Minimax Optimal Estimation of Transport-Growth Pairs in Unbalanced Optimal Transport

Donlapark Ponnoprat
Department of Statistics
Chiang Mai University
donlapark.p@cmu.ac.th

Noboru Isobe
RIKEN AIP
noboru.isobe@riken.jp

Masaaki Imaizumi
The University of Tokyo / RIKEN AIP / Kyoto University
imaizumi@g.ecc.u-tokyo.ac.jp

Abstract

Unbalanced optimal transport extends classical optimal transport to measures with different total masses, but statistical guarantees for estimation of the transport map and the growth map remain limited. We study unbalanced optimal transport with quadratic cost and Kullback-Leibler marginal penalties and argue that the natural target of estimation is a transport-growth pair. Consequently, we develop two estimators for the transport-growth pair under two setups: a nearest-neighbor estimator for transporting between general densities, and a kernel-based estimator for transporting between smooth densities. We prove that the kernel-based estimator achieves the minimax optimal rate up to a logarithmic factor by deriving a matching lower bound on the minimax risk. Our main technical contribution is a reduction of estimation to a stability bound on the unbalanced optimal transport value. This bound controls the transport map and growth map risks through the duality gap of the true dual potentials for the estimated measures. These results provide a statistical foundation for estimation of the transport-growth pair in unbalanced optimal transport.

1 Introduction

1.1 Background

Optimal transport (OT) provides a principled geometric language for comparing probability distributions and has become a standard tool in statistics and machine learning. In statistical applications, however, the underlying measures are rarely observed directly; instead, one must estimate transport maps and transport plans from finite samples. For balanced OT, this question is now supported by a substantial statistical theory. Smooth transport maps admit minimax analysis via semi-dual curvature and growth arguments [HR21], while plug-in, barycentric-projection, and entropic estimators have been analyzed in discrete, semi-discrete, and smooth regimes [DGS21, PNW21, PDNW23, MBNWW24]. More recently, the balanced theory has expanded to general function-space analyses and sharper stability reductions for plug-in estimators [DNWP25, BM26].

Many datasets are not naturally balanced: the total masses may differ, unmatched observations may be present, and mass creation or destruction may be an intrinsic feature of the phenomenon under study. Unbalanced optimal transport (UOT) addresses this issue by relaxing the hard marginal constraints of OT and penalizing deviations of the transport plan marginals from the reference

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

measures. This relaxation leads to the entropy-transport problems, which include the logarithmic entropy-transport problem and the Wasserstein-Fisher-Rao distance, also called the Hellinger-Kantorovich distance [CPSV18, LMS18, SS24, GGV25]. The relaxation has also generated a large computational and applied literature, including Sinkhorn algorithms and scalable parameterizations [PLH+20, SVP22, GABK24, YU19], with applications ranging from single-cell dynamics to growth modeling and shape analysis [SST+19, SQZN23, DC20, BHK22].

Despite this rapid development, the statistical theory of UOT remains less developed, especially for transport-growth pairs. For the UOT model of this paper, [VV22, VV23] derived semi-dual formulations and established global upper and lower bounds on the Bregman divergence of the semi-dual objective, yielding the first fast statistical rates for UOT semi-dual potentials. More recently, statistical properties of unbalanced Kantorovich-Rubinstein quantities have been analyzed for finitely supported measures and for spatio-temporal point-process models [HHK+25, SHSM26]. However, compared with the balanced case, minimax rates for estimating transport-growth pairs are still largely unknown.

The main difficulty is structural. In balanced OT, the specific form of the transport objective allows us to bound the transport error by exploiting the semi-dual form of the transport problem [HR21, MBNWW24, BM26]. UOT, by contrast, has a different objective where the hard marginal constraints are replaced with divergence penalties; this modification introduces a *growth map* that specifies how the masses will contract or expand after the transport. Consequently, a rigorous statistical theory for UOT must account for the risks of estimating the transport map and the growth map.

1.2 Our contribution

In this paper, we develop a statistical theory for estimation of the transport-growth pair in unbalanced optimal transport with KL marginal penalties. In particular, we develop two estimators for the transport-growth pair, then study their risks.

Concretely, our contributions are as follows.

- **Estimators for the transport-growth pair:** We introduce the minimax estimation theory for the transport-growth pair in the Gaussian-Hellinger case of the UOT problem, and develop two estimators for this pair.
- **Proved convergence rates and optimality:** We derive upper bounds on the risks of our estimators and a lower bound on the minimax risk. In particular, one of our estimators achieves the minimax optimal convergence rate up to a logarithmic factor. This rate improves upon the existing rates, as summarized in Table 1.
- **Proof technique by stability bound:** To analyze the transport-growth pair, we establish a stability bound and a first-order expansion of the UOT value under perturbations of the marginals, which separates the step specific to UOT from the choice of the estimators of the source and target measures.

Table 1: Comparison of convergence rates of the risk. $d \geq 5$ is the dimension of samples, $N = \min\{n, m\}$, where n and m are the sample sizes from the measures, $\widetilde{O}(\cdot)$ hides polylogarithmic factors, and $\alpha - 1$ is the smoothness of the densities of the measures. For [VV22], their regularity argument yields $(\alpha+1)$-smooth semi-dual potentials with $\alpha + 1 < d/2$.

Reference	Setup	transport estimation	growth estimation	Convergence rate	Lower bound
[HR21] [MBNWW24]	balanced (w/ smoothness)	✓	N/A	$\widetilde{O}\left(N^{-2\alpha/(2\alpha-2+d)}\right)$	✓
[VV22]	unbalanced (w/ smoothness)	✓		$O\left(N^{-(\alpha+1)/d}\right)$	
[VV23]	unbalanced	✓		$O\left(N^{-2/d}\right)$	
This paper (nearest-neighbor)	unbalanced	✓	✓	$\widetilde{O}(N^{-2/d})$	
This paper (kernel-based)	unbalanced (w/ smoothness)	✓	✓	$O\left(N^{-2\alpha/(2\alpha-2+d)}\right)$	✓

1.3 Notations

Throughout, $\Omega \subset \mathbb{R}^d$ denotes the ambient domain. We write $C(\Omega)$ for the space of continuous real-valued functions on Ω, and $C_b(\Omega)$ for the bounded continuous functions. The set of finite positive Radon measures on Ω is denoted by $\mathcal{M}_+(\Omega)$, and $\mathcal{P}(\Omega)$ denotes the Borel probability measures on Ω. For a measurable map $T : \Omega \to \Omega$ and a measure $\mu \in \mathcal{M}_+(\Omega)$, $T_\# \mu$ denotes the pushforward of μ. For $\gamma \in \mathcal{M}_+(\Omega \times \Omega)$, we write γ_0 and γ_1 for the first and second marginals of γ, respectively. If $\eta, \sigma \in \mathcal{M}_+(\Omega)$ have the same total mass, then $\Pi(\eta, \sigma)$ denotes the set of couplings between η and σ. We use $\|\cdot\|$ for the Euclidean norm, $\|\cdot\|_{\mathrm{op}}$ for the operator norm, and $\mathbf{1}\{\cdot\}$ for the indicator function. For $\eta, \sigma \in \mathcal{M}_+(\Omega)$ with the same total mass, we write $W_2(\eta, \sigma) := \left(\inf_{\gamma \in \Pi(\eta,\sigma)} \int_{\Omega\times\Omega} \|x - y\|^2 \, d\gamma(x, y)\right)^{1/2}$ for the 2-Wasserstein distance. For $\eta, \sigma \in \mathcal{M}_+(\Omega)$ with $\eta \ll \sigma$, we write $D_{\mathrm{KL}}(\eta \mid \sigma) := \int_\Omega F(d\eta/d\sigma) \, d\sigma$ for the KL divergence, where $F(r) := r \log r - r + 1$. Otherwise, $D_{\mathrm{KL}}(\eta \mid \sigma) := +\infty$. We write $a \lesssim b$ or $b \gtrsim a$ if $a \le Cb$ for some constant $C > 0$, and $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. For real numbers $a < b$, the clipping function is $\mathrm{clip}_{[a,b]}(t) := \min\{b, \max\{a, t\}\}$.

2 Unbalanced optimal transport with quadratic cost

We briefly review unbalanced optimal transport in the *Gaussian-Hellinger* case. For a complete treatment of the general case, see [LMS18, SS24, GGV25]. Let $\Omega \subset \mathbb{R}^d$ be a bounded convex domain, and let $\mu, \nu \in \mathcal{M}_+(\Omega)$ be finite positive Radon measures. We write $c(x, y) := \frac{1}{2}\|x - y\|^2$ for the quadratic cost. The *c-transform* of a function $\varphi : \Omega \to \mathbb{R}$ is defined as $\varphi^c(x) = \inf_{y\in\Omega}\{c(x, y) - \varphi(y)\}$. We say that φ is *c-concave* if $\varphi = \psi^c$ for some $\psi : \Omega \to \mathbb{R}$.

2.1 Transport formulations

Monge problem: We first formulate UOT as an optimization over a transport map and a growth map. A *transport-growth pair* consists of a measurable map $T : \Omega \to \Omega$ and a measurable weight $\lambda : \Omega \to [0, \infty)$, and maps μ to the weighted pushforward

$$(T, \lambda)_\# \mu := T_\#(\lambda^2 \mu).$$

The map λ is referred to as the *growth map*, since λ^2 rescales mass after the transport. In the Gaussian-Hellinger case, we define the cone cost

$$C((x, r), (y, s))^2 := r^2 + s^2 - 2rs\, e^{-\|x-y\|^2/4}, \qquad x, y \in \Omega,\ r, s \ge 0.$$

Using this cone cost, we obtain the *unbalanced Monge problem* for the Gaussian-Hellinger case:

$$\mathrm{UM}(\mu, \nu) := \inf_{(T,\lambda):\,(T,\lambda)_\#\mu=\nu} \int_\Omega \left(1 + \lambda(x)^2 - 2\lambda(x)\, e^{-\|x-T(x)\|^2/4}\right) d\mu(x). \tag{1}$$

In contrast to balanced OT, where only a transport map appears, UOT also requires a growth map to reconcile the discrepancy between the total masses of μ and ν.

Kantorovich problem: We consider another problem for unbalanced optimal transport by optimizing over transport plans $\gamma \in \mathcal{M}_+(\Omega \times \Omega)$ instead of deterministic transport-growth pairs for the Monge problem. Specifically, the Kantorovich problem associated with (1) is given by

$$\mathrm{UOT}(\mu, \nu) := \inf_{\gamma' \in \mathcal{M}_+(\Omega\times\Omega)} \left\{ \int_{\Omega\times\Omega} \frac{\|x - y\|^2}{2} \, d\gamma'(x, y) + D_{\mathrm{KL}}(\gamma'_0 \mid \mu) + D_{\mathrm{KL}}(\gamma'_1 \mid \nu) \right\}. \tag{2}$$

The corresponding dual problem is as follows:

$$\mathrm{UOT}(\mu, \nu) = \sup_{\substack{(\varphi,\psi)\in C_b(\Omega)^2 \\ \varphi(x)+\psi(y)\le\|x-y\|^2/2}} \left\{ \int_\Omega \left(1 - e^{-\varphi(x)}\right) d\mu(x) + \int_\Omega \left(1 - e^{-\psi(y)}\right) d\nu(y) \right\}. \tag{3}$$

Let (φ_0, ψ_0) denote an optimal dual pair for (3), and we refer to φ_0 and ψ_0 as *potentials*. Existence of the solutions of (1) and (3) is guaranteed by the following theorem:

Theorem 1 ([GGV25]). *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex, let $\mu = \rho_0\, dx$ with $\rho_0 > 0$ a.e., and let $\nu \ll dx$ be supported in Ω. Then there exists a μ-a.e. unique c-concave φ_0 such that (φ_0, ψ_0) solves the dual problem (3) with $\psi_0 = \varphi_0^c$, and φ_0 uniquely determines a solution (T_0, λ_0) of the Monge problem (1) via:*

$$T_0(x) = x - \nabla\varphi_0(x), \qquad \lambda_0(x) = \exp\left(-\varphi_0(x) + \frac{1}{4}\|\nabla\varphi_0(x)\|^2\right). \tag{4}$$

Useful form: Following Theorem 1, we define the *active marginal* $\gamma_0 = e^{-\varphi_0}\mu$ and the function $a_0 : \Omega \to (0, \infty)$, $a_0(x) := e^{-\varphi_0(x)/2}$, which is the square root of the density $d\gamma_0/d\mu$. It follows from complementary slackness that $\varphi_0(x) + \psi_0(T_0(x)) = \|x - T_0(x)\|^2/2$. We thus have the following gradient-free formulation of T_0 and λ_0:

$$T_0(x) \in \operatorname*{arg\,min}_{y\in\Omega}\left\{\frac{\|x-y\|^2}{2} - \psi_0(y)\right\}, \qquad \lambda_0(x) = a_0(x)^2 \exp\left(\frac{1}{4}\|x - T_0(x)\|^2\right).$$

2.2 Statistical estimation problem of the transport-growth pair

Given finite positive measures $\mu, \nu \in \mathcal{M}_+(\Omega)$, we consider total masses $M_\mu := \mu(\Omega)$ and $M_\nu := \nu(\Omega)$ and define the normalized measures $\bar\mu := \mu/M_\mu, \bar\nu := \nu/M_\nu$. Given sample sizes $n, m \in \mathbb{N}$, we observe mutually independent samples $X_1, \dots, X_n \sim \bar\mu$ and $Y_1, \dots, Y_m \sim \bar\nu$. Because the data are sampled from the normalized measures $\bar\mu$ and $\bar\nu$, the samples do not identify the total masses M_μ and M_ν. We therefore assume either that these masses are known, or that estimated masses $\widehat{M}_\mu$ and $\widehat{M}_\nu$ are available from external information. We then define atomic weights $\widehat\mu_i := \widehat{M}_\mu/n$, $\widehat\nu_j := \widehat{M}_\nu/m$ and the empirical measures

$$\widehat\mu_n := \frac{\widehat{M}_\mu}{n}\sum_{i=1}^n \delta_{X_i}, \qquad \widehat\nu_m := \frac{\widehat{M}_\nu}{m}\sum_{j=1}^m \delta_{Y_j}.$$

Our objective is to estimate the transport-growth pair (T_0, λ_0) associated with (μ, ν) from the samples $X_1, \dots, X_n, Y_1, \dots, Y_m$ and the estimated masses $\widehat{M}_\mu, \widehat{M}_\nu$. Our proofs also estimate a_0, since $\gamma_0 = a_0^2\mu$ expresses the active marginal through a_0.

Remark 1. The construction of $\widehat{M}_\mu$ and $\widehat{M}_\nu$ depends on the observation model. If external measurements of M_μ and M_ν are available, these measurements can serve directly as $\widehat{M}_\mu$ and $\widehat{M}_\nu$. In Poisson or more general point-process models, the observed counts contain information about the masses; see Appendix C.3 and the recent UOT analyses of [HHK+25, SHSM26].

3 Estimator design

We propose two estimators: (i) a nearest-neighbor estimator and (ii) a kernel-based estimator; the former can handle high-dimensional data, while the latter uses the smoothness of the densities to attain a faster rate.

We introduce some notation: for a nonnegative matrix $G = (G_{ij}) \in \mathbb{R}_+^{n\times m}$, define the discrete marginals $(G_0)_i := \sum_{j=1}^m G_{ij}$ and $(G_1)_j := \sum_{i=1}^n G_{ij}$. The corresponding discrete KL penalty between mass vectors is $D_{\mathrm{KL}}(G_0 \mid \widehat\mu_n) := \sum_{i=1}^n \widehat\mu_i\, F\big((G_0)_i/\widehat\mu_i\big)$.

3.1 Nearest-neighbor estimator

This estimator is constructed by solving the Kantorovich problem under the empirical measures $\widehat\mu_n$ and $\widehat\nu_m$ and using the discrete optimal transport plan, based on the following steps.

(i) Transport plan estimation: We define an estimator $\widehat\gamma$ of the unbalanced optimal transport plan as any solution of (2) with μ and ν replaced by $\widehat\mu_n$ and $\widehat\nu_m$, that is,

$$\widehat\gamma \in \operatorname*{arg\,min}_{\gamma'\in\mathbb{R}_+^{n\times m}}\left\{\sum_{i=1}^n\sum_{j=1}^m c(X_i, Y_j)\gamma'_{ij} + D_{\mathrm{KL}}(\gamma'_0 \mid \widehat\mu_n) + D_{\mathrm{KL}}(\gamma'_1 \mid \widehat\nu_m)\right\}.$$

(ii) Estimate discrete transport: Define the row masses $\widehat{r}_i := \sum_{j=1}^m \widehat{\gamma}_{ij}$, $i = 1, \dots, n$. We then define a point estimator $\widehat{T}_i$ of $T_0(X_i)$ by the barycentric projection: if $\widehat{r}_i > 0$,

$$\widehat{T}_i = \arg\min_{y \in \Omega} \sum_{j=1}^m \widehat{\gamma}_{ij} c(y, Y_j) = \frac{1}{\widehat{r}_i} \sum_{j=1}^m \widehat{\gamma}_{ij} Y_j \in \mathbb{R}^d. \tag{5}$$

If $\widehat{r}_i = 0$, then $\sum_{j=1}^m \widehat{\gamma}_{ij} c(y, Y_j) = 0$ for every $y \in \Omega$, and $\widehat{T}_i$ is not identifiable from $\widehat{\gamma}$. In this case, we set $\widehat{T}_i := X_i$ by convention. Motivated by the identity $\gamma_0 = a_0^2 \mu$ from Section 2, we define an estimator $\widehat{a}_i := \sqrt{\widehat{r}_i / \widehat{\mu}_i}$ for $a_0(X_i)$. To estimate the value $\lambda_0(X_i)$ of the growth map, we fix constants $0 < w_- < w_+$ such that $w_0 := e^{-\varphi_0}$ satisfies $w_- \le w_0 \le w_+$ on Ω, and set

$$\widehat{\lambda}_i := \mathrm{clip}_{[w_-, w_+]}(\widehat{a}_i^2) \exp\left(\frac{1}{4} \|X_i - \widehat{T}_i\|^2 \right). \tag{6}$$

By Lemma 7 in Appendix D, constants $0 < w_- < w_+$ with $w_- \le e^{-\varphi_0} \le w_+$ on Ω always exist. In particular, if M_μ and M_ν are known, the clipping constants may be chosen with $w_+ / w_- \le e^{\mathrm{diam}(\Omega)^2} (1 + M_\mu / M_\nu)(1 + M_\nu / M_\mu)$.

(iii) Estimate transport/growth map: Since $(\widehat{T}_i, \widehat{a}_i, \widehat{\lambda}_i)$ are defined only at the sample points $X_1, \dots, X_n$, we extend $(\widehat{T}_i, \widehat{a}_i, \widehat{\lambda}_i)_{i=1}^n$ to all of Ω. Here, we employ the nearest-neighbor method. Let $(V_i)_{i=1}^n$ be the Voronoi partition induced by $X_1, \dots, X_n$: $V_i := \{x \in \Omega : \|x - X_i\| \le \|x - X_k\| \ \forall k \ne i\}$. Setting $v_i(x) := \mathbf{1}\{x \in V_i\}$ yields the nearest-neighbor estimators:

$$\widehat{T}^{\mathrm{1NN}}(x) = \sum_{i=1}^n v_i(x) \widehat{T}_i, \ \ \widehat{a}^{\mathrm{1NN}}(x) = \sum_{i=1}^n v_i(x) \widehat{a}_i, \ \ \widehat{\lambda}^{\mathrm{1NN}}(x) = \sum_{i=1}^n v_i(x) \widehat{\lambda}_i, \tag{7}$$

where $\widehat{a}^{\mathrm{1NN}}$ and $\widehat{\lambda}^{\mathrm{1NN}}$ estimate a_0 and the growth map λ_0, respectively.

In addition to the nearest-neighbor estimator, the plan-based kernel estimator can also be considered; see Definition 2 in Appendix C.2.

3.2 Kernel-based estimator

We next present a kernel-based estimator for the transport-growth pair on the hypercube $\Omega = [0,1]^d$. This estimator is based on density estimates of the measures μ, ν and attains a faster rate when these densities are smooth.

(i) Prepare kernels: We introduce a new kernel function based on a cosine basis for estimating density functions on $[0,1]^d$. This method makes statistical use [Efr10, Tsy09] of the eigenfunctions of the Neumann Laplacian [Str99]. This kernel is useful to relax the constraint that a density function must lie on a torus, a requirement of the kernel estimators in the balanced case [MBNWW24].

We now define K_L. In preparation, we define a function $f(s) := e^{-1/s} \mathbf{1}\{s > 0\}$ for $s \in \mathbb{R}$, and the function $\tau(t) := f(2-t)/(f(2-t) + f(t-1))$ for $t \ge 0$. Then, $\tau \in C^\infty([0,\infty))$, $0 \le \tau \le 1$, $\tau(t) = 1$ for $t \in [0,1]$, and $\tau(t) = 0$ for $t \ge 2$. With $c_0 := 1, c_\ell := \sqrt{2}$ $(\ell \ge 1)$, and for $L \ge 1$, define the one-dimensional Neumann kernel

$$K_L^{(1)}(u,v) := 1 + \sum_{\ell=1}^\infty \tau\left(\frac{\pi^2 \ell^2}{L^2} \right) c_\ell^2 \cos(\pi \ell u) \cos(\pi \ell v), \qquad u, v \in [0,1]. \tag{8}$$

Because τ has compact support, the sum in (8) is finite for each L. We then define

$$K_L(x,y) := \prod_{r=1}^d K_L^{(1)}(x_r, y_r), \qquad x, y \in [0,1]^d. \tag{9}$$

Since each factor $K_L^{(1)}(x_r, y_r)$ contains only $O(L)$ nonzero cosine modes, one evaluation of $K_L(x,y)$ costs $O(dL)$ arithmetic operations. Moreover, we have $\int_{[0,1]^d} K_L(x,y)\, dy = 1$ for $x \in [0,1]^d$. This is the hypercube analogue of a smooth spectral cutoff, but expressed in a form that is computationally tractable in moderate and high dimension: the cosine basis gives the boundary adaptation, while the product form of K_L gives exact separability.

(ii) Estimate measures via densities: We first estimate the normalized densities p and q of $\bar{\mu}$ and $\bar{\nu}$, respectively. Then, we scale the estimates by the estimated masses $\widehat{M}_\mu$ and $\widehat{M}_\nu$ to obtain the measure estimates. In particular, for frequency cutoffs $L_n, L_m \geq 1$, we define the preliminary kernel density estimators

$$\widetilde{p}_n^{\ker}(x) := \frac{1}{n}\sum_{i=1}^{n} K_{L_n}(x, X_i), \qquad \widetilde{q}_m^{\ker}(y) := \frac{1}{m}\sum_{j=1}^{m} K_{L_m}(y, Y_j). \tag{10}$$

We then define the measure estimates, which have total masses $\widehat{M}_\mu$ and $\widehat{M}_\nu$, as

$$\widehat{\mu}_n^{\ker} := \frac{\widehat{M}_\mu(\widetilde{p}_n^{\ker}(x))_+}{\int_{[0,1]^d}(\widetilde{p}_n^{\ker}(u))_+\,du}\,dx, \qquad \widehat{\nu}_m^{\ker} := \frac{\widehat{M}_\nu(\widetilde{q}_m^{\ker}(y))_+}{\int_{[0,1]^d}(\widetilde{q}_m^{\ker}(v))_+\,dv}\,dy.$$

(iii) Estimate maps: Let $(\widehat{\varphi}_{nm}^{\ker}, \widehat{\psi}_{nm}^{\ker})$ be an optimal dual pair for $\mathrm{UOT}(\widehat{\mu}_n^{\ker}, \widehat{\nu}_m^{\ker})$. We define the estimator of a_0 as $\widehat{a}_{nm}^{\ker}(x) := e^{-\widehat{\varphi}_{nm}^{\ker}(x)/2}$. Then, we define estimators for the transport map and the growth map by

$$\widehat{T}_{nm}^{\ker}(x) \in \operatorname*{arg\,min}_{y\in[0,1]^d}\left\{\|x-y\|^2/2 - \widehat{\psi}_{nm}^{\ker}(y)\right\}, \tag{11}$$

$$\widehat{\lambda}_{nm}^{\ker}(x) := \mathrm{clip}_{[w_-,w_+]}(\widehat{a}_{nm}^{\ker}(x)^2)\exp\left(\frac{1}{4}\|x - \widehat{T}_{nm}^{\ker}(x)\|^2\right).$$

4 Minimax optimal rate

4.1 Basic assumptions

Assumption 1. *The potential φ_0 is twice continuously differentiable on Ω, and there exists a constant $\kappa \in (0,1)$ such that*

$$\left(1 - \frac{1}{\kappa}\right) I \preceq \nabla^2\varphi_0(x) \preceq (1-\kappa)I, \qquad x \in \Omega.$$

The bounds in Assumption 1 are equivalent to $\kappa I \preceq \nabla^2 f_0(x) \preceq \kappa^{-1} I$ for the Brenier potential $f_0(x) := \frac{1}{2}\|x\|^2 - \varphi_0(x)$ commonly used in convergence analysis of balanced OT map estimators [HR21, MBNWW24, DNWP25].

Assumption 2. *The domain $\Omega \subset \mathbb{R}^d$ is compact, convex, and satisfies the interior cone condition: there exist $\varepsilon_0, \delta_0 > 0$ such that for all $x \in \Omega$ and $\varepsilon \in (0, \varepsilon_0)$, $\mathrm{Vol}(B(x,\varepsilon)\cap\Omega) \geq \delta_0 \mathrm{Vol}(B(x,\varepsilon))$.*

Assumption 3. *The positive finite measures μ and ν have total masses M_μ and M_ν. Their normalized measures $\bar{\mu} = \mu/M_\mu$ and $\bar{\nu} = \nu/M_\nu$ admit Lebesgue densities p and q bounded away from zero and infinity on Ω: $0 < \beta_{\min} \leq p(x), q(x) \leq \beta_{\max} < \infty$. Equivalently, μ and ν have Lebesgue densities $M_\mu p$ and $M_\nu q$.*

Assumptions 1, 2, and 3 are standard in the statistical literature on transport map estimation: similar assumptions on the densities, on the transport potential or on the map itself appear in the balanced case [HR21, MBNWW24, BM26] and the unbalanced case [GGV25].

Under Assumptions 1, 2, and 3, the gap in the dual constraint $\varphi_0(x) + \psi_0(y) \leq \frac{1}{2}\|x-y\|^2$ is bounded below by a quadratic function of $\|y - T_0(x)\|$:

Lemma 2. *Suppose that Assumptions 1, 2, and 3 hold. Then the dual potentials (φ_0, ψ_0) satisfy:*

$$\frac{1}{2}\|x-y\|^2 - \varphi_0(x) - \psi_0(y) \geq \frac{\kappa}{2}\|y - T_0(x)\|^2, \qquad x, y \in \Omega.$$

Assumption 4. *There exist positive real sequences $a_n, b_m \to 0$ as $n, m \to \infty$ and a constant $\bar{c} \in (0,1]$ such that $\widehat{M}_\mu \geq \bar{c} M_\mu$ and $\widehat{M}_\nu \geq \bar{c} M_\nu$ almost surely for all $n, m \in \mathbb{N}$, and $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \leq a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \leq b_m$ for all sufficiently large n, m.*

Assumption 4 is mild and is satisfied, for example, when the masses are known or when additional measurements of M_μ and M_ν are available. A conditional version holds in point-process models, in which truncated counts estimate the masses [HHK+25, SHSM26]. For that reason we state Assumption 4 in an abstract form instead of focusing on any specific model of the masses.

4.2 Nearest-neighbor estimator

We study the risk of the nearest-neighbor estimator (7). In preparation, we define the following rate:

$$\mathfrak{R}_n^{\mathrm{emp}}(d) := \begin{cases} n^{-1/2}, & d \le 3, \\ (\log n) n^{-1/2}, & d = 4, \\ n^{-2/d}, & d \ge 5, \end{cases}$$

which is the upper bound of [FG15, Theorem 1] on the convergence rate of the empirical distribution in the squared Wasserstein distance W_2^2. Then, we obtain the following rate:

Theorem 3 (Error rate of nearest-neighbor estimator)**.** *Assume that Assumptions 1, 2, 3 and 4 hold. Let $(\widehat{T}^{1\mathrm{NN}}, \widehat{a}^{1\mathrm{NN}}, \widehat{\lambda}^{1\mathrm{NN}})$ be the nearest-neighbor estimators defined in Section 3.1, where the values $\widehat{\lambda}_i$ of $\widehat{\lambda}^{1\mathrm{NN}}$ are defined by (6). Let a_n, b_m be the sequences from Assumption 4. Then there exists a constant $C > 0$ such that, for all sufficiently large n, m,*

$$\mathbb{E}\Big[\int_\Omega \|\widehat{T}^{1\mathrm{NN}}(x) - T_0(x)\|^2 \, d\mu(x)\Big] \le C \log n \Big(M_\mu \mathfrak{R}_n^{\mathrm{emp}}(d) + M_\nu \mathfrak{R}_m^{\mathrm{emp}}(d)\Big), \tag{12}$$

$$\mathbb{E}\Big[\int_\Omega |\widehat{\lambda}^{1\mathrm{NN}}(x) - \lambda_0(x)|^2 \, d\mu(x)\Big] \le C \log n \Big(M_\mu \mathfrak{R}_n^{\mathrm{emp}}(d) + M_\nu \mathfrak{R}_m^{\mathrm{emp}}(d) + a_n + b_m\Big). \tag{13}$$

These bounds show that, for $d \ge 4$, the rates of the nearest-neighbor estimator nearly match the lower bound for discrete approximation of measures under the squared Wasserstein distance W_2^2, which is of order $n^{-2/t}$ for every $t < d$ (cf. Proposition 42). We note that the bound (12) does not have the mass estimation terms a_n, b_m because $\widehat{T}^{1\mathrm{NN}}$ does not depend on $\widehat{M}_\mu, \widehat{M}_\nu$. Rescaling the two masses multiplies the discrete optimal transport plan $\widehat{\gamma}$ by a constant, which cancels in the barycentric projection (5). In contrast, $\widehat{\lambda}^{1\mathrm{NN}}$ depends on the masses through $\widehat{a}_i^2 = \widehat{r}_i / \widehat{\mu}_i$. Furthermore, since a_n and b_m achieve the parametric rate in many cases, a_n and b_m do not affect the convergence rate of the right-hand side of (13).

4.3 Kernel-based estimator on the hypercube

We study the risk of the kernel-based estimator (11) on $[0,1]^d$. Here, we additionally introduce an assumption on the smoothness of the normalized densities of μ, ν.

Definition 1 (Neumann-compatible Hölder class)**.** Let $\mathbb{T}_2^d := (\mathbb{R}/2\mathbb{Z})^d$. For $f : [0,1]^d \to \mathbb{R}$, define the even 2-periodic reflection of f as $\mathcal{E}f : \mathbb{T}_2^d \to \mathbb{R}$ by $(\mathcal{E}f)(x_1, \dots, x_d) := f\big(\vartheta(x_1), \dots, \vartheta(x_d)\big)$, where $\vartheta(t) := \min_{m \in \mathbb{Z}} |t - 2m| \in [0,1]$. For $s > 0$, let $\|\cdot\|_{C^s(\mathbb{T}_2^d)}$ be the Hölder norm of order s (cf. Appendix C.1). For $M > 0$, we write

$$\mathcal{C}_N^s([0,1]^d; M) := \Big\{ f : [0,1]^d \to \mathbb{R} : \|\mathcal{E}f\|_{C^s(\mathbb{T}_2^d)} \le M \Big\},$$

which consists of functions whose coordinatewise even reflection is C^s-smooth on the doubled torus.

Assumption 5 (Smooth density)**.** *Let $\mathcal{C}_N^s([0,1]^d; M)$ be the Neumann-compatible Hölder class. For some $\alpha > 1$ and $M > 0$, the normalized densities satisfy $p, q \in \mathcal{C}_N^{\alpha-1}([0,1]^d; M)$.*

In preparation, we define a rate as follows:

$$\mathfrak{R}_n^{\mathrm{ker}}(\alpha) := \begin{cases} n^{-1}, & d = 1, \\ (\log n) n^{-1}, & d = 2, \\ n^{-2\alpha/(2(\alpha-1)+d)}, & d \ge 3 \end{cases}.$$

Theorem 4 (Error rate of kernel-based estimator)**.** *Assume that $\Omega = [0,1]^d$ and that Assumptions 1, 3, 4, and 5 hold. Assume that $L_n = n^{1/(d+2(\alpha-1))}$ and $L_m = m^{1/(d+2(\alpha-1))}$, and let a_n, b_m be the sequences from Assumption 4. Then there exist constants $C > 0$ and $N_0 \in \mathbb{N}$, depending only on $d, \alpha, M, \beta_{\min}, \beta_{\max}$, the constants κ and $\bar{c}$ of Assumptions 1 and 4, the clipping constants $w_\pm$, and the function τ, such that, for all $n, m \ge N_0$,*

$$\mathbb{E}\left[\int_{[0,1]^d} \|\widehat{T}_{nm}^{\mathrm{ker}}(x) - T_0(x)\|^2 \, d\mu(x)\right] \le C\Big(M_\mu \mathfrak{R}_n^{\mathrm{ker}}(\alpha) + M_\nu \mathfrak{R}_m^{\mathrm{ker}}(\alpha)\Big). \tag{14}$$

If, in addition, n and m are large enough that $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \le a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \le b_m$, then

$$\mathbb{E}\left[\int_{[0,1]^d} |\widehat{\lambda}^{\mathrm{ker}}_{nm}(x) - \lambda_0(x)|^2 \, d\mu(x)\right] \le C\Big(M_\mu \mathfrak{R}^{\mathrm{ker}}_n(\alpha) + M_\nu \mathfrak{R}^{\mathrm{ker}}_m(\alpha) + a_n + b_m\Big). \quad (15)$$

This result demonstrates that, under Assumption 5, the kernel-based estimator of the transport map achieves the rate $\mathfrak{R}^{\mathrm{ker}}_n(\alpha)$, which is faster than the rate $O((\log n) n^{-2/d})$ for $d \ge 5$ of the nearest-neighbor estimator in Theorem 3, potentially mitigating the curse of dimensionality through the smoothness. In fact, this rate is minimax optimal for estimation of smooth transport and growth maps, up to a logarithmic factor, as we will discuss in Section 4.5.

4.4 Proof outline for the upper bounds

We present the case of the nearest-neighbor estimator below, while the approach is similar for the kernel-based estimator (Appendix F).

Let $\widehat{\gamma}$ be an optimizer of $\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)$. Define the row masses $\widehat{r}_i := \sum_{j=1}^m \widehat{\gamma}_{ij}$ and column masses $\widehat{s}_j := \sum_{i=1}^n \widehat{\gamma}_{ij}$ of $\widehat{\gamma}$, $\widehat{r}^\star_i := e^{-\varphi_0(X_i)}\widehat{\mu}_i$ and $\widehat{s}^\star_j := e^{-\psi_0(Y_j)}\widehat{\nu}_j$, the barycentric projections $\widehat{T}_i := \frac{1}{\widehat{r}_i}\sum_{j=1}^m \widehat{\gamma}_{ij} Y_j$, and $\widehat{\Delta}^{\mathrm{bar}}_{nm} := \sum_{i=1}^n \widehat{r}_i \|\widehat{T}_i - T_0(X_i)\|^2$. The proof uses the following inequality, which bounds $\int_\Omega \|\widehat{T}^{\mathrm{1NN}} - T_0\|^2 \, d\gamma_0$, where $\gamma_0 := e^{-\varphi_0}\mu$ is the active marginal, in terms of $\widehat{\Delta}^{\mathrm{bar}}_{nm}$, $D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)$, and the Voronoi cells (proof of Theorem 19):

$$\int_\Omega \|\widehat{T}^{\mathrm{1NN}}(x) - T_0(x)\|^2 \, d\gamma_0(x) \lesssim n M_n\big(\widehat{\Delta}^{\mathrm{bar}}_{nm} + D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)\big) + R_n^2.$$

Here, M_n and R_n denote the maximum Voronoi cell mass and radius. By the Vapnik-Chervonenkis inequality (Lemma 16), $M_n \lesssim (\log n)/n$ with high probability and $\mathbb{E}[R_n^2] \lesssim ((\log n)/n)^{2/d}$.

To bound $\widehat{\Delta}^{\mathrm{bar}}_{nm}$ and $D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)$, we consider the empirical excess:

$$\widehat{E}_n := \mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) - \int (1 - e^{-\varphi_0(x)}) \, d\widehat{\mu}_n - \int (1 - e^{-\psi_0(y)}) \, d\widehat{\nu}_m.$$

We refer to this quantity as the *duality gap* of (φ_0, ψ_0) for the measures $\widehat{\mu}_n, \widehat{\nu}_m$, since (φ_0, ψ_0) is feasible for (3) with these measures. We show that $\widehat{E}_n$ can be expressed as follows (Proposition 15):

$$\widehat{E}_n = \sum_{i=1}^n \sum_{j=1}^m \widehat{\gamma}_{ij}\Big(\tfrac{1}{2}\|X_i - Y_j\|^2 - \varphi_0(X_i) - \psi_0(Y_j)\Big) + D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star) + D_{\mathrm{KL}}(\widehat{s} \mid \widehat{s}^\star).$$

By dual feasibility, all terms on the right-hand side are nonnegative. Hence, $D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star) \le \widehat{E}_n$. By Jensen's inequality and Lemma 2, we have $\widehat{\Delta}^{\mathrm{bar}}_{nm} \le \sum_{i=1}^n \sum_{j=1}^m \widehat{\gamma}_{ij} \|Y_j - T_0(X_i)\|^2 \le \frac{2}{\kappa}\widehat{E}_n$.

Taking the expectation, we then employ our stability bound (Proposition 11), which yields

$$\mathbb{E}[\widehat{E}_n] \lesssim \mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})] + a_n + b_m,$$

where a_n and b_m are the sequences from Assumption 4 and $\bar{\mu}_n, \bar{\nu}_m$ are the normalized empirical measures. Using any existing bound on $\mathbb{E}[W_2^2(\bar{\mu}_n, \bar{\mu})]$ and $\mathbb{E}[W_2^2(\bar{\nu}_m, \bar{\nu})]$, e.g., [FG15], we obtain (12). For $\widehat{T}^{\mathrm{1NN}}$, the terms a_n and b_m do not appear, because the argument applies with $\widehat{M}_\mu = M_\mu$ and $\widehat{M}_\nu = M_\nu$ without changing $\widehat{T}^{\mathrm{1NN}}$.

To bound $\mathbb{E}[\int_\Omega |\widehat{\lambda}^{\mathrm{1NN}} - \lambda_0|^2 \, d\mu]$, we relate the error of $\widehat{\lambda}_i$ to the errors of $\widehat{a}_i$ and $\widehat{T}_i$ using the following bound (Lemma 20):

$$|\widehat{\lambda}_i - \lambda_0(X_i)|^2 \le C_{\mathrm{gr}}\big(|\widehat{a}_i - a_0(X_i)|^2 + \|\widehat{T}_i - T_0(X_i)\|^2\big),$$

for some constant $C_{\mathrm{gr}} > 0$, where $\widehat{a}_i := \sqrt{\widehat{r}_i/\widehat{\mu}_i}$ is a plug-in estimate of $a_0(X_i)$. Multiplying both sides by $\widehat{\mu}_i$ and summing over i, we show in Lemma 23 that $\sum_{i=1}^n \widehat{\mu}_i |\widehat{a}_i - a_0(X_i)|^2 \le D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)$. The rest of the proof proceeds as for $\mathbb{E}[\int_\Omega \|\widehat{T}^{\mathrm{1NN}} - T_0\|^2 \, d\mu]$.

4.5 Minimax lower bound on the hypercube

To match Theorem 4 with a lower bound, we restrict to $\Omega = [0,1]^d$ and assume that the masses M_μ, M_ν are known and that $m = n$. We write $\mathbb{P}^n_{\mu,\nu}$ and $\mathbb{E}^n_{\mu,\nu}$ for the joint distribution and expectation, respectively.

Fix $\alpha > 1$, parameters $M, B < \infty$ and $R > 1$, and the constants κ, $\beta_{\min}$, and $\beta_{\max}$ in Assumptions 1 and 3. Let $\mathcal{U}_\alpha(M, B, R)$ denote the class of pairs $(\mu, \nu) \in \mathcal{M}_+([0,1]^d)^2$ with $\mu([0,1]^d) = 1$ that satisfy the two conditions: (i) Assumptions 1, 3, and 5 hold with these constants, with smoothness radius M in Assumption 5, and (ii) the unique c-concave solution φ_0 of the dual problem (3) and the associated solution (T_0, λ_0) of the Monge problem (1) obtained via Theorem 1 satisfy $\|T_0\|_{C^\alpha([0,1]^d)} + \|\lambda_0\|_{C^\alpha([0,1]^d)} \le B$ and $R^{-1} \le \lambda_0(x) \le R$ for all $x \in [0,1]^d$.

Our lower bound for the transport map follows the construction of the lower bound for balanced OT maps [HR21]. On the other hand, to the best of our knowledge, no minimax lower bound for the growth map has been proved before.

Theorem 5 (Minimax lower bound over $\mathcal{U}_\alpha(M, B, R)$)**.** *For every $\alpha > 1$, there exist constants $M_0, B_0 < \infty$, $R_0 > 1$, and $\kappa_0, b_0 \in (0,1)$ such that, for every $M \ge M_0$, $B \ge B_0$, $R \ge R_0$, $\kappa \le \kappa_0$, $\beta_{\min} \le b_0$, and $\beta_{\max} \ge 1/b_0$, there is a constant $c_{\mathrm{lb}} > 0$, depending only on d, α, M, B, R and the constants κ, $\beta_{\min}$, and $\beta_{\max}$ of Assumptions 1 and 3, such that for all sufficiently large n,*

$$\inf_{\widehat{T}} \sup_{(\mu,\nu)\in\mathcal{U}_\alpha(M,B,R)} \mathbb{E}^n_{\mu,\nu}\left[\int_{[0,1]^d} \|\widehat{T}(x) - T_0(x)\|^2 \, d\mu(x)\right] \ge c_{\mathrm{lb}} \left(n^{-2\alpha/(2\alpha-2+d)} \vee n^{-1}\right),$$

$$\inf_{\widehat{\lambda}} \sup_{(\mu,\nu)\in\mathcal{U}_\alpha(M,B,R)} \mathbb{E}^n_{\mu,\nu}\left[\int_{[0,1]^d} |\widehat{\lambda}(x) - \lambda_0(x)|^2 \, d\mu(x)\right] \ge c_{\mathrm{lb}} \left(n^{-2\alpha/(2\alpha-2+d)} \vee n^{-1}\right),$$

where the infima are taken over all measurable functions of $(X_1, \dots, X_n, Y_1, \dots, Y_n, M_\mu, M_\nu)$.

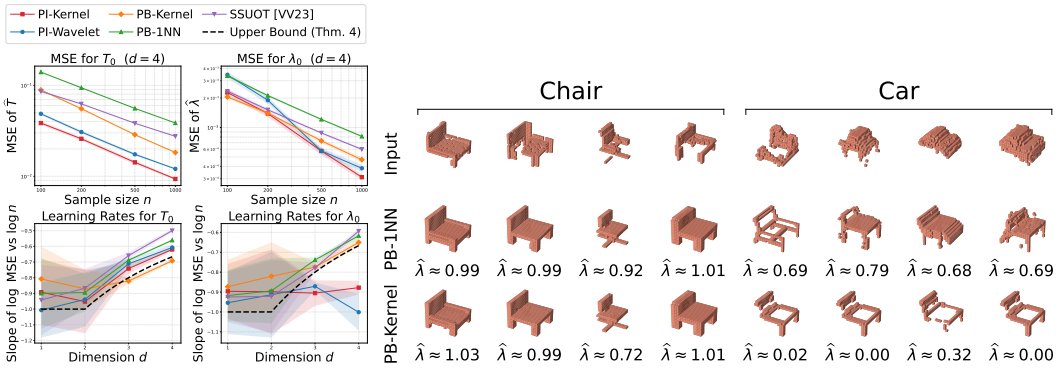
The set of pairs $(\mu, \nu) \in \mathcal{M}_+([0,1]^d)^2$ with $\mu([0,1]^d) = 1$ that satisfy condition (ii) and the assumptions of Theorem 4 on μ and ν, with the constants $\kappa, \beta_{\min}, \beta_{\max}, M$ fixed, is exactly $\mathcal{U}_\alpha(M, B, R)$. Moreover, $w_\pm$ can be taken independent of the pair (μ, ν) in this class (see Lemma 7 in Appendix D, with $M_\mu = 1$ and $M_\nu = \int \lambda_0^2 \, d\mu \in [R^{-2}, R^2]$), so the constants C and N_0 in the upper bounds (14) and (15) do not depend on (μ, ν). Hence, in view of Theorem 4, our lower bound indicates that our kernel-based estimator is minimax optimal up to the logarithmic factor.

5 Experiments

5.1 Simulation study

We sample data from source and target measures $\mu, \nu \in \mathcal{M}_+([0,1]^d)$ with $\mu([0,1]^d) = 1$ and $\nu([0,1]^d) = 2.5$. For $\theta = 0.1$, the normalized source and target densities are the mixtures $p_\theta = (1-\theta)p + \theta$ and $q_\theta = (1-\theta)q + \theta$, where $p(x) \propto \prod_{i=1}^d |\sin(\pi(x_i - 1/3))|$ and $q(x) \propto \prod_{i=1}^d |\sin(\pi(x_i - 2/3))|$. Thus, p_θ and q_θ are bounded below by θ, so Assumption 3 holds with $\beta_{\min} = \theta$. Moreover, p_θ and q_θ are 1-Hölder, which corresponds to $\alpha = 2$ in Assumption 5. We benchmark several estimators of the pair (T_0, λ_0): the nearest-neighbor estimator (PB-1NN, Section 3.1), the plan-based kernel estimator (PB-Kernel, Appendix C.2), the kernel-based estimator (PI-Kernel, Section 3.2), and the wavelet plug-in estimator (PI-Wavelet, Appendix G). We also include the SSUOT estimator of [VV23] as a competing baseline. Additional details are provided in Appendix L.2.

Figure 1a shows MSEs and empirical learning rates of the five methods. The bottom row shows the learning rates for T_0 and λ_0. For $d \le 3$, all the estimators learn at rates faster than the rate $O((\log n) n^{-1/2})$ of Theorem 3. At $d = 4$, the slowest estimator, SSUOT, learns T_0 at the rate $O(n^{-1/2})$. The dashed line is the rate $\mathfrak{R}^{\mathrm{ker}}_n(2)$ of Theorem 4. The learning rates of our plug-in estimators stay close to this line, except for λ_0 at $d \ge 3$, where the learning rates are faster than this line. Moreover, the top row shows that, at $d = 4$, both PI-Kernel and PI-Wavelet attain the smallest MSE for T_0 at every n. For λ_0, PI-Kernel and PI-Wavelet attain the smallest MSE for $n \ge 500$.



(a) (b)

Figure 1: (a) MSE of the four UOT estimators and SSUOT. Top: The MSEs of estimating T_0 and λ_0 vs. n at $d = 4$, averaged over 10 seeds, with one standard error. Bottom: Learning rates for T_0 and λ_0 vs. d, that is, the slopes of the least-squares fit of the log mean MSE on $\log n$, with one standard error of the fitted slope. (b) Top: Incomplete 3D shapes. Middle: Complete 3D shapes predicted by the nearest-neighbor estimator. Bottom: Complete 3D shapes predicted by the plan-based kernel estimator.

5.2 Application to 3D shape completion

To showcase PB-1NN and PB-Kernel for high-dimensional tasks, we apply them to an unpaired 3D shape completion task using the Completion3D dataset [CFG+15, YKH+18]. Specifically, we use the set of incomplete 3D point clouds of chairs as the source dataset and their complete versions as the target dataset. To test robustness, we introduce 30 incomplete cars into the source dataset as outliers. The point clouds are first transformed into a $16 \times 16 \times 16$ grid and subsequently projected via PCA into 1024 dimensions, yielding a source sample of size $n = 5780$ and a target sample of size $m = 5750$ in dimension $d = 1024$. In this task, both source and target measures have unit mass. Nonetheless, we can leverage UOT to selectively reject outliers through the growth map assignment.

We fit PB-1NN and PB-Kernel to this dataset pair and use them to predict the complete shapes of four chairs and four cars in a held-out test set. As Figure 1b illustrates, both estimators successfully recover the shapes of the test chairs. However, their behavior diverges on the outlier test cars: while PB-1NN attempts to reconstruct chairs from the cars, PB-Kernel predicts the mean in the latent PCA space and yields values of $\widehat{\lambda}$ that are close to zero. This indicates PB-Kernel implicitly discards the outliers, demonstrating that PB-Kernel can perform unpaired 3D shape completion in the presence of outliers.

We note that the plug-in estimators and SSUOT are omitted from this experiment due to computational intractability: the plug-in methods require a grid over $[0, 1]^d$, which is not feasible for $d = 1024$, while SSUOT requires solving, at each iteration, the quadratic program of [VV23, Prop. 5.1], which has $(2n + 1)(d + 1)$ variables and $O(n^2)$ constraints. This intractability is a known phenomenon in estimation of smooth transport maps, as discussed in [HR21, MBNWW24, DNWP25].

6 Conclusion

We studied statistical estimation in unbalanced optimal transport with quadratic cost, where the target of estimation is the transport-growth pair, which consists of the transport map and the growth map. We developed the nearest-neighbor estimator and the kernel-based estimator, then derived convergence rates of their risks. We also showed the optimality of the kernel-based estimator by deriving the lower bound. Our main technical contribution is a reduction of estimation to a stability bound and the proof of this bound. Overall, our results provide a statistical foundation for estimation of the transport-growth pair in UOT and clarify the distinct roles of active marginals and growth estimation in this problem. A limitation is that our setup does not cover entropic regularization.

Acknowledgments and Disclosure of Funding

D.P. was supported by JST CREST Grant Number JPMJCR21D2. N.I. was supported by JST ACT-X Grant Number JPMJAX25C2 and JSPS KAKENHI Grant Number JP25K24910. M.I. was supported by JSPS KAKENHI Grant Number JP24K02904, JST CREST Grant Number JPMJCR21D2, JST FOREST Grant Number JPMJFR216I, JST BOOST Grant Number JPMJBY24A9, and JST ERATO Grant Number JPMJER2601. The authors declare no competing interests.

References

[APP26] Francisco Andrade, Gabriel Peyré, and Clarice Poon. Sample complexity of unbalanced entropic ot. *arXiv preprint arXiv:2606.24987*, 2026.

[BHK22] Martin Bauer, Emmanuel Hartman, and Eric Klassen. The square root normal field distance and unbalanced optimal transport. *Applied Mathematics & Optimization*, 85(3), may 2022.

[BM26] Sivaraman Balakrishnan and Tudor Manole. Stability bounds for smooth optimal transport maps and their statistical implications. *Electronic Journal of Statistics*, 20(2):3078 – 3105, 2026.

[CCK24] Jaemoo Choi, Jaewoong Choi, and Myungjoo Kang. Scalable wasserstein gradient flow for generative modeling through unbalanced optimal transport. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[CDV93] Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, 1993.

[CFG+15] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015.

[CFW+21] Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23270–23282. Curran Associates, Inc., 2021.

[CPSV18] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.

[Dau92] Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1992.

[DB16] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[DC20] Di Marino, Simone and Chizat, Lénaïc. A tumor growth model of hele-shaw type as a gradient flow. *ESAIM: COCV*, 26:103, 2020.

[DGS21] Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. In *Advances in Neural Information Processing Systems*, volume 34, pages 29736–29753, 2021.

[DNWP25] Vincent Divol, Jonathan Niles-Weed, and Aram-Alexandre Pooladian. Optimal transport map estimation in general function spaces. *The Annals of Statistics*, 53(3):963–988, 2025.

[Efr10] Sam Efromovich. Orthogonal series density estimation. *WIREs Computational Statistics*, 2(4):467–476, 2010.

[FCG+21] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

[FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3–4):707–738, 2015.

[GABK24] Milena Gazdieva, Arip Asadulaev, Evgeny Burnaev, and Alexander Korotin. Light unbalanced optimal transport. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 93907–93938. Curran Associates, Inc., 2024.

[GGV25] Thomas Gallouët, Roberta Ghezzi, and François-Xavier Vialard. Regularity theory and geometry of unbalanced optimal transport. *Journal of Functional Analysis*, page 111042, 2025.

[HHK+25] Shayan Hundrieser, Florian Heinemann, Marcel Klatt, Marina Struleva, and Axel Munk. Unbalanced kantorovich–rubinstein distance, plan, and barycenter on finite spaces: A statistical perspective. *Journal of Machine Learning Research*, 26:37:1–37:70, 2025.

[HR21] Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166, 2021.

[LMS18] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

[LN89] Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, August 1989.

[MBNWW24] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *The Annals of Statistics*, 52(3):966–998, 2024.

[MOS24] MOSEK ApS. *MOSEK Optimizer API for Python*, 2024. Version 10.1.

[PDNW23] Aram-Alexandre Pooladian, Vincent Divol, and Jonathan Niles-Weed. Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28128–28150, 2023.

[PI25] Donlapark Ponnoprat and Masaaki Imaizumi. Minimax rates of estimation for optimal transport map between infinite-dimensional spaces. *arXiv preprint arXiv:2505.13570*, 2025.

[PLH+20] Khiem Pham, Khang Le, Nhat Ho, Tung Pham, and Hung Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7673–7682, 2020.

[PNW21] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *CoRR*, abs/2109.12004, 2021.

[San15] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Springer, Cham, 2015.

[SHSM26] Marina Struleva, Shayan Hundrieser, Dominic Schuhmacher, and Axel Munk. Sharp convergence rates of empirical unbalanced optimal transport for spatio-temporal point processes. *Stochastic Processes and their Applications*, 198:104938, 2026.

[SPV23] Thibault Séjourné, Gabriel Peyré, and François-Xavier Vialard. Chapter 12 - unbalanced optimal transport, from theory to numerics. In *Handbook of Numerical Analysis*, volume 24, pages 407–471. Elsevier, 2023.

[SQZN23] Yutong Sha, Yuchi Qiu, Peijie Zhou, and Qing Nie. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39, November 2023.

[SS24] Giuseppe Savaré and Giacomo Enrico Sodini. A relaxation viewpoint to unbalanced optimal transport: Duality, optimality and monge formulation. *Journal de Mathématiques Pures et Appliquées*, 188:114–178, 2024.

[SST+19] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019.

[Str99] Gilbert Strang. The discrete cosine transform. *SIAM Review*, 41(1):135–147, 1999.

[SVP22] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pages 4995–5021. PMLR, 2022.

[Tsy09] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, 2009.

[VV22] Adrien Vacher and François-Xavier Vialard. Stability and upper bounds for statistical estimation of unbalanced transport potentials. *arXiv preprint arXiv:2203.09143*, 2022.

[VV23] Adrien Vacher and François-Xavier Vialard. Semi-dual unbalanced quadratic optimal transport: fast statistical rates and convergent algorithm. In *International Conference on Machine Learning*, pages 34734–34758. PMLR, 2023.

[WB19] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

[YKH+18] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *3D Vision (3DV), 2018 International Conference on*, 2018.

[You82] Lai-Sang Young. Dimension, entropy and lyapunov exponents. *Ergodic Theory and Dynamical Systems*, 2(1):109–124, 1982.

[YU19] Karren D. Yang and Caroline Uhler. Scalable unbalanced optimal transport using generative adversarial networks. In *International Conference on Learning Representations*, 2019.

[YWR24] Yuling Yan, Kaizheng Wang, and Philippe Rigollet. Learning Gaussian mixtures using the Wasserstein–Fisher–Rao gradient flow. *The Annals of Statistics*, 52(4):1774 – 1795, 2024.

A Related work

Foundations of unbalanced optimal transport. UOT is largely built on entropy-transport formulations, which relax the marginal constraints by convex divergences and include the logarithmic entropy-transport problem and the Wasserstein-Fisher-Rao distance [CPSV18, LMS18]. Recent work has studied the Monge formulation, the primal-dual optimality conditions, and the regularity theory of these models [SS24, GGV25]. These results provide the structural background for our analysis.

Statistical map estimation in balanced OT. In the balanced setting, OT map estimation is now supported by a fairly rich statistical theory. Minimax-optimal rates for smooth maps were established via semi-dual curvature arguments by [HR21]. On the plug-in side, barycentric-projection estimators based on empirical or smoothed couplings were analyzed by [DGS21], while entropic estimators

were developed by [PNW21] and shown to be particularly effective in semi-discrete and discontinuous settings by [PDNW23]. Sharp smooth plug-in guarantees were obtained by [MBNWW24], and the stability-based perspective has recently been broadened both to more general function classes and to sharper reductions from map estimation to distribution estimation [DNWP25, BM26, PI25]. Our work can be viewed as an unbalanced counterpart to this line of work on stability and plug-in estimation, but with a different estimation target: the transport-growth pair induced by the active marginals of the UOT transport plan.

Statistical theory for UOT. The closest prior works are those of [VV22, VV23], who derive semi-dual formulations for quadratic UOT and global upper and lower bounds on the Bregman divergence of the semi-dual objective, leading to the first fast rates for UOT semi-dual potentials. Our contribution differs in both the estimation target and the technique. These works bound a pseudo-distance between the semi-dual potentials, which controls the L^2 error of the transport map, whereas we study the transport-growth pair (T_0, λ_0) and relate its risk directly to perturbations of the UOT value and to Wasserstein errors of measure estimates. Statistical results for other UOT models remain scarce. In particular, [HHK+25] study unbalanced Kantorovich-Rubinstein distances, transport plans, and barycenters on finite spaces, and [SHSM26] establish near-minimax rates of UOT under the Kantorovich-Rubinstein distance, which is used for comparing observations from spatio-temporal point processes. The statistical rates of entropically regularized UOT estimators are obtained by [APP26]. These works address different costs, observation models, and estimation targets from the estimation of the transport-growth pair considered here.

Algorithms and applications. A large parallel literature develops scalable algorithms and applications for UOT; see [SPV23] for a broad overview. On the computational side, generalized Sinkhorn methods and their complexity or acceleration for unbalanced problems were studied by [PLH+20, SVP22], and lightweight or neural parameterizations have been proposed for large-scale settings [GABK24, YU19, CCK24]. On the application side, UOT and the Wasserstein-Fisher-Rao distance have been used in single-cell trajectory inference and population dynamics [SST+19, SQZN23], tumor-growth and reaction-diffusion models [DC20], gradient-flow learning [YWR24], and geometric shape analysis [BHK22]. These works motivate statistical guarantees for UOT. A nonparametric estimation theory for transport-growth pairs is still missing from this literature.

B Notations

Table 2 lists the recurring symbols of the paper, each with its definition and the place where the symbol is introduced. In the proofs, C denotes a positive constant. The value of C may change from line to line, or even within a line, but it does not depend on n and m.

C Additional constructions and remarks

C.1 Hölder norms

In this subsection, we define the Hölder norms used in this paper. Let $s > 0$ and $k := \lceil s \rceil - 1$, so that $s - k \in (0, 1]$. Let $\mathcal{D}$ be one of the sets $\mathbb{T}_2^d$, $[0,1]^d$, and $\mathbb{R}$, with the distance $\mathrm{dist}_{\mathcal{D}} : \mathcal{D} \times \mathcal{D} \to [0, \infty)$ given by

$$\mathrm{dist}_{\mathbb{T}_2^d}(x, y) := \min_{z \in 2\mathbb{Z}^d} \|x - y - z\|, \qquad \mathrm{dist}_{[0,1]^d}(x, y) := \|x - y\|, \qquad \mathrm{dist}_{\mathbb{R}}(x, y) := |x - y|.$$

For $\flat \in \mathbb{N}_0^\ell$, where $\ell := 1$ if $\mathcal{D} = \mathbb{R}$ and $\ell := d$ otherwise, write $|\flat| := \flat_1 + \cdots + \flat_\ell$ and $\partial^\flat := \partial_1^{\flat_1} \cdots \partial_\ell^{\flat_\ell}$. For a function $f : \mathcal{D} \to \mathbb{R}$ that is k times continuously differentiable, with one-sided derivatives on the boundary of $[0,1]^d$,

$$\|f\|_{C^s(\mathcal{D})} := \sum_{|\flat| \le k} \sup_{x \in \mathcal{D}} |\partial^\flat f(x)| + \sum_{|\flat| = k} \sup_{\substack{x, y \in \mathcal{D} \\ x \neq y}} \frac{|\partial^\flat f(x) - \partial^\flat f(y)|}{\mathrm{dist}_{\mathcal{D}}(x, y)^{s-k}}.$$

Table 2: Recurring symbols.

Symbol	Meaning	Defined in
Ω, $\mathrm{diam}(\Omega)$	domain and its diameter	Sec. 2, Assm. 2
$c(x,y)$	quadratic cost $\frac{1}{2}\|x-y\|^2$	Sec. 2
μ,ν; M_μ, M_ν	source and target measures; their total masses	Sec. 2
$\bar\mu,\bar\nu$; p,q	normalized measures μ/M_μ, ν/M_ν; their normalized densities	Sec. 3.2, Assm. 3
φ^c	c-transform $\inf_y\{\frac{1}{2}\|x-y\|^2-\varphi(y)\}$	Sec. 2
(φ_0,ψ_0)	optimal dual pair, $\psi_0=\varphi_0^c$	(3), Thm. 1
(T_0,λ_0)	$T_0=\mathrm{id}-\nabla\varphi_0$, $\lambda_0=\exp(-\varphi_0+\frac{1}{4}\|\nabla\varphi_0\|^2)$	(4)
f_0	Brenier potential $\frac{1}{2}\|x\|^2-\varphi_0(x)$	after Assm. 1
a_0, w_0	$e^{-\varphi_0/2}=(d\gamma_0/d\mu)^{1/2}$; $w_0=a_0^2=e^{-\varphi_0}$	Secs. 2 and 3.1
γ_0,γ_1	active marginals $e^{-\varphi_0}\mu$ and $e^{-\psi_0}\nu$	Sec. 2, App. E
ζ_0,ξ_0	$1-e^{-\varphi_0}$ and $1-e^{-\psi_0}$	Prop. 8
κ	constant in the Hessian bounds on φ_0	Assm. 1
$\beta_{\min},\beta_{\max}$	lower and upper bounds on p,q	Assm. 3
$\widehat M_\mu,\widehat M_\nu$; a_n,b_m; $\bar c$	estimated masses; bounds on $\mathbb{E}\|\widehat M_\mu-M_\mu\|$, $\mathbb{E}\|\widehat M_\nu-M_\nu\|$; constant in $\widehat M_\mu\ge\bar c M_\mu$, $\widehat M_\nu\ge\bar c M_\nu$	Assm. 4
ρ_μ,ρ_ν	$\widehat M_\mu/M_\mu$ and $\widehat M_\nu/M_\nu$	App. D
α, M	smoothness index, $p,q\in\mathcal{C}_N^{\alpha-1}([0,1]^d;M)$, and radius	Assm. 5
$\|\cdot\|_{C^s(\mathcal{D})}$	Hölder norm on $\mathcal{D}\in\{\mathbb{T}_2^d,[0,1]^d,\mathbb{R}\}$	App. C.1
W_2	2-Wasserstein distance	Sec. 1.3
$w_\pm$	clipping constants, $w_-\le e^{-\varphi_0}\le w_+$	Sec. 3.1
M_φ, M_ψ	$\sup e^{-\varphi_0}$ and $\sup e^{-\psi_0}$	proof of Prop. 8
Λ	$\max(\mathrm{Lip}(\nabla\zeta_0),\mathrm{Lip}(\nabla\xi_0))$	proof of Prop. 8
C_Λ	stability constant $\max(M_\varphi,M_\psi)+\Lambda/2$	Prop. 8
$\mathcal{U}_\alpha(M,B,R)$	class of pairs (μ,ν) in Theorem 5, with $\mu([0,1]^d)=1$; B bounds $\|T_0\|_{C^\alpha}+\|\lambda_0\|_{C^\alpha}$, and $R^{-1}\le\lambda_0\le R$	Sec. 4.5
n,m; X_i,Y_j	sample sizes; samples from $\bar\mu$ and $\bar\nu$	Sec. 2
$\widehat\mu_n,\widehat\nu_m$	empirical measures with masses $\widehat M_\mu,\widehat M_\nu$	Sec. 2
$\bar\mu_n,\bar\nu_m$; $\widetilde\mu_n,\widetilde\nu_m$	normalized empirical measures; empirical measures with masses M_μ, M_ν	Sec. 4.4, App. D
$\widehat\gamma$; $\widehat r_i,\widehat s_j$; $\widehat r_i^\star,\widehat s_j^\star$	discrete optimal transport plan; its row and column masses; $e^{-\varphi_0(X_i)}\widehat\mu_i$, $e^{-\psi_0(Y_j)}\widehat\nu_j$	Secs. 3.1 and 4.4
$\widehat T^{\mathrm{1NN}},\widehat\lambda^{\mathrm{1NN}}$	nearest-neighbor estimators	(6), (7)
M_n, R_n	maximum Voronoi cell mass and radius	Sec. 4.4, Lem. 16
$\widehat T^{\mathrm{ker}}_{nm},\widehat\lambda^{\mathrm{ker}}_{nm}$; L_n; τ	kernel-based estimators; frequency cutoff; function in K_L	Sec. 3.2
$\mathfrak{R}_n^{\mathrm{emp}}(d)$, $\mathfrak{R}_n^{\mathrm{ker}}(\alpha)$	rates	before Thms. 3 and 4
$\mathfrak{L}_n(\alpha)$	rate $n^{-(\alpha-1)/(2(\alpha-1)+d)}$	(81)
$\mathcal{G}_{nm}$	event on which $\widetilde p_n^{\mathrm{ker}}\ge\underline p/2$ and $\widetilde q_m^{\mathrm{ker}}\ge\underline q/2$	App. F

For every other function $f:\mathcal{D}\to\mathbb{R}$, we set $\|f\|_{C^s(\mathcal{D})}:=+\infty$. We write $C^s(\mathcal{D}):=\{f:\mathcal{D}\to\mathbb{R}:\|f\|_{C^s(\mathcal{D})}<\infty\}$. For a map $T=(T_1,\dots,T_d):\mathcal{D}\to\mathbb{R}^d$, we write $\|T\|_{C^s(\mathcal{D})}:=\sum_{i=1}^d\|T_i\|_{C^s(\mathcal{D})}$. For a set $\mathcal{W}\subset\mathbb{R}^d$, $C^s(\mathcal{D};\mathcal{W})$ denotes the set of maps $T:\mathcal{D}\to\mathcal{W}$ with $\|T\|_{C^s(\mathcal{D})}<\infty$.

C.2 Plan-based kernel estimator

In this subsection, we define the *plan-based kernel estimator*, a second way to extend the discrete quantities $(\widehat T_i,\widehat a_i,\widehat\lambda_i)_{i=1}^n$ of Section 3.1 from the sample points to all of Ω. Step (iii) of Section 3.1 performs this extension with the Voronoi weights $v_i(x)=\mathbf{1}\{x\in V_i\}$. The resulting nearest-neighbor estimator (7) is piecewise constant on the cells $V_1,\dots,V_n$. In contrast, the plan-based kernel estimator uses kernel weights and therefore varies continuously with x whenever the kernel is continuous.

Fix a positive kernel $K : \mathbb{R}^d \to (0, \infty)$ and a bandwidth $h > 0$. The *Nadaraya-Watson weights* $v_{i,h} : \Omega \to [0, 1]$ are

$$v_{i,h}(x) := \frac{K\left((x - X_i)/h\right)}{\sum_{k=1}^{n} K\left((x - X_k)/h\right)}, \qquad i = 1, \ldots, n. \tag{16}$$

Like the Voronoi weights $v_i(x)$, these weights are nonnegative and sum to one. But unlike $v_i(x)$, which is positive for one sample point only, $v_{i,h}(x)$ is positive for every sample point X_i. We use these weights to define the plan-based kernel estimator.

Definition 2 (Plan-based kernel estimator)**.** Let $(\widehat{T}_i, \widehat{a}_i, \widehat{\lambda}_i)_{i=1}^n$ be the discrete quantities of Section 3.1, with $\widehat{T}_i$ given by (5) and $\widehat{\lambda}_i$ by (6). Let $v_{i,h}$ be the Nadaraya-Watson weights (16) of a positive kernel K with bandwidth $h > 0$. The plan-based kernel estimator is the triple of functions $\widehat{T}_h^{\mathrm{NW}} : \Omega \to \mathbb{R}^d$, $\widehat{a}_h^{\mathrm{NW}} : \Omega \to [0, \infty)$, and $\widehat{\lambda}_h^{\mathrm{NW}} : \Omega \to (0, \infty)$ given by

$$\widehat{T}_h^{\mathrm{NW}}(x) := \sum_{i=1}^{n} v_{i,h}(x)\widehat{T}_i, \qquad \widehat{a}_h^{\mathrm{NW}}(x) := \sum_{i=1}^{n} v_{i,h}(x)\widehat{a}_i, \qquad \widehat{\lambda}_h^{\mathrm{NW}}(x) := \sum_{i=1}^{n} v_{i,h}(x)\widehat{\lambda}_i.$$

In other words, we use the kernel K to smooth the estimates $\widehat{T}_i$, $\widehat{a}_i$, and $\widehat{\lambda}_i$ of $T_0(X_i)$, $a_0(X_i)$, and $\lambda_0(X_i)$. The three smoothed functions have the form of the nearest-neighbor estimator (7), with the Voronoi weights $v_i(x)$ replaced by $v_{i,h}(x)$. As in Section 3.1, the functions $\widehat{T}_h^{\mathrm{NW}}$, $\widehat{a}_h^{\mathrm{NW}}$, and $\widehat{\lambda}_h^{\mathrm{NW}}$ estimate T_0, a_0, and λ_0, respectively.

For the Gaussian kernel $K(u) = e^{-\|u\|^2/2}$, the plan-based kernel estimator converges to the nearest-neighbor estimator as $h \to 0$ at every x in the interior of some V_i. For this kernel, the weights (16) satisfy

$$\frac{v_{k,h}(x)}{v_{i,h}(x)} = \exp\left(-\frac{\|x - X_k\|^2 - \|x - X_i\|^2}{2h^2}\right), \qquad x \in \Omega,\ i, k \in \{1, \ldots, n\}.$$

Hence, if X_i is the unique sample point nearest to x, then $v_{i,h}(x) \to 1 = v_i(x)$ and $v_{k,h}(x) \to 0 = v_k(x)$ for every $k \in \{1, \ldots, n\} \setminus \{i\}$ as $h \to 0$.

C.3 Obtaining samples through Poisson point processes

For a parameter $t \geq 1$, we consider observations from two independent Poisson point processes,

$$\mathcal{X} = \{X_i\}_{i=1}^{N_\mu} \sim \mathrm{PPP}(t\mu), \qquad \mathcal{Y} = \{Y_j\}_{j=1}^{N_\nu} \sim \mathrm{PPP}(t\nu).$$

Here, the sample sizes satisfy $N_\mu \sim \mathrm{Poisson}(tM_\mu)$ and $N_\nu \sim \mathrm{Poisson}(tM_\nu)$. Conditional on N_μ, X_i independently and identically follows $(M_\mu)^{-1}\mu$. Then, we define the empirical measures

$$\widehat{\mu} := \frac{1}{t}\sum_{i=1}^{N_\mu} \delta_{X_i}, \qquad \widehat{\nu} := \frac{1}{t}\sum_{j=1}^{N_\nu} \delta_{Y_j}.$$

In this setup, we have $\mathbb{E}[\widehat{\mu}] = \mu$. This setup with slight modification is studied by [SHSM26].

The counts N_μ and N_ν are zero with positive probability, so we truncate them to estimate the masses. Let $0 < \underline{M}_\mu \leq M_\mu$ and $0 < \underline{M}_\nu \leq M_\nu$ be known constants, and define

$$\widehat{M}_\mu := \max\{N_\mu/t, \underline{M}_\mu\}, \qquad \widehat{M}_\nu := \max\{N_\nu/t, \underline{M}_\nu\}.$$

Since $\underline{M}_\mu \leq M_\mu$, the truncation does not increase the distance to M_μ. Hence, Jensen's inequality yields

$$\mathbb{E}\big[|\widehat{M}_\mu - M_\mu|\big] \leq \mathbb{E}\big[|N_\mu/t - M_\mu|\big] \leq \sqrt{M_\mu/t},$$

and the analogous bound holds for $\widehat{M}_\nu$. Conditional on $N_\mu = n$ and $N_\nu = m$, the samples are independent with distributions $\bar{\mu}$ and $\bar{\nu}$, as in Section 2, and $\widehat{M}_\mu, \widehat{M}_\nu$ are deterministic with $\widehat{M}_\mu \geq \bar{c}M_\mu$ and $\widehat{M}_\nu \geq \bar{c}M_\nu$, where $\bar{c} = \min\{\underline{M}_\mu/M_\mu, \underline{M}_\nu/M_\nu\}$. The bounds of Theorems 3 and 4 depend on a_n, b_m only through $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \leq a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \leq b_m$. Hence, the bounds of Theorems 3 and 4 hold conditionally on $N_\mu = n$ and $N_\nu = m$ for all sufficiently large n, m, with a_n and b_m replaced by $|\widehat{M}_\mu - M_\mu|$ and $|\widehat{M}_\nu - M_\nu|$.

C.4 Proof of Lemma 2

We first show that $T_0 : \Omega \to \Omega$ is a bijection and that, for every $y \in \Omega$, the infimum defining $\psi_0(y)$ is attained at $T_0^{-1}(y)$.

Lemma 6 (Bijectivity of T_0 and regularity of ψ_0)**.** *Suppose that Assumptions 1, 2, and 3 hold. Then $T_0 = \nabla f_0$ for the Brenier potential $f_0(x) = \frac{1}{2}\|x\|^2 - \varphi_0(x)$, which satisfies $\kappa I \preceq \nabla^2 f_0(x) \preceq \kappa^{-1} I$ on Ω, and $T_0 : \Omega \to \Omega$ is a bijection. Moreover, for every $y \in \Omega$, the infimum in the c-transform $\psi_0(y) = \inf_{x\in\Omega}\{\frac{1}{2}\|x - y\|^2 - \varphi_0(x)\}$ is attained at $x = T_0^{-1}(y)$, and*

$$\psi_0(y) = \frac{1}{2}\|T_0^{-1}(y) - y\|^2 - \varphi_0\big(T_0^{-1}(y)\big), \tag{17}$$

and

$$(1 - \kappa^{-1})I \preceq \nabla^2\psi_0(y) \preceq (1 - \kappa)I, \qquad y \in \Omega. \tag{18}$$

Proof of Lemma 6. By Assumption 1, $\nabla^2 f_0 = I - \nabla^2\varphi_0$ satisfies $\kappa I \preceq \nabla^2 f_0 \preceq \kappa^{-1} I$ on Ω. Since Ω is convex, f_0 is κ-strongly convex on Ω. Strong convexity yields

$$\langle T_0(x) - T_0(x'), x - x'\rangle \geq \kappa\|x - x'\|^2, \qquad x, x' \in \Omega,$$

so $\|T_0(x) - T_0(x')\| \geq \kappa\|x - x'\|$ by the Cauchy-Schwarz inequality; in particular T_0 is injective.

For surjectivity, we combine the primal-dual optimality conditions for (2) and (3) [LMS18] with Theorem 1: the optimal transport plan is $\gamma = (\mathrm{id}, T_0)_\#\gamma_0$ with $\gamma_0 = e^{-\varphi_0}\mu$ and $\gamma_1 = e^{-\psi_0}\nu$. The potential φ_0 is continuous by Assumption 1 and $\psi_0 = \varphi_0^c$ is continuous, so φ_0 and ψ_0 are bounded on the compact set Ω. By Assumption 3, ν has a density bounded below on Ω, so $\gamma_1 = e^{-\psi_0}\nu$ has full support: $\mathrm{supp}(\gamma_1) = \Omega$. On the other hand, $\gamma_1 = T_{0\#}\gamma_0$ is supported in the compact set $T_0(\Omega)$. Hence, $\Omega \subseteq T_0(\Omega)$. The reverse inclusion also holds: the transport plan γ is supported in $\Omega \times \Omega$, so $T_0(x) \in \Omega$ for γ_0-a.e. x, and since γ_0 has full support, T_0 is continuous, and Ω is closed, $T_0(\Omega) \subseteq \Omega$. The map T_0 is therefore onto.

Fix $y \in \Omega$ and set $x' \coloneqq T_0^{-1}(y)$. The function

$$u_y(x) \coloneqq \frac{1}{2}\|x - y\|^2 - \varphi_0(x) = f_0(x) - \langle x, y\rangle + \frac{1}{2}\|y\|^2$$

is κ-strongly convex on Ω with $\nabla u_y(x') = T_0(x') - y = 0$. A critical point of a convex function on a convex set is a global minimizer, so the infimum defining $\psi_0(y) = \varphi_0^c(y)$ is attained at x', which is (17). Since $\nabla^2 f_0 \succeq \kappa I$ is invertible and f_0 is twice continuously differentiable, the inverse function theorem shows that T_0^{-1} is continuously differentiable on Ω with $\nabla T_0^{-1}(y) = (\nabla^2 f_0(x'))^{-1}$. Differentiating $\psi_0(y) = u_y(T_0^{-1}(y))$ and using $\nabla u_y(x') = 0$ yields $\nabla\psi_0(y) = y - T_0^{-1}(y)$, hence

$$\nabla^2\psi_0(y) = I - \nabla T_0^{-1}(y) = I - \big(\nabla^2 f_0(x')\big)^{-1}.$$

Because the eigenvalues of $\nabla^2 f_0(x')$ lie in $[\kappa, \kappa^{-1}]$, those of $(\nabla^2 f_0(x'))^{-1}$ lie in the same interval, which proves (18). □

We now prove Lemma 2, which we restate below:

Lemma 2 (restated)**.** *Suppose that Assumptions 1, 2, and 3 hold. Then the dual potentials (φ_0, ψ_0) satisfy the following uniform lower bound on the gap:*

$$\frac{1}{2}\|x - y\|^2 - \varphi_0(x) - \psi_0(y) \geq \frac{\kappa}{2}\|y - T_0(x)\|^2, \qquad x, y \in \Omega.$$

Proof of Lemma 2. Fix $x, y \in \Omega$ and set $y' \coloneqq T_0(x)$. The function

$$g_x(z) \coloneqq \frac{1}{2}\|x - z\|^2 - \psi_0(z)$$

is κ-strongly convex on Ω with $\nabla g_x(y') = T_0^{-1}(y') - x = 0$, where we used $\nabla^2 g_x = I - \nabla^2\psi_0 \succeq \kappa I$ from the upper bound in (18) and $\nabla\psi_0(z) = z - T_0^{-1}(z)$ from the proof of Lemma 6. Moreover, (17) yields $\psi_0(y') = \frac{1}{2}\|x - y'\|^2 - \varphi_0(x)$, so $g_x(y') = \varphi_0(x)$. Hence,

$$\frac{1}{2}\|x - y\|^2 - \varphi_0(x) - \psi_0(y) = g_x(y) - g_x(y') \geq \frac{\kappa}{2}\|y - T_0(x)\|^2,$$

and this completes the proof. □

D Stability bound

We introduce two variants of the empirical measures $\widehat{\mu}_n$ and $\widehat{\nu}_m$. The normalized empirical measures are defined as

$$\bar{\mu}_n \coloneqq \widehat{\mu}_n/\widehat{M}_\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}, \qquad \bar{\nu}_m \coloneqq \widehat{\nu}_m/\widehat{M}_\nu = \frac{1}{m}\sum_{j=1}^{m}\delta_{Y_j}.$$

We also define the empirical measures with masses M_μ and M_ν as

$$\widetilde{\mu}_n \coloneqq \frac{M_\mu}{n}\sum_{i=1}^{n}\delta_{X_i} = M_\mu\bar{\mu}_n, \qquad \widetilde{\nu}_m \coloneqq \frac{M_\nu}{m}\sum_{j=1}^{m}\delta_{Y_j} = M_\nu\bar{\nu}_m.$$

The following lemma bounds the ranges of φ_0 and ψ_0 in terms of $\mathrm{diam}(\Omega)$ and the mass ratio M_μ/M_ν.

Lemma 7 (Two-sided bounds on the dual potentials)**.** *Let Ω be a compact set, and let (φ_0, ψ_0) be an optimal dual pair for (3) with φ_0 c-concave and $\psi_0 = \varphi_0^c$. Then*

$$\begin{aligned}
\sup_{x\in\Omega}\varphi_0(x) &\le \tfrac{1}{2}\mathrm{diam}(\Omega)^2 + \log(1+M_\mu/M_\nu),\\
\inf_{x\in\Omega}\varphi_0(x) &\ge -\tfrac{1}{2}\mathrm{diam}(\Omega)^2 - \log(1+M_\nu/M_\mu),
\end{aligned} \tag{19}$$

$$\begin{aligned}
\sup_{y\in\Omega}\psi_0(y) &\le \tfrac{1}{2}\mathrm{diam}(\Omega)^2 + \log(1+M_\nu/M_\mu),\\
\inf_{y\in\Omega}\psi_0(y) &\ge -\tfrac{1}{2}\mathrm{diam}(\Omega)^2 - \log(1+M_\mu/M_\nu).
\end{aligned} \tag{20}$$

Proof of Lemma 7. The pair $(0,0)$ is feasible for (3), so the dual value at the optimal pair (φ_0,ψ_0) is nonnegative:

$$\int_\Omega \left(1-e^{-\varphi_0}\right)d\mu + \int_\Omega \left(1-e^{-\psi_0}\right)d\nu \ge 0.$$

The first integral is at most M_μ. Suppose that $\sup_{y\in\Omega}\psi_0(y) < -\log(1+M_\mu/M_\nu)$. Then $1-e^{-\psi_0} < -M_\mu/M_\nu$ on Ω, so the second integral is smaller than $-M_\mu$. Consequently, the dual value is negative. This contradiction shows that

$$\sup_{y\in\Omega}\psi_0(y) \ge -\log(1+M_\mu/M_\nu). \tag{21}$$

Exchanging the roles of (φ_0,μ) and (ψ_0,ν) yields $\sup_{x\in\Omega}\varphi_0(x) \ge -\log(1+M_\nu/M_\mu)$.

For the upper bounds in (19) and (20), the dual constraint yields $\varphi_0(x)+\psi_0(y) \le \frac{1}{2}\|x-y\|^2 \le \frac{1}{2}\mathrm{diam}(\Omega)^2$ for all $x,y\in\Omega$. Taking the supremum over x and y and using (21), we obtain

$$\sup_{x\in\Omega}\varphi_0(x) \le \tfrac{1}{2}\mathrm{diam}(\Omega)^2 - \sup_{y\in\Omega}\psi_0(y) \le \tfrac{1}{2}\mathrm{diam}(\Omega)^2 + \log(1+M_\mu/M_\nu).$$

The upper bound in (20) follows in the same way.

For the lower bounds in (19) and (20), φ_0 is c-concave, so $\varphi_0 = \psi_0^c$. Hence, for every $x\in\Omega$,

$$\varphi_0(x) = \inf_{y\in\Omega}\left\{\tfrac{1}{2}\|x-y\|^2 - \psi_0(y)\right\} \ge -\sup_{y\in\Omega}\psi_0(y) \ge -\tfrac{1}{2}\mathrm{diam}(\Omega)^2 - \log(1+M_\nu/M_\mu),$$

where the last step uses (20). Similarly, $\psi_0 = \varphi_0^c$ yields the lower bound in (20). □

D.1 Stability bound with known masses

We are now ready to state the two-sample stability bound for the UOT problem with $\widetilde{\mu}_n$ and $\widetilde{\nu}_m$, in the case that the masses M_μ and M_ν are known.

Proposition 8 (Two-sample stability of UOT)**.** *Let $\mu, \nu, \widetilde{\mu}_n, \widetilde{\nu}_m \in \mathcal{M}_+(\Omega)$. Consider the UOT model introduced in Section 2. Let φ_0, ψ_0 be the optimal dual potentials for (μ, ν), and define $\zeta_0 := 1 - e^{-\varphi_0}$ and $\xi_0 := 1 - e^{-\psi_0}$. Assume that the problem setting satisfies Assumptions 1, 2, and 3. Then there exists a constant $C_\Lambda > 0$ such that:*

$$\begin{aligned}&\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) - \mathrm{UOT}(\mu, \nu)\\ &\leq \int \zeta_0 \, d(\widetilde{\mu}_n - \mu) + \int \xi_0 \, d(\widetilde{\nu}_m - \nu) + C_\Lambda \left(M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) + M_\nu W_2^2(\bar{\nu}_m, \bar{\nu}) \right). \end{aligned} \tag{22}$$

Moreover, the constant C_Λ can be taken to satisfy

$$C_\Lambda \leq e^{\mathrm{diam}(\Omega)^2/2} \Big(1 + \frac{M_\mu}{M_\nu} + \frac{M_\nu}{M_\mu}\Big)\Big(1 + \frac{\kappa^{-1} + \mathrm{diam}(\Omega)^2}{2}\Big). \tag{23}$$

Proof of Proposition 8. Let γ be the optimal transport plan for (μ, ν). By Assumption 3, the probability measures $\bar{\mu}$ and $\bar{\nu}$ are absolutely continuous. Hence, Brenier's theorem yields optimal transport maps $T_\mu, T_\nu : \Omega \to \Omega$ with $(T_\mu)_\# \bar{\mu} = \bar{\mu}_n$, $(T_\nu)_\# \bar{\nu} = \bar{\nu}_m$, and

$$\int \|T_\mu(x) - x\|^2 \, d\bar{\mu}(x) = W_2^2(\bar{\mu}_n, \bar{\mu}), \qquad \int \|T_\nu(y) - y\|^2 \, d\bar{\nu}(y) = W_2^2(\bar{\nu}_m, \bar{\nu}).$$

Since $\mu = M_\mu \bar{\mu}$ and $\widetilde{\mu}_n = M_\mu \bar{\mu}_n$, we obtain $(T_\mu)_\# \mu = \widetilde{\mu}_n$ and

$$\int \|T_\mu(x) - x\|^2 \, d\mu(x) = M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}).$$

The analogous identities hold for T_ν, ν, and $\widetilde{\nu}_m$. Define the candidate transport plan $\gamma' = (T_\mu, T_\nu)_\# \gamma$. By the data-processing inequality, D_{KL} contracts under deterministic maps:

$$D_{\mathrm{KL}}(\gamma_0' \mid \widetilde{\mu}_n) = D_{\mathrm{KL}}((T_\mu)_\# \gamma_0 \mid (T_\mu)_\# \mu) \leq D_{\mathrm{KL}}(\gamma_0 \mid \mu).$$

Similarly, $D_{\mathrm{KL}}(\gamma_1' \mid \widetilde{\nu}_m) \leq D_{\mathrm{KL}}(\gamma_1 \mid \nu)$. Thus, the UOT objective for the candidate γ' is bounded:

$$\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) \leq \int c \, d\gamma' + D_{\mathrm{KL}}(\gamma_0' \mid \widetilde{\mu}_n) + D_{\mathrm{KL}}(\gamma_1' \mid \widetilde{\nu}_m) \leq \int c \, d\gamma' + D_{\mathrm{KL}}(\gamma_0 \mid \mu) + D_{\mathrm{KL}}(\gamma_1 \mid \nu).$$

Subtracting $\mathrm{UOT}(\mu, \nu) = \int c \, d\gamma + D_{\mathrm{KL}}(\gamma_0 \mid \mu) + D_{\mathrm{KL}}(\gamma_1 \mid \nu)$ from this inequality, we get:

$$\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) - \mathrm{UOT}(\mu, \nu) \leq \int c \, d\gamma' - \int c \, d\gamma.$$

By Lemma 13, we have

$$\begin{aligned}\int c \, d\gamma' - \int c \, d\gamma \leq \int \|T_\mu(x) - x\|^2 \, d\gamma_0(x) + \int \|T_\nu(y) - y\|^2 \, d\gamma_1(y)\\ + \int \langle T_\mu(x) - x, x - y \rangle \, d\gamma(x, y)\\ + \int \langle T_\nu(y) - y, y - x \rangle \, d\gamma(x, y).\end{aligned} \tag{24}$$

We bound the two terms in (24) that involve $T_\mu(x)$. By Lemma 7, φ_0 and ψ_0 are bounded, so

$$\begin{aligned}M_\varphi &:= \sup_{x \in \Omega} e^{-\varphi_0(x)} \leq e^{\mathrm{diam}(\Omega)^2/2}(1 + M_\nu/M_\mu),\\ M_\psi &:= \sup_{y \in \Omega} e^{-\psi_0(y)} \leq e^{\mathrm{diam}(\Omega)^2/2}(1 + M_\mu/M_\nu).\end{aligned} \tag{25}$$

By the primal-dual optimality conditions [LMS18], the marginals of γ are $\gamma_0 = e^{-\varphi_0}\mu$ and $\gamma_1 = e^{-\psi_0}\nu$. Consequently,

$$\begin{aligned}\int \|T_\mu(x) - x\|^2 \, d\gamma_0(x) &= \int \|T_\mu(x) - x\|^2 e^{-\varphi_0(x)} \, d\mu(x)\\ &\leq M_\varphi \int \|T_\mu(x) - x\|^2 \, d\mu(x) = M_\varphi M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}).\end{aligned} \tag{26}$$

For $\int \langle T_\mu(x) - x, x - y\rangle \, d\gamma(x,y)$, we use the first-order optimality relation

$$x - y = \nabla\varphi_0(x) \qquad \text{for } \gamma\text{-a.e. } (x,y),$$

which follows from $c(x,y) = \frac{1}{2}\|x-y\|^2$ and the complementary slackness for the UOT dual problem: for γ-a.e. (x,y), $\varphi_0(x) + \psi_0(y) = c(x,y)$, while $\varphi_0(x') + \psi_0(y) \le c(x',y)$ for all $x' \in \Omega$. Differentiating in x at points where φ_0 is differentiable (which holds μ-a.e., hence γ_0-a.e. since $\gamma_0 \ll \mu$, by Theorem 1 under Assumption 1) yields $\nabla c(x,y) = \nabla\varphi_0(x)$, i.e., $x - y = \nabla\varphi_0(x)$. Therefore,

$$\begin{aligned}
\int \langle T_\mu(x) - x, x - y\rangle \, d\gamma(x,y) &= \int \langle T_\mu(x) - x, \nabla\varphi_0(x)\rangle \, d\gamma_0(x) \\
&= \int \langle T_\mu(x) - x, \nabla\varphi_0(x)\rangle e^{-\varphi_0(x)} \, d\mu(x) \\
&= \int \langle T_\mu(x) - x, \nabla\zeta_0(x)\rangle \, d\mu(x),
\end{aligned}$$

because $\nabla\zeta_0(x) = e^{-\varphi_0(x)}\nabla\varphi_0(x)$. Since both $\nabla\zeta_0$ and $\nabla\xi_0$ are Lipschitz on Ω by Lemma 12, by taking $\Lambda := \max\big(\mathrm{Lip}(\nabla\zeta_0), \mathrm{Lip}(\nabla\xi_0)\big) > 0$, we have the following inequality for all $x, y \in \Omega$:

$$\big|\zeta_0(y) - \zeta_0(x) - \langle \nabla\zeta_0(x), y - x\rangle\big| \le \frac{\Lambda}{2}\|y - x\|^2.$$

Applying this with $y = T_\mu(x)$ yields

$$\langle \nabla\zeta_0(x), T_\mu(x) - x\rangle \le \zeta_0(T_\mu(x)) - \zeta_0(x) + \frac{\Lambda}{2}\|T_\mu(x) - x\|^2.$$

Integrating with respect to μ and using $(T_\mu)_{\#}\mu = \widetilde{\mu}_n$, we obtain

$$\begin{aligned}
\int \langle T_\mu(x) - x, \nabla\zeta_0(x)\rangle \, d\mu(x) &\le \int (\zeta_0(T_\mu(x)) - \zeta_0(x)) \, d\mu(x) + \frac{\Lambda}{2}\int \|T_\mu(x) - x\|^2 \, d\mu(x) \\
&= \int \zeta_0 \, d(\widetilde{\mu}_n - \mu) + \frac{\Lambda}{2} M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}). \qquad (27)
\end{aligned}$$

The terms in (24) that involve $T_\nu(y)$ are handled in the same way, using $y - x = \nabla\psi_0(y)$ for γ-a.e. (x,y) and noting that $\nabla\xi_0$ is also Λ-Lipschitz by our choice of Λ. Combining (26) and (27) with their analogues for $T_\nu(y)$, we may take

$$C_\Lambda := \max(M_\varphi, M_\psi) + \frac{\Lambda}{2}. \qquad (28)$$

The desired inequality (22) then follows from this choice of C_Λ.

It remains to prove (23). To this end, we first bound Λ. Differentiating $\nabla\zeta_0 = e^{-\varphi_0}\nabla\varphi_0$ yields $\nabla^2\zeta_0 = e^{-\varphi_0}\big(\nabla^2\varphi_0 - \nabla\varphi_0\nabla\varphi_0^\top\big)$, where $\|\nabla^2\varphi_0\|_{\mathrm{op}} \le \kappa^{-1}$ by Assumption 1 and $\|\nabla\varphi_0(x)\| = \|x - T_0(x)\| \le \mathrm{diam}(\Omega)$. Moreover, (18) yields $\|\nabla^2\psi_0\|_{\mathrm{op}} \le \kappa^{-1}$, and $\|\nabla\psi_0(y)\| = \|y - T_0^{-1}(y)\| \le \mathrm{diam}(\Omega)$. Hence,

$$\Lambda \le \max(M_\varphi, M_\psi)\big(\kappa^{-1} + \mathrm{diam}(\Omega)^2\big).$$

Substituting this bound into (28) yields $C_\Lambda \le \max(M_\varphi, M_\psi)\big(1 + (\kappa^{-1} + \mathrm{diam}(\Omega)^2)/2\big)$. Moreover, (25) yields $\max(M_\varphi, M_\psi) \le e^{\mathrm{diam}(\Omega)^2/2}(1 + M_\mu/M_\nu + M_\nu/M_\mu)$, which proves (23). □

D.2 Extension to estimated masses

We now consider the empirical measures $\widehat{\mu}_n$ and $\widehat{\nu}_m$, whose total masses are $\widehat{M}_\mu$ and $\widehat{M}_\nu$. We have $\widehat{\mu}_n = \rho_\mu\widetilde{\mu}_n$ and $\widehat{\nu}_m = \rho_\nu\widetilde{\nu}_m$, where $\rho_\mu := \widehat{M}_\mu/M_\mu$ and $\rho_\nu := \widehat{M}_\nu/M_\nu$. By Assumption 4, $\rho_\mu, \rho_\nu \ge \bar{c}$ almost surely, and $\mathbb{E}[|\rho_\mu - 1|] \le a_n/M_\mu$ and $\mathbb{E}[|\rho_\nu - 1|] \le b_m/M_\nu$ for all sufficiently large n, m.

Thus, the following two lemmas yield, almost surely,

$$\begin{aligned}
&\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) - \mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) \\
&\le M_\mu|\rho_\mu - 1| + M_\nu|\rho_\nu - 1| + e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu| + |\log\rho_\nu|\big).
\end{aligned} \qquad (29)$$

Lemma 9 (D_{KL} under measure scaling)**.** *Let $\eta, \sigma \in \mathcal{M}_+(\Omega)$ with $\eta \ll \sigma$, and let $t > 0$. Then*

$$D_{\mathrm{KL}}(\eta \mid t\sigma) = D_{\mathrm{KL}}(\eta \mid \sigma) - \eta(\Omega)\log t + (t-1)\sigma(\Omega).$$

Proof of Lemma 9. Write $\eta = f\sigma$. Then $\eta = (f/t)(t\sigma)$, and therefore

$$\begin{aligned} D_{\mathrm{KL}}(\eta \mid t\sigma) &= \int \left(\frac{f}{t}\log\frac{f}{t} - \frac{f}{t} + 1\right) d(t\sigma) \\ &= \int \left(f\log f - f + 1\right) d\sigma - (\log t)\int f\, d\sigma + (t-1)\sigma(\Omega) \\ &= D_{\mathrm{KL}}(\eta \mid \sigma) - \eta(\Omega)\log t + (t-1)\sigma(\Omega). \end{aligned}$$

□

Lemma 10 (Bound on $\gamma(\Omega \times \Omega)$)**.** *Let $\eta, \sigma \in \mathcal{M}_+(\Omega)$ with $\eta(\Omega) = A > 0$ and $\sigma(\Omega) = B > 0$, and let γ be any optimal transport plan for $\mathrm{UOT}(\eta, \sigma)$ with quadratic cost. Then*

$$\gamma(\Omega \times \Omega) \le e\sqrt{AB}.$$

Proof of Lemma 10. Let $s := \gamma(\Omega\times\Omega) = \gamma_0(\Omega) = \gamma_1(\Omega)$. Since the zero transport plan is feasible, we have

$$\mathrm{UOT}(\eta, \sigma) \le \eta(\Omega) + \sigma(\Omega) = A + B.$$

Because $\int c\, d\gamma \ge 0$, it follows that

$$D_{\mathrm{KL}}(\gamma_0 \mid \eta) + D_{\mathrm{KL}}(\gamma_1 \mid \sigma) \le A + B. \tag{30}$$

Now let $f = d\gamma_0/d\eta$. Since $t \mapsto t\log t - t + 1$ is convex, Jensen's inequality yields

$$\begin{aligned} D_{\mathrm{KL}}(\gamma_0 \mid \eta) &= \int \left(f\log f - f + 1\right) d\eta \\ &\ge A\left(\frac{s}{A}\log\frac{s}{A} - \frac{s}{A} + 1\right) = s\log\frac{s}{A} - s + A. \end{aligned} \tag{31}$$

Similarly,

$$D_{\mathrm{KL}}(\gamma_1 \mid \sigma) \ge s\log\frac{s}{B} - s + B. \tag{32}$$

Summing (31) and (32) and using (30) yields

$$s\log\frac{s^2}{AB} - 2s + A + B \le A + B,$$

that is,

$$s\left(\log\frac{s^2}{AB} - 2\right) \le 0.$$

If $s = 0$ there is nothing to prove. Otherwise, $\log(s^2/(AB)) \le 2$, hence $s \le e\sqrt{AB}$. □

Proposition 11 (Stability of UOT under estimated masses)**.** *Assume that Assumptions 1, 2, 3, and 4 hold. Let C_Λ be the stability constant from Proposition 8. Then, almost surely,*

$$\begin{aligned} \mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) - \mathrm{UOT}(\mu, \nu) \le{}& \int \zeta_0\, d(\widetilde{\mu}_n - \mu) + \int \xi_0\, d(\widetilde{\nu}_m - \nu) \\ &+ C_\Lambda\Big(M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) + M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})\Big) \\ &+ \Big(M_\mu|\rho_\mu - 1| + M_\nu|\rho_\nu - 1| + e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu| + |\log\rho_\nu|\big)\Big). \end{aligned} \tag{33}$$

Consequently, there exists a constant $C_{M_\mu, M_\nu} > 0$ such that, for all sufficiently large n, m, the following bound holds:

$$\begin{aligned} \mathbb{E}\big[\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)\big] - \mathrm{UOT}(\mu, \nu) \le{}& C_\Lambda\Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})]\Big) \\ &+ C_{M_\mu, M_\nu}\big(a_n + b_m\big), \end{aligned} \tag{34}$$

where a_n, b_m are the sequences from Assumption 4.

Proof of Proposition 11. Let $\widetilde{\gamma}_{nm}$ be an optimal transport plan for $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$, and define $s_{nm} \coloneqq \widetilde{\gamma}_{nm}(\Omega \times \Omega)$. Since we have $\widehat{\mu}_n = \rho_\mu \widetilde{\mu}_n$ and $\widehat{\nu}_m = \rho_\nu \widetilde{\nu}_m$, the same transport plan $\widetilde{\gamma}_{nm}$ is admissible for $\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)$. Hence, we obtain

$$\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) \le \int c \, d\widetilde{\gamma}_{nm} + D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_0 \mid \widehat{\mu}_n) + D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_1 \mid \widehat{\nu}_m).$$

Applying Lemma 9 with $\eta = (\widetilde{\gamma}_{nm})_0$ and $\sigma = \widetilde{\mu}_n$, and then with $\eta = (\widetilde{\gamma}_{nm})_1$ and $\sigma = \widetilde{\nu}_m$, yields

$$\begin{aligned} D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_0 \mid \widehat{\mu}_n) &= D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_0 \mid \widetilde{\mu}_n) - s_{nm} \log \rho_\mu + (\rho_\mu - 1) M_\mu, \\ D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_1 \mid \widehat{\nu}_m) &= D_{\mathrm{KL}}((\widetilde{\gamma}_{nm})_1 \mid \widetilde{\nu}_m) - s_{nm} \log \rho_\nu + (\rho_\nu - 1) M_\nu. \end{aligned}$$

Therefore, we obtain

$$\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) - \mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) \le -s_{nm}(\log \rho_\mu + \log \rho_\nu) + (\rho_\mu - 1) M_\mu + (\rho_\nu - 1) M_\nu.$$

By Lemma 10, we also obtain

$$s_{nm} \le e \sqrt{M_\mu M_\nu}.$$

We then obtain (29). Next, Proposition 8 applies to $(\widetilde{\mu}_n, \widetilde{\nu}_m)$ because $\widetilde{\mu}_n(\Omega) = M_\mu$ and $\widetilde{\nu}_m(\Omega) = M_\nu$. Hence

$$\begin{aligned} &\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) - \mathrm{UOT}(\mu, \nu) \\ &\le \int \zeta_0 \, d(\widetilde{\mu}_n - \mu) + \int \xi_0 \, d(\widetilde{\nu}_m - \nu) + C_\Lambda \Big(M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) + M_\nu W_2^2(\bar{\nu}_m, \bar{\nu}) \Big). \end{aligned} \tag{35}$$

Adding (29) and (35) proves (33).

We now derive (34). Taking expectations in (33), the linear terms $\mathbb{E}[\int \zeta_0 \, d(\widetilde{\mu}_n - \mu)]$ and $\mathbb{E}[\int \xi_0 \, d(\widetilde{\nu}_m - \nu)]$ vanish by unbiasedness of $\widetilde{\mu}_n = M_\mu \bar{\mu}_n$ and $\widetilde{\nu}_m = M_\nu \bar{\nu}_m$. By the a.s. lower bound $\widehat{M}_\mu \ge \bar{c} M_\mu$ in Assumption 4, $\rho_\mu \ge \bar{c}$ a.s., so the mean-value theorem applied to $\log$ on $[\bar{c}, \infty)$ yields $|\log \rho_\mu| \le |\rho_\mu - 1| / \bar{c}$ a.s., and analogously $|\log \rho_\nu| \le |\rho_\nu - 1| / \bar{c}$ a.s. Hence, using the identities $M_\mu |\rho_\mu - 1| = |\widehat{M}_\mu - M_\mu|$ and $M_\nu |\rho_\nu - 1| = |\widehat{M}_\nu - M_\nu|$,

$$\begin{aligned} \mathbb{E}\big[M_\mu |\rho_\mu - 1| + e\sqrt{M_\mu M_\nu} \, |\log \rho_\mu| \big] &\le \mathbb{E}[|\widehat{M}_\mu - M_\mu|] + \frac{e\sqrt{M_\mu M_\nu}}{\bar{c} \, M_\mu} \mathbb{E}[|\widehat{M}_\mu - M_\mu|] \\ &\le \left(1 + \bar{c}^{-1} e \sqrt{M_\nu / M_\mu} \right) a_n, \end{aligned} \tag{36}$$

and analogously $\mathbb{E}\big[M_\nu |\rho_\nu - 1| + e\sqrt{M_\mu M_\nu} \, |\log \rho_\nu| \big] \le \left(1 + \bar{c}^{-1} e \sqrt{M_\mu / M_\nu} \right) b_m$. Setting $C_{M_\mu, M_\nu} \coloneqq 1 + \bar{c}^{-1} e \sqrt{M_\mu / M_\nu} + \bar{c}^{-1} e \sqrt{M_\nu / M_\mu}$ and bounding $1 + \bar{c}^{-1} e \sqrt{M_\nu / M_\mu}$ and $1 + \bar{c}^{-1} e \sqrt{M_\mu / M_\nu}$ by C_{M_μ, M_ν} yields (34). □

Remark 2 (Stability bound for general measures)**.** Assume Assumptions 1, 2, and 3. Let $\widetilde{\mu}, \widetilde{\nu} \in \mathcal{M}_+(\Omega)$ with $\widetilde{\mu}(\Omega) = M_\mu$ and $\widetilde{\nu}(\Omega) = M_\nu$, and let $\rho_\mu, \rho_\nu > 0$. Then, the proof of (33) yields

$$\begin{aligned} &\mathrm{UOT}(\rho_\mu \widetilde{\mu}, \rho_\nu \widetilde{\nu}) - \mathrm{UOT}(\mu, \nu) \\ &\le \int \zeta_0 \, d(\widetilde{\mu} - \mu) + \int \xi_0 \, d(\widetilde{\nu} - \nu) \\ &\quad + C_\Lambda \Big(M_\mu W_2^2(\widetilde{\mu}/M_\mu, \bar{\mu}) + M_\nu W_2^2(\widetilde{\nu}/M_\nu, \bar{\nu}) \Big) \\ &\quad + M_\mu |\rho_\mu - 1| + M_\nu |\rho_\nu - 1| + e\sqrt{M_\mu M_\nu} \big(|\log \rho_\mu| + |\log \rho_\nu| \big). \end{aligned}$$

D.3 Supportive results

The proof of the stability bound (22) uses the following two lemmas:

Lemma 12. *Assume Assumptions 1, 2, and 3, and define*

$$\zeta_0(x) \coloneqq 1 - e^{-\varphi_0(x)}, \qquad \xi_0(y) \coloneqq 1 - e^{-\psi_0(y)}.$$

Then $\nabla \zeta_0$ and $\nabla \xi_0$ are Lipschitz on Ω.

Proof of Lemma 12. We first prove this for ζ_0. The gradient is $\nabla\zeta_0(x) = e^{-\varphi_0(x)}\nabla\varphi_0(x)$. By the product rule,

$$\nabla^2\zeta_0(x) = -e^{-\varphi_0(x)}(\nabla\varphi_0(x))(\nabla\varphi_0(x))^\top + e^{-\varphi_0(x)}\nabla^2\varphi_0(x).$$

By Assumptions 1 and 2, φ_0 and $\nabla\varphi_0(x)$ are bounded on the compact set Ω, and $\|\nabla^2\varphi_0(x)\|_{\mathrm{op}} \le \kappa^{-1}$. Thus $\nabla^2\zeta_0$ is bounded on Ω, and $\nabla\zeta_0$ is Lipschitz on the convex set Ω.

For ξ_0, Lemma 6 yields $(1-\kappa^{-1})I \preceq \nabla^2\psi_0 \preceq (1-\kappa)I$ and the boundedness of $\nabla\psi_0(y) = y - T_0^{-1}(y)$ on Ω. The rest of the proof is identical to that for ζ_0. □

Lemma 13. *Let $\gamma \in \mathcal{M}_+(\Omega\times\Omega)$ have marginals γ_0 and γ_1, and let $T_\mu, T_\nu : \Omega \to \Omega$ be measurable maps. Define the candidate transport plan $\gamma' = (T_\mu, T_\nu)_\#\gamma$. Then*

$$\begin{aligned}\int c\,d\gamma' - \int c\,d\gamma \le \int \|T_\mu(x)-x\|^2\,d\gamma_0(x) + \int \|T_\nu(y)-y\|^2\,d\gamma_1(y)\\ + \int \langle T_\mu(x)-x, x-y\rangle\,d\gamma(x,y)\\ + \int \langle T_\nu(y)-y, y-x\rangle\,d\gamma(x,y).\end{aligned} \tag{37}$$

Proof of Lemma 13. By definition of the pushforward,

$$\int c(x,y)\,d\gamma'(x,y) = \frac{1}{2}\int \|T_\mu(x)-T_\nu(y)\|^2\,d\gamma(x,y).$$

For all $x, y \in \Omega$, expanding the square yields

$$\begin{aligned}\tfrac{1}{2}\|T_\mu(x)-T_\nu(y)\|^2 &= \tfrac{1}{2}\|(T_\mu(x)-x)-(T_\nu(y)-y)+(x-y)\|^2\\ &= \tfrac{1}{2}\|x-y\|^2 + \tfrac{1}{2}\|T_\mu(x)-x\|^2 + \tfrac{1}{2}\|T_\nu(y)-y\|^2\\ &\quad + \langle x-y, T_\mu(x)-x\rangle - \langle x-y, T_\nu(y)-y\rangle\\ &\quad - \langle T_\mu(x)-x, T_\nu(y)-y\rangle\\ &\le \tfrac{1}{2}\|x-y\|^2 + \|T_\mu(x)-x\|^2 + \|T_\nu(y)-y\|^2\\ &\quad + \langle T_\mu(x)-x, x-y\rangle + \langle T_\nu(y)-y, y-x\rangle.\end{aligned}$$

Integrating this inequality with respect to $d\gamma(x,y)$ and using the identities

$$\begin{aligned}\int \|T_\mu(x)-x\|^2\,d\gamma(x,y) &= \int \|T_\mu(x)-x\|^2\,d\gamma_0(x),\\ \int \|T_\nu(y)-y\|^2\,d\gamma(x,y) &= \int \|T_\nu(y)-y\|^2\,d\gamma_1(y),\end{aligned}$$

yields (37). □

E Proof of Theorem 3

We first decompose $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$, in the spirit of [MBNWW24, Proposition 13] for the balanced case. In the unbalanced case, the decomposition contains more than the deviations of Y_j from $T_0(X_i)$. The decomposition also measures how the row and column masses of an optimal transport plan for $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$ differ from the *active* marginals γ_0 and γ_1 transported by the transport-growth pair (T_0, λ_0). In Proposition 14, we derive the decomposition for the empirical measures $\widetilde{\mu}_n, \widetilde{\nu}_m$, whose total masses are M_μ, M_ν. We then derive in Proposition 15 this decomposition for the measures $\widehat{\mu}_n, \widehat{\nu}_m$, whose total masses are $\widehat{M}_\mu, \widehat{M}_\nu$.

Let $\widetilde{\gamma} = (\widetilde{\gamma}_{ij})$ be any optimizer of $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$, and write

$$\widetilde{r}_i := \sum_{j=1}^m \widetilde{\gamma}_{ij}, \qquad \widetilde{s}_j := \sum_{i=1}^n \widetilde{\gamma}_{ij}.$$

We also denote the atomic masses of $\widetilde{\mu}_n$ and $\widetilde{\nu}_m$ by $\widetilde{\mu}_i := M_\mu/n$ and $\widetilde{\nu}_j := M_\nu/m$. Finally, recall that for the optimal dual pair (φ_0, ψ_0) of (μ, ν) we have $\gamma_0 = e^{-\varphi_0}\mu$ and $\gamma_1 = e^{-\psi_0}\nu$, and $\zeta_0(x) = 1 - e^{-\varphi_0(x)}$, $\xi_0(y) = 1 - e^{-\psi_0(y)}$.

Proposition 14 (Empirical UOT excess identity). *Define* $r_i^\star := e^{-\varphi_0(X_i)}\widetilde{\mu}_i$ *and* $s_j^\star := e^{-\psi_0(Y_j)}\widetilde{\nu}_j$. *Then, we have*

$$\begin{aligned}&\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) - \int \zeta_0 \, d\widetilde{\mu}_n - \int \xi_0 \, d\widetilde{\nu}_m \\ &\quad = \sum_{i=1}^n \sum_{j=1}^m \widetilde{\gamma}_{ij} \Big(c(X_i, Y_j) - \varphi_0(X_i) - \psi_0(Y_j) \Big) + D_{\mathrm{KL}}(\widetilde{r} \mid r^\star) + D_{\mathrm{KL}}(\widetilde{s} \mid s^\star).\end{aligned} \tag{38}$$

In particular, the right-hand side is nonnegative.

Proof of Proposition 14. Since $\widetilde{\gamma}$ is optimal for $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$, we have

$$\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m) = \sum_{i=1}^n \sum_{j=1}^m c(X_i, Y_j)\widetilde{\gamma}_{ij} + D_{\mathrm{KL}}(\widetilde{r} \mid \widetilde{\mu}_n) + D_{\mathrm{KL}}(\widetilde{s} \mid \widetilde{\nu}_m). \tag{39}$$

By the definition of $r_i^\star = e^{-\varphi_0(X_i)}\widetilde{\mu}_i$, we obtain

$$\begin{aligned}D_{\mathrm{KL}}(\widetilde{r} \mid \widetilde{\mu}_n) &= \sum_{i=1}^n \left[\widetilde{r}_i \log\left(\frac{\widetilde{r}_i}{r_i^\star} \right) - \widetilde{r}_i + r_i^\star \right] + \sum_{i=1}^n \widetilde{r}_i \log\left(\frac{r_i^\star}{\widetilde{\mu}_i} \right) + \sum_{i=1}^n (\widetilde{\mu}_i - r_i^\star) \\ &= D_{\mathrm{KL}}(\widetilde{r} \mid r^\star) - \sum_{i=1}^n \widetilde{r}_i \varphi_0(X_i) + \sum_{i=1}^n \zeta_0(X_i)\widetilde{\mu}_i,\end{aligned} \tag{40}$$

where the second equality follows from

$$\log\left(\frac{r_i^\star}{\widetilde{\mu}_i} \right) = -\varphi_0(X_i), \qquad \left(1 - e^{-\varphi_0(X_i)} \right) = \zeta_0(X_i).$$

Similarly, we obtain

$$D_{\mathrm{KL}}(\widetilde{s} \mid \widetilde{\nu}_m) = D_{\mathrm{KL}}(\widetilde{s} \mid s^\star) - \sum_{j=1}^m \widetilde{s}_j \psi_0(Y_j) + \sum_{j=1}^m \xi_0(Y_j)\widetilde{\nu}_j. \tag{41}$$

Substituting (40) and (41) into (39), and using the relations

$$\sum_{i=1}^n \widetilde{r}_i \varphi_0(X_i) = \sum_{i=1}^n \sum_{j=1}^m \widetilde{\gamma}_{ij} \varphi_0(X_i), \qquad \sum_{j=1}^m \widetilde{s}_j \psi_0(Y_j) = \sum_{i=1}^n \sum_{j=1}^m \widetilde{\gamma}_{ij} \psi_0(Y_j),$$

we obtain (38). Nonnegativity follows because the dual constraint yields

$$c(X_i, Y_j) - \varphi_0(X_i) - \psi_0(Y_j) \geq 0$$

for every i, j, and $D_{\mathrm{KL}}(\widetilde{r} \mid r^\star)$ and $D_{\mathrm{KL}}(\widetilde{s} \mid s^\star)$ are nonnegative. □

E.1 From the discrete optimal transport plan to $\widehat{\Delta}_{nm}^{\mathrm{tr}}$ and $\widehat{\Delta}_{nm}^{\mathrm{bar}}$

We now pass from the discrete optimal transport plan $\widehat{\gamma}$ for $\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)$ to the barycentric projections $\widehat{T}_i$ and the nearest-neighbor estimator $\widehat{T}^{\mathrm{1NN}}$.

Proposition 15 (Bounds on $\widehat{\Delta}_{nm}^{\mathrm{tr}}$ and $\widehat{\Delta}_{nm}^{\mathrm{bar}}$). *Assume that Assumptions 1, 2, 3, and 4 hold. Let* C_Λ *be the stability constant from Proposition 8. Define*

$$\widehat{r}_i^\star := e^{-\varphi_0(X_i)}\widehat{\mu}_i, \qquad \widehat{s}_j^\star := e^{-\psi_0(Y_j)}\widehat{\nu}_j,$$

and

$$\widehat{\Delta}_{nm}^{\mathrm{tr}} := \sum_{i=1}^n \sum_{j=1}^m \widehat{\gamma}_{ij} \|T_0(X_i) - Y_j\|^2,$$

$$\widehat{\Delta}_{nm}^{\mathrm{bar}} := \sum_{i=1}^n \widehat{r}_i \|\widehat{T}_i - T_0(X_i)\|^2.$$

Then, almost surely,

$$\begin{aligned}&\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m\\&\quad=\sum_{i=1}^{n}\sum_{j=1}^{m}\widehat{\gamma}_{ij}\Big(c(X_i,Y_j)-\varphi_0(X_i)-\psi_0(Y_j)\Big)+D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^{\star})+D_{\mathrm{KL}}(\widehat{s}\mid\widehat{s}^{\star}).\end{aligned}\tag{42}$$

In particular,

$$\widehat{\Delta}^{\mathrm{bar}}_{nm}\le\widehat{\Delta}^{\mathrm{tr}}_{nm}\le\frac{2}{\kappa}\left[\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m\right].\tag{43}$$

Moreover, there exists a constant $\widetilde{C}_{\mathrm{mass}}>0$ such that, for all sufficiently large n,m,

$$\mathbb{E}[\widehat{\Delta}^{\mathrm{bar}}_{nm}]\le\frac{2C_\Lambda}{\kappa}\left(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n,\bar{\mu})]+\mathbb{E}[M_\nu W_2^2(\bar{\nu}_m,\bar{\nu})]\right)+\frac{2\widetilde{C}_{\mathrm{mass}}}{\kappa}\big(a_n+b_m\big).\tag{44}$$

Proof of Proposition 15. The identities (39), (40), and (41) hold for arbitrary positive masses $\widetilde{\mu}_i$ and $\widetilde{\nu}_j$. Therefore, replacing $(\widetilde{\mu}_n,\widetilde{\nu}_m,\widetilde{\gamma},\widetilde{r},\widetilde{s},r^\star,s^\star)$ by $(\widehat{\mu}_n,\widehat{\nu}_m,\widehat{\gamma},\widehat{r},\widehat{s},\widehat{r}^\star,\widehat{s}^\star)$ in these identities yields (42).

Lemma 2 and the nonnegativity of the discrete D_{KL} terms in (42) imply

$$\frac{\kappa}{2}\widehat{\Delta}^{\mathrm{tr}}_{nm}\le\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m.$$

Next, fix i with $\widehat{r}_i>0$. By (5) and $\sum_{j=1}^m\widehat{\gamma}_{ij}=\widehat{r}_i$,

$$\widehat{T}_i-T_0(X_i)=\frac{1}{\widehat{r}_i}\sum_{j=1}^{m}\widehat{\gamma}_{ij}\big(Y_j-T_0(X_i)\big).$$

Since $z\mapsto\|z\|^2$ is convex, Jensen's inequality yields

$$\widehat{r}_i\|\widehat{T}_i-T_0(X_i)\|^2\le\sum_{j=1}^{m}\widehat{\gamma}_{ij}\|Y_j-T_0(X_i)\|^2.\tag{45}$$

When $\widehat{r}_i=0$, (45) holds trivially. Summing over i yields

$$\widehat{\Delta}^{\mathrm{bar}}_{nm}\le\widehat{\Delta}^{\mathrm{tr}}_{nm},$$

which proves (43).

Next, by duality and (33),

$$\begin{aligned}&\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m\\&\quad=\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\mathrm{UOT}(\mu,\nu)-\int\zeta_0\,d(\widehat{\mu}_n-\mu)-\int\xi_0\,d(\widehat{\nu}_m-\nu)\\&\quad\le C_\Lambda\Big(M_\mu W_2^2(\bar{\mu}_n,\bar{\mu})+M_\nu W_2^2(\bar{\nu}_m,\bar{\nu})\Big)+\big(|\widehat{M}_\mu-M_\mu|+|\widehat{M}_\nu-M_\nu|\big)\\&\qquad+e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu|+|\log\rho_\nu|\big)\\&\qquad-\int\zeta_0\,d(\widehat{\mu}_n-\widetilde{\mu}_n)-\int\xi_0\,d(\widehat{\nu}_m-\widetilde{\nu}_m),\end{aligned}\tag{46}$$

where we used $\mathrm{UOT}(\mu,\nu)=\int\zeta_0\,d\mu+\int\xi_0\,d\nu$, the identity $\int\zeta_0\,d(\widetilde{\mu}_n-\mu)-\int\zeta_0\,d(\widehat{\mu}_n-\mu)=-\int\zeta_0\,d(\widehat{\mu}_n-\widetilde{\mu}_n)$, and the analogous identity for ξ_0. Since $\widehat{\mu}_n-\widetilde{\mu}_n$ and $\widehat{\nu}_m-\widetilde{\nu}_m$ are signed atomic measures supported on $\{X_1,\dots,X_n\}$ and $\{Y_1,\dots,Y_m\}$, respectively,

$$\left|\int\zeta_0\,d(\widehat{\mu}_n-\widetilde{\mu}_n)\right|\le\|\zeta_0\|_{L^\infty(\Omega)}|\widehat{M}_\mu-M_\mu|,\qquad\left|\int\xi_0\,d(\widehat{\nu}_m-\widetilde{\nu}_m)\right|\le\|\xi_0\|_{L^\infty(\Omega)}|\widehat{M}_\nu-M_\nu|.$$

Substituting these bounds into (46), multiplying by $2/\kappa$ via (43), and absorbing constants, we obtain constants $C_{\mathrm{mass}} > 0$ and $C_{\log} := e\sqrt{M_\mu M_\nu}/\kappa$, depending only on $M_\mu, M_\nu, \kappa, \|\zeta_0\|_{L^\infty(\Omega)}, \|\xi_0\|_{L^\infty(\Omega)}$, such that, almost surely,

$$\widehat{\Delta}^{\mathrm{bar}}_{nm} \le \frac{2C_\Lambda}{\kappa}\left(M_\mu W_2^2(\bar\mu_n,\bar\mu) + M_\nu W_2^2(\bar\nu_m,\bar\nu)\right) + \frac{2C_{\mathrm{mass}}}{\kappa}\left(|\widehat{M}_\mu - M_\mu| + |\widehat{M}_\nu - M_\nu|\right) + 2C_{\log}\big(|\log\rho_\mu| + |\log\rho_\nu|\big). \tag{47}$$

It remains to take expectations of (47). By Assumption 4, $\widehat{M}_\mu \ge \bar c M_\mu$ a.s. yields $|\log\rho_\mu| \le |\rho_\mu - 1|/\bar c$ a.s., hence $M_\mu\mathbb{E}[|\log\rho_\mu|] \le \mathbb{E}[|\widehat{M}_\mu - M_\mu|]/\bar c \le a_n/\bar c$, and analogously for the ν-side, so $e\sqrt{M_\mu M_\nu}\,\mathbb{E}[|\log\rho_\mu|] \le \bar c^{-1} e\sqrt{M_\nu/M_\mu}\,a_n$ and similarly $e\sqrt{M_\mu M_\nu}\mathbb{E}[|\log\rho_\nu|] \le \bar c^{-1} e\sqrt{M_\mu/M_\nu}\,b_m$. Setting $\widetilde{C}_{\mathrm{mass}} := C_{\mathrm{mass}} + \bar c^{-1} e\sqrt{M_\mu/M_\nu} + \bar c^{-1} e\sqrt{M_\nu/M_\mu}$ proves (44). □

E.2 From $\widehat{\Delta}^{\mathrm{bar}}_{nm}$ to the transport map risk under γ_0

To pass from $\widehat{\Delta}^{\mathrm{bar}}_{nm}$ to $\int_\Omega \|\widehat{T}^{1\mathrm{NN}} - T_0\|^2\, d\gamma_0$, we compare the masses $\gamma_0(V_i)$ of the Voronoi cells with $\widehat{r}_i^\star$, keeping the term $D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^\star)$ explicit in the argument.

Lemma 16 ([MBNWW24, Lemma 40])**.** *Let $V_1,\dots,V_n$ be the Voronoi cells induced by $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \bar\mu$, and define*

$$M_n := \max_{1\le i\le n} \bar\mu(V_i), \qquad R_n := \max_{1\le i\le n}\sup_{x\in V_i}\|x - X_i\|.$$

Assume that the density p of $\bar\mu$ satisfies $\beta_{\min} \le p \le \beta_{\max}$ on Ω, and that Ω is compact, convex, and satisfies the interior cone condition (Assumption 2). Then there exist constants $C_1, C_2 > 0$, depending only on d, $\beta_{\min}$, $\beta_{\max}$, ε_0, and δ_0, such that, for every integer $n \ge 2$:

(i) for every $\delta \in (0,1)$,

$$\mathbb{P}\left(M_n \ge \frac{C_1}{n}\big[d\log n + \log(1/\delta)\big]\right) \le \delta;$$

(ii)

$$\mathbb{E}[R_n^2] \le C_2\left((\log n)/n\right)^{2/d}.$$

To compare $\widehat{r}_i$ with $\widehat{r}_i^\star$ through $D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^\star)$, we bound each summand $\widehat{r}_i\log(\widehat{r}_i/\widehat{r}_i^\star) - \widehat{r}_i + \widehat{r}_i^\star$ from below by $(\sqrt{\widehat{r}_i} - \sqrt{\widehat{r}_i^\star})^2$, using the following lemma.

Lemma 17. *For every $a, b \ge 0$, with the conventions $0\log 0 = 0$ and $a\log(a/0) = +\infty$ for $a > 0$, we have*

$$a\log\left(\frac{a}{b}\right) - a + b \ge \left(\sqrt{a} - \sqrt{b}\right)^2. \tag{48}$$

Proof of Lemma 17. If $b = 0$, then (48) is immediate from the stated conventions, so we may assume that $b > 0$. Writing $a = tb$, it suffices to prove that

$$t\log t - t + 1 \ge (\sqrt{t} - 1)^2, \qquad t \ge 0.$$

Set $s = \sqrt{t}$. Then

$$t\log t - t + 1 - (\sqrt{t} - 1)^2 = 2s^2\log s - 2s^2 + 2s = 2s(s\log s - s + 1) \ge 0,$$

because $u \mapsto u\log u - u + 1$ is nonnegative on $[0,\infty)$. □

Lemma 18. *Assume that Assumptions 1, 2, and 3 hold and that $\widehat{M}_\mu, \widehat{M}_\nu > 0$ almost surely. Define*

$$\widehat{b}_i := \|\widehat{T}_i - T_0(X_i)\|^2, \qquad C_\Omega := \mathrm{diam}(\Omega)^2.$$

Then, almost surely,

$$\sum_{i=1}^n \widehat{r}_i^\star \widehat{b}_i \le 3\widehat{\Delta}^{\mathrm{bar}}_{nm} + 2C_\Omega D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^\star). \tag{49}$$

Proof of Lemma 18. Set

$$B_n^\star := \sum_{i=1}^n \widehat{r}_i^\star \widehat{b}_i, \qquad B_n := \sum_{i=1}^n \widehat{r}_i \widehat{b}_i = \widehat{\Delta}_{nm}^{\mathrm{bar}}.$$

By Lemma 17, applied with $a = \widehat{r}_i$ and $b = \widehat{r}_i^\star$ for each i, we have

$$\begin{aligned}\sum_{i=1}^n \frac{(\widehat{r}_i - \widehat{r}_i^\star)^2}{\widehat{r}_i + \widehat{r}_i^\star} &= \sum_{i=1}^n \frac{(\sqrt{\widehat{r}_i} + \sqrt{\widehat{r}_i^\star})^2}{\widehat{r}_i + \widehat{r}_i^\star} \big(\sqrt{\widehat{r}_i} - \sqrt{\widehat{r}_i^\star}\big)^2 \\ &\le 2\sum_{i=1}^n \big(\sqrt{\widehat{r}_i} - \sqrt{\widehat{r}_i^\star}\big)^2 \le 2D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star). \end{aligned} \tag{50}$$

If $\widehat{r}_i = 0$, then $\widehat{T}_i = X_i \in \Omega$, while if $\widehat{r}_i > 0$, then $\widehat{T}_i$ is a convex combination of $Y_1, \dots, Y_m \in \Omega$ and Ω is convex. Since also $T_0(X_i) \in \Omega$, we have $0 \le \widehat{b}_i \le C_\Omega$. Hence, the Cauchy-Schwarz inequality and (50) imply

$$\begin{aligned} B_n^\star - B_n &= \sum_{i=1}^n (\widehat{r}_i^\star - \widehat{r}_i)\widehat{b}_i \\ &\le \left(\sum_{i=1}^n \frac{(\widehat{r}_i - \widehat{r}_i^\star)^2}{\widehat{r}_i + \widehat{r}_i^\star}\right)^{1/2} \left(\sum_{i=1}^n (\widehat{r}_i + \widehat{r}_i^\star)\widehat{b}_i^2\right)^{1/2} \\ &\le \sqrt{2D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)} \left(C_\Omega \sum_{i=1}^n (\widehat{r}_i + \widehat{r}_i^\star)\widehat{b}_i\right)^{1/2} \\ &= \sqrt{2C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)}\sqrt{B_n^\star + B_n}. \end{aligned}$$

Hence,

$$B_n^\star - B_n \le C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star) + \tfrac{1}{2}(B_n^\star + B_n).$$

Rearranging yields $B_n^\star \le 3B_n + 2C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)$, which is (49).

□

Theorem 19 (Transport map risk of $\widehat{T}^{\mathrm{1NN}}$)**.** *Assume that Assumptions 1, 2, and 3 hold and that $\widehat{M}_\mu, \widehat{M}_\nu > 0$ almost surely. Then there exists a constant $C > 0$, depending only on $d, \beta_{\min}, \beta_{\max}, \varepsilon_0, \delta_0, M_\mu, M_\nu, \kappa, \Omega$ and the clipping constants $w_\pm$ of Section 3.1, such that for all sufficiently large n,*

$$\begin{aligned}\mathbb{E}\Big[\int_\Omega \|\widehat{T}^{\mathrm{1NN}}(x) - T_0(x)\|^2 \, d\gamma_0(x)\Big] &\le C \log n \Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})]\Big) \\ &\quad + C\left((\log n)/n\right)^{2/d} + C(\log n)/n. \end{aligned} \tag{51}$$

Proof of Theorem 19. By Assumptions 1 and 2, the map $T_0(x) = x - \nabla\varphi_0(x)$ is Lipschitz with constant $L_T := \kappa^{-1}$. Let $0 < w_- < w_+$ be the clipping constants of Section 3.1, so that

$$w_- \le w_0(x) \le w_+, \qquad x \in \Omega.$$

Define the event

$$\mathcal{A}_{nm} := \left\{ |\widehat{M}_\mu - M_\mu| \le \frac{M_\mu}{2},\ |\widehat{M}_\nu - M_\nu| \le \frac{M_\nu}{2} \right\}.$$

We first show that, on $\mathcal{A}_{nm}$, almost surely,

$$\begin{aligned}\int_\Omega \|\widehat{T}^{\mathrm{1NN}}(x) - T_0(x)\|^2 \, d\gamma_0(x) &\le 4\frac{w_+}{w_-} n M_n \left(3\widehat{\Delta}_{nm}^{\mathrm{bar}} + 2C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)\right) \\ &\quad + 2L_T^2 \gamma_0(\Omega) R_n^2. \end{aligned} \tag{52}$$

For $x \in V_i$, one has $\widehat{T}^{1\mathrm{NN}}(x) = \widehat{T}_i$, hence

$$\|\widehat{T}^{1\mathrm{NN}}(x) - T_0(x)\|^2 \le 2\widehat{b}_i + 2\|T_0(X_i) - T_0(x)\|^2 \le 2\widehat{b}_i + 2L_T^2\|X_i - x\|^2.$$

Integrating over V_i with respect to γ_0 and summing over i yields

$$\int_\Omega \|\widehat{T}^{1\mathrm{NN}}(x) - T_0(x)\|^2 \, d\gamma_0(x) \le 2\sum_{i=1}^n \gamma_0(V_i)\widehat{b}_i + 2L_T^2\gamma_0(\Omega)R_n^2. \tag{53}$$

On $\mathcal{A}_{nm}$, we have $\widehat{M}_\mu \ge M_\mu/2$. Moreover,

$$\gamma_0(V_i) = \int_{V_i} w_0 \, d\mu \le w_+\mu(V_i) = M_\mu w_+ \bar{\mu}(V_i) \le M_\mu w_+ M_n,$$

whereas

$$\widehat{r}_i^\star = w_0(X_i)\widehat{M}_\mu/n \ge w_- \frac{M_\mu}{2n}.$$

Hence, on $\mathcal{A}_{nm}$,

$$\gamma_0(V_i) \le 2\frac{w_+}{w_-} n M_n \widehat{r}_i^\star.$$

Substituting this into (53) yields

$$\int_\Omega \|\widehat{T}^{1\mathrm{NN}}(x) - T_0(x)\|^2 \, d\gamma_0(x) \le 4\frac{w_+}{w_-} n M_n \sum_{i=1}^n \widehat{r}_i^\star \widehat{b}_i + 2L_T^2\gamma_0(\Omega)R_n^2.$$

Applying Lemma 18 proves (52).

For (51), we first show that we may take $\widehat{M}_\mu = M_\mu$ and $\widehat{M}_\nu = M_\nu$. Let $G \in \mathbb{R}_+^{n\times m}$ have marginals $G_0 \ll \widetilde{\mu}_n$ and $G_1 \ll \widetilde{\nu}_m$. For $\eta \ll \sigma$, the definition of D_{KL} yields

$$D_{\mathrm{KL}}(\sqrt{\rho_\mu\rho_\nu}\,\eta \mid \sqrt{\rho_\mu\rho_\nu}\,\sigma) = \sqrt{\rho_\mu\rho_\nu}\, D_{\mathrm{KL}}(\eta \mid \sigma).$$

Hence, Lemma 9 with $t = \sqrt{\rho_\mu/\rho_\nu}$ and with $t = \sqrt{\rho_\nu/\rho_\mu}$ yields

$$\begin{aligned}
D_{\mathrm{KL}}(\sqrt{\rho_\mu\rho_\nu}\,G_0 \mid \widehat{\mu}_n) &= \sqrt{\rho_\mu\rho_\nu}\Big(D_{\mathrm{KL}}(G_0 \mid \widetilde{\mu}_n) - G_0(\Omega)\log\sqrt{\rho_\mu/\rho_\nu} \\
&\qquad + \big(\sqrt{\rho_\mu/\rho_\nu} - 1\big)M_\mu\Big), \\
D_{\mathrm{KL}}(\sqrt{\rho_\mu\rho_\nu}\,G_1 \mid \widehat{\nu}_m) &= \sqrt{\rho_\mu\rho_\nu}\Big(D_{\mathrm{KL}}(G_1 \mid \widetilde{\nu}_m) - G_1(\Omega)\log\sqrt{\rho_\nu/\rho_\mu} \\
&\qquad + \big(\sqrt{\rho_\nu/\rho_\mu} - 1\big)M_\nu\Big).
\end{aligned} \tag{54}$$

Since $G_0(\Omega) = G_1(\Omega)$, the two logarithmic terms in (54) cancel. Thus, the objectives of $\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)$ and $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$ satisfy

$$\begin{aligned}
&\sum_{i=1}^n\sum_{j=1}^m c(X_i, Y_j)\sqrt{\rho_\mu\rho_\nu}\,G_{ij} + D_{\mathrm{KL}}(\sqrt{\rho_\mu\rho_\nu}\,G_0 \mid \widehat{\mu}_n) + D_{\mathrm{KL}}(\sqrt{\rho_\mu\rho_\nu}\,G_1 \mid \widehat{\nu}_m) \\
&\quad = \sqrt{\rho_\mu\rho_\nu}\Big(\sum_{i=1}^n\sum_{j=1}^m c(X_i, Y_j)G_{ij} + D_{\mathrm{KL}}(G_0 \mid \widetilde{\mu}_n) + D_{\mathrm{KL}}(G_1 \mid \widetilde{\nu}_m)\Big) \\
&\qquad + \big(\rho_\mu - \sqrt{\rho_\mu\rho_\nu}\big)M_\mu + \big(\rho_\nu - \sqrt{\rho_\mu\rho_\nu}\big)M_\nu.
\end{aligned}$$

The map $G \mapsto \sqrt{\rho_\mu\rho_\nu}\,G$ is a bijection of $\mathbb{R}_+^{n\times m}$. Therefore, the transport plan $\widehat{\gamma}/\sqrt{\rho_\mu\rho_\nu}$ is optimal for $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$. This transport plan has the same barycentric projections (5) as $\widehat{\gamma}$ and the same indices i with $\widehat{r}_i = 0$. Thus, $\widehat{T}^{1\mathrm{NN}}$ is also the nearest-neighbor estimator of T_0 built from an optimal transport plan for $\mathrm{UOT}(\widetilde{\mu}_n, \widetilde{\nu}_m)$.

It therefore suffices to consider the case $\widehat{M}_\mu = M_\mu$ and $\widehat{M}_\nu = M_\nu$ with an arbitrary optimal transport plan $\widehat{\gamma}$. In this case, the event $\mathcal{A}_{nm}$ always holds. Moreover, Assumption 4 holds with $\bar{c} = 1$ and arbitrary positive sequences $a_n, b_m \to 0$, so Proposition 15 applies. Now set

$$\widehat{Z}_n \coloneqq 3\widehat{\Delta}^{\mathrm{bar}}_{nm} + 2C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star), \qquad \widehat{E}_n \coloneqq \mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m) - \int \zeta_0 \, d\widehat{\mu}_n - \int \xi_0 \, d\widehat{\nu}_m.$$

By (42) and dual feasibility,

$$D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star) \le \widehat{E}_n. \tag{55}$$

By (43),

$$\widehat{\Delta}^{\mathrm{bar}}_{nm} \le \frac{2}{\kappa}\widehat{E}_n. \tag{56}$$

Since $\mathrm{UOT}(\mu,\nu) = \int \zeta_0 \, d\mu + \int \xi_0 \, d\nu$, (22) yields

$$\widehat{E}_n \le C_\Lambda \Big(M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) + M_\nu W_2^2(\bar{\nu}_m, \bar{\nu}) \Big). \tag{57}$$

Combining (55), (56), and (57), we obtain, with $C_3 \coloneqq (6/\kappa + 2C_\Omega)C_\Lambda$,

$$\widehat{Z}_n \le C_3 \Big(M_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) + M_\nu W_2^2(\bar{\nu}_m, \bar{\nu}) \Big) \le C_0 \qquad \text{almost surely,} \tag{58}$$

where $C_0 \coloneqq C_3 C_\Omega (M_\mu + M_\nu)$ and the last inequality uses $W_2^2 \le C_\Omega$ on Ω.

Let C_1, C_2 be the constants of Lemma 16, and set $t_n \coloneqq C_1(d+2)(\log n)/n$. By part (i) of Lemma 16 with $\delta = n^{-2}$, we have $\mathbb{P}(M_n \ge t_n) \le n^{-2}$. On the event $\{M_n < t_n\}$, we have $nM_n\widehat{Z}_n \le nt_n\widehat{Z}_n$; on the event $\{M_n \ge t_n\}$, the bound $nM_n \le n$ and (58) yield $nM_n\widehat{Z}_n \le C_0 n$. Thus, by splitting into these two events, we obtain

$$\mathbb{E}[nM_n\widehat{Z}_n] \le nt_n\mathbb{E}[\widehat{Z}_n] + C_0 n\,\mathbb{P}(M_n \ge t_n) \le C_1(d+2)\log n\,\mathbb{E}[\widehat{Z}_n] + C_0/n. \tag{59}$$

Substituting $\gamma_0(\Omega) \le w_+ M_\mu$, the bound $\mathbb{E}[R_n^2] \le C_2((\log n)/n)^{2/d}$ from Lemma 16(ii), (59), and (58) into (52) proves (51). □

E.3 Active-source and growth estimation for the nearest-neighbor estimator

The row masses of the discrete optimal transport plan $\widehat{\gamma}$ directly encode $w_0 = a_0^2 = e^{-\varphi_0}$. The growth map $\lambda_0 = w_0 \exp(\|x - T_0(x)\|^2/4)$ also depends on the displacement $x - T_0(x)$. Hence, we proceed in two steps: first we estimate a_0 from the row masses $\widehat{r}_i$, and then we bound $\int_\Omega |\widehat{\lambda}^{\mathrm{1NN}} - \lambda_0|^2 \, d\mu$ in terms of the errors of $\widehat{a}^{\mathrm{1NN}}$ and $\widehat{T}^{\mathrm{1NN}}$. Our bound relies on Lemma 17 and on the following pointwise bound.

Lemma 20 (Pointwise bound on $|\widehat{\lambda}_i - \lambda_0(X_i)|^2$)**.** *Let $x \in \Omega$, let $T, T^0 \in \Omega$, and let $w^0 \in [w_-, w_+]$. Define*

$$a^0 = \sqrt{w^0}, \qquad \lambda^0 = w^0 \exp\left(\frac{1}{4}\|x - T^0\|^2\right).$$

For any $u \ge 0$, set

$$\bar{w} = \mathrm{clip}_{[w_-, w_+]}(u), \qquad \bar{\lambda} = \bar{w} \exp\left(\frac{1}{4}\|x - T\|^2\right).$$

Then, with $C_{\mathrm{gr}} \coloneqq e^{\mathrm{diam}(\Omega)^2/2}\big(8w_+ + w_+^2 \mathrm{diam}(\Omega)^2\big)$,

$$|\bar{\lambda} - \lambda^0|^2 \le C_{\mathrm{gr}} \left(|\sqrt{u} - a^0|^2 + \|T - T^0\|^2 \right). \tag{60}$$

Proof of Lemma 20. Since $w^0 \in [w_-, w_+]$, clipping can only move u closer to w^0; hence

$$|\sqrt{\bar{w}} - \sqrt{w^0}| \le |\sqrt{u} - \sqrt{w^0}|.$$

Moreover, $|\bar{w} - w^0| \le 2\sqrt{w_+}|\sqrt{\bar{w}} - \sqrt{w^0}|$. For $x, T \in \Omega$, the map $T \mapsto \exp(\|x - T\|^2/4)$ is bounded by $e^{\mathrm{diam}(\Omega)^2/4}$ and is Lipschitz with constant $\frac{1}{2}\mathrm{diam}(\Omega)e^{\mathrm{diam}(\Omega)^2/4}$. Since $w^0 \le w_+$, we therefore have

$$\begin{aligned}|\bar{\lambda} - \lambda^0| &\le e^{\mathrm{diam}(\Omega)^2/4}|\bar{w} - w^0| + \tfrac{1}{2}w_+\mathrm{diam}(\Omega)e^{\mathrm{diam}(\Omega)^2/4}\|T - T^0\| \\ &\le 2\sqrt{w_+}\, e^{\mathrm{diam}(\Omega)^2/4}|\sqrt{u} - a^0| + \tfrac{1}{2}w_+\mathrm{diam}(\Omega)e^{\mathrm{diam}(\Omega)^2/4}\|T - T^0\|.\end{aligned}$$

Squaring yields

$$\begin{aligned}|\bar{\lambda} - \lambda^0|^2 &\le 8w_+e^{\mathrm{diam}(\Omega)^2/2}|\sqrt{u} - a^0|^2 + \tfrac{1}{2}w_+^2\mathrm{diam}(\Omega)^2e^{\mathrm{diam}(\Omega)^2/2}\|T - T^0\|^2 \\ &\le C_{\mathrm{gr}}\left(|\sqrt{u} - a^0|^2 + \|T - T^0\|^2\right).\end{aligned}$$

This proves (60). □

Lemma 21 (Bound on $\int |\widehat{\lambda} - \lambda_0|^2\, d\widehat{\eta}$)**.** *Let $\widehat{\eta}$ be a finite positive measure on a bounded set Ω, let $\widehat{a}, a_0 : \Omega \to [0, \infty)$, and let $\widehat{T}, T_0 : \Omega \to \Omega$ be measurable. Suppose $a_0^2 = w_0 \in [w_-, w_+]$ and set $\widehat{\gamma}_0 = \widehat{a}^2\widehat{\eta}$. Define*

$$\widehat{\lambda}(x) \coloneqq \mathrm{clip}_{[w_-, w_+]}\big(\widehat{a}(x)^2\big)\exp\left(\frac{1}{4}\|x - \widehat{T}(x)\|^2\right), \qquad \lambda_0(x) \coloneqq w_0(x)\exp\left(\frac{1}{4}\|x - T_0(x)\|^2\right).$$

Then, with $C \coloneqq C_{\mathrm{gr}}\big(1 + 4\max(1, \mathrm{diam}(\Omega)^2)/w_-\big)$ and C_{gr} the constant of Lemma 20,

$$\int_\Omega |\widehat{\lambda}(x) - \lambda_0(x)|^2\, d\widehat{\eta}(x) \le C\int_\Omega |\widehat{a}(x) - a_0(x)|^2\, d\widehat{\eta}(x) + C\int_\Omega \|\widehat{T}(x) - T_0(x)\|^2\, d\widehat{\gamma}_0(x) \quad (61)$$

Proof of Lemma 21. The pointwise argument in Lemma 20 yields

$$|\widehat{\lambda}(x) - \lambda_0(x)|^2 \le C_{\mathrm{gr}}|\widehat{a}(x) - a_0(x)|^2 + C_{\mathrm{gr}}\|\widehat{T}(x) - T_0(x)\|^2. \quad (62)$$

To integrate the last term in (62) with respect to $\widehat{\gamma}_0$ instead of $\widehat{\eta}$, we prove the pointwise inequality

$$\|\widehat{T}(x) - T_0(x)\|^2 \le C'\widehat{a}(x)^2\|\widehat{T}(x) - T_0(x)\|^2 + C'|\widehat{a}(x) - a_0(x)|^2, \qquad x \in \Omega, \quad (63)$$

for $C' \coloneqq \max\Big(\frac{4}{w_-}, \frac{4\,\mathrm{diam}(\Omega)^2}{w_-}\Big)$. Fix $x \in \Omega$ and split into two cases.

Case 1: $\widehat{a}(x)^2 \ge w_-/4$. Then $\widehat{a}(x)^2\|\widehat{T}(x) - T_0(x)\|^2 \ge (w_-/4)\|\widehat{T}(x) - T_0(x)\|^2$, hence $\|\widehat{T}(x) - T_0(x)\|^2 \le (4/w_-)\widehat{a}(x)^2\|\widehat{T}(x) - T_0(x)\|^2$.

Case 2: $\widehat{a}(x)^2 < w_-/4$. Since $a_0(x)^2 = w_0(x) \ge w_-$, $a_0(x) \ge \sqrt{w_-}$, while $\widehat{a}(x) < \sqrt{w_-}/2$, so $|\widehat{a}(x) - a_0(x)| \ge a_0(x) - \widehat{a}(x) \ge \sqrt{w_-}/2$, i.e. $|\widehat{a}(x) - a_0(x)|^2 \ge w_-/4$. Combined with the trivial bound $\|\widehat{T}(x) - T_0(x)\|^2 \le \mathrm{diam}(\Omega)^2$, we obtain $\|\widehat{T} - T_0\|^2 \le (4\,\mathrm{diam}(\Omega)^2/w_-)|\widehat{a} - a_0|^2$.

Thus, (63) holds in either case. Substituting (63) into (62) yields

$$|\widehat{\lambda}(x) - \lambda_0(x)|^2 \le C_{\mathrm{gr}}(1 + C')|\widehat{a}(x) - a_0(x)|^2 + C_{\mathrm{gr}}C'\widehat{a}(x)^2\|\widehat{T}(x) - T_0(x)\|^2. \quad (64)$$

Since $C' = 4\max(1, \mathrm{diam}(\Omega)^2)/w_-$, both coefficients are at most C. Integrating (64) against $\widehat{\eta}$ and using $\widehat{a}(x)^2\, d\widehat{\eta}(x) = d\widehat{\gamma}_0(x)$, which holds by definition of $\widehat{\gamma}_0$, yields (61). □

Lemma 22. *Assume that Assumptions 1, 2, 3, and 4 hold. Let $\widehat{\gamma}$ be any optimizer of $\mathrm{UOT}(\widehat{\mu}_n, \widehat{\nu}_m)$, define*

$$\widehat{r}_i \coloneqq \sum_{j=1}^m \widehat{\gamma}_{ij}, \qquad \widehat{\mu}_i \coloneqq \widehat{M}_\mu/n, \qquad \widehat{r}_i^\star \coloneqq e^{-\varphi_0(X_i)}\widehat{\mu}_i,$$

and write $\widehat{r} = (\widehat{r}_i)_{i=1}^n$ and $\widehat{r}^\star = (\widehat{r}_i^\star)_{i=1}^n$. Then there exists a constant $C > 0$, depending only on $M_\mu, M_\nu, \|\zeta_0\|_{L^\infty(\Omega)}, \|\xi_0\|_{L^\infty(\Omega)}$, the constant $\bar{c}$ of Assumption 4, and C_Λ from Proposition 8, such that for all sufficiently large n, m,

$$\mathbb{E}[D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star)] \le C\Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})] + a_n + b_m\Big), \quad (65)$$

where a_n, b_m are the sequences from Assumption 4.

Proof of Lemma 22. By (42) and the nonnegativity of $\sum_{i=1}^{n}\sum_{j=1}^{m}\widehat{\gamma}_{ij}\big(c(X_i,Y_j)-\varphi_0(X_i)-\psi_0(Y_j)\big)$ and of $D_{\mathrm{KL}}(\widehat{s}\mid\widehat{s}^{\star})$, the inequality

$$D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^{\star})\le \mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m \tag{66}$$

holds. By the stability bound (33),

$$\begin{aligned}&\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\mathrm{UOT}(\mu,\nu)\\&\quad\le\int\zeta_0\,d(\widetilde{\mu}_n-\mu)+\int\xi_0\,d(\widetilde{\nu}_m-\nu)+C_\Lambda\Big(M_\mu W_2^2(\bar{\mu}_n,\bar{\mu})+M_\nu W_2^2(\bar{\nu}_m,\bar{\nu})\Big)\\&\qquad+\big(|\widehat{M}_\mu-M_\mu|+|\widehat{M}_\nu-M_\nu|\big)+e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu|+|\log\rho_\nu|\big),\end{aligned}$$

using $M_\mu|\rho_\mu-1|=|\widehat{M}_\mu-M_\mu|$ and $M_\nu|\rho_\nu-1|=|\widehat{M}_\nu-M_\nu|$. Subtracting $\int\zeta_0\,d(\widehat{\mu}_n-\mu)+\int\xi_0\,d(\widehat{\nu}_m-\nu)$ from both sides and using $\mathrm{UOT}(\mu,\nu)=\int\zeta_0\,d\mu+\int\xi_0\,d\nu$ yields

$$\begin{aligned}\mathrm{UOT}(\widehat{\mu}_n,\widehat{\nu}_m)-\int\zeta_0\,d\widehat{\mu}_n-\int\xi_0\,d\widehat{\nu}_m\le{}&C_\Lambda\Big(M_\mu W_2^2(\bar{\mu}_n,\bar{\mu})+M_\nu W_2^2(\bar{\nu}_m,\bar{\nu})\Big)\\&+\big(|\widehat{M}_\mu-M_\mu|+|\widehat{M}_\nu-M_\nu|\big)\\&+e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu|+|\log\rho_\nu|\big)\\&-\int\zeta_0\,d(\widehat{\mu}_n-\widetilde{\mu}_n)-\int\xi_0\,d(\widehat{\nu}_m-\widetilde{\nu}_m).\end{aligned}\tag{67}$$

The last two integrals are bounded in absolute value by

$$\|\zeta_0\|_{L^\infty(\Omega)}|\widehat{M}_\mu-M_\mu|+\|\xi_0\|_{L^\infty(\Omega)}|\widehat{M}_\nu-M_\nu|. \tag{68}$$

We now take expectations of (67). The a.s. lower bound $\widehat{M}_\mu\ge\bar{c}M_\mu$ in Assumption 4 yields $|\log\rho_\mu|\le|\rho_\mu-1|/\bar{c}$ a.s. Since Assumption 4 also gives $\mathbb{E}[|\widehat{M}_\mu-M_\mu|]\le a_n$, $M_\mu\mathbb{E}[|\log\rho_\mu|]\le a_n/\bar{c}$. Hence, $e\sqrt{M_\mu M_\nu}\,\mathbb{E}[|\log\rho_\mu|]\le\bar{c}^{-1}e\sqrt{M_\nu/M_\mu}\,a_n$. The analogous bounds hold on the ν-side with b_m. Setting $C:=C_\Lambda+\|\zeta_0\|_{L^\infty(\Omega)}+\|\xi_0\|_{L^\infty(\Omega)}+1+\bar{c}^{-1}e\sqrt{M_\nu/M_\mu}+\bar{c}^{-1}e\sqrt{M_\mu/M_\nu}$ and combining (66), (67), and (68) with these expectation bounds proves (65). □

Lemma 23 (Bound on $\sum_i\widehat{\mu}_i|\widehat{a}_i-a_0(X_i)|^2$)**.** *Under the assumptions of Lemma 22, we have*

$$\sum_{i=1}^{n}\widehat{\mu}_i|\widehat{a}_i-a_0(X_i)|^2\le D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^{\star}). \tag{69}$$

Consequently, with a_n,b_m as in Assumption 4 and the constant $C>0$ of Lemma 22, for all sufficiently large n,m,

$$\mathbb{E}\left[\sum_{i=1}^{n}\widehat{\mu}_i|\widehat{a}_i-a_0(X_i)|^2\right]\le C\Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n,\bar{\mu})]+\mathbb{E}[M_\nu W_2^2(\bar{\nu}_m,\bar{\nu})]+a_n+b_m\Big). \tag{70}$$

Proof of Lemma 23. Since $\widehat{r}_i=\widehat{\mu}_i\widehat{a}_i^2$ and $\widehat{r}_i^{\star}=\widehat{\mu}_i a_0(X_i)^2$, we have

$$D_{\mathrm{KL}}(\widehat{r}\mid\widehat{r}^{\star})=\sum_{i=1}^{n}\widehat{\mu}_i\left[\widehat{a}_i^2\log\left(\frac{\widehat{a}_i^2}{a_0(X_i)^2}\right)-\widehat{a}_i^2+a_0(X_i)^2\right].$$

Lemma 17 yields (69) termwise, and (70) follows from Lemma 22. □

Lemma 24 (Bound on $\int|\widehat{\lambda}^{1\mathrm{NN}}-\lambda_0|^2\,d\mu$)**.** *Assume Assumptions 1, 2, and 3. Let V_i,M_n,R_n be as in Lemma 16, and let $\mathcal{A}_{nm}:=\{|\widehat{M}_\mu-M_\mu|\le M_\mu/2,\ |\widehat{M}_\nu-M_\nu|\le M_\nu/2\}$. Then there is a constant $C>0$, depending only on M_μ and $\mathrm{Lip}(\lambda_0)$, such that, on $\mathcal{A}_{nm}$,*

$$\int_\Omega|\widehat{\lambda}^{1\mathrm{NN}}(x)-\lambda_0(x)|^2\,d\mu(x)\le CnM_n\sum_{i=1}^{n}\widehat{\mu}_i|\widehat{\lambda}_i-\lambda_0(X_i)|^2+CR_n^2. \tag{71}$$

Proof of Lemma 24. We first check that $\lambda_0 = w_0 \exp(\|x - T_0(x)\|^2/4)$ is Lipschitz on Ω. The potential φ_0 is c-concave, so φ_0 is an infimum of functions with a common Lipschitz constant on the bounded set Ω. Hence, φ_0 is Lipschitz and bounded, and so is $w_0 = e^{-\varphi_0}$. Moreover, T_0 is κ^{-1}-Lipschitz by Assumption 1, so $x \mapsto \exp(\|x - T_0(x)\|^2/4)$ is Lipschitz and bounded on the compact set Ω. Consequently, the product λ_0 is Lipschitz. We write $L_\lambda := \mathrm{Lip}(\lambda_0)$. For $x \in V_i$,

$$|\widehat{\lambda}_i - \lambda_0(x)|^2 \le 2|\widehat{\lambda}_i - \lambda_0(X_i)|^2 + 2L_\lambda^2 \|x - X_i\|^2. \tag{72}$$

On $\mathcal{A}_{nm}$, we have $\widehat{\mu}_i = \widehat{M}_\mu/n \ge M_\mu/(2n)$. Combining this lower bound with $\mu(V_i) = M_\mu \bar{\mu}(V_i) \le M_\mu M_n$, we obtain $\mu(V_i) \le 2nM_n\widehat{\mu}_i$. Integrating (72) over V_i with respect to μ, we obtain

$$\int_{V_i} |\widehat{\lambda}^{\mathrm{1NN}}(x) - \lambda_0(x)|^2 \, d\mu(x) \le 4nM_n\widehat{\mu}_i |\widehat{\lambda}_i - \lambda_0(X_i)|^2 + 2L_\lambda^2 \int_{V_i} \|x - X_i\|^2 \, d\mu(x).$$

Summing over i and using $\sum_{i=1}^n \int_{V_i} \|x - X_i\|^2 \, d\mu(x) \le M_\mu R_n^2$ yields (71) with $C = \max(4, 2L_\lambda^2 M_\mu)$. □

Theorem 25 (Growth map risk of $\widehat{\lambda}^{\mathrm{1NN}}$)**.** *Assume that Assumptions 1, 2, 3, and 4 hold. Let $\widehat{\lambda}_i$ and $\widehat{\lambda}^{\mathrm{1NN}}$ be defined by (6) and (7). Then there exists a constant $C > 0$ such that for all sufficiently large n, m,*

$$\mathbb{E}\Big[\int_\Omega |\widehat{\lambda}^{\mathrm{1NN}}(x) - \lambda_0(x)|^2 \, d\mu(x)\Big] \le C \log n \Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})] + a_n + b_m\Big) + C\left((\log n)/n\right)^{2/d} + C(\log n)/n, \tag{73}$$

where a_n, b_m are the sequences from Assumption 4.

Proof of Theorem 25. We first prove that, for all sufficiently large n, m,

$$\mathbb{E}\Big[\int_\Omega |\widehat{\lambda}^{\mathrm{1NN}}(x) - \lambda_0(x)|^2 \, d\widehat{\mu}_n(x)\Big] \le C\Big(\mathbb{E}[M_\mu W_2^2(\bar{\mu}_n, \bar{\mu})] + \mathbb{E}[M_\nu W_2^2(\bar{\nu}_m, \bar{\nu})] + a_n + b_m\Big), \tag{74}$$

and then pass to (73) through Lemma 24. Apply Lemma 20 with

$$u_i = \widehat{r}_i/\widehat{\mu}_i, \qquad w_i^0 = w_0(X_i) = e^{-\varphi_0(X_i)}, \qquad T_i = \widehat{T}_i, \qquad T_i^0 = T_0(X_i).$$

This yields

$$\sum_{i=1}^n \widehat{\mu}_i |\widehat{\lambda}_i - \lambda_0(X_i)|^2 \le C_{\mathrm{gr}} \sum_{i=1}^n \widehat{\mu}_i |\widehat{a}_i - a_0(X_i)|^2 + C_{\mathrm{gr}} \sum_{i=1}^n \widehat{\mu}_i \|\widehat{T}_i - T_0(X_i)\|^2. \tag{75}$$

Lemma 23 bounds the first term on the right-hand side of (75). For the second term, since $w_0 \ge w_-$,

$$\sum_{i=1}^n \widehat{\mu}_i \|\widehat{T}_i - T_0(X_i)\|^2 \le w_-^{-1} \sum_{i=1}^n \widehat{r}_i^\star \|\widehat{T}_i - T_0(X_i)\|^2.$$

Lemma 18 gives

$$\sum_{i=1}^n \widehat{r}_i^\star \|\widehat{T}_i - T_0(X_i)\|^2 \le 3\widehat{\Delta}_{nm}^{\mathrm{bar}} + 2C_\Omega D_{\mathrm{KL}}(\widehat{r} \mid \widehat{r}^\star).$$

Since $\widehat{\lambda}^{\mathrm{1NN}}(X_i) = \widehat{\lambda}_i$, the left-hand side of (75) is $\int_\Omega |\widehat{\lambda}^{\mathrm{1NN}} - \lambda_0|^2 \, d\widehat{\mu}_n$. Taking expectations and using Proposition 15 and Lemma 22 proves (74).

We now prove the bound (73) under μ. Let $\mathcal{A}_{nm}$ be the event from Lemma 24, and write

$$S_n := \sum_{i=1}^n \widehat{\mu}_i |\widehat{\lambda}_i - \lambda_0(X_i)|^2.$$

Since $\mathrm{clip}_{[w_-,w_+]}(\widehat{a}_i^2) \le w_+$, $e^{-\varphi_0} \le w_+$, and $\|x - T\| \le \mathrm{diam}(\Omega)$ for all $x, T \in \Omega$, both $\widehat{\lambda}_i$ and λ_0 take values in $[0, w_+ e^{\mathrm{diam}(\Omega)^2/4}]$. Hence, $|\widehat{\lambda}_i - \lambda_0(x)| \le C_\lambda := 2w_+ e^{\mathrm{diam}(\Omega)^2/4}$ for all i and x. Since $\widehat{M}_\mu \le \frac{3}{2} M_\mu$ on $\mathcal{A}_{nm}$, it follows that

$$\int_\Omega |\widehat{\lambda}^{1\mathrm{NN}} - \lambda_0|^2 \, d\mu \le C_\lambda^2 M_\mu \quad \text{almost surely}, \qquad S_n \mathbf{1}_{\mathcal{A}_{nm}} \le C_\lambda^2 \widehat{M}_\mu \mathbf{1}_{\mathcal{A}_{nm}} \le \tfrac{3}{2} C_\lambda^2 M_\mu. \tag{76}$$

The complement $\mathcal{A}_{nm}^c$ is the union of the events $\{|\widehat{M}_\mu - M_\mu| > M_\mu/2\}$ and $\{|\widehat{M}_\nu - M_\nu| > M_\nu/2\}$. Hence, using the first bound in (76), Markov's inequality, the union bound, and Assumption 4, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{1}_{\mathcal{A}_{nm}^c} \int_\Omega |\widehat{\lambda}^{1\mathrm{NN}} - \lambda_0|^2 \, d\mu\Big] &\le C_\lambda^2 M_\mu \mathbb{P}(\mathcal{A}_{nm}^c) \\
&\le C_\lambda^2 M_\mu \Big(\frac{2}{M_\mu} \mathbb{E}[|\widehat{M}_\mu - M_\mu|] + \frac{2}{M_\nu} \mathbb{E}[|\widehat{M}_\nu - M_\nu|] \Big) \\
&\le C(a_n + b_m).
\end{aligned} \tag{77}$$

On $\mathcal{A}_{nm}$, Lemma 24 gives

$$\mathbf{1}_{\mathcal{A}_{nm}} \int_\Omega |\widehat{\lambda}^{1\mathrm{NN}} - \lambda_0|^2 \, d\mu \le C n M_n S_n \mathbf{1}_{\mathcal{A}_{nm}} + C R_n^2. \tag{78}$$

Let C_1, C_2 be the constants of Lemma 16. By part (i) of Lemma 16 with $\delta = n^{-2}$, the threshold $t_n := C_1(d+2)(\log n)/n$ satisfies $\mathbb{P}(M_n \ge t_n) \le n^{-2}$. On the event $\{M_n < t_n\}$, we have $n M_n S_n \mathbf{1}_{\mathcal{A}_{nm}} \le n t_n S_n$; on the event $\{M_n \ge t_n\}$, the bound $n M_n \le n$ and the second bound in (76) yield $n M_n S_n \mathbf{1}_{\mathcal{A}_{nm}} \le \frac{3}{2} C_\lambda^2 M_\mu n$. Thus, by splitting into these two events, we obtain

$$\begin{aligned}
\mathbb{E}[n M_n S_n \mathbf{1}_{\mathcal{A}_{nm}}] &\le C_1(d+2) \log n \, \mathbb{E}[S_n] + \tfrac{3}{2} C_\lambda^2 M_\mu \, n \mathbb{P}(M_n \ge t_n) \\
&\le C_1(d+2) \log n \, \mathbb{E}[S_n] + \tfrac{3}{2} C_\lambda^2 M_\mu / n,
\end{aligned} \tag{79}$$

where $\mathbb{E}[S_n]$ is the left-hand side of (74). Moreover, Lemma 16(ii) gives $\mathbb{E}[R_n^2] \le C_2((\log n)/n)^{2/d}$. Combining (74) and this bound on $\mathbb{E}[R_n^2]$ with (77), (78), and (79) proves (73). □

Proof of Theorem 3. We first reduce the expectations in (12) and (13) to (51) and (73). For $\mathbb{E}[\int_\Omega \|\widehat{T}^{1\mathrm{NN}} - T_0\|^2 \, d\mu]$, we have $d\gamma_0 = w_0 \, d\mu$, and $w_0 = e^{-\varphi_0} \ge w_- > 0$ on Ω by the choice of the clipping constants in Section 3.1. Hence,

$$\int_\Omega \|\widehat{T}^{1\mathrm{NN}} - T_0\|^2 \, d\mu = \int_\Omega \|\widehat{T}^{1\mathrm{NN}} - T_0\|^2 w_0^{-1} \, d\gamma_0 \le w_-^{-1} \int_\Omega \|\widehat{T}^{1\mathrm{NN}} - T_0\|^2 \, d\gamma_0,$$

so Theorem 19 bounds this expectation after absorbing w_-^{-1} into the constant. The expectation $\mathbb{E}[\int_\Omega |\widehat{\lambda}^{1\mathrm{NN}} - \lambda_0|^2 \, d\mu]$ is bounded by (73) in Theorem 25.

It remains to bound $\mathbb{E}[W_2^2(\bar{\mu}_n, \bar{\mu})]$ and $\mathbb{E}[W_2^2(\bar{\nu}_m, \bar{\nu})]$. By Assumptions 2 and 3, the normalized measures $\bar{\mu}$ and $\bar{\nu}$ are supported on a bounded subset of $\mathbb{R}^d$ and have finite moments of every order. Theorem 1 of [FG15], applied with $p = 2$, yields

$$\mathbb{E}[W_2^2(\bar{\mu}_n, \bar{\mu})] \le C_\mu \mathfrak{R}_n^{\mathrm{emp}}(d), \qquad \mathbb{E}[W_2^2(\bar{\nu}_m, \bar{\nu})] \le C_\nu \mathfrak{R}_m^{\mathrm{emp}}(d).$$

Substituting these rates into (51), we obtain a bound of the form

$$C \log n \Big(M_\mu \mathfrak{R}_n^{\mathrm{emp}}(d) + M_\nu \mathfrak{R}_m^{\mathrm{emp}}(d) \Big) + C\big((\log n)/n\big)^{2/d} + C(\log n)/n. \tag{80}$$

Substituting these rates and Assumption 4 into (73), we obtain (80) with the additional term $C \log n \, (a_n + b_m)$. Finally, $C((\log n)/n)^{2/d}$ and $C(\log n)/n$ are absorbed into the first term in (80), since $((\log n)/n)^{2/d} \le \log n \, \mathfrak{R}_n^{\mathrm{emp}}(d)$ and $(\log n)/n \le \log n \, \mathfrak{R}_n^{\mathrm{emp}}(d)$ for all sufficiently large n. This proves (12) and (13). □

Remark 3 (Order of the constants in Theorem 3)**.** Write $D := \mathrm{diam}(\Omega)$ and $\mathfrak{m} := 1 + M_\mu/M_\nu + M_\nu/M_\mu$, and take the clipping constants $w_+ := e^{D^2/2}(1 + M_\nu/M_\mu)$ and $w_- := e^{-D^2/2}(1 + M_\mu/M_\nu)^{-1}$, which satisfy $w_- \le e^{-\varphi_0} \le w_+$ on Ω by Lemma 7. By tracking the constants in the proofs of Theorems 19 and 25, the constant in (12) is $O\big(e^{3D^2/2} \mathfrak{m}^2 (C_\Lambda \kappa^{-1} + \kappa^{-2})\big)$. The constant in (13) is $O\big(e^{2D^2} \mathfrak{m}^3 (C_\Lambda \kappa^{-1} + \kappa^{-2})\big)$.

F Proof of Theorem 4

Here, we additionally define

$$\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm} \coloneqq \mathrm{UOT}(\widehat{\mu}^{\mathrm{ker}}_n, \widehat{\nu}^{\mathrm{ker}}_m).$$

Let $\widehat{p}^{\mathrm{ker}}_n$ and $\widehat{q}^{\mathrm{ker}}_m$ be the densities of the probability measures $\widehat{\mu}^{\mathrm{ker}}_n / \widehat{M}_\mu$ and $\widehat{\nu}^{\mathrm{ker}}_m / \widehat{M}_\nu$, that is,

$$\widehat{p}^{\mathrm{ker}}_n(x) \coloneqq \frac{(\widetilde{p}^{\mathrm{ker}}_n(x))_+}{\int_{[0,1]^d} (\widetilde{p}^{\mathrm{ker}}_n(u))_+ \, du}, \qquad \widehat{q}^{\mathrm{ker}}_m(y) \coloneqq \frac{(\widetilde{q}^{\mathrm{ker}}_m(y))_+}{\int_{[0,1]^d} (\widetilde{q}^{\mathrm{ker}}_m(v))_+ \, dv}.$$

We then define

$$\widetilde{\mu}^{\mathrm{ker}}_n \coloneqq M_\mu \widehat{p}^{\mathrm{ker}}_n(x)\, dx, \qquad \widetilde{\nu}^{\mathrm{ker}}_m \coloneqq M_\nu \widehat{q}^{\mathrm{ker}}_m(y)\, dy.$$

We also define the rate

$$\mathfrak{L}_n(\alpha) \coloneqq n^{-(\alpha-1)/(2(\alpha-1)+d)}. \tag{81}$$

The proof follows the stability argument used for Theorem 3, with the empirical measures $\widehat{\mu}_n, \widehat{\nu}_m$ replaced by the measure estimates $\widehat{\mu}^{\mathrm{ker}}_n, \widehat{\nu}^{\mathrm{ker}}_m$. The overall idea is to decompose $\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm} - \int \zeta_0 \, d\widehat{\mu}^{\mathrm{ker}}_n - \int \xi_0 \, d\widehat{\nu}^{\mathrm{ker}}_m$ into an integral against $\widehat{\gamma}^{\mathrm{ker}}_{nm}$ and two D_{KL} terms relative to $e^{-\varphi_0}\widehat{\mu}^{\mathrm{ker}}_n$ and $e^{-\psi_0}\widehat{\nu}^{\mathrm{ker}}_m$, which is analogous to what we did in the proof of Theorem 3. We then bound the risks of $\widehat{T}^{\mathrm{ker}}_{nm}$ and $\widehat{\lambda}^{\mathrm{ker}}_{nm}$ in terms of $\mathbb{E}[W_2^2(\widetilde{\mu}^{\mathrm{ker}}_n, \mu)]$, $\mathbb{E}[W_2^2(\widetilde{\nu}^{\mathrm{ker}}_m, \nu)]$, and the sequences a_n and b_m of Assumption 4. The upper bound therefore depends on the convergence rate of $\mathbb{E}[W_2^2(\widetilde{\mu}^{\mathrm{ker}}_n, \mu)]$ and $\mathbb{E}[W_2^2(\widetilde{\nu}^{\mathrm{ker}}_m, \nu)]$.

F.1 Convergence rates of the kernel density estimators

For $k \in \mathbb{N}_0^d$, let $\eta_k(x) \coloneqq \prod_{r=1}^d c_{k_r} \cos(\pi k_r x_r)$, with the constants c_ℓ of Section 3.2, and let $\varrho_k \coloneqq \pi^2 \|k\|^2$. The functions $(\eta_k)_{k \in \mathbb{N}_0^d}$ form an orthonormal basis of $L^2([0,1]^d)$, and η_k is the eigenfunction of the Neumann Laplacian $-\Delta$ with eigenvalue ϱ_k. For a zero-mean $h \in L^2([0,1]^d)$, we write

$$\|h\|^2_{H_N^{-1}([0,1]^d)} \coloneqq \sum_{k \in \mathbb{N}_0^d \setminus \{0\}} \frac{|\langle h, \eta_k \rangle_{L^2([0,1]^d)}|^2}{\varrho_k}. \tag{82}$$

Lemma 26 (Bound on W_2 by the H^{-1} norm [San15, Theorem 5.34])**.** *Let f and g be bounded probability densities on $[0,1]^d$ satisfying*

$$0 < \underline{\beta} \le f(x), g(x), \qquad x \in [0,1]^d.$$

Then

$$W_2^2(f\,dx, g\,dx) \le \underline{\beta}^{-1} \|f - g\|^2_{H_N^{-1}([0,1]^d)}. \tag{83}$$

Proof of Lemma 26. Set $h \coloneqq g - f$. Because $\int_{[0,1]^d} h = 0$, there exists a unique weak solution u of the Neumann problem

$$-\Delta u = h, \qquad \partial_{\mathbf{n}} u = 0 \text{ on } \partial[0,1]^d, \qquad \int_{[0,1]^d} u = 0.$$

Writing $u = \sum_{k \neq 0} u_k \eta_k$ yields $u_k = \langle h, \eta_k \rangle / \varrho_k$, and therefore

$$\int_{[0,1]^d} \|\nabla u(x)\|^2 \, dx = \sum_{k \neq 0} \varrho_k |u_k|^2 = \sum_{k \neq 0} \frac{|\langle h, \eta_k \rangle|^2}{\varrho_k} = \|h\|^2_{H_N^{-1}([0,1]^d)}.$$

Now set $f_t = (1-t)f + tg$ for $t \in [0,1]$ and define

$$v_t(x) \coloneqq \frac{\nabla u(x)}{f_t(x)}.$$

Since $f_t \geq \underline{\beta}$, the vector field v_t is well defined. Moreover, $f_t v_t = \nabla u$ and $-\Delta u = g - f$ yield

$$\partial_t f_t + \nabla \cdot (f_t v_t) = (g - f) + \nabla \cdot (\nabla u) = 0,$$

so (f_t, v_t) is an admissible Benamou-Brenier path from $f\,dx$ to $g\,dx$. Hence, the Benamou-Brenier formula [San15, Theorem 5.28] yields

$$\begin{aligned} W_2^2(f\,dx, g\,dx) &\leq \int_0^1 \int_{[0,1]^d} f_t(x) \|v_t(x)\|^2 \,dx\,dt \\ &= \int_0^1 \int_{[0,1]^d} \frac{\|\nabla u(x)\|^2}{f_t(x)} \,dx\,dt \\ &\leq \underline{\beta}^{-1} \int_{[0,1]^d} \|\nabla u(x)\|^2 \,dx = \underline{\beta}^{-1} \|f - g\|^2_{H_N^{-1}([0,1]^d)}, \end{aligned}$$

which proves (83). □

Proposition 27 (Kernel density rates on the hypercube)**.** *Set $s := \alpha - 1 > 0$. Assume that $\Omega = [0,1]^d$, that Assumption 5 holds, and that*

$$\beta_{\min} \leq p(x), q(x) \leq \beta_{\max} \quad \text{for all } x \in [0,1]^d.$$

Let

$$L_n = n^{1/(d+2s)}, \qquad L_m = m^{1/(d+2s)}.$$

Then there exist constants $C > 0$ and $N_0 \in \mathbb{N}$, depending only on $d, \alpha, M, \beta_{\min}, \beta_{\max}$, and the function τ such that, for all $n, m \geq N_0$,

$$\mathbb{E}\big[W_2^2(\widetilde{\mu}_n^{\ker}, \mu)\big] \leq C M_\mu \mathfrak{R}_n^{\ker}(\alpha), \tag{84}$$

$$\mathbb{E}\big[W_2^2(\widetilde{\nu}_m^{\ker}, \nu)\big] \leq C M_\nu \mathfrak{R}_m^{\ker}(\alpha), \tag{85}$$

and

$$\mathbb{E}\big[\|\widehat{p}_n^{\ker} - p\|_{L^1([0,1]^d)}\big] \leq C \mathfrak{L}_n(\alpha), \tag{86}$$

$$\mathbb{E}\big[\|\widehat{q}_m^{\ker} - q\|_{L^1([0,1]^d)}\big] \leq C \mathfrak{L}_m(\alpha). \tag{87}$$

Moreover, let $\underline{p} := \inf_{x \in [0,1]^d} p(x)$ and $\underline{q} := \inf_{y \in [0,1]^d} q(y)$. For every $A > 0$, there exists a constant $C_A > 0$, depending only on A and the quantities above, such that, for all $n, m \geq N_0$,

$$\mathbb{P}\Big(\inf_{x \in [0,1]^d} \widetilde{p}_n^{\ker}(x) < \frac{\underline{p}}{2}\Big) \leq C_A n^{-A}, \qquad \mathbb{P}\Big(\inf_{y \in [0,1]^d} \widetilde{q}_m^{\ker}(y) < \frac{\underline{q}}{2}\Big) \leq C_A m^{-A}. \tag{88}$$

Proof of Proposition 27. We prove (84), (86), and the first bound in (88). The same argument applied to q and $Y_1, \ldots, Y_m$ proves (85), (87), and the second bound in (88). For brevity write

$$\tau_n(k) := \prod_{r=1}^d \tau\left(\frac{\pi^2 k_r^2}{L_n^2}\right), \qquad k \in \mathbb{N}_0^d.$$

Let

$$\theta_k := \int_{[0,1]^d} p(x) \eta_k(x)\,dx, \qquad \widehat{\theta}_k := \frac{1}{n} \sum_{i=1}^n \eta_k(X_i),$$

so that $\theta_0 = 1$ and

$$\widetilde{p}_n^{\ker}(x) = \sum_{k \in \mathbb{N}_0^d} \tau_n(k) \widehat{\theta}_k \eta_k(x). \tag{89}$$

Write

$$p_{L_n}(x) := \mathbb{E}\big[\widetilde{p}_n^{\ker}(x)\big] = \sum_{k \in \mathbb{N}_0^d} \tau_n(k) \theta_k \eta_k(x).$$

Equivalently, p_{L_n} is the restriction to $[0,1]^d$ of $\mathcal{K}_n \mathcal{E} p$, where the multiplier operator $\mathcal{K}_n$ on $C(\mathbb{T}_2^d)$ is $\mathcal{K}_n f := \sum_{k \in \mathbb{Z}^d} \tau_n(|k_1|, \dots, |k_d|) \widehat{f}(k) e^{i\pi k \cdot (\,\cdot\,)}$, with $\widehat{f}(k)$ the kth Fourier coefficient of f. By Assumption 5 and Definition 1, the even reflection $\mathcal{E} p$ belongs to $C^s(\mathbb{T}_2^d)$ with $\|\mathcal{E} p\|_{C^s(\mathbb{T}_2^d)} \le M$.

We bound the bias in the sup norm through the moments of the kernel of $\mathcal{K}_n$. Let $\upsilon : \mathbb{R} \to [0,1]$ be $\upsilon(u) := \tau(u^2)$. Since τ is smooth with $\tau \equiv 1$ on $[0,1]$ and $\tau \equiv 0$ on $[2, \infty)$, we have $\upsilon \in C_c^\infty(\mathbb{R})$ and $\upsilon \equiv 1$ on $[-1,1]$. Define the function $\varpi : \mathbb{R} \to \mathbb{R}$ by

$$\varpi(y) := \frac{1}{2\pi} \int_{\mathbb{R}} \upsilon(u) e^{iuy} \, du.$$

Then ϖ is a Schwartz function, and

$$\int_{\mathbb{R}} \varpi = \upsilon(0) = 1, \qquad \int_{\mathbb{R}} y^j \varpi(y) \, dy = 0 \quad \text{for every integer } j \ge 1.$$

Let $\varpi_d : \mathbb{R}^d \to \mathbb{R}$ be $\varpi_d(y) := \prod_{r=1}^d \varpi(y_r)$. Since $\tau_n(k) = \prod_{r=1}^d \upsilon(\pi k_r / L_n)$, the operator $\mathcal{K}_n$ is the convolution

$$\mathcal{K}_n f(x) = \int_{\mathbb{R}^d} f\big(x - y/L_n\big) \varpi_d(y) \, dy, \qquad x \in \mathbb{R}^d, \tag{90}$$

where we identify $f \in C(\mathbb{T}_2^d)$ with its periodic extension to $\mathbb{R}^d$. Both sides of (90) are continuous in f for the sup norm and agree on the functions $e^{i\pi k \cdot (\,\cdot\,)}$, whose span is dense in $C(\mathbb{T}_2^d)$.

Let $\ell := \lceil s \rceil - 1$. The derivatives of $\mathcal{E} p$ of order ℓ are $(s-\ell)$-Hölder with constant at most M. Thus, Taylor's formula yields, for all $x, y \in \mathbb{R}^d$,

$$\Big| (\mathcal{E} p)\big(x - y/L_n\big) - \sum_{|\flat| \le \ell} \frac{\partial^\flat (\mathcal{E} p)(x)}{\flat!} \Big(-\frac{y}{L_n}\Big)^\flat \Big| \le C M \, \|y\|^s L_n^{-s},$$

where the sum is over multi-indices $\flat \in \mathbb{N}_0^d$. The moments of ϖ_d of positive order vanish and $\int_{\mathbb{R}^d} \varpi_d = 1$. Hence, integrating against ϖ_d in (90) removes the Taylor polynomial except for $(\mathcal{E} p)(x)$. Since $\int_{\mathbb{R}^d} \|y\|^s |\varpi_d(y)| \, dy < \infty$, we obtain

$$\|p_{L_n} - p\|_{L^\infty([0,1]^d)} \le \|\mathcal{E} p - \mathcal{K}_n \mathcal{E} p\|_{L^\infty(\mathbb{T}_2^d)} \le C L_n^{-s}. \tag{91}$$

Since $[0,1]^d$ has volume one, it follows that

$$\|p_{L_n} - p\|_{L^2([0,1]^d)} \le C L_n^{-s}. \tag{92}$$

Moreover, if the kth coefficient of $p_{L_n} - p$ is nonzero, then $\tau_n(k) \ne 1$, so for some coordinate r one has $\pi^2 k_r^2 / L_n^2 \ge 1$ and therefore $\varrho_k \ge \pi^2 k_r^2 \ge L_n^2$. Hence, by (82),

$$\begin{aligned} \|p_{L_n} - p\|^2_{H_N^{-1}([0,1]^d)} &= \sum_{k \ne 0} \frac{|1 - \tau_n(k)|^2 |\theta_k|^2}{\varrho_k} \\ &\le L_n^{-2} \sum_{k \ne 0} |1 - \tau_n(k)|^2 |\theta_k|^2 \\ &= L_n^{-2} \|p_{L_n} - p\|^2_{L^2([0,1]^d)} \le C L_n^{-2(s+1)} = C L_n^{-2\alpha}, \end{aligned} \tag{93}$$

where the last inequality uses (92).

For the stochastic term $\widetilde{p}_n^{\ker} - p_{L_n}$, orthonormality of (η_k) and the bound $p \le \beta_{\max}$ imply

$$\begin{aligned} \mathbb{E}\big[\|\widetilde{p}_n^{\ker} - p_{L_n}\|^2_{L^2([0,1]^d)}\big] &= \sum_{k \ne 0} \tau_n(k)^2 \mathrm{Var}(\widehat{\theta}_k) \\ &\le \frac{C}{n} \# \Big\{ k \in \mathbb{N}_0^d : \tau_n(k) \ne 0 \Big\} \\ &\le C \frac{L_n^d}{n}. \end{aligned} \tag{94}$$

Likewise,

$$\begin{aligned}\mathbb{E}\big[\|\widetilde{p}_n^{\ker} - p_{L_n}\|^2_{H_N^{-1}([0,1]^d)}\big] &= \sum_{k\neq 0} \frac{\tau_n(k)^2}{\varrho_k}\mathrm{Var}(\widehat{\theta}_k) \\ &\le \frac{C}{n} \sum_{\substack{k\in\mathbb{N}_0^d\setminus\{0\} \\ 0\le k_r\lesssim L_n}} \|k\|^{-2} \\ &\le C\begin{cases} n^{-1}, & d=1, \\ (\log L_n)n^{-1}, & d=2, \\ L_n^{d-2}n^{-1}, & d\ge 3. \end{cases}\end{aligned} \tag{95}$$

Combining (92), (93), (94), and (95), and using the Cauchy-Schwarz and Jensen inequalities, we obtain

$$\mathbb{E}\big[\|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)}\big] \le C\left(L_n^{-s} + \sqrt{\frac{L_n^d}{n}}\right), \tag{96}$$

$$\mathbb{E}\big[\|\widetilde{p}_n^{\ker} - p\|^2_{H_N^{-1}([0,1]^d)}\big] \le C\left(L_n^{-2\alpha} + \begin{cases} n^{-1}, & d=1, \\ (\log L_n)n^{-1}, & d=2, \\ L_n^{d-2}n^{-1}, & d\ge 3 \end{cases}\right). \tag{97}$$

We now derive the corresponding L^1 bound for the density estimate $\widehat{p}_n^{\ker}$. Because $\eta_0 \equiv 1$ on $[0,1]^d$, the orthonormality of $(\eta_k)_{k\in\mathbb{N}_0^d}$ yields $\int_{[0,1]^d}\eta_k(x)\,dx = 0$ for every $k \neq 0$. Moreover, $\tau(0) = 1$, so the spectral form (89) and $\widehat{\theta}_0 = \frac{1}{n}\sum_{i=1}^n \eta_0(X_i) = 1$ yield

$$\int_{[0,1]^d} \widetilde{p}_n^{\ker}(x)\,dx = \tau_n(0)\widehat{\theta}_0 = 1.$$

Set $\widetilde{p}_n^{+} := (\widetilde{p}_n^{\ker})_+$ and $Z := \int_{[0,1]^d}\widetilde{p}_n^{+}$. Since $\int_{[0,1]^d}\widetilde{p}_n^{\ker} = 1$, we have

$$Z - 1 = \int_{[0,1]^d} (\widetilde{p}_n^{\ker})_-.$$

Moreover, on the set $\{\widetilde{p}_n^{\ker} < 0\}$ we have $p - \widetilde{p}_n^{\ker} \ge -\widetilde{p}_n^{\ker} = (\widetilde{p}_n^{\ker})_-$, hence

$$Z - 1 = \int_{[0,1]^d} (\widetilde{p}_n^{\ker})_- \le \|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)}.$$

Also,

$$\|\widetilde{p}_n^{+} - \widetilde{p}_n^{\ker}\|_{L^1([0,1]^d)} = \int_{[0,1]^d}(\widetilde{p}_n^{\ker})_- = Z - 1,$$

so by the triangle inequality,

$$\|\widetilde{p}_n^{+} - p\|_{L^1([0,1]^d)} \le \|\widetilde{p}_n^{+} - \widetilde{p}_n^{\ker}\|_{L^1([0,1]^d)} + \|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)} \le 2\|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)}. \tag{98}$$

Finally,

$$\left\|\frac{\widetilde{p}_n^{+}}{Z} - \widetilde{p}_n^{+}\right\|_{L^1([0,1]^d)} = |1 - Z| = Z - 1 \le \|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)},$$

which, combined with (98), implies

$$\|\widehat{p}_n^{\ker} - p\|_{L^1([0,1]^d)} \le 3\|\widetilde{p}_n^{\ker} - p\|_{L^1([0,1]^d)}. \tag{99}$$

Taking expectations and combining with (96) yields

$$\mathbb{E}\big[\|\widehat{p}_n^{\ker} - p\|_{L^1([0,1]^d)}\big] \le C\left(L_n^{-s} + \sqrt{\frac{L_n^d}{n}}\right). \tag{100}$$

It remains to prove (84). We have $\underline{p} \geq \beta_{\min} > 0$. By (91), there exists an integer $N_0 \geq 3$, depending only on $d, \alpha, M, \beta_{\min}, \beta_{\max}$, and the function τ, such that $\|p_{L_n} - p\|_{L^\infty([0,1]^d)} \leq \beta_{\min}/4$ for all $n \geq N_0$. Moreover, the product form (9) implies

$$\sup_{x\in[0,1]^d} K_{L_n}(x,x) = \sup_{x\in[0,1]^d} \prod_{r=1}^{d} K_{L_n}^{(1)}(x_r,x_r) \leq CL_n^d.$$

Define

$$Pf \coloneqq \int f(z)p(z)\,dz, \qquad P_n f \coloneqq \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Then

$$\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} = \sup_{x\in[0,1]^d} |(P_n - P)K_{L_n}(x,\cdot)|.$$

For every $x, z \in [0,1]^d$, the Cauchy-Schwarz inequality in the cosine expansion yields

$$|K_{L_n}(x,z)| \leq K_{L_n}(x,x)^{1/2} K_{L_n}(z,z)^{1/2} \leq CL_n^d,$$

so the class $\mathcal{F}_n \coloneqq \{K_{L_n}(x,\cdot) : x \in [0,1]^d\}$ has envelope $U_n \leq CL_n^d$. Likewise, using $p \leq \beta_{\max}$, orthonormality of (η_k), and $0 \leq \tau_n(k) \leq 1$,

$$\begin{aligned}
\operatorname{Var}(K_{L_n}(x,X_1)) &\leq \mathbb{E}[K_{L_n}(x,X_1)^2] \\
&= \int K_{L_n}(x,z)^2 p(z)\,dz \\
&\leq \beta_{\max} \int K_{L_n}(x,z)^2\,dz \\
&= \beta_{\max} \sum_{k\in\mathbb{N}_0^d} \tau_n(k)^2 \eta_k(x)^2 \leq \beta_{\max} K_{L_n}(x,x) \leq CL_n^d.
\end{aligned}$$

Hence, $\sigma_n^2 \coloneqq \sup_x \operatorname{Var}(K_{L_n}(x,X_1)) \leq CL_n^d$. We next pass from a fixed x to the supremum through a finite grid. Since $|K_{L_n}^{(1)}(u,v)| \leq CL_n$ and $|\partial_u K_{L_n}^{(1)}(u,v)| \leq CL_n^2$ for $u, v \in [0,1]$, the product form (9) yields

$$\|K_{L_n}(x,\cdot) - K_{L_n}(x',\cdot)\|_{L^\infty([0,1]^d)} \leq CL_n^{d+1}\|x - x'\|, \qquad x, x' \in [0,1]^d.$$

Hence, there exists a finite set $\mathcal{X}_n \subset [0,1]^d$ with $|\mathcal{X}_n| \leq CL_n^{d(d+1)}$ such that

$$\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} = \sup_{x\in[0,1]^d} |(P_n - P)K_{L_n}(x,\cdot)| \leq \max_{x\in\mathcal{X}_n} |(P_n - P)K_{L_n}(x,\cdot)| + \frac{\beta_{\min}}{8}.$$

Consequently, the event $\{\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} > \beta_{\min}/4\}$ is contained in the union of the events $\{|(P_n - P)K_{L_n}(x,\cdot)| > \beta_{\min}/8\}$ over $x \in \mathcal{X}_n$. For each fixed x, the variables $K_{L_n}(x,X_i) - PK_{L_n}(x,\cdot)$ are i.i.d., centered, bounded by $2U_n$, and have variance at most σ_n^2. Therefore, Bernstein's inequality and the union bound over $\mathcal{X}_n$ yield, for a constant $c > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} > \frac{\beta_{\min}}{4}\Big) &\leq 2|\mathcal{X}_n| \exp\Big(-\frac{n(\beta_{\min}/8)^2}{2\sigma_n^2 + (4/3)U_n(\beta_{\min}/8)}\Big) \\
&\leq CL_n^{d(d+1)} \exp\Big(-\frac{cn}{L_n^d}\Big).
\end{aligned}$$

Since $L_n = n^{1/(d+2s)}$, we have $n/L_n^d = n^{2s/(d+2s)}$ with $s > 0$. Thus, for every $A > 0$, there exists a constant $C_A > 0$ such that

$$\mathbb{P}\Big(\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} > \frac{\beta_{\min}}{4}\Big) \leq C_A n^{-A}. \tag{101}$$

For $n \geq N_0$, we have $\|p_{L_n} - p\|_{L^\infty([0,1]^d)} \leq \beta_{\min}/4$, so the event $\{\|\widetilde{p}_n^{\ker} - p\|_{L^\infty([0,1]^d)} > \beta_{\min}/2\}$ is contained in the event $\{\|\widetilde{p}_n^{\ker} - p_{L_n}\|_{L^\infty([0,1]^d)} > \beta_{\min}/4\}$. Therefore, by (101), for all $n \geq N_0$, the event

$$\mathcal{G}_n \coloneqq \Big\{\|\widetilde{p}_n^{\ker} - p\|_{L^\infty([0,1]^d)} \leq \frac{\beta_{\min}}{2}\Big\}$$

satisfies

$$\mathbb{P}(\mathcal{G}_n^c) \le C_A n^{-A}. \tag{102}$$

Since $\beta_{\min} \le \underline{p}$, we have $\widetilde{p}_n^{\ker} \ge \underline{p} - \beta_{\min}/2 \ge \underline{p}/2$ on $\mathcal{G}_n$. This proves the first bound in (88). On $\mathcal{G}_n$ we have $\widetilde{p}_n^{\ker} \ge \underline{p}/2 > 0$, so $(\widetilde{p}_n^{\ker})_+ = \widetilde{p}_n^{\ker}$. Since $\int_{[0,1]^d} \widetilde{p}_n^{\ker} = 1$, it follows that

$$\widehat{p}_n^{\ker} = \widetilde{p}_n^{\ker}.$$

Since $p \ge \underline{p}$ and $\widetilde{p}_n^{\ker} \ge \underline{p}/2$ on $\mathcal{G}_n$, Lemma 26 yields

$$W_2^2\big(\widehat{p}_n^{\ker}\,dx, p\,dx\big) = W_2^2\big(\widetilde{p}_n^{\ker}\,dx, p\,dx\big) \le \frac{2}{\underline{p}} \|\widetilde{p}_n^{\ker} - p\|^2_{H_N^{-1}([0,1]^d)} \qquad \text{on } \mathcal{G}_n.$$

On $\mathcal{G}_n^c$, both $\widehat{p}_n^{\ker}\,dx$ and $p\,dx$ are supported in $[0,1]^d$, so

$$W_2^2\big(\widehat{p}_n^{\ker}\,dx, p\,dx\big) \le d.$$

Taking expectations and using (97) and (102) with $A = 2$ therefore yields

$$\mathbb{E}\big[W_2^2(\widehat{p}_n^{\ker}\,dx, p\,dx)\big] \le C\left(L_n^{-2\alpha} + \begin{cases} n^{-1}, & d = 1, \\ (\log L_n) n^{-1}, & d = 2, \\ L_n^{d-2} n^{-1}, & d \ge 3 \end{cases}\right). \tag{103}$$

Since $L_n = n^{1/(d+2s)}$, (100) yields

$$\mathbb{E}\big[\|\widehat{p}_n^{\ker} - p\|_{L^1([0,1]^d)}\big] \le C n^{-s/(2s+d)} = C\mathfrak{L}_n(\alpha),$$

while (103) becomes

$$\mathbb{E}\big[W_2^2(\widehat{p}_n^{\ker}\,dx, p\,dx)\big] \le C\mathfrak{R}_n^{\ker}(\alpha).$$

Finally,

$$W_2^2(\widetilde{\mu}_n^{\ker}, \mu) = M_\mu W_2^2(\widehat{p}_n^{\ker}\,dx, p\,dx),$$

which proves (84) and (86). The same argument applied to q and $Y_1, \dots, Y_m$ proves (85), (87), and the second bound in (88). □

F.2 Proof of the upper bound

We are now in a position to prove (14) and (15).

Corollary 28 (Risk of the kernel-based estimator)**.** *Assume that $\Omega = [0,1]^d$, $c(x,y) = \frac{1}{2}\|x-y\|^2$, that Assumptions 1, 3, and 4 hold, and that the hypotheses of Proposition 27 hold. Let*

$$\widehat{\mu}_n^{\ker},\ \widehat{\nu}_m^{\ker},\ \widehat{T}_{nm}^{\ker},\ \widehat{\lambda}_{nm}^{\ker},\ \widehat{\mathrm{UOT}}_{nm}^{\ker}$$

be as in Section 3.2 and Appendix F, and let a_n, b_m be the sequences from Assumption 4. Suppose that, for some deterministic sequences $r_n^\mu, r_m^\nu, \ell_n, \ell_m \ge 0$, where $r_n^\mu \ge M_\mu/n$ and $r_m^\nu \ge M_\nu/m$,

$$\mathbb{E}\big[W_2^2(\widetilde{\mu}_n^{\ker}, \mu)\big] \le r_n^\mu, \qquad \mathbb{E}\big[W_2^2(\widetilde{\nu}_m^{\ker}, \nu)\big] \le r_m^\nu, \tag{104}$$

$$\mathbb{E}\big[\|\widehat{p}_n^{\ker} - p\|_{L^1([0,1]^d)}\big] \le \ell_n, \qquad \mathbb{E}\big[\|\widehat{q}_m^{\ker} - q\|_{L^1([0,1]^d)}\big] \le \ell_m.$$

Then there exist constants $C > 0$ and $N_0 \in \mathbb{N}$, depending only on d, α, M, $\beta_{\min}$, $\beta_{\max}$, $w_\pm$, the function τ, and the constants κ and $\bar{c}$ of Assumptions 1 and 4, such that, for all $n, m \ge N_0$,

$$\mathbb{E}\left[\int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2\, d\mu(x)\right] \le C(r_n^\mu + r_m^\nu). \tag{105}$$

If, in addition, n and m are large enough that $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \le a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \le b_m$, then

$$\mathbb{E}\left[\int_{[0,1]^d} |\widehat{\lambda}_{nm}^{\ker}(x) - \lambda_0(x)|^2\, d\mu(x)\right] \le C(r_n^\mu + r_m^\nu + a_n + b_m), \tag{106}$$

and

$$\begin{aligned}\mathbb{E}\left[\left|\widehat{\mathrm{UOT}}_{nm}^{\ker} - \mathrm{UOT}(\mu,\nu)\right|\right] \le{}& M_\mu \|\zeta_0\|_{L^\infty([0,1]^d)} \ell_n + M_\nu \|\xi_0\|_{L^\infty([0,1]^d)} \ell_m \\ &+ C(r_n^\mu + r_m^\nu + a_n + b_m).\end{aligned} \tag{107}$$

Proof of Corollary 28. Let $\widehat{\gamma}^{\mathrm{ker}}_{nm}$ be an optimal transport plan for $\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm}$, with marginals $\widehat{\gamma}^{\mathrm{ker}}_{0,nm}$ and $\widehat{\gamma}^{\mathrm{ker}}_{1,nm}$.

We first derive an algebraic identity that decomposes $\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm}$ into an integral against $\widehat{\gamma}^{\mathrm{ker}}_{nm}$ and two D_{KL} terms relative to $e^{-\varphi_0}\widehat{\mu}^{\mathrm{ker}}_n$ and $e^{-\psi_0}\widehat{\nu}^{\mathrm{ker}}_m$. By optimality of $\widehat{\gamma}^{\mathrm{ker}}_{nm}$ for $\mathrm{UOT}(\widehat{\mu}^{\mathrm{ker}}_n, \widehat{\nu}^{\mathrm{ker}}_m)$,

$$\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm} = \int \tfrac{1}{2}\|x-y\|^2 \, d\widehat{\gamma}^{\mathrm{ker}}_{nm}(x,y) + D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \widehat{\mu}^{\mathrm{ker}}_n\big) + D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{1,nm} \mid \widehat{\nu}^{\mathrm{ker}}_m\big). \tag{108}$$

We rewrite $D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \widehat{\mu}^{\mathrm{ker}}_n)$ with $\gamma^{\mathrm{or}}_0 := e^{-\varphi_0}\widehat{\mu}^{\mathrm{ker}}_n$. If $f = d\widehat{\gamma}^{\mathrm{ker}}_{0,nm}/d\widehat{\mu}^{\mathrm{ker}}_n$, then $d\widehat{\gamma}^{\mathrm{ker}}_{0,nm}/d\gamma^{\mathrm{or}}_0 = fe^{\varphi_0}$, and therefore

$$\begin{aligned} D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \gamma^{\mathrm{or}}_0\big) &= \int \Big[f\log\big(fe^{\varphi_0}\big) - f + e^{-\varphi_0}\Big] \, d\widehat{\mu}^{\mathrm{ker}}_n \\ &= D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \widehat{\mu}^{\mathrm{ker}}_n\big) + \int \varphi_0 \, d\widehat{\gamma}^{\mathrm{ker}}_{0,nm} - \int \zeta_0 \, d\widehat{\mu}^{\mathrm{ker}}_n, \end{aligned}$$

where we used $\zeta_0 = 1 - e^{-\varphi_0}$. Hence

$$D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \widehat{\mu}^{\mathrm{ker}}_n\big) = D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid \gamma^{\mathrm{or}}_0\big) - \int \varphi_0 \, d\widehat{\gamma}^{\mathrm{ker}}_{0,nm} + \int \zeta_0 \, d\widehat{\mu}^{\mathrm{ker}}_n. \tag{109}$$

The same computation with $\gamma^{\mathrm{or}}_1 := e^{-\psi_0}\widehat{\nu}^{\mathrm{ker}}_m$ yields

$$D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{1,nm} \mid \widehat{\nu}^{\mathrm{ker}}_m\big) = D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{1,nm} \mid \gamma^{\mathrm{or}}_1\big) - \int \psi_0 \, d\widehat{\gamma}^{\mathrm{ker}}_{1,nm} + \int \xi_0 \, d\widehat{\nu}^{\mathrm{ker}}_m. \tag{110}$$

Substituting (109) and (110) into (108) and using

$$\int \varphi_0 \, d\widehat{\gamma}^{\mathrm{ker}}_{0,nm} + \int \psi_0 \, d\widehat{\gamma}^{\mathrm{ker}}_{1,nm} = \int \big(\varphi_0(x) + \psi_0(y)\big) \, d\widehat{\gamma}^{\mathrm{ker}}_{nm}(x,y),$$

we obtain

$$\begin{aligned} &\widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm} - \int \zeta_0 \, d\widehat{\mu}^{\mathrm{ker}}_n - \int \xi_0 \, d\widehat{\nu}^{\mathrm{ker}}_m \\ &\quad = \int_{[0,1]^d \times [0,1]^d} \Big(\tfrac{1}{2}\|x-y\|^2 - \varphi_0(x) - \psi_0(y) \Big) \, d\widehat{\gamma}^{\mathrm{ker}}_{nm}(x,y) \\ &\qquad + D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{0,nm} \mid e^{-\varphi_0}\widehat{\mu}^{\mathrm{ker}}_n\big) + D_{\mathrm{KL}}\big(\widehat{\gamma}^{\mathrm{ker}}_{1,nm} \mid e^{-\psi_0}\widehat{\nu}^{\mathrm{ker}}_m\big). \end{aligned} \tag{111}$$

We define

$$V^{\mathrm{ker}}_{nm} := \widehat{\mathrm{UOT}}^{\mathrm{ker}}_{nm} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0 \, d(\widehat{\mu}^{\mathrm{ker}}_n - \mu) - \int \xi_0 \, d(\widehat{\nu}^{\mathrm{ker}}_m - \nu). \tag{112}$$

Assumption 2 holds on $\Omega = [0,1]^d$. The measures $\widehat{\mu}^{\mathrm{ker}}_n$ and $\widehat{\nu}^{\mathrm{ker}}_m$ satisfy $\widehat{\mu}^{\mathrm{ker}}_n = \rho_\mu \widetilde{\mu}^{\mathrm{ker}}_n$ and $\widehat{\nu}^{\mathrm{ker}}_m = \rho_\nu \widetilde{\nu}^{\mathrm{ker}}_m$ with $\widetilde{\mu}^{\mathrm{ker}}_n(\Omega) = M_\mu$ and $\widetilde{\nu}^{\mathrm{ker}}_m(\Omega) = M_\nu$. Hence, (33) holds almost surely with $\widehat{\mu}_n, \widehat{\nu}_m, \widetilde{\mu}_n, \widetilde{\nu}_m$ replaced by $\widehat{\mu}^{\mathrm{ker}}_n, \widehat{\nu}^{\mathrm{ker}}_m, \widetilde{\mu}^{\mathrm{ker}}_n, \widetilde{\nu}^{\mathrm{ker}}_m$, by Remark 2. Moreover, $|\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{ker}}_n - \widetilde{\mu}^{\mathrm{ker}}_n)| \le \|\zeta_0\|_{L^\infty([0,1]^d)} |\widehat{M}_\mu - M_\mu|$. The analogous bound holds for ξ_0. Thus, subtracting $\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{ker}}_n - \mu) + \int \xi_0 \, d(\widehat{\nu}^{\mathrm{ker}}_m - \nu)$ from both sides of (33), we obtain, almost surely,

$$\begin{aligned} V^{\mathrm{ker}}_{nm} \le\ & C_\Lambda \Big(W_2^2(\widetilde{\mu}^{\mathrm{ker}}_n, \mu) + W_2^2(\widetilde{\nu}^{\mathrm{ker}}_m, \nu) \Big) \\ &+ M_\mu |\rho_\mu - 1| + e\sqrt{M_\mu M_\nu}\, |\log \rho_\mu| + \|\zeta_0\|_{L^\infty([0,1]^d)} |\widehat{M}_\mu - M_\mu| \\ &+ M_\nu |\rho_\nu - 1| + e\sqrt{M_\mu M_\nu}\, |\log \rho_\nu| + \|\xi_0\|_{L^\infty([0,1]^d)} |\widehat{M}_\nu - M_\nu|. \end{aligned}$$

For all n, m sufficiently large such that $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \le a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \le b_m$, taking expectations in the previous display and applying (104) and (36), with its analogue for $\widehat{M}_\nu$, we obtain

$$\mathbb{E}\big[V^{\mathrm{ker}}_{nm}\big] \le C(r^\mu_n + r^\nu_m + a_n + b_m). \tag{113}$$

The constant C in (113) depends on the masses only through the ratio M_ν/M_μ. Since $M_\nu = \int \lambda_0^2 \, d\mu$ and $w_- \le \lambda_0 \le w_+ e^{d/4}$, we have

$$w_-^2 \le \frac{M_\nu}{M_\mu} \le w_+^2 e^{d/2}.$$

Moreover, $\psi_0 = \varphi_0^c \ge \log w_-$, so

$$\|\zeta_0\|_{L^\infty([0,1]^d)} \le 1 + w_+, \qquad \|\xi_0\|_{L^\infty([0,1]^d)} \le 1 + \frac{1}{w_-}.$$

Finally, (23) bounds C_Λ.

Next, using the notation from Proposition 27, we define the event $\mathcal{G}_{nm}$ by

$$\mathcal{G}_{nm} := \left\{ \inf_{x\in[0,1]^d} \widetilde{p}_n^{\ker}(x) \ge \frac{\underline{p}}{2} \quad \text{and} \quad \inf_{y\in[0,1]^d} \widetilde{q}_m^{\ker}(y) \ge \frac{\underline{q}}{2} \right\},$$

where $\widetilde{p}_n^{\ker}$ is defined in (10), $\underline{p} = \inf_{x\in[0,1]^d} p(x)$, and $\widetilde{q}_m^{\ker}$ and $\underline{q}$ are defined analogously. On $\mathcal{G}_{nm}$, the measure estimates have Lebesgue densities bounded below by $\widehat{M}_\mu \underline{p}/2$ and $\widehat{M}_\nu \underline{q}/2$. Hence, Theorem 1 applies to the pair $(\widehat{\mu}_n^{\ker}, \widehat{\nu}_m^{\ker})$. Moreover, replacing $\widehat{\psi}_{nm}^{\ker}$ by $(\widehat{\varphi}_{nm}^{\ker})^c \ge \widehat{\psi}_{nm}^{\ker}$ preserves feasibility in (3) and does not decrease the dual objective. Since $\widehat{\nu}_m^{\ker}$ has a positive density, optimality and continuity yield

$$\widehat{\psi}_{nm}^{\ker} = (\widehat{\varphi}_{nm}^{\ker})^c \qquad \text{on } [0,1]^d.$$

The same argument yields $\widehat{\varphi}_{nm}^{\ker} = (\widehat{\psi}_{nm}^{\ker})^c$. Thus, $(\widehat{\varphi}_{nm}^{\ker}, \widehat{\psi}_{nm}^{\ker})$ is the optimal dual pair given by Theorem 1. Using the same argument as in the proof of Proposition 8, we obtain

$$\widehat{\gamma}_{nm}^{\ker} = (\mathrm{id}, \widehat{T}_{nm}^{\ker})_\# \widehat{\gamma}_{0,nm}^{\ker}, \qquad \widehat{\gamma}_{0,nm}^{\ker} = e^{-\widehat{\varphi}_{nm}^{\ker}} \widehat{\mu}_n^{\ker} = (\widehat{a}_{nm}^{\ker})^2 \widehat{\mu}_n^{\ker} \qquad \text{on } \mathcal{G}_{nm}.$$

Since $\mathrm{UOT}(\mu,\nu) = \int \zeta_0 \, d\mu + \int \xi_0 \, d\nu$, $V_{nm}^{\ker}$ is the left-hand side of (111). Applying Lemma 2 to the integrand in (111), we obtain, on $\mathcal{G}_{nm}$,

$$\frac{\kappa}{2} \int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \, d\widehat{\gamma}_{0,nm}^{\ker}(x) \le V_{nm}^{\ker}. \tag{114}$$

Directly calculating $D_{\mathrm{KL}}\big(\widehat{\gamma}_{0,nm}^{\ker} \mid e^{-\varphi_0} \widehat{\mu}_n^{\ker}\big)$, applying Lemma 17 pointwise with $a = (\widehat{a}_{nm}^{\ker})^2$ and $b = a_0^2$, and using (111) yields the following bound on $\mathcal{G}_{nm}$ for $\widehat{a}_{nm}^{\ker}$ and a_0:

$$\int_{[0,1]^d} |\widehat{a}_{nm}^{\ker}(x) - a_0(x)|^2 \, d\widehat{\mu}_n^{\ker}(x) \le D_{\mathrm{KL}}\big(\widehat{\gamma}_{0,nm}^{\ker} \mid e^{-\varphi_0} \widehat{\mu}_n^{\ker}\big) \le V_{nm}^{\ker}. \tag{115}$$

By the choice of the clipping constants in Section 3.1, $a_0(x)^2 = e^{-\varphi_0(x)} \ge w_-$ for all $x \in [0,1]^d$. Also, since the images of both $\widehat{T}_{nm}^{\ker}$ and T_0 are contained in $[0,1]^d$, we have $\|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \le \mathrm{diam}([0,1]^d)^2 = d$. Consequently,

$$\begin{aligned}
\|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 &\le \frac{1}{w_-} a_0(x)^2 \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \\
&\le \frac{2}{w_-} \widehat{a}_{nm}^{\ker}(x)^2 \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \\
&\quad + \frac{2}{w_-} |\widehat{a}_{nm}^{\ker}(x) - a_0(x)|^2 \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \\
&\le C_{\mathrm{tr}} \, \widehat{a}_{nm}^{\ker}(x)^2 \, \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 + C_{\mathrm{tr}} \, |\widehat{a}_{nm}^{\ker}(x) - a_0(x)|^2,
\end{aligned}$$

with $C_{\mathrm{tr}} := 2\max(1,d)/w_-$. We integrate this bound with respect to $d\widehat{\mu}_n^{\ker}(x)$ and use $d\widehat{\gamma}_{0,nm}^{\ker}(x) = \widehat{a}_{nm}^{\ker}(x)^2 \, d\widehat{\mu}_n^{\ker}(x)$ on $\mathcal{G}_{nm}$. Hence, by (114) and (115), on $\mathcal{G}_{nm}$,

$$\begin{aligned}
\int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \, d\widehat{\mu}_n^{\ker}(x) &\le C_{\mathrm{tr}} \int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \, d\widehat{\gamma}_{0,nm}^{\ker}(x) \\
&\quad + C_{\mathrm{tr}} \int_{[0,1]^d} |\widehat{a}_{nm}^{\ker}(x) - a_0(x)|^2 \, d\widehat{\mu}_n^{\ker}(x) \\
&\le C_{\mathrm{tr}} \big(2/\kappa + 1\big) V_{nm}^{\ker}.
\end{aligned} \tag{116}$$

On $\mathcal{G}_{nm}$, we have $\widehat{p}_n^{\ker} = \widetilde{p}_n^{\ker}$ and $\widehat{q}_m^{\ker} = \widetilde{q}_m^{\ker}$. Hence, $\widehat{p}_n^{\ker} \geq \underline{p}/2$ and $\widehat{q}_m^{\ker} \geq \underline{q}/2$. Moreover, Assumptions 4 and 3 give $\widehat{M}_\mu \geq \bar{c} M_\mu$ and $p \leq \beta_{\max}$. Consequently, on $\mathcal{G}_{nm}$,

$$d\mu(x) = M_\mu p(x)\,dx \leq \frac{2\beta_{\max}}{\bar{c}\,\underline{p}}\,\widehat{M}_\mu \widetilde{p}_n^{\ker}(x)\,dx = \frac{2\beta_{\max}}{\bar{c}\,\underline{p}}\,d\widetilde{\mu}_n^{\ker}(x). \tag{117}$$

Combining (117) with (116) yields, on $\mathcal{G}_{nm}$,

$$\begin{aligned}\int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2\,d\mu(x) &\leq \frac{2\beta_{\max}}{\bar{c}\,\underline{p}} \int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2\,d\widetilde{\mu}_n^{\ker}(x) \\ &\leq \frac{2\beta_{\max}}{\bar{c}\,\underline{p}}\,C_{\mathrm{tr}}\big(2/\kappa + 1\big)V_{nm}^{\ker}.\end{aligned} \tag{118}$$

On the complement $\mathcal{G}_{nm}^c$, the density of $\widehat{\mu}_n^{\ker}$ can vanish on a set of positive Lebesgue measure, so Theorem 1 need not apply. We therefore construct an optimal dual pair for $\mathrm{UOT}(\widehat{\mu}_n^{\ker}, \widehat{\nu}_m^{\ker})$. Let $(\varphi_j, \psi_j)_{j\geq 1}$ be a maximizing sequence for the dual problem (3) with the measures $\widehat{\mu}_n^{\ker}$ and $\widehat{\nu}_m^{\ker}$. Replacing ψ_j by φ_j^c and then φ_j by ψ_j^c preserves feasibility and does not decrease the dual objective, so we may assume that

$$\psi_j = \varphi_j^c, \qquad \varphi_j = \psi_j^c.$$

Since the pair $(0,0)$ is feasible, we may also assume that the dual objective at (φ_j, ψ_j) is nonnegative. Assumption 4 yields $\widehat{M}_\mu, \widehat{M}_\nu > 0$. Hence, the proof of Lemma 7 applies to (φ_j, ψ_j) and yields

$$\sup_{j\geq 1}\big(\|\varphi_j\|_{L^\infty([0,1]^d)} + \|\psi_j\|_{L^\infty([0,1]^d)}\big) < \infty.$$

Moreover, the function $x \mapsto \frac{1}{2}\|x-y\|^2$ is $\sqrt{d}$-Lipschitz on $[0,1]^d$ for every $y \in [0,1]^d$, so φ_j and ψ_j are $\sqrt{d}$-Lipschitz. By the Arzelà-Ascoli theorem, a subsequence of $(\varphi_j, \psi_j)_{j\geq 1}$ converges uniformly to a pair $(\varphi, \psi) \in C_b([0,1]^d)^2$. This pair is feasible. Uniform convergence also shows that this pair attains the supremum in (3), so (φ, ψ) is an optimal dual pair.

Thus, on the event $\mathcal{G}_{nm}^c$, the pair $(\widehat{\varphi}_{nm}^{\ker}, \widehat{\psi}_{nm}^{\ker})$ exists. The set of minimizers in (11) is nonempty because $\widehat{\psi}_{nm}^{\ker}$ is continuous. Both $\widehat{T}_{nm}^{\ker}$ and T_0 take values in the hypercube $[0,1]^d$, so $\|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2 \leq \mathrm{diam}([0,1]^d)^2 = d$ for every $x \in [0,1]^d$. We take $A = 2$ in (88). Hence, (88) and the union bound yield, for all $n, m \geq N_0$, where N_0 is the integer of Proposition 27,

$$\begin{aligned}\mathbb{E}\left[\int_{[0,1]^d} \|\widehat{T}_{nm}^{\ker}(x) - T_0(x)\|^2\,d\mu(x)\mathbf{1}_{\mathcal{G}_{nm}^c}\right] &\leq dM_\mu C_A(n^{-2} + m^{-2}) \\ &\leq dC_A\Big(1 + \frac{M_\mu}{M_\nu}\Big)(r_n^\mu + r_m^\nu),\end{aligned} \tag{119}$$

where the last inequality uses $r_n^\mu \geq M_\mu/n$ and $r_m^\nu \geq M_\nu/m$.

We now prove (105), whose right-hand side has no terms a_n, b_m. Let $\varsigma_{nm} \coloneqq \frac{1}{2}\log(\rho_\mu/\rho_\nu) \in \mathbb{R}$, so that

$$\rho_\mu e^{-\varsigma_{nm}} = \rho_\nu e^{\varsigma_{nm}} = \sqrt{\rho_\mu \rho_\nu}.$$

A pair $(\varphi, \psi) \in C_b([0,1]^d)^2$ satisfies the constraint in (3) if and only if $(\varphi + \varsigma_{nm}, \psi - \varsigma_{nm})$ does. Since $\widehat{\mu}_n^{\ker} = \rho_\mu \widetilde{\mu}_n^{\ker}$ and $\widehat{\nu}_m^{\ker} = \rho_\nu \widetilde{\nu}_m^{\ker}$, the dual objectives satisfy

$$\begin{aligned}&\int \big(1 - e^{-\varphi - \varsigma_{nm}}\big)\,d\widehat{\mu}_n^{\ker} + \int \big(1 - e^{-\psi + \varsigma_{nm}}\big)\,d\widehat{\nu}_m^{\ker} \\ &\quad= \sqrt{\rho_\mu\rho_\nu}\Big(\int \big(1 - e^{-\varphi}\big)\,d\widetilde{\mu}_n^{\ker} + \int \big(1 - e^{-\psi}\big)\,d\widetilde{\nu}_m^{\ker}\Big) \\ &\qquad+ \big(\rho_\mu - \sqrt{\rho_\mu\rho_\nu}\big)M_\mu + \big(\rho_\nu - \sqrt{\rho_\mu\rho_\nu}\big)M_\nu.\end{aligned} \tag{120}$$

The map $(\varphi, \psi) \mapsto (\varphi + \varsigma_{nm}, \psi - \varsigma_{nm})$ is a bijection of the feasible pairs. Thus, the pair $(\widehat{\varphi}_{nm}^{\ker} - \varsigma_{nm}, \widehat{\psi}_{nm}^{\ker} + \varsigma_{nm})$ is an optimal dual pair for $\mathrm{UOT}(\widetilde{\mu}_n^{\ker}, \widetilde{\nu}_m^{\ker})$. For every $x \in [0,1]^d$, the functions

$y \mapsto \|x-y\|^2/2 - \widehat{\psi}^{\ker}_{nm}(y)$ and $y \mapsto \|x-y\|^2/2 - \widehat{\psi}^{\ker}_{nm}(y) - \varsigma_{nm}$ have the same minimizers. Therefore, $\widehat{T}^{\ker}_{nm}$ is also the estimator of T_0 built from this pair.

Consequently, we may take $\widehat{M}_\mu = M_\mu$ and $\widehat{M}_\nu = M_\nu$ when bounding the risk of $\widehat{T}^{\ker}_{nm}$. In this case, $\rho_\mu = \rho_\nu = 1$. Moreover, Assumption 4 holds with $\bar{c} = 1$ and arbitrary positive sequences $a_n, b_m \to 0$. Let $\widetilde{V}^{\ker}_{nm}$ denote the quantity defined in (112) with $\widehat{\mu}^{\ker}_n, \widehat{\nu}^{\ker}_m$ replaced by $\widetilde{\mu}^{\ker}_n, \widetilde{\nu}^{\ker}_m$. Hence, (33) yields

$$\widetilde{V}^{\ker}_{nm} \le C_\Lambda\Big(W_2^2(\widetilde{\mu}^{\ker}_n, \mu) + W_2^2(\widetilde{\nu}^{\ker}_m, \nu)\Big).$$

The functions $\widetilde{p}^{\ker}_n$ and $\widetilde{q}^{\ker}_m$ that define $\mathcal{G}_{nm}$ do not involve $\widehat{M}_\mu$ and $\widehat{M}_\nu$. Thus, (118), with $\bar{c} = 1$ and with $V^{\ker}_{nm}$ replaced by $\widetilde{V}^{\ker}_{nm}$, and (104) yield

$$\mathbb{E}\left[\int_{[0,1]^d} \|\widehat{T}^{\ker}_{nm}(x) - T_0(x)\|^2 \, d\mu(x) \mathbf{1}_{\mathcal{G}_{nm}}\right] \le \frac{2\beta_{\max}}{\underline{p}} \, C_{\mathrm{tr}}\big(2/\kappa + 1\big) C_\Lambda(r^\mu_n + r^\nu_m).$$

Adding (119) yields (105).

For the growth map, Lemma 21, (115), and (114) provide the following bound under $\widehat{\mu}^{\ker}_n$ on $\mathcal{G}_{nm}$, where $C > 0$ is the constant of Lemma 21:

$$\begin{aligned}\int_{[0,1]^d} |\widehat{\lambda}^{\ker}_{nm}(x) - \lambda_0(x)|^2 \, d\widehat{\mu}^{\ker}_n(x) &\le C D_{\mathrm{KL}}\big(\widehat{\gamma}^{\ker}_{0,nm} \mid e^{-\varphi_0}\widehat{\mu}^{\ker}_n\big) \\ &\quad + C\int_{[0,1]^d} \|\widehat{T}^{\ker}_{nm}(x) - T_0(x)\|^2 \, d\widehat{\gamma}^{\ker}_{0,nm}(x) \\ &\le C(1 + 2/\kappa) V^{\ker}_{nm}.\end{aligned}$$

Combining this bound with (117) yields, on $\mathcal{G}_{nm}$,

$$\begin{aligned}\int_{[0,1]^d} |\widehat{\lambda}^{\ker}_{nm}(x) - \lambda_0(x)|^2 \, d\mu(x) &\le \frac{2\beta_{\max}}{\bar{c}\,\underline{p}} \int_{[0,1]^d} |\widehat{\lambda}^{\ker}_{nm}(x) - \lambda_0(x)|^2 \, d\widehat{\mu}^{\ker}_n(x) \\ &\le \frac{2\beta_{\max}}{\bar{c}\,\underline{p}} \, C(1 + 2/\kappa) V^{\ker}_{nm}.\end{aligned}$$

On $\mathcal{G}^c_{nm}$, both $\widehat{\lambda}^{\ker}_{nm}$ and λ_0 are bounded by $w_+ e^{d/4}$. Hence, the same union-bound argument as in (119) absorbs the contribution of $\mathcal{G}^c_{nm}$. Moreover, $V^{\ker}_{nm} \ge 0$, because the right-hand side of (111) is nonnegative and $\mathrm{UOT}(\mu,\nu) = \int \zeta_0 \, d\mu + \int \xi_0 \, d\nu$. Taking expectations and applying (113) therefore proves (106).

Finally, $V^{\ker}_{nm} \ge 0$, the definition (112), the bound on $|\int \zeta_0 \, d(\widehat{\mu}^{\ker}_n - \widetilde{\mu}^{\ker}_n)|$, and the analogous bound for ξ_0 yield

$$\begin{aligned}\left|\widehat{\mathrm{UOT}}^{\ker}_{nm} - \mathrm{UOT}(\mu,\nu)\right| &\le \left|\int \zeta_0 \, d(\widetilde{\mu}^{\ker}_n - \mu)\right| + \left|\int \xi_0 \, d(\widetilde{\nu}^{\ker}_m - \nu)\right| \\ &\quad + \|\zeta_0\|_{L^\infty([0,1]^d)} |\widehat{M}_\mu - M_\mu| + \|\xi_0\|_{L^\infty([0,1]^d)} |\widehat{M}_\nu - M_\nu| + V^{\ker}_{nm}.\end{aligned}$$

Moreover,

$$\begin{aligned}\left|\int \zeta_0 \, d(\widetilde{\mu}^{\ker}_n - \mu)\right| &\le M_\mu \|\zeta_0\|_{L^\infty([0,1]^d)} \|\widehat{p}^{\ker}_n - p\|_{L^1([0,1]^d)}, \\ \left|\int \xi_0 \, d(\widetilde{\nu}^{\ker}_m - \nu)\right| &\le M_\nu \|\xi_0\|_{L^\infty([0,1]^d)} \|\widehat{q}^{\ker}_m - q\|_{L^1([0,1]^d)},\end{aligned}$$

so taking expectations and applying (113) proves (107). □

Proof of Theorem 4. Apply Corollary 28 with

$$r^\mu_n = C M_\mu \mathfrak{R}^{\ker}_n(\alpha), \qquad r^\nu_m = C M_\nu \mathfrak{R}^{\ker}_m(\alpha),$$

and

$$\ell_n = C\mathfrak{L}_n(\alpha), \qquad \ell_m = C\mathfrak{L}_m(\alpha),$$

which are available from Proposition 27, whose hypotheses follow from Assumptions 3 and 5 and the choice of L_n, L_m. We enlarge C so that $C \geq 1$. Since $\mathfrak{N}_n^{\ker}(\alpha) \geq 1/n$ for all $n \geq 3$, this choice yields $r_n^\mu \geq M_\mu/n$ and $r_m^\nu \geq M_\nu/m$. For all sufficiently large n, m, Assumption 4 gives the bounds $\mathbb{E}[|\widehat{M}_\mu - M_\mu|] \leq a_n$ and $\mathbb{E}[|\widehat{M}_\nu - M_\nu|] \leq b_m$, so the corollary yields (14) and (15). □

Remark 4 (Order of the constants in Theorem 4)**.** Let $D = \sqrt{d}$, $\mathfrak{m}$, and $w_\pm$ be as in Remark 3. By tracking the constants in the proof of Corollary 28, the constant in (14) is $O\big(e^{D^2/2}\mathfrak{m}\, C_\Lambda \kappa^{-1}\big)$. The constant in (15) is $O\big(e^{2D^2}\mathfrak{m}^3 C_\Lambda \kappa^{-1}\big)$.

G Wavelet plug-in estimator on the hypercube

G.1 Definition

In this section, we define the wavelet plug-in estimator, which uses a boundary-corrected wavelet basis on the hypercube $\Omega = [0,1]^d$. This basis allows the wavelet plug-in estimator to operate directly on nonperiodic data.

Definition 3 (Boundary-corrected wavelet basis)**.** Fix an integer $N \geq 2$ and a base scale $j_0 \geq \lceil \log_2 N \rceil + 1$. Let $\zeta^{\mathrm{db}}, \xi^{\mathrm{db}}$ denote the compactly supported Daubechies scaling and wavelet functions on $\mathbb{R}$ such that ξ^{db} has N vanishing moments, shifted so that their supports are $[-N+1, N]$. We follow the construction of [CDV93], which is also used in the analysis of plug-in wavelet estimators for balanced OT maps in [MBNWW24].

For each level $j \geq j_0$, define the *interior* one-dimensional translates

$$\zeta_{j,k}^{\mathrm{int}}(t) \coloneqq 2^{j/2}\zeta^{\mathrm{db}}(2^j t - k), \qquad \xi_{j,k}^{\mathrm{int}}(t) \coloneqq 2^{j/2}\xi^{\mathrm{db}}(2^j t - k),$$

for $N \leq k \leq 2^j - N - 1$. The boundary-corrected construction replaces the first and last N translates at each level by the functions

$$\zeta_{j,k}^L,\ \xi_{j,k}^L, \qquad \zeta_{j,k}^R,\ \xi_{j,k}^R, \qquad k = 0, \ldots, N-1,$$

satisfying

$$\begin{aligned}
&\mathrm{supp}(\zeta_{j,k}^L), \mathrm{supp}(\xi_{j,k}^L) \subset [0, (2N-1)2^{-j}], \\
&\mathrm{supp}(\zeta_{j,k}^R), \mathrm{supp}(\xi_{j,k}^R) \subset [1-(2N-1)2^{-j}, 1].
\end{aligned}$$

We then define the one-dimensional boundary-corrected scaling and wavelet families by

$$\zeta_{j,k}^{\mathrm{bc},1}(t) \coloneqq \begin{cases} \zeta_{j,k}^L(t), & 0 \leq k \leq N-1, \\ \zeta_{j,k}^{\mathrm{int}}(t), & N \leq k \leq 2^j - N - 1, \\ \zeta_{j,2^j-1-k}^R(t), & 2^j - N \leq k \leq 2^j - 1, \end{cases}$$

$$\xi_{j,k}^{\mathrm{bc},1}(t) \coloneqq \begin{cases} \xi_{j,k}^L(t), & 0 \leq k \leq N-1, \\ \xi_{j,k}^{\mathrm{int}}(t), & N \leq k \leq 2^j - N - 1, \\ \xi_{j,2^j-1-k}^R(t), & 2^j - N \leq k \leq 2^j - 1. \end{cases}$$

Now write $K(j) \coloneqq \{0, \ldots, 2^j - 1\}^d$. For $k = (k_1, \ldots, k_d)$, $x = (x_1, \ldots, x_d) \in \Omega$, and $\ell \in \{0,1\}^d \setminus \{0\}$, define

$$\begin{aligned}
\zeta_{j_0,k}^{\mathrm{bc}}(x) &\coloneqq \prod_{r=1}^d \zeta_{j_0,k_r}^{\mathrm{bc},1}(x_r), \qquad k \in K(j_0), \\
\xi_{j,k,\ell}^{\mathrm{bc}}(x) &\coloneqq \prod_{r:\ell_r=0} \zeta_{j,k_r}^{\mathrm{bc},1}(x_r) \prod_{r:\ell_r=1} \xi_{j,k_r}^{\mathrm{bc},1}(x_r), \qquad j \geq j_0,\ k \in K(j).
\end{aligned}$$

Set

$$\Phi^{\mathrm{bc}} \coloneqq \left\{\zeta_{j_0,k}^{\mathrm{bc}} : k \in K(j_0)\right\},$$

$$\Psi_j^{\mathrm{bc}} := \left\{\xi_{j,k,\ell}^{\mathrm{bc}} : k \in K(j),\ \ell \in \{0,1\}^d \setminus \{0\}\right\}, \qquad j \geq j_0.$$

The resulting family

$$\Psi^{\mathrm{bc}} := \Phi^{\mathrm{bc}} \cup \bigcup_{j \geq j_0} \Psi_j^{\mathrm{bc}}$$

is the boundary-corrected wavelet basis on $\Omega = [0,1]^d$. By this construction, Ψ^{bc} is an orthonormal basis of $L^2(\Omega)$, and the scaling space $\mathrm{span}(\Phi^{\mathrm{bc}})$ contains all tensor-product polynomials of coordinatewise degree at most $N-1$; in particular, $\mathrm{span}(\Phi^{\mathrm{bc}})$ contains the constants.

Definition 4 (Wavelet plug-in estimator)**.** Let $\Psi^{\mathrm{bc}} = \Phi^{\mathrm{bc}} \cup \bigcup_{j \geq j_0} \Psi_j^{\mathrm{bc}}$ be the boundary-corrected wavelet basis from Definition 3. For $\zeta \in \Phi^{\mathrm{bc}}$ and $\xi \in \Psi_j^{\mathrm{bc}}$, define the empirical coefficients $\widehat{\alpha}_\zeta := \frac{1}{n}\sum_{i=1}^n \zeta(X_i), \widehat{\beta}_\zeta := \frac{1}{m}\sum_{i=1}^m \zeta(Y_i)$, and $\widehat{\alpha}_\xi := \frac{1}{n}\sum_{i=1}^n \xi(X_i), \widehat{\beta}_\xi := \frac{1}{m}\sum_{i=1}^m \xi(Y_i)$. Given resolution levels $J_n, J_m \geq j_0$, the preliminary boundary-corrected wavelet estimators of the normalized densities p and q are

$$\widetilde{p}_n^{\mathrm{wav}}(x) := \sum_{\zeta \in \Phi^{\mathrm{bc}}} \widehat{\alpha}_\zeta \zeta(x) + \sum_{j=j_0}^{J_n} \sum_{\xi \in \Psi_j^{\mathrm{bc}}} \widehat{\alpha}_\xi \xi(x), \quad \widetilde{q}_m^{\mathrm{wav}}(y) := \sum_{\zeta \in \Phi^{\mathrm{bc}}} \widehat{\beta}_\zeta \zeta(y) + \sum_{j=j_0}^{J_m} \sum_{\xi \in \Psi_j^{\mathrm{bc}}} \widehat{\beta}_\xi \xi(y).$$

Since these preliminary estimators may take negative values, we define the density estimates

$$\widehat{p}_n^{\mathrm{wav}}(x) := \frac{(\widetilde{p}_n^{\mathrm{wav}}(x))_+}{\int_\Omega (\widetilde{p}_n^{\mathrm{wav}}(u))_+ \, du}, \qquad \widehat{q}_m^{\mathrm{wav}}(y) := \frac{(\widetilde{q}_m^{\mathrm{wav}}(y))_+}{\int_\Omega (\widetilde{q}_m^{\mathrm{wav}}(v))_+ \, dv}.$$

We then define the measure estimates, which have total masses $\widehat{M}_\mu$ and $\widehat{M}_\nu$, as $\widehat{\mu}_n^{\mathrm{wav}} := \widehat{M}_\mu \widehat{p}_n^{\mathrm{wav}} \, dx$ and $\widehat{\nu}_m^{\mathrm{wav}} := \widehat{M}_\nu \widehat{q}_m^{\mathrm{wav}} \, dx$. Let $(\widehat{\varphi}_{nm}^{\mathrm{wav}}, \widehat{\psi}_{nm}^{\mathrm{wav}})$ be an optimal dual pair for $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$. As for the kernel-based estimator in Section 3.2, the *wavelet plug-in estimator* consists of

$$\widehat{a}_{nm}^{\mathrm{wav}}(x) := e^{-\widehat{\varphi}_{nm}^{\mathrm{wav}}(x)/2}, \qquad \widehat{T}_{nm}^{\mathrm{wav}}(x) \in \operatorname*{arg\,min}_{y \in \Omega} \left\{ \tfrac{1}{2}\|x-y\|^2 - \widehat{\psi}_{nm}^{\mathrm{wav}}(y) \right\},$$

$$\widehat{\lambda}_{nm}^{\mathrm{wav}}(x) := \mathrm{clip}_{[w_-, w_+]}\big(\widehat{a}_{nm}^{\mathrm{wav}}(x)^2\big) \exp\left(\frac{1}{4}\|x - \widehat{T}_{nm}^{\mathrm{wav}}(x)\|^2\right),$$

where $w_\pm$ are the clipping constants of Section 3.1. We also write $\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} := \mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$ for the plug-in estimator of $\mathrm{UOT}(\mu,\nu)$, and $\widehat{\gamma}_{0,nm}^{\mathrm{wav}} := (\widehat{a}_{nm}^{\mathrm{wav}})^2 \widehat{\mu}_n^{\mathrm{wav}} = e^{-\widehat{\varphi}_{nm}^{\mathrm{wav}}} \widehat{\mu}_n^{\mathrm{wav}}$ for the active marginal of $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$. By the primal-dual optimality conditions [LMS18], $\widehat{\gamma}_{0,nm}^{\mathrm{wav}}$ is the first marginal of every optimal transport plan for $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$. We say that the problem $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$ is *Monge-type* if $(\mathrm{id}, \widehat{T}_{nm}^{\mathrm{wav}})_\# \widehat{\gamma}_{0,nm}^{\mathrm{wav}}$ is an optimal transport plan for this problem. By Theorem 1 and the primal-dual optimality conditions, $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$ is Monge-type whenever $\widehat{M}_\mu, \widehat{M}_\nu > 0$ and the density estimates $\widehat{p}_n^{\mathrm{wav}}$ and $\widehat{q}_m^{\mathrm{wav}}$ are bounded below by a positive constant.

The density estimates in Definition 4 are well defined. By Definition 3, the scaling space $\mathrm{span}(\Phi^{\mathrm{bc}})$ contains the constants, so there exist coefficients $(c_\zeta)_{\zeta \in \Phi^{\mathrm{bc}}}$ such that $1 = \sum_{\zeta \in \Phi^{\mathrm{bc}}} c_\zeta \zeta$. The orthonormality of Ψ^{bc} yields $\int_\Omega \zeta(x)\,dx = c_\zeta$ for every $\zeta \in \Phi^{\mathrm{bc}}$ and $\int_\Omega \xi(x)\,dx = 0$ for every $\xi \in \Psi_j^{\mathrm{bc}}$ with $j \geq j_0$. Hence,

$$\int_\Omega \widetilde{p}_n^{\mathrm{wav}}(x)\,dx = \sum_{\zeta \in \Phi^{\mathrm{bc}}} c_\zeta \widehat{\alpha}_\zeta = \frac{1}{n}\sum_{i=1}^n \sum_{\zeta \in \Phi^{\mathrm{bc}}} c_\zeta \zeta(X_i) = \frac{1}{n}\sum_{i=1}^n 1 = 1.$$

The analogous identity holds for $\widetilde{q}_m^{\mathrm{wav}}$. Since $(\widetilde{p}_n^{\mathrm{wav}})_+ \geq \widetilde{p}_n^{\mathrm{wav}}$ and $(\widetilde{q}_m^{\mathrm{wav}})_+ \geq \widetilde{q}_m^{\mathrm{wav}}$, the denominators of $\widehat{p}_n^{\mathrm{wav}}$ and $\widehat{q}_m^{\mathrm{wav}}$ are at least 1.

G.2 Theoretical analysis

Write

$$\mathfrak{R}_n^{\mathrm{wav}}(\alpha) := \begin{cases} n^{-1}, & d = 1, \\ (\log n)^2 n^{-1}, & d = 2, \\ n^{-2\alpha/(2(\alpha-1)+d)}, & d \geq 3, \end{cases}$$

and let $\mathfrak{L}_n(\alpha)$ be as in (81).

Theorem 29 (Error rate of wavelet plug-in estimator)**.** *Assume that $\Omega = [0,1]^d$, and let*

$$\widehat{\mu}_n^{\mathrm{wav}}, \qquad \widehat{\nu}_m^{\mathrm{wav}}, \qquad \widehat{T}_{nm}^{\mathrm{wav}}, \qquad \widehat{\lambda}_{nm}^{\mathrm{wav}}, \qquad \widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}}$$

be as in Definition 4. Assume Assumptions 1, 3, and 4. Assume moreover that $\alpha > 1$, that p and q satisfy

$$\|p\|_{C^{\alpha-1}(\Omega)} + \|q\|_{C^{\alpha-1}(\Omega)} \le M,$$

and that the functions ζ^{db} and ξ^{db} in Definition 3 belong to $C^r(\mathbb{R})$ for some $r > \lceil \alpha - 1 \rceil$. Assume also that $2^{J_n} \asymp n^{1/(d+2(\alpha-1))}$ and $2^{J_m} \asymp m^{1/(d+2(\alpha-1))}$. Let a_n, b_m be the sequences from Assumption 4. Then there exists a constant $C > 0$ such that, for all sufficiently large n, m,

$$\mathbb{E}\left[\int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\mu(x)\right] \le C\Big(M_\mu \mathfrak{R}_n^{\mathrm{wav}}(\alpha) + M_\nu \mathfrak{R}_m^{\mathrm{wav}}(\alpha)\Big),$$

$$\mathbb{E}\left[\int_\Omega |\widehat{\lambda}_{nm}^{\mathrm{wav}}(x) - \lambda_0(x)|^2 \, d\mu(x)\right] \le C\Big(M_\mu \mathfrak{R}_n^{\mathrm{wav}}(\alpha) + M_\nu \mathfrak{R}_m^{\mathrm{wav}}(\alpha) + a_n + b_m\Big),$$

and

$$\begin{aligned}\mathbb{E}\left[\left|\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \mathrm{UOT}(\mu,\nu)\right|\right] \le C\Big(&\mathfrak{L}_n(\alpha) + \mathfrak{L}_m(\alpha) + M_\mu \mathfrak{R}_n^{\mathrm{wav}}(\alpha) \\ &+ M_\nu \mathfrak{R}_m^{\mathrm{wav}}(\alpha) + a_n + b_m\Big).\end{aligned} \tag{121}$$

H Proof of Theorem 29

H.1 Convergence rates of the wavelet density estimators

Throughout this subsection we write

$$s := \alpha - 1 > 0,$$

and we assume that the densities p and q satisfy

$$\beta_{\min} \le p(x), q(x) \le \beta_{\max}, \qquad \|p\|_{C^s(\Omega)} + \|q\|_{C^s(\Omega)} \le M,$$

for some $M > 0$ and $0 < \beta_{\min} \le \beta_{\max} < \infty$. The Wasserstein distance is defined for measures of equal total mass, so we compare μ and ν with

$$\widetilde{\mu}_n^{\mathrm{wav}} := M_\mu \widehat{p}_n^{\mathrm{wav}}(x)\, dx, \qquad \widetilde{\nu}_m^{\mathrm{wav}} := M_\nu \widehat{q}_m^{\mathrm{wav}}(y)\, dy.$$

The following theorem provides bounds for the errors of density estimation by the plug-in wavelet estimators. For balanced OT, [MBNWW24] use only the W_2 bound to derive their rate for one-sample map estimation. In contrast, we also use the two L^1 bounds to derive our rate (121) for estimation of $\mathrm{UOT}(\mu,\nu)$.

Theorem 30 (Wavelet density rates)**.** *Assume $\Omega = [0,1]^d$ and $\alpha > 1$. Assume moreover that*

$$\beta_{\min} \le p(x), q(x) \le \beta_{\max}, \qquad \|p\|_{C^{\alpha-1}(\Omega)} + \|q\|_{C^{\alpha-1}(\Omega)} \le M,$$

that the functions ζ^{db} and ξ^{db} in Definition 3 belong to $C^r(\mathbb{R})$ for some $r > \lceil \alpha - 1 \rceil$, and choose

$$2^{J_n} \asymp n^{1/(d+2(\alpha-1))}, \qquad 2^{J_m} \asymp m^{1/(d+2(\alpha-1))}.$$

Then there exists a constant $C > 0$, depending only on $d, \alpha, M, \beta_{\min}, \beta_{\max}$ and the basis Ψ^{bc}, such that, for all sufficiently large n, m,

$$\begin{aligned}\mathbb{E}\big[W_2^2(\widetilde{\mu}_n^{\mathrm{wav}}, \mu)\big] &\le C M_\mu \mathfrak{R}_n^{\mathrm{wav}}(\alpha), \\ \mathbb{E}\big[W_2^2(\widetilde{\nu}_m^{\mathrm{wav}}, \nu)\big] &\le C M_\nu \mathfrak{R}_m^{\mathrm{wav}}(\alpha),\end{aligned} \tag{122}$$

and

$$\begin{aligned}\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] &\le C \mathfrak{L}_n(\alpha), \\ \mathbb{E}\big[\|\widehat{q}_m^{\mathrm{wav}} - q\|_{L^1(\Omega)}\big] &\le C \mathfrak{L}_m(\alpha).\end{aligned} \tag{123}$$

Proof of Theorem 30. We first prove the L^1 bounds. Let

$$p_{J_n}^{\mathrm{bc}}(x) \coloneqq \sum_{\zeta\in\Phi^{\mathrm{bc}}} \alpha_\zeta \zeta(x) + \sum_{j=j_0}^{J_n} \sum_{\xi\in\Psi_j^{\mathrm{bc}}} \alpha_\xi \xi(x),$$

where $\alpha_\zeta = \int_\Omega \zeta(x)p(x)\,dx$ and $\alpha_\xi = \int_\Omega \xi(x)p(x)\,dx$. Let

$$V_{J_n}^{\mathrm{bc}} \coloneqq \mathrm{span}(\Phi^{\mathrm{bc}}) \oplus \bigoplus_{j=j_0}^{J_n} \mathrm{span}(\Psi_j^{\mathrm{bc}}),$$

so that $p_{J_n}^{\mathrm{bc}}$ is exactly the L^2-orthogonal projection of p onto $V_{J_n}^{\mathrm{bc}}$. Fix $j \geq j_0$ and $\xi \in \Psi_j^{\mathrm{bc}}$, let $I_\xi \subset \Omega$ be a rectangle of diameter at most $C2^{-j}$ that contains the support of ξ, and let π_ξ be the Taylor polynomial of p of degree $\lceil s\rceil - 1$ at a point of I_ξ. Since $\xi^{\mathrm{db}} \in C^r(\mathbb{R})$ with $r > \lceil s\rceil \geq s$, the wavelet ξ^{db} has more than s vanishing moments [Dau92, Corollary 5.5.2], so $N > s$. By [MBNWW24, Lemma 24(iv)], every polynomial of degree at most $N-1$ over Ω, including π_ξ, lies in $\mathrm{span}(\Phi^{\mathrm{bc}})$. Thus, by the orthonormality of the wavelet basis, ξ is orthogonal to π_ξ. Therefore,

$$|\alpha_\xi| = \left|\int_{I_\xi} \xi(x)\big(p(x) - \pi_\xi(x)\big)\,dx\right| \leq \|\xi\|_{L^1(\Omega)} \|p - \pi_\xi\|_{L^\infty(I_\xi)} \leq C2^{-j(s+d/2)}\|p\|_{C^s(\Omega)}, \tag{124}$$

where the last inequality uses $\|p - \pi_\xi\|_{L^\infty(I_\xi)} \leq C2^{-js}\|p\|_{C^s(\Omega)}$, by Taylor's theorem on the rectangle I_ξ of diameter at most $C2^{-j}$, and $\|\xi\|_{L^1(\Omega)} \leq |I_\xi|\,\|\xi\|_{L^\infty(\Omega)} \leq C2^{-jd/2}$, by the bound $\|\xi\|_{L^\infty(\Omega)} \leq C2^{jd/2}$ of [MBNWW24, Lemma 24(ii) and (v)]. By [MBNWW24, Lemma 24(ii) and (v)], we have $\|\xi\|_{L^\infty(\Omega)} \leq C2^{jd/2}$ for $\xi \in \Psi_j^{\mathrm{bc}}$, while every $x \in \Omega$ belongs to at most C of the rectangles I_ξ. Hence, for every $j > J_n$ and every $x \in \Omega$,

$$\begin{aligned}\Big|\sum_{\xi\in\Psi_j^{\mathrm{bc}}} \alpha_\xi \xi(x)\Big| &\leq \sum_{\xi\in\Psi_j^{\mathrm{bc}}} |\alpha_\xi||\xi(x)| \leq C \max_{\xi\in\Psi_j^{\mathrm{bc}}} |\alpha_\xi| \sup_{\xi\in\Psi_j^{\mathrm{bc}}} \|\xi\|_{L^\infty(\Omega)} \\ &\leq C2^{-j(s+d/2)}\|p\|_{C^s(\Omega)} 2^{jd/2} = C2^{-js}\|p\|_{C^s(\Omega)},\end{aligned}$$

so, since $s > 0$ and the expansion converges to p in $L^2(\Omega)$, the wavelet expansion of p converges to p absolutely and uniformly on Ω. In particular, the boundary-corrected projection $p_{J_n}^{\mathrm{bc}} = P_{V_{J_n}^{\mathrm{bc}}} p$ satisfies

$$\|p - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} = \|p - P_{V_{J_n}^{\mathrm{bc}}} p\|_{L^\infty(\Omega)} \leq C\|p\|_{C^s(\Omega)} \sum_{j>J_n} 2^{-js} \leq C2^{-J_n s}, \tag{125}$$

after absorbing $\|p\|_{C^s(\Omega)} \leq M$ into C, and therefore also, since $|\Omega| = 1$,

$$\|p - p_{J_n}^{\mathrm{bc}}\|_{L^1(\Omega)} \leq \|p - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} \leq C2^{-J_n s}.$$

For the stochastic part, the orthonormality of Ψ^{bc} and Parseval's identity imply

$$\begin{aligned}\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^2(\Omega)}^2\big] &= \sum_{\zeta\in\Phi^{\mathrm{bc}}} \mathrm{Var}(\widehat{\alpha}_\zeta) + \sum_{j=j_0}^{J_n} \sum_{\xi\in\Psi_j^{\mathrm{bc}}} \mathrm{Var}(\widehat{\alpha}_\xi) \\ &\leq \frac{C}{n}\left(\#\Phi^{\mathrm{bc}} + \sum_{j=j_0}^{J_n} \#\Psi_j^{\mathrm{bc}}\right) \\ &= \frac{C}{n}\left(2^{dj_0} + (2^d - 1)\sum_{j=j_0}^{J_n} 2^{dj}\right) \leq C\frac{2^{dJ_n}}{n}.\end{aligned}$$

Since $|\Omega| = 1$, Jensen's inequality yields

$$\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^1(\Omega)}\big] \leq \Big(\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^2(\Omega)}^2\big]\Big)^{1/2} \leq C\sqrt{\frac{2^{dJ_n}}{n}}.$$

Therefore

$$\mathbb{E}\big[\|\widetilde{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] \le C\left(2^{-J_n s} + \sqrt{\frac{2^{dJ_n}}{n}}\right). \tag{126}$$

Balancing the two terms in (126) with $2^{J_n} \asymp n^{1/(d+2s)}$ yields

$$\mathbb{E}\big[\|\widetilde{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] \le Cn^{-s/(2s+d)} = C\mathfrak{L}_n(\alpha).$$

Since $\int_\Omega \widetilde{p}_n^{\mathrm{wav}}(x)\,dx = 1$, as shown after Definition 4, and $p(x) \ge 0$ for all $x \in \Omega$, the derivation of (99) in the proof of Proposition 27, with $\widetilde{p}_n^{\ker}$ replaced by $\widetilde{p}_n^{\mathrm{wav}}$, yields

$$\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] \le 3\mathbb{E}\big[\|\widetilde{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] \le C\mathfrak{L}_n(\alpha). \tag{127}$$

The same argument applied to $\widetilde{q}_m^{\mathrm{wav}}$ and q yields the analogue of (127) for $\widehat{q}_m^{\mathrm{wav}}$.

For the W_2 bounds, we use [MBNWW24, Lemma 30], which requires $p, q \ge \beta_{\min}$ and $\|p\|_{\mathcal{B}^s_{\infty,\infty}(\Omega)}, \|q\|_{\mathcal{B}^s_{\infty,\infty}(\Omega)} < \infty$, where $\|\cdot\|_{\mathcal{B}^s_{\infty,\infty}(\Omega)}$ is a Besov norm (cf. [MBNWW24, Appendix A.2.4]). By (124) and $|\alpha_\zeta| \le \|\zeta\|_{L^1(\Omega)}\|p\|_{L^\infty(\Omega)}$, we have $\|p\|_{\mathcal{B}^s_{\infty,\infty}(\Omega)} \le C'\|p\|_{C^s(\Omega)}$, where $C' \coloneqq C + \max_{\zeta\in\Phi^{\mathrm{bc}}}\|\zeta\|_{L^1(\Omega)}$ and C is the constant in (124). The same argument yields $\|q\|_{\mathcal{B}^s_{\infty,\infty}(\Omega)} \le C'\|q\|_{C^s(\Omega)}$. Therefore,

$$\mathbb{E}\big[W_2^2(\widehat{p}_n^{\mathrm{wav}}\,dx, p\,dx)\big] \le C\mathfrak{R}_n^{\mathrm{wav}}(\alpha), \qquad \mathbb{E}\big[W_2^2(\widehat{q}_m^{\mathrm{wav}}\,dx, q\,dx)\big] \le C\mathfrak{R}_m^{\mathrm{wav}}(\alpha).$$

Thus, since M_μ and M_ν are deterministic,

$$W_2^2(\widetilde{\mu}_n^{\mathrm{wav}}, \mu) = M_\mu W_2^2(\widehat{p}_n^{\mathrm{wav}}\,dx, p\,dx), \qquad W_2^2(\widetilde{\nu}_m^{\mathrm{wav}}, \nu) = M_\nu W_2^2(\widehat{q}_m^{\mathrm{wav}}\,dx, q\,dx),$$

which proves (122)–(123). □

Lemma 31 (High-probability lower bound on $\widetilde{p}_n^{\mathrm{wav}}$ and $\widetilde{q}_m^{\mathrm{wav}}$)**.** *Under the assumptions of Theorem 30, write* $\underline{p} \coloneqq \inf_{x\in\Omega} p(x) > 0$ *and* $\underline{q} \coloneqq \inf_{y\in\Omega} q(y) > 0$*, and define the event*

$$\mathcal{H}_{nm} \coloneqq \left\{\inf_{x\in\Omega} \widetilde{p}_n^{\mathrm{wav}}(x) \ge \frac{\underline{p}}{2} \quad \text{and} \quad \inf_{y\in\Omega} \widetilde{q}_m^{\mathrm{wav}}(y) \ge \frac{\underline{q}}{2}\right\}.$$

Then for every $A > 0$ there exists $C_A > 0$ such that, for all sufficiently large n, m,

$$\mathbb{P}\big(\mathcal{H}_{nm}^c\big) \le C_A\big(n^{-A} + m^{-A}\big). \tag{128}$$

On $\mathcal{H}_{nm}$, we have $\widehat{p}_n^{\mathrm{wav}} = \widetilde{p}_n^{\mathrm{wav}}$, $\widehat{q}_m^{\mathrm{wav}} = \widetilde{q}_m^{\mathrm{wav}}$, and

$$d\mu \le \frac{2\beta_{\max}}{\underline{p}}\,d\widetilde{\mu}_n^{\mathrm{wav}}, \qquad d\nu \le \frac{2\beta_{\max}}{\underline{q}}\,d\widetilde{\nu}_m^{\mathrm{wav}}. \tag{129}$$

Proof of Lemma 31. With the rectangles I_ξ from the proof of Theorem 30, let $C_\Psi \ge 1$ be a constant, depending only on d and Ψ^{bc}, such that, for all $j \ge j_0$ and all $x \in \Omega$,

$$\#\Phi^{\mathrm{bc}} \le C_\Psi, \qquad \#\Psi_j^{\mathrm{bc}} \le C_\Psi 2^{dj}, \qquad \max_{\zeta\in\Phi^{\mathrm{bc}}} \|\zeta\|_{L^\infty(\Omega)} \le C_\Psi, \tag{130}$$

$$\max_{\xi\in\Psi_j^{\mathrm{bc}}} \|\xi\|_{L^\infty(\Omega)} \le C_\Psi 2^{jd/2}, \qquad \sum_{\xi\in\Psi_j^{\mathrm{bc}}} \mathbf{1}\{x \in I_\xi\} \le C_\Psi. \tag{131}$$

Recall the coefficients $\widehat{\alpha}_\xi, \widehat{\alpha}_\zeta$ from Definition 4 and α_ξ, α_ζ from the proof of Theorem 30:

$$\widehat{\alpha}_\xi = \frac{1}{n}\sum_{i=1}^n \xi(X_i), \quad \alpha_\xi = \int_\Omega \xi(x)p(x)\,dx, \quad \widehat{\alpha}_\zeta = \frac{1}{n}\sum_{i=1}^n \zeta(X_i), \quad \alpha_\zeta = \int_\Omega \zeta(x)p(x)\,dx.$$

Note that $\beta_{\max} \ge \int_\Omega p(x)\,dx = 1$. Since $\|\xi\|_{L^2(\Omega)} = 1$ and $\beta_{\min} \le p(x) \le \beta_{\max}$ for all $x \in \Omega$,

$$\mathrm{Var}(\xi(X_1)) \le \mathbb{E}\big[\xi(X_1)^2\big] = \int_\Omega \xi(x)^2 p(x)\,dx \le \beta_{\max}\int_\Omega \xi(x)^2\,dx = \beta_{\max}\|\xi\|_{L^2(\Omega)}^2 = \beta_{\max},$$

and $|\xi(X_1) - \alpha_\xi| \le 2\|\xi\|_{L^\infty(\Omega)} \le 2C_\Psi 2^{jd/2}$. Similarly, we have $|\zeta(X_1) - \alpha_\zeta| \le 2C_\Psi$. Hence, applying Bernstein's inequality, we obtain for all $t > 0$,

$$\mathbb{P}\big(|\widehat{\alpha}_\xi - \alpha_\xi| \ge t\big) \le 2\exp\Big(-\frac{nt^2}{2\beta_{\max} + \frac{4}{3}C_\Psi 2^{jd/2} t}\Big), \qquad \xi \in \Psi_j^{\mathrm{bc}}, \tag{132}$$

and

$$\mathbb{P}\big(|\widehat{\alpha}_\zeta - \alpha_\zeta| \ge t\big) \le 2\exp\Big(-\frac{nt^2}{2\beta_{\max} + \frac{4}{3}C_\Psi t}\Big), \qquad \zeta \in \Phi^{\mathrm{bc}}. \tag{133}$$

Since $\int_\Omega p(x)\,dx = 1$, we have $\underline{p} \le 1$. Let $u \in (0,1]$, which we will specify later, and let $t_j := u2^{-jd/2}/(4C_\Psi^2 J_n)$ for $j_0 \le j \le J_n$ and $t_\Phi := u/(4C_\Psi^2)$. Since $\frac{4}{3}C_\Psi 2^{jd/2} t_j = u/(3C_\Psi J_n) \le 1$, $\frac{4}{3}C_\Psi t_\Phi = u/(3C_\Psi) \le 1$ and $\beta_{\max} \ge 1$, setting $t = t_j$ and $t = t_\Phi$ in (132) and (133), respectively, yields the following bounds:

$$\mathbb{P}\Big(\max_{\xi \in \Psi_j^{\mathrm{bc}}} |\widehat{\alpha}_\xi - \alpha_\xi| \ge t_j\Big) \le 2C_\Psi 2^{dj} \exp\Big(-\frac{nu^2 2^{-jd}}{48\beta_{\max} C_\Psi^4 J_n^2}\Big), \tag{134}$$

and

$$\mathbb{P}\Big(\max_{\zeta \in \Phi^{\mathrm{bc}}} |\widehat{\alpha}_\zeta - \alpha_\zeta| \ge t_\Phi\Big) \le 2C_\Psi \exp\Big(-\frac{nu^2}{48\beta_{\max} C_\Psi^4}\Big). \tag{135}$$

Recall that $\widetilde{p}_n^{\mathrm{wav}}(x) - p_{J_n}^{\mathrm{bc}}(x) = \sum_{\zeta \in \Phi^{\mathrm{bc}}} (\widehat{\alpha}_\zeta - \alpha_\zeta)\zeta(x) + \sum_{j=j_0}^{J_n} \sum_{\xi \in \Psi_j^{\mathrm{bc}}} (\widehat{\alpha}_\xi - \alpha_\xi)\xi(x)$. On the event that $|\widehat{\alpha}_\zeta - \alpha_\zeta| < t_\Phi$ for every $\zeta \in \Phi^{\mathrm{bc}}$ and $|\widehat{\alpha}_\xi - \alpha_\xi| < t_j$ for every $\xi \in \Psi_j^{\mathrm{bc}}$ and every $j_0 \le j \le J_n$, the bounds (130) and (131) in the definition of C_Ψ yield, for every $x \in \Omega$,

$$\big|\widetilde{p}_n^{\mathrm{wav}}(x) - p_{J_n}^{\mathrm{bc}}(x)\big| \le C_\Psi^2 t_\Phi + \sum_{j=j_0}^{J_n} C_\Psi^2 2^{jd/2} t_j = \frac{u}{4} + \frac{(J_n - j_0 + 1)u}{4J_n} < u,$$

where the last inequality uses $j_0 \ge 1$. Hence, the event $\{\|\widetilde{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} \ge u\}$ is contained in the union of the events $\{\max_{\zeta \in \Phi^{\mathrm{bc}}} |\widehat{\alpha}_\zeta - \alpha_\zeta| \ge t_\Phi\}$ and $\{\max_{\xi \in \Psi_j^{\mathrm{bc}}} |\widehat{\alpha}_\xi - \alpha_\xi| \ge t_j\}$ for $j_0 \le j \le J_n$. Thus, the union bound, (135), and (134) yield

$$\begin{aligned}\mathbb{P}\big(\|\widetilde{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} \ge u\big) &\le 2C_\Psi \exp\Big(-\frac{nu^2}{48\beta_{\max} C_\Psi^4}\Big) + \sum_{j=j_0}^{J_n} 2C_\Psi 2^{dj} \exp\Big(-\frac{nu^2 2^{-jd}}{48\beta_{\max} C_\Psi^4 J_n^2}\Big) \\ &\le 2(J_n - j_0 + 2) C_\Psi 2^{dJ_n} \exp\Big(-\frac{nu^2 2^{-dJ_n}}{48\beta_{\max} C_\Psi^4 J_n^2}\Big).\end{aligned}$$

Since $2^{J_n} \asymp n^{1/(d+2s)}$, the exponent is at least $cn^{2s/(d+2s)}/\log^2 n$ for a constant $c > 0$ depending only on d, $\beta_{\min}$, $\beta_{\max}$, and C_Ψ, while the prefactor is at most $Cn^{d/(d+2s)} \log n$ for a constant $C > 0$. Hence, letting $u = \underline{p}/4$, we have for every $A > 0$ and all sufficiently large n,

$$\mathbb{P}\left(\|\widetilde{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} \ge \frac{\underline{p}}{4}\right) \le C_A n^{-A}.$$

By (125), we have $\|p_{J_n}^{\mathrm{bc}} - p\|_{L^\infty(\Omega)} \le C2^{-J_n s}$ for some $C > 0$. Thus, $\|p_{J_n}^{\mathrm{bc}} - p\|_{L^\infty(\Omega)} \le \underline{p}/4$ for all sufficiently large n. Therefore, for such n, the event $\{\|\widetilde{p}_n^{\mathrm{wav}} - p\|_{L^\infty(\Omega)} > \underline{p}/2\}$ is contained in the event $\{\|\widetilde{p}_n^{\mathrm{wav}} - p_{J_n}^{\mathrm{bc}}\|_{L^\infty(\Omega)} \ge \underline{p}/4\}$. Hence, $\mathbb{P}(\|\widetilde{p}_n^{\mathrm{wav}} - p\|_{L^\infty(\Omega)} > \underline{p}/2) \le C_A n^{-A}$. The complement of this event implies $\inf_{x \in \Omega} \widetilde{p}_n^{\mathrm{wav}}(x) \ge \underline{p}/2$. The same argument applied to q and $Y_1, \dots, Y_m$ yields the corresponding bound for $\widetilde{q}_m^{\mathrm{wav}}$. Hence, the union bound proves (128).

On $\mathcal{H}_{nm}$, we have $\widetilde{p}_n^{\mathrm{wav}}(x) \ge \underline{p}/2 > 0$ for all $x \in \Omega$, so $(\widetilde{p}_n^{\mathrm{wav}})_+ = \widetilde{p}_n^{\mathrm{wav}}$ and consequently $\widehat{p}_n^{\mathrm{wav}} = \widetilde{p}_n^{\mathrm{wav}}$. Hence, $\widetilde{\mu}_n^{\mathrm{wav}} = M_\mu \widehat{p}_n^{\mathrm{wav}}\,dx \ge M_\mu(\underline{p}/2)\,dx$, while $\mu = M_\mu p\,dx \le M_\mu \beta_{\max}\,dx$, which combine to yield the first bound in (129). The same argument applied to $\widetilde{\nu}_m^{\mathrm{wav}}$ and ν yields the second bound in (129). $\square$

H.2 Proof of the upper bound

Throughout this section we work on the hypercube $\Omega = [0,1]^d$ and write

$$\widehat{\mu}_n^{\mathrm{wav}}, \qquad \widehat{\nu}_m^{\mathrm{wav}}, \qquad \widehat{T}_{nm}^{\mathrm{wav}}, \qquad \widehat{\lambda}_{nm}^{\mathrm{wav}}, \qquad \widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}}$$

for the objects of Definition 4.

Lemma 32 (Plug-in UOT excess identity)**.** *Work on $\Omega = [0,1]^d$, and let*

$$\widehat{\mu}_n^{\mathrm{wav}},\ \widehat{\nu}_m^{\mathrm{wav}},\ \widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}}$$

be as in Definition 4. Let (φ_0, ψ_0) be an optimal dual pair for $\mathrm{UOT}(\mu,\nu)$, define

$$\zeta_0(x) := 1 - e^{-\varphi_0(x)}, \qquad \xi_0(y) := 1 - e^{-\psi_0(y)},$$

and let $\widehat{\gamma}_{nm}^{\mathrm{wav}}$ be any optimal transport plan for $\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}}$ with marginals $(\widehat{\gamma}_{0,nm}^{\mathrm{wav}}, \widehat{\gamma}_{1,nm}^{\mathrm{wav}})$. Define

$$\gamma_{0,nm}^{\mathrm{wav,or}} := e^{-\varphi_0}\widehat{\mu}_n^{\mathrm{wav}},$$
$$\gamma_{1,nm}^{\mathrm{wav,or}} := e^{-\psi_0}\widehat{\nu}_m^{\mathrm{wav}}.$$

Then

$$\begin{aligned}\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \int \zeta_0\, d\widehat{\mu}_n^{\mathrm{wav}} - \int \xi_0\, d\widehat{\nu}_m^{\mathrm{wav}} \\ = \int_{\Omega\times\Omega} \Big(\tfrac{1}{2}\|x-y\|^2 - \varphi_0(x) - \psi_0(y)\Big)\, d\widehat{\gamma}_{nm}^{\mathrm{wav}}(x,y) \\ + D_{\mathrm{KL}}(\widehat{\gamma}_{0,nm}^{\mathrm{wav}} \mid \gamma_{0,nm}^{\mathrm{wav,or}}) + D_{\mathrm{KL}}(\widehat{\gamma}_{1,nm}^{\mathrm{wav}} \mid \gamma_{1,nm}^{\mathrm{wav,or}}).\end{aligned} \tag{136}$$

In particular,

$$\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0\, d(\widehat{\mu}_n^{\mathrm{wav}} - \mu) - \int \xi_0\, d(\widehat{\nu}_m^{\mathrm{wav}} - \nu) \geq 0. \tag{137}$$

Proof of Lemma 32. Since $\widehat{\gamma}_{nm}^{\mathrm{wav}}$ is optimal for $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$,

$$\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} = \int \tfrac{1}{2}\|x-y\|^2\, d\widehat{\gamma}_{nm}^{\mathrm{wav}}(x,y) + D_{\mathrm{KL}}\big(\widehat{\gamma}_{0,nm}^{\mathrm{wav}} \mid \widehat{\mu}_n^{\mathrm{wav}}\big) + D_{\mathrm{KL}}\big(\widehat{\gamma}_{1,nm}^{\mathrm{wav}} \mid \widehat{\nu}_m^{\mathrm{wav}}\big). \tag{138}$$

The computation of (109) and (110) in the proof of Corollary 28, with $(\widehat{\mu}_n^{\mathrm{ker}}, \widehat{\nu}_m^{\mathrm{ker}}, \widehat{\gamma}_{nm}^{\mathrm{ker}})$ replaced by $(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}}, \widehat{\gamma}_{nm}^{\mathrm{wav}})$, yields

$$\begin{aligned} D_{\mathrm{KL}}\big(\widehat{\gamma}_{0,nm}^{\mathrm{wav}} \mid \widehat{\mu}_n^{\mathrm{wav}}\big) &= D_{\mathrm{KL}}\big(\widehat{\gamma}_{0,nm}^{\mathrm{wav}} \mid \gamma_{0,nm}^{\mathrm{wav,or}}\big) - \int \varphi_0\, d\widehat{\gamma}_{0,nm}^{\mathrm{wav}} + \int \zeta_0\, d\widehat{\mu}_n^{\mathrm{wav}}, \\ D_{\mathrm{KL}}\big(\widehat{\gamma}_{1,nm}^{\mathrm{wav}} \mid \widehat{\nu}_m^{\mathrm{wav}}\big) &= D_{\mathrm{KL}}\big(\widehat{\gamma}_{1,nm}^{\mathrm{wav}} \mid \gamma_{1,nm}^{\mathrm{wav,or}}\big) - \int \psi_0\, d\widehat{\gamma}_{1,nm}^{\mathrm{wav}} + \int \xi_0\, d\widehat{\nu}_m^{\mathrm{wav}}. \end{aligned} \tag{139}$$

Substituting (139) into (138) and using $\int \varphi_0\, d\widehat{\gamma}_{0,nm}^{\mathrm{wav}} + \int \psi_0\, d\widehat{\gamma}_{1,nm}^{\mathrm{wav}} = \int (\varphi_0(x)+\psi_0(y))\, d\widehat{\gamma}_{nm}^{\mathrm{wav}}(x,y)$ proves (136).

Finally, duality yields

$$\mathrm{UOT}(\mu,\nu) = \int \zeta_0\, d\mu + \int \xi_0\, d\nu,$$

and (137) follows from $\frac{1}{2}\|x-y\|^2 - \varphi_0(x) - \psi_0(y) \geq 0$ for all $x, y \in \Omega$ and the nonnegativity of the D_{KL} terms. □

Theorem 33 (Plug-in UOT stability bound)**.** *Assume Assumptions 1, 3, and 4, and work on $\Omega = [0,1]^d$. Let C_Λ be the stability constant from Proposition 8, write $\rho_\mu := \widehat{M}_\mu / M_\mu$ and $\rho_\nu := \widehat{M}_\nu / M_\nu$, and define*

$$\begin{aligned} E_{nm}^{\mathrm{wav}} := C_\Lambda\Big(W_2^2(\widetilde{\mu}_n^{\mathrm{wav}}, \mu) + W_2^2(\widetilde{\nu}_m^{\mathrm{wav}}, \nu)\Big) \\ + \big(1 + \|\zeta_0\|_{L^\infty(\Omega)}\big)|\widehat{M}_\mu - M_\mu| + \big(1 + \|\xi_0\|_{L^\infty(\Omega)}\big)|\widehat{M}_\nu - M_\nu| \\ + e\sqrt{M_\mu M_\nu}\big(|\log\rho_\mu| + |\log\rho_\nu|\big). \end{aligned} \tag{140}$$

Then, almost surely,

$$0 \le \widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \mu) - \int \xi_0 \, d(\widehat{\nu}^{\mathrm{wav}}_m - \nu) \le E^{\mathrm{wav}}_{nm}. \tag{141}$$

In particular, almost surely,

$$\begin{aligned}\left|\widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu)\right| \le{}& M_\mu \|\zeta_0\|_{L^\infty(\Omega)} \|\widehat{p}^{\mathrm{wav}}_n - p\|_{L^1(\Omega)} \\ &+ M_\nu \|\xi_0\|_{L^\infty(\Omega)} \|\widehat{q}^{\mathrm{wav}}_m - q\|_{L^1(\Omega)} \\ &+ \|\zeta_0\|_{L^\infty(\Omega)} |\widehat{M}_\mu - M_\mu| + \|\xi_0\|_{L^\infty(\Omega)} |\widehat{M}_\nu - M_\nu| + E^{\mathrm{wav}}_{nm}.\end{aligned} \tag{142}$$

Proof of Theorem 33. The lower bound in (141) is exactly (137) from Lemma 32. Assumption 2 holds on $\Omega = [0,1]^d$. The measures $\widehat{\mu}^{\mathrm{wav}}_n$ and $\widehat{\nu}^{\mathrm{wav}}_m$ satisfy $\widehat{\mu}^{\mathrm{wav}}_n = \rho_\mu \widetilde{\mu}^{\mathrm{wav}}_n$ and $\widehat{\nu}^{\mathrm{wav}}_m = \rho_\nu \widetilde{\nu}^{\mathrm{wav}}_m$ with $\widetilde{\mu}^{\mathrm{wav}}_n(\Omega) = M_\mu$ and $\widetilde{\nu}^{\mathrm{wav}}_m(\Omega) = M_\nu$. Hence, (33) holds almost surely with $\widehat{\mu}_n, \widehat{\nu}_m, \widetilde{\mu}_n, \widetilde{\nu}_m, \bar{\mu}_n, \bar{\nu}_m$ replaced by $\widehat{\mu}^{\mathrm{wav}}_n, \widehat{\nu}^{\mathrm{wav}}_m, \widetilde{\mu}^{\mathrm{wav}}_n, \widetilde{\nu}^{\mathrm{wav}}_m, \widehat{p}^{\mathrm{wav}}_n \, dx, \widehat{q}^{\mathrm{wav}}_m \, dx$, by Remark 2. Moreover, the proof of Theorem 30 yields $M_\mu W_2^2(\widehat{p}^{\mathrm{wav}}_n \, dx, p \, dx) = W_2^2(\widetilde{\mu}^{\mathrm{wav}}_n, \mu)$. We also have $M_\mu |\rho_\mu - 1| = |\widehat{M}_\mu - M_\mu|$. Since $\int_\Omega \widehat{p}^{\mathrm{wav}}_n(x) \, dx = 1$,

$$\left|\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \widetilde{\mu}^{\mathrm{wav}}_n)\right| = |\widehat{M}_\mu - M_\mu| \left|\int_\Omega \zeta_0(x) \widehat{p}^{\mathrm{wav}}_n(x) \, dx\right| \le \|\zeta_0\|_{L^\infty(\Omega)} |\widehat{M}_\mu - M_\mu|. \tag{143}$$

The analogous bounds hold for $\widehat{\nu}^{\mathrm{wav}}_m$, $\widetilde{\nu}^{\mathrm{wav}}_m$, and ξ_0. Subtracting $\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \mu) + \int \xi_0 \, d(\widehat{\nu}^{\mathrm{wav}}_m - \nu)$ from both sides of (33) yields the upper bound in (141).

For (142), write

$$\begin{aligned} A^{\mathrm{wav}}_{nm} &:= \int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \mu) + \int \xi_0 \, d(\widehat{\nu}^{\mathrm{wav}}_m - \nu), \\ V^{\mathrm{wav}}_{nm} &:= \widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu) - A^{\mathrm{wav}}_{nm}. \end{aligned}$$

By (141), $0 \le V^{\mathrm{wav}}_{nm} \le E^{\mathrm{wav}}_{nm}$. Therefore

$$\left|\widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu)\right| \le |A^{\mathrm{wav}}_{nm}| + E^{\mathrm{wav}}_{nm}.$$

Since $\widetilde{\mu}^{\mathrm{wav}}_n = M_\mu \widehat{p}^{\mathrm{wav}}_n \, dx$ and $\mu = M_\mu p \, dx$, (143) yields

$$\begin{aligned}\left|\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \mu)\right| &\le \left|\int \zeta_0 \, d(\widehat{\mu}^{\mathrm{wav}}_n - \widetilde{\mu}^{\mathrm{wav}}_n)\right| + M_\mu \left|\int_\Omega \zeta_0(x)(\widehat{p}^{\mathrm{wav}}_n(x) - p(x)) \, dx\right| \\ &\le \|\zeta_0\|_{L^\infty(\Omega)} |\widehat{M}_\mu - M_\mu| + M_\mu \|\zeta_0\|_{L^\infty(\Omega)} \|\widehat{p}^{\mathrm{wav}}_n - p\|_{L^1(\Omega)}.\end{aligned}$$

The term $|\int \xi_0 \, d(\widehat{\nu}^{\mathrm{wav}}_m - \nu)|$ is treated in the same way, yielding (142). □

Theorem 34 (Plug-in transport map risk)**.** *Assume Assumptions 1, 3, and 4, and work on $\Omega = [0,1]^d$. Suppose that $\mathrm{UOT}(\widehat{\mu}^{\mathrm{wav}}_n, \widehat{\nu}^{\mathrm{wav}}_m)$ is Monge-type, with optimal transport plan*

$$\widehat{\gamma}^{\mathrm{wav}}_{nm} = (\mathrm{id}, \widehat{T}^{\mathrm{wav}}_{nm})_\# \widehat{\gamma}^{\mathrm{wav}}_{0,nm}, \qquad \widehat{\gamma}^{\mathrm{wav}}_{0,nm} = (\widehat{a}^{\mathrm{wav}}_{nm})^2 \widehat{\mu}^{\mathrm{wav}}_n.$$

Then

$$\int_\Omega \|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2 \, d\widehat{\gamma}^{\mathrm{wav}}_{0,nm}(x) \le \frac{2}{\kappa} E^{\mathrm{wav}}_{nm}. \tag{144}$$

Moreover, we have

$$D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{0,nm} \mid e^{-\varphi_0} \widehat{\mu}^{\mathrm{wav}}_n) + D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{1,nm} \mid e^{-\psi_0} \widehat{\nu}^{\mathrm{wav}}_m) \le E^{\mathrm{wav}}_{nm}, \tag{145}$$

and $\widehat{a}^{\mathrm{wav}}_{nm}$ satisfies

$$\int_\Omega |\widehat{a}^{\mathrm{wav}}_{nm}(x) - a_0(x)|^2 \, d\widehat{\mu}^{\mathrm{wav}}_n(x) \le D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{0,nm} \mid e^{-\varphi_0} \widehat{\mu}^{\mathrm{wav}}_n). \tag{146}$$

Finally, there exists a constant $C_{\mathrm{tr}} > 0$, depending only on w_- and $\mathrm{diam}(\Omega)$, such that

$$\int_\Omega \|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2 \, d\widehat{\mu}^{\mathrm{wav}}_n(x) \le C_{\mathrm{tr}}(2/\kappa + 1) E^{\mathrm{wav}}_{nm}. \tag{147}$$

Proof of Theorem 34. Since $\mathrm{UOT}(\mu,\nu) = \int \zeta_0\,d\mu + \int \xi_0\,d\nu$ and $\widehat{\gamma}^{\mathrm{wav}}_{nm} = (\mathrm{id}, \widehat{T}^{\mathrm{wav}}_{nm})_{\#}\widehat{\gamma}^{\mathrm{wav}}_{0,nm}$, Lemma 32 yields

$$\begin{aligned}\widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0\,d(\widehat{\mu}^{\mathrm{wav}}_n - \mu) - \int \xi_0\,d(\widehat{\nu}^{\mathrm{wav}}_m - \nu)\\ = \int \Big(\tfrac{1}{2}\|x - \widehat{T}^{\mathrm{wav}}_{nm}(x)\|^2 - \varphi_0(x) - \psi_0(\widehat{T}^{\mathrm{wav}}_{nm}(x))\Big)\,d\widehat{\gamma}^{\mathrm{wav}}_{0,nm}(x)\\ + D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{0,nm} \mid e^{-\varphi_0}\widehat{\mu}^{\mathrm{wav}}_n) + D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{1,nm} \mid e^{-\psi_0}\widehat{\nu}^{\mathrm{wav}}_m).\end{aligned} \tag{148}$$

Assumption 2 holds on $\Omega = [0,1]^d$. Hence, Lemma 2 yields

$$\tfrac{1}{2}\|x - \widehat{T}^{\mathrm{wav}}_{nm}(x)\|^2 - \varphi_0(x) - \psi_0(\widehat{T}^{\mathrm{wav}}_{nm}(x)) \geq \frac{\kappa}{2}\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2$$

for all $x \in \Omega$. Substituting this lower bound into (148) and discarding $D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{0,nm} \mid e^{-\varphi_0}\widehat{\mu}^{\mathrm{wav}}_n) + D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{1,nm} \mid e^{-\psi_0}\widehat{\nu}^{\mathrm{wav}}_m) \geq 0$ yields the first inequality below. The second inequality is exactly Theorem 33:

$$\begin{aligned}\frac{\kappa}{2}\int_\Omega \|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2\,d\widehat{\gamma}^{\mathrm{wav}}_{0,nm}(x) &\leq \widehat{\mathrm{UOT}}^{\mathrm{wav}}_{nm} - \mathrm{UOT}(\mu,\nu)\\ &\quad - \int \zeta_0\,d(\widehat{\mu}^{\mathrm{wav}}_n - \mu) - \int \xi_0\,d(\widehat{\nu}^{\mathrm{wav}}_m - \nu)\\ &\leq E^{\mathrm{wav}}_{nm}.\end{aligned}$$

Hence, (144) follows immediately. Keeping the D_{KL} terms in (148) and using once again the upper bound in (141) proves (145).

For (146), the identities $\widehat{\gamma}^{\mathrm{wav}}_{0,nm} = (\widehat{a}^{\mathrm{wav}}_{nm})^2\widehat{\mu}^{\mathrm{wav}}_n$ and $e^{-\varphi_0}\widehat{\mu}^{\mathrm{wav}}_n = a_0^2\widehat{\mu}^{\mathrm{wav}}_n$ allow us to apply Lemma 17 pointwise with $a = (\widehat{a}^{\mathrm{wav}}_{nm})^2$ and $b = a_0^2$. Integrating against $\widehat{\mu}^{\mathrm{wav}}_n$ yields

$$\int_\Omega |\widehat{a}^{\mathrm{wav}}_{nm}(x) - a_0(x)|^2\,d\widehat{\mu}^{\mathrm{wav}}_n(x) \leq D_{\mathrm{KL}}(\widehat{\gamma}^{\mathrm{wav}}_{0,nm} \mid e^{-\varphi_0}\widehat{\mu}^{\mathrm{wav}}_n).$$

To prove (147), we bound $\int \|\widehat{T}^{\mathrm{wav}}_{nm} - T_0\|^2\,d\widehat{\mu}^{\mathrm{wav}}_n$ in terms of $\int \|\widehat{T}^{\mathrm{wav}}_{nm} - T_0\|^2\,d\widehat{\gamma}^{\mathrm{wav}}_{0,nm}$. By the choice of the clipping constants in Section 3.1, $a_0(x)^2 = e^{-\varphi_0(x)} \geq w_-$ for all $x \in \Omega$. Because $\widehat{T}^{\mathrm{wav}}_{nm}$ and T_0 both take values in $\Omega = [0,1]^d$, we also have $\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2 \leq \mathrm{diam}(\Omega)^2 = d$. Consequently,

$$\begin{aligned}\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2 &\leq \frac{1}{w_-}a_0(x)^2\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2\\ &\leq \frac{2}{w_-}\widehat{a}^{\mathrm{wav}}_{nm}(x)^2\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2\\ &\quad + \frac{2}{w_-}|\widehat{a}^{\mathrm{wav}}_{nm}(x) - a_0(x)|^2\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2\\ &\leq C_{\mathrm{tr}}\,\widehat{a}^{\mathrm{wav}}_{nm}(x)^2\,\|\widehat{T}^{\mathrm{wav}}_{nm}(x) - T_0(x)\|^2 + C_{\mathrm{tr}}\,|\widehat{a}^{\mathrm{wav}}_{nm}(x) - a_0(x)|^2,\end{aligned}$$

for a constant $C_{\mathrm{tr}} > 0$ depending only on w_- and $\mathrm{diam}(\Omega)$. Integrating against $d\widehat{\mu}^{\mathrm{wav}}_n$ and using $d\widehat{\gamma}^{\mathrm{wav}}_{0,nm} = \widehat{a}^{\mathrm{wav}}_{nm}(x)^2\,d\widehat{\mu}^{\mathrm{wav}}_n(x)$, (144), (145), and (146) yields (147). □

Corollary 35 (Plug-in growth map risk)**.** *Assume that $\Omega = [0,1]^d$ and that Assumptions 1, 3, and 4 hold. Let*

$$\widehat{\mu}^{\mathrm{wav}}_n,\ \widehat{\nu}^{\mathrm{wav}}_m,\ \widehat{T}^{\mathrm{wav}}_{nm},\ \widehat{a}^{\mathrm{wav}}_{nm},\ \widehat{\lambda}^{\mathrm{wav}}_{nm}$$

be as in Definition 4. Suppose that $\mathrm{UOT}(\widehat{\mu}^{\mathrm{wav}}_n, \widehat{\nu}^{\mathrm{wav}}_m)$ is Monge-type, with optimal transport plan

$$\widehat{\gamma}^{\mathrm{wav}}_{nm} = (\mathrm{id}, \widehat{T}^{\mathrm{wav}}_{nm})_{\#}\widehat{\gamma}^{\mathrm{wav}}_{0,nm}, \qquad \widehat{\gamma}^{\mathrm{wav}}_{0,nm} = (\widehat{a}^{\mathrm{wav}}_{nm})^2\widehat{\mu}^{\mathrm{wav}}_n.$$

Then, with $C > 0$ the constant of Lemma 21,

$$\int_\Omega |\widehat{\lambda}^{\mathrm{wav}}_{nm}(x) - \lambda_0(x)|^2\,d\widehat{\mu}^{\mathrm{wav}}_n(x) \leq C(1 + 2/\kappa)E^{\mathrm{wav}}_{nm}. \tag{149}$$

Proof of Corollary 35. By Lemma 21 applied with $\widehat{\eta} = \widehat{\mu}_n^{\mathrm{wav}}$, $\widehat{a} = \widehat{a}_{nm}^{\mathrm{wav}}$, $\widehat{T} = \widehat{T}_{nm}^{\mathrm{wav}}$, and $\widehat{\gamma}_0 = \widehat{\gamma}_{0,nm}^{\mathrm{wav}}$,

$$\int_\Omega |\widehat{\lambda}_{nm}^{\mathrm{wav}} - \lambda_0|^2 \, d\widehat{\mu}_n^{\mathrm{wav}} \le C \int_\Omega |\widehat{a}_{nm}^{\mathrm{wav}} - a_0|^2 \, d\widehat{\mu}_n^{\mathrm{wav}} + C \int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}} - T_0\|^2 \, d\widehat{\gamma}_{0,nm}^{\mathrm{wav}}. \quad (150)$$

The first term on the right-hand side of (150) is at most CE_{nm}^{wav} by (146) and (145). The second term is at most $(2C/\kappa)E_{nm}^{\mathrm{wav}}$ by (144), as claimed. □

Corollary 36 (Risk of the wavelet plug-in estimator)**.** *Assume that $\Omega = [0,1]^d$, that $c(x,y) = \frac{1}{2}\|x-y\|^2$, that Assumptions 1, 3, and 4 hold, and that the hypotheses of Theorem 30 hold. Let*

$$\widehat{\mu}_n^{\mathrm{wav}}, \ \widehat{\nu}_m^{\mathrm{wav}}, \ \widehat{T}_{nm}^{\mathrm{wav}}, \ \widehat{\lambda}_{nm}^{\mathrm{wav}}, \ \widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}}$$

be as in Definition 4, and let a_n, b_m be the sequences from Assumption 4. Assume that for some deterministic sequences $r_{n,\mathrm{wav}}^{\mu}, r_{m,\mathrm{wav}}^{\nu}, \ell_{n,\mathrm{wav}}, \ell_{m,\mathrm{wav}} \ge 0$, where $r_{n,\mathrm{wav}}^{\mu} \ge M_\mu/n$ and $r_{m,\mathrm{wav}}^{\nu} \ge M_\nu/m$,

$$\mathbb{E}\big[W_2^2(\widetilde{\mu}_n^{\mathrm{wav}}, \mu)\big] \le r_{n,\mathrm{wav}}^{\mu}, \qquad \mathbb{E}\big[W_2^2(\widetilde{\nu}_m^{\mathrm{wav}}, \nu)\big] \le r_{m,\mathrm{wav}}^{\nu}. \quad (151)$$

Then there exists a constant $C > 0$ such that, for all sufficiently large n, m,

$$\mathbb{E}\left[\int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\mu(x)\right] \le C(r_{n,\mathrm{wav}}^{\mu} + r_{m,\mathrm{wav}}^{\nu}), \quad (152)$$

$$\mathbb{E}\left[\int_\Omega |\widehat{\lambda}_{nm}^{\mathrm{wav}}(x) - \lambda_0(x)|^2 \, d\mu(x)\right] \le C(r_{n,\mathrm{wav}}^{\mu} + r_{m,\mathrm{wav}}^{\nu} + a_n + b_m), \quad (153)$$

and

$$\mathbb{E}\left[\left|\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0 \, d(\widehat{\mu}_n^{\mathrm{wav}} - \mu) - \int \xi_0 \, d(\widehat{\nu}_m^{\mathrm{wav}} - \nu)\right|\right] \le C(r_{n,\mathrm{wav}}^{\mu} + r_{m,\mathrm{wav}}^{\nu} + a_n + b_m). \quad (154)$$

If, in addition,

$$\mathbb{E}\big[\|\widehat{p}_n^{\mathrm{wav}} - p\|_{L^1(\Omega)}\big] \le \ell_{n,\mathrm{wav}}, \qquad \mathbb{E}\big[\|\widehat{q}_m^{\mathrm{wav}} - q\|_{L^1(\Omega)}\big] \le \ell_{m,\mathrm{wav}}, \quad (155)$$

then, for all sufficiently large n, m,

$$\mathbb{E}\left[\left|\widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \mathrm{UOT}(\mu,\nu)\right|\right] \le M_\mu \|\zeta_0\|_{L^\infty(\Omega)} \ell_{n,\mathrm{wav}} + M_\nu \|\xi_0\|_{L^\infty(\Omega)} \ell_{m,\mathrm{wav}} + C(r_{n,\mathrm{wav}}^{\mu} + r_{m,\mathrm{wav}}^{\nu} + a_n + b_m). \quad (156)$$

Proof of Corollary 36. On the event $\mathcal{H}_{nm}$ of Lemma 31, the density estimates $\widehat{p}_n^{\mathrm{wav}}$ and $\widehat{q}_m^{\mathrm{wav}}$ are bounded below by $\underline{p}/2$ and $\underline{q}/2$. Assumption 4 also yields $\widehat{M}_\mu, \widehat{M}_\nu > 0$. As in the proof of Corollary 28, $(\widehat{\varphi}_{nm}^{\mathrm{wav}}, \widehat{\psi}_{nm}^{\mathrm{wav}})$ is therefore the optimal dual pair given by Theorem 1. Hence, $\mathrm{UOT}(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$ is Monge-type by the last sentence of Definition 4, so (147) and (149) hold on $\mathcal{H}_{nm}$. Moreover, Assumption 4 yields $\rho_\mu, \rho_\nu \ge \bar{c}$, so $d\widetilde{\mu}_n^{\mathrm{wav}} = \rho_\mu^{-1} \, d\widehat{\mu}_n^{\mathrm{wav}} \le \bar{c}^{-1} \, d\widehat{\mu}_n^{\mathrm{wav}}$. Combining this bound with (129), we obtain, on $\mathcal{H}_{nm}$,

$$d\mu \le \frac{2\beta_{\max}}{\bar{c}\,\underline{p}} \, d\widehat{\mu}_n^{\mathrm{wav}}, \qquad d\nu \le \frac{2\beta_{\max}}{\bar{c}\,\underline{q}} \, d\widehat{\nu}_m^{\mathrm{wav}}. \quad (157)$$

We next bound the expectation of E_{nm}^{wav}. Since $\rho_\mu \ge \bar{c}$ almost surely, the mean-value theorem applied to $\log$ on $[\bar{c}, \infty)$ yields $|\log \rho_\mu| \le |\rho_\mu - 1|/\bar{c}$ almost surely, and therefore $e\sqrt{M_\mu M_\nu}\,\mathbb{E}[|\log \rho_\mu|] \le \bar{c}^{-1} e\sqrt{M_\nu/M_\mu}\, a_n$. The analogous bound holds for ρ_ν and b_m. Combining these bounds with (151) and Assumption 4, we obtain, for all sufficiently large n, m,

$$\mathbb{E}\big[E_{nm}^{\mathrm{wav}}\big] \le C_\Lambda(r_{n,\mathrm{wav}}^{\mu} + r_{m,\mathrm{wav}}^{\nu}) + C(a_n + b_m). \quad (158)$$

We now prove (152), whose right-hand side has no terms a_n, b_m. We have $\widehat{\mu}_n^{\mathrm{wav}} = \rho_\mu \widetilde{\mu}_n^{\mathrm{wav}}$, $\widehat{\nu}_m^{\mathrm{wav}} = \rho_\nu \widetilde{\nu}_m^{\mathrm{wav}}$, $\widetilde{\mu}_n^{\mathrm{wav}}(\Omega) = M_\mu$, and $\widetilde{\nu}_m^{\mathrm{wav}}(\Omega) = M_\nu$. Thus, the identity (120) in the proof of

Corollary 28, with $(\widehat{\mu}_n^{\mathrm{ker}}, \widehat{\nu}_m^{\mathrm{ker}}, \widetilde{\mu}_n^{\mathrm{ker}}, \widetilde{\nu}_m^{\mathrm{ker}})$ replaced by $(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}}, \widetilde{\mu}_n^{\mathrm{wav}}, \widetilde{\nu}_m^{\mathrm{wav}})$, shows that the pair $(\widehat{\varphi}_{nm}^{\mathrm{wav}} - \frac{1}{2}\log(\rho_\mu/\rho_\nu), \widehat{\psi}_{nm}^{\mathrm{wav}} + \frac{1}{2}\log(\rho_\mu/\rho_\nu))$ is an optimal dual pair for $\mathrm{UOT}(\widetilde{\mu}_n^{\mathrm{wav}}, \widetilde{\nu}_m^{\mathrm{wav}})$.

For every $x \in \Omega$, the functions $y \mapsto \frac{1}{2}\|x-y\|^2 - \widehat{\psi}_{nm}^{\mathrm{wav}}(y)$ and $y \mapsto \frac{1}{2}\|x-y\|^2 - \widehat{\psi}_{nm}^{\mathrm{wav}}(y) - \frac{1}{2}\log(\rho_\mu/\rho_\nu)$ have the same minimizers. Therefore, $\widehat{T}_{nm}^{\mathrm{wav}}$ is also the estimator of T_0 built from this pair.

Consequently, we may take $\widehat{M}_\mu = M_\mu$ and $\widehat{M}_\nu = M_\nu$ when bounding the risk of $\widehat{T}_{nm}^{\mathrm{wav}}$. In this case, Assumption 4 holds with $\bar{c} = 1$ and arbitrary positive sequences $a_n, b_m \to 0$. Moreover, $\widehat{\mu}_n^{\mathrm{wav}} = \widetilde{\mu}_n^{\mathrm{wav}}$. With these masses, the remainder (140) reduces to

$$\widetilde{E}_{nm}^{\mathrm{wav}} := C_\Lambda\Big(W_2^2(\widetilde{\mu}_n^{\mathrm{wav}}, \mu) + W_2^2(\widetilde{\nu}_m^{\mathrm{wav}}, \nu)\Big).$$

On $\mathcal{H}_{nm}$, (129) and (147), with E_{nm}^{wav} replaced by $\widetilde{E}_{nm}^{\mathrm{wav}}$, yield

$$\begin{aligned}\int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\mu(x) &\le \frac{2\beta_{\max}}{\underline{p}} \int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\widetilde{\mu}_n^{\mathrm{wav}}(x) \\ &\le \frac{2\beta_{\max}}{\underline{p}} \, C_{\mathrm{tr}}(2/\kappa + 1)\widetilde{E}_{nm}^{\mathrm{wav}}.\end{aligned}$$

The functions $\widetilde{p}_n^{\mathrm{wav}}$ and $\widetilde{q}_m^{\mathrm{wav}}$ that define $\mathcal{H}_{nm}$ do not involve $\widehat{M}_\mu$ and $\widehat{M}_\nu$. Since $\widetilde{E}_{nm}^{\mathrm{wav}} \ge 0$, taking expectations and using (151) therefore yields

$$\mathbb{E}\left[\int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\mu(x) \mathbf{1}_{\mathcal{H}_{nm}}\right] \le C(r_{n,\mathrm{wav}}^\mu + r_{m,\mathrm{wav}}^\nu). \tag{159}$$

On $\mathcal{H}_{nm}^c$, the argument used on $\mathcal{G}_{nm}^c$ in the proof of Corollary 28, applied to $\widehat{\mu}_n^{\mathrm{wav}}$ and $\widehat{\nu}_m^{\mathrm{wav}}$, yields an optimal dual pair. Thus, $\widehat{T}_{nm}^{\mathrm{wav}}$ is defined on this event. Both $\widehat{T}_{nm}^{\mathrm{wav}}$ and T_0 take values in $\Omega = [0,1]^d$, so $\|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \le d$, and (128) yields

$$\mathbb{E}\left[\int_\Omega \|\widehat{T}_{nm}^{\mathrm{wav}}(x) - T_0(x)\|^2 \, d\mu(x) \mathbf{1}_{\mathcal{H}_{nm}^c}\right] \le dM_\mu \mathbb{P}(\mathcal{H}_{nm}^c) \le dM_\mu C_A(n^{-A} + m^{-A}). \tag{160}$$

We take $A = 2$. Since $r_{n,\mathrm{wav}}^\mu \ge M_\mu/n$ and $r_{m,\mathrm{wav}}^\nu \ge M_\nu/m$, the right-hand side of (160) is at most $dC_A(1 + M_\mu/M_\nu)(r_{n,\mathrm{wav}}^\mu + r_{m,\mathrm{wav}}^\nu)$. Adding (159) and (160) proves (152).

For the growth map, (157), combined with (149), yields

$$\int_\Omega |\widehat{\lambda}_{nm}^{\mathrm{wav}}(x) - \lambda_0(x)|^2 \, d\mu(x) \le \frac{2\beta_{\max}}{\bar{c}\,\underline{p}} \, C(1 + 2/\kappa) E_{nm}^{\mathrm{wav}} \quad \text{on } \mathcal{H}_{nm}.$$

On $\mathcal{H}_{nm}^c$, both $\widehat{\lambda}_{nm}^{\mathrm{wav}}$ and λ_0 are bounded by $w_+ e^{\mathrm{diam}(\Omega)^2/4}$, so $|\widehat{\lambda}_{nm}^{\mathrm{wav}} - \lambda_0|^2$ is uniformly bounded. Hence, the same argument as in (160) absorbs the contribution of the complement of $\mathcal{H}_{nm}$ into $C(r_{n,\mathrm{wav}}^\mu + r_{m,\mathrm{wav}}^\nu)$. Taking expectations on $\mathcal{H}_{nm}$ and applying (158) proves (153).

For (154), (141) implies

$$\begin{aligned}0 &\le \widehat{\mathrm{UOT}}_{nm}^{\mathrm{wav}} - \mathrm{UOT}(\mu,\nu) - \int \zeta_0 \, d(\widehat{\mu}_n^{\mathrm{wav}} - \mu) - \int \xi_0 \, d(\widehat{\nu}_m^{\mathrm{wav}} - \nu) \\ &\le E_{nm}^{\mathrm{wav}},\end{aligned}$$

so taking expectations and applying (158) proves (154). Finally, taking expectations in (142) and using (155), Assumption 4, and (158) yields (156). □

Proof of Theorem 29. By Assumption 3 and the hypotheses of Theorem 29, the hypotheses of Theorem 30 hold. Hence, we apply Corollary 36 with

$$r_{n,\mathrm{wav}}^\mu = CM_\mu \mathfrak{R}_n^{\mathrm{wav}}(\alpha), \quad r_{m,\mathrm{wav}}^\nu = CM_\nu \mathfrak{R}_m^{\mathrm{wav}}(\alpha), \quad \ell_{n,\mathrm{wav}} = C\mathfrak{L}_n(\alpha), \quad \ell_{m,\mathrm{wav}} = C\mathfrak{L}_m(\alpha),$$

as supplied by Theorem 30, where we enlarge C so that $C \ge 1$. Since $\mathfrak{R}_n^{\mathrm{wav}}(\alpha) \ge 1/n$ for all $n \ge 3$, we obtain $r_{n,\mathrm{wav}}^\mu \ge M_\mu/n$ and $r_{m,\mathrm{wav}}^\nu \ge M_\nu/m$ for all $n, m \ge 3$. The constants $M_\mu\|\zeta_0\|_{L^\infty(\Omega)}$ and $M_\nu\|\xi_0\|_{L^\infty(\Omega)}$ are absorbed into the generic constant C. □

I Proof of Theorem 5

I.1 Statistical model and the class $\mathcal{U}_\alpha(M, B, R)$

Let $\Omega = [0,1]^d$. We assume that M_μ, M_ν are known and observe two independent samples $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \bar{\mu} := \mu/M_\mu$ and $Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \bar{\nu} := \nu/M_\nu$. We write $\mathbb{P}^n_{\mu,\nu}$ and $\mathbb{E}^n_{\mu,\nu}$ for the joint distribution and expectation. The KL divergence between the joint distributions induced by two pairs of source and target measures (μ, ν) and (μ', ν') is

$$D_{\mathrm{KL}}(\mathbb{P}^n_{\mu,\nu} \mid \mathbb{P}^n_{\mu',\nu'}) = n\big[D_{\mathrm{KL}}(\bar{\mu} \mid \bar{\mu}') + D_{\mathrm{KL}}(\bar{\nu} \mid \bar{\nu}')\big]. \tag{161}$$

For the appendix proof, write $\mathcal{U}_\alpha$ for $\mathcal{U}_\alpha(M, B, R)$ from Section 4.5. For $(\mu, \nu) \in \mathcal{U}_\alpha$, let $(T_{\mu,\nu}, \lambda_{\mu,\nu})$ denote the transport-growth pair (T_0, λ_0) given by Theorem 1 for (μ, ν). We also write the maximum risks of estimators $\widehat{T}$ and $\widehat{\lambda}$ as

$$\mathfrak{M}_n^T(\widehat{T}) := \sup_{(\mu,\nu)\in\mathcal{U}_\alpha} \mathbb{E}^n_{\mu,\nu}\left[\int_{[0,1]^d} \|\widehat{T}(x) - T_{\mu,\nu}(x)\|^2 \, d\mu(x)\right],$$

$$\mathfrak{M}_n^\lambda(\widehat{\lambda}) := \sup_{(\mu,\nu)\in\mathcal{U}_\alpha} \mathbb{E}^n_{\mu,\nu}\left[\int_{[0,1]^d} |\widehat{\lambda}(x) - \lambda_{\mu,\nu}(x)|^2 \, d\mu(x)\right].$$

I.2 Tools for minimax lower bounds

Throughout, we use standard tools for minimax lower bounds. See, e.g., [Tsy09, Theorems 2.2 and 2.5], restated here for convenience.

External result 37 (Lower bound from two hypotheses, [Tsy09])**.** *If probability measures P_0, P_1 satisfy $D_{\mathrm{KL}}(P_1 \mid P_0) \le \varrho < \infty$ and $d(\theta_0, \theta_1) \ge 2s$, then there exists a constant $c_\varrho > 0$ such that $\inf_{\widehat{\theta}} \max_{i=0,1} P_i(d(\widehat{\theta}, \theta_i) \ge s) \ge c_\varrho$.*

External result 38 (Lower bound from multiple hypotheses, [Tsy09])**.** *Let (Θ, d) be a pseudometric space and $\{P_\theta\}_{\theta\in\Theta}$ a family of probability measures. Suppose there exist $\theta_0, \theta_1, \dots, \theta_K \in \Theta$ and $s > 0$ such that $P_{\theta_k} \ll P_{\theta_0}$ for $k = 1, \dots, K$, $d(\theta_k, \theta_{k'}) \ge 2s$ for $0 \le k \ne k' \le K$, and $K^{-1}\sum_{k=1}^K D_{\mathrm{KL}}(P_{\theta_k} \mid P_{\theta_0}) \le (\log K)/9$, with $K \ge 2$. Then there exists a universal constant $c > 0$ such that $\inf_{\widehat{\theta}} \sup_k P_{\theta_k}(d(\widehat{\theta}, \theta_k) \ge s) \ge c$.*

External result 39 (Varshamov-Gilbert bound, [Tsy09, Lemma 2.9])**.** *For every integer $N_J \ge 8$ there exist binary strings $\omega^{(0)}, \omega^{(1)}, \dots, \omega^{(K)} \in \{0,1\}^{N_J}$ with $\omega^{(0)} = \mathbf{0}$, $K \ge 2^{N_J/8}$, and pairwise Hamming separation*

$$\|\omega^{(k)} - \omega^{(k')}\|^2 = \#\{i : \omega_i^{(k)} \ne \omega_i^{(k')}\} \ge N_J/8 \qquad \text{for all } 0 \le k \ne k' \le K.$$

In the proof of Theorem 5, we define each hypothesis through its potential, so we also record a lemma that identifies the transport-growth pair associated with a given c-concave potential.

Lemma 40 (Identifiability of the transport-growth pair)**.** *Let $\Omega \subset \mathbb{R}^d$ be compact and convex, and let $\mu = \rho_0\,dx$ with $\rho_0 > 0$ a.e. on Ω. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable and such that $f := \frac{1}{2}\|\cdot\|^2 - \varphi$ satisfies $\nabla^2 f(x) \succ 0$ for all $x \in \mathbb{R}^d$ and $T := \mathrm{id} - \nabla\varphi$ maps Ω into Ω. Define $\psi := \varphi^c$ and*

$$\lambda := \exp\Big(-\varphi + \tfrac{1}{4}\|\nabla\varphi\|^2\Big), \qquad \nu := T_\#(\lambda^2\mu).$$

Then, the potential φ is c-concave, the pair (φ, ψ) is optimal for the dual problem (3) associated with (μ, ν), the transport plan $\gamma := (\mathrm{id}, T)_\#(e^{-\varphi}\mu)$ is optimal for the Kantorovich problem (2), and (T, λ) coincides μ-a.e. with the solution (T_0, λ_0) of the Monge problem (1) provided by Theorem 1 for the pair (μ, ν).

Proof of Lemma 40. We first show that, for every $x \in \Omega$,

$$\psi(T(x)) = \tfrac{1}{2}\|x - T(x)\|^2 - \varphi(x). \tag{162}$$

Fix $x \in \Omega$ and write $y = T(x)$. By definition of the c-transform, $\psi(y) = \inf_{x' \in \Omega}\{\frac{1}{2}\|x' - y\|^2 - \varphi(x')\}$, and the objective is $f(x') - \langle x', y\rangle + \frac{1}{2}\|y\|^2$, a strictly convex function of x' whose gradient vanishes at $x' = x$ because $\nabla f(x) = x - \nabla\varphi(x) = T(x) = y$. A critical point of a convex function is a global minimizer, so the infimum is attained at $x' = x$, which is (162). Since φ is twice continuously differentiable, φ is continuous and bounded on the compact set Ω. The function $\psi = \varphi^c$ is an infimum of functions with a common Lipschitz constant on the bounded set Ω. Thus, ψ is Lipschitz and bounded on Ω. Hence, the pair (φ, ψ) belongs to $C_b(\Omega)^2$. Since $\psi = \varphi^c$, the inequality $\varphi(x) + \psi(y) \le \frac{1}{2}\|x - y\|^2$ holds for all $x, y \in \Omega$. Therefore, the pair (φ, ψ) is feasible for (3). Moreover, (162) yields equality at $y = T(x)$. The potential φ is also c-concave: $\psi = \varphi^c$ implies $\varphi^{cc} \ge \varphi$ pointwise, while (162) yields $\varphi^{cc}(x) \le \frac{1}{2}\|x - T(x)\|^2 - \psi(T(x)) = \varphi(x)$.

Next, we verify the marginal relations of $\gamma = (\mathrm{id}, T)_\#(e^{-\varphi}\mu)$. The first marginal is $\gamma_0 = e^{-\varphi}\mu$ by construction. For the second, (162) and $\|x - T(x)\| = \|\nabla\varphi(x)\|$ yield

$$e^{-\psi(T(x))}\lambda(x)^2 = \exp\Big(\varphi(x) - \tfrac{1}{2}\|\nabla\varphi(x)\|^2\Big)\exp\Big(-2\varphi(x) + \tfrac{1}{2}\|\nabla\varphi(x)\|^2\Big) = e^{-\varphi(x)},$$

so that, for every bounded measurable h,

$$\int h\, e^{-\psi}\, d\nu = \int h(T(x))e^{-\psi(T(x))}\lambda(x)^2\, d\mu(x) = \int h(T(x))e^{-\varphi(x)}\, d\mu(x) = \int h\, d\gamma_1.$$

Hence, $\gamma_1 = e^{-\psi}\nu$.

It remains to show that the objective of (2) at γ is the same as the objective of (3) at (φ, ψ). With $F(r) = r\log r - r + 1$ evaluated at $r = e^{-\varphi}$ and $r = e^{-\psi}$,

$$D_{\mathrm{KL}}(\gamma_0 \mid \mu) = \int \big(-\varphi e^{-\varphi} - e^{-\varphi} + 1\big)\, d\mu = -\int \varphi\, d\gamma_0 + \int (1 - e^{-\varphi})\, d\mu, \tag{163}$$

$$D_{\mathrm{KL}}(\gamma_1 \mid \nu) = -\int \psi\, d\gamma_1 + \int (1 - e^{-\psi})\, d\nu, \tag{164}$$

while (162) yields

$$\int \tfrac{1}{2}\|x - y\|^2\, d\gamma(x, y) = \int \big(\varphi(x) + \psi(T(x))\big)\, d\gamma_0(x) = \int \varphi\, d\gamma_0 + \int \psi\, d\gamma_1. \tag{165}$$

Summing (163), (164), and (165), the potential terms cancel and

$$\int \tfrac{1}{2}\|x - y\|^2\, d\gamma + D_{\mathrm{KL}}(\gamma_0 \mid \mu) + D_{\mathrm{KL}}(\gamma_1 \mid \nu) = \int (1 - e^{-\varphi})\, d\mu + \int (1 - e^{-\psi})\, d\nu.$$

The left-hand side is an upper bound for $\mathrm{UOT}(\mu, \nu)$ and, since (φ, ψ) is feasible for (3), the right-hand side is a lower bound. Hence, both sides coincide with $\mathrm{UOT}(\mu, \nu)$, the transport plan γ is optimal, and (φ, ψ) is an optimal dual pair.

Finally, $\nu = T_\#(\lambda^2\mu) \ll dx$ because $T = \nabla f$ is injective with $\det \nabla T = \det \nabla^2 f > 0$ by assumption, and λ^2 is bounded on Ω. Since T maps Ω into Ω, ν is supported in Ω. Theorem 1 therefore applies to (μ, ν), and the μ-a.e. unique c-concave dual optimizer φ_0 coincides with φ μ-a.e. Hence, the pair (T_0, λ_0) in (4) satisfies $(T_0, \lambda_0) = (T, \lambda)$ μ-a.e., as claimed. □

I.3 Proof of Theorem 5

Proof. To establish the minimax lower bound, we will appeal to both External results 37 and 38. Both results involve constructing hypotheses $(\mu, \nu_k) \in \mathcal{U}_\alpha$ for which estimation is hard. Throughout, $O_\infty(r)$ denotes a function $f : \Omega \to \mathbb{R}$ with $\|f\|_{L^\infty(\Omega)} \le Cr$, where C is a constant independent of k. For any matrix-valued function $A : [0,1]^d \to \mathbb{R}^{d\times d}$, we denote $\|A\|_{\mathrm{op},\infty} := \sup_{x\in[0,1]^d}\|A(x)\|_{\mathrm{op}}$.

We start off by fixing $\mu := \mathrm{Unif}([0,1]^d)$ throughout the entire proof, so that $d\mu(x) = dx$ on $[0,1]^d$.

Setting up multiple hypotheses

We construct a single family of hypotheses $(\mu, \nu_k) \in \mathcal{U}_\alpha$, $k = 0, \dots, K$, whose transport maps and growth maps are both well separated in $L^2(\mu)$ while the distributions $\mathbb{P}^n_{\mu,\nu_k}$ stay close in KL divergence.

Let $\xi \in C^\infty(\mathbb{R})$ be a nonzero smooth bump function whose support is a compact subset of $(0,1)$, and let $x_* \in (0,1)$ satisfy $\xi(x_*) \neq 0$; define

$$g(x) = \prod_{i=1}^{d} \xi(x_i), \qquad x = (x_1, \dots, x_d).$$

Fix a small constant $a > 0$ and a smooth cutoff $\chi \in C_c^\infty((0,1)^d)$ with $\chi \equiv 1$ on $[\delta_0, 1-\delta_0]^d$ for some $\delta_0 \in (0, 1/4)$ such that $\mathrm{supp}(\xi) \subset [\delta_0, 1-\delta_0]$. Define

$$\varphi_0(x) \coloneqq a\chi(x)x_1. \tag{166}$$

In particular, we have $\nabla\varphi_0(x) = (a, 0, \dots, 0)$ for all $x \in [\delta_0, 1-\delta_0]^d$. Since $\varphi_0 \in C_c^\infty((0,1)^d)$, the transport-growth pair (T_0, λ_0) via (4) satisfies $T_0 = \mathrm{id}$ and $\lambda_0 = 1$ on a neighborhood of $\partial[0,1]^d$. Since $\|\nabla^2\varphi_0\|_{\mathrm{op},\infty} \le Ca$, the function $\frac{1}{2}\|x\|^2 - \varphi_0$ remains strictly convex and T_0 is a diffeomorphism for a sufficiently small $a > 0$.

Let $J = \lceil \theta n^{1/(2\alpha-2+d)} \rceil$ for a constant $\theta > 0$ to be determined later, and let $\{x^{(j)}\}_{j\in\mathcal{J}}$ be a grid with $x_i^{(j)} = (j_i - 1)/J$, indexed by the set $\mathcal{J}$ of all $j \in \{1, \dots, J\}^d$ such that $\prod_{i=1}^d [x_i^{(j)}, x_i^{(j)} + 1/J] \subset [\delta_0, 1-\delta_0]^d$; the cardinality $N_J \coloneqq |\mathcal{J}| \asymp J^d$. Define

$$g_j(x) \coloneqq \frac{\varepsilon_{\mathrm{b}}}{J^{\alpha+1}} g(J(x - x^{(j)})),$$

so that $\{\mathrm{supp}(g_j)\}_{j\in\mathcal{J}}$ are pairwise disjoint and contained in $[\delta_0, 1-\delta_0]^d$. Since $\partial^\flat g_j(\cdot) = \varepsilon_{\mathrm{b}} J^{|\flat|-\alpha-1} \partial^\flat g(J(\cdot - x^{(j)}))$ for any multi-index $\flat$, and $\alpha > 1$, one may fix ε_{b} small enough and $J \ge J_0$ large enough that uniformly in j, J:

$$\|g_j\|_\infty \le J^{-\alpha-1}, \quad \|\nabla g_j\|_\infty \le J^{-\alpha}, \quad \|\nabla^2 g_j\|_{\mathrm{op},\infty} \le 1/2, \quad \|g_j\|_{C^{\alpha+1}} \le C_g, \tag{167}$$

for some constant $C_g > 0$. Applying External result 39 with $N_J \ge 8$, we obtain binary strings $\omega^{(0)}, \dots, \omega^{(K)} \in \{0,1\}^{\mathcal{J}}$ with $\omega^{(0)} = \mathbf{0}$, $K \ge 2^{N_J/8}$, and $\|\omega^{(k)} - \omega^{(k')}\|^2 \ge N_J/8$ for $k \neq k'$. Define

$$\begin{aligned} \varphi_k(x) &\coloneqq \varphi_0(x) + \sum_{j\in\mathcal{J}} \omega_j^{(k)} g_j(x), \qquad k = 0, \dots, K, \\ T_k(x) &\coloneqq x - \nabla\varphi_k(x), \quad \lambda_k(x) \coloneqq \exp\Big(-\varphi_k(x) + \frac{1}{4}\|\nabla\varphi_k(x)\|^2 \Big), \\ \nu_k &\coloneqq (T_k)_\#(\lambda_k^2 \mu). \end{aligned} \tag{168}$$

Write $f_k(x) \coloneqq \frac{1}{2}\|x\|^2 - \varphi_k(x)$, so that $T_k = \nabla f_k$.

Notice that the masses of ν_k are not the same across $k = 0, \dots, K$, and we shall denote:

$$H_k \coloneqq \nu_k(\Omega).$$

Since $\nu_k = (T_k)_\#(\lambda_k^2\mu)$, we have $H_k = \int \lambda_k^2 \, d\mu$.

Preliminary analysis of the construction

Since $\nabla^2\varphi_k = \nabla^2\varphi_0 + \nabla^2(\varphi_k - \varphi_0)$ and $\|\nabla^2\varphi_0\|_{\mathrm{op},\infty} \le Ca$, $\|\nabla^2(\varphi_k - \varphi_0)\|_{\mathrm{op},\infty} \le 1/2$ by (167), we have $\nabla^2 f_k \succeq (1/2 - Ca) I \succ 0$ for a small enough, so f_k is strictly convex on $\mathbb{R}^d$.

Moreover, the bumps g_j are translates of the function $x \mapsto \frac{\varepsilon_{\mathrm{b}}}{J^{\alpha+1}} g(Jx)$ with disjoint supports. Every derivative of g_j vanishes on the boundary of $\mathrm{supp}(g_j)$. Hence, $\varphi_k \in C^{\alpha+1}(\Omega)$ with

$$\|\varphi_k\|_{C^{\alpha+1}(\Omega)} \le \|\varphi_0\|_{C^{\alpha+1}} + 2C_g, \tag{169}$$

so $T_k \in C^\alpha(\Omega; \Omega)$ and $\lambda_k \in C^\alpha(\Omega)$ with uniform bounds. Because $\varphi_0 \in C_c^\infty((0,1)^d)$ and $\varphi_k - \varphi_0$ is supported strictly inside $[\delta_0, 1-\delta_0]^d$, we have $\nabla\varphi_k \equiv 0$ on a neighborhood of $\partial[0,1]^d$, hence $T_k \equiv \mathrm{id}$ there. Since $T_k = \nabla f_k$ is a diffeomorphism of $\mathbb{R}^d$, it maps $\partial[0,1]^d$ onto itself and $(0,1)^d$ onto $(0,1)^d$, so $T_k([0,1]^d) = [0,1]^d$. The assumptions of Lemma 40 are therefore satisfied, and the lemma identifies (T_k, λ_k) as the μ-a.e. unique transport-growth pair associated with the pair (μ, ν_k).

Because $\mu = \mathrm{Unif}([0,1]^d)$ and each perturbation is supported strictly inside $[0,1]^d$, the densities of ν_k remain uniformly positive and uniformly bounded and are identically 1 near the boundary. By the change-of-variables formula, the density of ν_k is

$$h_k := \frac{d\nu_k}{d\mu} = \Big(\frac{\lambda_k^2}{\det \nabla T_k}\Big) \circ T_k^{-1}.$$

Thus, (169) yields a uniform bound on $\|h_k\|_{C^{\alpha-1}(\Omega)}$. Since $h_k \equiv 1$ near the boundary of $[0,1]^d$, we have $\|\mathcal{E}h_k\|_{C^{\alpha-1}(\mathbb{T}_2^d)} \le C\|h_k\|_{C^{\alpha-1}(\Omega)}$ for the even reflection $\mathcal{E}$ of Definition 1. Moreover, $\nabla^2 f_k = I - \nabla^2\varphi_k$ satisfies $(1/2 - Ca)I \preceq \nabla^2 f_k \preceq (3/2 + Ca)I$. Combined with the uniform C^α bounds on T_k and λ_k from (169), these properties give constants M_0, B_0, R_0 and κ_0, b_0 such that each pair (μ, ν_k) belongs to $\mathcal{U}_\alpha$ for every $M \ge M_0$, $B \ge B_0/2$, $R \ge R_0/2$, $\kappa \le \kappa_0$, $\beta_{\min} \le b_0$, and $\beta_{\max} \ge 1/b_0$.

Throughout these estimates we compare the hypothesis k with the base hypothesis $k = 0$. Denote $\ell_0 := \log\lambda_0 = -\varphi_0 + \frac{1}{4}\|\nabla\varphi_0\|^2$ and $\ell_k := \log\lambda_k = -\varphi_k + \frac{1}{4}\|\nabla\varphi_k\|^2$. Since $\nabla\varphi_0(x) = (a, 0, \dots, 0)$ for all $x \in [\delta_0, 1-\delta_0]^d$, we compute

$$\begin{aligned}
\ell_k - \ell_0 &= \big(-\varphi_k + \frac{1}{4}\|\nabla\varphi_k\|^2\big) - \big(-\varphi_0 + \frac{1}{4}\|\nabla\varphi_0\|^2\big) \\
&= -(\varphi_k - \varphi_0) + \frac{1}{2}\nabla\varphi_0 \cdot \nabla(\varphi_k - \varphi_0) + \frac{1}{4}\|\nabla(\varphi_k - \varphi_0)\|^2 \\
&= \frac{a}{2}\partial_1(\varphi_k - \varphi_0) - (\varphi_k - \varphi_0) + \frac{1}{4}\|\nabla(\varphi_k - \varphi_0)\|^2. \qquad (170)
\end{aligned}$$

Due to (167) and the definitions of φ_k and φ_0 in (168), the three terms on the right-hand side of (170) have pointwise values of $O(aJ^{-\alpha})$, $O(J^{-\alpha-1})$, and $O(J^{-2\alpha})$; here, $(a/2)\partial_1(\varphi_k - \varphi_0)$ is the largest of the three. By $\lambda_k - \lambda_0 = \lambda_0(e^{\ell_k - \ell_0} - 1)$ and $|e^u - 1 - u| \le u^2 e^{|u|}$,

$$\lambda_k - \lambda_0 = \lambda_0\Big(\frac{a}{2}\partial_1(\varphi_k - \varphi_0) + \mathcal{R}_k\Big), \qquad (171)$$

where $|\mathcal{R}_k| \le |\varphi_k - \varphi_0| + \frac{1}{4}\|\nabla(\varphi_k - \varphi_0)\|^2 + e^{|\ell_k - \ell_0|}(\ell_k - \ell_0)^2 \le CJ^{-\alpha-1}$ for $J \ge J_0$.

Set $\lambda_{\min} := \inf_{x\in[0,1]^d}\lambda_0(x) > 0$ and $\lambda_{\max} := \sup_{x\in[0,1]^d}\lambda_0(x) < \infty$. We next record the total masses, which are needed for the comparison of the distributions.

By (170) and the bump bounds (167), $\|\log\lambda_k - \log\lambda_0\|_\infty = O(aJ^{-\alpha}) \le \log 2$ for all sufficiently large J, so the growth map λ_k satisfies $\lambda_{\min}/2 \le \lambda_k \le 2\lambda_{\max}$ uniformly in k. Combined with $\mu(\Omega) = 1$ and $H_k = \int \lambda_k^2\, d\mu$, this yields

$$\lambda_{\min}^2/4 \le H_k \le 4\lambda_{\max}^2 \qquad \text{for all } k = 0, \dots, K.$$

We shall also need the stability of the inverse maps. To bound $T_k^{-1}(y) - T_0^{-1}(y)$ for a fixed $y \in \Omega$, we set $z_k := T_k^{-1}(y)$, $z_0 := T_0^{-1}(y)$. From $y = T_k(z_k) = z_k - \nabla\varphi_k(z_k)$ and $y = T_0(z_0) = z_0 - \nabla\varphi_0(z_0)$, we have

$$(z_k - z_0) - \big(\nabla\varphi_0(z_k) - \nabla\varphi_0(z_0)\big) = \nabla\varphi_k(z_k) - \nabla\varphi_0(z_k).$$

By the fundamental theorem of calculus, $\nabla\varphi_0(z_k) - \nabla\varphi_0(z_0) = A_y(z_k - z_0)$ with $A_y := \int_0^1 \nabla^2\varphi_0(z_0 + t(z_k - z_0))\, dt$ and $\|A_y\|_{\mathrm{op}} \le \|\nabla^2\varphi_0\|_{\mathrm{op},\infty} \le Ca$. Hence

$$(I - A_y)(z_k - z_0) = \nabla\varphi_k(z_k) - \nabla\varphi_0(z_k).$$

For a sufficiently small $a > 0$, $I - A_y$ is invertible with $\|(I - A_y)^{-1}\|_{\mathrm{op}} \le (1 - Ca)^{-1}$. Consequently, using $\|\nabla(\varphi_k - \varphi_0)\|_\infty \le J^{-\alpha}$ from (167), we obtain:

$$\|T_k^{-1}(y) - T_0^{-1}(y)\| \le \frac{\|\nabla(\varphi_k - \varphi_0)\|_\infty}{1 - Ca} = O(J^{-\alpha}). \qquad (172)$$

Reduction to hypotheses of equal total mass

Another preliminary step is to rescale each ν_k to the total mass $\nu_0(\Omega) = H_0$. For $k = 0, \dots, K$, we set $c_k := \frac{1}{2}\log(H_k/H_0) \in \mathbb{R}$ and define

$$\check{\varphi}_k := \varphi_k + c_k, \quad \check{\lambda}_k := e^{-c_k}\lambda_k, \quad \check{\nu}_k := (T_k)_\#(\check{\lambda}_k^2\mu) = \frac{H_0}{H_k}\nu_k, \quad \bar{\nu}_k := \frac{\nu_k}{H_k}.$$

Thus, $\check{\nu}_k$ has the total mass H_0, and $\bar{\nu}_k$ is its normalized measure. Hence, $(M_\mu, M_\nu) = (1, H_0)$ under every hypothesis $(\mu, \check{\nu}_k)$. Under these hypotheses, each measurable function of $(X_1, \dots, X_n, Y_1, \dots, Y_n, M_\mu, M_\nu)$ is a measurable function of $(X_1, \dots, X_n, Y_1, \dots, Y_n)$, whose distribution is $\mathbb{P}^n_{\mu,\check{\nu}_k}$.

Since $\check{\varphi}_k - \varphi_k = c_k$ is a constant, the function $\check{\varphi}_k$ satisfies the assumptions of Lemma 40 with the same map T_k, and

$$\check{\lambda}_k = \exp\Big(-\check{\varphi}_k + \frac{1}{4}\|\nabla\check{\varphi}_k\|^2\Big).$$

Thus, $(T_k, \check{\lambda}_k)$ is the μ-a.e. unique transport-growth pair associated with the pair $(\mu, \check{\nu}_k)$.

We next bound c_k, which determines the difference between $\check{\lambda}_k$ and λ_k. On $[\delta_0, 1-\delta_0]^d$, we have $\lambda_0^2 = \exp(-2ax_1 + a^2/2)$, so $\partial_1\lambda_0^2 = -2a\lambda_0^2$. Since $\varphi_k - \varphi_0$ is supported in this set, integration by parts yields

$$a\int \lambda_0^2\,\partial_1(\varphi_k - \varphi_0)\,d\mu = 2a^2\int \lambda_0^2(\varphi_k - \varphi_0)\,d\mu. \tag{173}$$

By (167), the right-hand side of (173) is $O(J^{-\alpha-1})$, whereas the integrand on the left-hand side is of order $J^{-\alpha}$. Moreover, $\|\ell_k - \ell_0\|_\infty \le \log 2$ for all sufficiently large J. Since $|e^{2u} - 1 - 2u| \le Cu^2$ for $|u| \le \log 2$, (170) and (173) yield

$$\begin{aligned}|H_k - H_0| &= \Big|\int \lambda_0^2\big(e^{2(\ell_k-\ell_0)} - 1\big)\,d\mu\Big| \\ &\le 2|1-a^2|\Big|\int \lambda_0^2(\varphi_k - \varphi_0)\,d\mu\Big| + CJ^{-2\alpha} \le CJ^{-\alpha-1},\end{aligned}$$

where $C > 0$ is a constant independent of k and J. The last inequality uses (167) and $\alpha > 1$. Since $H_k \ge \lambda_{\min}^2/4$, it follows that $|c_k| \le CJ^{-\alpha-1}$.

Consequently, $e^{|c_k|} \le 2$ for all sufficiently large J. Hence,

$$\|\check{\lambda}_k\|_{C^\alpha([0,1]^d)} \le 2\|\lambda_k\|_{C^\alpha([0,1]^d)}, \qquad \frac{\lambda_k}{2} \le \check{\lambda}_k \le 2\lambda_k.$$

Since $\nabla^2\check{\varphi}_k = \nabla^2\varphi_k$ and $\check{\nu}_k$ and ν_k have the same normalized measure $\bar{\nu}_k$, the pair $(\mu, \check{\nu}_k)$ satisfies condition (i) of $\mathcal{U}_\alpha$ with the same constants as (μ, ν_k). For condition (ii), the transport-growth pair of $(\mu, \check{\nu}_k)$ is $(T_k, \check{\lambda}_k)$. By the choice of B_0 and R_0, we have $\|T_k\|_{C^\alpha([0,1]^d)} + \|\lambda_k\|_{C^\alpha([0,1]^d)} \le B_0/2$ and $2R_0^{-1} \le \lambda_k \le R_0/2$. Combining these bounds with the previous display yields

$$\begin{aligned}\|T_k\|_{C^\alpha([0,1]^d)} + \|\check{\lambda}_k\|_{C^\alpha([0,1]^d)} &\le 2\big(\|T_k\|_{C^\alpha([0,1]^d)} + \|\lambda_k\|_{C^\alpha([0,1]^d)}\big) \le B_0, \\ R_0^{-1} \le \frac{\lambda_k}{2} \le \check{\lambda}_k &\le 2\lambda_k \le R_0.\end{aligned}$$

Hence, each pair $(\mu, \check{\nu}_k)$ belongs to $\mathcal{U}_\alpha$ for every $M \ge M_0$, $B \ge B_0$, $R \ge R_0$, $\kappa \le \kappa_0$, $\beta_{\min} \le b_0$, and $\beta_{\max} \ge 1/b_0$.

Separation bounds for the transport and growth maps

To utilize the multiple hypotheses in External result 38, we calculate the $L^2(\mu)$-separation between the $\check{\lambda}_k$'s. For the separation of the growth maps, we write

$$\check{\lambda}_k - \check{\lambda}_{k'} = (\lambda_k - \lambda_{k'}) + (e^{-c_k} - 1)\lambda_k - (e^{-c_{k'}} - 1)\lambda_{k'},$$

where the last two terms are $O_\infty(J^{-\alpha-1})$ because $|c_k| \le CJ^{-\alpha-1}$ and $\lambda_k \le 2\lambda_{\max}$.

Combined with (171), applied to k and to k' and subtracted, this yields, with $\check{\mathcal{R}}_k := \mathcal{R}_k + (e^{-c_k} - 1)\lambda_k/\lambda_0$,

$$\check{\lambda}_k - \check{\lambda}_{k'} = \lambda_0\Big(\frac{a}{2}\partial_1(\varphi_k - \varphi_{k'}) + \check{\mathcal{R}}_k - \check{\mathcal{R}}_{k'}\Big), \tag{174}$$

where $\|\check{\mathcal{R}}_k\|_\infty \le CJ^{-\alpha-1}$ for $J \ge J_0$, because $|e^{-c_k}-1| \le 2|c_k| \le CJ^{-\alpha-1}$ and $\lambda_k \le 2\lambda_0$. We use the cross term $(a/2)\partial_1(\varphi_k - \varphi_{k'})$ in (174) and the disjointness of $\{\mathrm{supp}(\nabla g_j)\}_{j\in\mathcal{J}}$. We claim

$$c_g \;:=\; \int_{[0,1]^d} |\partial_1 g(y)|^2\, dy \;>\; 0. \tag{175}$$

Recall that $g(x) = \prod_i \xi(x_i)$. If $\partial_1 g \equiv 0$ on $[0,1]^d$, then g is constant in x_1. Since $\mathrm{supp}(g) \subset [\delta_0, 1-\delta_0]^d$, it follows that $g \equiv 0$, contradicting $g(x_*,\dots,x_*) = \xi(x_*)^d \neq 0$. Hence, $\partial_1 g$ is nonzero on a set of positive Lebesgue measure, which proves (175). By definition, $\partial_1 g_j(x) = \varepsilon_{\mathrm{b}} J^{-\alpha}(\partial_1 g)(J(x-x^{(j)}))$. Thus, the substitution $y = J(x-x^{(j)})$ yields

$$\int |\partial_1 g_j|^2\, d\mu = \varepsilon_{\mathrm{b}}^2\, c_g\, J^{-2\alpha-d}.$$

Since $N_J \asymp J^d$, there is a constant $c_l > 0$ with $N_J \ge c_l J^d$ for all $J \ge J_0$. Set $c_\lambda := (1/128)a^2\lambda_{\min}^2 \varepsilon_{\mathrm{b}}^2 c_g c_l > 0$. By (174), the disjoint supports of $\{\partial_1 g_j\}_{j\in\mathcal{J}}$, and $\|\omega^{(k)} - \omega^{(k')}\|^2 \ge N_J/8 \ge c_l J^d/8$, we obtain

$$\begin{aligned}
\|\check{\lambda}_k - \check{\lambda}_{k'}\|_{L^2(\mu)}^2 &\ge \frac{a^2\lambda_{\min}^2}{8} \sum_{j\in\mathcal{J}} (\omega_j^{(k)} - \omega_j^{(k')})^2 \int |\partial_1 g_j|^2\, d\mu - \lambda_{\max}^2 \int (\check{\mathcal{R}}_k - \check{\mathcal{R}}_{k'})^2\, d\mu \\
&\ge \frac{a^2\lambda_{\min}^2}{8} \cdot \frac{c_l J^d}{8} \cdot \varepsilon_{\mathrm{b}}^2 c_g J^{-2\alpha-d} - CJ^{-2\alpha-2} \\
&\ge c_\lambda J^{-2\alpha},
\end{aligned} \tag{176}$$

for J large enough that the remainder $CJ^{-2\alpha-2}$ is absorbed; the constant c_λ is independent of J, k, k'.

The hypothesis $(\mu, \check{\nu}_k)$ has the transport map T_k, because Lemma 40 applies to $\check{\varphi}_k$ with the map T_k. Since $T_k - T_{k'} = -\nabla(\varphi_k - \varphi_{k'}) = -\sum_{j\in\mathcal{J}} (\omega_j^{(k)} - \omega_j^{(k')}) \nabla g_j$ and the supports of $\{\nabla g_j\}_{j\in\mathcal{J}}$ are pairwise disjoint, we have

$$\|T_k - T_{k'}\|_{L^2(\mu)}^2 = \sum_{j\in\mathcal{J}} (\omega_j^{(k)} - \omega_j^{(k')})^2 \int \|\nabla g_j\|^2\, d\mu.$$

By definition, we have $\nabla g_j(x) = \varepsilon_{\mathrm{b}} J^{-\alpha}(\nabla g)(J(x-x^{(j)}))$. Thus, the substitution $y = J(x-x^{(j)})$ yields

$$\int \|\nabla g_j\|^2\, d\mu = \varepsilon_{\mathrm{b}}^2\, c_\nabla\, J^{-2\alpha-d}, \qquad c_\nabla := \int_{[0,1]^d} \|\nabla g(y)\|^2\, dy \ge c_g > 0,$$

where the last inequality follows from $\|\nabla g\|^2 \ge |\partial_1 g|^2$ and (175). Combining this identity with $\|\omega^{(k)} - \omega^{(k')}\|^2 \ge N_J/8 \ge c_l J^d/8$, we obtain the separation

$$\|T_k - T_{k'}\|_{L^2(\mu)}^2 \ge \frac{c_l \varepsilon_{\mathrm{b}}^2 c_\nabla}{8}\, J^{-2\alpha} \qquad \text{for all } 0 \le k \neq k' \le K. \tag{177}$$

KL bounds for the hypotheses

Next, we bound the KL divergence $D_{\mathrm{KL}}(\bar{\nu}_k \mid \bar{\nu}_0)$ between the normalized measures. We do so by the χ^2-divergence in two steps. First, the Cauchy-Schwarz inequality applied to $(H_k - H_0) = \int (h_k - h_0)\, d\mu = \int \frac{h_k - h_0}{\sqrt{h_0}} \cdot \sqrt{h_0}\, d\mu$ yields

$$(H_k - H_0)^2 \le \Big(\int \frac{(h_k-h_0)^2}{h_0}\, d\mu\Big)\Big(\int h_0\, d\mu\Big) = H_0 \int \frac{(h_k-h_0)^2}{h_0}\, d\mu. \tag{178}$$

Writing $\bar{h}_k := h_k/H_k$ for the density of $\bar{\nu}_k$ with respect to μ, and splitting $\bar{h}_k - \bar{h}_0 = \frac{h_k - h_0}{H_k} - \frac{h_0(H_k - H_0)}{H_0 H_k}$, we have

$$D_{\mathrm{KL}}(\bar{\nu}_k \mid \bar{\nu}_0) \le \int \frac{(\bar{h}_k - \bar{h}_0)^2}{\bar{h}_0}\, d\mu \le \frac{2H_0}{H_k^2} \int \frac{(h_k-h_0)^2}{h_0}\, d\mu + \frac{2(H_k-H_0)^2}{H_k^2 H_0} \int h_0\, d\mu$$

$$= \frac{2H_0}{H_k^2}\int \frac{(h_k-h_0)^2}{h_0}\,d\mu + \frac{2(H_k-H_0)^2}{H_k^2}$$
$$\overset{(178)}{\le} \frac{4H_0}{H_k^2}\int \frac{(h_k-h_0)^2}{h_0}\,d\mu \;\le\; C\int \frac{(h_k-h_0)^2}{h_0}\,d\mu, \tag{179}$$

where the last inequality uses $\lambda_{\min}^2/4 \le H_k \le 4\lambda_{\max}^2$ and $H_0 = \int \lambda_0^2\,d\mu \le \lambda_{\max}^2$ to bound $4H_0/H_k^2 \le 64\lambda_{\max}^2/\lambda_{\min}^4$. To bound the integral, we consider $r_k(y) := h_k(y)/h_0(y)$. If $\|\log r_k\|_\infty \le 1/2$ for a sufficiently large J, we have:

$$h_k(y) - h_0(y) = h_0(y)\log r_k(y) + Q_k(y), \qquad |Q_k(y)| \le h_0(y)(\log r_k(y))^2. \tag{180}$$

Consequently,

$$\int \frac{(h_k-h_0)^2}{h_0}\,d\mu \le \frac{9}{4}\|h_0\|_\infty \int (\log r_k)^2\,d\mu. \tag{181}$$

It thus suffices to bound $\int(\log r_k)^2\,d\mu$. By the change-of-variables formula for $\nu_k = (T_k)_\#(\lambda_k^2\mu)$, $\lambda_k = \exp(-\varphi_k + \frac{1}{4}\|\nabla\varphi_k\|^2)$, and $\nabla T_k = I - \nabla^2\varphi_k$,

$$\log h_k(y) = \Big[2\log\lambda_k(x) - \log\det\nabla T_k(x)\Big]\Big|_{x=T_k^{-1}(y)}$$
$$= \underbrace{\Big[-2\varphi_k(x) + \frac{1}{2}\|\nabla\varphi_k(x)\|^2 - \log\det(I-\nabla^2\varphi_k(x))\Big]}_{L_k(x)}\Big|_{x=T_k^{-1}(y)}.$$

Thus $\log r_k(y) = L_k(T_k^{-1}(y)) - L_0(T_0^{-1}(y))$, which we split as

$$\log r_k(y) = (L_k - L_0)(T_k^{-1}(y)) + \big[L_0(T_k^{-1}(y)) - L_0(T_0^{-1}(y))\big]. \tag{182}$$

We first bound $L_k(x) - L_0(x)$. By definition,

$$L_k(x) - L_0(x) = -2(\varphi_k-\varphi_0)(x) + \frac{1}{2}\big(\|\nabla\varphi_k(x)\|^2 - \|\nabla\varphi_0(x)\|^2\big)$$
$$+ \big[-\log\det(I-\nabla^2\varphi_k(x)) + \log\det(I-\nabla^2\varphi_0(x))\big]. \tag{183}$$

We bound the three terms on the right-hand side of (183) in turn, using (166), (167), and (168).

(i) *Zeroth-order term.* $\varphi_k - \varphi_0 = O_\infty(J^{-\alpha-1})$.

(ii) *Gradient-squared term.* Since $\nabla\varphi_0 = (a,0,\dots,0)$ on $[\delta_0, 1-\delta_0]^d$,

$$\frac{1}{2}\|\nabla\varphi_k\|^2 - \frac{1}{2}\|\nabla\varphi_0\|^2 = a\partial_1(\varphi_k-\varphi_0) + \frac{1}{2}\|\nabla(\varphi_k-\varphi_0)\|^2 = O_\infty(aJ^{-\alpha}) + O_\infty(J^{-2\alpha}).$$

(iii) *Log-determinant term.* We claim that pointwise on Ω,

$$-\log\det(I-\nabla^2\varphi_k(x)) + \log\det(I-\nabla^2\varphi_0(x)) = -\log\det(I-\nabla^2(\varphi_k-\varphi_0)(x)). \tag{184}$$

To see this, let $\mathcal{S} := \bigcup_{j\in\mathcal{J}}\operatorname{supp}(g_j)$ and split Ω into $\mathcal{S}$ and $\Omega\setminus\mathcal{S}$. For $x\in\mathcal{S}$, that is $x\in\operatorname{supp}(g_j)$ for some $j\in\mathcal{J}$, we have $\chi(x)=1$, so $\varphi_0(x) = ax_1$ is affine and $\nabla^2\varphi_0(x) = 0$; hence $\nabla^2\varphi_k(x) = \nabla^2(\varphi_k-\varphi_0)(x)$ and $\log\det(I-\nabla^2\varphi_0(x)) = 0$, which yields (184). For $x\in\Omega\setminus\mathcal{S}$, $\varphi_k(x) - \varphi_0(x) = \sum_{j\in\mathcal{J}}\omega_j^{(k)}g_j(x) = 0$, so $\nabla^2\varphi_k(x) = \nabla^2\varphi_0(x)$ and both sides of (184) vanish. The claim thus follows.

Denote $A := \nabla^2(\varphi_k-\varphi_0)$. Since $\|\nabla^2(\varphi_k-\varphi_0)\|_{\mathrm{op},\infty} \le 1/2$ by (167), we can apply the Taylor expansion to obtain $\log\det(I-A) = -\operatorname{tr}(A) + O(\|A\|_{\mathrm{op}}^2)$. Since $\|\nabla^2(\varphi_k-\varphi_0)\|_{\mathrm{op},\infty} = O(J^{1-\alpha})$, we have

$$-\log\det(I-\nabla^2(\varphi_k-\varphi_0)) = \Delta(\varphi_k-\varphi_0) + O_\infty(J^{2-2\alpha}).$$

Combining (i)–(iii) with $\varphi_k - \varphi_0 = O_\infty(J^{-\alpha-1})$, we obtain, for $J \geq 1/a$,

$$L_k(x) - L_0(x) = \Delta\varphi_k(x) - \Delta\varphi_0(x) + O_\infty(aJ^{-\alpha}) + O_\infty(J^{2-2\alpha}). \tag{185}$$

For the second term in (182), the function L_0 is Lipschitz on $[0,1]^d$ with constant at most Ca, since $\varphi_0(x) = a\chi(x)x_1$ and a is small. Hence, for every $y \in \Omega$,

$$\left|L_0(T_k^{-1}(y)) - L_0(T_0^{-1}(y))\right| \leq Ca\,\|T_k^{-1}(y) - T_0^{-1}(y)\|. \tag{186}$$

Substituting (185), (186) and (172) into (182) yields

$$\log r_k(y) = \Delta\varphi_k(T_k^{-1}(y)) - \Delta\varphi_0(T_k^{-1}(y)) + O_\infty(aJ^{-\alpha}) + O_\infty(J^{2-2\alpha}). \tag{187}$$

In particular $\|\log r_k\|_\infty = O(J^{1-\alpha}) \leq 1/2$ for a sufficiently large J, justifying the condition used in (180).

Combining (179) and (181) with (187) yields

$$D_{\mathrm{KL}}(\bar\nu_k \mid \bar\nu_0) \leq C\int(\Delta\varphi_k \circ T_k^{-1} - \Delta\varphi_0 \circ T_k^{-1})^2\,d\mu + O(a^2J^{-2\alpha}) + O(J^{4-4\alpha}). \tag{188}$$

Since $N_J \asymp J^d$, there is a constant $c_u > 0$ with $N_J \leq c_u J^d$ for all $J \geq J_0$. We now bound the leading term $\int(\Delta(\varphi_k-\varphi_0)\circ T_k^{-1})^2\,d\mu$. Since $\Delta(\varphi_k-\varphi_0)$ is supported in $\bigcup_{j\in\mathcal{J}}\mathrm{supp}(g_j) \subset [\delta_0, 1-\delta_0]^d$, where $\nabla^2\varphi_0 \equiv 0$, the change of variables $x = T_k^{-1}(y)$ has Jacobian $\det\nabla T_k(x) = \det(I - \nabla^2(\varphi_k-\varphi_0)) = 1+O(J^{1-\alpha})$, so $\|\Delta(\varphi_k-\varphi_0)\circ T_k^{-1}\|_{L^2(\mu)}^2 \leq C\|\Delta(\varphi_k-\varphi_0)\|_{L^2(\mu)}^2$. Using disjoint supports of $\{g_j\}_{j\in\mathcal{J}}$, $(\omega_j^{(k)})^2 = \omega_j^{(k)} \in \{0,1\}$, and the rescaling $\Delta g_j(x) = \varepsilon_{\mathrm{b}}J^{1-\alpha}(\Delta g)(J(x - x^{(j)}))$,

$$\int(\Delta(\varphi_k - \varphi_0))^2\,d\mu = \sum_{j\in\mathcal{J}}\omega_j^{(k)} \cdot \frac{\varepsilon_{\mathrm{b}}^2}{J^{2\alpha-2+d}}\int(\Delta g)^2\,dy \leq C\varepsilon_{\mathrm{b}}^2J^{2-2\alpha},$$

using $\sum_{j\in\mathcal{J}}\omega_j^{(k)} \leq N_J \leq c_uJ^d$. The two remainders in (188) contribute $O(a^2J^{-2\alpha})$ and $O(J^{4-4\alpha})$. Both remainders are $O(J^{2-2\alpha})$ for $\alpha \geq 1$. Hence,

$$D_{\mathrm{KL}}(\bar\nu_k \mid \bar\nu_0) \leq CJ^{2-2\alpha}.$$

By (161) and μ fixed across hypotheses, for a constant $C_{\mathrm{KL}} > 0$ independent of n and J,

$$D_{\mathrm{KL}}(\mathbb{P}_{\mu,\nu_k}^n \mid \mathbb{P}_{\mu,\nu_0}^n) = nD_{\mathrm{KL}}(\bar\nu_k \mid \bar\nu_0) \leq C_{\mathrm{KL}}nJ^{2-2\alpha}.$$

From External result 39 and $N_J \geq c_lJ^d$, $\log K \geq (N_J/8)\log 2 \geq (c_lJ^d\log 2)/8$, so

$$(\log K)/9 \geq (c_lJ^d\log 2)/72.$$

Using $J \geq \theta n^{1/(2\alpha-2+d)}$ from the definition $J = \lceil\theta n^{1/(2\alpha-2+d)}\rceil$,

$$\theta^{2\alpha-2+d} \geq 72C_{\mathrm{KL}}/(c_l\log 2) \quad\Longrightarrow\quad C_{\mathrm{KL}}nJ^{2-2\alpha} \leq (\log K)/9.$$

Therefore, by choosing

$$\theta > (72C_{\mathrm{KL}}/(c_l\log 2))^{1/(2\alpha-2+d)},$$

we have $D_{\mathrm{KL}}(\mathbb{P}_{\mu,\nu_k}^n \mid \mathbb{P}_{\mu,\nu_0}^n) \leq (\log K)/9$, as required by External result 38.

Moreover, the right-hand side of (161) depends only on the normalized measures. Therefore,

$$D_{\mathrm{KL}}(\mathbb{P}_{\mu,\check\nu_k}^n \mid \mathbb{P}_{\mu,\check\nu_0}^n) \leq (\log K)/9. \tag{189}$$

Applying Fano's inequality

We now apply External result 38 to the hypotheses $(\mu, \check\nu_k)$. Let $\widehat{\lambda}$ be an estimator, and let $s > 0$. For each k, the pair $(\mu, \check\nu_k)$ belongs to $\mathcal{U}_\alpha$, and $\lambda_{\mu,\check\nu_k} = \check\lambda_k$. Hence, the definition of $\mathfrak{M}_n^\lambda$ and Markov's inequality yield

$$\mathfrak{M}_n^\lambda(\widehat{\lambda}) \geq \max_{0\leq k\leq K}\mathbb{E}_{\mu,\check\nu_k}^n\left[\|\widehat{\lambda} - \check\lambda_k\|_{L^2(\mu)}^2\right]$$

$$\geq s^2 \max_{0\leq k\leq K} \mathbb{P}^n_{\mu,\check{\nu}_k}\big(\|\widehat{\lambda}-\check{\lambda}_k\|_{L^2(\mu)}\geq s\big). \tag{190}$$

Since $T_{\mu,\check{\nu}_k}=T_k$, the same argument yields the analogue of (190) for every estimator $\widehat{T}$, with $\mathfrak{M}^\lambda_n$, $\widehat{\lambda}$, and $\check{\lambda}_k$ replaced by $\mathfrak{M}^T_n$, $\widehat{T}$, and T_k. By (189), (176), and $K\geq 2^{N_J/8}\geq 2$, External result 38, applied with $d_\lambda(\lambda,\lambda'):=\|\lambda-\lambda'\|_{L^2(\mu)}$ and $2s=\sqrt{c_\lambda}J^{-\alpha}$, yields $\inf_{\widehat{\lambda}}\max_{0\leq k\leq K}\mathbb{P}^n_{\mu,\check{\nu}_k}(\|\widehat{\lambda}-\check{\lambda}_k\|_{L^2(\mu)}\geq s)\geq c$ for a constant $c>0$. Hence, (190) yields

$$\inf_{\widehat{\lambda}}\mathfrak{M}^\lambda_n(\widehat{\lambda})\gtrsim J^{-2\alpha}\asymp n^{-2\alpha/(2\alpha-2+d)}. \tag{191}$$

Similarly, by (189), (177), and $K\geq 2$, External result 38, applied with $d_T(T,T'):=\|T-T'\|_{L^2(\mu)}$ and $2s=(c_l\varepsilon_{\mathrm{b}}^2 c_\nabla/8)^{1/2}J^{-\alpha}$, yields $\inf_{\widehat{T}}\max_{0\leq k\leq K}\mathbb{P}^n_{\mu,\check{\nu}_k}(\|\widehat{T}-T_k\|_{L^2(\mu)}\geq s)\geq c$ for a constant $c>0$. Hence, the analogue of (190) for $\widehat{T}$ yields

$$\inf_{\widehat{T}}\mathfrak{M}^T_n(\widehat{T})\gtrsim J^{-2\alpha}\asymp n^{-2\alpha/(2\alpha-2+d)}. \tag{192}$$

Lower bounds based on two hypotheses

Define φ_0 as in (166) and $\mu=\mathrm{Unif}([0,1]^d)$. For $\widetilde{\theta}>0$ small, define

$$\varphi_*(x):=\varphi_0(x)+(\widetilde{\theta}/\sqrt{n})g(x),$$

and let (T_*,λ_*,ν_*) be defined by (168) with φ_k replaced by φ_*. For $\widetilde{\theta}$ small and n large, $f_*=\frac{1}{2}\|\cdot\|^2-\varphi_*$ satisfies $\nabla^2 f_*\succ 0$ on $\mathbb{R}^d$. Moreover, T_* maps $[0,1]^d$ onto itself, by the same argument as for T_k.

As in the multiple hypotheses proof, we rescale ν_* to the total mass $H_0=\nu_0(\Omega)$. Let $H_*:=\nu_*(\Omega)=\int\lambda_*^2\,d\mu$ and $c_*:=\frac{1}{2}\log(H_*/H_0)\in\mathbb{R}$, and define

$$\check{\lambda}_*:=e^{-c_*}\lambda_*,\qquad \check{\nu}_*:=(T_*)_\#(\check{\lambda}_*^2\mu)=\frac{H_0}{H_*}\,\nu_*,\qquad \bar{\nu}_*:=\frac{\nu_*}{H_*}.$$

Thus, $\check{\nu}_*$ has the normalized measure $\bar{\nu}_*$ and the total mass H_0. Moreover, Lemma 40 applied to φ_*+c_* identifies $(T_*,\check{\lambda}_*)$ as the μ-a.e. unique transport-growth pair associated with the pair $(\mu,\check{\nu}_*)$.

For all sufficiently large n, the function $\varphi_*-\varphi_0=(\widetilde{\theta}/\sqrt{n})g$ satisfies the last two bounds in (167):

$$\|\nabla^2((\widetilde{\theta}/\sqrt{n})g)\|_{\mathrm{op},\infty}\leq 1/2,\qquad \|(\widetilde{\theta}/\sqrt{n})g\|_{C^{\alpha+1}}\leq C_g.$$

Since $c_*\to 0$ as $n\to\infty$, we have $\check{\lambda}_*\to\lambda_0$ in $C^\alpha(\Omega)$ and the density of $\check{\nu}_*$ converges to h_0 in $C^{\alpha-1}(\Omega)$, so the constants M_0,B_0,R_0,κ_0,b_0 chosen as in the multiple hypotheses proof also place the pair $(\mu,\check{\nu}_*)$ in $\mathcal{U}_\alpha$ for all sufficiently large n.

For $\|T_0-T_*\|_{L^2(\mu)}$, we have

$$\|T_0-T_*\|^2_{L^2(\mu)}=(\widetilde{\theta}^2/n)\int\|\nabla g\|^2\,d\mu\gtrsim 1/n. \tag{193}$$

For the growth maps, let $v:[0,1]^d\to\mathbb{R}$ be $v:=(a/2)\partial_1 g-g$. Since $\nabla\varphi_0=(a,0,\dots,0)$ on $\mathrm{supp}(g)\subset[\delta_0,1-\delta_0]^d$, the computation in (170) with $\varphi_k-\varphi_0$ replaced by $(\widetilde{\theta}/\sqrt{n})g$ yields

$$\lambda_*(x)-\lambda_0(x)=\lambda_0(x)\,\frac{\widetilde{\theta}}{\sqrt{n}}\,v(x)+\mathcal{R}(x), \tag{194}$$

with $\|\mathcal{R}\|_\infty=O(n^{-1})$. Let $\bar{v}:=H_0^{-1}\int\lambda_0^2 v\,d\mu\in\mathbb{R}$. By (194),

$$H_*-H_0=2\frac{\widetilde{\theta}}{\sqrt{n}}\int\lambda_0^2 v\,d\mu+O(n^{-1}),\qquad\text{so}\qquad c_*=\frac{\widetilde{\theta}}{\sqrt{n}}\,\bar{v}+O(n^{-1}).$$

Hence,

$$\check{\lambda}_*-\lambda_0=\lambda_0\,\frac{\widetilde{\theta}}{\sqrt{n}}\,(v-\bar{v})+\check{\mathcal{R}},\qquad \|\check{\mathcal{R}}\|_\infty=O(n^{-1}). \tag{195}$$

We claim that the following quantity is strictly positive:

$$c_g^{\mathrm{par}} := \int_{[0,1]^d} (v - \bar{v})^2 \, dx.$$

Since $\mathrm{supp}(g) \subset [\delta_0, 1-\delta_0]^d$, v vanishes on $\partial[0,1]^d$. Thus, it suffices to show that v is not identically zero. Suppose for contradiction that $v \equiv 0$ on $[0,1]^d$. Then $\frac{a}{2}\partial_1 g = g$ on $[0,1]^d$. For every $(x_2, \dots, x_d) \in [0,1]^{d-1}$, solving the linear ODE $\frac{a}{2}\partial_1 g = g$ in x_1 yields

$$g(x_1, x_2, \dots, x_d) = g(0, x_2, \dots, x_d)\, e^{2x_1/a} = 0, \qquad x_1 \in [0,1],$$

contradicting $g(x_*, \dots, x_*) = \xi(x_*)^d \neq 0$. Thus, $c_g^{\mathrm{par}} > 0$ as claimed.

Squaring (195) and integrating against $d\mu = dx$ yields

$$\|\lambda_0 - \check{\lambda}_*\|_{L^2(\mu)}^2 \geq \lambda_{\min}^2 \frac{\widetilde{\theta}^2}{n} c_g^{\mathrm{par}} - O(n^{-3/2}) \gtrsim 1/n. \tag{196}$$

We now bound $D_{\mathrm{KL}}(\mathbb{P}^n_{\mu,\check{\nu}_*} \mid \mathbb{P}^n_{\mu,\nu_0}) = n D_{\mathrm{KL}}(\bar{\nu}_* \mid \bar{\nu}_0)$. Let $r_* := h_*/h_0$, where h_* is the density of ν_* with respect to μ, and let L_* be the function L_k with φ_k replaced by φ_*. Then (182) holds with (r_k, L_k, T_k) replaced by (r_*, L_*, T_*). Since $\|\varphi_* - \varphi_0\|_{C^2([0,1]^d)} \leq C\widetilde{\theta}/\sqrt{n}$ and the eigenvalues of $I - \nabla^2\varphi_0$ and $I - \nabla^2\varphi_*$ are at least $1/4$, we have

$$\|L_* - L_0\|_\infty \leq \frac{C\widetilde{\theta}}{\sqrt{n}}.$$

Moreover, (186) holds with T_k replaced by T_*. The argument for (172) also yields

$$\|T_*^{-1}(y) - T_0^{-1}(y)\| \leq \frac{C\widetilde{\theta}}{\sqrt{n}}, \qquad y \in \Omega.$$

Thus, for all sufficiently large n,

$$\|\log r_*\|_\infty \leq \frac{C\widetilde{\theta}}{\sqrt{n}} \leq \frac{1}{2}.$$

Since $\|\log \lambda_* - \log \lambda_0\|_\infty \leq C\widetilde{\theta}/\sqrt{n} \leq \log 2$ for all sufficiently large n by (194), we have $\lambda_{\min}/2 \leq \lambda_* \leq 2\lambda_{\max}$, and $H_* = \int \lambda_*^2 \, d\mu$ implies $\lambda_{\min}^2/4 \leq H_* \leq 4\lambda_{\max}^2$. Hence, (179) and (181) hold with $(\bar{\nu}_k, h_k, H_k, r_k)$ replaced by $(\bar{\nu}_*, h_*, H_*, r_*)$ and imply that

$$D_{\mathrm{KL}}(\bar{\nu}_* \mid \bar{\nu}_0) \leq C \int (\log r_*)^2 \, d\mu \leq \frac{C\widetilde{\theta}^2}{n}.$$

Hence, $D_{\mathrm{KL}}(\mathbb{P}^n_{\mu,\check{\nu}_*} \mid \mathbb{P}^n_{\mu,\nu_0}) = n D_{\mathrm{KL}}(\bar{\nu}_* \mid \bar{\nu}_0) \leq C\widetilde{\theta}^2$. For all sufficiently large n, the pairs (μ, ν_0) and $(\mu, \check{\nu}_*)$ belong to $\mathcal{U}_\alpha$, with the transport-growth pairs (T_0, λ_0) and $(T_*, \check{\lambda}_*)$, respectively. Thus, (190) and the analogue of (190) for $\widehat{T}$ hold with the maximum taken over the hypotheses (μ, ν_0) and $(\mu, \check{\nu}_*)$. By External result 37, applied with $\varrho = C\widetilde{\theta}^2$, we obtain

$$\inf_{\widehat{T}} \max\Big\{ \mathbb{P}^n_{\mu,\nu_0}\big(\|\widehat{T} - T_0\|_{L^2(\mu)} \geq s_T\big),\, \mathbb{P}^n_{\mu,\check{\nu}_*}\big(\|\widehat{T} - T_*\|_{L^2(\mu)} \geq s_T\big) \Big\} \geq c_\varrho,$$

$$\inf_{\widehat{\lambda}} \max\Big\{ \mathbb{P}^n_{\mu,\nu_0}\big(\|\widehat{\lambda} - \lambda_0\|_{L^2(\mu)} \geq s_\lambda\big),\, \mathbb{P}^n_{\mu,\check{\nu}_*}\big(\|\widehat{\lambda} - \check{\lambda}_*\|_{L^2(\mu)} \geq s_\lambda\big) \Big\} \geq c_\varrho,$$

where $s_T := \frac{1}{2}\|T_0 - T_*\|_{L^2(\mu)}$ and $s_\lambda := \frac{1}{2}\|\lambda_0 - \check{\lambda}_*\|_{L^2(\mu)}$, which, by (193) and (196), satisfy $s_T^2 \gtrsim 1/n$ and $s_\lambda^2 \gtrsim 1/n$. Combining the previous display with the analogue of (190) for $\widehat{T}$ at $s = s_T$ and with (190) at $s = s_\lambda$, we obtain

$$\inf_{\widehat{T}} \mathfrak{M}_n^T(\widehat{T}) \gtrsim 1/n, \tag{197}$$

$$\inf_{\widehat{\lambda}} \mathfrak{M}_n^\lambda(\widehat{\lambda}) \gtrsim 1/n. \tag{198}$$

Combining (192), (191), (197), and (198) yields Theorem 5.

□

J Extensions to other costs

The analysis of this paper uses the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$ of Section 2. In this section, we identify the steps of the analysis that use the quadratic form of c and state what a general cost would need at each step, with the Hellinger-Kantorovich cost as an example.

Remark 5 (General costs)**.** Let $\widetilde{c} : \Omega \times \Omega \to \mathbb{R}$ be a cost function, and consider the problem (2) with c replaced by $\widetilde{c}$. An extension of the upper bounds of Theorems 3 and 4 to $\widetilde{c}$ would require at least five ingredients at the steps that use the quadratic form of c:

(i) $\widetilde{c} \in C^2(\Omega \times \Omega)$ with bounded second derivatives, for an analogue of the bound (37) in Lemma 13, which is used in the proof of Proposition 8;

(ii) an analogue of Theorem 1 for $\widetilde{c}$ under the twist condition, for the transport-growth pair (T_0, λ_0);

(iii) for every $x \in \Omega$, the strong convexity of the map $y \mapsto \widetilde{c}(x,y) - \psi_0(y)$ on Ω, for the gap bound of Lemma 2;

(iv) constants $0 < \sigma \le L < \infty$ such that $\sigma I \preceq \nabla_y^2 \widetilde{c}(y,y') \preceq LI$ and $\nabla_y \widetilde{c}(y,y')|_{y=y'} = 0$ for all $y, y' \in \Omega$, for the first inequality in (43) in Proposition 15;

(v) an analogue of (18), whose proof uses $T_0 = \nabla f_0$, for the Lipschitz gradient of ξ_0 in Lemma 12, which is used in the proof of Proposition 8.

Under (iv), let $\widehat{T}_i$ be the minimizer in (5) with c replaced by $\widetilde{c}$. From (iv), we have $0 \le \widetilde{c}(y, Y_j) - \widetilde{c}(Y_j, Y_j) \le \frac{L}{2}\|y - Y_j\|^2$ for all $y \in \Omega$ and $j \in \{1, \dots, m\}$. For every $i \in \{1, \dots, n\}$ with $\widehat{r}_i > 0$, combining the upper bound at $y = T_0(X_i)$ and the lower bound at $y = \widehat{T}_i$ with the $\sigma\widehat{r}_i$-strong convexity of the map $y \mapsto \sum_{j=1}^m \widehat{\gamma}_{ij}\widetilde{c}(y, Y_j)$ on Ω yields

$$\widehat{r}_i\|\widehat{T}_i - T_0(X_i)\|^2 \le \frac{L}{\sigma}\sum_{j=1}^m \widehat{\gamma}_{ij}\|Y_j - T_0(X_i)\|^2. \tag{199}$$

For the quadratic cost, $\sigma = L = 1$. Hence, summing (199) over the indices i with $\widehat{r}_i > 0$ yields the first inequality in (43) in Proposition 15. In contrast, the lower bound of Theorem 5 needs more than (i)–(v). The construction behind this lower bound uses the explicit forms (4) of T_0 and λ_0, which hold for the quadratic cost. For a general cost $\widetilde{c}$, both forms change, so the separation estimates in the proof of Theorem 5 would have to be derived again.

Remark 6 (Hellinger-Kantorovich cost)**.** The Hellinger-Kantorovich problem [LMS18] is the UOT problem with KL penalties and the cost $c_{\mathrm{HK}} : \Omega \times \Omega \to [0, \infty]$, $c_{\mathrm{HK}}(x,y) := -\log\cos^2(\|x-y\| \wedge \pi/2)$. Suppose that $D := \operatorname{diam}(\Omega) < \pi/2$. Then c_{HK} is finite and smooth on $\Omega \times \Omega$ and satisfies

$$\|x-y\|^2 \le c_{\mathrm{HK}}(x,y) \le C_D\|x-y\|^2, \qquad x, y \in \Omega,$$

with the constant $C_D := -2\log\cos(D)/D^2 > 1$. These bounds hold because $-\log\cos^2 t = t^2 + t^4/6 + \cdots$ has nonnegative Taylor coefficients for $t \in [0, \pi/2)$. Moreover, $2I \preceq \nabla_y^2 c_{\mathrm{HK}}(y,y') \preceq 2\sec^2(D)\, I$ and $\nabla_y c_{\mathrm{HK}}(y,y')|_{y=y'} = 0$ for all $y, y' \in \Omega$. At $t = \|y - y'\| \in [0, D]$, the eigenvalues of $\nabla_y^2 c_{\mathrm{HK}}(y,y')$ are $2\sec^2 t$ in the direction $y - y'$ and $2\tan(t)/t$ in the orthogonal directions. Both eigenvalues lie in $[2, 2\sec^2(D)]$, which yields the Hessian bounds. Thus, c_{HK} satisfies (i) and (iv) in Remark 5 with $\sigma = 2$ and $L = 2\sec^2(D)$. In addition, [GGV25] proves the analogue of Theorem 1 for c_{HK} on a smooth compact Riemannian manifold, which is (ii) in that setting. On the other hand, (iii) and (v) in Remark 5 remain open. We expect these two properties to follow from the Hellinger-Kantorovich analogue of Assumption 1, in which case the upper bounds of Theorems 3 and 4 would hold with the same rates. For the lower bound, the forms (4) change for c_{HK} as well. However, the construction behind Theorem 5 moves mass over distances much smaller than $\pi/2$, where $c_{\mathrm{HK}}(x,y) = \|x-y\|^2 + O(\|x-y\|^4)$, so we expect the same minimax rate.

K The Wasserstein distance between measures of equal total mass

We present here the definition of the Wasserstein distance between finite positive measures of equal total mass and the effect of mass scaling on the Wasserstein distances.

Definition 5. Let $p \in [1, \infty)$ and let $\eta, \sigma \in \mathcal{M}_+(\Omega)$ satisfy $\eta(\Omega) = \sigma(\Omega) < \infty$. We define the p-Wasserstein distance between η and σ by

$$W_p(\eta, \sigma) := \left(\inf_{\gamma \in \Pi(\eta,\sigma)} \int_{\Omega \times \Omega} \|x - y\|^p \, d\gamma(x, y) \right)^{1/p},$$

where $\Pi(\eta, \sigma)$ denotes the set of couplings of η and σ. When η and σ are probability measures, this is the usual Wasserstein distance.

Lemma 41 (Scaling in the total mass)**.** *Let $m > 0$ and let η, σ be probability measures on Ω. Then*

$$W_p^p(m\eta, m\sigma) = m W_p^p(\eta, \sigma), \qquad W_p(m\eta, m\sigma) = m^{1/p} W_p(\eta, \sigma). \tag{200}$$

Proof of Lemma 41. The map $\gamma \mapsto m\gamma$ is a bijection from $\Pi(\eta, \sigma)$ onto $\Pi(m\eta, m\sigma)$. Therefore,

$$\inf_{\gamma \in \Pi(m\eta, m\sigma)} \int \|x - y\|^p \, d\gamma(x, y) = m \inf_{\gamma \in \Pi(\eta,\sigma)} \int \|x - y\|^p \, d\gamma(x, y).$$

Taking the pth root yields the second identity in (200). □

To give some context on the rates of our nearest-neighbor estimator in Theorem 3, we provide below a Wasserstein lower bound for discrete approximation of measures, which is expressed in the notion of the *lower Wasserstein dimension* of [You82, WB19]: for a probability measure η on Ω and $\tau \in [0, 1)$, let $N_\varepsilon(S)$ be the minimum number of closed balls of diameter ε needed to cover S, and define

$$N_\varepsilon(\eta, \tau) := \inf\{N_\varepsilon(S) : S \subset \Omega,\ \eta(S) \geq 1 - \tau\}, \qquad d_\varepsilon(\eta, \tau) := \frac{\log N_\varepsilon(\eta, \tau)}{-\log \varepsilon},$$
$$d_*(\eta) := \lim_{\tau \downarrow 0} \liminf_{\varepsilon \downarrow 0} d_\varepsilon(\eta, \tau).$$

Here, $d_*(\bar{\mu})$ is referred to as the lower Wasserstein dimension of $\bar{\mu}$.

Proposition 42 (Lower bound for discrete approximations)**.** *Let $\mu \in \mathcal{M}_+(\Omega)$ with $\mu(\Omega) = M_\mu \in (0, \infty)$, and let $\bar{\mu} := \mu / M_\mu$. If $t < d_*(\bar{\mu})$ and $\eta_n \in \mathcal{M}_+(\Omega)$ is any discrete measure with $\eta_n(\Omega) = M_\mu$ and $|\mathrm{supp}(\eta_n)| \leq n$, then*

$$W_p(\mu, \eta_n) \gtrsim (M_\mu)^{1/p} n^{-1/t}. \tag{201}$$

In particular, the same lower bound holds for the empirical measure $\eta_n = \widetilde{\mu}_n$ with mass M_μ. Under Assumption 3, (201) holds for every $t < d$.

Proof of Proposition 42. Let $\bar{\eta}_n := \eta_n / M_\mu$. Then $\bar{\eta}_n$ is a probability measure supported on at most n points. The lower bound in Theorem 1 of [WB19] and the remark after the theorem imply that for every $t < d_*(\bar{\mu})$,

$$W_p(\bar{\mu}, \bar{\eta}_n) \gtrsim n^{-1/t}.$$

Applying Lemma 41 yields

$$W_p(\mu, \eta_n) = (M_\mu)^{1/p} W_p(\bar{\mu}, \bar{\eta}_n) \gtrsim (M_\mu)^{1/p} n^{-1/t}.$$

Taking $\eta_n = \widetilde{\mu}_n$ proves the claim for $\widetilde{\mu}_n$. For the final claim, Assumption 3 gives $\bar{\mu}$ a Lebesgue density on Ω. Hence, every Borel set $S \subset \Omega$ with $\bar{\mu}(S) > 0$ has positive Lebesgue measure, so the Hausdorff dimension of S is d. Consequently, the Hausdorff dimension of $\bar{\mu}$ is d, and [WB19, Proposition 2] yields $d_*(\bar{\mu}) \geq d$. Hence, (201) holds for every $t < d$. □

L Details of the simulation

L.1 Computational resources

The experiments were run on a Linux computing node with two 4-core Intel(R) Xeon(R) Gold 5222 processors and a single NVIDIA Tesla V100-SXM2 GPU, utilizing 32 GB of VRAM.

L.2 Simulation and implementation details

We generate source and target samples of size $n \in \{100, 200, 500, 1000\}$ over 10 random seeds. For PB-1NN and PB-Kernel, we solve the discrete UOT problem, framed as nonnegative regularized linear regression as described in [CFW+21], using the UOT solver from the Python Optimal Transport library [FCG+21]. For PI-Kernel and PI-Wavelet, we solve the dual problem (3) for $(\widehat{\mu}_n^{\mathrm{ker}}, \widehat{\nu}_m^{\mathrm{ker}})$ and $(\widehat{\mu}_n^{\mathrm{wav}}, \widehat{\nu}_m^{\mathrm{wav}})$ using the L-BFGS-B algorithm [LN89] over the discretized grid with resolution $128, 64, 32, 32$ for $d = 1, 2, 3, 4$, respectively. For each $n \in \{100, 200, 500, 1000\}$, we select the frequency cutoff L_n by 10-fold least-squares cross-validation on a fresh sample of size n, over the range $[6, L_{\max}]$ with $L_{\max} = 100, 60, 30, 20$ for $d = 1, 2, 3, 4$. In contrast, we set the wavelet resolution level by the choice $2^{J_n} \asymp n^{1/(d+2(\alpha-1))}$ as specified in Theorem 29, namely $J_n = \lceil \log_2 n/(d + 2(\alpha - 1)) \rceil$ with $\alpha = 2$. We implement PI-Wavelet with the db4 stationary wavelet transform, even reflection at the boundary, and soft thresholding of the detail coefficients $\widehat{\alpha}_\xi$ and $\widehat{\beta}_\xi$. Finally, the cone program in the SSUOT estimator was solved using the CVXPY library [DB16] with MOSEK solver [MOS24].

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Introduction follows the technical result of the main body of this paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It is stated in the conclusion section.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide rigorous proof in Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The experimental details are shown in the experiments section and Appendix. We also attach a code as a supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The data is open and we refer to the link. The code will be opened.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: The experimental details are shown in the experiments section and Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We have included the error bars in our experimental plot, while it is too small and hard to see it because our result is very stable.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The experiment ran on a small laptop, so there was no need to set up a large-scale environment.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We have followed the code of ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: These results are fundamental in nature and focus on theory; they do not have a direct impact on society.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Our models and methods do not have such a high risk of misuse, since it is very fundamental and theoretical.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We have cited every papers and works used in this paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We mention our new code and attach them as supplementary materials.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: We do not use crowdsourcing and research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The use of LLMs in this paper was strictly limited to writing, editing, and formatting assistance, which does not require a declaration under the current policy.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.