Diagonal Exact Theory, Shrinking Matched-Template Phase Diagrams,

and Finite-Template Classes for High-Dimensional Wasserstein Distance Estimation

Keywords: Wasserstein distance; minimax estimation; optimal transport; high-dimensional statistics; dyadic trees; mixed-scale neighborhoods; matched-template phase diagrams; positive-distance neighborhoods; thin annuli; nonsmooth functional estimation.

Abstract

We study the open minimax problem of estimating the two-sample Wasserstein distance $W_p(P,Q)$ over the unrestricted class of Borel probability measures on $[0,1]^d$. For fixed dimension in the supercritical regime $d > 2p$, the published theorem-level picture still leaves a logarithmic gap between the balanced lower bound $(N \log N)^{-1/d}$ of Niles–Weed and Rigollet and the empirical plug-in upper bound $N^{-1/d}$, where $N = n \wedge m$.

Our first main theorem remains the local exact result at the center of that gap. We prove that the lower-envelope scale $(N \log N)^{-1/d}$ is already the exact minimax rate on the *full Euclidean neighborhood of the diagonal*

$$\mathcal{L}_{A,N} = \{(P,Q) : \ W_p(P,Q) \le A(N \log N)^{-1/d}\}$$

for a suitable fixed constant $A = A(d,p)$, and that the empirical plug-in estimator is still strictly suboptimal there.

We then push the off-diagonal analysis in five rigorous directions.

First, for every fixed positive baseline radius r, every $p > 1$, and every $d \ge 5$, we prove a theorem-level minimax statement on the *full Euclidean product neighborhood* $\mathcal{P}(X_r) \times \mathcal{P}(Y_r)$ of two small separated balls at distance $\asymp r$: the transport cost W_p^p has minimax absolute risk between $c_{d,p} r^p (N \log N)^{-2/d}$ and $C_{d,p} r^p N^{-2/d}$, and therefore W_p itself has minimax absolute risk between $c_{d,p} r (N \log N)^{-2/d}$ and $C_{d,p} r N^{-2/d}$.

Second, on the *full two-ball mixed-scale neighborhoods*

$$\mathcal{C}_{r,s} = \mathcal{P}(z_\diamond + sB) \times \mathcal{P}(z_\diamond + re_1 + sB), \qquad 0 < s \le \kappa_\star r,$$

we prove an exact first-plus-second-order decomposition

$$W_p(P,Q)^p = r^p + r^{p-1} s\, L(P,Q) + r^{p-2} s^2\, U_{\kappa,p}(P,Q),$$

where the residual functional $U_{\kappa,p}$ is uniformly comparable to a fixed quadratic transport cost. From this we derive a mixed-scale phase diagram on the *entire product class*: the critical width is $s_{\mathrm{crit}}(N,r) \asymp r N^{2/d-1/2}$; below it, the class is generically parametric with absolute distance risk $\asymp sN^{-1/2}$; above it, the class is generically nonparametric with absolute distance risk between $c_{d,p}(s^2/r)(N \log N)^{-2/d}$ and $C_{d,p}(s^2/r)N^{-2/d}$.

Third, we construct exact *translated paired thin-annulus classes* away from the diagonal. On these classes we prove an exact Euclidean identity

$$W_p(P_\theta, Q_\eta)^p = r^p + a_{r,\delta,p}\ \mathrm{TV}(q_\theta, q_\eta), \qquad a_{r,\delta,p} = (r^2 + \delta^2)^{p/2} - r^p,$$

which shows that direct estimation still beats empirical plug-in by a logarithmic factor far from the diagonal.

Fourth, we prove a new theorem for *all fixed finite-template multicluster families at positive baseline*. For any fixed finite source template and target template whose cross-distances stay uniformly away from zero, the exact Euclidean problem can be relabeled as a bounded smooth-cost transport problem on a fixed compact class. Consequently every such one-sided or two-sided multicluster family already enjoys the upper bound $rN^{-2/d}$ for W_p, hence lies at or below the global target scale $(N \log N)^{-1/d}$. Moreover, every such class contains a two-ball subclass, so it always inherits the lower bound $\max\{sN^{-1/2}, (s^2/r)(N \log N)^{-2/d}\}$. In particular, whenever $s \asymp r$, the minimax order on the whole finite-template class is again exact up to the same logarithm as in the smooth-cost theory.

Fifth, we quantify the first genuinely growing-complexity regime that survives the fixed-template analysis. For shrinking matched-template families with arbitrary weight profile $\alpha = (\alpha_1, \ldots, \alpha_K)$, we prove the explicit upper bound

$$\sup \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim sN^{-1/2} \sum_{i=1}^{K} \alpha_i^{1/2} + \frac{s^2}{r} N^{-2/d} \sum_{i=1}^{K} \alpha_i^{1-2/d}.$$

For equal weights this becomes

$$\sup \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim s\sqrt{\frac{K}{N}} + \frac{s^2}{r}\Big(\frac{K}{N}\Big)^{2/d}.$$

Thus even a growing matched-template family is already below the global target scale provided its effective branching complexity stays below the displayed thresholds.

Taken together with the critical-grid and linearization reductions preserved from the previous versions, these results further sharpen the unresolved core of the problem. After eliminating the diagonal neighborhood, all fixed positive-distance neighborhoods, the full two-ball mixed-scale geometry, translated thin annuli, and every fixed finite-template positive- baseline multicluster family, the remaining difficulty must involve either *growing macro-complexity*, *collapsing cross-baselines on more than one macro-edge*, or both. That is substantially closer to the actual center of the unrestricted problem than any earlier version of the paper.

Contents

1 Introduction

Let $d \in \mathbb{N}$, $p \geq 1$, and let $\mathcal{P}_d$ denote the set of all Borel probability measures on $[0,1]^d$. Given $P, Q \in \mathcal{P}_d$, the p-Wasserstein distance is

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

From independent samples

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P, \qquad Y_1, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} Q,$$

one wants to estimate the scalar functional $W_p(P,Q)$ itself.

The empirical approximation problem and the direct functional-estimation problem are not the same. By the triangle inequality,

$$|W_p(P_n, Q_m) - W_p(P,Q)| \leq W_p(P_n, P) + W_p(Q_m, Q),$$

so the empirical plug-in estimator inherits the classical curse of dimensionality from one-sample Wasserstein approximation. On compact d-dimensional spaces this yields the benchmark rate $N^{-1/d}$ in the supercritical regime $d > 2p$, where $N = n \wedge m$, and the sharp one-sample theory of Weed and Bach shows that this behavior is essentially optimal for estimating a measure in Wasserstein distance. What is *not* settled by these facts is whether the scalar functional $W_p(P,Q)$ itself can be estimated strictly better than empirical plug-in over the unrestricted class $\mathcal{P}_d$.

The strongest general lower bound currently available in fixed dimension is due to Niles–Weed and Rigollet, who proved in the balanced supercritical case $n = m = N$ that

$$\inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}_d} \mathbb{E}|\widehat{W} - W_p(P,Q)| \gtrsim (N \log N)^{-1/d} \qquad (d > 2p).$$

This leaves a logarithmic gap to the plug-in upper bound $N^{-1/d}$. Recent expositions, including the survey chapter of Chewi, Niles–Weed, and Rigollet, continue to list the unrestricted fixed-dimensional high-dimensional minimax problem as open.

The earlier versions of the present project established several exact Euclidean subclass facts: paired-block geometries on which direct estimation beats empirical plug-in; rigidity of exact one-scale total-variation reductions; and a dyadic-tree surrogate with the right logarithmic gap. The criticism that remained valid was also clear: those results still stopped short of the actual center of the unrestricted Euclidean problem.

The main new theorem of this paper attacks that center directly. Beyond the diagonal theorem, the off-diagonal part of the paper now shows that fixed positive-baseline multicluster classes are already easy, that rigid shrinking matched-template classes have the same phase diagram as a single two-ball pair, and that even growing matched-template families admit an explicit complexity-sensitive upper bound driven by the profiles $\sum_i \alpha_i^{1/2}$ and $\sum_i \alpha_i^{1-2/d}$.

Theorem (Informal main theorem)**.** *Assume $d > 2p$ and $n = m = N$. There exists a constant $A_0(d,p) > 0$ such that on the full Euclidean local class*

$$\mathcal{L}_{A_0,N} = \{(P,Q) \in \mathcal{P}_d^2 : \ W_p(P,Q) \le A_0(N \log N)^{-1/d}\},$$

the exact minimax rates are

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \asymp (N \log N)^{-1/d},$$

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \asymp (N \log N)^{-2/d}.$$

Moreover, on the same full local Euclidean class, the empirical plug-in estimator still has worst-case squared risk $\gtrsim N^{-2/d}$.

This theorem changes the status of the supercritical problem in a concrete way. The scale $(N \log N)^{-1/d}$ is no longer only a lower bound and no longer only a surrogate-tree phenomenon. It is the exact minimax scale on the *actual Euclidean model* near the diagonal.

The present version then pushes the off-diagonal analysis much further. We prove that, up to the critical scale, the full balanced supercritical continuum problem is equivalent to the same Euclidean estimation problem on the level-J dyadic grid with $2^{Jd} \asymp N \log N$ atoms, and we preserve the piecewise-affine linearization reduction of the p-cost from the previous version.

More importantly, the present manuscript now contains five additional theorem-level layers beyond the diagonal theorem.

First, for every fixed positive radius r, every $p > 1$, and every $d \ge 5$, we prove sharp upper and lower minimax envelopes (up to the same logarithm as in the smooth-cost lower theory) on the full Euclidean product class $\mathcal{P}(X_r) \times \mathcal{P}(Y_r)$ of two separated balls at distance $\asymp r$.

Second, we solve, up to that same logarithm on the nonparametric side, the complete mixed-scale picture on the full two-ball product neighborhoods

$$\mathcal{C}_{r,s} = \mathcal{P}(z_\diamond + sB) \times \mathcal{P}(z_\diamond + re_1 + sB), \qquad 0 < s \le \kappa_\star r.$$

There the transport cost admits an exact first-plus-second-order decomposition, the residual term is uniformly comparable to a quadratic transport cost, and a genuine phase transition appears at the critical width $s_{\mathrm{crit}}(N,r) \asymp rN^{2/d-1/2}$: below it, the whole class is generically parametric; above it, the whole class is generically nonparametric.

Third, we construct exact translated paired thin-annulus Euclidean classes away from the diagonal on which direct estimation still improves on empirical optimal transport by a logarithmic

factor. The mixed-scale theorem shows that these thin-annulus examples are nongeneric: their block symmetry cancels the first-order term which dominates on the full two-ball class.

Fourth, we prove that every fixed finite-template positive-baseline multicluster family, one-sided or two-sided, is already estimable at or below the global target scale. We then sharpen this further for shrinking two-sided matched-template families with rigid macro-assignment: such classes obey the same mixed-scale phase diagram as the one-pair two-ball problem.

Fifth, we quantify the first genuinely growing-complexity regime that survives the fixed-template analysis. For matched-template families with arbitrary weight profile $\alpha = (\alpha_1, \ldots, \alpha_K)$, we prove an explicit complexity-sensitive upper bound driven by $\sum_i \alpha_i^{1/2}$ and $\sum_i \alpha_i^{1-2/d}$. In particular, even a growing matched family is already below the global target scale provided its effective branching complexity stays below the corresponding thresholds.

These additions change the qualitative picture. The unresolved part of the unrestricted problem is no longer the whole off-diagonal region, no longer every positive-distance neighborhood, no longer every one-baseline mixed-scale family, no longer any fixed finite-template positive-baseline multicluster geometry, and no longer every rigid matched shrinking family. What remains is now concentrated in a genuinely two-sided branching globalization problem, in which microscopic diagonal nonsmoothness must be reconciled with collapsing or proliferating macro-geometry beyond the fixed-template and rigid-matching regimes.

Structure of the paper

(1) In Section 2, we record the current unrestricted theorem-level picture.

(2) In Sections 3 and 5, we preserve the exact Euclidean paired reduction, the direct paired estimator, and the same-class plug-in lower bound.

(3) In Section 6, we preserve the one-sided collapse theorem, the rigidity of exact one-scale Euclidean TV reductions, and the fixed-size equilateral obstruction.

(4) In Sections 7 and 8, we preserve the exact multiscale dyadic-tree surrogate and its supercritical minimax law.

(5) In Section 9, we prove the exact local Euclidean theorem on the full diagonal neighborhood.

(6) In Section 10, we record the globalization gap as it stood immediately after the diagonal theorem.

(7) In Section 11, we reduce the unrestricted balanced supercritical problem to the critical dyadic grid with $|X_J| \asymp N \log N$ atoms.

(8) In Section 12, we preserve the coarse separated-class upper bound and the first annulus reduction for absolute loss.

(9) In Section 13, we preserve the piecewise-affine linearization reduction and the splitting obstruction.

(10) In Section 14, we prove theorem-level minimax envelopes on full positive-distance Euclidean product neighborhoods.

(11) In Section 15, we prove a mixed-scale phase diagram on the full two-ball product neighborhoods $\mathcal{C}_{r,s}$, including the critical width $s_{\mathrm{crit}}(N,r) \asymp rN^{2/d-1/2}$.

(12) In Section 16, we construct exact translated paired thin-annulus classes away from the diagonal and prove an off-diagonal logarithmic plug-in gap there.

(13) In Section 17, we prove that every fixed finite-template positive-baseline multicluster family, one-sided or two-sided, is already estimable at or below the global target scale.

(14) In Section 18, we prove exact matched-pair localization and the full mixed-scale phase diagram for rigid shrinking matched-template classes.

(15) In Section 19, we quantify growing matched-template complexity through an explicit complexity-sensitive upper bound.

(16) In Section 20, we summarize what remains open after this stronger finite-template and matched-template theory.

(17) The appendices preserve the exact Gaussian kernel of the local paired Poisson channel and the almost-equilateral transport-cost transfer principle.

2 The current unrestricted theorem-level picture

For $n, m \geq 1$, define the minimax absolute and squared risks

$$\mathfrak{M}^{\mathrm{abs}}_{n,m,d,p} := \inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}_d} \mathbb{E}_{P^{\otimes n}\otimes Q^{\otimes m}} \big|\widehat{W} - W_p(P,Q)\big|,$$

$$\mathfrak{M}^{\mathrm{sq}}_{n,m,d,p} := \inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}_d} \mathbb{E}_{P^{\otimes n}\otimes Q^{\otimes m}} \big(\widehat{W} - W_p(P,Q)\big)^2.$$

Theorem 2.1 (Current theorem-level unrestricted envelopes)**.** *Fix $d \geq 1$, $p \geq 1$, and let $N := n \wedge m$.*

(i) *If $d < 2p$, then*

$$\mathfrak{M}^{\mathrm{abs}}_{n,m,d,p} \asymp N^{-1/2}, \qquad \mathfrak{M}^{\mathrm{sq}}_{n,m,d,p} \asymp N^{-1}.$$

(ii) *If $d = 2p$, then*

$$N^{-1/2} \lesssim \mathfrak{M}^{\mathrm{abs}}_{n,m,d,p} \lesssim N^{-1/2}(\log N)^{1/p},$$
$$N^{-1} \lesssim \mathfrak{M}^{\mathrm{sq}}_{n,m,d,p} \lesssim N^{-1}(\log N)^{2/p}.$$

(iii) *If $d > 2p$, then*

$$N^{-1/2} \lesssim \mathfrak{M}^{\mathrm{abs}}_{n,m,d,p} \lesssim N^{-1/d},$$
$$N^{-1} \lesssim \mathfrak{M}^{\mathrm{sq}}_{n,m,d,p} \lesssim N^{-2/d}.$$

(iv) *If $d > 2p$ and $n = m = N$, then*

$$(N\log N)^{-1/d} \lesssim \mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} \lesssim N^{-1/d},$$

$$(N\log N)^{-2/d} \lesssim \mathfrak{M}^{\mathrm{sq}}_{N,N,d,p} \lesssim N^{-2/d}.$$

Proof. The upper bounds come from the empirical plug-in estimator

$$\widehat{W}_{\mathrm{emp}} := W_p(P_n, Q_m).$$

By the triangle inequality,

$$|W_p(P_n, Q_m) - W_p(P, Q)| \le W_p(P_n, P) + W_p(Q_m, Q).$$

Taking expectations and using the sharp one-sample empirical Wasserstein bounds of Weed–Bach [8], together with the compact-support L_2 control obtained from Fournier–Guillin [3], yields the displayed upper bounds. The logarithmic boundary at $d = 2p$ is exactly the usual empirical boundary.

For the parametric lower bound, fix $z \in [0,1]^d$ and choose $a, b \in [0,1]^d$ with $\|a-z\|_2^p \neq \|b-z\|_2^p$. For $t \in [1/4, 3/4]$, define

$$P_t := t\delta_a + (1-t)\delta_b, \qquad Q := \delta_z.$$

Then

$$W_p(P_t, Q)^p = t\|a-z\|_2^p + (1-t)\|b-z\|_2^p = \alpha + \beta t$$

with $\beta \neq 0$. Hence $t \mapsto W_p(P_t, Q)$ is a smooth bi-Lipschitz function on $[1/4, 3/4]$. Estimating $W_p(P_t, Q)$ is therefore equivalent, up to constants, to estimating the Bernoulli mean t, which has minimax absolute risk of order $n^{-1/2}$ and minimax squared risk of order n^{-1}. By symmetry the same lower bound holds with m in place of n, and therefore with $N = n \wedge m$.

Finally, in the balanced supercritical regime $d > 2p$, Niles-Weed and Rigollet proved that

$$\mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} \gtrsim (N \log N)^{-1/d}$$

[6, Theorem 3 and the discussion immediately after it]. The squared lower bound then follows from Jensen's inequality. □

Remark 2.2. The unresolved part of the unrestricted fixed-dimensional theory is therefore sharply localized. Outside the balanced supercritical regime $d > 2p$, the theorem-level picture is already complete or differs only by the critical boundary logarithm. In the balanced supercritical regime, the remaining ambiguity is purely logarithmic.

3 Exact Euclidean paired reductions

3.1 Geometric paired codebooks

We begin with the geometric packing lemma needed throughout.

Lemma 3.1 (Paired packing)**.** *There exists a constant $c_d > 0$ such that the following holds. For every $M \ge 1$ and every $0 < \delta \le c_d M^{-1/d}$, one can choose $2M$ points*

$$\{x_1^+, x_1^-, \dots, x_M^+, x_M^-\} \subset [0,1]^d$$

such that

$$\|x_j^+ - x_j^-\|_2 = \delta, \qquad j = 1, \dots, M,$$

and for every distinct pairs $i \neq j$,

$$\min\{\|u-v\|_2 : \ u \in \{x_i^+, x_i^-\},\, v \in \{x_j^+, x_j^-\}\} \geq 5\delta.$$

Proof. Fix a small absolute constant $a_d > 0$ and assume $\delta \leq a_d M^{-1/d}$. Choose a grid of points $z \in [0,1]^d$ with spacing 8δ in every coordinate. For a_d small enough, the grid contains at least M points. For each selected grid point z_j, define

$$x_j^- := z_j + 2\delta e_1, \qquad x_j^+ := z_j + 3\delta e_1.$$

Then $x_j^\pm \in [0,1]^d$ and $\|x_j^+ - x_j^-\|_2 = \delta$. If $i \neq j$, the two grid anchors differ by at least 8δ in some coordinate, while the offsets inside each cell are at most 3δ along e_1, hence any point from pair i is at distance at least 5δ from any point in pair j. □

The next proposition is the exact transport identity underlying the entire paper.

Proposition 3.2 (Exact paired Euclidean reduction)**.** *Fix $M \geq 1$, $p \geq 1$, and a paired configuration as in Theorem 3.1. For $\theta = (\theta_1, \ldots, \theta_M) \in [-1,1]^M$, define*

$$P_\theta := \frac{1}{2M} \sum_{j=1}^{M} \Big[(1+\theta_j)\delta_{x_j^+} + (1-\theta_j)\delta_{x_j^-}\Big].$$

Then for every $\theta, \eta \in [-1,1]^M$,

$$W_p(P_\theta, P_\eta)^p = \frac{\delta^p}{2M} \sum_{j=1}^{M} |\theta_j - \eta_j|. \tag{3.1}$$

Equivalently, if

$$q_\theta(j,\pm) = \frac{1 \pm \theta_j}{2M}, \qquad q_\eta(j,\pm) = \frac{1 \pm \eta_j}{2M},$$

then

$$W_p(P_\theta, P_\eta)^p = \delta^p \,\mathrm{TV}(q_\theta, q_\eta) = \frac{\delta^p}{2}\|q_\theta - q_\eta\|_1. \tag{3.2}$$

Proof. Both P_θ and P_η place total mass $1/M$ on every pair $\{x_j^+, x_j^-\}$. Let π be any coupling of P_θ and P_η, and collapse each pair to one macro-node. The induced off-diagonal pair-to-pair flow is a circulation because every pair has the same total source mass and the same total target mass. Any circulation decomposes into directed cycles. Consider one cycle $j_1 \to j_2 \to \cdots \to j_r \to j_1$ carrying mass $\varepsilon > 0$. Each unit of this cycle pays at least $(5\delta)^p$ on every edge, hence its total cost is at least $r\varepsilon(5\delta)^p$.

Now remove this cycle and reroute the same mass inside each participating pair. Within pair j_k, the cycle removes ε units from one site and delivers ε units to the other site, so it can be replaced by an internal transfer of at most ε units across the two points of the pair. The cost of this local replacement in pair j_k is at most $\varepsilon\delta^p$. Thus the whole cycle can be replaced by a localized coupling of cost at most $r\varepsilon\delta^p$, which is strictly smaller because $(5\delta)^p > \delta^p$.

Eliminating cycles one by one shows that some optimal coupling is completely localized within the pairs. Once localization is known, the transport problem decouples into M independent

two-point problems. In pair j, the source masses are $\frac{1+\theta_j}{2M}$ and $\frac{1-\theta_j}{2M}$, and the target masses are $\frac{1+\eta_j}{2M}$ and $\frac{1-\eta_j}{2M}$. Hence the amount of mass that must cross the edge of length δ is

$$\frac{1}{2M}|\theta_j - \eta_j|.$$

Summing the costs over j proves (3.1). The identity (3.2) is immediate from

$$\|q_\theta - q_\eta\|_1 = \sum_{j=1}^{M}\Big(\frac{|\theta_j - \eta_j|}{2M} + \frac{|\theta_j - \eta_j|}{2M}\Big) = \frac{1}{M}\sum_{j=1}^{M}|\theta_j - \eta_j|. \qquad \square$$

Remark 3.3. The paired class is exact because the block masses are preserved pair by pair. This is the feature that allows one to import discrete L_1-functional estimation without any approximation loss whatsoever.

4 A direct paired estimator on the exact Euclidean class

4.1 Transfer of the discrete L1-distance estimator

The exact reduction above converts the Euclidean transport problem into a discrete total-variation estimation problem on an alphabet of size $2M$. We now transfer the minimax-optimal L_1-distance estimator of Jiao–Han–Weissman.

For convenience, we isolate the external input we use.

Theorem 4.1 (Jiao–Han–Weissman, unknown Q case)**.** *There exists a universal constant $C_0 > 0$ with the following property. Let $S \geq 2$, let $N \geq 2$, and assume $\log N \leq C_0 \log S$. Then there exists an estimator $\widehat{L} = \widehat{L}(X_1, \ldots, X_N, Y_1, \ldots, Y_N)$ such that for every pair of distributions r, s on an S-point alphabet,*

$$\mathbb{E}_{r^{\otimes N} \otimes s^{\otimes N}}\Big[\big(\widehat{L} - \|r - s\|_1\big)^2\Big] \leq C\,\frac{S}{N \log N}$$

for a universal constant $C < \infty$.

Theorem 4.2 (Direct paired estimator)**.** *Fix $p \geq 1$, $d \in \mathbb{N}$, $M \geq 2$, and $0 < \delta \leq c_d M^{-1/d}$. Let $\mathcal{G}^{\mathrm{pair}}_{M,\delta}$ be the Euclidean paired class*

$$\mathcal{G}^{\mathrm{pair}}_{M,\delta} = \{P_\theta : \ \theta \in [-1,1]^M\} \subset \mathcal{P}_d$$

built from Theorem 3.1. Let $n, m \geq 2$, set $N := \min(n, m)$, and assume $\log N \leq C_0 \log(2M)$, where C_0 is the constant from Theorem 4.1. Then there exists an estimator $\widehat{W}$ such that

$$\sup_{P,Q \in \mathcal{G}^{\mathrm{pair}}_{M,\delta}} \mathbb{E}(\widehat{W}^p - W_p(P,Q)^p)^2 \leq C_p\,\delta^{2p}\frac{M}{N \log N}, \tag{4.1}$$

and consequently

$$\sup_{P,Q \in \mathcal{G}^{\mathrm{pair}}_{M,\delta}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \leq C'_p\,\delta\Big(\frac{M}{N \log N}\Big)^{1/(2p)}, \tag{4.2}$$

$$\sup_{P,Q \in \mathcal{G}^{\mathrm{pair}}_{M,\delta}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \leq C''_p\,\delta^2\Big(\frac{M}{N \log N}\Big)^{1/p}. \tag{4.3}$$

Proof. Discard extra samples if $n \neq m$, and keep only N samples from each distribution. By Theorem 3.2, every $P_\theta \in \mathcal{G}^{\text{pair}}_{M,\delta}$ corresponds exactly to a distribution q_θ on the $2M$-point alphabet

$$\mathcal{A}_M := \{(j,+),(j,-) : j = 1,\dots,M\}$$

through

$$q_\theta(j,\pm) = \frac{1 \pm \theta_j}{2M}.$$

The observed X-sample and Y-sample are therefore simply i.i.d. samples from two unknown alphabet distributions q_θ and q_η belonging to a subset of Δ_{2M}.

Apply Theorem 4.1 with $S = 2M$ to obtain an estimator $\widehat{L}$ of $\|q_\theta - q_\eta\|_1$ satisfying

$$\sup_{\theta,\eta} \mathbb{E}\big(\widehat{L} - \|q_\theta - q_\eta\|_1\big)^2 \le C\,\frac{2M}{N\log N}.$$

Define

$$\widehat{V} := \frac{\delta^p}{2}\,\widehat{L}, \qquad V(\theta,\eta) := W_p(P_\theta,P_\eta)^p.$$

Then Equation (3.2) yields

$$V(\theta,\eta) = \frac{\delta^p}{2}\|q_\theta - q_\eta\|_1,$$

and therefore

$$\sup_{\theta,\eta} \mathbb{E}(\widehat{V} - V(\theta,\eta))^2 \le C_p\,\delta^{2p}\frac{M}{N\log N}.$$

This proves (4.1) after clipping $\widehat{V}$ to the interval $[0,\delta^p]$, which can only reduce the risk because $0 \le V(\theta,\eta) \le \delta^p$.

Now define

$$\widehat{W} := \widehat{V}^{1/p}.$$

For $x, y \ge 0$, the map $t \mapsto t^{1/p}$ is $1/p$-Hölder:

$$|x^{1/p} - y^{1/p}| \le |x-y|^{1/p}.$$

Hence

$$|\widehat{W} - W_p(P_\theta,P_\eta)| \le |\widehat{V} - V(\theta,\eta)|^{1/p}.$$

Using Lyapunov's inequality with exponents $1/p \le 2$ and $2/p \le 2$, we obtain

$$\mathbb{E}|\widehat{W} - W_p| \le \Big(\mathbb{E}|\widehat{V} - V|^2\Big)^{1/(2p)}$$

and

$$\mathbb{E}|\widehat{W} - W_p|^2 = \mathbb{E}|\widehat{V} - V|^{2/p} \le \Big(\mathbb{E}|\widehat{V} - V|^2\Big)^{1/p}.$$

Substituting the bound for $\widehat{V}$ proves (4.2) and (4.3). □

4.2 Geometric tuning

Two tunings are especially important.

Corollary 4.3 (Strict logarithmic improvement at the geometric scale $M \asymp N$)**.** *Fix $d > 2p$ and $p \geq 1$, and consider the balanced case $n = m = N$. Then there exists an exact Euclidean paired subclass $\mathcal{G}_N^{\mathrm{geom}} \subset \mathcal{P}_d$ such that*

$$\sup_{P,Q \in \mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim_{d,p} N^{-1/d} (\log N)^{-1/(2p)},$$

and

$$\sup_{P,Q \in \mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \lesssim_{d,p} N^{-2/d} (\log N)^{-1/p}.$$

Proof. Choose $M \asymp N$, and then choose $\delta \asymp M^{-1/d} \asymp N^{-1/d}$ using Theorem 3.1. The displayed bounds are exactly (4.2) and (4.3). □

Corollary 4.4 (The lower-envelope tuning $M \asymp N \log N$)**.** *Fix $d > 2p$, $p \geq 1$, and consider the balanced case $n = m = N$. Then there exists an exact Euclidean paired subclass $\mathcal{G}_N^{\mathrm{crit}} \subset \mathcal{P}_d$ such that*

$$\sup_{P,Q \in \mathcal{G}_N^{\mathrm{crit}}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim_{d,p} (N \log N)^{-1/d},$$

and

$$\sup_{P,Q \in \mathcal{G}_N^{\mathrm{crit}}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \lesssim_{d,p} (N \log N)^{-2/d}.$$

Proof. Choose $M \asymp N \log N$ and $\delta \asymp M^{-1/d}$. Then $M/(N \log N) \asymp 1$, so (4.2) and (4.3) reduce to the claimed bounds. □

Remark 4.5. Theorem 4.4 reaches the same $(N \log N)^{-1/d}$ scale as the strongest presently known unrestricted lower-envelope in the balanced supercritical regime. This does *not* settle the unrestricted minimax rate, because the theorem concerns an explicit exact Euclidean subclass. But it does show that the currently known lower-envelope scale already appears inside a completely rigorous Euclidean model for direct functional estimation.

5 The empirical plug-in estimator is strictly worse on the same exact Euclidean class

The direct estimator above would be far less significant if the empirical plug-in estimator enjoyed the same risk on the paired class. The next theorem shows that this is not the case.

Lemma 5.1 (A lower bound on empirical total variation at the uniform point)**.** *Let U_M be the uniform distribution on an alphabet of size $2M$, and let $\widehat{U}_N, \widehat{U}'_N$ be two independent empirical measures based on N samples from U_M. Then*

$$\mathbb{E}\,\mathrm{TV}(\widehat{U}_N, \widehat{U}'_N) = \frac{M}{N}\,\mathbb{E}|B - B'|,$$

where B, B' are i.i.d. $\mathrm{Bin}(N, \frac{1}{2M})$*. In particular,*

$$\mathbb{E}\,\mathrm{TV}(\widehat{U}_N, \widehat{U}'_N) \geq \left(1 - \frac{1}{2M}\right)^{2N-1}. \tag{5.1}$$

Hence, whenever $M \geq cN$ for a fixed constant $c > 0$,

$$\mathbb{E}\,\mathrm{TV}(\widehat{U}_N, \widehat{U}'_N) \geq c' > 0$$

for a constant c' depending only on c.

Proof. Let X_i, Y_i denote the counts of the i-th symbol in the two empirical samples. Then

$$\mathrm{TV}(\widehat{U}_N, \widehat{U}'_N) = \frac{1}{2}\sum_{i=1}^{2M}\left|\frac{X_i}{N} - \frac{Y_i}{N}\right|.$$

By linearity of expectation and symmetry of the symbols,

$$\mathbb{E}\,\mathrm{TV}(\widehat{U}_N, \widehat{U}'_N) = \frac{2M}{2N}\mathbb{E}|B - B'| = \frac{M}{N}\mathbb{E}|B - B'|,$$

where B, B' are i.i.d. $\mathrm{Bin}(N, \frac{1}{2M})$.

For the lower bound, note that

$$\mathbb{E}|B - B'| \geq \mathbb{P}(B = 1, B' = 0) + \mathbb{P}(B = 0, B' = 1) = 2\mathbb{P}(B = 1)\mathbb{P}(B' = 0).$$

Since

$$\mathbb{P}(B = 1) = N\frac{1}{2M}\Big(1 - \frac{1}{2M}\Big)^{N-1}, \qquad \mathbb{P}(B = 0) = \Big(1 - \frac{1}{2M}\Big)^{N},$$

we obtain

$$\mathbb{E}|B - B'| \geq \frac{N}{M}\Big(1 - \frac{1}{2M}\Big)^{2N-1}.$$

Multiplying by M/N proves (5.1). If $M \geq cN$, then $\frac{1}{2M} \leq \frac{1}{2cN}$, and

$$\Big(1 - \frac{1}{2M}\Big)^{2N-1} \geq \Big(1 - \frac{1}{2cN}\Big)^{2N-1} \to e^{-1/c},$$

so the expectation stays bounded away from zero uniformly in N. □

Theorem 5.2 (Plug-in lower bound on the exact paired class)**.** *Fix $d \in \mathbb{N}$, $p \geq 1$, and consider the balanced case $n = m = N$. Let $\mathcal{G}^{\mathrm{pair}}_{M,\delta}$ be the paired class from Theorem 4.2, and let $Q_0 \in \mathcal{G}^{\mathrm{pair}}_{M,\delta}$ denote the uniform paired point*

$$Q_0 = \frac{1}{2M}\sum_{j=1}^{M}(\delta_{x_j^+} + \delta_{x_j^-}).$$

Then the empirical plug-in estimator obeys

$$\mathbb{E}_{Q_0^{\otimes N}\otimes Q_0^{\otimes N}} W_p(P_N, Q_N) \geq \delta\left(1 - \frac{1}{2M}\right)^{2N-1}, \tag{5.2}$$

and therefore

$$\mathbb{E}_{Q_0^{\otimes N}\otimes Q_0^{\otimes N}} W_p(P_N, Q_N)^2 \geq \delta^2\left(1 - \frac{1}{2M}\right)^{4N-2}. \tag{5.3}$$

In particular, if $M \geq cN$, then

$$\mathbb{E}W_p(P_N, Q_N) \gtrsim_c \delta, \qquad \mathbb{E}W_p(P_N, Q_N)^2 \gtrsim_c \delta^2.$$

Proof. The minimal nonzero distance among the support points of Q_0 is δ. Hence, for any two measures μ, ν supported on this $2M$-point set,

$$W_p(\mu,\nu)^p \geq \delta^p \,\mathrm{TV}(\mu,\nu).$$

Since $0 \leq \mathrm{TV}(\mu,\nu) \leq 1$ and $p \geq 1$, we also have $\mathrm{TV}(\mu,\nu)^{1/p} \geq \mathrm{TV}(\mu,\nu)$, so

$$W_p(\mu,\nu) \geq \delta \;\mathrm{TV}(\mu,\nu).$$

Apply this to $\mu = P_N$ and $\nu = Q_N$, where P_N, Q_N are the two empirical measures built from independent samples from Q_0. Taking expectations and invoking Theorem 5.1 yields (5.2). Finally,

$$\mathbb{E}W_p(P_N,Q_N)^2 \geq \big(\mathbb{E}W_p(P_N,Q_N)\big)^2,$$

which proves (5.3). □

Corollary 5.3 (Strict plug-in separation on a fixed-dimensional Euclidean subclass)**.** *Fix $d > 2p$ and $p \geq 1$, and consider the balanced case $n = m = N$. Then there exists an exact Euclidean paired subclass $\mathcal{G}_N^{\mathrm{geom}} \subset \mathcal{P}_d$ such that:*

(i) *there exists a direct estimator with*

$$\sup_{P,Q\in\mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim_{d,p} N^{-1/d}(\log N)^{-1/(2p)},$$

$$\sup_{P,Q\in\mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \lesssim_{d,p} N^{-2/d}(\log N)^{-1/p};$$

(ii) *the empirical plug-in estimator satisfies*

$$\sup_{P,Q\in\mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}|W_p(P_N,Q_N) - W_p(P,Q)| \gtrsim_{d,p} N^{-1/d},$$

$$\sup_{P,Q\in\mathcal{G}_N^{\mathrm{geom}}} \mathbb{E}\big(W_p(P_N,Q_N) - W_p(P,Q)\big)^2 \gtrsim_{d,p} N^{-2/d}.$$

Therefore direct functional estimation is strictly better than empirical optimal transport on the same exact Euclidean high-dimensional subclass by factors

$$(\log N)^{-1/(2p)} \quad \textit{in absolute risk,} \qquad (\log N)^{-1/p} \quad \textit{in squared risk.}$$

Proof. Use Theorem 4.3 for the upper bound and Theorem 5.2 with $M \asymp N$ and $\delta \asymp N^{-1/d}$ for the lower bound on the plug-in estimator at the point $(Q_0, Q_0) \in \mathcal{G}_N^{\mathrm{geom}} \times \mathcal{G}_N^{\mathrm{geom}}$. □

6 One-sided cones, exact one-scale reductions, and a statistical obstruction

This section records two short exact facts that remain correct and useful.

6.1 One-sided paired cones are parametric

The improvement in Theorem 4.2 relies on the genuinely sign-indefinite local geometry. If the difference is constrained to one side, the problem collapses.

Proposition 6.1 (One-sided paired cones collapse)**.** *In the paired model of Theorem 3.2, assume that $\theta_j \geq \eta_j$ for every j. Then*

$$W_p(P_\theta, P_\eta)^p = \delta^p \left(\sum_{j=1}^{M} q_\theta(j,+) - \sum_{j=1}^{M} q_\eta(j,+) \right),$$

so the target depends only on the total plus-mass. Consequently, in this one-sided cone, $W_p(P_\theta, P_\eta)^p$ is estimable at the parametric scale $N^{-1/2}$ from a single difference of binomial proportions.

Proof. If $\theta_j \geq \eta_j$ for all j, then $|\theta_j - \eta_j| = \theta_j - \eta_j$, so

$$W_p(P_\theta, P_\eta)^p = \frac{\delta^p}{2M} \sum_{j=1}^{M} (\theta_j - \eta_j).$$

But

$$q_\theta(j,+) - q_\eta(j,+) = \frac{\theta_j - \eta_j}{2M},$$

hence the displayed identity follows after summation over j. The total plus-count in each sample is binomial, so estimating the difference of plus-masses is a one-dimensional parametric problem. □

6.2 Rigidity of exact one-scale TV reductions

Earlier codebook constructions led to the following exact rigidity theorem.

Proposition 6.2 (Rigidity of exact one-scale Euclidean TV reductions)**.** *Let $X = \{x_1, \ldots, x_S\} \subset \mathbb{R}^d$, and write*

$$F_X(q) := \sum_{i=1}^{S} q_i \, \delta_{x_i}, \qquad q \in \Delta_S.$$

Assume that there exists $c > 0$ such that

$$W_p(F_X(q), F_X(r))^p = c \ \mathrm{TV}(q,r) \qquad \textit{for every } q, r \in \Delta_S.$$

Then X is equilateral:

$$\|x_i - x_j\|_2^p = c \qquad \textit{for all } i \neq j.$$

In particular, $S \leq d+1$.

Proof. Take $q = \delta_i$ and $r = \delta_j$ for $i \neq j$. Then

$$W_p(F_X(q), F_X(r))^p = \|x_i - x_j\|_2^p, \qquad \mathrm{TV}(q,r) = 1.$$

Hence $\|x_i - x_j\|_2^p = c$ for every $i \neq j$, so the support is equilateral. It is classical that an equilateral set in $\mathbb{R}^d$ has size at most $d+1$. □

Remark 6.3. Theorem 6.2 explains why exact large-support Euclidean reductions cannot come from a single deterministic one-scale support. The paired construction avoids this obstruction by using a block-preserving geometry rather than a universal support identity.

6.3 Equilateral simplex classes and a new one-scale obstruction

For a finite codebook $X = \{x_1, \dots, x_M\} \subset [0,1]^d$, write

$$F_X(q) := \sum_{i=1}^{M} q_i \, \delta_{x_i}, \qquad q \in \Delta_M,$$

and

$$\mathcal{P}(X) := \{F_X(q) : \, q \in \Delta_M\} \subset \mathcal{P}_d.$$

Proposition 6.4 (Exact equilateral reduction)**.** *Assume that $X = \{x_1, \dots, x_M\} \subset [0,1]^d$ is equilateral with common separation Δ, meaning that*

$$\|x_i - x_j\|_2 = \Delta \qquad (i \neq j).$$

Then for every $q, r \in \Delta_M$,

$$W_p(F_X(q), F_X(r))^p = \Delta^p \, \mathrm{TV}(q,r) = \frac{\Delta^p}{2} \|q - r\|_1.$$

Proof. Let $P := F_X(q)$ and $Q := F_X(r)$, and write

$$m_i := \min\{q_i, r_i\}, \qquad m := \sum_{i=1}^{M} m_i.$$

The overlapping mass $\sum_i m_i \delta_{x_i}$ can be coupled at zero cost. The remaining unmatched mass is

$$1 - m = \frac{1}{2} \sum_{i=1}^{M} |q_i - r_i| = \mathrm{TV}(q,r).$$

Every unit of this residual mass must move between two distinct support points, and every such move costs exactly Δ^p. Hence every feasible transport plan has cost at least $\Delta^p \, \mathrm{TV}(q,r)$, and the obvious plan which keeps the overlap fixed and moves all residual mass between distinct atoms attains this bound. □

Theorem 6.5 (Fixed-size equilateral codebooks have exact rate N to the minus one over p)**.** *Let $X \subset [0,1]^d$ be equilateral of size $M \geq 2$ and common separation Δ, and let $N := n \wedge m$. Then*

$$c_p \, \Delta^2 N^{-1/p} \leq \inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}(X)} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \leq C_{M,p} \, \Delta^2 N^{-1/p}.$$

Moreover,

$$c'_p \, \Delta N^{-1/(2p)} \leq \inf_{\widehat{W}} \sup_{P,Q \in \mathcal{P}(X)} \mathbb{E}|\widehat{W} - W_p(P,Q)| \leq C'_{M,p} \, \Delta N^{-1/(2p)}.$$

Proof. We begin with the upper bound. Discard extra samples if necessary and retain N observations from each distribution. From the samples, form the empirical histograms $\widehat{q}, \widehat{r}$ on the support X, and define

$$\widehat{T} := \mathrm{TV}(\widehat{q}, \widehat{r}), \qquad T := \mathrm{TV}(q, r), \qquad \widehat{W} := \Delta \widehat{T}^{1/p}.$$

By Theorem 6.4,

$$W_p(F_X(q), F_X(r)) = \Delta T^{1/p}.$$

Now

$$|\widehat{T} - T| \leq \mathrm{TV}(\widehat{q}, q) + \mathrm{TV}(\widehat{r}, r).$$

Also,

$$\mathrm{TV}(\widehat{q}, q)^2 = \frac{1}{4}\|\widehat{q} - q\|_1^2 \leq \frac{M}{4}\|\widehat{q} - q\|_2^2.$$

Since $\widehat{q}$ is the empirical histogram of N i.i.d. samples from q,

$$\mathbb{E}\|\widehat{q} - q\|_2^2 = \sum_{i=1}^{M} \mathrm{Var}(\widehat{q}_i) = \frac{1 - \sum_i q_i^2}{N} \leq \frac{1}{N}.$$

Hence

$$\sup_q \mathbb{E}\,\mathrm{TV}(\widehat{q}, q)^2 \leq \frac{M}{4N}, \qquad \sup_r \mathbb{E}\,\mathrm{TV}(\widehat{r}, r)^2 \leq \frac{M}{4N},$$

and therefore

$$\sup_{q,r} \mathbb{E}|\widehat{T} - T|^2 \leq \frac{M}{N}. \tag{6.1}$$

For $\alpha := 1/p \in (0, 1]$, the map $x \mapsto x^\alpha$ is subadditive on $[0, \infty)$, so

$$|\widehat{T}^\alpha - T^\alpha| \leq |\widehat{T} - T|^\alpha.$$

Consequently,

$$|\widehat{W} - W_p(F_X(q), F_X(r))|^2 \leq \Delta^2 |\widehat{T} - T|^{2/p}.$$

Since $2/p \leq 2$, Lyapunov's inequality and (6.1) give

$$\sup_{q,r} \mathbb{E}(\widehat{W} - W_p(F_X(q), F_X(r)))^2 \leq \Delta^2 \Big(\sup_{q,r} \mathbb{E}|\widehat{T} - T|^2\Big)^{1/p} \leq \Delta^2 \Big(\frac{M}{N}\Big)^{1/p}.$$

This proves the squared upper bound. The absolute upper bound follows from Jensen:

$$\mathbb{E}|\widehat{W} - W_p| \leq \big(\mathbb{E}(\widehat{W} - W_p)^2\big)^{1/2}.$$

For the lower bound, it suffices to fix one of the two samples. Choose two distinct support points, relabel them as x_1, x_2, and define

$$r_0 := \Big(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0\Big), \qquad q_\varepsilon := \Big(\frac{1}{2} + \varepsilon, \frac{1}{2} - \varepsilon, 0, \ldots, 0\Big)$$

for $0 < \varepsilon \leq 1/4$. Then

$$\mathrm{TV}(q_\varepsilon, r_0) = \varepsilon, \qquad W_p(F_X(q_\varepsilon), F_X(r_0)) = \Delta \varepsilon^{1/p}.$$

If the Y-sample is drawn from the fixed distribution $F_X(r_0)$, the problem reduces to estimating the scalar parameter $\Delta\varepsilon^{1/p}$ from N i.i.d. draws of a Bernoulli-type experiment.

Let P_0 denote the law of the X-sample when $q = r_0$, and let P_1 denote the law when $q = q_\varepsilon$. Then

$$\mathrm{KL}(P_1 \| P_0) = N \ \mathrm{KL}\Big(\mathrm{Bern}\Big(\frac{1}{2} + \varepsilon\Big) \Big\| \mathrm{Bern}\Big(\frac{1}{2}\Big)\Big).$$

For $0 < \varepsilon \le 1/4$, the Bernoulli divergence is bounded by $4\varepsilon^2$, hence

$$\mathrm{KL}(P_1 \| P_0) \le 4N\varepsilon^2.$$

Choose $\varepsilon = c_0 N^{-1/2}$ with $c_0 > 0$ small enough that $\mathrm{KL}(P_1 \| P_0) \le 1/4$. By Pinsker's inequality, $\mathrm{TV}(P_0, P_1) \le 1/2$. A standard two-point testing argument therefore gives

$$\inf_{\widehat{W}} \sup_{i \in \{0,1\}} \mathbb{E}_i(\widehat{W} - W_i)^2 \ge \frac{(W_1 - W_0)^2}{16} = \frac{\Delta^2 \varepsilon^{2/p}}{16} \gtrsim_p \Delta^2 N^{-1/p},$$

where $W_0 = 0$ and $W_1 = \Delta\varepsilon^{1/p}$. This proves the squared lower bound.

For absolute loss, use the same two-point pair P_0, P_1. If ϕ is the test that chooses the closer of W_0 and W_1, then on the event of testing error the absolute estimation error is at least $(W_1 - W_0)/2$. Hence

$$\frac{1}{2}\sum_{i=0}^{1} \mathbb{E}_i|\widehat{W} - W_i| \ge \frac{W_1 - W_0}{4}\big(\mathbb{P}_0(\phi = 1) + \mathbb{P}_1(\phi = 0)\big) \ge \frac{W_1 - W_0}{4}\big(1 - \mathrm{TV}(P_0, P_1)\big).$$

Since $\mathrm{TV}(P_0, P_1) \le 1/2$, we obtain

$$\inf_{\widehat{W}} \sup_{i \in \{0,1\}} \mathbb{E}_i|\widehat{W} - W_i| \gtrsim_p \Delta\varepsilon^{1/p} \gtrsim_p \Delta N^{-1/(2p)}.$$

This proves the absolute lower bound. □

Corollary 6.6 (Exact deterministic one-scale reductions are statistically too easy)**.** *Assume that $d > 2p$ and that a deterministic support family $X = \{x_1, \dots, x_S\} \subset [0,1]^d$ satisfies*

$$W_p(F_X(q), F_X(r))^p = c \ \mathrm{TV}(q, r) \qquad \textit{for all } q, r \in \Delta_S.$$

Then $S \le d + 1$ by Theorem 6.2, and

$$\inf_{\widehat{W}} \sup_{q,r \in \Delta_S} \mathbb{E}\big(\widehat{W} - W_p(F_X(q), F_X(r))\big)^2 \le C_{d,p} N^{-1/p}.$$

Since $d > 2p$ implies $1/p > 2/d$, we have

$$N^{-1/p} = o(N^{-2/d}).$$

Therefore no exact deterministic one-scale Euclidean identity of the form $W_p^p = c$ TV can, in fixed dimension, realize the supercritical unrestricted difficulty suggested by the empirical upper envelope.

Proof. By Theorem 6.2, the support must be equilateral, with size $S \le d+1$ and common separation $\Delta = c^{1/p} \le \operatorname{diam}([0,1]^d) \le \sqrt{d}$. The upper bound therefore follows from Theorem 6.5:

$$\inf_{\widehat{W}} \sup_{q,r \in \Delta_S} \mathbb{E}(\widehat{W} - W_p(F_X(q), F_X(r)))^2 \le \Delta^2 \Big(\frac{S}{N}\Big)^{1/p} \le d\,\Big(\frac{d+1}{N}\Big)^{1/p}.$$

The asymptotic comparison with $N^{-2/d}$ is immediate from $d > 2p$. □

Remark 6.7. Theorem 6.6 is stronger than the geometric rigidity statement alone. It says that fixed-dimensional exact one-scale deterministic support reductions are not merely too small to generate large alphabets; even on their own natural statistical scale they are faster than the supercritical unrestricted plug-in envelope. This pushes the hard part of the fixed-d problem toward genuinely multiscale or genuinely nonlocal transport families.

7 Dyadic cubes, the ultrametric surrogate, and the Euclidean comparison

For $j \ge 0$, let $\mathcal{D}_j$ denote the standard dyadic partition of $[0,1]^d$ into 2^{jd} half-open cubes of side length 2^{-j}. Thus $\mathcal{D}_0 = \{[0,1]^d\}$.

For distinct $x, y \in [0,1]^d$, let

$$k(x,y) := \min\{j \ge 1 : \ x \text{ and } y \text{ lie in distinct cubes of } \mathcal{D}_j\},$$

and define the dyadic ultrametric

$$\rho(x,y) := 2^{-k(x,y)}, \qquad \rho(x,x) := 0.$$

The corresponding p-Wasserstein distance is denoted

$$W_{\rho,p}(P,Q) = \Big(\inf_{\pi \in \Pi(P,Q)} \int \rho(x,y)^p \, d\pi(x,y)\Big)^{1/p}.$$

It is convenient to work with the p-th power

$$T_\rho(P,Q) := W_{\rho,p}(P,Q)^p.$$

For a cube $Q \in \mathcal{D}_j$, write $\Delta_Q(P,Q) = P(Q) - Q(Q)$. Set

$$V_j(P,Q) := \sum_{R \in \mathcal{D}_j} |\Delta_R(P,Q)|.$$

Finally define

$$a_j := \frac{1}{2}\Big(2^{-jp} - 2^{-(j+1)p}\Big) = \frac{1-2^{-p}}{2}\, 2^{-jp}.$$

Proposition 7.1 (Exact dyadic tree formula)**.** *For every $P, Q \in \mathcal{P}_d$,*

$$T_\rho(P,Q) = \sum_{j \ge 1} a_j \, V_j(P,Q).$$

Moreover, for every $J \geq 1$,

$$0 \leq T_\rho(P,Q) - \sum_{j=1}^{J} a_j V_j(P,Q) \leq 2 \sum_{j>J} a_j = 2^{-(J+1)p}.$$

Proof. The metric ρ^p is exactly the path metric on the complete rooted dyadic tree whose level-j edges carry weight a_j: if two leaves first split at level k, then the tree path between them contains two copies of every edge weight a_j for $j \geq k$, and therefore

$$2 \sum_{j \geq k} a_j = 2^{-kp} = \rho(x,y)^p.$$

For weighted tree metrics, the 1-Wasserstein cost equals the sum, over edges, of edge length times absolute subtree imbalance. Grouping edges by level gives

$$T_\rho(P,Q) = \sum_{j \geq 1} a_j \sum_{R \in \mathcal{D}_j} |P(R) - Q(R)| = \sum_{j \geq 1} a_j V_j(P,Q).$$

Since $0 \leq V_j(P,Q) \leq 2$, the tail bound follows immediately:

$$0 \leq \sum_{j>J} a_j V_j(P,Q) \leq 2 \sum_{j>J} a_j = 2^{-(J+1)p}.$$ □

The Euclidean geometry of the dyadic grid is uniformly comparable to ρ.

For $J \geq 1$, let X_J be the set of centers of the cubes in $\mathcal{D}_J$.

Proposition 7.2 (Pointwise comparison on the dyadic grid)**.** *Let $J \geq 1$, and let $x, y \in X_J$. Then*

$$\rho(x,y) \leq \|x-y\|_2 \leq 2\sqrt{d}\,\rho(x,y).$$

Consequently, for all $P, Q \in \mathcal{P}(X_J)$,

$$W_{\rho,p}(P,Q) \leq W_p(P,Q) \leq 2\sqrt{d}\,W_{\rho,p}(P,Q).$$

Proof. Let $k = k(x,y)$, so x and y lie in the same cube of $\mathcal{D}_{k-1}$ but in distinct cubes of $\mathcal{D}_k$. Since x and y are centers of level-J dyadic cubes, some coordinate of $x - y$ differs by at least 2^{-k}, hence $\|x-y\|_2 \geq 2^{-k} = \rho(x,y)$. Also x, y lie in one common level-$(k-1)$ cube, whose diameter is $\sqrt{d}\,2^{-(k-1)}$, so

$$\|x-y\|_2 \leq \sqrt{d}\,2^{-(k-1)} = 2\sqrt{d}\,2^{-k} = 2\sqrt{d}\,\rho(x,y).$$

The Wasserstein comparison is obtained by integrating the pointwise ground-metric inequalities over any coupling and then taking the infimum. □

8 The exact supercritical picture on the dyadic tree surrogate

Throughout this section we focus on the balanced case $n = m = N$, and we assume

$$d > 2p.$$

8.1 External discrete input

We again use the Jiao–Han–Weissman theorem stated earlier as Theorem 4.1.

Lemma 8.1 (Empirical L_1-bound on a finite alphabet)**.** *Let r, s be distributions on an S-point alphabet, and let $\widehat{r}_N, \widehat{s}_N$ be their empirical measures based on N samples each. Then*

$$\mathbb{E}\Big(\|\widehat{r}_N - \widehat{s}_N\|_1 - \|r - s\|_1\Big)^2 \le \frac{4S}{N}.$$

Proof. By the triangle inequality,

$$\big|\|\widehat{r}_N - \widehat{s}_N\|_1 - \|r - s\|_1\big| \le \|\widehat{r}_N - r\|_1 + \|\widehat{s}_N - s\|_1.$$

Hence

$$\mathbb{E}(\cdots)^2 \le 2\mathbb{E}\|\widehat{r}_N - r\|_1^2 + 2\mathbb{E}\|\widehat{s}_N - s\|_1^2.$$

By Cauchy–Schwarz,

$$\|\widehat{r}_N - r\|_1^2 \le S\|\widehat{r}_N - r\|_2^2,$$

and therefore

$$\mathbb{E}\|\widehat{r}_N - r\|_1^2 \le S\sum_{i=1}^{S} \mathrm{Var}(\widehat{r}_{N,i}) \le \frac{S}{N}.$$

The same bound holds for $\widehat{s}_N$, giving the claim. □

8.2 A direct multiscale estimator

Let $J_1 = J_1(N)$ be defined by

$$2^{J_1 d} \le N \log N < 2^{(J_1+1)d}.$$

Let

$$J_\star := \left\lceil \frac{J_1}{2} \right\rceil.$$

For $j \in \{1, \dots, J_1\}$, the cell masses of P, Q on $\mathcal{D}_j$ form two distributions on an alphabet of size $S_j := 2^{jd}$. Let $V_j(P, Q)$ be as in Theorem 7.1. We estimate V_j as follows:

- for $1 \le j < J_\star$, use the empirical estimator

$$\widehat{V}_j^{\mathrm{emp}} := V_j(P_N, Q_N);$$

- for $J_\star \le j \le J_1$, use the Jiao–Han–Weissman estimator on the level-j alphabet, denoted $\widehat{V}_j^{\mathrm{JHW}}$, and clip it to $[0, 2]$.

Define

$$\widehat{T}_\rho := \sum_{j=1}^{J_\star - 1} a_j\, \widehat{V}_j^{\mathrm{emp}} + \sum_{j=J_\star}^{J_1} a_j\, \widehat{V}_j^{\mathrm{JHW}}, \qquad \widehat{W}_\rho := (\widehat{T}_\rho)_+^{1/p}.$$

Theorem 8.2 (Direct dyadic-tree upper bound)**.** *Assume $d > 2p$ and $n = m = N$. Then*

$$\sup_{P,Q \in \mathcal{P}_d} \mathbb{E}\big(\widehat{T}_\rho - T_\rho(P, Q)\big)^2 \lesssim_{d,p} (N \log N)^{-2p/d},$$

and therefore

$$\sup_{P,Q\in\mathcal{P}_d} \mathbb{E}\big(\widehat{W}_\rho - W_{\rho,p}(P,Q)\big)^2 \lesssim_{d,p} (N\log N)^{-2/d}.$$

Proof. Let

$$\Delta_j := \widehat{V}_j - V_j(P,Q),$$

where $\widehat{V}_j$ denotes the estimator used at level j. By Theorem 7.1,

$$\widehat{T}_\rho - T_\rho(P,Q) = \sum_{j=1}^{J_1} a_j\Delta_j - \sum_{j>J_1} a_j V_j(P,Q).$$

Since $V_j(P,Q)\le 2$, Minkowski's inequality in L_2 gives

$$\Big(\mathbb{E}(\widehat{T}_\rho - T_\rho)^2\Big)^{1/2} \le \sum_{j=1}^{J_1} a_j\big(\mathbb{E}\Delta_j^2\big)^{1/2} + 2\sum_{j>J_1} a_j. \tag{8.1}$$

For the coarse levels $j<J_\star$, Theorem 8.1 yields

$$\big(\mathbb{E}\Delta_j^2\big)^{1/2} \le 2\sqrt{\frac{2^{jd}}{N}}.$$

Hence

$$\sum_{j=1}^{J_\star-1} a_j\big(\mathbb{E}\Delta_j^2\big)^{1/2} \lesssim_p \frac{1}{\sqrt{N}}\sum_{j=1}^{J_\star-1} 2^{j(d/2-p)}.$$

Since $d/2-p>0$, the geometric sum is dominated by its last term, so

$$\sum_{j=1}^{J_\star-1} a_j\big(\mathbb{E}\Delta_j^2\big)^{1/2} \lesssim_{d,p} N^{-1/2}2^{J_\star(d/2-p)}.$$

Using $J_\star \le J_1/2+1$ and $2^{J_1 d}\asymp N\log N$, we obtain

$$N^{-1/2}2^{J_\star(d/2-p)} \lesssim_{d,p} 2^{-J_1 p}\,2^{-(J_1/2-1)(d/2-p)} \lesssim_{d,p} 2^{-J_1 p}. \tag{8.2}$$

For the fine levels $J_\star\le j\le J_1$, observe that

$$\log S_j = jd\log 2 \ge \frac{J_1 d}{2}\log 2 \ge \frac{1}{2}\log(N\log N) + O_d(1),$$

so for all large N,

$$\log N \le 3\log S_j.$$

Thus Theorem 4.1 applies with a universal constant $C_\star = 3$, and

$$\big(\mathbb{E}\Delta_j^2\big)^{1/2} \lesssim \sqrt{\frac{2^{jd}}{N\log N}}.$$

Therefore

$$\sum_{j=J_\star}^{J_1} a_j\big(\mathbb{E}\Delta_j^2\big)^{1/2} \lesssim_p \frac{1}{\sqrt{N\log N}}\sum_{j=J_\star}^{J_1} 2^{j(d/2-p)} \lesssim_{d,p} \frac{2^{J_1(d/2-p)}}{\sqrt{N\log N}} \lesssim_{d,p} 2^{-J_1 p}.$$

Finally,

$$2\sum_{j>J_1} a_j = 2^{-(J_1+1)p} \lesssim_p 2^{-J_1 p}.$$

Returning to (8.1), we conclude

$$\Big(\mathbb{E}(\widehat{T}_\rho - T_\rho)^2\Big)^{1/2} \lesssim_{d,p} 2^{-J_1 p} \asymp_{d,p} (N\log N)^{-p/d},$$

which proves the first claim.

For the second claim, note that $0 \le T_\rho, \widehat{T}_\rho \le 1$ after clipping. Hence

$$|\widehat{W}_\rho - W_{\rho,p}|^2 = |\widehat{T}_\rho^{1/p} - T_\rho^{1/p}|^2 \le |\widehat{T}_\rho - T_\rho|^{2/p}.$$

By monotonicity of L_r-norms,

$$\mathbb{E}|\widehat{T}_\rho - T_\rho|^{2/p} \le \big(\mathbb{E}|\widehat{T}_\rho - T_\rho|^2\big)^{1/p} \lesssim_{d,p} 2^{-2J_1} \asymp_{d,p} (N\log N)^{-2/d}. \qquad \square$$

8.3 The exact dyadic minimax rate and the plug-in gap

The matching minimax lower bound on trees is already known in the general weighted-tree setting.

Theorem 8.3 (Exact dyadic-tree minimax rate)**.** *Assume $d > 2p$ and $n = m = N$. Then*

$$\inf_{\widehat{W}} \sup_{P,Q\in\mathcal{P}_d} \mathbb{E}(\widehat{W} - W_{\rho,p}(P,Q))^2 \asymp_{d,p} (N\log N)^{-2/d}.$$

Source and specialization. The upper bound is Theorem 8.2. The matching lower bound is the specialization of the minimax lower theory of Wang, Cai, and Li for Wasserstein estimation on weighted trees [7] to the complete dyadic tree with branching factor 2^d, depth $J_1 \asymp \log(N\log N)$, and edge weights $a_j \asymp 2^{-jp}$. $\square$

We now analyze the empirical plug-in estimator on the same dyadic model. Since empirical measures are still measures on $[0,1]^d$, the exact tree identity of Theorem 7.1 applies verbatim to P_N, Q_N.

Theorem 8.4 (Exact plug-in rate on the dyadic tree surrogate)**.** *Assume $d > 2p$ and $n = m = N$. Then*

$$\sup_{P,Q\in\mathcal{P}_d} \mathbb{E}\big(W_{\rho,p}(P_N,Q_N) - W_{\rho,p}(P,Q)\big)^2 \asymp_{d,p} N^{-2/d}.$$

Proof. Let

$$T_{\rho,N} := W_{\rho,p}(P_N,Q_N)^p.$$

By Theorem 7.1,

$$T_{\rho,N} - T_\rho(P,Q) = \sum_{j\ge1} a_j\Big(V_j(P_N,Q_N) - V_j(P,Q)\Big).$$

Arguing exactly as in the proof of Theorem 8.2, but now using only Theorem 8.1 on every level, we obtain

$$\Big(\mathbb{E}(T_{\rho,N} - T_\rho)^2\Big)^{1/2} \lesssim_p \sum_{j\ge1} 2^{-jp} \min\left\{1, \sqrt{\frac{2^{jd}}{N}}\right\}.$$

Let J_0 be defined by

$$2^{J_0 d} \le N < 2^{(J_0+1)d}.$$

Splitting at J_0, we get

$$\sum_{j \le J_0} 2^{-jp} \sqrt{\frac{2^{jd}}{N}} \lesssim_{d,p} N^{-1/2} 2^{J_0(d/2-p)} \lesssim_{d,p} N^{-p/d},$$

and

$$\sum_{j > J_0} 2^{-jp} \lesssim_p 2^{-J_0 p} \asymp_d N^{-p/d}.$$

Thus

$$\Big(\mathbb{E}(T_{\rho,N} - T_\rho)^2\Big)^{1/2} \lesssim_{d,p} N^{-p/d},$$

and therefore

$$\sup_{P,Q} \mathbb{E}(W_{\rho,p}(P_N, Q_N) - W_{\rho,p}(P,Q))^2 \lesssim_{d,p} N^{-2/d}$$

by the same $L_{2/p}$-to-L_2 argument as before.

For the lower bound, let J_0 be as above and let $Q^{(0)}$ be the uniform distribution on the $2^{J_0 d}$ centers X_{J_0}. Then $W_{\rho,p}(Q^{(0)}, Q^{(0)}) = 0$, so the squared risk at this point equals

$$\mathbb{E} W_{\rho,p}(P_N, Q_N)^2,$$

where now P_N, Q_N are two independent empirical measures based on $Q^{(0)}$.

The minimal nonzero ρ-distance among points of X_{J_0} is 2^{-J_0}. Hence for any two measures μ, ν supported on X_{J_0},

$$W_{\rho,p}(\mu,\nu)^p \ge 2^{-J_0 p} \operatorname{TV}(\mu,\nu).$$

Since $0 \le \operatorname{TV}(\mu,\nu) \le 1$, we have $\operatorname{TV}(\mu,\nu)^{1/p} \ge \operatorname{TV}(\mu,\nu)$, and therefore

$$W_{\rho,p}(\mu,\nu) \ge 2^{-J_0} \operatorname{TV}(\mu,\nu).$$

Applying this to P_N, Q_N and using the uniform total-variation lower bound from Theorem 5.1 below with $M = 2^{J_0 d - 1} \asymp N$, we obtain

$$\mathbb{E} W_{\rho,p}(P_N, Q_N) \gtrsim_d 2^{-J_0} \asymp_d N^{-1/d},$$

hence

$$\mathbb{E} W_{\rho,p}(P_N, Q_N)^2 \ge \big(\mathbb{E} W_{\rho,p}(P_N, Q_N)\big)^2 \gtrsim_d N^{-2/d}.$$

This yields the lower bound. □

Remark 8.5. Theorem 8.3 together with Theorem 8.4 isolates the exact multiscale phenomenon one expects in the Euclidean supercritical problem: the canonical direct rate is $(N \log N)^{-1/d}$, while the empirical plug-in remains at $N^{-1/d}$. This is already a theorem on the dyadic tree surrogate. The Euclidean problem is therefore not blocked by a lack of hard scales; it is blocked by a lack of a global Euclidean transfer principle.

9 A full Euclidean theorem at the diagonal scale

This section contains the main new result of the present paper. Instead of working on a surrogate class or on a grid-supported family, we study the *full* Euclidean neighborhood of the diagonal at the critical supercritical scale.

Throughout this section we assume

$$d > 2p, \qquad n = m = N \geq 2,$$

and we set

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

For $A > 0$, define the local Euclidean class

$$\mathcal{L}_{A,N} := \{(P, Q) \in \mathcal{P}_d^2 : \ W_p(P, Q) \leq A\eta_N\}.$$

9.1 Quantization to the fine dyadic grid

Let $J = J(N)$ be the unique integer such that

$$2^{Jd} \leq N \log N < 2^{(J+1)d}.$$

Let X_J be the set of centers of the cubes in $\mathcal{D}_J$, and let

$$\kappa_J : [0, 1]^d \to X_J$$

map each point to the center of the unique cube of $\mathcal{D}_J$ that contains it. For $P \in \mathcal{P}_d$, write

$$P^{(J)} := \kappa_{J\#} P.$$

Lemma 9.1 (Deterministic quantization bound)**.** *For every $P \in \mathcal{P}_d$,*

$$W_p(P, P^{(J)}) \leq \sqrt{d}\, 2^{-J}.$$

Consequently, for every $P, Q \in \mathcal{P}_d$,

$$|W_p(P^{(J)}, Q^{(J)}) - W_p(P, Q)| \leq 2\sqrt{d}\, 2^{-J}.$$

Proof. Couple x with $\kappa_J(x)$ deterministically. The Euclidean diameter of every level-J cube is $\sqrt{d}\, 2^{-J}$, hence

$$\|x - \kappa_J(x)\|_2 \leq \sqrt{d}\, 2^{-J} \qquad \text{for all } x \in [0, 1]^d.$$

Therefore

$$W_p(P, P^{(J)})^p \leq \int \|x - \kappa_J(x)\|_2^p \, dP(x) \leq d^{p/2} 2^{-Jp},$$

which proves the first claim. The second follows from the triangle inequality:

$$W_p(P^{(J)}, Q^{(J)}) \leq W_p(P^{(J)}, P) + W_p(P, Q) + W_p(Q, Q^{(J)}),$$

and symmetrically with $(P^{(J)}, Q^{(J)})$ and (P, Q) exchanged. □

9.2 A direct estimator on the full local class

We now quantize the samples and apply the dyadic-tree estimator from Theorem 8.2 to the resulting grid-supported measures.

Theorem 9.2 (Direct local upper bound on the full Euclidean neighborhood)**.** *Fix $A > 0$. There exists an estimator $\widehat{W}_{\mathrm{loc}}$ such that*

$$\sup_{(P,Q)\in\mathcal{L}_{A,N}} \mathbb{E}\big(\widehat{W}_{\mathrm{loc}} - W_p(P,Q)\big)^2 \lesssim_{A,d,p} \eta_N^2 = (N\log N)^{-2/d}.$$

Consequently,

$$\sup_{(P,Q)\in\mathcal{L}_{A,N}} \mathbb{E}\big|\widehat{W}_{\mathrm{loc}} - W_p(P,Q)\big| \lesssim_{A,d,p} \eta_N = (N\log N)^{-1/d}.$$

Proof. Given the original samples $X_1,\dots,X_N$ and $Y_1,\dots,Y_N$, replace them by the quantized samples

$$\kappa_J(X_1),\dots,\kappa_J(X_N), \qquad \kappa_J(Y_1),\dots,\kappa_J(Y_N).$$

These are i.i.d. samples from $P^{(J)}$ and $Q^{(J)}$, respectively. Apply the dyadic-tree estimator of Theorem 8.2 to these quantized samples, and denote the resulting estimator by $\widehat{W}_{\mathrm{loc}}$.

By Theorem 8.2,

$$\sup_{P,Q\in\mathcal{P}_d} \mathbb{E}\big(\widehat{W}_{\mathrm{loc}} - W_{\rho,p}(P^{(J)},Q^{(J)})\big)^2 \lesssim_{d,p} \eta_N^2. \tag{9.1}$$

Now fix $(P,Q)\in\mathcal{L}_{A,N}$. By Theorem 9.1,

$$W_p(P^{(J)},Q^{(J)}) \le W_p(P,Q) + 2\sqrt{d}\,2^{-J} \lesssim_{A,d} \eta_N.$$

Since $P^{(J)}$ and $Q^{(J)}$ are supported on the fine dyadic grid X_J, Theorem 7.2 gives

$$0 \le W_p(P^{(J)},Q^{(J)}) - W_{\rho,p}(P^{(J)},Q^{(J)}) \le (2\sqrt{d}-1)W_{\rho,p}(P^{(J)},Q^{(J)}) \le (2\sqrt{d}-1)W_p(P^{(J)},Q^{(J)}).$$

Hence

$$\big|W_p(P^{(J)},Q^{(J)}) - W_{\rho,p}(P^{(J)},Q^{(J)})\big| \lesssim_{A,d} \eta_N. \tag{9.2}$$

Together with Theorem 9.1, this yields

$$\begin{aligned}\big|W_p(P,Q) - W_{\rho,p}(P^{(J)},Q^{(J)})\big| &\le \big|W_p(P,Q) - W_p(P^{(J)},Q^{(J)})\big| \\ &\quad + \big|W_p(P^{(J)},Q^{(J)}) - W_{\rho,p}(P^{(J)},Q^{(J)})\big| \\ &\lesssim_{A,d} \eta_N.\end{aligned}$$

Therefore

$$|\widehat{W}_{\mathrm{loc}} - W_p(P,Q)| \le |\widehat{W}_{\mathrm{loc}} - W_{\rho,p}(P^{(J)},Q^{(J)})| + C_{A,d}\eta_N.$$

Squaring, taking expectations, and using Equation (9.1) proves the squared bound. The absolute bound follows from Jensen's inequality. □

9.3 The hard lower-envelope scale already lives inside the local neighborhood

The next theorem extracts a stronger conclusion from the proof of Niles-Weed–Rigollet's supercritical lower bound: their hard instances already lie inside a shrinking Euclidean neighborhood of the diagonal.

Theorem 9.3 (Local lower bound from the Niles–Weed–Rigollet construction)**.** *Assume $d > 2p$. There exist constants $A_0 = A_0(d,p)$, $c_0 = c_0(d,p) > 0$, and $N_0 = N_0(d,p)$ such that for every $N \geq N_0$ and every estimator $\widehat{W}$,*

$$\sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{P}_{P^{\otimes N}\otimes Q^{\otimes N}}\big(|\widehat{W} - W_p(P,Q)| \geq c_0\eta_N\big) \geq c_0.$$

Consequently,

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \gtrsim_{d,p} \eta_N,$$

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}\big(\widehat{W} - W_p(P,Q)\big)^2 \gtrsim_{d,p} \eta_N^2.$$

Proof. We follow the proof of the supercritical lower bound in Niles-Weed and Rigollet [6, Proof of Theorem 11], but we keep track of the size of the hard instances.

Let u be the uniform distribution on $[m]$, where m will be chosen below. Let $c_* > 0$ and $C_{d,p} > 0$ denote the constants from [6, Proposition 9], and let $\delta_{d,p} > 0$ be the small constant appearing in the proof of [6, Theorem 11]. Fix $0 < \delta \leq \delta_{d,p}$, and set

$$m := \left\lceil C\delta^{-1} N \log N \right\rceil,$$

with C large enough that [6, Proposition 10] applies. Let D_m, $D^-_{m,\delta}$, and D^+_m be the same composite hypotheses as in [6, Proposition 10]: namely

$$D_m := \{q : \chi^2(q,u) \leq 9\},$$

$$D^-_{m,\delta} := \{q \in D_m : \ \mathrm{TV}(q,u) \leq \delta\}, \qquad D^+_m := \{q \in D_m : \ \mathrm{TV}(q,u) \geq 1/4\}.$$

Let $F : [m] \to [0,1]^d$ be the random map from [6, Proposition 9], and write

$$\mu_{F,q} := F_\# q, \qquad \nu_F := F_\# u.$$

Define

$$\Delta_N := \frac{c_*}{16} m^{-1/d}.$$

Because $m \geq N \log N$, we have $m^{-1/d} \leq \eta_N$. Also, by [6, Proposition 9], every $q \in D_m$ satisfies

$$W_p(\mu_{F,q}, \nu_F) \leq C_{d,p} m^{-1/d} \chi^2(q,u)^{1/d}\, \mathrm{TV}(q,u)^{1/p-2/d} \leq C_{d,p} 9^{1/d} m^{-1/d} \leq C_{d,p} 9^{1/d} \eta_N$$

with probability at least 0.9 over the random map F. We therefore fix

$$A_0 := C_{d,p} 9^{1/d}.$$

Now define the "good local" events

$$G^-_q := \{W_p(\mu_{F,q}, \nu_F) \leq \Delta_N\}, \qquad q \in D^-_{m,\delta},$$

and

$$G_q^+ := \{3\Delta_N \le W_p(\mu_{F,q}, \nu_F) \le A_0\eta_N\}, \qquad q \in D_m^+.$$

The proof of [6, Theorem 11] shows that

$$\mathbb{P}_F(G_q^-) \ge 0.9 \qquad (q \in D_{m,\delta}^-),$$

$$\mathbb{P}_F(G_q^+) \ge 0.9 \qquad (q \in D_m^+).$$

In particular, on both G_q^- and G_q^+, the pair $(\mu_{F,q}, \nu_F)$ belongs to the local class $\mathcal{L}_{A_0,N}$.

Let $\widehat{W}$ be any estimator, and define the threshold test

$$\psi := \mathbf{1}\{\widehat{W} \le 2\Delta_N\}.$$

Also define the estimation-error event

$$A_q(F) := \Big\{|\widehat{W} - W_p(\mu_{F,q}, \nu_F)| \ge \Delta_N\Big\}.$$

If $q \in D_{m,\delta}^-$, then on G_q^- the implication

$$\{\widehat{W} > 2\Delta_N\} \subset A_q(F)$$

holds. Therefore

$$\mathbb{E}_F\Big[\mathbb{P}_{\mu_{F,q}^{\otimes N}\otimes\nu_F^{\otimes N}}\big(A_q(F)\big)\,\mathbf{1}_{G_q^-}\Big] \ge \mathbb{P}_q(\psi = 0) - 0.1.$$

Similarly, if $q \in D_m^+$, then on G_q^+ the implication

$$\{\widehat{W} \le 2\Delta_N\} \subset A_q(F)$$

holds, because $W_p(\mu_{F,q}, \nu_F) \ge 3\Delta_N$. Hence

$$\mathbb{E}_F\Big[\mathbb{P}_{\mu_{F,q}^{\otimes N}\otimes\nu_F^{\otimes N}}\big(A_q(F)\big)\,\mathbf{1}_{G_q^+}\Big] \ge \mathbb{P}_q(\psi = 1) - 0.1.$$

For every fixed realization of F, the indicator $\mathbf{1}_{G_q^\pm}$ forces $(\mu_{F,q}, \nu_F) \in \mathcal{L}_{A_0,N}$, so

$$\mathbb{P}_{\mu_{F,q}^{\otimes N}\otimes\nu_F^{\otimes N}}\big(A_q(F)\big)\,\mathbf{1}_{G_q^\pm} \le \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{P}_{P^{\otimes N}\otimes Q^{\otimes N}}\big(|\widehat{W} - W_p(P,Q)| \ge \Delta_N\big).$$

Taking expectations in F and combining the last three displays yields

$$\sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{P}\big(|\widehat{W} - W_p(P,Q)| \ge \Delta_N\big) \ge \frac{1}{2}\Big(\sup_{q\in D_m^+} \mathbb{P}_q(\psi = 1) + \sup_{q\in D_{m,\delta}^-} \mathbb{P}_q(\psi = 0)\Big) - 0.1.$$

By [6, Proposition 10], the quantity in parentheses is at least 0.9 for all large N. Therefore

$$\sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{P}\big(|\widehat{W} - W_p(P,Q)| \ge \Delta_N\big) \ge 0.35$$

for all large N. Since

$$\Delta_N = \frac{c_*}{16} m^{-1/d} \asymp_{d,p} (N\log N)^{-1/d} = \eta_N,$$

the first claim follows by absorbing constants into a smaller $c_0 > 0$.

The absolute-risk lower bound follows from

$$\mathbb{E}|X| \geq a\,\mathbb{P}(|X| \geq a),$$

applied with $a = c_0\eta_N$. The squared lower bound follows analogously from

$$\mathbb{E}X^2 \geq a^2\,\mathbb{P}(|X| \geq a). \qquad \square$$

Corollary 9.4 (Exact local minimax law at the diagonal scale)**.** *Assume $d > 2p$ and $n = m = N$.*

(i) *For every fixed $A > 0$,*

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A,N}} \mathbb{E}\big(\widehat{W} - W_p(P,Q)\big)^2 \lesssim_{A,d,p} (N\log N)^{-2/d},$$

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A,N}} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \lesssim_{A,d,p} (N\log N)^{-1/d}.$$

(ii) *There exists $A_0 = A_0(d,p)$ such that*

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}\big(\widehat{W} - W_p(P,Q)\big)^2 \asymp_{d,p} (N\log N)^{-2/d},$$

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{L}_{A_0,N}} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \asymp_{d,p} (N\log N)^{-1/d}.$$

Proof. Combine Theorems 9.2 and 9.3. $\square$

9.4 The empirical plug-in estimator is still locally suboptimal on the full class

The local exact law above concerns the minimax benchmark. We now show that the empirical plug-in estimator remains worse by a full logarithmic factor *on the same full local Euclidean class.*

Theorem 9.5 (Local plug-in lower bound on the full Euclidean neighborhood)**.** *Assume $d > 2p$ and $n = m = N$. For every fixed $A > 0$,*

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

Proof. Fix $M := N$, and choose

$$\delta := \frac{c_d}{2} N^{-1/d},$$

where c_d is the packing constant from Theorem 3.1. Let $\mathcal{G}^{\text{pair}}_{M,\delta}$ be the corresponding paired class, and let

$$Q_0 = \frac{1}{2M}\sum_{j=1}^{M}(\delta_{x_j^+} + \delta_{x_j^-}) \in \mathcal{G}^{\text{pair}}_{M,\delta}$$

be the uniform paired point from Theorem 5.2. Then

$$W_p(Q_0,Q_0) = 0,$$

so $(Q_0, Q_0) \in \mathcal{L}_{A,N}$ for every $A > 0$.

Now apply Theorem 5.2 with $M = N$. Since $M \geq cN$ for $c = 1$, it yields

$$\mathbb{E}_{Q_0^{\otimes N} \otimes Q_0^{\otimes N}} W_p(P_N, Q_N)^2 \gtrsim \delta^2 \asymp_d N^{-2/d}.$$

Because the true target at (Q_0, Q_0) is zero, the left-hand side is exactly the squared risk of the empirical plug-in estimator at a point of $\mathcal{L}_{A,N}$. □

Corollary 9.6 (A full Euclidean local logarithmic separation)**.** *Assume $d > 2p$ and $n = m = N$. Then there exists $A_0 = A_0(d, p)$ such that*

$$\inf_{\widehat{W}} \sup_{(P,Q) \in \mathcal{L}_{A_0,N}} \mathbb{E}\big(\widehat{W} - W_p(P, Q)\big)^2 \asymp_{d,p} (N \log N)^{-2/d},$$

whereas the empirical plug-in estimator satisfies

$$\sup_{(P,Q) \in \mathcal{L}_{A_0,N}} \mathbb{E}\big(W_p(P_N, Q_N) - W_p(P, Q)\big)^2 \gtrsim_{d,p} N^{-2/d}.$$

Hence, even on the full Euclidean neighborhood of the diagonal*, direct estimation beats the empirical plug-in estimator by a genuine logarithmic factor.*

Proof. Combine Theorems 9.4 and 9.5. □

Remark 9.7. The key conceptual point is that the supercritical lower-envelope scale $(N \log N)^{-1/d}$ is now fully resolved on the *actual Euclidean model*, not merely on a tree surrogate and not merely on a specially engineered grid-supported subclass. What remains open is the extension from the diagonal neighborhood $\mathcal{L}_{A_0,N}$ to the full unrestricted class.

10 What remained open after the diagonal theorem?

The new theorem of Theorem 9.4 settles the exact minimax scale $(N \log N)^{-1/d}$ on the *full Euclidean neighborhood of the diagonal*. Combined with the exact dyadic-tree theory of Theorems 8.3 and 8.4, this changes the status of the supercritical problem in a substantial way.

Discussion 10.1 (What is now settled)*.* For fixed $d > 2p$ and $n = m = N$, the lower-envelope scale $(N \log N)^{-1/d}$ is no longer only a lower bound and no longer only a surrogate-tree phenomenon. It is the exact minimax absolute rate, and $(N \log N)^{-2/d}$ is the exact minimax squared rate, on the full Euclidean class

$$\mathcal{L}_{A_0,N} = \{(P, Q) : \ W_p(P, Q) \leq A_0 (N \log N)^{-1/d}\}.$$

Moreover, even on that full class, the empirical plug-in estimator remains stuck at the slower squared scale $N^{-2/d}$.

What is still missing is a globalization step. The unrestricted balanced supercritical problem would be solved if one could prove a matching upper bound on all pairs $P, Q \in \mathcal{P}_d$, not only on the local diagonal neighborhood. At present, the remaining open region is therefore genuinely *off-diagonal*.

There are two natural paths.

Discussion 10.2 (Two remaining routes). Any full solution of the unrestricted balanced supercritical problem would follow from either of the next two advances.

(a) A *nonlocalized Euclidean multiscale transfer theorem*: an additive approximation principle showing that after quantization to the fine dyadic grid, the Euclidean and dyadic-tree Wasserstein functionals differ by at most $O((N \log N)^{-1/d})$ uniformly over all $P, Q \in \mathcal{P}_d$.

(b) An *intrinsically Euclidean direct estimator*: a construction that estimates $W_p(P, Q)$ at scale $(N \log N)^{-1/d}$ without passing through a surrogate tree metric, for example by combining a diagonal debiasing mechanism with an off-diagonal regularity expansion.

The first route is conceptually attractive because Theorem 8.3 already gives the exact multiscale answer on the canonical dyadic surrogate. The second route is attractive because the sharp off-diagonal empirical theory with smooth costs developed by Manole and Niles-Weed suggests that once the transport cost is sufficiently regular, plug-in errors can be much smaller than the diagonal lower-envelope scale.

Remark 10.3. The present paper therefore does not close the unrestricted minimax problem completely. But it does move the center of gravity. The hard part is no longer "finding the right lower-envelope scale" and no longer "building a fixed-dimensional Euclidean class that exhibits a logarithmic gap." Both of those tasks are now solved theorem-wise. The remaining difficulty is to interpolate from the exact diagonal theory to the full global Euclidean geometry.

The next two sections sharpen this globalization picture in two orthogonal directions. First, we show that the unrestricted balanced supercritical problem is already equivalent, up to the critical scale, to a finite-state Euclidean problem on the level-J dyadic grid with $|X_J| \asymp N \log N$ atoms. Second, for absolute loss we combine the local diagonal theorem with the smooth-cost empirical theory of Manole and Niles-Weed to prove that sufficiently separated pairs already admit the desired η_N-scale upper bound. This isolates an explicit middle annulus as the only distance regime not yet covered by current theorem-level arguments.

11 A critical-grid reduction of the unrestricted problem

We now show that the unrestricted balanced supercritical problem is equivalent, up to the critical scale, to the same Euclidean estimation problem on the level-J dyadic grid with $2^{Jd} \asymp N \log N$ atoms.

Throughout this section we assume

$$d > 2p, \qquad n = m = N \geq 2.$$

Fix $L \geq 1$, and let $J = J_L(N)$ be the unique integer such that

$$2^{Jd} \leq LN \log N < 2^{(J+1)d}.$$

Let X_J be the set of centers of the cubes in $\mathcal{D}_J$, and let $\mathcal{P}(X_J)$ denote the class of all probability measures supported on X_J. Define the corresponding grid minimax absolute and squared risks by

$$\mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} := \inf_{\widehat{W}} \sup_{\mu,\nu \in \mathcal{P}(X_J)} \mathbb{E}_{\mu^{\otimes N} \otimes \nu^{\otimes N}} \big|\widehat{W} - W_p(\mu, \nu)\big|,$$

$$\mathfrak{G}^{\mathrm{sq}}_{N,L,d,p} := \inf_{\widehat{W}} \sup_{\mu,\nu \in \mathcal{P}(X_J)} \mathbb{E}_{\mu^{\otimes N} \otimes \nu^{\otimes N}} \big(\widehat{W} - W_p(\mu, \nu)\big)^2.$$

Theorem 11.1 (Critical-grid reduction)**.** *Under the assumptions above,*

$$\mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} \le \mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} \le \mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} + 2\sqrt{d}\,2^{-J},$$

and

$$\mathfrak{G}^{\mathrm{sq}}_{N,L,d,p} \le \mathfrak{M}^{\mathrm{sq}}_{N,N,d,p} \le 2\mathfrak{G}^{\mathrm{sq}}_{N,L,d,p} + 8d\,2^{-2J}.$$

Proof. The left-hand inequalities are immediate because $\mathcal{P}(X_J) \subset \mathcal{P}_d$, so the grid problem is a restriction of the unrestricted one.

For the right-hand inequalities, let $\widehat{W}_J$ be any estimator based on N samples from each of two unknown measures in $\mathcal{P}(X_J)$. Given arbitrary $P, Q \in \mathcal{P}_d$, quantize the samples:

$$\kappa_J(X_1), \dots, \kappa_J(X_N), \qquad \kappa_J(Y_1), \dots, \kappa_J(Y_N).$$

These are i.i.d. samples from $P^{(J)}$ and $Q^{(J)}$, both supported on X_J. Define the lifted estimator

$$\widetilde{W} := \widehat{W}_J\big(\kappa_J(X_1), \dots, \kappa_J(X_N), \kappa_J(Y_1), \dots, \kappa_J(Y_N)\big).$$

By Theorem 9.1,

$$|W_p(P^{(J)}, Q^{(J)}) - W_p(P, Q)| \le 2\sqrt{d}\,2^{-J}.$$

Therefore

$$|\widetilde{W} - W_p(P, Q)| \le |\widetilde{W} - W_p(P^{(J)}, Q^{(J)})| + 2\sqrt{d}\,2^{-J}.$$

Taking expectations and then the supremum over $P, Q \in \mathcal{P}_d$ yields

$$\sup_{P,Q \in \mathcal{P}_d} \mathbb{E}|\widetilde{W} - W_p(P, Q)| \le \sup_{\mu,\nu \in \mathcal{P}(X_J)} \mathbb{E}|\widehat{W}_J - W_p(\mu, \nu)| + 2\sqrt{d}\,2^{-J}.$$

Infimizing over $\widehat{W}_J$ proves the absolute-loss upper bound.

For squared loss, write

$$\widetilde{W} - W_p(P, Q) = (\widetilde{W} - W_p(P^{(J)}, Q^{(J)})) + (W_p(P^{(J)}, Q^{(J)}) - W_p(P, Q)),$$

so by $(a + b)^2 \le 2a^2 + 2b^2$,

$$(\widetilde{W} - W_p(P, Q))^2 \le 2\big(\widetilde{W} - W_p(P^{(J)}, Q^{(J)})\big)^2 + 8d\,2^{-2J}.$$

Taking expectations, suprema, and then infima gives the claimed squared-loss bound. □

Since $2^{-J} \asymp L^{-1/d}(N \log N)^{-1/d}$, the additive loss in Theorem 11.1 is exactly of the same order as the balanced supercritical lower-envelope scale.

Corollary 11.2 (Equivalent finite-state formulation at the critical scale)**.** *Let*

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

There exist constants $L_0 = L_0(d, p) \ge 1$ and $N_0 = N_0(d, p)$ such that for every fixed $L \ge L_0$ and every $N \ge N_0$,

$$\mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} \asymp_{L,d,p} \eta_N \qquad \Longleftrightarrow \qquad \mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} \asymp_{d,p} \eta_N,$$

and

$$\mathfrak{G}^{\mathrm{sq}}_{N,L,d,p} \asymp_{L,d,p} \eta_N^2 \qquad \Longleftrightarrow \qquad \mathfrak{M}^{\mathrm{sq}}_{N,N,d,p} \asymp_{d,p} \eta_N^2.$$

In particular, the unrestricted balanced supercritical continuum problem is equivalent, up to the critical scale, to the corresponding Euclidean problem on the finite grid X_J with $|X_J| = 2^{Jd} \asymp LN \log N$.

Proof. By Theorem 11.1,

$$\mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} \le \mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} \le \mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} + 2\sqrt{d}\, 2^{-J}.$$

Since $2^{-J} < 2L^{-1/d}\eta_N$, the rightmost additive term is at most $4\sqrt{d}\, L^{-1/d}\eta_N$.

On the other hand, Theorem 9.4 implies that for all large N,

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

Choose $L_0 = L_0(d,p)$ large enough that

$$4\sqrt{d}\, L_0^{-1/d} \le \frac{1}{2} c_{d,p}, \qquad 16d\, L_0^{-2/d} \le \frac{1}{4} c'_{d,p},$$

where $c_{d,p}, c'_{d,p} > 0$ are the implied lower-bound constants above. Then for every fixed $L \ge L_0$ and all sufficiently large N,

$$\mathfrak{G}^{\mathrm{abs}}_{N,L,d,p} \ge \mathfrak{M}^{\mathrm{abs}}_{N,N,d,p} - 4\sqrt{d}\, L^{-1/d}\eta_N \gtrsim_{L,d,p} \eta_N,$$

and similarly

$$\mathfrak{G}^{\mathrm{sq}}_{N,L,d,p} \ge \frac{1}{2}\mathfrak{M}^{\mathrm{sq}}_{N,N,d,p} - 4d\, 2^{-2J} \gtrsim_{L,d,p} \eta_N^2.$$

The forward implications then follow from the trivial bounds $\mathfrak{G} \le \mathfrak{M}$, while the reverse implications follow from the upper bounds in Theorem 11.1. □

Remark 11.3. Theorem 11.2 identifies the correct finite alphabet size behind the unrestricted balanced supercritical problem: the relevant Euclidean discrete model is not fixed-size and not of order N, but rather of order $N \log N$. This is the same alphabet size at which the classical sample-size enlargement phenomenon appears for discrete L_1-functional estimation.

12 A coarse separated-class upper bound for absolute loss

A second sharpening of the globalization question comes from the smooth-cost empirical optimal transport theory of Manole and Niles-Weed. Throughout this section we assume

$$d > 2p, \qquad n = m = N \ge 2,$$

and we set

$$\eta_N := (N \log N)^{-1/d}, \qquad \beta_p := p \wedge 2.$$

For $r > 0$, define the separated class

$$\mathcal{S}_r := \{(P,Q) \in \mathcal{P}_d^2 : \ W_p(P,Q) \ge r\}.$$

Proposition 12.1 (Smooth-cost upper bound for the empirical p-cost)**.** *For every $p \geq 1$,*

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

Proof. This is exactly [5, Corollary 3(i)] specialized to the Euclidean cost $c(x,y) = \|x-y\|_2^p$ on the compact class $[0,1]^d$, after a fixed affine normalization of the ambient cube into the unit-ball framework used there. The homogeneity of the transport cost under dilations changes only the multiplicative constant, which remains a function of d and p alone. □

The next elementary lemma converts an error bound for W_p^p into an error bound for W_p, uniformly over classes separated away from the diagonal.

Lemma 12.2 (Power-to-distance conversion off the diagonal)**.** *Let $p \geq 1$, $r > 0$, $y \geq r$, and $x \geq 0$. Then*

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

Proof. If $x \geq y$, then

$$x^p - y^p = p\int_y^x t^{p-1}\,dt \geq pr^{p-1}(x-y) \geq r^{p-1}(x-y),$$

because $t \geq y \geq r$ on the interval of integration.

If $0 \leq x \leq y$, define

$$g(x) := y^p - x^p - r^{p-1}(y-x), \qquad x \in [0,y].$$

Then

$$g''(x) = -p(p-1)x^{p-2} \leq 0,$$

so g is concave on $[0,y]$. Also,

$$g(y) = 0, \qquad g(0) = y^p - r^{p-1}y = y(y^{p-1} - r^{p-1}) \geq 0,$$

because $y \geq r$. A concave function that is nonnegative at both endpoints of an interval is nonnegative throughout the interval, so $g(x) \geq 0$ for all $x \in [0,y]$. Equivalently,

$$y^p - x^p \geq r^{p-1}(y-x).$$

Combining the two cases proves the claim. □

Theorem 12.3 (Separated-class upper bound in absolute loss)**.** *For every $r > 0$,*

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{S}_r} \mathbb{E}|\widehat{W} - W_p(P,Q)| \leq \sup_{(P,Q)\in\mathcal{S}_r} \mathbb{E}|W_p(P_N,Q_N) - W_p(P,Q)| \lesssim_{d,p} r^{1-p}N^{-\beta_p/d}.$$

Proof. The first inequality is trivial, since it suffices to consider the empirical plug-in estimator. Now fix $(P,Q) \in \mathcal{S}_r$. Set

$$x := W_p(P_N,Q_N), \qquad y := W_p(P,Q).$$

Because $(P,Q) \in \mathcal{S}_r$, we have $y \geq r$. Applying Theorem 12.2 pointwise gives

$$|W_p(P_N,Q_N) - W_p(P,Q)| \leq r^{1-p}|W_p(P_N,Q_N)^p - W_p(P,Q)^p|.$$

Taking expectations and then the supremum over $(P,Q) \in \mathcal{S}_r$, and finally invoking Theorem 12.1, proves the claim. $\square$

For $p > 1$, define the separated threshold

$$\tau_{N,p} := \left(N^{-\beta_p/d}\eta_N^{-1}\right)^{1/(p-1)} = N^{-(\beta_p-1)/(d(p-1))}(\log N)^{1/(d(p-1))}.$$

Explicitly,

$$\tau_{N,p} = \begin{cases} N^{-1/d}(\log N)^{1/((p-1)d)}, & 1 < p < 2, \\ N^{-1/d}(\log N)^{1/d}, & p = 2, \\ N^{-1/((p-1)d)}(\log N)^{1/((p-1)d)}, & p > 2. \end{cases}$$

Corollary 12.4 (The target scale is already achieved on sufficiently separated classes)**.** *Assume $p > 1$. There exists $B = B(d,p) \geq 1$ such that*

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{S}_{B\tau_{N,p}}} \mathbb{E}\left|\widehat{W} - W_p(P,Q)\right| \lesssim_{d,p} \eta_N.$$

In particular, every pair satisfying $W_p(P,Q) \gtrsim_{d,p} \tau_{N,p}$ is already estimable at the balanced supercritical lower-envelope scale in absolute loss.

Proof. By Theorem 12.3,

$$\sup_{(P,Q)\in\mathcal{S}_{B\tau_{N,p}}} \mathbb{E}\left|W_p(P_N,Q_N) - W_p(P,Q)\right| \lesssim_{d,p} B^{1-p}\tau_{N,p}^{1-p}N^{-\beta_p/d}.$$

By the definition of $\tau_{N,p}$,

$$\tau_{N,p}^{1-p}N^{-\beta_p/d} = \eta_N.$$

Choosing B large enough absorbs the implicit constant. $\square$

Discussion 12.5 (A first remaining annulus for absolute loss)*.* Let $A_0 = A_0(d,p)$ be the constant from Theorem 9.4. For $p > 1$, Theorems 9.4 and 12.4 show that current theorem-level arguments already attain the target absolute scale η_N on the full diagonal neighborhood

$$\mathcal{L}_{A_0,N} = \{(P,Q) : \ W_p(P,Q) \leq A_0\eta_N\}$$

and on every sufficiently separated class $\mathcal{S}_{B\tau_{N,p}}$. Thus the only distances for which no theorem-level $O((N\log N)^{-1/d})$ upper bound is presently available are those in the middle annulus

$$A_0\eta_N \lesssim W_p(P,Q) \lesssim \tau_{N,p}.$$

When $p = 2$, this annulus has only logarithmic width in the functional value:

$$(N\log N)^{-1/d} \lesssim W_2(P,Q) \lesssim N^{-1/d}(\log N)^{1/d}.$$

For $1 < p < 2$, the upper endpoint is $N^{-1/d}(\log N)^{1/((p-1)d)}$, while for $p > 2$ it becomes $N^{-1/((p-1)d)}(\log N)^{1/((p-1)d)}$. The case $p = 1$ remains exceptional: the smooth-cost off-diagonal theory does not improve the generic $N^{-1/d}$ absolute plug-in scale.

Remark 12.6. Combined with Theorem 11.2, the preceding discussion reduces the current absolute-loss ambiguity to a finite-state Euclidean problem on a dyadic grid with $|X_J| \asymp N\log N$

atoms and to transport distances lying in the annulus of Theorem 12.5. In this precise sense, the remaining obstruction is no longer the whole continuum geometry and no longer the whole off-diagonal region.

13 Piecewise-affine linearization of the p-cost and the splitting obstruction

The critical-grid reduction of Theorem 11.1 shows that the unrestricted problem is, at the target scale, equivalent to a finite Euclidean grid problem. The next question is whether one can sharpen that reduction further by replacing the nonlinear kernel $(x, y) \mapsto \|x - y\|_2^p$ with a simpler local model.

For $p > 1$, the answer is yes at the level of the *p-cost*

$$T_p(P, Q) := W_p(P, Q)^p.$$

The kernel $\|x - y\|_2^p$ is globally $C^{1,\beta_p - 1}$ on bounded sets, where

$$\beta_p := p \wedge 2.$$

This yields a dyadic cellwise linearization with deterministic error $O(2^{-J\beta_p})$. The theorem below upgrades the vague "intrinsically Euclidean" route from Theorems 11.1 and 12.5 to a precise reduced target.

Lemma 13.1 (Uniform first-order remainder for $\|z\|_2^p$)**.** *Assume $p > 1$, and set $\beta_p := p \wedge 2$. There exists a constant $C_{d,p} < \infty$ such that for all $z, u \in \mathbb{R}^d$ satisfying $\|z\|_2 \le \sqrt{d}$ and $\|z + u\|_2 \le \sqrt{d}$,*

$$\big|\|z + u\|_2^p - \|z\|_2^p - \nabla h_p(z) \cdot u\big| \le C_{d,p} \|u\|_2^{\beta_p},$$

where $h_p(z) := \|z\|_2^p$ and $\nabla h_p(0) := 0$.

Proof. Let

$$g(t) := \|z + tu\|_2^p, \qquad t \in [0, 1].$$

Then

$$g'(t) = \nabla h_p(z + tu) \cdot u,$$

and whenever $z + tu \neq 0$,

$$g''(t) = u^\top \nabla^2 h_p(z + tu)\, u.$$

If $p \ge 2$, then for $w \neq 0$,

$$\nabla^2 h_p(w) = p\|w\|_2^{p-2} I_d + p(p-2)\|w\|_2^{p-4} w w^\top.$$

Hence

$$\|\nabla^2 h_p(w)\|_{\mathrm{op}} \le p(p-1)\|w\|_2^{p-2} \le p(p-1) d^{(p-2)/2}$$

on the segment $\{z + tu : \ 0 \le t \le 1\}$. Taylor's theorem with integral remainder therefore gives

$$|g(1) - g(0) - g'(0)| \le \int_0^1 (1 - t)|g''(t)|\, dt \le \frac{1}{2} p(p-1) d^{(p-2)/2} \|u\|_2^2.$$

This proves the claim when $p \ge 2$.

Now assume $1 < p < 2$. For $w \neq 0$,

$$\|\nabla^2 h_p(w)\|_{\mathrm{op}} \leq p(p-1)\|w\|_2^{p-2}.$$

Thus

$$\|\nabla h_p(z+u) - \nabla h_p(z)\|_2 \leq p(p-1)\|u\|_2 \int_0^1 \|z+tu\|_2^{p-2}\, dt.$$

Write $u = \|u\|_2 v$ with $\|v\|_2 = 1$, and set $\alpha := z/\|u\|_2$. Then

$$\int_0^1 \|z+tu\|_2^{p-2}\, dt = \|u\|_2^{p-2} \int_0^1 \|\alpha + tv\|_2^{p-2}\, dt.$$

Since $\|\alpha + tv\|_2 \geq |\alpha \cdot v + t|$ and $p - 2 < 0$,

$$\int_0^1 \|\alpha + tv\|_2^{p-2}\, dt \leq \int_0^1 |\alpha \cdot v + t|^{p-2}\, dt.$$

The latter is bounded uniformly over $\alpha \cdot v \in \mathbb{R}$ by a constant depending only on p, because every interval of length one has

$$\int_I |r|^{p-2}\, dr \leq 2\int_0^1 r^{p-2}\, dr = \frac{2}{p-1}.$$

Hence

$$\|\nabla h_p(z+u) - \nabla h_p(z)\|_2 \leq 2p\, \|u\|_2^{p-1}.$$

Applying this with tu in place of u and integrating once more,

$$\begin{aligned}
|g(1) - g(0) - g'(0)| &= \left| \int_0^1 \big(\nabla h_p(z+tu) - \nabla h_p(z)\big) \cdot u\, dt \right| \\
&\leq \int_0^1 \|\nabla h_p(z+tu) - \nabla h_p(z)\|_2\, \|u\|_2\, dt \\
&\leq 2p\|u\|_2^p \int_0^1 t^{p-1}\, dt = 2\|u\|_2^p.
\end{aligned}$$

This proves the claim for $1 < p < 2$ as well. □

For a dyadic cube $R \in \mathcal{D}_J$, write c_R for its center. Define the piecewise affine linearized cost $c_{J,p}^{\mathrm{lin}}$ by

$$c_{J,p}^{\mathrm{lin}}(x,y) := \|c_R - c_S\|_2^p + \nabla h_p(c_R - c_S) \cdot \big[(x - c_R) - (y - c_S)\big], \qquad x \in R,\ y \in S.$$

Let

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

Proposition 13.2 (Dyadic linearization of the p-cost)**.** *Assume $p > 1$, and set $\beta_p := p \wedge 2$. For every level $J \geq 1$,*

$$\sup_{x,y \in [0,1]^d} \big| \|x-y\|_2^p - c_{J,p}^{\mathrm{lin}}(x,y) \big| \lesssim_{d,p} 2^{-J\beta_p}.$$

Consequently, for all $P, Q \in \mathcal{P}_d$,

$$\big| W_p(P,Q)^p - T_{J,p}^{\mathrm{lin}}(P,Q) \big| \lesssim_{d,p} 2^{-J\beta_p}.$$

Proof. Fix $x, y \in [0,1]^d$, let $R, S \in \mathcal{D}_J$ be the cells containing them, and write

$$z := c_R - c_S, \qquad u := (x - c_R) - (y - c_S).$$

Then

$$x - y = z + u.$$

Since z and $z + u = x - y$ are both differences of points in $[0,1]^d$, their Euclidean norms are at most $\sqrt{d}$. Moreover,

$$\|u\|_2 \le \|x - c_R\|_2 + \|y - c_S\|_2 \le \sqrt{d}\, 2^{-J}.$$

Applying Theorem 13.1 gives

$$\big| \|x - y\|_2^p - \|c_R - c_S\|_2^p - \nabla h_p(c_R - c_S) \cdot u \big| \lesssim_{d,p} 2^{-J\beta_p}.$$

This is exactly the displayed pointwise bound.

For the transport-cost statement, let

$$\delta_J := \sup_{x,y} \big| \|x - y\|_2^p - c_{J,p}^{\mathrm{lin}}(x,y) \big|.$$

For any coupling $\pi \in \Pi(P,Q)$,

$$\left| \int \|x - y\|_2^p \, d\pi - \int c_{J,p}^{\mathrm{lin}}(x,y)\, d\pi \right| \le \delta_J.$$

Taking the infimum over π on both sides yields

$$\big| T_p(P,Q) - T_{J,p}^{\mathrm{lin}}(P,Q) \big| \le \delta_J \lesssim_{d,p} 2^{-J\beta_p}. \qquad \square$$

The next corollary identifies the *critical linearization scale* at which the reduction error matches the target p-cost scale η_N^p, where

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

Corollary 13.3 (Critical linearization reduction)**.** *Assume $d > 2p$, $p > 1$, and $n = m = N$. Let $J_{\mathrm{lin}} = J_{\mathrm{lin}}(N)$ be any level satisfying*

$$2^{-J_{\mathrm{lin}}\beta_p} \asymp \eta_N^p.$$

Equivalently,

$$2^{J_{\mathrm{lin}} d} \asymp (N \log N)^{p/\beta_p}.$$

Then

$$\sup_{P,Q \in \mathcal{P}_d} \big| W_p(P,Q)^p - T_{J_{\mathrm{lin}},p}^{\mathrm{lin}}(P,Q) \big| \lesssim_{d,p} \eta_N^p.$$

In particular, for $1 < p \le 2$ one has

$$2^{J_{\mathrm{lin}} d} \asymp N \log N,$$

so the linearized reduction occurs on the same critical grid size as the diagonal theorem and the critical-grid reduction.

Proof. This is immediate from Theorem 13.2 after choosing J_{lin} so that $2^{-J_{\mathrm{lin}}\beta_p} \asymp \eta_N^p$. The identity for the grid size is obtained by taking d-th powers. If $1 < p \le 2$, then $\beta_p = p$, so $2^{J_{\mathrm{lin}}d} \asymp N \log N$. □

Remark 13.4 (A sharpened intrinsically Euclidean route). For $p > 1$, any estimator of $T^{\mathrm{lin}}_{J_{\mathrm{lin}},p}(P,Q)$ with absolute error $O(\eta_N^p)$ would immediately yield an estimator of $W_p(P,Q)$ with absolute error $O(\eta_N)$, since $T^{\mathrm{lin}}_{J_{\mathrm{lin}},p}(P,Q) \ge -C\eta_N^p$ by Theorem 13.3 and the map $t \mapsto t^{1/p}$ is $1/p$-Hölder on $[0,\infty)$. Thus the unrestricted balanced supercritical problem reduces, for $p > 1$, to a specific *piecewise affine transport-cost estimation problem.* This is substantially sharper than the earlier formulation "find an intrinsically Euclidean estimator".

The reduction above removes the kernel singularity, but not the combinatorics of how mass is split *inside a cell* between multiple destinations. The next proposition shows that this reduced functional does not collapse even to the full dyadic cell masses and the cellwise first moments.

Proposition 13.5 (The linearized reduced functional is not determined by masses and means)**.** *Assume $p > 1$, and work in dimension $d = 1$ at dyadic level $J = 2$. Let*

$$R = [0, \tfrac{1}{4}), \qquad S_1 = [\tfrac{1}{2}, \tfrac{3}{4}), \qquad S_2 = [\tfrac{3}{4}, 1],$$

whose centers are

$$c_R = \tfrac{1}{8}, \qquad c_{S_1} = \tfrac{5}{8}, \qquad c_{S_2} = \tfrac{7}{8}.$$

For $0 < a < b < \frac{1}{8}$, define

$$P_a := \frac{1}{2}\delta_{c_R - a} + \frac{1}{2}\delta_{c_R + a}, \qquad P_b := \frac{1}{2}\delta_{c_R - b} + \frac{1}{2}\delta_{c_R + b},$$

and

$$Q := \frac{1}{2}\delta_{c_{S_1}} + \frac{1}{2}\delta_{c_{S_2}}.$$

Then:

(i) *P_a and P_b have the same masses on every dyadic cell of $\mathcal{D}_2$;*

(ii) *for every $D \in \mathcal{D}_2$,*

$$\int_D x\, dP_a(x) = \int_D x\, dP_b(x);$$

(iii) *nevertheless,*

$$T^{\mathrm{lin}}_{2,p}(P_a, Q) \ne T^{\mathrm{lin}}_{2,p}(P_b, Q).$$

More precisely,

$$T^{\mathrm{lin}}_{2,p}(P_b, Q) - T^{\mathrm{lin}}_{2,p}(P_a, Q) = -\frac{p}{2}\Big[\Big(\frac{3}{4}\Big)^{p-1} - \Big(\frac{1}{2}\Big)^{p-1}\Big](b - a) < 0.$$

Proof. Claims (i) and (ii) are immediate: both P_a and P_b place total mass 1 in the single cell R and zero mass elsewhere, and their first moment inside R is

$$\int_R x\, dP_a(x) = \int_R x\, dP_b(x) = c_R = \frac{1}{8}.$$

Now compute the linearized cost. Since Q is supported on the two destination centers, the relevant coefficients are

$$g_1 := \nabla h_p(c_R - c_{S_1}) = -p\Big(\frac{1}{2}\Big)^{p-1}, \qquad g_2 := \nabla h_p(c_R - c_{S_2}) = -p\Big(\frac{3}{4}\Big)^{p-1}.$$

Thus

$$g_2 < g_1 < 0.$$

Also set

$$A_1 := \Big(\frac{1}{2}\Big)^p, \qquad A_2 := \Big(\frac{3}{4}\Big)^p.$$

For $t \in (-\frac{1}{8}, \frac{1}{8})$,

$$c^{\mathrm{lin}}_{2,p}(c_R + t, c_{S_i}) = A_i + g_i t, \qquad i \in \{1, 2\}.$$

Consider the 2×2 transport matrix between the two atoms of P_a and the two atoms of Q, each of mass $1/2$. Every feasible coupling is a convex combination of the two permutations. The permutation sending $c_R - a$ to c_{S_1} and $c_R + a$ to c_{S_2} has cost

$$\frac{1}{2}(A_1 - g_1 a) + \frac{1}{2}(A_2 + g_2 a) = \frac{A_1 + A_2}{2} + \frac{a}{2}(g_2 - g_1).$$

The opposite permutation has cost

$$\frac{1}{2}(A_1 + g_1 a) + \frac{1}{2}(A_2 - g_2 a) = \frac{A_1 + A_2}{2} + \frac{a}{2}(g_1 - g_2).$$

Since $g_2 - g_1 < 0$, the first permutation is strictly cheaper, so

$$T^{\mathrm{lin}}_{2,p}(P_a, Q) = \frac{A_1 + A_2}{2} + \frac{a}{2}(g_2 - g_1).$$

The same argument with a replaced by b gives

$$T^{\mathrm{lin}}_{2,p}(P_b, Q) = \frac{A_1 + A_2}{2} + \frac{b}{2}(g_2 - g_1).$$

Subtracting the two identities yields

$$T^{\mathrm{lin}}_{2,p}(P_b, Q) - T^{\mathrm{lin}}_{2,p}(P_a, Q) = \frac{b - a}{2}(g_2 - g_1),$$

which is exactly the displayed formula and is strictly negative. This proves (iii). □

Discussion 13.6 (What the linearization reduction isolates). Theorems 13.3 and 13.5 sharpen the remaining difficulty in a precise way.

For $p > 1$, the singular Euclidean kernel can already be removed at the target cost scale: the unrestricted problem reduces to estimating a piecewise affine transport functional on a critical dyadic grid. However, Theorem 13.5 shows that this reduced functional is not encoded by the dyadic cell masses and not even by the cellwise first moments. The remaining obstruction is a genuinely *intra-cell splitting phenomenon*: how the source mass inside one cell is partitioned across several destinations with different linearized gradients.

Thus the next credible full attack is no longer "find a better coarse statistic". It must instead control these splitting statistics, either by proving that they are estimable at the critical scale,

or by showing that they can be recursively localized until one falls back into the exact diagonal regime.

Remark 13.7. The case $p = 1$ again stands apart. The kernel $z \mapsto \|z\|_2$ is only Lipschitz, so there is no first-order linearization with remainder $O(2^{-J\beta})$ for any $\beta > 1$. This is perfectly consistent with the earlier fact that the off-diagonal smooth-cost argument does not improve the generic $N^{-1/d}$ absolute scale when $p = 1$.

14 Full positive-distance Euclidean product neighborhoods

The coarse separated-class upper bound of Section 12 still leaves open whether large off-diagonal regions of the Euclidean model admit theorem-level minimax statements. For $p > 1$, the answer is yes on broad fixed positive-distance neighborhoods.

Throughout this section we assume

$$d \geq 5, \qquad p > 1, \qquad n = m = N \geq 2.$$

Fix

$$\gamma := \frac{1}{20}, \qquad z_\star := \left(\frac{1}{4}, \frac{1}{2}, \ldots, \frac{1}{2}\right) \in [0,1]^d.$$

Let

$$X_\star := \{u \in \mathbb{R}^d : \ \|u\|_2 \leq \gamma\}, \qquad Y_\star := e_1 + X_\star,$$

and for $0 < r \leq \frac{1}{4}$, define the scaled Euclidean neighborhoods

$$X_r := z_\star + rX_\star, \qquad Y_r := z_\star + rY_\star.$$

Finally set

$$\mathcal{C}_r := \mathcal{P}(X_r) \times \mathcal{P}(Y_r).$$

Lemma 14.1 (Template geometry)**.** *For every $0 < r \leq \frac{1}{4}$ and every $x \in X_r$, $y \in Y_r$,*

$$\frac{9}{10} r \leq \|x - y\|_2 \leq \frac{11}{10} r.$$

Consequently, for every $(P, Q) \in \mathcal{C}_r$,

$$\frac{9}{10} r \leq W_p(P, Q) \leq \frac{11}{10} r.$$

Proof. Write

$$x = z_\star + ru, \qquad y = z_\star + rv,$$

with $u \in X_\star$ and $v \in Y_\star = e_1 + X_\star$. Then

$$x - y = r(u - v), \qquad u - v \in -e_1 + \{w : \ \|w\|_2 \leq 2\gamma\}.$$

Hence

$$1 - 2\gamma \leq \|u - v\|_2 \leq 1 + 2\gamma,$$

which is exactly the displayed pointwise bound because $\gamma = 1/20$. Integrating the pointwise inequalities over any coupling of P and Q, and then taking the infimum over couplings, proves the same bounds for $W_p(P, Q)$. □

Let

$$Z_\star := X_\star - Y_\star = -e_1 + \{w : \ \|w\|_2 \le 2\gamma\}.$$

Lemma 14.2 (Uniform smoothness and strong convexity on $Z_\star$)**.** *Assume $p > 1$, and define $h_p(z) := \|z\|_2^p$. Then $h_p \in C^2(U_\star)$ on an open neighborhood $U_\star \supset Z_\star$, and there exist constants $\Lambda_{d,p}, \lambda_{d,p} > 0$ such that*

$$\|h_p\|_{C^2(U_\star)} \le \Lambda_{d,p},$$

and

$$h_p(z) - h_p(z_0) - \nabla h_p(z_0)\cdot (z - z_0) \ge \lambda_{d,p}\|z - z_0\|_2^2 \qquad (z, z_0 \in U_\star).$$

In particular, the template cost $c_\star(u, v) := \|u - v\|_2^p$ satisfies the assumptions (H0)*,* (H1)*, and* (H4) *of [5] with $\alpha = 2$ on $X_\star \times Y_\star$.*

Proof. Because $Z_\star \subset \{z : \ \frac{9}{10} \le \|z\|_2 \le \frac{11}{10}\}$, the set $Z_\star$ lies in a compact annulus bounded away from the origin. Choose any open neighborhood $U_\star$ of $Z_\star$ still contained in such an annulus. On $U_\star$, the function $h_p(z) = \|z\|_2^p$ is C^∞, so its C^2 norm is finite.

For $z \ne 0$,

$$\nabla^2 h_p(z) = p\|z\|_2^{p-2} I_d + p(p-2)\|z\|_2^{p-4} zz^\top.$$

Its eigenvalues are

$$p\|z\|_2^{p-2} \quad \text{(multiplicity } d-1\text{)}, \qquad p(p-1)\|z\|_2^{p-2} \quad \text{(radial direction)}.$$

Since $\|z\|_2$ stays in a compact subset of $(0, \infty)$ on $U_\star$, all eigenvalues are bounded above and below by positive constants depending only on d and p. The uniform lower bound on $\nabla^2 h_p$ implies the displayed quadratic inequality by the standard integral form of Taylor's theorem. The final sentence is exactly the specialization of the smoothness and curvature hypotheses in [5] to the compact separated template $X_\star \times Y_\star$. □

For $0 < r \le \frac{1}{4}$, define the dilation

$$S_r(u) := z_\star + ru, \qquad u \in \mathbb{R}^d.$$

For $P \in \mathcal{P}(X_r)$ and $Q \in \mathcal{P}(Y_r)$, write

$$\overline{P} := (S_r^{-1})_\# P \in \mathcal{P}(X_\star), \qquad \overline{Q} := (S_r^{-1})_\# Q \in \mathcal{P}(Y_\star).$$

Then every coupling between P and Q is the S_r-pushforward of a coupling between $\overline{P}$ and $\overline{Q}$, so

$$W_p(P, Q)^p = r^p T_\star(\overline{P}, \overline{Q}), \qquad T_\star(\mu, \nu) := T_{c_\star}(\mu, \nu).$$

Theorem 14.3 (Full product neighborhoods away from the diagonal: the p-cost)**.** *Assume $d \ge 5$ and $p > 1$. There exist constants $0 < c_{d,p} \le C_{d,p} < \infty$ such that for every $0 < r \le \frac{1}{4}$,*

$$c_{d,p}\, r^p (N \log N)^{-2/d} \le \inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_r} \mathbb{E}|\widehat{T} - W_p(P, Q)^p| \le C_{d,p}\, r^p N^{-2/d}.$$

Proof. By the exact scaling relation above, estimating $W_p(P, Q)^p$ on $\mathcal{C}_r$ is equivalent, after division by r^p, to estimating the template transport cost $T_\star(\mu, \nu)$ on $\mathcal{P}(X_\star) \times \mathcal{P}(Y_\star)$.

By Theorem 14.2, the template problem satisfies the hypotheses of [5, Theorem 22] with $\alpha = 2$, so

$$\inf_{\widehat{T}_\star} \sup_{\mu\in\mathcal{P}(X_\star),\,\nu\in\mathcal{P}(Y_\star)} \mathbb{E}\big|\widehat{T}_\star - T_\star(\mu,\nu)\big| \gtrsim_{d,p} (N\log N)^{-2/d}.$$

Multiplying by r^p gives the lower bound.

For the upper bound, $X_\star$ and $Y_\star$ are disjoint compact convex sets, so [5, Corollary 3(ii)] yields

$$\sup_{\mu\in\mathcal{P}(X_\star),\,\nu\in\mathcal{P}(Y_\star)} \mathbb{E}\big|T_\star(\mu_N,\nu_N) - T_\star(\mu,\nu)\big| \lesssim_{d,p} N^{-2/d}.$$

Scaling back by r^p proves the upper bound, and the empirical plug-in estimator $W_p(P_N,Q_N)^p$ attains it up to constants. □

Corollary 14.4 (Full product neighborhoods away from the diagonal: the distance itself)**.** *Assume $d \geq 5$ and $p > 1$. There exist constants $0 < c_{d,p} \leq C_{d,p} < \infty$ such that for every $0 < r \leq \frac{1}{4}$,*

$$c_{d,p}\, r(N\log N)^{-2/d} \leq \inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_r} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \leq C_{d,p}\, rN^{-2/d}.$$

Proof. By Theorem 14.1, all values $W_p(P,Q)$ on $\mathcal{C}_r$ lie in the interval $[\frac{9}{10}r, \frac{11}{10}r]$. On this interval the map $g(x) = x^p$ is bi-Lipschitz:

$$\Big(\frac{9}{10}r\Big)^{p-1}|x-y| \leq |x^p - y^p| \leq p\Big(\frac{11}{10}r\Big)^{p-1}|x-y|.$$

For the upper bound, let $\widehat{T}$ be any estimator of W_p^p on $\mathcal{C}_r$, clip it to $[\big(\frac{9}{10}r\big)^p, \big(\frac{11}{10}r\big)^p]$, and define $\widehat{W} := \widehat{T}^{1/p}$. Then

$$|\widehat{W} - W_p(P,Q)| \lesssim_p r^{1-p}|\widehat{T} - W_p(P,Q)^p|.$$

Applying Theorem 14.3 gives the upper bound.

For the lower bound, let $\widehat{W}$ be any estimator of W_p, clip it to $[\frac{9}{10}r, \frac{11}{10}r]$, and define $\widehat{T} := \widehat{W}^p$. Then

$$|\widehat{T} - W_p(P,Q)^p| \lesssim_p r^{p-1}|\widehat{W} - W_p(P,Q)|.$$

Taking expectations and suprema, and invoking the lower bound in Theorem 14.3, yields the claim. □

Remark 14.5. For $p = 1$, the same separated product neighborhoods still enjoy the empirical upper bound

$$\sup_{(P,Q)\in\mathcal{C}_r} \mathbb{E}\big|W_1(P_N,Q_N) - W_1(P,Q)\big| \lesssim_d rN^{-2/d}$$

by the smooth-cost theory for the transport cost itself. What fails in that case is the lower-bound input with $\alpha = 2$: the Euclidean norm is smooth away from the origin but not strongly convex there, so the minimax lower theory of [5] does not presently force the same off-diagonal exponent for W_1.

15 A mixed-scale phase diagram on full two-ball neighborhoods

The positive-distance neighborhood theorem of Section 14 treats the case of two separated r-scale neighborhoods in a coarse way: it records the r^p-scale smooth-cost behavior, but it does not yet distinguish the much thinner regime in which each marginal is confined to a ball of radius $s \ll r$. On such classes the transport cost admits an exact first-plus-second-order decomposition, and the residual nonlinear piece is uniformly comparable to a fixed quadratic transport cost. This yields a full mixed-scale phase diagram on the *entire* product neighborhood.

Throughout this section we assume

$$d \geq 5, \qquad p > 1, \qquad n = m = N \geq 2,$$

and we set

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

Let

$$B := \{u \in \mathbb{R}^d : \ \|u\|_2 \leq 1\}, \qquad z_\diamond := \Big(\frac{1}{4}, \frac{1}{2}, \ldots, \frac{1}{2}\Big) \in [0,1]^d.$$

Fix parameters

$$0 < r \leq \frac{1}{4}, \qquad 0 < s \leq \frac{r}{20}, \qquad \kappa := \frac{s}{r} \in \Big(0, \frac{1}{20}\Big].$$

Define the two balls

$$X_{r,s} := z_\diamond + sB, \qquad Y_{r,s} := z_\diamond + re_1 + sB,$$

and the full product neighborhood

$$\mathcal{C}_{r,s} := \mathcal{P}(X_{r,s}) \times \mathcal{P}(Y_{r,s}).$$

Lemma 15.1 (Two-ball geometry)**.** *For every $x \in X_{r,s}$ and $y \in Y_{r,s}$,*

$$\frac{9}{10} r \leq \|x - y\|_2 \leq \frac{11}{10} r.$$

Consequently, for every $(P, Q) \in \mathcal{C}_{r,s}$,

$$\frac{9}{10} r \leq W_p(P, Q) \leq \frac{11}{10} r.$$

Proof. Write

$$x = z_\diamond + su, \qquad y = z_\diamond + re_1 + sv, \qquad u, v \in B.$$

Then

$$x - y = -re_1 + s(u - v),$$

so

$$r - 2s \leq \|x - y\|_2 \leq r + 2s.$$

Because $s \leq r/20$, this gives

$$\frac{9}{10} r \leq r - 2s \leq \|x - y\|_2 \leq r + 2s \leq \frac{11}{10} r.$$

Integrating the pointwise inequalities over any coupling and taking the infimum proves the same bounds for $W_p(P, Q)$. □

Let

$$S_{r,s}(u) := z_\diamond + su, \qquad T_{r,s}(v) := z_\diamond + re_1 + sv, \qquad u, v \in \mathbb{R}^d.$$

For $(P, Q) \in \mathcal{C}_{r,s}$, define the rescaled marginals on B by

$$\mu := (S_{r,s}^{-1})_\# P \in \mathcal{P}(B), \qquad \nu := (T_{r,s}^{-1})_\# Q \in \mathcal{P}(B).$$

Let

$$h_p(z) := \|z\|_2^p, \qquad b_p := \nabla h_p(-e_1) = -pe_1.$$

For $t \in 2B$, define the residual kernel

$$g_{\kappa,p}(t) := \frac{h_p(-e_1 + \kappa t) - h_p(-e_1) - \kappa\, b_p \cdot t}{\kappa^2}.$$

Finally write

$$U_{\kappa,p}(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int g_{\kappa,p}(u - v)\, d\pi(u, v).$$

15.1 Exact first-plus-second-order decomposition

The first key fact is exact: on the full class $\mathcal{C}_{r,s}$, the transport cost splits into a constant baseline term, a one-dimensional parametric mean term, and a residual transport functional of size $r^{p-2}s^2$.

Proposition 15.2 (Exact decomposition on $\mathcal{C}_{r,s}$)**.** *For every $(P, Q) \in \mathcal{C}_{r,s}$, with rescaled marginals (μ, ν) as above,*

$$W_p(P, Q)^p = r^p + r^{p-1} s\, b_p \cdot \Big(\int u\, d\mu(u) - \int v\, d\nu(v) \Big) + r^{p-2} s^2\, U_{\kappa,p}(\mu, \nu).$$

Proof. Let $\pi \in \Pi(\mu, \nu)$, and let $\widetilde{\pi} := (S_{r,s}, T_{r,s})_\# \pi \in \Pi(P, Q)$. For $(u, v) \in B \times B$,

$$S_{r,s}(u) - T_{r,s}(v) = -re_1 + s(u - v) = r\big(-e_1 + \kappa(u - v)\big).$$

Therefore

$$\begin{aligned} \int \|x - y\|_2^p\, d\widetilde{\pi}(x, y) &= r^p \int h_p(-e_1 + \kappa(u - v))\, d\pi(u, v) \\ &= r^p + r^p \kappa\, b_p \cdot \int (u - v)\, d\pi(u, v) \\ &\qquad + r^p \kappa^2 \int g_{\kappa,p}(u - v)\, d\pi(u, v). \end{aligned}$$

Since the marginals of π are μ and ν,

$$\int (u - v)\, d\pi(u, v) = \int u\, d\mu(u) - \int v\, d\nu(v).$$

Taking the infimum over all $\pi \in \Pi(\mu, \nu)$ proves the identity, because the constant and linear terms are independent of the coupling while the residual term is exactly $U_{\kappa,p}(\mu, \nu)$. □

The residual kernel is a genuine second-order remainder around the baseline direction $-e_1$. Its Hessian is therefore the ambient Euclidean Hessian evaluated away from the singularity, which makes the remainder uniformly smooth and uniformly comparable to a fixed quadratic form.

15.2 The residual kernel and the quadratic comparison

Let

$$H_p := \nabla^2 h_p(-e_1) = pI_d + p(p-2)e_1 e_1^\top,$$

and define the positive-definite quadratic form

$$q_p^\circ(t) := \frac{1}{2}\, t^\top H_p t.$$

Equivalently,

$$q_p^\circ(t) = \frac{p}{2}\sum_{i=2}^{d} t_i^2 + \frac{p(p-1)}{2}\, t_1^2.$$

Lemma 15.3 (Uniform integral remainder formula)**.** *For every $0 < \kappa \le 1/20$ and every $t \in 2B$,*

$$g_{\kappa,p}(t) = \int_0^1 (1-\tau)\, t^\top \nabla^2 h_p(-e_1 + \tau\kappa t)\, t\, d\tau.$$

In particular, $g_{\kappa,p}(t) \ge 0$.

Proof. Apply Taylor's theorem with integral remainder to the one-variable function

$$\phi(\lambda) := h_p(-e_1 + \lambda t), \qquad 0 \le \lambda \le \kappa.$$

Then

$$h_p(-e_1 + \kappa t) = h_p(-e_1) + \kappa\, \phi'(0) + \int_0^\kappa (\kappa - \lambda)\phi''(\lambda)\, d\lambda.$$

After the change of variables $\lambda = \kappa\tau$, this becomes exactly the displayed formula. Nonnegativity follows because h_p is convex for $p > 1$. □

Lemma 15.4 (Quadratic comparison for the residual kernel)**.** *There exists a constant $\kappa_0 = \kappa_0(d,p) \in (0, 1/20]$ such that for every $0 < \kappa \le \kappa_0$ and every $t \in 2B$,*

$$\frac{1}{2}\, q_p^\circ(t) \le g_{\kappa,p}(t) \le \frac{3}{2}\, q_p^\circ(t).$$

Consequently, for all $\mu, \nu \in \mathcal{P}(B)$,

$$\frac{1}{2}\, T_{q_p^\circ}(\mu,\nu) \le U_{\kappa,p}(\mu,\nu) \le \frac{3}{2}\, T_{q_p^\circ}(\mu,\nu),$$

where

$$T_{q_p^\circ}(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int q_p^\circ(u-v)\, d\pi(u,v).$$

Proof. Because $p > 1$, the map h_p is C^∞ on

$$\mathcal{A} := \{z \in \mathbb{R}^d :\ 1/2 \le \|z\|_2 \le 3/2\}.$$

For $0 < \kappa \le 1/20$, $t \in 2B$, and $0 \le \tau \le 1$, one has

$$\| -e_1 + \tau\kappa t\|_2 \in [1 - 2\kappa, 1 + 2\kappa] \subset [9/10, 11/10] \subset (1/2, 3/2),$$

so all points $-e_1 + \tau\kappa t$ lie in the compact annulus $\mathcal{A}$.

Since $\nabla^2 h_p$ is continuous on $\mathcal{A}$ and

$$H_p = \nabla^2 h_p(-e_1)$$

is positive definite, there exists $\kappa_0 = \kappa_0(d,p)$ such that for every $0 < \kappa \le \kappa_0$, $t \in 2B$, and $0 \le \tau \le 1$,

$$\frac{1}{2} H_p \preceq \nabla^2 h_p(-e_1 + \tau\kappa t) \preceq \frac{3}{2} H_p$$

in the Loewner order. Multiplying by t on both sides and integrating τ against $(1-\tau)\,d\tau$, Theorem 15.3 yields

$$\frac{1}{2}\int_0^1 (1-\tau)\, t^\top H_p t\, d\tau \le g_{\kappa,p}(t) \le \frac{3}{2}\int_0^1 (1-\tau)\, t^\top H_p t\, d\tau.$$

Since $\int_0^1 (1-\tau) d\tau = 1/2$, this is exactly

$$\frac{1}{2} q_p^\circ(t) \le g_{\kappa,p}(t) \le \frac{3}{2} q_p^\circ(t).$$

The transport-cost comparison is obtained by integrating the pointwise inequalities over any coupling $\pi \in \Pi(\mu,\nu)$ and then taking infima. □

The next proposition extracts a theorem-level lower bound for the quadratic reference cost $T_{q_p^\circ}$ from the diagonal exact theorem already proved earlier in the paper.

Proposition 15.5 (Local exact lower bound for a quadratic transport cost)**.** *There exist constants $a_{d,p}, c_{d,p} > 0$ such that for all sufficiently large N,*

$$\inf_{\widehat{U}} \sup_{\substack{\mu,\nu \in \mathcal{P}(B):\\ T_{q_p^\circ}(\mu,\nu) \le a_{d,p}\eta_N^2}} \mathbb{E}\big|\widehat{U} - T_{q_p^\circ}(\mu,\nu)\big| \ge c_{d,p}\eta_N^2.$$

Proof. Let A_p be the unique positive-definite matrix such that

$$q_p^\circ(t) = \|A_p t\|_2^2.$$

Then

$$T_{q_p^\circ}(\mu,\nu)^{1/2} = W_2\big((A_p)_\#\mu, (A_p)_\#\nu\big).$$

The image $A_p(B)$ is a fixed ellipsoid in $\mathbb{R}^d$. Choose any affine bijection L_p from a neighborhood of $A_p(B)$ into $[0,1]^d$. Because L_p is bi-Lipschitz with constants depending only on d and p, estimating $W_2((A_p)_\#\mu, (A_p)_\#\nu)$ on $A_p(B)$ is equivalent, up to fixed multiplicative constants, to estimating the standard Euclidean W_2-distance on a fixed compact subset of $[0,1]^d$.

Apply Theorem 9.4 with $p = 2$ to the pushed-forward problem. Since $d \ge 5 > 4 = 2p$ in the quadratic model, there exist constants $a'_{d,p}, c'_{d,p} > 0$ such that

$$\inf_{\widehat{W}} \sup_{\substack{\mu,\nu \in \mathcal{P}(B):\\ T_{q_p^\circ}(\mu,\nu) \le a'_{d,p}\eta_N^2}} \mathbb{E}\big(\widehat{W} - T_{q_p^\circ}(\mu,\nu)^{1/2}\big)^2 \ge c'_{d,p}\eta_N^2.$$

Now let $\widehat{U}$ be any estimator of $T_{q_p^\circ}(\mu,\nu)$, and define

$$\widehat{W} := (\widehat{U}_+)^{1/2}.$$

For all $a, b \geq 0$,

$$\left|\sqrt{a} - \sqrt{b}\right|^2 \leq |a - b|.$$

Hence

$$\left(\widehat{W} - T_{q_p^\circ}(\mu,\nu)^{1/2}\right)^2 \leq \left|\widehat{U} - T_{q_p^\circ}(\mu,\nu)\right|.$$

Taking expectations and then suprema shows that the minimax absolute risk for the cost dominates the minimax squared risk for its square-root distance. This yields the claim after relabeling constants. □

Proposition 15.6 (Residual lower bound on the local quadratic scale)**.** *There exist constants $\kappa_1 = \kappa_1(d,p) \in (0, \kappa_0]$, $a^\sharp_{d,p} > 0$, and $c^\sharp_{d,p} > 0$ such that for every $0 < \kappa \leq \kappa_1$ and all sufficiently large N,*

$$\inf_{\widehat{U}} \sup_{\substack{\mu,\nu \in \mathcal{P}(B): \\ T_{q_p^\circ}(\mu,\nu) \leq a^\sharp_{d,p}\eta_N^2}} \mathbb{E}\left|\widehat{U} - U_{\kappa,p}(\mu,\nu)\right| \geq c^\sharp_{d,p}\eta_N^2.$$

Proof. Fix $\kappa \leq \kappa_0$, and let $\widehat{U}$ be any estimator of $U_{\kappa,p}$. On the class

$$\mathcal{G}_N := \{(\mu,\nu) \in \mathcal{P}(B)^2 : \; T_{q_p^\circ}(\mu,\nu) \leq a_{d,p}\eta_N^2\},$$

Theorem 15.4 gives

$$\left|U_{\kappa,p}(\mu,\nu) - T_{q_p^\circ}(\mu,\nu)\right| \leq \frac{1}{2} T_{q_p^\circ}(\mu,\nu) \leq \frac{1}{2} a_{d,p}\eta_N^2.$$

Therefore

$$\left|\widehat{U} - T_{q_p^\circ}(\mu,\nu)\right| \leq \left|\widehat{U} - U_{\kappa,p}(\mu,\nu)\right| + \frac{1}{2} a_{d,p}\eta_N^2.$$

Taking expectations, suprema over $\mathcal{G}_N$, and infima over $\widehat{U}$, and then using Theorem 15.5, yields

$$\inf_{\widehat{U}} \sup_{(\mu,\nu)\in\mathcal{G}_N} \mathbb{E}\left|\widehat{U} - U_{\kappa,p}(\mu,\nu)\right| \geq \Big(c_{d,p} - \frac{1}{2} a_{d,p}\Big)\eta_N^2.$$

If necessary, shrink $a_{d,p}$ so that $c_{d,p} - a_{d,p}/2 > 0$, and rename the resulting constants as $a^\sharp_{d,p}$ and $c^\sharp_{d,p}$. □

15.3 Upper and lower envelopes on the full two-ball class

We first record the parametric lower bound coming from the exact one-dimensional first-order term.

Lemma 15.7 (Parametric lower bound on $\mathcal{C}_{r,s}$)**.** *There exists a constant $c_{d,p} > 0$ such that*

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\left|\widehat{T} - W_p(P,Q)^p\right| \geq c_{d,p}\, r^{p-1} s\, N^{-1/2},$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\left|\widehat{W} - W_p(P,Q)\right| \geq c_{d,p}\, s\, N^{-1/2}.$$

Proof. Fix

$$x_\pm := z_\diamond \pm s e_1, \qquad y_0 := z_\diamond + r e_1.$$

These points belong to $X_{r,s}$ and $Y_{r,s}$, respectively, because $s \le r/20 \le 1/20$. For $t \in [1/4, 3/4]$, define

$$P_t := t\,\delta_{x_+} + (1-t)\,\delta_{x_-}, \qquad Q := \delta_{y_0}.$$

Then $(P_t, Q) \in \mathcal{C}_{r,s}$, and

$$W_p(P_t, Q)^p = t(r-s)^p + (1-t)(r+s)^p = \alpha_{r,s,p} + \beta_{r,s,p} t,$$

where

$$\beta_{r,s,p} := (r-s)^p - (r+s)^p.$$

By the mean value theorem, for some $\xi \in (r-s, r+s)$,

$$|\beta_{r,s,p}| = 2p\,\xi^{p-1} s \asymp_p r^{p-1} s,$$

uniformly over $0 < s \le r/20$. Hence $t \mapsto W_p(P_t, Q)^p$ is bi-Lipschitz on $[1/4, 3/4]$ with Lipschitz constants of order $r^{p-1}s$.

Estimating $W_p(P_t, Q)^p$ over this subfamily is therefore equivalent, up to constants, to estimating the Bernoulli mean t from N i.i.d. samples, which has minimax absolute risk of order $N^{-1/2}$. This proves the first displayed lower bound.

For the distance itself, the map $x \mapsto x^{1/p}$ is bi-Lipschitz on $[(r-s)^p, (r+s)^p]$ with derivative comparable to r^{1-p}, so $t \mapsto W_p(P_t, Q)$ is bi-Lipschitz on $[1/4, 3/4]$ with slope of order s. The same Bernoulli argument then yields the second lower bound. □

We now prove the corresponding upper bound by estimating the mean term directly and the residual term by plug-in for the smooth bounded cost $g_{\kappa,p}$.

Proposition 15.8 (Direct estimator on $\mathcal{C}_{r,s}$)**.** *There exists an estimator $\widehat{T}_{r,s}$ such that*

$$\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{T}_{r,s} - W_p(P,Q)^p| \lesssim_{d,p} r^{p-1} s\, N^{-1/2} + r^{p-2} s^2 N^{-2/d}.$$

Consequently, there exists an estimator $\widehat{W}_{r,s}$ such that

$$\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{W}_{r,s} - W_p(P,Q)| \lesssim_{d,p} s\, N^{-1/2} + \frac{s^2}{r} N^{-2/d}.$$

Proof. Let

$$U_i := S_{r,s}^{-1}(X_i) \in B, \qquad V_i := T_{r,s}^{-1}(Y_i) \in B,$$

and let

$$\widehat{\mu}_N := \frac{1}{N}\sum_{i=1}^N \delta_{U_i}, \qquad \widehat{\nu}_N := \frac{1}{N}\sum_{i=1}^N \delta_{V_i}.$$

These are the empirical measures of the rescaled marginals $\mu, \nu \in \mathcal{P}(B)$.

Define

$$\overline{U}_N := \frac{1}{N}\sum_{i=1}^N U_i, \qquad \overline{V}_N := \frac{1}{N}\sum_{i=1}^N V_i,$$

and

$$\widehat{U}_{\kappa,p} := T_{g_{\kappa,p}}(\widehat{\mu}_N, \widehat{\nu}_N).$$

Set

$$\widehat{T}_{r,s} := r^p + r^{p-1} s\, b_p \cdot (\overline{U}_N - \overline{V}_N) + r^{p-2} s^2\, \widehat{U}_{\kappa,p}.$$

By Theorem 15.2,

$$\begin{aligned}\widehat{T}_{r,s} - W_p(P,Q)^p = r^{p-1} s\, b_p \cdot \Big[(\overline{U}_N - \overline{V}_N) - \Big(\int u\, d\mu - \int v\, d\nu\Big)\Big] \\ + r^{p-2} s^2 \Big[\widehat{U}_{\kappa,p} - U_{\kappa,p}(\mu,\nu)\Big].\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}|\widehat{T}_{r,s} - W_p(P,Q)^p| \le r^{p-1} s\, |b_p|\, \mathbb{E}\left\|(\overline{U}_N - \overline{V}_N) - \Big(\int u\, d\mu - \int v\, d\nu\Big)\right\|_2 \\ + r^{p-2} s^2 \mathbb{E}|\widehat{U}_{\kappa,p} - U_{\kappa,p}(\mu,\nu)|. \qquad (15.1)\end{aligned}$$

Because $U_i, V_i \in B$, each coordinate has variance at most 1, so

$$\mathbb{E}\left\|(\overline{U}_N - \overline{V}_N) - \Big(\int u\, d\mu - \int v\, d\nu\Big)\right\|_2 \lesssim_d N^{-1/2}.$$

This controls the first term in Equation (15.1) by

$$\lesssim_{d,p} r^{p-1} s\, N^{-1/2}.$$

For the residual term, note that $g_{\kappa,p}$ is a bounded C^2 function on $2B$ with C^2-norm bounded uniformly over $0 < \kappa \le 1/20$; indeed

$$\nabla g_{\kappa,p}(t) = \kappa^{-1}\Big(\nabla h_p(-e_1 + \kappa t) - \nabla h_p(-e_1)\Big), \qquad \nabla^2 g_{\kappa,p}(t) = \nabla^2 h_p(-e_1 + \kappa t),$$

and all arguments $-e_1 + \kappa t$ stay in a fixed compact annulus away from the origin. The upper-bound part of [5, Theorem 2], which the authors explicitly note remains valid for general bounded smooth costs h, therefore gives

$$\sup_{\mu,\nu \in \mathcal{P}(B)} \mathbb{E}|\widehat{U}_{\kappa,p} - U_{\kappa,p}(\mu,\nu)| \lesssim_{d,p} N^{-2/d}.$$

Substituting this into Equation (15.1) proves the cost bound.

For the distance bound, clip $\widehat{T}_{r,s}$ to the interval $[(\frac{9}{10}r)^p, (\frac{11}{10}r)^p]$ and denote the clipped estimator by $\widetilde{T}_{r,s}$. Then define

$$\widehat{W}_{r,s} := \widetilde{T}_{r,s}^{1/p}.$$

By Theorem 15.1, the true value $W_p(P,Q)^p$ also lies in this interval. On that interval, $x \mapsto x^{1/p}$ is Lipschitz with constant $\lesssim_p r^{1-p}$, hence

$$|\widehat{W}_{r,s} - W_p(P,Q)| \lesssim_p r^{1-p}\, |\widetilde{T}_{r,s} - W_p(P,Q)^p| \le r^{1-p}\, |\widehat{T}_{r,s} - W_p(P,Q)^p|.$$

Taking expectations and using the cost bound yields

$$\mathbb{E}|\widehat{W}_{r,s} - W_p(P,Q)| \lesssim_{d,p} s\, N^{-1/2} + \frac{s^2}{r}\, N^{-2/d}. \qquad \square$$

The lower bound is obtained by combining the parametric family of Theorem 15.7 with an oracle reduction to the residual local problem.

Theorem 15.9 (Mixed-scale lower envelope on the full two-ball class)**.** *There exist constants $\kappa_\star = \kappa_\star(d,p) \in (0,\kappa_1]$ and $c_{d,p} > 0$ such that the following holds. For every $0 < r \le 1/4$ and every $0 < s \le \kappa_\star r$,*

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\big|\widehat{T} - W_p(P,Q)^p\big| \ge c_{d,p} \max\Big\{r^{p-1} s\, N^{-1/2},\, r^{p-2} s^2\, \eta_N^2\Big\},$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \ge c_{d,p} \max\left\{s\, N^{-1/2},\, \frac{s^2}{r}\, \eta_N^2\right\}.$$

Proof. The parametric part is exactly Theorem 15.7. It remains to prove the residual lower bound.

Fix $0 < s \le \kappa_\star r$, let $\kappa = s/r$, and consider only those pairs $(P,Q) \in \mathcal{C}_{r,s}$ whose rescaled marginals (μ,ν) satisfy

$$T_{q_p^\circ}(\mu,\nu) \le a_{d,p}^\sharp \eta_N^2.$$

This is a restriction of the full class $\mathcal{C}_{r,s}$, so any lower bound on this subclass is also a lower bound on $\mathcal{C}_{r,s}$.

Now reveal to the estimator, as side information, the exact value of the first-order term

$$L(P,Q) := r^p + r^{p-1} s\, b_p \cdot \Big(\int u\, d\mu - \int v\, d\nu\Big).$$

This oracle information can only make the estimation problem easier. After revealing $L(P,Q)$, Theorem 15.2 shows that estimating $W_p(P,Q)^p$ is equivalent to estimating

$$r^{p-2} s^2\, U_{\kappa,p}(\mu,\nu).$$

By Theorem 15.6,

$$\inf_{\widehat{U}} \sup_{\substack{\mu,\nu\in\mathcal{P}(B):\\ T_{q_p^\circ}(\mu,\nu)\le a_{d,p}^\sharp \eta_N^2}} \mathbb{E}\big|\widehat{U} - U_{\kappa,p}(\mu,\nu)\big| \ge c_{d,p}^\sharp \eta_N^2.$$

Multiplying by $r^{p-2}s^2$ yields

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\big|\widehat{T} - W_p(P,Q)^p\big| \ge c_{d,p}\, r^{p-2} s^2 \eta_N^2.$$

Combining this with the parametric lower bound proves the cost statement.

For the distance lower bound, first convert the parametric lower bound in Theorem 15.7. For the residual part, suppose $\widehat{W}$ estimates $W_p(P,Q)$. By Theorem 15.1, on $\mathcal{C}_{r,s}$ one has

$$W_p(P,Q) \in \left[\frac{9}{10} r, \frac{11}{10} r\right].$$

Clip $\widehat{W}$ to this interval and denote the clipped estimator by $\widetilde{W}$. Then define

$$\widehat{T} := \widetilde{W}^p.$$

Because the true value already lies in the same interval, clipping cannot increase the absolute error:

$$|\widetilde{W} - W_p(P,Q)| \le |\widehat{W} - W_p(P,Q)|.$$

Moreover, on $[\frac{9}{10}r, \frac{11}{10}r]$ the map $x \mapsto x^p$ is Lipschitz with constant $\lesssim_p r^{p-1}$, so the mean value theorem gives

$$|\widehat{T} - W_p(P,Q)^p| \lesssim_p r^{p-1}\,|\widetilde{W} - W_p(P,Q)| \le r^{p-1}\,|\widehat{W} - W_p(P,Q)|.$$

Therefore any lower bound of size $r^{p-2}s^2\eta_N^2$ for the cost translates into a lower bound of size $(s^2/r)\eta_N^2$ for the distance. Taking the maximum with the parametric term $sN^{-1/2}$ proves the claim. □

Combining Theorems 15.8 and 15.9 yields the promised phase diagram.

Corollary 15.10 (Mixed-scale phase diagram on $\mathcal{C}_{r,s}$)**.** *There exist constants $\kappa_\star = \kappa_\star(d,p) \in (0, 1/20]$ and $0 < c_{d,p} \le C_{d,p} < \infty$ such that for every $0 < r \le 1/4$ and $0 < s \le \kappa_\star r$,*

$$c_{d,p}\,\max\Big\{r^{p-1}s\,N^{-1/2},\, r^{p-2}s^2\,\eta_N^2\Big\} \le \inf_{\widehat{T}}\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{T} - W_p(P,Q)^p|$$

$$\le C_{d,p}\,\left(r^{p-1}s\,N^{-1/2} + r^{p-2}s^2N^{-2/d}\right),$$

and

$$c_{d,p}\,\max\left\{s\,N^{-1/2},\, \frac{s^2}{r}\,\eta_N^2\right\} \le \inf_{\widehat{W}}\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{W} - W_p(P,Q)|$$

$$\le C_{d,p}\,\left(s\,N^{-1/2} + \frac{s^2}{r}\,N^{-2/d}\right).$$

Corollary 15.11 (Exact parametric regime)**.** *There exists $c_{d,p} > 0$ such that if*

$$0 < s \le c_{d,p}\,r\,N^{2/d-1/2},$$

then

$$\inf_{\widehat{T}}\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{T} - W_p(P,Q)^p| \asymp_{d,p} r^{p-1}s\,N^{-1/2},$$

and

$$\inf_{\widehat{W}}\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{W} - W_p(P,Q)| \asymp_{d,p} s\,N^{-1/2}.$$

Proof. Under the displayed condition,

$$r^{p-2}s^2N^{-2/d} \le c_{d,p}\,r^{p-1}sN^{-1/2}, \qquad \frac{s^2}{r}N^{-2/d} \le c_{d,p}\,sN^{-1/2}.$$

Choose $c_{d,p}$ small enough, and apply Theorem 15.10. □

Corollary 15.12 (Generic nonparametric regime)**.** *There exists $C_{d,p} > 0$ such that if*

$$C_{d,p}\,r\,N^{2/d-1/2} \le s \le \kappa_\star r,$$

then

$$r^{p-2}s^2\,\eta_N^2 \lesssim_{d,p} \inf_{\widehat{T}}\sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}|\widehat{T} - W_p(P,Q)^p| \lesssim_{d,p} r^{p-2}s^2\,N^{-2/d},$$

and

$$\frac{s^2}{r}\,\eta_N^2 \lesssim_{d,p} \inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_{r,s}} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \lesssim_{d,p} \frac{s^2}{r}\,N^{-2/d}.$$

In particular, for every fixed eccentricity ratio $s/r \in (0,\kappa_\star]$, the full two-ball class $\mathcal{C}_{r,s}$ lies in the nonparametric regime for all sufficiently large N.

Proof. If $s \geq C_{d,p} r N^{2/d-1/2}$, then

$$r^{p-1} s N^{-1/2} \leq C_{d,p}^{-1}\, r^{p-2} s^2 N^{-2/d}, \qquad s N^{-1/2} \leq C_{d,p}^{-1}\,\frac{s^2}{r} N^{-2/d}.$$

Choose $C_{d,p}$ large enough and apply Theorem 15.10. The final sentence follows because $2/d - 1/2 < 0$ when $d \geq 5$. □

Discussion 15.13 (Why the thin-annulus classes are genuinely nongeneric). Theorems 15.11 and 15.12 solve, up to the unavoidable logarithm on the nonparametric side, the complete mixed-scale picture on every full two-ball product neighborhood $\mathcal{C}_{r,s}$. The critical width is

$$s_{\mathrm{crit}}(N,r) \asymp r\, N^{2/d-1/2}.$$

Below that width, the full class is generically parametric because the first-order mean term in Theorem 15.2 dominates. Above it, the full class is generically nonparametric and behaves like a quadratic transport cost.

This also clarifies the meaning of the exact translated thin-annulus classes proved later in the paper. Those classes are much smaller than $\mathcal{C}_{r,s}$: their block symmetry annihilates the generic first-order term and leaves only a sign-indefinite second-order fluctuation. The logarithmic direct-versus-plug-in gap on thin annuli is therefore a genuine cancellation phenomenon, not the generic behavior of the whole two-ball neighborhood.

16 Exact translated thin-annulus classes away from the diagonal

The product-neighborhood theorem above shows that broad positive-distance regions are already under theorem-level control up to the logarithm inherited from the smooth-cost minimax lower bound. We now add a complementary exact Euclidean construction showing that the logarithmic advantage of direct estimation over empirical plug-in is *not* confined to the diagonal.

Throughout this section we assume

$$d \geq 2, \qquad p \geq 1, \qquad n = m = N \geq 2.$$

Lemma 16.1 (Translated paired packing)**.** *There exists a constant $c_d > 0$ such that the following holds. For every $M \geq 1$, every $0 < r \leq \frac{1}{10}$, and every*

$$0 < \delta \leq c_d\, r\, M^{-1/(d-1)},$$

one can choose points $z_1,\dots,z_M \in [0,1]^d$ such that, for

$$x_j^{\pm} := z_j \pm \frac{\delta}{2} e_2, \qquad y_j^{\pm} := z_j + r e_1 \pm \frac{\delta}{2} e_2,$$

all four points $x_j^{\pm}, y_j^{\pm}$ lie in $[0,1]^d$, and for all distinct $i \neq j$,

$$\|z_i - z_j\|_2 \geq 8\delta.$$

Proof. Choose a small absolute constant $a_d > 0$ and assume $\delta \leq a_d r M^{-1/(d-1)}$. Consider the $(d-1)$-dimensional box

$$B_r := \Big[\frac{1}{4}, \frac{1}{4} + \frac{r}{4}\Big]^{d-1}.$$

A grid in B_r with spacing 8δ in every coordinate contains at least M points when a_d is sufficiently small. Let $\zeta_1, \ldots, \zeta_M \in B_r$ be M such grid points and define

$$z_j := \frac{1}{4}e_1 + \sum_{k=2}^{d} (\zeta_j)_{k-1} e_k.$$

Then $z_j \in [0,1]^d$, the points z_i and z_j differ only in coordinates $2, \ldots, d$, and $\|z_i - z_j\|_2 \geq 8\delta$ for $i \neq j$. Because $r \leq 1/10$ and $\delta \leq a_d r$ with a_d sufficiently small, the translated points $x_j^{\pm}$ and $y_j^{\pm}$ remain inside $[0,1]^d$. □

Fix such a translated paired configuration, and for $\theta, \eta \in [-1,1]^M$, define

$$P_\theta := \frac{1}{2M} \sum_{j=1}^{M} \Big((1+\theta_j)\delta_{x_j^+} + (1-\theta_j)\delta_{x_j^-} \Big),$$

$$Q_\eta := \frac{1}{2M} \sum_{j=1}^{M} \Big((1+\eta_j)\delta_{y_j^+} + (1-\eta_j)\delta_{y_j^-} \Big).$$

As before, let q_θ, q_η denote the corresponding distributions on the alphabet

$$\mathcal{A}_M := \{(j,+), (j,-) : \ j = 1, \ldots, M\}, \qquad q_\theta(j,\pm) = \frac{1 \pm \theta_j}{2M}, \qquad q_\eta(j,\pm) = \frac{1 \pm \eta_j}{2M}.$$

Proposition 16.2 (Exact translated paired reduction)**.** *Let*

$$D_{r,\delta} := \sqrt{r^2 + \delta^2}, \qquad a_{r,\delta,p} := D_{r,\delta}^p - r^p.$$

Then for every $\theta, \eta \in [-1,1]^M$,

$$W_p(P_\theta, Q_\eta)^p = r^p + a_{r,\delta,p}\ \mathrm{TV}(q_\theta, q_\eta) = r^p + \frac{a_{r,\delta,p}}{2M} \sum_{j=1}^{M} |\theta_j - \eta_j|.$$

Proof. Within a fixed block j, the four relevant source-target distances are

$$\|x_j^+ - y_j^+\|_2 = \|x_j^- - y_j^-\|_2 = r,$$

$$\|x_j^+ - y_j^-\|_2 = \|x_j^- - y_j^+\|_2 = D_{r,\delta}.$$

If $i \neq j$, then for any $u \in \{x_i^+, x_i^-\}$ and $v \in \{y_j^+, y_j^-\}$,

$$u - v = -re_1 + (z_i - z_j) + \sigma e_2,$$

for some $\sigma \in \{0, \pm\delta\}$. Because $z_i - z_j$ is orthogonal to e_1, $\|z_i - z_j\|_2 \geq 8\delta$, and $\|\sigma e_2\|_2 \leq \delta$, the orthogonal component has norm at least 7δ. Hence

$$\|u - v\|_2 \geq \sqrt{r^2 + (7\delta)^2} > D_{r,\delta}.$$

Both P_θ and Q_η assign total mass $1/M$ to every block j, so any coupling induces a balanced flow on the block graph and therefore decomposes into directed cycles. Each unit of mass running along a cycle through distinct blocks pays at least $(r^2 + 49\delta^2)^{p/2}$ on every edge. Removing that cycle and replacing it by blockwise couplings costs at most $D^p_{r,\delta}$ per participating block, which is strictly smaller. Eliminating cycles one by one shows that some optimal coupling is fully localized blockwise.

Inside block j, the same-sign mass can be coupled at cost r^p, while the unmatched amount

$$\frac{|\theta_j - \eta_j|}{2M}$$

must be coupled cross-sign at cost $D^p_{r,\delta}$. Thus the optimal blockwise cost equals

$$\frac{r^p}{M} + \frac{a_{r,\delta,p}}{2M}|\theta_j - \eta_j|.$$

Summing over j proves

$$W_p(P_\theta, Q_\eta)^p = r^p + \frac{a_{r,\delta,p}}{2M}\sum_{j=1}^{M}|\theta_j - \eta_j|.$$

Since

$$\mathrm{TV}(q_\theta, q_\eta) = \frac{1}{2M}\sum_{j=1}^{M}|\theta_j - \eta_j|,$$

the displayed identity follows. □

Lemma 16.3 (Quadratic annulus thickness)**.** *Assume* $0 < \delta \leq r$*. Then*

$$a_{r,\delta,p} \asymp_p r^{p-2}\delta^2.$$

Moreover, for every $\theta, \eta \in [-1, 1]^M$*,*

$$r \leq W_p(P_\theta, Q_\eta) \leq r + C_p\frac{\delta^2}{r}.$$

Proof. By the mean value theorem applied to $t \mapsto (r^2 + t)^{p/2}$,

$$a_{r,\delta,p} = \frac{p}{2}(r^2 + \xi)^{p/2-1}\delta^2$$

for some $\xi \in [0, \delta^2]$. Since $\delta \leq r$, one has $r^2 \leq r^2 + \xi \leq 2r^2$, which gives

$$a_{r,\delta,p} \asymp_p r^{p-2}\delta^2.$$

By Theorem 16.2, $W_p(P_\theta, Q_\eta)^p \in [r^p, r^p + a_{r,\delta,p}]$. Hence $W_p(P_\theta, Q_\eta) \geq r$. For the upper bound, the mean value theorem applied to $x \mapsto x^{1/p}$ on $[r^p, r^p + a_{r,\delta,p}]$ gives

$$W_p(P_\theta, Q_\eta) - r \leq \frac{1}{p} r^{1-p} a_{r,\delta,p} \lesssim_p \frac{\delta^2}{r}.$$

□

Theorem 16.4 (Direct estimation on exact translated thin annuli)**.** *Assume*

$$\log N \leq C_0 \log(2M), \qquad 0 < \delta \leq r,$$

where C_0 is the constant from Theorem 4.1. Then there exists an estimator $\widehat{W}$ such that

$$\sup_{\theta,\eta \in [-1,1]^M} \mathbb{E}\big(\widehat{W} - W_p(P_\theta, Q_\eta)\big)^2 \leq C_p\, r^{2-2p} a_{r,\delta,p}^2 \frac{M}{N \log N}.$$

In particular, when $\delta \leq r$,

$$\sup_{\theta,\eta} \mathbb{E}\big(\widehat{W} - W_p(P_\theta, Q_\eta)\big)^2 \lesssim_p \frac{\delta^4}{r^2} \cdot \frac{M}{N \log N}.$$

Proof. Apply Theorem 4.1 on the alphabet $\mathcal{A}_M$ of size $2M$. There exists an estimator $\widehat{V}$ of $\mathrm{TV}(q_\theta, q_\eta)$ such that

$$\sup_{\theta,\eta} \mathbb{E}\big(\widehat{V} - \mathrm{TV}(q_\theta, q_\eta)\big)^2 \leq C \frac{M}{N \log N}.$$

Clip $\widehat{V}$ to $[0,1]$, which can only reduce the risk. Now define

$$\widehat{T} := r^p + a_{r,\delta,p} \widehat{V}, \qquad \widehat{W} := \widehat{T}^{1/p}.$$

By Theorem 16.2,

$$W_p(P_\theta, Q_\eta)^p = r^p + a_{r,\delta,p}\, \mathrm{TV}(q_\theta, q_\eta).$$

Since $\widehat{T}$ and $W_p(P_\theta, Q_\eta)^p$ both lie in $[r^p, r^p + a_{r,\delta,p}]$, the map $x \mapsto x^{1/p}$ is $p^{-1} r^{1-p}$-Lipschitz on that interval. Therefore

$$|\widehat{W} - W_p(P_\theta, Q_\eta)| \leq \frac{1}{p} r^{1-p} a_{r,\delta,p} |\widehat{V} - \mathrm{TV}(q_\theta, q_\eta)|.$$

Squaring and taking expectations yields the first displayed bound. The second follows from Theorem 16.3. □

Proposition 16.5 (The empirical plug-in remains logarithmically worse off the diagonal)**.** *Assume $M \geq cN$ for a fixed constant $c > 0$, and $0 < \delta \leq r$. Let $\theta = \eta = 0$, and write $P_0 := P_\theta$, $Q_0 := Q_\eta$. Then*

$$\mathbb{E}[W_p((P_0)_N, (Q_0)_N) - W_p(P_0, Q_0)] \gtrsim_{c,p} r^{1-p} a_{r,\delta,p},$$

and consequently

$$\mathbb{E}\big(W_p((P_0)_N, (Q_0)_N) - W_p(P_0, Q_0)\big)^2 \gtrsim_{c,p} r^{2-2p} a_{r,\delta,p}^2.$$

When $\delta \leq r$, this is

$$\mathbb{E}\big(W_p((P_0)_N, (Q_0)_N) - W_p(P_0, Q_0)\big)^2 \gtrsim_{c,p} \frac{\delta^4}{r^2}.$$

Proof. The distributions P_0 and Q_0 are both uniform on $2M$ atoms, so $W_p(P_0, Q_0) = r$ by Theorem 16.2.

Let $\widehat{q}$ and $\widehat{r}$ be the empirical histograms of the two samples on the alphabet $\mathcal{A}_M$. Subtract the baseline r^p from every source-target edge cost. The resulting modified cost matrix has cost 0 on same-label edges $(j, \pm) \to (j, \pm)$, and cost at least $a_{r,\delta,p}$ on every other edge. Therefore the optimal modified transport cost is at least $a_{r,\delta,p}\,\mathrm{TV}(\widehat{q}, \widehat{r})$, that is,

$$W_p((P_0)_N, (Q_0)_N)^p - r^p \geq a_{r,\delta,p}\,\mathrm{TV}(\widehat{q}, \widehat{r}).$$

Since $\delta \leq r$, we have $D_{r,\delta} \leq \sqrt{2}\,r$, so $a_{r,\delta,p} \leq (2^{p/2} - 1)r^p$. Hence the map $t \mapsto (r^p + t)^{1/p}$ is bounded below on $[0, a_{r,\delta,p}]$ by a positive multiple of $r^{1-p}t$, namely

$$(r^p + t)^{1/p} - r \geq \frac{1}{p}(\sqrt{2}\,r)^{1-p}t \qquad (0 \leq t \leq a_{r,\delta,p}).$$

Applying this with $t = a_{r,\delta,p}\,\mathrm{TV}(\widehat{q}, \widehat{r})$ yields

$$W_p((P_0)_N, (Q_0)_N) - r \gtrsim_p r^{1-p} a_{r,\delta,p}\,\mathrm{TV}(\widehat{q}, \widehat{r}).$$

Taking expectations and invoking Theorem 5.1 gives

$$\mathbb{E}[W_p((P_0)_N, (Q_0)_N) - r] \gtrsim_{c,p} r^{1-p} a_{r,\delta,p}.$$

The squared lower bound follows from Jensen's inequality. The final display is again Theorem 16.3. □

Corollary 16.6 (A geometric thin-annulus Euclidean class with an off-diagonal logarithmic gap)**.** *Fix $0 < r \leq \frac{1}{10}$. Assume $d \geq 2$, $p \geq 1$, and $n = m = N$. Then for all sufficiently large N there exists an exact Euclidean class*

$$\mathcal{A}^{\mathrm{off}}_{N,r} \subset \mathcal{P}^2_d$$

such that every $(P, Q) \in \mathcal{A}^{\mathrm{off}}_{N,r}$ satisfies

$$r \leq W_p(P, Q) \leq r + C_{d,p} r N^{-2/(d-1)},$$

while

$$\inf_{\widehat{W}} \sup_{(P,Q) \in \mathcal{A}^{\mathrm{off}}_{N,r}} \mathbb{E}\big(\widehat{W} - W_p(P, Q)\big)^2 \lesssim_{d,p} r^2 N^{-4/(d-1)} (\log N)^{-1},$$

and the empirical plug-in estimator obeys

$$\sup_{(P,Q) \in \mathcal{A}^{\mathrm{off}}_{N,r}} \mathbb{E}\big(W_p(P_N, Q_N) - W_p(P, Q)\big)^2 \gtrsim_{d,p} r^2 N^{-4/(d-1)}.$$

Proof. Choose $M \asymp N$ and

$$\delta \asymp_d r M^{-1/(d-1)} \asymp_d r N^{-1/(d-1)}$$

as permitted by Theorem 16.1. Because $M \asymp N$, the condition $\log N \leq C_0 \log(2M)$ holds for all large N. Let $\mathcal{A}^{\mathrm{off}}_{N,r}$ be the resulting translated paired class $\{(P_\theta, Q_\eta) : \theta, \eta \in [-1, 1]^M\}$.

By Theorem 16.3,

$$W_p(P_\theta, Q_\eta) - r \lesssim_p \frac{\delta^2}{r} \asymp_{d,p} rN^{-2/(d-1)},$$

which proves the annulus containment. The upper bound follows from Theorem 16.4:

$$\sup_{(P,Q)\in\mathcal{A}^{\mathrm{off}}_{N,r}} \mathbb{E}(\widehat{W} - W_p(P,Q))^2 \lesssim_p \frac{\delta^4}{r^2} \cdot \frac{M}{N\log N} \asymp_{d,p} r^2 N^{-4/(d-1)}(\log N)^{-1}.$$

The empirical lower bound follows from Theorem 16.5 and the fact that

$$\frac{\delta^4}{r^2} \asymp_{d,p} r^2 N^{-4/(d-1)}.$$

□

Remark 16.7. Theorem 16.6 is qualitatively different from the original diagonal paired class. The functional value now stays in a shrinking annulus around a *fixed positive radius* r, yet direct estimation still beats empirical plug-in by a logarithmic factor. Thus sample-size enlargement for Wasserstein functional estimation is not merely a diagonal phenomenon. It persists in genuinely off-diagonal Euclidean geometry.

17 Fixed finite-template multicluster families at positive baseline

The strongest remaining off-diagonal obstruction after Sections 15 and 16 appears, at first sight, to be a genuinely two-sided multicluster geometry: several shrinking source regions may interact simultaneously with several shrinking target regions. The purpose of this section is to show that *every fixed finite-template positive-baseline family of that kind is already below the global target scale.* This rigorously subsumes the one-sided star geometries considered earlier, but avoids any termwise separation of a single optimization problem, which would be non-rigorous in that setting.

17.1 Template setup and geometry

Fix finite families

$$\Xi = \{\xi_1, \dots, \xi_K\}, \qquad \Upsilon = \{\upsilon_1, \dots, \upsilon_L\} \subset \left(-\frac{1}{4}, \frac{1}{4}\right)^d,$$

and assume

$$\Delta_\Xi := \min_{i\neq i'} \|\xi_i - \xi_{i'}\|_2 > 0, \qquad \Delta_\Upsilon := \min_{j\neq j'} \|\upsilon_j - \upsilon_{j'}\|_2 > 0,$$

and

$$0 < \rho_- := \min_{i,j} \|\xi_i - \upsilon_j\|_2 \le \rho_+ := \max_{i,j} \|\xi_i - \upsilon_j\|_2 < \infty.$$

Let

$$z_\heartsuit := \frac{1}{2}(1, \dots, 1), \qquad \kappa_{\Xi,\Upsilon} := \frac{1}{40}\min\{\Delta_\Xi, \Delta_\Upsilon, \rho_-\}.$$

For $0 < r \le \frac{1}{4}$ and $0 < s \le \kappa_{\Xi,\Upsilon} r$, define the source and target balls

$$X_{i,r,s} := z_\heartsuit + r\xi_i + sB, \qquad Y_{j,r,s} := z_\heartsuit + r\upsilon_j + sB,$$

and the full finite-template class

$$\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon) := \mathcal{P}\Big(\bigcup_{i=1}^{K} X_{i,r,s}\Big) \times \mathcal{P}\Big(\bigcup_{j=1}^{L} Y_{j,r,s}\Big).$$

Lemma 17.1 (Finite-template geometry)**.** *For every $0 < r \le \frac{1}{4}$ and every $0 < s \le \kappa_{\Xi,\Upsilon} r$, the balls $X_{1,r,s}, \dots, X_{K,r,s}$ are pairwise disjoint and the balls $Y_{1,r,s}, \dots, Y_{L,r,s}$ are pairwise disjoint. Moreover, for every $x \in X_{i,r,s}$ and every $y \in Y_{j,r,s}$,*

$$r(\rho_- - 2\kappa_{\Xi,\Upsilon}) \le \|x - y\|_2 \le r(\rho_+ + 2\kappa_{\Xi,\Upsilon}).$$

Consequently, for every $(P, Q) \in \mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)$,

$$c_{\Xi,\Upsilon}\, r \le W_p(P, Q) \le C_{\Xi,\Upsilon}\, r$$

for suitable constants $0 < c_{\Xi,\Upsilon} \le C_{\Xi,\Upsilon} < \infty$ depending only on (Ξ,Υ).

Proof. If $x \in X_{i,r,s}$ and $x' \in X_{i',r,s}$, then

$$\|x - x'\|_2 \ge r\|\xi_i - \xi_{i'}\|_2 - 2s \ge r\Delta_\Xi - 2\kappa_{\Xi,\Upsilon} r > 0,$$

because $2\kappa_{\Xi,\Upsilon} \le \Delta_\Xi/20$. Hence the source balls are pairwise disjoint. The same argument proves pairwise disjointness of the target balls.

Now write

$$x = z_\heartsuit + r\xi_i + su, \qquad y = z_\heartsuit + r\upsilon_j + sv, \qquad u, v \in B.$$

Then

$$x - y = r(\xi_i - \upsilon_j) + s(u - v),$$

so

$$r\|\xi_i - \upsilon_j\|_2 - 2s \le \|x - y\|_2 \le r\|\xi_i - \upsilon_j\|_2 + 2s.$$

Using $\rho_- \le \|\xi_i - \upsilon_j\|_2 \le \rho_+$ and $s \le \kappa_{\Xi,\Upsilon} r$ yields the displayed pointwise bound. Integrating over any coupling and taking the infimum gives the same bounds for $W_p(P, Q)$. □

17.2 Exact relabeling to a bounded smooth cost

For $1 \le i \le K$ and $1 \le j \le L$, define the local affine maps

$$S_{i,r,s}(u) := z_\heartsuit + r\xi_i + su, \qquad T_{j,r,s}(v) := z_\heartsuit + r\upsilon_j + sv, \qquad u, v \in B.$$

Every $P \in \mathcal{P}(\bigcup_i X_{i,r,s})$ admits a unique decomposition

$$P = \sum_{i=1}^{K} (S_{i,r,s})_\# \mu_i,$$

where each μ_i is a finite nonnegative measure on B, and $\sum_i \mu_i(B) = 1$. Likewise every $Q \in \mathcal{P}(\bigcup_j Y_{j,r,s})$ admits a unique decomposition

$$Q = \sum_{j=1}^{L} (T_{j,r,s})_{\#}\nu_j.$$

Set

$$\kappa := \frac{s}{r}, \qquad d_{ij} := \xi_i - \upsilon_j.$$

The exact local cost between the i-th source copy and the j-th target copy is

$$c^{\text{full}}_{\kappa,p,\Xi,\Upsilon}(i,u,j,v) := \|d_{ij} + \kappa(u-v)\|_2^p, \qquad u,v \in B.$$

Proposition 17.2 (Exact smooth relabeling for finite-template classes)**.** *Assume $d \geq 5$ and $p > 1$. There exist translations $a_1, \dots, a_K, b_1, \dots, b_L \in \mathbb{R}^d$, compact convex sets $\widetilde{X}_\Xi, \widetilde{Y}_\Upsilon \subset \mathbb{R}^d$, and for every $0 < \kappa \leq \kappa_{\Xi,\Upsilon}$ a bounded C^2 cost*

$$c^\sharp_{\kappa,p,\Xi,\Upsilon} : \widetilde{X}_\Xi \times \widetilde{Y}_\Upsilon \to \mathbb{R}$$

such that

$$(a_i + B) \cap (a_{i'} + B) = \varnothing \quad (i \neq i'), \qquad (b_j + B) \cap (b_{j'} + B) = \varnothing \quad (j \neq j'),$$

$$\bigcup_{i=1}^{K} (a_i + B) \subset \widetilde{X}_\Xi, \qquad \bigcup_{j=1}^{L} (b_j + B) \subset \widetilde{Y}_\Upsilon,$$

$$c^\sharp_{\kappa,p,\Xi,\Upsilon}(a_i + u, b_j + v) = \|d_{ij} + \kappa(u-v)\|_2^p \qquad (u,v \in B),$$

and

$$\sup_{0<\kappa\leq\kappa_{\Xi,\Upsilon}} \|c^\sharp_{\kappa,p,\Xi,\Upsilon}\|_{C^2(\widetilde{X}_\Xi \times \widetilde{Y}_\Upsilon)} \leq C_{\Xi,\Upsilon,d,p}.$$

If

$$\Gamma_X := \sum_{i=1}^{K} (\sigma_i)_{\#}\mu_i, \qquad \Gamma_Y := \sum_{j=1}^{L} (\tau_j)_{\#}\nu_j,$$

with $\sigma_i(u) := a_i + u$ and $\tau_j(v) := b_j + v$, then

$$W_p(P,Q)^p = r^p\, T_{c^\sharp_{\kappa,p,\Xi,\Upsilon}}(\Gamma_X, \Gamma_Y).$$

Moreover, if $\widehat{\Gamma}_{X,N}$ and $\widehat{\Gamma}_{Y,N}$ are the empirical measures of the relabeled source and target observations, then

$$\sup_{(P,Q)\in\mathcal{C}^{\text{temp}}_{r,s}(\Xi,\Upsilon)} \mathbb{E}\big|T_{c^\sharp_{\kappa,p,\Xi,\Upsilon}}(\widehat{\Gamma}_{X,N}, \widehat{\Gamma}_{Y,N}) - r^{-p}W_p(P,Q)^p\big| \lesssim_{\Xi,\Upsilon,d,p} N^{-2/d}.$$

Proof. Choose the translations $a_1, \dots, a_K, b_1, \dots, b_L$ so far apart that the enlarged balls $a_i + 2B$ are pairwise disjoint and the enlarged balls $b_j + 2B$ are pairwise disjoint. Let $\widetilde{X}_\Xi$ and $\widetilde{Y}_\Upsilon$ be compact convex sets containing the corresponding finite unions. Because the sets $a_i + 2B$ and $b_j + 2B$ are disjoint and finite in number, a standard partition-of-unity construction extends the prescription

$$(a_i + u, b_j + v) \longmapsto \|d_{ij} + \kappa(u-v)\|_2^p$$

from $(\bigcup_i(a_i+B))\times(\bigcup_j(b_j+B))$ to a bounded C^2 function $c^\sharp_{\kappa,p,\Xi,\Upsilon}$ on $\widetilde{X}_\Xi\times\widetilde{Y}_\Upsilon$. The uniform C^2 bound follows because every vector $d_{ij}+\kappa(u-v)$ stays in the fixed annulus

$$\left\{z\in\mathbb{R}^d:\frac{1}{2}\rho_-\le\|z\|_2\le\rho_++1\right\}$$

for $0<\kappa\le\kappa_{\Xi,\Upsilon}$, while $h_p(z)=\|z\|_2^p$ is C^∞ on that annulus.

Now let $\pi\in\Pi(\Gamma_X,\Gamma_Y)$. Because the copies a_i+B and b_j+B are disjoint, π decomposes uniquely as the sum of its restrictions to the rectangles $(a_i+B)\times(b_j+B)$. Pushing these restrictions forward by $(\sigma_i^{-1},\tau_j^{-1})$ produces a coupling between the local pieces μ_i and ν_j, hence between P and Q. Conversely, every coupling between P and Q decomposes into couplings between the rectangles $X_{i,r,s}\times Y_{j,r,s}$, and after the inverse affine changes of variables these pieces become a coupling between Γ_X and Γ_Y. Under the change of variables,

$$S_{i,r,s}(u)-T_{j,r,s}(v)=r(d_{ij}+\kappa(u-v)),$$

so the transport cost scales by r^p. Therefore

$$W_p(P,Q)^p=r^pT_{c^\sharp_{\kappa,p,\Xi,\Upsilon}}(\Gamma_X,\Gamma_Y).$$

Finally, the relabeled observations are i.i.d. with laws Γ_X and Γ_Y. The upper-bound part of [5, Theorem 2], which the authors explicitly note remains valid for general bounded smooth costs on compact classes, yields

$$\mathbb{E}|T_{c^\sharp_{\kappa,p,\Xi,\Upsilon}}(\widehat{\Gamma}_{X,N},\widehat{\Gamma}_{Y,N})-T_{c^\sharp_{\kappa,p,\Xi,\Upsilon}}(\Gamma_X,\Gamma_Y)|\lesssim_{\Xi,\Upsilon,d,p}N^{-2/d}.$$

Multiplying by r^p proves the claim. □

Theorem 17.3 (Fixed finite-template families are already below the global target scale)**.** *Assume $d\ge5$ and $p>1$, and fix any finite template pair (Ξ,Υ) satisfying the assumptions above. Then there exists a constant $C_{\Xi,\Upsilon,d,p}<\infty$ such that for every $0<r\le\frac{1}{4}$ and every $0<s\le\kappa_{\Xi,\Upsilon}r$,*

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)}\mathbb{E}|W_p(P_N,Q_N)^p-W_p(P,Q)^p|\le C_{\Xi,\Upsilon,d,p}\,r^pN^{-2/d},$$

and

$$\sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)}\mathbb{E}|W_p(P_N,Q_N)-W_p(P,Q)|\le C_{\Xi,\Upsilon,d,p}\,rN^{-2/d}.$$

In particular,

$$\sup_{0<r\le\frac{1}{4}}\ \sup_{0<s\le\kappa_{\Xi,\Upsilon}r}\ \inf_{\widehat{W}}\ \sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)}\mathbb{E}|\widehat{W}-W_p(P,Q)|\lesssim_{\Xi,\Upsilon,d,p}\eta_N,\qquad\eta_N:=(N\log N)^{-1/d}.$$

Proof. The cost bound follows immediately from Theorem 17.2. By Theorem 17.1, every value $W_p(P,Q)$ on the class satisfies $W_p(P,Q)\ge c_{\Xi,\Upsilon}r$. Therefore Theorem 12.2, applied with $x=W_p(P_N,Q_N)$, $y=W_p(P,Q)$, and r replaced by $c_{\Xi,\Upsilon}r$, yields

$$|W_p(P_N,Q_N)-W_p(P,Q)|\lesssim_{\Xi,\Upsilon,p}r^{1-p}|W_p(P_N,Q_N)^p-W_p(P,Q)^p|.$$

Taking expectations and using the first display proves the distance bound.

Since $d \geq 5$, one has $N^{-2/d} \leq (N \log N)^{-1/d} = \eta_N$ for all $N \geq 2$, so the final display follows. $\square$

17.3 Every finite-template class still contains a hard two-ball core

Although the upper theorem is completely general, the lower theory depends on how thin the template balls are relative to their baseline radius. The key point is that every finite-template class contains a one-pair two-ball subclass.

Corollary 17.4 (Universal lower bound inside every finite-template class)**.** *Assume $d > 2p$ and fix a template pair (Ξ, Υ) as above. Then there exists $c_{\Xi,\Upsilon,d,p} > 0$ such that for every $0 < r \leq \frac{1}{4}$ and every $0 < s \leq \kappa_{\Xi,\Upsilon} r$,*

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \geq c_{\Xi,\Upsilon,d,p} \max\left\{ sN^{-1/2},\ \frac{s^2}{r}\eta_N^2 \right\}.$$

Likewise,

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)} \mathbb{E}\big|\widehat{T} - W_p(P,Q)^p\big| \geq c_{\Xi,\Upsilon,d,p} \max\left\{ r^{p-1} sN^{-1/2},\ r^{p-2}s^2\eta_N^2 \right\}.$$

Proof. Fix any pair (i_0, j_0). Restrict attention to the subclass of $\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)$ defined by

$$P \in \mathcal{P}(X_{i_0,r,s}), \qquad Q \in \mathcal{P}(Y_{j_0,r,s}).$$

After translation and rigid rotation, this is exactly the two-ball class $\mathcal{C}_{\rho_{i_0 j_0} r,s}$ from Section 15, where

$$\rho_{i_0 j_0} := \|\xi_{i_0} - \upsilon_{j_0}\|_2 \in [\rho_-, \rho_+].$$

Apply Theorem 15.9. Because $\rho_{i_0 j_0}$ is bounded above and below by constants depending only on (Ξ, Υ), the resulting lower bound has exactly the displayed form after absorbing $\rho_{i_0 j_0}$ into the constants. $\square$

Corollary 17.5 (Thick finite-template classes are solved up to the same logarithm)**.** *Assume $d > 2p$, $p > 1$, and fix (Ξ, Υ) as above. For every $c_0 \in (0, \kappa_{\Xi,\Upsilon}]$, there exist constants $0 < c'_{\Xi,\Upsilon,d,p,c_0} \leq C'_{\Xi,\Upsilon,d,p,c_0} < \infty$ such that whenever*

$$c_0 r \leq s \leq \kappa_{\Xi,\Upsilon} r,$$

one has

$$c'_{\Xi,\Upsilon,d,p,c_0}\, r^p \eta_N^2 \leq \inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)} \mathbb{E}\big|\widehat{T} - W_p(P,Q)^p\big| \leq C'_{\Xi,\Upsilon,d,p,c_0}\, r^p N^{-2/d},$$

and

$$c'_{\Xi,\Upsilon,d,p,c_0}\, r \eta_N^2 \leq \inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi,\Upsilon)} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \leq C'_{\Xi,\Upsilon,d,p,c_0}\, r N^{-2/d}.$$

Proof. If $s \geq c_0 r$, then

$$\frac{s^2}{r}\eta_N^2 \geq c_0^2\, r\eta_N^2, \qquad r^{p-2}s^2\eta_N^2 \geq c_0^2\, r^p \eta_N^2.$$

Combine Theorems 17.3 and 17.4. $\square$

Remark 17.6 (One-sided stars are a special case, but not the whole story). The one-sided multibaseline star geometries considered earlier correspond to the special choice $K = 1$, $\Xi = \{0\}$, and $\Upsilon = \{v_1, \dots, v_L\} \subset \mathbb{S}^{d-1}$ with positive mutual separation. Theorem 17.3 therefore rigorously implies that every such star family is already estimable at or below the global target scale. More importantly, the same theorem now covers *all fixed finite two-sided* positive-baseline template families as well. The real remaining difficulty is thus not merely "multibaseline" or even merely "two-sided". It must involve either a growing number of active macro-regions, collapsing source-target baselines on more than one macro-edge, or both.

18 Permutation-stable shrinking matched-template neighborhoods

The finite-template theorem of Section 17 shows that every fixed positive-baseline multicluster family is already below the global target scale. What it does *not* yet resolve is whether a genuinely shrinking two-sided multicluster geometry can remain hard when several source-target pairs approach the diagonal simultaneously but the macro-assignment itself is rigid. The purpose of this section is to show that this regime is also solved theorem-level.

18.1 Matched template setup

Fix $K \geq 1$, distinct template points

$$\Xi = \{\xi_1, \dots, \xi_K\}, \qquad \Upsilon = \{\upsilon_1, \dots, \upsilon_K\} \subset \left(-\frac{1}{4}, \frac{1}{4}\right)^d,$$

and a weight vector $\alpha = (\alpha_1, \dots, \alpha_K)$ with

$$\alpha_i > 0, \qquad \sum_{i=1}^{K} \alpha_i = 1.$$

Assume

$$\rho_i := \|\xi_i - \upsilon_i\|_2, \qquad 0 < \rho_- := \min_i \rho_i \leq \rho_+ := \max_i \rho_i < \infty,$$

$$\Delta_\Xi := \min_{i \neq i'} \|\xi_i - \xi_{i'}\|_2 > 0, \qquad \Delta_\Upsilon := \min_{j \neq j'} \|\upsilon_j - \upsilon_{j'}\|_2 > 0,$$

and, crucially, the *row-wise matched gap*

$$\Delta_{\text{match}} := \min_{i \neq j} \Big(\|\xi_i - \upsilon_j\|_2^p - \|\xi_i - \upsilon_i\|_2^p \Big) > 0.$$

Set

$$z_\clubsuit := \frac{1}{2}(1, \dots, 1), \qquad L_{\Xi,\Upsilon,p} := p(\rho_+ + 1)^{p-1},$$

and define

$$\kappa_{\text{match}} := \frac{1}{40} \min \left\{ \Delta_\Xi,\ \Delta_\Upsilon,\ \rho_-,\ \frac{\Delta_{\text{match}}}{L_{\Xi,\Upsilon,p}} \right\}.$$

For $0 < r \leq \frac{1}{4}$ and $0 < s \leq \kappa_{\text{match}} r$, define

$$X_{i,r,s} := z_\clubsuit + r\xi_i + sB, \qquad Y_{i,r,s} := z_\clubsuit + r\upsilon_i + sB,$$

and the shrinking matched-template class

$$\mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha) := \left\{ \Big(\sum_{i=1}^{K} \alpha_i P_i,\ \sum_{i=1}^{K} \alpha_i Q_i\Big) : P_i \in \mathcal{P}(X_{i,r,s}),\ Q_i \in \mathcal{P}(Y_{i,r,s}) \right\}.$$

Lemma 18.1 (Matched geometry and row-wise pointwise gap)**.** *For every $0 < r \le \frac{1}{4}$ and every $0 < s \le \kappa_{\mathrm{match}} r$, the source balls $X_{1,r,s}, \dots, X_{K,r,s}$ are pairwise disjoint, the target balls $Y_{1,r,s}, \dots, Y_{K,r,s}$ are pairwise disjoint, and*

$$c_{\Xi,\Upsilon}\, r \le W_p(P,Q) \le C_{\Xi,\Upsilon}\, r$$

for all $(P,Q) \in \mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha)$, with constants $0 < c_{\Xi,\Upsilon} \le C_{\Xi,\Upsilon} < \infty$ depending only on (Ξ,Υ).

Moreover, for every $i \ne j$,

$$\sup_{x \in X_{i,r,s},\, y \in Y_{i,r,s}} \|x-y\|_2^p < \inf_{x \in X_{i,r,s},\, y \in Y_{j,r,s}} \|x-y\|_2^p.$$

Proof. Pairwise disjointness of the source balls and target balls is identical to the argument in Theorem 17.1, because $2\kappa_{\mathrm{match}} \le \Delta_\Xi/20$ and $2\kappa_{\mathrm{match}} \le \Delta_\Upsilon/20$.

Now fix $i \ne j$, and write

$$x = z_\clubsuit + r\xi_i + su, \qquad y = z_\clubsuit + r\upsilon_i + sv, \qquad y' = z_\clubsuit + r\upsilon_j + sv', \qquad u, v, v' \in B.$$

Then

$$x - y = r(\xi_i - \upsilon_i) + s(u - v), \qquad x - y' = r(\xi_i - \upsilon_j) + s(u - v').$$

Since $s/r \le \kappa_{\mathrm{match}} \le \rho_-/40$, all vectors $\xi_i - \upsilon_i + (s/r)(u-v)$ and $\xi_i - \upsilon_j + (s/r)(u-v')$ stay in the ball of radius $\rho_+ + 1$, where $z \mapsto \|z\|_2^p$ is $L_{\Xi,\Upsilon,p}$-Lipschitz. Hence

$$\|x-y\|_2^p \le r^p \|\xi_i - \upsilon_i\|_2^p + L_{\Xi,\Upsilon,p}\, r^{p-1} s$$

and

$$\|x-y'\|_2^p \ge r^p \|\xi_i - \upsilon_j\|_2^p - L_{\Xi,\Upsilon,p}\, r^{p-1} s.$$

Subtracting yields

$$\|x-y'\|_2^p - \|x-y\|_2^p \ge r^p \Delta_{\mathrm{match}} - 2L_{\Xi,\Upsilon,p} r^{p-1} s \ge \frac{1}{2} r^p \Delta_{\mathrm{match}} > 0,$$

because $s \le \kappa_{\mathrm{match}} r$ and $2L_{\Xi,\Upsilon,p}\kappa_{\mathrm{match}} \le \Delta_{\mathrm{match}}/20$. This proves the pointwise row-wise gap.

Finally, the lower and upper bounds on $W_p(P,Q)$ follow exactly as in Theorem 17.1, using the facts that every matched distance lies in $[\rho_-, \rho_+]$ and $s \le \kappa_{\mathrm{match}} r \le \rho_- r/40$. □

18.2 Exact localization to the matched pairs

Theorem 18.2 (Exact matched-pair localization)**.** *For every $(P,Q) \in \mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha)$, with*

$$P = \sum_{i=1}^{K} \alpha_i P_i, \qquad Q = \sum_{i=1}^{K} \alpha_i Q_i, \qquad P_i \in \mathcal{P}(X_{i,r,s}), \qquad Q_i \in \mathcal{P}(Y_{i,r,s}),$$

one has the exact identity

$$W_p(P,Q)^p = \sum_{i=1}^{K} \alpha_i W_p(P_i,Q_i)^p.$$

Equivalently, some optimal coupling transports all mass from $X_{i,r,s}$ to $Y_{i,r,s}$ for every i, and no optimal coupling uses an off-diagonal macro-edge.

Proof. Let π be any coupling of P and Q. For $1 \le i, j \le K$, let

$$\pi_{ij} := \pi(X_{i,r,s} \times Y_{j,r,s}).$$

Because P places mass α_i on $X_{i,r,s}$ and Q places mass α_i on $Y_{i,r,s}$, the matrix (π_{ij}) has row sums and column sums equal to $(\alpha_1, \dots, \alpha_K)$. Hence its off-diagonal part is a circulation and therefore decomposes into directed cycles.

Consider one such cycle

$$i_1 \to i_2 \to \cdots \to i_\ell \to i_1$$

carrying mass $\varepsilon > 0$. Every unit of mass sent along the edge $i_a \to i_{a+1}$ pays at least

$$m_{i_a i_{a+1}} := \inf_{x \in X_{i_a,r,s},\, y \in Y_{i_{a+1},r,s}} \|x-y\|_2^p,$$

whereas every unit of mass sent along the diagonal edge $i_a \to i_a$ pays at most

$$M_{i_a i_a} := \sup_{x \in X_{i_a,r,s},\, y \in Y_{i_a,r,s}} \|x-y\|_2^p.$$

By Theorem 18.1,

$$m_{i_a i_{a+1}} > M_{i_a i_a} \qquad \text{for every } a.$$

Therefore the whole cycle costs at least

$$\varepsilon \sum_{a=1}^{\ell} m_{i_a i_{a+1}} > \varepsilon \sum_{a=1}^{\ell} M_{i_a i_a}.$$

But replacing the cycle by ε units of mass on each diagonal block $X_{i_a,r,s} \times Y_{i_a,r,s}$ preserves all macro-marginals and strictly lowers the cost. Eliminating cycles one by one shows that some optimal coupling has no off-diagonal macro-flow:

$$\pi(X_{i,r,s} \times Y_{j,r,s}) = 0 \qquad (i \neq j).$$

Once localization is known, the coupling decomposes into K independent within-pair couplings. Since the transport cost scales linearly with the total mass,

$$\inf_{\pi_i \in \Pi(\alpha_i P_i, \alpha_i Q_i)} \int \|x-y\|_2^p \, d\pi_i(x,y) = \alpha_i W_p(P_i,Q_i)^p.$$

Summing over i gives the claimed identity. □

18.3 A full matched-template phase diagram

Each localized pair $X_{i,r,s} \times Y_{i,r,s}$ is, after translation and rigid rotation, exactly the two-ball product neighborhood $\mathcal{C}_{\rho_i r,s}$ from Section 15. The only new issue is that the empirical sample sizes inside the i-th pair are random.

Lemma 18.3 (Inverse moments for binomial block counts)**.** *Fix $\alpha_0 \in (0,1)$ and $q > 0$. If $M \sim \mathrm{Bin}(N, \alpha)$ with $\alpha \in [\alpha_0, 1]$, then*

$$\mathbb{P}\Big(M \le \frac{\alpha_0 N}{2}\Big) \le e^{-c_{\alpha_0} N}$$

for some $c_{\alpha_0} > 0$, and

$$\mathbb{E}\big[M^{-q}\mathbf{1}_{\{M\ge 1\}}\big] \lesssim_{\alpha_0,q} N^{-q}.$$

Proof. The large-deviation bound is the standard Chernoff estimate. For the inverse moment, split the expectation according to the events $\{M \ge \alpha_0 N/2\}$ and $\{1 \le M < \alpha_0 N/2\}$. On the first event,

$$M^{-q} \le \Big(\frac{\alpha_0 N}{2}\Big)^{-q}.$$

On the second event,

$$M^{-q}\mathbf{1}_{\{1\le M<\alpha_0 N/2\}} \le \mathbb{P}\Big(M < \frac{\alpha_0 N}{2}\Big),$$

which is exponentially small. This yields the claimed bound. □

Theorem 18.4 (Upper bound on the full matched-template class)**.** *Assume $d \ge 5$, $p > 1$, and $n = m = N$. Then there exists an estimator $\widehat{T}^{\mathrm{match}}_{r,s}$ such that*

$$\sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}\big|\widehat{T}^{\mathrm{match}}_{r,s} - W_p(P,Q)^p\big| \lesssim_{\Xi,\Upsilon,\alpha,d,p} r^{p-1} s\, N^{-1/2} + r^{p-2}s^2 N^{-2/d}.$$

Consequently, there exists an estimator $\widehat{W}^{\mathrm{match}}_{r,s}$ such that

$$\sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}\big|\widehat{W}^{\mathrm{match}}_{r,s} - W_p(P,Q)\big| \lesssim_{\Xi,\Upsilon,\alpha,d,p} s\, N^{-1/2} + \frac{s^2}{r} N^{-2/d}.$$

Proof. For a sample $X_1, \dots, X_N \sim P$, let N_i^X be the number of observations falling in $X_{i,r,s}$. Since the source balls are pairwise disjoint, the counts N_i^X and the corresponding within-ball samples are observable. Define N_i^Y analogously for the target sample, and set

$$M_i := N_i^X \wedge N_i^Y.$$

Because the i-th source and target weights are both equal to α_i, one has $N_i^X, N_i^Y \overset{d}{=} \mathrm{Bin}(N, \alpha_i)$, independently across the two samples.

If $M_i \ge 1$, discard all but M_i observations from each side of the i-th pair and, after translating and rotating to the canonical geometry, apply the two-ball estimator from Theorem 15.8 with baseline radius $\rho_i r$, thickness s, and sample size M_i. Call the resulting estimator $\widehat{T}_i$. If $M_i = 0$, set $\widehat{T}_i := (\rho_i r)^p$.

Now define

$$\widehat{T}^{\mathrm{match}}_{r,s} := \sum_{i=1}^{K} \alpha_i \widehat{T}_i.$$

By Theorem 18.2,

$$W_p(P,Q)^p = \sum_{i=1}^{K} \alpha_i W_p(P_i,Q_i)^p.$$

Hence, conditioning on the counts and using Theorem 15.8,

$$\begin{aligned}
\mathbb{E}|\widehat{T}^{\mathrm{match}}_{r,s} - W_p(P,Q)^p| &\le \sum_{i=1}^{K} \alpha_i\, \mathbb{E}|\widehat{T}_i - W_p(P_i,Q_i)^p| \\
&\lesssim_{\Xi,\Upsilon,\alpha,d,p} \sum_{i=1}^{K} \alpha_i (\rho_i r)^{p-1} s\, \mathbb{E}(M_i^{-1/2}\mathbf{1}_{\{M_i\ge 1\}}) \\
&\quad + \sum_{i=1}^{K} \alpha_i (\rho_i r)^{p-2} s^2\, \mathbb{E}(M_i^{-2/d}\mathbf{1}_{\{M_i\ge 1\}}) \\
&\quad + C_{\Xi,\Upsilon,p} r^p \sum_{i=1}^{K} \alpha_i \mathbb{P}(M_i = 0).
\end{aligned}$$

Because $\rho_i \in [\rho_-,\rho_+]$ and $\alpha_i \ge \alpha_- := \min_i \alpha_i > 0$, Theorem 18.3 yields

$$\mathbb{E}(M_i^{-1/2}\mathbf{1}_{\{M_i\ge 1\}}) \lesssim_{\alpha_-} N^{-1/2}, \qquad \mathbb{E}(M_i^{-2/d}\mathbf{1}_{\{M_i\ge 1\}}) \lesssim_{\alpha_-} N^{-2/d},$$

and $\mathbb{P}(M_i = 0) \le e^{-c_{\alpha_-}N}$. Summing over i proves

$$\sup_{(P,Q)} \mathbb{E}|\widehat{T}^{\mathrm{match}}_{r,s} - W_p(P,Q)^p| \lesssim_{\Xi,\Upsilon,\alpha,d,p} r^{p-1} s\, N^{-1/2} + r^{p-2}s^2 N^{-2/d}.$$

Finally, by Theorem 18.1, the true value $W_p(P,Q)^p$ lies in an interval $[c^p_{\Xi,\Upsilon} r^p, C^p_{\Xi,\Upsilon} r^p]$ on which $x \mapsto x^{1/p}$ is Lipschitz with constant $\lesssim_{\Xi,\Upsilon,p} r^{1-p}$. Clip $\widehat{T}^{\mathrm{match}}_{r,s}$ to that interval and take the p-th root. This gives the distance estimator and the displayed bound. □

Theorem 18.5 (Lower bound on the full matched-template class)**.** *Assume $d > 2p$ and $n = m = N$. Then*

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}|\widehat{T} - W_p(P,Q)^p| \gtrsim_{\Xi,\Upsilon,\alpha,d,p} \max\left\{r^{p-1} s\, N^{-1/2},\ r^{p-2}s^2 (N\log N)^{-2/d}\right\},$$

and

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}|\widehat{W} - W_p(P,Q)| \gtrsim_{\Xi,\Upsilon,\alpha,d,p} \max\left\{s\, N^{-1/2},\ \frac{s^2}{r}(N\log N)^{-2/d}\right\}.$$

Proof. Fix one index $i_0 \in \{1,\dots,K\}$. Restrict attention to the subclass on which

$$P_i = \delta_{z_\clubsuit + r\xi_i}, \qquad Q_i = \delta_{z_\clubsuit + r\upsilon_i} \qquad (i \ne i_0),$$

while (P_{i_0}, Q_{i_0}) varies arbitrarily in $\mathcal{P}(X_{i_0,r,s}) \times \mathcal{P}(Y_{i_0,r,s})$. After translation and rigid rotation, this active block is exactly the two-ball class $\mathcal{C}_{\rho_{i_0} r,s}$ from Section 15.

By Theorem 18.2, on this subclass

$$W_p(P,Q)^p = C_0 + \alpha_{i_0} W_p(P_{i_0},Q_{i_0})^p,$$

where

$$C_0 := \sum_{i \neq i_0} \alpha_i (\rho_i r)^p \asymp_{\Xi,\Upsilon,\alpha} r^p.$$

Therefore the active block enters through the smooth affine map

$$t \longmapsto C_0 + \alpha_{i_0} t$$

for the p-th power, and through the smooth bi-Lipschitz map

$$t \longmapsto (C_0 + \alpha_{i_0} t)^{1/p}$$

for the distance itself.

Reveal to the estimator all samples outside the active block, and reveal the block labels of all observations. This can only make estimation easier. The active sample counts are then $M_X, M_Y \overset{d}{=} \mathrm{Bin}(N, \alpha_{i_0})$, independently. On the event

$$E := \{M_X \leq 2\alpha_{i_0} N,\ M_Y \leq 2\alpha_{i_0} N\},$$

which has probability bounded below by a positive constant depending only on α_{i_0}, one may further reveal the values of M_X, M_Y and discard the excess samples to obtain a balanced active experiment with at most $2\alpha_{i_0} N$ samples from each side. By monotonicity of minimax risk under sample size enlargement, Theorem 15.9 yields

$$\inf_{\widehat{T}} \sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}} \mathbb{E}\big|\widehat{T} - W_p(P,Q)^p\big| \gtrsim_{\Xi,\Upsilon,\alpha,d,p} (\rho_{i_0} r)^{p-1} s\, N^{-1/2} \vee (\rho_{i_0} r)^{p-2} s^2 (N \log N)^{-2/d}.$$

and similarly

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| \gtrsim_{\Xi,\Upsilon,\alpha,d,p} \max\left\{ s\, N^{-1/2},\ \frac{s^2}{\rho_{i_0} r} (N \log N)^{-2/d} \right\}.$$

Because $\rho_{i_0} \in [\rho_-, \rho_+]$, the constants are uniform in the template. □

Corollary 18.6 (Matched-template phase diagram)**.** *Assume $d > 2p$ and $n = m = N$. Then*

$$\max\left\{ s\, N^{-1/2},\ \frac{s^2}{r} (N \log N)^{-2/d} \right\} \lesssim_{\Xi,\Upsilon,\alpha,d,p} \inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big|$$

$$\lesssim_{\Xi,\Upsilon,\alpha,d,p} s\, N^{-1/2} + \frac{s^2}{r} N^{-2/d}.$$

In particular, the critical width is still

$$s_{\mathrm{crit}}(N, r) \asymp r\, N^{2/d - 1/2}.$$

Below that width, the whole matched-template class is generically parametric; above it, the whole class is generically nonparametric up to the same logarithm as in the one-pair theorem.

Proof. Combine Theorems 18.4 and 18.5. The final sentence follows exactly as in Theorems 15.11 and 15.12. □

Discussion 18.7 (What the matched-template theorem rules out). The importance of Theorem 18.6 is structural, not merely incremental. It shows that even a *two-sided* shrinking multicluster geometry remains statistically tame as long as the macro-assignment is rigid and the number of active matched pairs stays fixed. In other words, simultaneous baseline collapse along finitely many *nonbranching* macro-edges is not enough to sustain the unrestricted lower-envelope scale.

Combined with the diagonal theorem, the two-ball phase diagram, the translated thin-annulus counterexamples, and the positive-distance finite-template theorem, this leaves a much narrower candidate core for the unrestricted problem: the remaining obstruction must involve either genuinely *growing* macro-complexity, or a *branching* two-sided collapse in which the same shrinking source region can interact competitively with more than one shrinking target region (and vice versa).

19 Growing matched-template complexity and an entropy barrier

The fixed-K matched-template theorem leaves open the possibility that the first genuinely hard regime might already arise when the number of rigid matched pairs grows with N. The goal of this section is to quantify that possibility. We keep the matched geometry of Section 18, but we no longer insist that the constants be uniform over the weight profile. Instead we track the complexity contribution of the weights explicitly.

Lemma 19.1 (Explicit inverse moments for sparse matched block counts)**.** *Fix $q > 0$. Let B, B' be independent* $\mathrm{Bin}(N, \alpha)$ *random variables with $\alpha \in (0, 1]$, and set $M := B \wedge B'$. Then*

$$\mathbb{E}[M^{-q}\mathbf{1}_{\{M\geq 1\}}] \lesssim_q (N\alpha)^{-q}, \qquad \mathbb{P}(M = 0) \lesssim e^{-N\alpha/8}.$$

In particular,

$$\alpha\,\mathbb{E}[M^{-1/2}\mathbf{1}_{\{M\geq 1\}}] \lesssim N^{-1/2}\alpha^{1/2}, \qquad \alpha\,\mathbb{E}[M^{-2/d}\mathbf{1}_{\{M\geq 1\}}] \lesssim_d N^{-2/d}\alpha^{1-2/d},$$

and

$$\alpha\,\mathbb{P}(M = 0) \lesssim N^{-1/2}\alpha^{1/2}, \qquad \alpha\,\mathbb{P}(M = 0) \lesssim_d N^{-2/d}\alpha^{1-2/d}.$$

Proof. By independence,

$$\mathbb{P}(M = 0) = 1 - \mathbb{P}(B \geq 1)^2 \leq 2\mathbb{P}(B = 0) \leq 2e^{-N\alpha}.$$

For the inverse moment, put $t := N\alpha$. On the event $\{B \geq t/2,\ B' \geq t/2\}$, one has $M \geq t/2$, hence

$$M^{-q}\mathbf{1}_{\{M\geq 1\}} \leq (2/t)^q.$$

Therefore

$$\mathbb{E}[M^{-q}\mathbf{1}_{\{M\geq 1\}}] \leq (2/t)^q + \mathbb{P}(B < t/2) + \mathbb{P}(B' < t/2).$$

The Chernoff bound yields

$$\mathbb{P}(B < t/2) \leq e^{-t/8}, \qquad \mathbb{P}(B' < t/2) \leq e^{-t/8}.$$

Since $e^{-u/8} \lesssim_q u^{-q}$ for all $u > 0$,

$$\mathbb{E}\big[M^{-q}\mathbf{1}_{\{M\geq 1\}}\big] \lesssim_q t^{-q} = (N\alpha)^{-q}.$$

Multiplying by α gives the displayed weighted bounds. Finally, $e^{-u/8} \lesssim u^{-1/2}$ and $e^{-u/8} \lesssim_d u^{-2/d}$ imply

$$\alpha\,\mathbb{P}(M=0) \lesssim \alpha(N\alpha)^{-1/2} = N^{-1/2}\alpha^{1/2},$$

and

$$\alpha\,\mathbb{P}(M=0) \lesssim_d \alpha(N\alpha)^{-2/d} = N^{-2/d}\alpha^{1-2/d}. \qquad \square$$

Theorem 19.2 (Complexity-sensitive upper bound for matched-template families)**.** *Assume $d \geq 5$, $p > 1$, $n = m = N$, and let*

$$(P,Q) = \Big(\sum_{i=1}^{K}\alpha_i P_i,\ \sum_{i=1}^{K}\alpha_i Q_i\Big) \in \mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha).$$

Then there exists an estimator $\widehat{T}_{r,s}^{\mathrm{cmp}}$ such that

$$\sup_{(P,Q)\in\mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha)} \mathbb{E}\big|\widehat{T}_{r,s}^{\mathrm{cmp}} - W_p(P,Q)^p\big| \lesssim_{\Xi,\Upsilon,d,p} r^{p-1}s\,N^{-1/2}\sum_{i=1}^{K}\alpha_i^{1/2} + r^{p-2}s^2N^{-2/d}\sum_{i=1}^{K}\alpha_i^{1-2/d}.$$

Consequently, there exists an estimator $\widehat{W}_{r,s}^{\mathrm{cmp}}$ such that

$$\sup_{(P,Q)\in\mathcal{M}_{r,s}^{\mathrm{match}}(\Xi,\Upsilon,\alpha)} \mathbb{E}\big|\widehat{W}_{r,s}^{\mathrm{cmp}} - W_p(P,Q)\big| \lesssim_{\Xi,\Upsilon,d,p} s\,N^{-1/2}\sum_{i=1}^{K}\alpha_i^{1/2} + \frac{s^2}{r}N^{-2/d}\sum_{i=1}^{K}\alpha_i^{1-2/d}.$$

Proof. Repeat the construction from the proof of Theorem 18.4. For the i-th matched pair, let N_i^X, N_i^Y be the observable block counts in $X_{i,r,s}$ and $Y_{i,r,s}$, let $M_i := N_i^X \wedge N_i^Y$, and define the two-ball estimator $\widehat{T}_i$ exactly as before. Set

$$\widehat{T}_{r,s}^{\mathrm{cmp}} := \sum_{i=1}^{K}\alpha_i\widehat{T}_i.$$

By Theorem 18.2,

$$W_p(P,Q)^p = \sum_{i=1}^{K}\alpha_i W_p(P_i,Q_i)^p.$$

For the i-th matched pair, write $\rho_i = \|\xi_i - \upsilon_i\|_2$. Since (P_i,Q_i), after translation and rotation, lies in the two-ball class $\mathcal{C}_{\rho_i r,s}$, Theorem 15.2 and the boundedness of the rescaled class imply the deterministic baseline approximation

$$\big|W_p(P_i,Q_i)^p - (\rho_i r)^p\big| \lesssim_{\Xi,\Upsilon,d,p} r^{p-1}s + r^{p-2}s^2.$$

Therefore, conditioning on the counts and using Theorem 15.8, we obtain

$$\mathbb{E}\big|\widehat{T}_{r,s}^{\mathrm{cmp}} - W_p(P,Q)^p\big| \leq \sum_{i=1}^{K}\alpha_i\,\mathbb{E}\big|\widehat{T}_i - W_p(P_i,Q_i)^p\big|$$

$$\lesssim_{\Xi,\Upsilon,d,p} \sum_{i=1}^{K} \alpha_i \Big[r^{p-1} s\, \mathbb{E}(M_i^{-1/2} \mathbf{1}_{\{M_i \geq 1\}}) + r^{p-2} s^2\, \mathbb{E}(M_i^{-2/d} \mathbf{1}_{\{M_i \geq 1\}}) \Big]$$
$$+ C_{\Xi,\Upsilon,d,p} (r^{p-1} s + r^{p-2} s^2) \sum_{i=1}^{K} \alpha_i \mathbb{P}(M_i = 0).$$

Applying Theorem 19.1 to each M_i proves

$$\mathbb{E}|\widehat{T}_{r,s}^{\mathrm{cmp}} - W_p(P,Q)^p| \lesssim_{\Xi,\Upsilon,d,p} r^{p-1} s\, N^{-1/2} \sum_{i=1}^{K} \alpha_i^{1/2} + r^{p-2} s^2 N^{-2/d} \sum_{i=1}^{K} \alpha_i^{1-2/d},$$

where the zero-count contribution is split using the two probability bounds from Theorem 19.1. Finally, as in the proof of Theorem 18.4, Theorem 18.1 places $W_p(P,Q)^p$ in a deterministic interval $[c^p_{\Xi,\Upsilon} r^p, C^p_{\Xi,\Upsilon} r^p]$, on which $x \mapsto x^{1/p}$ is $O_{\Xi,\Upsilon,p}(r^{1-p})$-Lipschitz. Clipping and taking the p-th root yields the displayed distance bound. □

Corollary 19.3 (Equal-weight growing matched families)**.** *Assume the setup of Theorem 19.2 and, additionally, $\alpha_i = 1/K$ for all i. Then*

$$\sup_{(P,Q)\in\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi,\Upsilon,\alpha)} \mathbb{E}|\widehat{W}_{r,s}^{\mathrm{cmp}} - W_p(P,Q)| \lesssim_{\Xi,\Upsilon,d,p} s\sqrt{\frac{K}{N}} + \frac{s^2}{r}\Big(\frac{K}{N}\Big)^{2/d}.$$

In particular, a growing equal-weight matched-template family is already below the global target scale whenever

$$s\sqrt{\frac{K}{N}} + \frac{s^2}{r}\Big(\frac{K}{N}\Big)^{2/d} \lesssim_{\Xi,\Upsilon,d,p} (N \log N)^{-1/d}.$$

Proof. If $\alpha_i = 1/K$, then

$$\sum_{i=1}^{K} \alpha_i^{1/2} = K^{1/2}, \qquad \sum_{i=1}^{K} \alpha_i^{1-2/d} = K^{2/d}.$$

Substitute these identities into Theorem 19.2. The final sentence is immediate. □

20 What remains open after the finite-template theorem?

The cumulative picture is now significantly sharper than in the diagonal-only version.

Discussion 20.1 (What is genuinely under theorem-level control)*.* Fix $d > 2p$ and $n = m = N$.

(a) On the full diagonal neighborhood $\mathcal{L}_{A_0,N}$, the exact minimax absolute and squared rates are $(N \log N)^{-1/d}$ and $(N \log N)^{-2/d}$ (Theorem 9.4).

(b) On the canonical dyadic-tree surrogate, the exact supercritical minimax law is already known, and empirical plug-in is strictly slower there (Theorems 8.3 and 8.4).

(c) For $p > 1$ and $d \geq 5$, every full Euclidean product neighborhood $\mathcal{C}_r = \mathcal{P}(X_r) \times \mathcal{P}(Y_r)$ at a fixed positive radius r is solved up to the same logarithm as the smooth-cost minimax lower theory:

$$r(N \log N)^{-2/d} \lesssim_{d,p} \inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{C}_r} \mathbb{E}|\widehat{W} - W_p(P,Q)| \lesssim_{d,p} r N^{-2/d}$$

by Theorem 14.4.

(d) For every one-baseline two-ball class

$$\mathcal{C}_{r,s} = \mathcal{P}(z_\diamond + sB) \times \mathcal{P}(z_\diamond + re_1 + sB), \qquad 0 < s \le \kappa_\star r,$$

the mixed-scale phase diagram is understood up to the same logarithm on the nonparametric side: the critical width is $s_{\mathrm{crit}}(N, r) \asymp rN^{2/d-1/2}$, below it the whole class is parametric, and above it the whole class is nonparametric (Theorems 15.11 and 15.12).

(e) Exact Euclidean translated thin-annulus classes far from the diagonal still exhibit a logarithmic direct-versus-plug-in separation (Theorem 16.6).

(f) Every fixed finite-template positive-baseline multicluster family $\mathcal{C}^{\mathrm{temp}}_{r,s}(\Xi, \Upsilon)$, one-sided or two-sided, is already estimable at or below the global target scale (Theorem 17.3). Moreover, when $s \asymp r$, the whole class is solved up to the same logarithm as the positive-distance smooth-cost theory (Theorem 17.5).

(g) Every fixed shrinking matched-template class $\mathcal{M}^{\mathrm{match}}_{r,s}(\Xi, \Upsilon, \alpha)$ with rigid macro-assignment has the same full mixed-scale phase diagram as the one-pair two-ball problem (Theorem 18.6).

(h) Even for growing rigid matched-template families, the risk is controlled by the explicit complexity profile

$$sN^{-1/2} \sum_i \alpha_i^{1/2} + \frac{s^2}{r} N^{-2/d} \sum_i \alpha_i^{1-2/d},$$

and for equal weights by

$$s\sqrt{\frac{K}{N}} + \frac{s^2}{r}\Big(\frac{K}{N}\Big)^{2/d}$$

(Theorems 19.2 and 19.3).

Conjecture 20.2 (Residual branching conjecture). Fix $d > 2p$ and $n = m = N$. If the unrestricted minimax rate in the balanced supercritical regime is strictly larger than the lower-envelope scale $(N \log N)^{-1/d}$, then the hard instances must necessarily involve two-sided shrinking configurations with *non-rigid branching* on a set of macro-edges whose effective complexity is at least of order

$$\sum_i \alpha_i^{1/2} \gtrsim N^{1/2-1/d} \qquad \text{or} \qquad \sum_i \alpha_i^{1-2/d} \gtrsim N^{1/d},$$

up to logarithmic factors. Equivalently, no family with smaller effective branching complexity can sustain the unrestricted supercritical barrier.

These eight theorem-level layers localize the remaining obstruction more sharply than any of the earlier versions.

Discussion 20.3 (The residual globalization bottleneck). What is still missing is no longer an analysis of arbitrary positive-distance neighborhoods, no longer every one-baseline shrinking family, and no longer any fixed finite-template positive-baseline multicluster geometry. The new finite-template theorem shows that even genuinely two-sided multicluster classes cease to be dangerous as long as their macro-template complexity stays fixed and their cross-baselines remain uniformly positive.

The remaining unrestricted bottleneck must therefore involve at least one of two new features.

(i) *Growing macro-complexity:* the number of statistically relevant source and target macro-regions must itself increase with N, rather than stay bounded.

(ii) *Branching multi-edge baseline collapse:* more than one source-target macro-edge must approach the diagonal simultaneously *and* the macro-assignment must branch, so that no rigid matched decomposition is available.

The critical-grid reduction from Section 11 already points toward a scale $|X_J| \asymp N \log N$. Combined with Theorems 17.3 and 18.6, this suggests a stronger organizing principle: fixed macro-template complexity is too small to sustain the unrestricted lower-envelope scale, and any eventual hard instance must exhibit either genuinely growing two-sided macro-combinatorics or a branching collapse that destroys rigid matched localization.

That principle is not yet a theorem, but it is now backed by several rigorous eliminations: the diagonal class, the full positive-distance product neighborhoods, the one-baseline two-ball family, translated thin annuli, all fixed finite-template positive-baseline multicluster families, and now all fixed shrinking matched-template classes with rigid macro-assignment. What remains is therefore not a vague "globalization step" in the old sense. It is a much more specific task: control two-sided shrinking families whose effective macro-template complexity itself grows toward the critical-grid scale while several cross-baselines simultaneously approach the diagonal.

Conjecture 20.4 (Critical-complexity principle). Fix $d > 2p$. Let $\mathcal{F}_N \subset \mathcal{P}_d \times \mathcal{P}_d$ be any sequence of classes admitting a representation by at most M_N source macro-regions and at most M_N target macro-regions, with all relevant source-target baselines bounded below by a fixed positive multiple of the ambient shrinking scale. If $M_N = O(1)$, then

$$\inf_{\widehat{W}} \sup_{(P,Q)\in\mathcal{F}_N} \mathbb{E}\big|\widehat{W} - W_p(P,Q)\big| = o(\eta_N).$$

More ambitiously, the true unrestricted lower-envelope scale $\eta_N = (N \log N)^{-1/d}$ should require $M_N \to \infty$, and plausibly M_N of order $N \log N$ after the critical-grid reduction.

The conjecture above is intentionally stronger than what is proved here. Its value is organizational: it states in one sentence the common message of the positive-distance, two-ball, thin-annulus, and finite-template results. The only configurations not yet eliminated are those in which the Euclidean transport geometry and the combinatorics of many active macro-cells become critical *simultaneously*.

A An exact Gaussian kernel in the local paired Poisson channel

The paired lower-bound architecture anchored at the balanced reference

$$Q_0 := \frac{1}{2M} \sum_{j=1}^{M} (\delta_{x_j^+} + \delta_{x_j^-})$$

leads, after Poissonization, to independent local blocks. For a single block, fix $\lambda > 0$ and consider the experiment

$$N^+ \sim \mathrm{Poi}(\lambda(1+\theta)), \qquad N^- \sim \mathrm{Poi}(\lambda(1-\theta)),$$

with N^+ and N^- independent and $|\theta| \le 1$. Introduce the local rescaling

$$\mu := \sqrt{2\lambda}\,\theta, \qquad |\mu| \le \sqrt{2\lambda},$$

and let $\mathbb{P}_\mu$ denote the resulting law. The next proposition is not needed for the subclass separation theorem, but it gives a rigorous bridge from the paired channel to the Gaussian shift model.

Proposition A.1 (Exact Gaussian kernel)**.** *For $|\mu|, |\nu|, |t|, |u| \le \sqrt{2\lambda}$, let L_μ denote the likelihood ratio of $\mathbb{P}_\mu$ with respect to $\mathbb{P}_0$. Then*

$$L_\mu(N^+, N^-) = \left(1 + \frac{\mu}{\sqrt{2\lambda}}\right)^{N^+} \left(1 - \frac{\mu}{\sqrt{2\lambda}}\right)^{N^-},$$

and

$$\mathbb{E}_0[L_\mu L_\nu] = e^{\mu\nu}, \qquad \mathbb{E}_\mu[L_t L_u] = e^{\mu(t+u)+tu}.$$

If

$$\psi_r := (2\lambda)^{r/2}\varphi_r,$$

where φ_r is the unbiased local estimator of θ^r obtained from the paired channel, then

$$L_z = \sum_{r=0}^{\infty} \frac{z^r}{r!}\psi_r \qquad (|z| \le \sqrt{2\lambda}),$$

and therefore

$$\mathbb{E}_\mu[\psi_r] = \mu^r.$$

Moreover, if H_r denotes the probabilists' Hermite polynomial, then for every $r, s \ge 0$,

$$\mathbb{E}_\mu[\psi_r\psi_s] = \mathbb{E}_{X\sim N(\mu,1)}[H_r(X)H_s(X)].$$

In particular,

$$\mathbb{E}_0[\psi_r\psi_s] = r!\,\mathbf{1}\{r = s\}.$$

Proof. The Radon–Nikodym derivative of $\mathrm{Poi}(\lambda(1+\theta))$ with respect to $\mathrm{Poi}(\lambda)$ is

$$e^{-\lambda\theta}(1+\theta)^{N^+},$$

and the corresponding derivative for N^- is

$$e^{\lambda\theta}(1-\theta)^{N^-}.$$

Multiplying these expressions and substituting $\theta = \mu/\sqrt{2\lambda}$ gives the formula for L_μ.

Under $\mu = 0$, independence of N^+ and N^- yields

$$\mathbb{E}_0[L_\mu L_\nu] = \exp\left(\lambda\left[\left(1 + \frac{\mu}{\sqrt{2\lambda}}\right)\left(1 + \frac{\nu}{\sqrt{2\lambda}}\right) - 1\right] + \lambda\left[\left(1 - \frac{\mu}{\sqrt{2\lambda}}\right)\left(1 - \frac{\nu}{\sqrt{2\lambda}}\right) - 1\right]\right) = e^{\mu\nu}.$$

Similarly,

$$\mathbb{E}_\mu[L_t L_u] = \mathbb{E}_0[L_\mu L_t L_u].$$

Expanding the N^+- and N^--factors shows that the odd monomials cancel and only the pairwise terms remain, which yields

$$\mathbb{E}_\mu[L_t L_u] = e^{\mu(t+u)+tu}.$$

For the generating function, write

$$L_z = \left(1 + \frac{z}{\sqrt{2\lambda}}\right)^{N^+} \left(1 - \frac{z}{\sqrt{2\lambda}}\right)^{N^-} = \sum_{a,b\geq 0} \frac{(N^+)_a}{a!} \frac{(N^-)_b}{b!} \left(\frac{z}{\sqrt{2\lambda}}\right)^a \left(-\frac{z}{\sqrt{2\lambda}}\right)^b,$$

where $(N)_a$ denotes the falling factorial. Collecting coefficients of z^r gives

$$L_z = \sum_{r=0}^{\infty} \frac{z^r}{r!} \psi_r.$$

Taking $\mathbb{E}_\mu$ of both sides and comparing coefficients implies $\mathbb{E}_\mu[\psi_r] = \mu^r$.

Finally, the Hermite generating function is

$$e^{zX - z^2/2} = \sum_{r=0}^{\infty} \frac{z^r}{r!} H_r(X).$$

If $X \sim N(\mu, 1)$, then

$$\mathbb{E}\big[e^{tX - t^2/2} e^{uX - u^2/2}\big] = e^{\mu(t+u)+tu}.$$

Comparing this with the identity for $\mathbb{E}_\mu[L_t L_u]$ proves the second-moment equality and, in particular, the orthogonality relation at $\mu = 0$. □

Corollary A.2 (Polynomial moment transfer)**.** *Let*

$$P(x) = \sum_{r=0}^{K} a_r x^r$$

be a polynomial, and define

$$P^\sharp := \sum_{r=0}^{K} a_r \psi_r, \qquad P^{\mathrm{Her}}(X) := \sum_{r=0}^{K} a_r H_r(X).$$

If $X \sim N(\mu, 1)$, then

$$\mathbb{E}_\mu[P^\sharp] = \mathbb{E}[P^{\mathrm{Her}}(X)], \qquad \mathbb{E}_\mu[(P^\sharp)^2] = \mathbb{E}[(P^{\mathrm{Her}}(X))^2].$$

Proof. Expand both sides and apply Theorem A.1. □

Lemma A.3 (Subgaussian score tails)**.** *Let*

$$Z := \psi_1 = \frac{N^+ - N^-}{\sqrt{2\lambda}}.$$

Then $\mathbb{E}_\mu[Z] = \mu$, and for every $|s| \leq \sqrt{2\lambda}$,

$$\mathbb{E}_\mu e^{s(Z-\mu)} \leq e^{s^2}.$$

Consequently, for every $0 \le t \le 2\sqrt{2\lambda}$,

$$\mathbb{P}_\mu(|Z-\mu| \ge t) \le 2e^{-t^2/4}.$$

Proof. The identity $\mathbb{E}_\mu[Z] = \mu$ follows from $\mathbb{E}_\mu[\psi_1] = \mu$. For the moment generating function,

$$\log \mathbb{E}_\mu e^{s(Z-\mu)} = \lambda(1+\theta)\left(e^{s/\sqrt{2\lambda}} - 1 - \frac{s}{\sqrt{2\lambda}}\right) + \lambda(1-\theta)\left(e^{-s/\sqrt{2\lambda}} - 1 + \frac{s}{\sqrt{2\lambda}}\right),$$

where $\theta = \mu/\sqrt{2\lambda}$. If $|s| \le \sqrt{2\lambda}$, then $|s|/\sqrt{2\lambda} \le 1$, and the elementary bounds $e^u - 1 - u \le u^2$ and $e^{-u} - 1 + u \le u^2$ for $|u| \le 1$ imply

$$\log \mathbb{E}_\mu e^{s(Z-\mu)} \le \lambda(1+\theta)\frac{s^2}{2\lambda} + \lambda(1-\theta)\frac{s^2}{2\lambda} = s^2.$$

The tail bound follows by Chernoff's inequality with $s = t/2$. □

Remark A.4. Theorems A.1 to A.3 show that the local paired Poisson channel has exactly the same reproducing kernel and exactly the same polynomial moment calculus as the Gaussian shift model. This is a rigorous explanation for why Cai–Low type nonsmooth-functional methods are a natural candidate on the dense side of the paired architecture.

B Almost-equilateral supports and transport-cost transfer

This appendix preserves one more correct theme from the earlier codebook analysis: exact equilateral reductions are rigid, but approximate equilateral geometry still transfers the transport *cost* W_p^p to total variation up to a controlled multiplicative distortion.

Let $X = \{x_1, \dots, x_M\} \subset [0,1]^d$, and write

$$F_X(q) := \sum_{i=1}^{M} q_i\, \delta_{x_i}, \qquad U_X(q,r) := W_p(F_X(q), F_X(r))^p.$$

Definition B.1 (Almost-equilateral codebook)**.** A finite set $X = \{x_1, \dots, x_M\} \subset [0,1]^d$ is called (Δ, ε)-equilateral if

$$\Delta \le \|x_i - x_j\|_2 \le (1+\varepsilon)\Delta \qquad (i \ne j).$$

Proposition B.2 (Approximate reduction to total variation)**.** *If X is (Δ, ε)-equilateral, then for every $q, r \in \Delta_M$,*

$$\Delta^p \,\mathrm{TV}(q,r) \le U_X(q,r) \le (1+\varepsilon)^p \Delta^p \,\mathrm{TV}(q,r).$$

Consequently, with

$$\eta_{p,\varepsilon} := (1+\varepsilon)^p - 1,$$

one has

$$\sup_{q,r \in \Delta_M} \left| \frac{U_X(q,r)}{\Delta^p} - \mathrm{TV}(q,r) \right| \le \eta_{p,\varepsilon}.$$

Proof. Fix $q, r \in \Delta_M$, and let $P = F_X(q)$, $Q = F_X(r)$. After keeping the overlap fixed, the amount of unmatched mass equals

$$\mathrm{TV}(q, r) = \frac{1}{2}\|q - r\|_1.$$

Every unit of unmatched mass must move between two distinct support points, and every such move costs at least Δ^p and at most $(1+\varepsilon)^p\Delta^p$. This gives the two-sided bound. The last display follows immediately because $\mathrm{TV}(q, r) \le 1$. □

Corollary B.3 (Minimax transfer for the transport cost)**.** *Let*

$$R^{(p\text{-}cost)}_{n,m,p}(X) := \inf_{\widehat{U}} \sup_{q,r\in\Delta_M} \mathbb{E}\big(\widehat{U} - U_X(q, r)\big)^2,$$

and let

$$R^{\mathrm{TV}}_{n,m,M} := \inf_{\widehat{T}} \sup_{q,r\in\Delta_M} \mathbb{E}\big(\widehat{T} - \mathrm{TV}(q, r)\big)^2.$$

If X is (Δ, ε)-equilateral, then

$$\frac{\Delta^{2p}}{2}\Big(R^{\mathrm{TV}}_{n,m,M} - 2\eta^2_{p,\varepsilon}\Big) \le R^{(p\text{-}cost)}_{n,m,p}(X) \le 2\Delta^{2p}\Big(R^{\mathrm{TV}}_{n,m,M} + \eta^2_{p,\varepsilon}\Big).$$

In particular, whenever $\eta^2_{p,\varepsilon} \lesssim R^{\mathrm{TV}}_{n,m,M}$,

$$R^{(p\text{-}cost)}_{n,m,p}(X) \asymp \Delta^{2p} R^{\mathrm{TV}}_{n,m,M}.$$

Proof. Let $\widehat{U}$ be any estimator of $U_X(q, r)$, and define

$$\widehat{T} := \frac{\widehat{U}}{\Delta^p}.$$

By Theorem B.2,

$$|\widehat{T} - \mathrm{TV}(q, r)| \le \Delta^{-p}|\widehat{U} - U_X(q, r)| + \eta_{p,\varepsilon}.$$

Squaring and averaging gives

$$R^{\mathrm{TV}}_{n,m,M} \le 2\Delta^{-2p} R^{(p\text{-cost})}_{n,m,p}(X) + 2\eta^2_{p,\varepsilon},$$

which rearranges to the lower bound.

Conversely, if $\widehat{T}$ is any estimator of $\mathrm{TV}(q, r)$, define

$$\widehat{U} := \Delta^p \widehat{T}.$$

Then

$$|\widehat{U} - U_X(q, r)| \le \Delta^p|\widehat{T} - \mathrm{TV}(q, r)| + \Delta^p\eta_{p,\varepsilon}.$$

Squaring and averaging yields

$$R^{(p\text{-cost})}_{n,m,p}(X) \le 2\Delta^{2p}\Big(R^{\mathrm{TV}}_{n,m,M} + \eta^2_{p,\varepsilon}\Big).$$ □

Remark B.4. The corollary preserves the earlier "almost-equilateral codebook" mechanism. Its real value lies in regimes where one can build large deterministic supports with uniformly small ε. In fixed dimension such constructions cannot be exact one-scale solutions by Theorems 6.2

and 6.6, but they remain a useful geometric bridge between Euclidean transport costs and discrete nonsmooth functionals.

References

[1] T. T. Cai and M. G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Annals of Statistics*, 39(2):1012–1041, 2011.

[2] S. Chewi, J. Niles-Weed, and P. Rigollet. Estimation of Wasserstein distances. In *Statistical Optimal Transport*, Lecture Notes in Mathematics 2364, pages 37–76. 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(3–4):707–738, 2015.

[4] J. Jiao, Y. Han, and T. Weissman. Minimax estimation of the L_1 distance. *IEEE Transactions on Information Theory*, 64(10):6672–6706, 2018.

[5] T. Manole and J. Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *Annals of Applied Probability*, 34(1A):1–43, 2024.

[6] J. Niles-Weed and P. Rigollet. Estimation of Wasserstein distances in the spiked transport model. *Bernoulli*, 28(4):2663–2688, 2022.

[7] S. Wang, T. T. Cai, and H. Li. Optimal estimation of Wasserstein distance on a tree with an application to microbiome studies. *Journal of the American Statistical Association*, 116(535):1237–1253, 2021.

[8] 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.