ell\le K} S_N(q_\ell\lambda)\right].$$

Proof. By the defining moment equations for w_ℓ,

$$\sum_{\ell=0}^{K} w_\ell T_{q_\ell\lambda} = T + \sum_{j=1}^{K} b_j\lambda^j\left(\sum_{\ell=0}^{K} w_\ell q_\ell^j\right) + \sum_{\ell=0}^{K} w_\ell R_{K+1,q_\ell\lambda} = T + O_{K,q}(A\lambda^{K+1}).$$

The stochastic term is bounded by

$$\sum_{\ell=0}^{K} |w_\ell|\, \mathbb{E}|\widehat{T}_{q_\ell\lambda} - T_{q_\ell\lambda}|,$$

which gives the claim. □

21.2 Sampling-bias jackknife

Regularization is only one source of bias. A smooth plug-in estimator has a sampling bias whose leading terms may scale with the alphabet size. The next criterion is the finite-sample version of the von Mises/jackknife cancellation needed in the critical-grid problem.

Let $\widehat{F}_n$ be an estimator of a smooth functional $F(\theta)$ from n observations, where θ ranges over a finite-dimensional simplex. Suppose $n_\ell = \lfloor \alpha_\ell n \rfloor$, $0 < \alpha_0 < \cdots < \alpha_K \le 1$, and let $v_0, \ldots, v_K$ solve

$$\sum_{\ell=0}^{K} v_\ell = 1, \qquad \sum_{\ell=0}^{K} v_\ell \alpha_\ell^{-j} = 0 \quad (1 \le j \le K).$$

Theorem 21.2 (Finite jackknife criterion)**.** *Assume that uniformly over θ*

$$\mathbb{E}_\theta \widehat{F}_n = F(\theta) + \sum_{j=1}^{K} a_j(\theta) n^{-j} + R_{K+1,n}(\theta), \qquad |R_{K+1,n}(\theta)| \le A_N n^{-(K+1)},$$

and

$$\mathbb{E}_\theta |\widehat{F}_n - \mathbb{E}_\theta \widehat{F}_n| \le V_N(n).$$

Construct independent subsample estimators $\widehat{F}^{(\ell)}_{n_\ell}$ and set

$$\widehat{F}^{\text{jack}}_n := \sum_{\ell=0}^{K} v_\ell \widehat{F}^{(\ell)}_{n_\ell}.$$

Then

$$\sup_\theta \mathbb{E}_\theta |\widehat{F}^{\text{jack}}_n - F(\theta)| \le C_{K,\alpha} \left[A_N n^{-(K+1)} + \max_{0 \le \ell \le K} V_N(n_\ell) \right].$$

Proof. The proof is identical to Richardson cancellation. Ignoring integer parts first, the equations defining v_ℓ cancel every power n^{-j}, $1 \le j \le K$, because $n_\ell^{-j} = \alpha_\ell^{-j} n^{-j}$. Integer rounding changes n_ℓ^{-j} by $O(n^{-j-1})$, which is absorbed into the $R_{K+1,n}$ term after enlarging A_N. The stochastic part is controlled by the triangle inequality and the displayed deviation bound. □

21.3 The combined finite target

The two cancellation mechanisms can be composed. A regularized LP estimator for the critical grid should be judged against the following concrete target, not against an informal statement that "debiasing is needed."

Corollary 21.3 (Combined debiasing target)**.** *Let $h_N = (N \log N)^{-1/d}$, $\varepsilon_N = h_N^p$, and $|G_N| \asymp N \log N$. Suppose there are regularized empirical estimators $\widehat{T}_{\lambda,n}$ for which:*

(i) $T_\lambda = T + \sum_{j=1}^{K} b_j \lambda^j + O(\lambda^{K+1})$ *uniformly;*

(ii) *after a K'-th order sample jackknife, the centered stochastic error is $O(\varepsilon_N)$;*

(iii) *the remaining sample-bias remainder is $O(\varepsilon_N)$;*

(iv) *one can choose λ_N with $\lambda_N^{K+1} \lesssim \varepsilon_N$.*

Then the critical-grid powered value is estimable with risk $O(\varepsilon_N)$, and hence W_p is estimable with risk $O(h_N)$.

Proof. Apply Theorem 21.2 to the estimator of each $T_{q_\ell \lambda_N}$, then apply Theorem 21.1. The four assumptions make every term $O(\varepsilon_N)$. The powered-to-distance conversion is the same as in Theorem 20.3. □

Remark 21.4 (Why plain regularized plug-in is not enough)**.** For smooth functionals on an M-point simplex, second-order sampling bias typically contains a trace term of size comparable to $M/(N\lambda)$, where λ is the regularization curvature scale. At the critical alphabet size $M \asymp N \log N$, this term is far too large if λ is chosen small enough to make the deterministic regularization bias $O(h_N^p)$. Thus the essential missing step is not the existence of a regularization, but a uniform high-dimensional cancellation of its trace bias. Theorems 21.2 and 21.3 state the exact form in which such a cancellation would close the problem.

22 Root accuracy and scale-adaptive powered loss

The preceding finite criterion was deliberately stated in the stronger powered form

$$\mathbb{E}|\widehat{T} - T| \lesssim h_N^p, \qquad T = W_p^p,$$

because this is the cleanest sufficient condition near the diagonal. It is, however, stronger than the root-distance risk actually required by the minimax problem. This distinction matters for a full solution: far from the diagonal, estimating W_p^p to order h_N^p is wasteful. The correct global powered target is scale-adaptive.

Throughout this section $h \in (0,1)$, $T \ge 0$, $w = T^{1/p}$, and $\widehat{T}$ is any real estimator. We write $\widehat{W} = (\widehat{T})_+^{1/p}$.

Lemma 22.1 (Powered-to-root deterministic inequality)**.** *For every $p \ge 1$ there is $C_p < \infty$ such that, for all $T \ge 0$, all $h \in (0,1)$, and all real S,*

$$\left| (S)_+^{1/p} - T^{1/p} \right| \le C_p h + C_p \frac{|S-T|}{(T^{1/p}+h)^{p-1}}.$$

Consequently, if

$$\mathbb{E}|S - T| \le A h (T^{1/p} + h)^{p-1},$$

then

$$\mathbb{E} \left| (S)_+^{1/p} - T^{1/p} \right| \le C_p (1 + A) h.$$

Proof. Let $w = T^{1/p}$ and $u = (S)_+^{1/p}$. If $w \le 2h$, then

$$|u - w| \le u + w \le |u^p - w^p|^{1/p} + 2w \le |S - T|^{1/p} + 4h.$$

Since $a^{1/p} \le h + ah^{1-p}$ for $a \ge 0$, this is bounded by

$$C_p h + C_p |S - T| h^{1-p} \le C_p h + C_p \frac{|S - T|}{(w + h)^{p-1}}.$$

If $w > 2h$ and $u \ge w/2$, the mean-value theorem gives

$$|u - w| \le C_p \frac{|u^p - w^p|}{w^{p-1}} \le C_p \frac{|S - T|}{(w + h)^{p-1}}.$$

If $w > 2h$ and $u < w/2$, then $|S - T| \ge (1 - 2^{-p})w^p$, and hence

$$|u - w| \le w \le C_p \frac{|S - T|}{w^{p-1}} \le C_p \frac{|S - T|}{(w + h)^{p-1}}.$$

The expectation bound follows immediately. ☐

Theorem 22.2 (Scale-adaptive critical-grid criterion). *Let $G_N \subset [0,1]^d$ have mesh $h_N \asymp (N \log N)^{-1/d}$, and set*

$$T_N(r, s) = W_p(r, s)^p, \qquad w_N(r, s) = W_p(r, s).$$

Assume that an estimator $\widehat{T}_N$, based on N samples from each of $r, s \in \mathcal{P}(G_N)$, satisfies the adaptive powered risk bound

$$\sup_{r,s \in \mathcal{P}(G_N)} \frac{\mathbb{E}|\widehat{T}_N - T_N(r, s)|}{h_N \big(w_N(r, s) + h_N\big)^{p-1}} \le C.$$

Then

$$\sup_{r,s \in \mathcal{P}(G_N)} \mathbb{E} \left| (\widehat{T}_N)_+^{1/p} - W_p(r, s) \right| \le C'_{p,C}\, h_N.$$

By Theorem 2.2, the same conclusion on the critical grid implies the unrestricted continuum upper bound

$$\inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}([0,1]^d)} \mathbb{E}|\widehat{W} - W_p(P, Q)| \lesssim_{d,p} (N \log N)^{-1/d}.$$

Proof. Apply Theorem 22.1 with $S = \widehat{T}_N$ and $T = T_N(r, s)$, uniformly over r, s. The continuum implication is exactly the quantization and lifting theorem proved in Section 2. ☐

Remark 22.3 (Why this correction is important). The powered target h_N^p remains the right target on the diagonal $W_p(r, s) \lesssim h_N$. On the annulus $W_p(r, s) \asymp a \gg h_N$, the root-risk problem only requires powered accuracy $h_N a^{p-1}$. A full proof should therefore be judged against the adaptive denominator in Theorem 22.2, not only against the stronger uniform h_N^p denominator.

23 Four-sample centering and the finite curvature defect

The full Euclidean grid cannot be solved by a purely uncentered empirical supremum over near-active dual potentials. The diagonal $r = s$ already has a huge optimal dual face, and the empirical maximum over that face is precisely the $N^{-1/d}$ plug-in bias. The next statistic removes that diagonal width before any geometric localization is attempted.

Let $G \subset [0,1]^d$ be finite and $c(x, y) = \|x - y\|_2^p$. Write

$$T(r, s) = W_p(r, s)^p = \min_{\pi \in \Pi(r,s)} \sum_{x,y \in G} c(x, y)\pi_{xy}.$$

Let

$$\widehat{r}, \widehat{r}', \widehat{s}, \widehat{s}'$$

be four independent empirical measures, where $\widehat{r}, \widehat{r}'$ use N samples from r, and $\widehat{s}, \widehat{s}'$ use N samples from s. Define the centered powered statistic

$$\widehat{T}_N^{\mathrm{ctr}} := T(\widehat{r}, \widehat{s}) - \frac{1}{2} T(\widehat{r}, \widehat{r}') - \frac{1}{2} T(\widehat{s}, \widehat{s}').$$

In applications one clips it to $[0, d^{p/2}]$, but clipping is immaterial for the estimates below. Its deterministic bias is the centered curvature defect

$$\Gamma_N(r, s) := \mathbb{E}\widehat{T}_N^{\mathrm{ctr}} - T(r, s).$$

Lemma 23.1 (Exact diagonal cancellation). *For every $r \in \mathcal{P}(G)$,*

$$\Gamma_N(r, r) = 0.$$

Equivalently,

$$\mathbb{E}T(\widehat{r}, \widehat{r}') - \frac{1}{2}\mathbb{E}T(\widehat{r}, \widehat{r}'') - \frac{1}{2}\mathbb{E}T(\widehat{r}', \widehat{r}''') = T(r, r) = 0,$$

where all empirical measures are independent copies from r.

Proof. If $r = s$, then the three random variables

$$T(\widehat{r}, \widehat{s}), \qquad T(\widehat{r}, \widehat{r}'), \qquad T(\widehat{s}, \widehat{s}')$$

have the same distribution. Therefore their expectations cancel with coefficients $1, -1/2, -1/2$. Since $T(r, r) = 0$, the claim follows. ☐

Lemma 23.2 (Universal bounded-difference variance). *Let $D = \mathrm{diam}(G)$. For all $r, s \in \mathcal{P}(G)$,*

$$\mathrm{Var}\Big(\widehat{T}_N^{\mathrm{ctr}}\Big) \le \frac{C_p D^{2p}}{N}.$$

Consequently, in the supercritical regime $d > 2p$, for the critical mesh

$$h_N = (N \log N)^{-1/d},$$

one has

$$\left(\operatorname{Var} \widehat{T}_N^{\mathrm{ctr}}\right)^{1/2} = o(h_N^p).$$

Proof. The map $(a, b) \mapsto T(a, b)$ is D^p-Lipschitz in each marginal with respect to total variation:

$$|T(a, b) - T(a', b)| \le D^p \ \mathrm{TV}(a, a').$$

Indeed, keep the common part of a and a', and move only the excess mass; the latter has total mass $\mathrm{TV}(a, a')$, and every unit of mass costs at most D^p. Replacing one observation in an empirical measure changes that empirical measure by total variation $1/N$. Hence changing one of the $4N$ observations changes $\widehat{T}_N^{\mathrm{ctr}}$ by at most $C_p D^p/N$. Efron–Stein gives

$$\operatorname{Var}\left(\widehat{T}_N^{\mathrm{ctr}}\right) \le \frac{1}{2}(4N)\frac{C_p D^{2p}}{N^2} \le \frac{C_p D^{2p}}{N}.$$

Finally,

$$\frac{N^{-1/2}}{h_N^p} = N^{p/d-1/2}(\log N)^{p/d} \to 0$$

because $d > 2p$. □

Theorem 23.3 (Centered curvature criterion for the full minimax law)**.** *Assume that for the critical Euclidean grids G_N there is a constant C such that*

$$\sup_{r,s\in\mathcal{P}(G_N)} \frac{|\Gamma_N(r, s)|}{h_N\left(W_p(r, s) + h_N\right)^{p-1}} \le C. \tag{CC}$$

Then the clipped centered estimator

$$\widehat{W}_N^{\mathrm{ctr}} := \left[\left(\widehat{T}_N^{\mathrm{ctr}} \vee 0\right) \wedge d^{p/2}\right]^{1/p}$$

satisfies

$$\sup_{r,s\in\mathcal{P}(G_N)} \mathbb{E}|\widehat{W}_N^{\mathrm{ctr}} - W_p(r, s)| \lesssim_{d,p,C} h_N.$$

Consequently, (CC) *implies the full unrestricted continuum upper bound*

$$\inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}([0,1]^d)} \mathbb{E}|\widehat{W} - W_p(P, Q)| \lesssim_{d,p} (N \log N)^{-1/d}.$$

Proof. By Theorem 23.2,

$$\mathbb{E}\left|\widehat{T}_N^{\mathrm{ctr}} - \mathbb{E}\widehat{T}_N^{\mathrm{ctr}}\right| \le C_{d,p}N^{-1/2} \le C_{d,p}h_N^p \le C_{d,p}h_N\left(W_p(r, s) + h_N\right)^{p-1}.$$

Together with (CC), this gives the adaptive powered risk bound of Theorem 22.2. Clipping cannot increase the distance to $T(r, s) \in [0, d^{p/2}]$. The grid-to-continuum implication is Theorem 2.2. □

Remark 23.4 (What centering does and does not prove). The original finite LP problem asks for an estimator. Theorem 23.3 gives a sufficient route: if the signed deterministic curvature defect satisfies

$$\begin{aligned}\mathbb{E}T(\widehat{r}, \widehat{s}) - \frac{1}{2}\mathbb{E}T(\widehat{r}, \widehat{r}') - \frac{1}{2}\mathbb{E}T(\widehat{s}, \widehat{s}') - T(r, s)\\ = O\Big(h_N\left(W_p(r, s) + h_N\right)^{p-1}\Big),\end{aligned}$$

then the full upper bound follows. However, Theorem 7.3 shows that raw diagonal centering is not, by itself, a universal $N \log N$-gain mechanism for transport LPs. Thus the displayed curvature inequality should be viewed as a special Euclidean possibility, not as a generic consequence of centering. The robust finite-grid target is adaptive polynomialized LP debiasing; centering is one component of that target because it annihilates the diagonal width and gives $N^{-1/2}$ stochastic fluctuations.

24 Finite curvature calculus for the centered statistic

The centered statistic isolates a deterministic curvature defect. This section records the finite-dimensional calculus behind that defect. The point is not that the Kantorovich LP is globally smooth; it is not. Rather, any proposed full proof must show that after the diagonal self-noise has been subtracted, the remaining nonsmoothness has only critical size. The following identities make this requirement explicit.

Let

$$\mathsf{T}(a, b) := W_p(a, b)^p, \qquad a, b \in \mathcal{P}(G),$$

and let

$$\mathsf{U}_G := \{u \in \mathbb{R}^G : \langle u, \mathbf{1}\rangle = 0\}$$

be the tangent space of the simplex. For $r \in \mathcal{P}(G)$, write

$$\Sigma_r := \operatorname{diag}(r) - rr^\top.$$

If $\widehat{r}$ is the empirical law of N samples from r, then

$$\mathbb{E}(\widehat{r} - r) = 0, \qquad \mathbb{E}[(\widehat{r} - r) \otimes (\widehat{r} - r)] = \frac{1}{N}\Sigma_r.$$

Definition 24.1 (Local smooth chart)**.** Let $(a,b) \in \mathcal{P}(G)^2$, let $\rho > 0$, and let $\|\cdot\|_*$ be a norm on U_G. We say that T has a $C^3(M,\rho,\|\cdot\|_*)$-chart at (a,b) if there is a C^3 function

$$F_{a,b} : \mathsf{U}_G \times \mathsf{U}_G \to \mathbb{R}$$

such that

$$F_{a,b}(u,v) = \mathsf{T}(a+u, b+v)$$

whenever $a+u, b+v \in \mathcal{P}(G)$ and

$$\|u\|_* + \|v\|_* \le \rho,$$

and such that

$$\sup_{\|u\|_*+\|v\|_*\le\rho} \|D^3 F_{a,b}(u,v)\|_{*,op} \le M.$$

For three base pairs (r,s), (r,r), and (s,s), define the formal centered trace curvature

$$\begin{aligned}\mathfrak{K}_N(r,s) := {}& \mathrm{Tr}\Big(D^2_{11}F_{r,s}(0,0)\Sigma_r\Big) + \mathrm{Tr}\Big(D^2_{22}F_{r,s}(0,0)\Sigma_s\Big) \\ & - \frac{1}{2}\,\mathrm{Tr}\Big(\big[D^2_{11}F_{r,r}(0,0) + D^2_{22}F_{r,r}(0,0)\big]\Sigma_r\Big) \\ & - \frac{1}{2}\,\mathrm{Tr}\Big(\big[D^2_{11}F_{s,s}(0,0) + D^2_{22}F_{s,s}(0,0)\big]\Sigma_s\Big),\end{aligned}$$

whenever the three charts exist. Notice that no mixed derivative appears, because the two empirical measures in every term are independent.

Theorem 24.2 (Finite von Mises identity for centering)**.** *Assume that T has $C^3(M,\rho,\|\cdot\|_*)$-charts at (r,s), (r,r), and (s,s). Let*

$$Z_r = \widehat{r} - r, \quad Z'_r = \widehat{r}' - r, \quad Z_s = \widehat{s} - s, \quad Z'_s = \widehat{s}' - s$$

be independent empirical fluctuations. Let $\mathcal{E}$ be the event on which all pairs of fluctuations appearing in

$$(Z_r, Z_s), \quad (Z_r, Z'_r), \quad (Z_s, Z'_s)$$

have $\|\cdot\|_*$-sum at most ρ. Then

$$\Gamma_N(r,s) = \frac{1}{2N}\mathfrak{K}_N(r,s) + R_N(r,s),$$

with

$$\begin{aligned}|R_N(r,s)| \le CM\,\mathbb{E}\Big[&(\|Z_r\|_* + \|Z_s\|_*)^3 + (\|Z_r\|_* + \|Z'_r\|_*)^3 \\ &+ (\|Z_s\|_* + \|Z'_s\|_*)^3\Big] \\ &+ Cd^{p/2}\,\mathbb{P}(\mathcal{E}^c).\end{aligned}$$

Proof. On the chart event, Taylor expansion at the origin gives

$$\begin{aligned}F_{a,b}(U,V) = {}& F_{a,b}(0,0) + D_1F_{a,b}(0,0)[U] + D_2F_{a,b}(0,0)[V] \\ & + \frac{1}{2}D^2_{11}F_{a,b}(0,0)[U,U] + D^2_{12}F_{a,b}(0,0)[U,V] + \frac{1}{2}D^2_{22}F_{a,b}(0,0)[V,V] + \mathrm{Rem}_{a,b}(U,V),\end{aligned}$$

where

$$|\,\mathrm{Rem}_{a,b}(U,V)| \le CM(\|U\|_* + \|V\|_*)^3.$$

Taking expectations with U,V independent empirical fluctuations eliminates the first-order terms and the mixed second-order term. The remaining second-order expectations are

$$\mathbb{E}D^2_{11}F_{a,b}(0,0)[Z_a,Z_a] = \frac{1}{N}\,\mathrm{Tr}(D^2_{11}F_{a,b}(0,0)\Sigma_a),$$

and similarly for the second marginal. Applying this expansion to (r,s), (r,r), and (s,s), with coefficients $1, -1/2, -1/2$, gives the trace term. Outside the chart event, all three transport costs are bounded by $d^{p/2}$, giving the displayed tail contribution. ☐

Corollary 24.3 (Stable-face centering)**.** *Suppose that, in neighborhoods of (r,s), (r,r), and (s,s) of $\|\cdot\|_*$-radius ρ, the Kantorovich value is affine on each of the three neighborhoods. Then $\mathfrak{K}_N(r,s) = 0$, $M = 0$, and*

$$|\Gamma_N(r,s)| \le Cd^{p/2}\,\mathbb{P}(\mathcal{E}^c).$$

In particular, if

$$\mathbb{P}(\mathcal{E}^c) \lesssim h_N\big(W_p(r,s) + h_N\big)^{p-1},$$

then the centered estimator achieves the critical root rate on this stable-face class.

Proof. On an affine chart all second and third derivatives vanish. The conclusion is the previous theorem. ☐

Corollary 24.4 (Trace-curvature sufficient condition)**.** *Assume the hypotheses of Theorem 24.2 on the critical grid and suppose*

$$\frac{|\mathfrak{K}_N(r,s)|}{N} + M\,\mathbb{E}\Big[(\|Z_r\|_* + \|Z_s\|_*)^3 + (\|Z_r\|_* + \|Z'_r\|_*)^3 + (\|Z_s\|_* + \|Z'_s\|_*)^3\Big] + d^{p/2}\mathbb{P}(\mathcal{E}^c)$$

is bounded by

$$Ch_N\big(W_p(r,s) + h_N\big)^{p-1}$$

uniformly over a class $\mathcal{C}_N \subset \mathcal{P}(G_N)^2$. Then the clipped centered estimator satisfies

$$\sup_{(r,s)\in\mathcal{C}_N} \mathbb{E}|\widehat{W}_N^{\mathrm{ctr}} - W_p(r,s)| \lesssim h_N.$$

Proof. The displayed condition is precisely the centered curvature bound (CC). Apply Theorem 23.3. ☐

Remark 24.5 (Interpretation). For a genuinely stable LP face, centering leaves no bias except the probability of leaving the face. For the quadratic smooth-density theory, the same identity becomes the weighted H^{-1} trace calculation in Section 15: the chart is not affine, but the trace difference is controlled by elliptic regularity and cross-fitting. For the unrestricted Euclidean grid, the obstacle is exactly to prove that the nonsmooth charts left after diagonal centering have trace curvature no larger than the adaptive critical scale.

25 A diagonal obstruction to uncentered localized contact entropy

The localized contact method of Section 18 remains useful for structured subclasses, but it cannot be the full proof without centering. The reason is not technical: the full diagonal already violates the entropy budget required by the uncentered gap-peeling theorem.

Let $G_N \subset [0,1]^d$ be a grid with separation h_N and cardinality $m_N \asymp h_N^{-d} \asymp N \log N$. Let u_N be the uniform law on G_N. Normalize dual potentials by $\phi_{x_0} = 0$ only after the construction; the normalization changes no packing statement in the quotient by constants.

Theorem 25.1 (Exponential diagonal near-face)**.** *Fix $p \ge 1$. There exist constants $a, c > 0$, depending only on d, p, such that the optimal dual face for $T(u_N, u_N) = 0$ contains a set $\mathcal{S}_N$ of feasible dual pairs with the following properties:*

$$|\mathcal{S}_N| \ge \exp(cm_N),$$

every $a = (\phi, \psi) \in \mathcal{S}_N$ satisfies

$$L_a(u_N, u_N) = 0,$$

and distinct elements of $\mathcal{S}_N$, after quotienting by additive constants, are separated by at least ah_N^p in the $\ell_\infty(G_N)$ norm of the first potential. In particular, at powered accuracy $\varepsilon_N = h_N^p$, the near-active dual shell at the diagonal has metric entropy at least cm_N.

Proof. Choose $0 < a < 1/4$. For each sign vector $\sigma \in \{-1, 1\}^{G_N}$, set

$$\phi_x^\sigma = ah_N^p \sigma_x, \qquad \psi_x^\sigma = -ah_N^p \sigma_x.$$

If $x = y$, then $\phi_x^\sigma + \psi_y^\sigma = 0 = c(x, x)$. If $x \ne y$, grid separation gives $c(x, y) = \|x - y\|_2^p \ge h_N^p$, while

$$\phi_x^\sigma + \psi_y^\sigma = ah_N^p(\sigma_x - \sigma_y) \le 2ah_N^p \le h_N^p \le c(x, y).$$

Thus each pair is dual feasible. Moreover,

$$L_{(\phi^\sigma, \psi^\sigma)}(u_N, u_N) = m_N^{-1} \sum_x \phi_x^\sigma + m_N^{-1} \sum_x \psi_x^\sigma = 0 = T(u_N, u_N),$$

so all these pairs are optimal. By the Varshamov–Gilbert bound, there is a subset of $\{-1, 1\}^{G_N}$ of size at least $\exp(cm_N)$ whose Hamming distances are at least $m_N/4$. After quotienting by constants, two such sign fields remain separated by a constant multiple of h_N^p in ℓ_∞: otherwise all coordinate differences would be nearly constant, which is impossible when two sign fields disagree and agree on positive fractions of sites. This gives the claimed packing. ☐

Corollary 25.2 (Failure of the uncentered contact-entropy route)**.** *The uncentered gap-shell entropy hypothesis of Theorem 18.1, with $\varepsilon_N = h_N^p$, cannot hold uniformly on the full Euclidean critical grid.*

Proof. At the diagonal (u_N, u_N), Theorem 25.1 gives a near-active shell with entropy at least

$$cm_N \asymp cN \log N.$$

The budget in Theorem 18.1 is

$$N\varepsilon_N^2 = Nh_N^{2p} \asymp N^{1-2p/d}(\log N)^{-2p/d},$$

which is $o(N \log N)$ because $d > 2p$. Hence the required entropy inequality fails by a polynomial factor. ☐

Remark 25.3 (What survives). This obstruction does not invalidate the contact theorems proved earlier. It says exactly where they must be used: after a centering, margin, mass-contact, smooth-phase, or skeleton reduction has removed the diagonal sign cube. The old conjecture that raw localized contact entropy alone proves the unrestricted theorem is therefore removed; the viable full route is the centered curvature inequality (CC) or an equivalent growing-alphabet LP debiasing theorem.

26 The central finite theorem equivalent to the unrestricted law

The unrestricted continuum problem has now been reduced to a finite adaptive theorem on the critical Euclidean grid. The paper proves that theorem on several large families: exact L_1-skeletons, complete additive tree transports, sparse-shortcut geometries, catalog-compressible dual phases, low-dimensional dual manifolds, smooth semidual phase classes, sparse mass-contact plans, active-face margin classes, smooth quadratic tangent classes, and continuum lifts of the critical lower-bound cores. The two obstruction theorems identify what is still missing. The diagonal sign-cube theorem rules out a naive entropy bound on the uncentered near-active dual class. The tree-level obstruction rules out raw four-sample centering as a generic replacement for large-alphabet debiasing. A full proof must therefore combine diagonal cancellation with polynomial or regularized LP debiasing at the active Euclidean scales.

Problem 26.1 (Adaptive critical Euclidean grid LP)**.** *Let $G_N \subset [0,1]^d$ be a grid of mesh*

$$h_N \asymp (N \log N)^{-1/d}, \qquad |G_N| \asymp N \log N.$$

Given two independent samples of size N from arbitrary laws $r, s \in \mathcal{P}(G_N)$, construct an estimator $\widehat{T}_N$ of

$$T_N(r, s) := W_p(r, s)^p$$

such that

$$\sup_{r,s \in \mathcal{P}(G_N)} \frac{\mathbb{E}|\widehat{T}_N - T_N(r, s)|}{h_N \big(W_p(r, s) + h_N\big)^{p-1}} < \infty. \tag{AGrid}$$

By Theorem 22.2, a positive answer to (AGrid) gives the unrestricted continuum upper bound

$$\inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}([0,1]^d)} \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim_{d,p} (N \log N)^{-1/d}.$$

Conversely, the finite grid is a subclass of the continuum problem, and the local diagonal theorem shows that this is the correct scale at the diagonal. Thus the missing upper bound is not a continuum approximation issue; it is the adaptive estimation theory of one growing-alphabet Euclidean Kantorovich LP.

The work above leaves three precise routes.

(i) *Polynomialized multiscale LP debiasing.* The complete tree theorem proves that the correct logarithmic gain is obtained by estimating each active cut discrepancy with a large-alphabet polynomial L_1 estimator. The Euclidean analogue would decompose the Kantorovich value, up to $O(h_N(W_p + h_N)^{p-1})$, into a controlled family of signed cut, contact, or flux discrepancies and apply the same polynomial debiasing to those discrepancies. This is the most robust route because it survives the raw-centering obstruction.

(ii) *Regularized LP Richardson debiasing.* Prove uniform growing-alphabet versions of the expansion and sampling bounds in Theorem 17.4 or Theorem 21.3, at alphabet size $N \log N$, with enough orders of debiasing to match the adaptive denominator in (AGrid). This would realize polynomial debiasing implicitly through the regularized LP value.

(iii) *Special Euclidean centered curvature.* Prove the stronger signed curvature bound

$$\left|\mathbb{E}T(\widehat{r},\widehat{s}) - \frac{1}{2}\mathbb{E}T(\widehat{r},\widehat{r}') - \frac{1}{2}\mathbb{E}T(\widehat{s},\widehat{s}') - T(r,s)\right| \lesssim_{d,p} h_N \big(W_p(r,s) + h_N\big)^{p-1}. \qquad \text{(CGrid)}$$

By Theorem 23.3, this would solve (AGrid). The tree obstruction shows that such a proof cannot rely only on scalar diagonal centering; it would have to exploit genuinely Euclidean cycle geometry or hidden polynomial cancellation in the Kantorovich LP.

Conjecture 26.2 (Adaptive polynomialized Euclidean LP debiasing)**.** *For every $d > 2p$, the adaptive critical-grid bound* (AGrid) *holds uniformly over $r, s \in \mathcal{P}(G_N)$.*

Conjecture 26.3 (Uniform growing-alphabet regularized LP debiasing)**.** *There exists a regularized/debiased estimator of the critical-grid Kantorovich LP value satisfying the adaptive powered risk criterion of* Theorem 22.2 *uniformly over all $r, s \in \mathcal{P}(G_N)$.*

Either conjecture implies the full high-dimensional minimax upper law. The matching lower law is the Niles–Weed–Rigollet large-alphabet mechanism, already realized inside the paired-grid and tree lower cores.

Conjecture 26.4 (Sharp supercritical minimax law)**.** *For every $d > 2p$,*

$$\inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}([0,1]^d)} \mathbb{E}|\widehat{W} - W_p(P,Q)| \asymp_{d,p} (N \log N)^{-1/d},$$

and

$$\inf_{\widehat{W}} \sup_{P,Q} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \asymp_{d,p} (N \log N)^{-2/d}.$$

The key advance in the present reduction is that the missing theorem is now finite, adaptive, and falsifiable by explicit multiscale tests. It cannot be replaced by uncentered entropy, unweighted H^{-1} geometry, or raw centering. It must reproduce, inside the Euclidean Kantorovich LP, the same $N \log N$ polynomial debiasing that is fully proved for the dyadic additive tree transport in Theorem 7.1.

A Structural limitations of insufficient routes

Several tempting routes are insufficient as proofs of the unrestricted minimax theorem. The manuscript keeps only the parts that are either proved, converted into explicit finite criteria, or recorded as obstructions that prevent the same mistake from reappearing.

A.1 Convex-function entropy

A global covering argument for the full quadratic dual class cannot close the critical logarithm. The invalid exponent $(d-1)/2$ belongs to Hausdorff entropy of convex bodies, not to uniformly bounded convex functions. Bronshtein's theorem for convex functions gives the exponent $d/2$:

$$\log N(\mathcal{F}, \varepsilon, \|\cdot\|_\infty) \lesssim_d \varepsilon^{-d/2}.$$

Dudley's integral at exponent $d/2$ yields the smooth-cost plug-in scale $N^{-2/d}$ for W_2^2, not the critical scale $h_N^2 = (N \log N)^{-2/d}$. Therefore this route is useful for the far annulus but not for the critical finite-grid LP.

A.2 Unweighted H^{-1} geometry

The unweighted H^{-1} method is valid at the flat background but false at non-uniform densities. Theorem 16.1 gives a smooth one-mode counterexample where the deterministic tangent-metric mismatch is larger than the target strip accuracy. Consequently all quadratic strip arguments must be genuinely weighted or must use a different centered finite-LP mechanism.

A.3 Raw diagonal centering

Four-sample centering is useful because it annihilates the diagonal empirical width and leaves only $N^{-1/2}$ stochastic fluctuations. It is nevertheless not a complete large-alphabet debiasing method. Theorem 7.3 gives a single-scale obstruction inside a completely solvable tree model: at alphabet size $\asymp N$, the scalar absolute-value bias remains of constant order, and the corresponding transport scale is $N^{-1/d}$, whereas the target scale is $(N \log N)^{-1/d}$. Thus any full Euclidean proof based on centering must contain additional polynomial or regularized LP bias cancellation.

References

[1] E. M. Bronshtein. ε-entropy of convex sets and functions. *Siberian Mathematical Journal*, 17:393–398, 1976.

[2] S. Chewi, J. Niles-Weed, and P. Rigollet. *Statistical Optimal Transport.* Lecture Notes in Mathematics, Springer, 2025.

[3] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738, 2015.

[4] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485–497, 2004.

[5] J. Jiao, Y. Han, and T. Weissman. Minimax estimation of the L_1 distance. *IEEE Transactions on Information Theory*, 64(10):6672–6706, 2018.

[6] A. N. Kolmogorov and V. M. Tikhomirov. ε-entropy and ε-capacity of sets in functional spaces. *Uspekhi Matematicheskikh Nauk*, 14(2):3–86, 1959.

[7] S. Liu, F. Bunea, and J. Niles-Weed. Beyond entropic regularization: debiased Gaussian estimators for discrete optimal transport and general linear programs. *arXiv:2505.04312*, 2025.

[8] T. Manole and J. Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *Annals of Applied Probability*, 34(1B):1108–1135, 2024.

[9] J. Niles-Weed and Q. Berthet. Minimax estimation of smooth densities in Wasserstein distance. *Annals of Statistics*, 50(3):1519–1540, 2022.

[10] J. Niles-Weed and P. Rigollet. Estimation of Wasserstein distances in the spiked transport model. *Bernoulli*, 28(4):2663–2688, 2022.

[11] T. P. Papp and C. Sherlock. Centered plug-in estimation of Wasserstein distances. *arXiv:2203.11627*, revised 2025.

[12] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes.* Springer Series in Statistics, second edition material, 2023.

[13] J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.