

SPARSE REGULARIZED OPTIMAL TRANSPORT WITHOUT CURSE OF DIMENSIONALITY

BY ALBERTO GONZÁLEZ-SANZ^{1,a}, STEPHAN ECKSTEIN^{2,b} AND MARCEL NUTZ^{3,c}

¹*Department of Statistics, Columbia University* ^aag4855@columbia.edu

²*Department of Mathematics, University of Tübingen* ^bstephan.eckstein@uni-tuebingen.de

³*Departments of Statistics and Mathematics, Columbia University* ^cmmutz@columbia.edu

Entropic optimal transport—the optimal transport problem regularized by KL divergence—is highly successful in statistical applications. Thanks to the smoothness of the entropic coupling, its sample complexity avoids the curse of dimensionality suffered by unregularized optimal transport. The flip side of smoothness is overspreading: the entropic coupling always has full support, whereas the unregularized coupling that it approximates is usually sparse, even given by a map. Regularizing optimal transport by less-smooth f -divergences such as Tsallis divergence (i.e., L^p -regularization) is known to allow for sparse approximations, but is often thought to suffer from the curse of dimensionality as the couplings have limited differentiability and the dual is not strongly concave. We refute this conventional wisdom and show, for a broad family of divergences, that the key empirical quantities converge at the parametric rate, independently of the dimension. More precisely, we provide central limit theorems for the optimal cost, the optimal coupling, and the dual potentials induced by i.i.d. samples from the marginals. These results are obtained by a powerful yet elementary approach that is of broader interest for Z-estimation in function classes that are not Donsker.

1. Introduction. Optimal transport has become ubiquitous following computational advances enabling applications in statistics, machine learning, image processing, and other domains where distributions or data sets need to be compared (e.g., [37, 43]). Given probability measures P and Q on \mathbb{R}^d , and a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, the Monge–Kantorovich optimal transport problem is

$$(1) \quad \text{OT}(P, Q) = \inf_{\pi \in \Pi(P, Q)} \int c(x, y) d\pi(x, y)$$

where $\Pi(P, Q)$ denotes the set of couplings; i.e., joint distributions π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals P and Q . A key bottleneck for statistical applications is that optimal transport suffers from the curse of dimensionality in terms of sample complexity (see [4, 12, 16] among many others). More specifically, let X_1, \dots, X_n and Y_1, \dots, Y_n be i.i.d. samples from P and Q , respectively, and consider their empirical measures $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $Q_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$. Then $\text{OT}(P_n, Q_n)$, the value of (1) computed with marginals (P_n, Q_n) instead of (P, Q) , converges to the population value $\text{OT}(P, Q)$ at a rate that deteriorates exponentially with the dimension d . For instance, for the important cost $c(x, y) = \|x - y\|^2$ defining the 2-Wasserstein distance, $\mathbb{E}[|\text{OT}(P_n, Q_n) - \text{OT}(P, Q)|] \sim n^{-2/d}$ for dimension $d \geq 5$, under regularity conditions on the population marginals P and Q (see [33] for this particular result).

MSC2020 subject classifications: Primary: 62G05 Secondary: 62R10, 62G30.

Keywords and phrases: Optimal Transport, Regularization, Sample Complexity, Central Limit Theorem.

By contrast, the celebrated entropy-regularized optimal transport (EOT) avoids the curse of dimensionality. The EOT problem penalizes (1) with the Kullback–Leibler (KL) divergence $\text{KL}(\pi|P \otimes Q)$ between the coupling π and the product $P \otimes Q$ of the marginals,

$$(2) \quad \text{EOT}_\varepsilon(P, Q) := \inf_{\pi \in \Pi(P, Q)} \int cd\pi + \varepsilon \text{KL}(\pi|P \otimes Q)$$

where $\varepsilon > 0$ is a fixed parameter determining the strength of regularization. Indeed, a series of works starting with [18] has shown that $\text{EOT}_\varepsilon(P_n, Q_n)$ converges to $\text{EOT}_\varepsilon(P, Q)$ at the parametric rate $n^{-1/2}$, thus the dimension d (and the parameter ε) only affect the constants. Literature is reviewed in Section 1.1 below. The prevailing thinking is that EOT overcomes the curse of dimensionality thanks to its smoothness. Indeed, the entropic penalty leads to optimal couplings whose density is smooth (or at least as smooth as the cost c), and this regularity holds independently of the marginals. In particular, it holds uniformly over the empirical measures (P_n, Q_n) , and this fact drives the aforementioned result of [18].

The flip side of smoothness is that the optimal coupling of EOT always has full support (i.e., the same support as $P \otimes Q$), for any value of the regularization parameter $\varepsilon > 0$. By contrast, the unregularized optimal transport coupling that it approximates, typically has sparse support (the graph of a function, namely the Monge map). This disconnect can be undesirable depending on the application; for instance, the large support (or “overspreading”) can amount to blurrier images in an image processing task as shown in [2], bias in a manifold learning task as in [45], or unfaithful approximation of barycenters [29]. In such cases, sparse approximations are desirable.

It is known that less-smooth f -divergence penalties give rise to sparse approximations. Indeed, consider the divergence-regularized optimal transport problem with the divergence $D_\varphi(\pi|P \otimes Q)$ defined by a function φ ,

$$(3) \quad \text{ROT}_\varepsilon(P, Q) := \inf_{\pi \in \Pi(P, Q)} \int cd\pi + \varepsilon D_\varphi(\pi|P \otimes Q),$$

$$(4) \quad D_\varphi(\pi|P \otimes Q) := \int \varphi \left(\frac{d\pi}{d(P \otimes Q)} \right) d(P \otimes Q),$$

where $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is convex and satisfies certain conditions (see Assumption 2.1). While the KL divergence of EOT is recovered for $\varphi(t) = t \log t$, replacing this by a power $\varphi(t) = (p-1)^{-1}(t^p - 1)$ for $p > 1$ (Tsallis divergence, including the quadratic or χ^2 divergence for $p = 2$) was studied starting with [35, 2, 15] and empirically seen to have sparse solutions. Recent investigations underline this finding with theoretical results; see the literature review below. The main objection against such regularizations is that they lack the smoothness of EOT. Indeed, the density of the optimal coupling is only as smooth as the derivative of the convex conjugate $\psi(s) = \sup_{t \geq 0} \{st - \varphi(t)\}$ of φ , which for $\varphi(t) = (p-1)^{-1}(t^p - 1)$ is k -times differentiable only for $k < p/(p-1)$. In particular, it does not enjoy the C^∞ -smoothness crucially used in EOT, where the proof of sample complexity uses derivatives of higher and higher order as the dimension d grows. Following the aforementioned prevailing thinking, it is often assumed that (3) thus suffers from the curse of dimensionality when it is not smooth. This idea is in line with the recent work [1] which gives an upper bound for the sample complexity that deteriorates exponentially with the dimension d (see Section 1.1).

In the present paper, our goal is to refute this prevailing thinking. For a broad class of regularizations φ , we show that (3) overcomes the curse of dimensionality and converges at the parametric rate, despite a lack of smoothness (and also of a PL inequality, cf. Section 1.1). In fact, our results are much more detailed: we establish central limit theorems for all key objects. As a consequence, (3) provides approximate solutions to optimal transport with precise statistical guarantees.

Synopsis of main results. We consider marginals P, Q with bounded supports, at least one of them connected. The transport cost c is a general \mathcal{C}^1 function. The main assumption on the divergence is that the convex conjugate ψ of φ is \mathcal{C}^2 . This includes the KL divergence where ψ is \mathcal{C}^∞ , but also Tsallis divergences that yield sparse approximations of unregularized optimal transport. There are three main objects under consideration. The first is the optimal cost $\text{ROT}_\varepsilon(P_n, Q_n)$; here the result is simple to state as the object is scalar. In Theorem 3.5, we show that the optimal cost is asymptotically normal at rate \sqrt{n} . More precisely, $\sqrt{n}(\text{ROT}(P_n, Q_n) - \text{ROT}(P, Q)) \rightarrow N(0, \sigma^2)$ where $N(0, \sigma^2)$ is the centered normal distribution with a variance σ^2 detailed in Theorem 3.5. The second key object is the optimal coupling $\pi_n \in \Pi(P_n, Q_n)$. Here the asymptotic normality can be stated by integrating a test function: Theorem 3.6 shows that for any bounded measurable function η , we have $\sqrt{n}(\int \eta d(\pi_n - \pi)) \rightarrow N(0, \sigma^2(\eta))$, where again the variance is detailed in the theorem. The third object is the pair of dual potentials; that is, functions (f_*, g_*) solving the dual problem of (3) in the sense of convex analysis,

(5)

$$\sup_{(f,g) \in L^\infty(P) \times L^\infty(Q)} \int \left\{ f(x) + g(y) - \varepsilon \cdot \psi \left(\frac{f(x) + g(y) - c(x,y)}{\varepsilon} \right) \right\} d(P \otimes Q)(x, y).$$

The potentials determine the optimal coupling via

$$d\pi = \psi' \left(\frac{f_*(x) + g_*(y) - c(x,y)}{\varepsilon} \right) d(P \otimes Q)$$

where ψ' is the derivative of the conjugate of φ . Unlike in the special case of EOT, ψ' is generally not invertible, and hence the potentials are the main object in our study. Denoting by (f_*, g_*) the population potentials and by (f_n, g_n) the empirical counterparts, our central limit theorem for the potentials states that $\sqrt{n}(f_n - f_*, g_n - g_*)$ converges (weakly wrt. the uniform norm) to the Gaussian random process detailed in Theorem 3.4. Of course this result can only hold if (f_*, g_*) is unique. Indeed, we obtain uniqueness (up to additive constants) as a by-product of our analysis; this is a new result of its own interest and holds even for merely continuous transport costs c (Theorem 3.2).

1.1. *Related literature.* Next, we review literature on optimal transport regularized by an f -divergence. Following the main focus of the present work, we begin with divergences other than the entropic one. After that, we review the existing sample complexity results for EOT.

1.1.1. *Non-entropic divergences.* Optimal transport problems regularized by Tsallis divergence were first considered in the discrete setting, starting with [35] for applications in ecological inference. Focusing on the particular case of quadratic (i.e., χ^2) divergence, [2] highlighted the sparsity of the solution for use in image processing and [15] studied minimum-cost flow problems on graphs. The earlier work [10] considered optimal transport with a more general convex regularization. In the continuous setting, divergence-regularized optimal transport was first explored in the computational literature. Several works including [13, 18, 26, 29, 40] approach the dual problem by optimization techniques. For instance, [29] computes regularized Wasserstein barycenters using neural networks. From a computational point of view, Tsallis divergences are attractive to mitigate a well-known issue of EOT for small values of the regularization parameter ε , namely that optimization methods struggle with the exponentially small values of the density (e.g., [39]).

On the theoretical side, [31] established duality results for the case of quadratic divergence. Also for quadratic divergence, [30] showed convergence $\text{ROT}_\varepsilon \rightarrow \text{OT}$ as the regularization parameter ε tends to zero. More recently, [14] derives a rate for this convergence, for general

f -divergences. For the special case of quadratic divergence, [17] shows that this rate corresponds to the exact leading order and identifies the multiplicative constant, whereas [24] focuses on the discrete case where convergence is stationary. The recent works [25, 44] study the size of the optimal support in the regime $\varepsilon \rightarrow 0$, thus quantifying the sparsity of the solution (qualitatively proved in [36] but empirically known long before, as mentioned above).

The only previous work on sample complexity for non-entropic divergences is [1]. The authors bound the expected absolute difference between the population and empirical optimal costs. The bound depends on the smoothness and the dimension of the problem. Recall that ψ denotes the conjugate of the convex function φ defining the divergence in (4). Let $k \in \mathbb{N}$ be such that $\psi \in \mathcal{C}^k$ as well as $c \in \mathcal{C}^k$. Moreover, let d be the dimension of the marginals. The main result [1, Theorem 5.3] states that for $d > 2k$,

$$(6) \quad \mathbb{E}[|\text{ROT}_\varepsilon(P, Q) - \text{ROT}_\varepsilon(P_n, Q_n)|] \lesssim n^{-\frac{k}{d}}.$$

For most applications, the bottleneck in this bound is the smoothness of ψ , rather than c . Assuming merely $\psi \in \mathcal{C}^2$ as in the present work, the rate (6) is $n^{-\frac{2}{d}}$ for $d \geq 5$, suggesting that the curse of dimensionality is exactly *the same as for the unregularized* optimal transport problem mentioned at the beginning of this introduction. The bound (6) is obtained by adapting the classical approach going back to [18] for EOT: estimating the regularity of the empirical potentials to define a function space with controlled covering number and then applying empirical process theory. Because the potentials are only as smooth as ψ' , this yields a bound (6) that in general deteriorates exponentially with the dimension (whereas when $\psi \in \mathcal{C}^\infty$ as for EOT, the dimension can be eliminated).

The present work indicates that the bound (6) is loose, even to the extent that its implicit message regarding the curse of dimensionality is misleading. Indeed, we show that the actual rate is $n^{-\frac{1}{2}}$, *independently of the dimension d* , and in particular that the influence of the dimension is more similar to EOT than to unregularized optimal transport. One may consider this result highly surprising given the lack of smoothness (and failure of strong concavity, cf. Section 1.1.2). As we derive central limit theorems with the usual \sqrt{n} scaling for all key quantities, this rate is sharp. (Except possibly in degenerate examples where the Gaussian limit has vanishing variance; then, the actual rate is likely faster.)

1.1.2. Entropic divergence. The literature on EOT is extremely large, and fairly well known. We only review the literature on sample complexity, focusing on continuous population marginals (and with a view towards explaining why the existing approaches fail in our setting). As mentioned above, parametric rates for the optimal EOT cost (2) were first obtained by [18]. Assuming a smooth (i.e., \mathcal{C}^∞) transport cost c and compactly supported marginals, the authors show that the potentials are smooth with \mathcal{C}^k norm bounded independently of the marginal measures, for any $k \in \mathbb{N}$. Thus the empirical potentials all belong to a function class with controlled covering number. The result then follows by applying empirical process theory to the dual problem of EOT. A similar bound with improved constants and more general (sub-Gaussian) marginals was obtained in [34] using a refinement of the same approach. Moreover, [34] provided the first central limit-type theorem on EOT, namely for the optimal cost. In this result, the centering is by the mean of the empirical cost, instead of the population cost as in a usual central limit theorem. The proof follows an approach based on the Efron–Stein inequality that is adapted from (unregularized) optimal transport; cf. [7]. A central limit theorem for the EOT cost in the classical sense (centered at the population value) was first derived in [9]. The proof combines the result of [34] with an analysis of the bias–variance decomposition showing that the bias tends to zero faster than $n^{-1/2}$. This is based on the fact that the empirical EOT potentials belong to a Donsker class. Namely, the \mathcal{C}^k -norm of the potentials is uniformly bounded for any k , and choosing $k > d/2$ implies

the Donsker property [42, p. 157]. The authors also show that the potentials converge at the parametric rate (in the norm of \mathcal{C}^k , for any $k \in \mathbb{N}$); this is based on the strong convexity of the dual problem (see [9, Lemma 4.6]). In concurrent work, [20] provided a very similar central limit theorem for the EOT cost, using a different proof technique based on the functional delta method for supremum-type functionals (see [6]). This argument again rests on the potentials belonging to a Donsker class.

While the aforementioned central limit theorems were all about the optimal cost, central limit theorems for the potentials and couplings were first derived in [23] and [21]. Both works use the same methodology of Z-estimation and the delta method (see also Section 1.2 below). Specifically, [21, Theorem 3] shows that the potentials are Hadamard differentiable as functions of the marginal measures, tangentially to perturbations in $\mathcal{C}^k(\Omega)'$ for any $k \in \mathbb{N}$. As the unit ball in $\mathcal{C}^k(\Omega)$ is a Donsker class for $k > d/2$, the delta method then yields the central limit theorem for the potentials, and the theorems for the couplings are derived from there.

The aforementioned works all deal with a transport cost function $c \in \mathcal{C}^\infty$, and exploit the fact that the EOT potentials are as smooth as that cost. A substantially different approach was taken in [38]. Assuming only that c is bounded, the approach is based on the fact that the dual problem of EOT is strongly concave, and hence satisfies a PL inequality, uniformly over the empirical marginals (see [38, Lemma 16]). The authors then obtain the parametric rate for the convergence of the empirical EOT cost; more precisely, the mean squared error and bias are both bounded by $1/n$, with constants that are fully dimension-independent. In the setting of non-smooth cost, central limit theorems for the optimal costs and the optimal couplings were established in [22]. The authors linearize the potentials in the empirical L^2 norms and, unlike [21, 23], do not use empirical processes theory but instead approximate the optimal couplings by infinite order U-statistics. This approximation holds by a uniform contraction argument over the linearized Schrödinger system, which is deeply related to the PL inequality in [38].

All of the above approaches fail in our setting because the empirical potentials are not smooth (and do not belong to a Donsker class) while the dual problem is not strongly concave and does not satisfy a uniform PL inequality.

While our main interest is in divergences other than the entropic one, let us mention that we add to the literature even in the special case of EOT. Namely, we provide a central limit theorem for the potentials under the assumption that the cost c is \mathcal{C}^1 , where such a result previously existed only for $c \in \mathcal{C}^\infty$ (see [21, 23]). Our proof technique is substantially different, as we explain next.

1.2. Methodology. We follow the approach of Z-estimation in deriving central limit theorems; see [42, Chapter 3.3]. While the usual argument via Donsker classes is doomed to fail due to the missing regularity of the empirical potentials, one key methodological innovation is to overcome this issue with a novel line of argument. As our approach may be useful for Z-estimation problems in other areas, we sketch the approach in general terms.

In Z-estimation, the empirical quantities θ_n of interest are described by an equation of the form $\Phi_n(\theta_n) = 0$ with a random operator Φ_n while the population counterpart θ_* is described by $\Phi(\theta_*) = 0$ with a deterministic Φ . The goal is to show a central limit theorem for the convergence $\theta_n \rightarrow \theta_*$. More specifically, θ_n, θ_* are elements of the parameter set Θ which is contained in a Banach space $(\mathcal{B}, \|\cdot\|)$. We assume for simplicity that $\Phi_n, \Phi : \mathcal{B} \rightarrow \mathcal{B}$; in general the image may be in another Banach space. In our particular problem, θ_n are the empirical potentials (f_n, g_n) and θ_* is the population counterpart (f_*, g_*) . The nonlinear operators Φ_n, Φ represent the first-order conditions of optimality in the dual problem. As in

many applications of Z-estimation, these operators have the integral form

$$(7) \quad \Phi_n(\theta) = \int \phi(\theta) d\mathbb{P}_n, \quad \Phi(\theta) = \int \phi(\theta) d\mathbb{P}$$

where ϕ is deterministic and \mathbb{P}_n are empirical measures derived from i.i.d. samples of \mathbb{P} .

The basic theorem of Z-estimation (see [42, Theorem 3.3.1]) separates the conditions for the desired asymptotic normality of $\sqrt{n}(\theta_n - \theta_*)$ into analytical conditions on the population quantities θ_* , Φ and a stochastic condition stating that the remainder in a Taylor expansion is negligible. Roughly, these conditions are

- (i) $[\Phi_n - \Phi](\theta_*)$ satisfies a central limit theorem in \mathcal{B} ,
- (ii) Φ is Fréchet differentiable at θ_* with invertible derivative $\mathbb{L} := D\Phi(\theta_*) \in \mathcal{L}(\mathcal{B}, \mathcal{B})$,
- (iii) $\|\theta_n - \theta_*\| \xrightarrow{\mathbb{P}} 0$,
- (iv) the following expansion holds,

$$(8) \quad \Delta_n := [\Phi_n - \Phi](\theta_n) - [\Phi_n - \Phi](\theta_*) = o_{\mathbb{P}}(n^{-1/2} + \|\theta_n - \theta_*\|).$$

While obtaining invertibility has its own challenges in our setting, we defer that discussion to Section 3.1. Like in most applications, the main issue is to establish (8). The standard approach (see [42, Lemma 3.3.5]) is to use a sufficient condition whose main part is that the random functions $\phi(\theta_n) - \phi(\theta_*)$ form a \mathbb{P} -Donsker class (for $\|\theta_n - \theta_*\|$ sufficiently small). In the setting of EOT, the smoothness of the empirical potentials indeed yields the Donsker property, and together with the aforementioned Hadamard differentiability, this allowed [21, 23] to infer the desired central limit theorem for the potentials. In the present case, this approach is a nonstarter because the regularity of the potentials is too poor to give a Donsker class.

We take a different route. While our aim is, as before, a central limit theorem in \mathcal{B} , we also use an auxiliary Banach space $\mathcal{B}_s \subset \mathcal{B}$ equipped with a stronger norm $\|\cdot\|_s$ such that the unit ball of \mathcal{B}_s is compactly embedded in \mathcal{B} . In our case, \mathcal{B} is (up to some details) the space of continuous functions on a compact set endowed with the uniform norm whereas \mathcal{B}_s is (up to an isomorphism) the space of Hölder continuous functions with a fixed Hölder exponent $\beta \in (0, 1)$. The stronger Hölder norm guarantees the compact embedding. While we keep the conditions (ii), (iii) above, we strengthen (i) to

- (i') $[\Phi_n - \Phi](\theta_*)$ satisfies a central limit theorem in \mathcal{B}_s

as the stronger topology will be used to help establish (8). We verify (i') by applying a general central limit theorem for Hölder spaces detailed in Appendix B. The key intermediate step towards (8) is that

- (iv') there are a compact $\mathcal{K} \subset \mathcal{B}$, random elements $U_n \in \mathcal{K}$, and random elements $V_n \in \mathcal{B}$ with $\|V_n\| \xrightarrow{\mathbb{P}} 0$, such that

$$\Delta_n := [\Phi_n - \Phi](\theta_n) - [\Phi_n - \Phi](\theta_*) = (U_n + V_n)\|\theta_n - \theta_*\|.$$

This is verified using the specifics of the problem at hand (cf. Lemma 5.4), without requiring smoothness. Next, we outline how these conditions lead to (8). Note that the differentiability condition (ii) implies

$$(\theta_* - \theta_n) = \mathbb{L}^{-1}(\Phi(\theta_*) - \Phi(\theta_n)) + R_n$$

with $\|R_n\| = o_{\mathbb{P}}(\|\theta_* - \theta_n\|)$. Using $\Phi(\theta_*) = 0 = \Phi_n(\theta_n)$, this can be written as

$$\theta_* - \theta_n = \mathbb{L}^{-1}(\Phi_n(\theta_n) - \Phi(\theta_n)) + R_n = \mathbb{L}^{-1}(\Delta_n + [\Phi_n - \Phi](\theta_*)) + R_n.$$

By (i'), the sequence $\sqrt{n}[\Phi_n - \Phi](\theta_*)$ is tight in \mathcal{B}_s . Using also (iv') and the continuity of \mathbb{L}^{-1} , we deduce a key decomposition (cf. Proposition 5.3)

$$(9) \quad \theta_* - \theta_n = \|\theta_n - \theta_*\| (U'_n + V'_n) + n^{-\frac{1}{2}} W_n$$

with random functions U'_n taking values in a compact $\mathcal{K}' \subset \mathcal{B}$, random functions $\|V'_n\| \xrightarrow{\mathbb{P}} 0$, and $(\mathbb{L}W_n)$ tight in \mathcal{B}_s . From here, we make (mild) use of the particular properties the functionals (7) at hand. Namely, ϕ is differentiable, hence the fundamental theorem of calculus allows us to write

$$\Delta_n = \int (\phi(\theta) - \phi(\theta_n)) d(\mathbb{P} - \mathbb{P}_n) = \int (\theta_* - \theta_n) \phi_n d(\mathbb{P} - \mathbb{P}_n)$$

for $\phi_n := \int_0^1 \phi'(\lambda\theta_* + (1-\lambda)\theta_n) d\lambda$. Inserting (9) yields

$$\Delta_n = \|\theta_* - \theta_n\| \int (U'_n + V'_n) \phi_n d(\mathbb{P} - \mathbb{P}_n) + n^{-\frac{1}{2}} \int W_n \phi_n d(\mathbb{P} - \mathbb{P}_n).$$

The stated properties of U'_n, V'_n enable us to verify that first term is $o_{\mathbb{P}}(\|\theta_n - \theta_*\|)$. Similarly, the tightness of $(\mathbb{L}W_n)$ in \mathcal{B}_s and the fact that $\|\cdot\|_s$ -bounded sets are relatively compact in \mathcal{B} allow us to verify that the second term is $o_{\mathbb{P}}(n^{-1/2})$. This completes the derivation of (8). The central limit theorem for the potentials then follows by standard arguments. The proofs of the central limit theorems for the optimal cost and the optimal couplings are based on the result for the potentials, (8) and the central limit theorem for two-sample U -statistics.

Finally, we comment on the uniqueness of the population potentials, shown in Theorem 3.2 when the cost c is continuous and one marginal has connected support. The usual technique to obtain uniqueness—familiar in unregularized optimal transport and adapted in [36] for the regularized problem with quadratic divergence—is to argue that the gradient of the potential equals the partial derivative of the cost c on the support of the optimal coupling. This argument is a nonstarter if the cost c is not differentiable. In the present work, uniqueness is instead deduced from the invertibility of \mathbb{L} . Clearly uniqueness of the population potentials is a precondition for any result on the convergence to them. On the other hand, we mention that the empirical potentials are in general not unique. They are unique in the special case of EOT and more generally when the optimal coupling has sufficiently large support, but not in general. The family of all potentials can be parametrized by the components (in the sense of connectedness of graphs) of the support, as shown in [36].

Organization. The remainder of this paper is organized as follows. Section 2 details the problem formulation and notation, and summarizes basic facts about the regularized optimal transport problem. Section 3 first states the main results, namely the uniqueness of the population potentials and the three central limit theorems, then continues with an overview of the proof strategy. The proofs start in Section 4 by analyzing the linearization of the first-order condition of the potentials, here the key result is the invertibility of the derivative \mathbb{L} (Proposition 4.4). The proofs of the central limit theorems are presented in Section 5. We first show in Section 5.1 that the empirical potentials are a consistent estimator. Section 5.2 contains the technical core of the proof, namely the aforementioned decomposition into three terms, each with a different compactness property. Section 5.3 concludes with the proofs of the central limit theorems. In Appendix A we provide a numerical analysis of a particular example where the population problem can be solved in closed form and hence the convergence rate can be accurately observed. Appendix B provides an abstract central limit theorem for the Banach space of Hölder functions that is used in the proof of our main results. Omitted proofs are collected in Appendix C.

2. Problem statement and preliminaries.

2.1. *Notation.* Let Ω be a compact subset of \mathbb{R}^d . For $f : \Omega \rightarrow \mathbb{R}$, we write $\|f\|_\infty = \sup_{x \in \Omega} |f(x)|$. The space of continuous functions $\Omega \rightarrow \mathbb{R}$ is denoted $\mathcal{C}(\Omega)$ and endowed with $\|\cdot\|_\infty$. Moreover, $\mathcal{C}^k(\Omega)$ denotes k -times continuously differentiable functions. For $\alpha \in (0, 1]$, the α -Hölder semi-norm of $f : \Omega \rightarrow \mathbb{R}$ is

$$[f]_\alpha = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|^\alpha}.$$

We set $\mathcal{C}^{0,\alpha}(\Omega) = \{f \in \mathcal{C}(\Omega) : [f]_\alpha < \infty\}$, a Banach space under the norm $\|f\|_{0,\alpha} := \|f\|_\infty + [f]_\alpha$. If \mathcal{B}_1 and \mathcal{B}_2 are Banach spaces, the space of bounded linear operators $F : \mathcal{B}_1 \rightarrow \mathcal{B}_2$ is denoted $\mathcal{L}(\mathcal{B}_1, \mathcal{B}_2)$ and endowed with the norm topology.

The open unit ball in \mathbb{R}^d is denoted by \mathbb{B} and its closure by $\overline{\mathbb{B}}$. Thus $x + r\mathbb{B}$ is the ball centered at x with radius r . The gradient of $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is denoted ∇f and the gradients of a function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ with respect to the first and second coordinates are denoted $\nabla_x c$ and $\nabla_y c$, respectively. The derivative of a univariate function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is denoted ψ' . Given functions $x \mapsto f(x)$ and $y \mapsto g(y)$, we denote by $f \oplus g$ the function $(x, y) \mapsto f(x) + g(y)$.

We fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ where all random variables are defined. For a measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and a random vector $X : \Omega \rightarrow \mathbb{R}^d$ with distribution P , the expectation of $f(X)$ is denoted $\mathbb{E}[f(X)] = \int f(x) dP(x) = \int f dP$. Almost-sure convergence is denoted by $\xrightarrow{\text{a.s.}}$, convergence in probability by $\xrightarrow{\mathbb{P}}$, and convergence in distribution (or weak convergence) by \xrightarrow{w} . The latter refers to convergence induced by continuous and bounded test functions. For scalar random variables X_n, Y_n , we write $X_n = \mathcal{O}_{\mathbb{P}}(Y_n)$ if X_n/Y_n is stochastically bounded and $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \xrightarrow{\mathbb{P}} 0$.

More terminology, related to probability in Banach spaces, can be found in Appendix B.

2.2. *Divergence.* Given measures μ, ν on the same space, the divergence $D_\varphi(\mu|\nu)$ is determined by the function $\varphi : [0, \infty) \rightarrow \mathbb{R}$ via

$$D_\varphi(\mu|\nu) = \int \varphi \left(\frac{d\mu}{d\nu} \right) d\mu,$$

with the convention that $D_\varphi(\mu|\nu) = \infty$ if $\mu \not\ll \nu$. The following assumption on φ is in force throughout the paper.

ASSUMPTION 2.1 (Divergence). The function $\varphi : [0, \infty) \rightarrow \mathbb{R}$ is strictly convex with $\varphi(1) = 0$, $\lim_{x \rightarrow \infty} \varphi(x)/x = +\infty$ and such that the conjugate

$$y \mapsto \psi(y) := \varphi^*(y) := \sup_{x \geq 0} \{xy - \varphi(x)\}$$

is in $\mathcal{C}^2(\mathbb{R})$. Moreover, there exists $C > 0$ such that $\psi'(x) \geq x$ for $x \geq C$, and there exist $t_0 > 0$ and $\delta > 0$ such that $\psi'(t_0) = 1$ and ψ is strictly convex on $[t_0 - \delta, \infty)$.

The detailed assumptions about the shape of ψ are useful to derive basic estimates for the potentials; cf. Proposition 2.3. While they are a bit clumsy, they are verified in all examples of our interest. The more restrictive assumption is that ψ is \mathcal{C}^2 .

EXAMPLE 2.2. For the Kullback–Leibler divergence of EOT, we take $\varphi(x) = x \log(x)$ which yields $\psi(y) = e^{y-1}$. Note that $\psi \in \mathcal{C}^\infty$ and $\psi'(y) > 0$ for all y , corresponding to the

fact that the optimal coupling always has full support. Next, consider the polynomial (Tsallis) divergence

$$\varphi(x) = \frac{x^\alpha - 1}{\alpha}$$

for $\alpha \in (1, \infty)$. Then $\psi'(y) = y_+^{\beta-1}$ where $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ and $y_+ = \max\{y, 0\}$. This truncation at zero corresponds to the fact that the optimal coupling does not have full support in general. Note that ψ is strictly convex on \mathbb{R}_+ and in \mathcal{C}^1 . Moreover, ψ is \mathcal{C}^2 for $\beta > 2$; i.e., for any $\alpha \in (1, 2)$.

We note that the quadratic case $\alpha = 2$ is not covered. For this boundary case, we expect to provide similar results as in the present paper in separate work, assuming that c is the quadratic cost and the marginals have additional regularity (see also Appendix A for numerical hints). That setting implies additional convexity properties for the potentials. In the present work, we focus on giving a general result for a broad class of divergences, costs and marginals, thus showing that the conclusions do not hinge on a particular algebraic structure.

2.3. Regularized optimal transport. For brevity, we treat the regularized optimal transport problem for regularization parameter $\varepsilon = 1$. The general case is recovered by a simple scaling argument detailed in Remark 2.5 below. Thus the primal problem is

$$(10) \quad \text{ROT}(P, Q) := \inf_{\pi \in \Pi(P, Q)} \int cd\pi + \int \varphi \left(\frac{d\pi}{d(P \otimes Q)} \right) d(P \otimes Q).$$

The associated dual problem is

$$(11) \quad \text{DUAL}(P, Q) := \sup_{(f, g) \in L^\infty(P) \times L^\infty(Q)} \int (f \oplus g - \psi(f \oplus g - c)) d(P \otimes Q)$$

and its first-order condition for optimality is

$$(12) \quad \begin{cases} \int \psi'(f_*(\cdot) + g_*(y) - c(\cdot, y)) dQ(y) = 1 & P\text{-a.s.}, \\ \int \psi'(f_*(x) + g_*(\cdot) - c(x, \cdot)) dP(x) = 1 & Q\text{-a.s.} \end{cases}$$

The following proposition summarizes some basic facts to be used throughout the paper. While the proofs are essentially known and largely contained in [1], we give a self-contained and simpler proof in Appendix C.1, for the convenience of the reader and also to rectify minor inaccuracies in [1].

PROPOSITION 2.3. *Let P, Q be probability measures on \mathbb{R}^d with compact supports Ω, Ω' . Moreover, let $c \in \mathcal{C}(\Omega \times \Omega')$.*

- (i) *The strong duality $\text{ROT}(P, Q) = \text{DUAL}(P, Q)$ holds.*
- (ii) *The primal problem (10) admits a unique optimizer $\pi \in \Pi(P, Q)$.*
- (iii) *The dual problem (11) admits a (non-unique) optimizer $(f_*, g_*) \in L^\infty(P) \times L^\infty(Q)$.*
- (iv) *A pair $(f_*, g_*) \in L^\infty(P) \times L^\infty(Q)$ is an optimizer of the dual problem (11) if and only if it satisfies the first-order condition (12).*
- (v) *Any dual optimizer (f_*, g_*) yields a primal optimizer via $d\pi := \psi'(f_* \oplus g_* - c)d(P \otimes Q)$.*
- (vi) *Let $(f_*, g_*) \in L^\infty(P) \times L^\infty(Q)$ satisfy (12). Then one can choose a version of (f_*, g_*) satisfying (12) everywhere; that is,*

$$(13) \quad \begin{cases} \int \psi'(f_*(x) + g_*(y) - c(x, y)) dQ(y) = 1 & \text{for all } x \in \Omega, \\ \int \psi'(f_*(x) + g_*(y) - c(x, y)) dP(x) = 1 & \text{for all } y \in \Omega'. \end{cases}$$

For such a version (f_, g_*) , the uniform bound $\|f_* \oplus g_*\|_\infty \leq 5\|c\|_\infty + (\psi')^{-1}(1) < \infty$ holds. Moreover, if ρ is a modulus of uniform continuity for c on $\Omega \times \Omega'$, then f_* and g_* are also uniformly continuous with modulus ρ .*

(vii) Let (f_*, g_*) solve (13) and abbreviate $\xi_*(x, y) := f_*(x) + g_*(y) - c(x, y)$. Then there exists $\delta > 0$ such that

$$\int \psi''(\xi_*(\cdot, y)) dQ(y) \geq \delta \quad \text{on } \Omega, \quad \int \psi''(\xi_*(x, \cdot)) dP(x) \geq \delta \quad \text{on } \Omega'.$$

Functions (f_*, g_*) that are optimal for the dual problem (11), or equivalently solve the first-order condition (12), will be called *potentials* of the regularized optimal transport problem.

REMARK 2.4 (Limited Smoothness). Proposition 2.3 (vi) may suggest that the potentials f_* and g_* are “as smooth as the cost c .” While this holds for the KL divergence of EOT, ψ is in general only C^2 for the polynomial divergences included in our setting, and this implies that f_* and g_* are only C^1 in general. Indeed, the formula

$$\nabla f_*(\cdot) = \frac{\int \psi''(f_*(\cdot) + g_*(y) - c(\cdot, y)) \nabla_x c(\cdot, y) dQ(y)}{\int \psi''(f_*(\cdot) + g_*(y) - c(\cdot, y)) dQ(y)}$$

highlights how the smoothness of both ψ and c affects the regularity of f_* . (See also [36] for a concrete counterexample.) As mentioned in the introduction, the limited smoothness of the empirical potentials is a major obstacle for deriving our main results.

REMARK 2.5 (Regularization parameter). The regularized optimal transport problem is often considered with a regularization parameter $\varepsilon > 0$, which was taken to be $\varepsilon = 1$ above. In general,

$$\text{ROT}_\varepsilon(P, Q) := \inf_{\pi \in \Pi(P, Q)} \int cd\pi + \varepsilon \int \varphi \left(\frac{d\pi}{d(P \otimes Q)} \right) d(P \otimes Q),$$

the associated dual problem is

$$(14) \quad \text{DUAL}_\varepsilon(P, Q) := \sup_{(f, g) \in L^\infty(P) \times L^\infty(Q)} \int \left\{ f \oplus g - \varepsilon \cdot \psi \left(\frac{f \oplus g(y) - c}{\varepsilon} \right) \right\} d(P \otimes Q)$$

and the first-order condition for optimality reads

$$\begin{cases} \int \psi' \left(\frac{f(\cdot) + g(y) - c(\cdot, y)}{\varepsilon} \right) dQ(y) = 1, \\ \int \psi' \left(\frac{f(x) + g(\cdot) - c(x, \cdot)}{\varepsilon} \right) dP(x) = 1. \end{cases}$$

The results for the general problem can be deduced from the special case $\varepsilon = 1$ by a simple scaling. Namely, defining $\bar{c} = c/\varepsilon$, we have

$$\text{ROT}_\varepsilon(P, Q) = \varepsilon \text{ROT}(P, Q, \bar{c})$$

where $\text{ROT}(P, Q, \bar{c})$ is the problem with $\varepsilon = 1$ and cost \bar{c} . Moreover, the optimal couplings of these two problems coincide. Similarly for the dual: $(f_\varepsilon, g_\varepsilon)$ is optimal for (14) if and only if $(\varepsilon f_\varepsilon, \varepsilon g_\varepsilon)$ is optimal for the problem $\text{DUAL}(P, Q, \bar{c})$ with $\varepsilon = 1$ and cost \bar{c} . For convenience, we detail in Remark 3.7 below how the main results translate to a general parameter $\varepsilon > 0$.

3. Main results. Our main results are central limit theorems for the potentials, the optimal costs, and the optimal couplings. The following condition on the population marginals P and Q is imposed throughout this section.

ASSUMPTION 3.1 (Marginals). The probability measures P and Q on \mathbb{R}^d have compact supports Ω and Ω' , respectively, and Ω is connected.

A central limit theorem for the potentials can only hold if the limiting (population) potentials (f_*, g_*) are unique in a suitable sense. Indeed we obtain a uniqueness result under quite general conditions as by-product of our approach to the central limit theorem. Note that potentials always admit a degree of freedom: if (f_*, g_*) are potentials, then so are $(f_* + a, g_* - a)$ for any $a \in \mathbb{R}$. Additionally, there is some freedom in changing potentials on nullsets, which can be mended by using continuous versions as in Proposition 2.3. To eliminate those degrees of freedom, we define the Banach space

$$\mathcal{C}_\oplus = (\mathcal{C}(\Omega) \times \mathcal{C}(\Omega')) / \sim_\oplus$$

where $(f, g) \sim_\oplus (u, v)$ if $f \oplus g = u \oplus v$, endowed with the quotient norm

$$\|(f, g)\|_\oplus = \inf_{a \in \mathbb{R}} \{\|f - a\|_\infty + \|g + a\|_\infty\}.$$

The equivalence class of (f, g) will be denoted by $[(f, g)]_\oplus$ when we want to emphasize the identification; however, we often just write (f, g) for the sake of brevity. We can now state the uniqueness of the population potentials. We emphasize that this result holds for costs c that are merely continuous, in contrast to related results in the literature that rely on differentiating c (cf. Section 1.2).

THEOREM 3.2. *Let $c \in \mathcal{C}(\Omega \times \Omega')$ and let (P, Q) satisfy Assumption 3.1. Then the associated potentials (f_*, g_*) are unique in \mathcal{C}_\oplus . That is, there exist $(f_*, g_*) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ solving (13), and any pair solving (13) is of the form $(f_* + a, g_* - a)$ for some $a \in \mathbb{R}$.*

Next, we present the central limit theorems. Given (population) marginals P, Q satisfying Assumption 3.1, we consider the two-sample case. Let X_1, \dots, X_n be i.i.d. samples from P and let Y_1, \dots, Y_n be i.i.d. samples from Q ; more precisely, the two sequences are independent and defined on the common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The random samples give rise to the empirical measures $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $Q_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$, which in turn lead to empirical potentials (f_n, g_n) by Proposition 2.3. More precisely, Proposition 2.3 yields (f_n, g_n) satisfying (12) on the supports $\text{supp}(P_n)$ and $\text{supp}(Q_n)$, respectively, but we can extend (f_n, g_n) to continuous functions on $\Omega \times \Omega'$. While (f_n, g_n) are not unique, we choose and fix them. Note that all these objects are random; that is, depend on the parameter $\omega \in \Omega$ determining the realization of the random samples. We choose (f_n, g_n) such that the dependence on ω is measurable. To simplify the notation we also introduce

$$(15) \quad \xi_*(x, y) := f_*(x) + g_*(y) - c(x, y) \quad \text{and} \quad \xi_n(x, y) := f_n(x) + g_n(y) - c(x, y).$$

ASSUMPTION 3.3 (Cost). The transport cost c is in $\mathcal{C}^1(\Omega \times \Omega')$.

This assumption is in force for the three central limit theorems below, in addition to Assumption 2.1 on the divergence and Assumption 3.1 on the population marginals (P, Q) . We can now state our main result.

THEOREM 3.4 (CLT for potentials). *Let $\mathbf{G}_P = (\mathbf{G}_P(y))_{y \in \Omega'}$ be a centered Gaussian process in $\mathcal{C}(\Omega')$ with covariance function*

$$\begin{aligned} & \mathbb{E}[\mathbf{G}_P(y)\mathbf{G}_P(y')] \\ &= \mathbb{E} \left\{ \left(\psi'(\xi_*(X, y)) - \mathbb{E}[\psi'(\xi_*(X, y))] \right) \left(\psi'(\xi_*(X, y')) - \mathbb{E}[\psi'(\xi_*(X, y'))] \right) \right\} \end{aligned}$$

for $(y, y') \in \Omega'$, where $X \sim P$. Let $\mathbf{G}_Q = (\mathbf{G}_Q(x))_{x \in \Omega}$ be a centered Gaussian process in $\mathcal{C}(\Omega)$ with analogous covariance, independent of \mathbf{G}_P . Then the weak limit

$$\sqrt{n} \begin{pmatrix} f_n - f_* \\ g_n - g_* \end{pmatrix} \xrightarrow{w} -\mathbb{L}^{-1} \left[\begin{pmatrix} \mathbf{G}_Q \\ \frac{\int \psi''(\xi_*(\cdot, x)) dQ(y)}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \end{pmatrix} \right]_{\oplus}$$

holds in \mathcal{C}_{\oplus} , where $\mathbb{L}^{-1} \in \mathcal{L}(\mathcal{C}_{\oplus}, \mathcal{C}_{\oplus})$ is the inverse of the (bounded, bijective) linear operator

$$\mathbb{L} : \mathcal{C}_{\oplus} \rightarrow \mathcal{C}_{\oplus}, \quad (f, g) \mapsto \begin{pmatrix} f(\cdot) + \frac{\int \psi''(\xi_*(\cdot, y)) g(y) dQ(y)}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \\ g(\cdot) + \frac{\int \psi''(\xi_*(x, \cdot)) f(x) dP(x)}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \end{pmatrix}.$$

THEOREM 3.5 (CLT for costs). *We have*

$$\sqrt{n}(\text{ROT}(P_n, Q_n) - \text{ROT}(P, Q)) \xrightarrow{w} N(0, \sigma^2)$$

where the variance σ^2 is that of the random variable

$$(16) \quad f_*(X) + g_*(Y) - \left(\int \psi(\xi_*(x, Y)) dP(x) + \int \psi(\xi_*(X, y)) dQ(y) \right)$$

for $(X, Y) \sim P \otimes Q$.

Let $\pi \in \Pi(P, Q)$ denote the optimal coupling for the population marginals and let $\pi_n \in \Pi(P_n, Q_n)$ denote the empirical counterpart. To state the central limit theorem for $\pi_n \rightarrow \pi$, recall that $[(f, g)]_{\oplus}$ denotes the equivalence class of $(f, g) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ in \mathcal{C}_{\oplus} . To streamline the notation, we also define the operator

$$[(f, g)]_{\oplus} \mapsto \oplus(f, g) := f \oplus g \in \mathcal{C}(\Omega \times \Omega').$$

THEOREM 3.6 (CLT for couplings). *For any bounded and measurable $\eta : \Omega \times \Omega' \rightarrow \mathbb{R}$,*

$$\sqrt{n} \left(\int \eta d(\pi_n - \pi) \right) \xrightarrow{w} N(0, \sigma^2(\eta))$$

with $\sigma^2(\eta) = \text{Var}(V_X + V_Y)$ where, for $Z \in \{X, Y\}$,

$$V_Z := \mathbb{E} \left[\int U(x, y, X, Y) \bar{\eta}(x, y) \psi'(\xi_*(x, y)) d(P \otimes Q)(x, y) - \bar{\eta}(X, Y) \psi'(\xi_*(X, Y)) \middle| Z \right]$$

for $(X, Y) \sim P \otimes Q$ and $\bar{\eta} := \eta - \int \eta \psi''(\xi_*) d(P \otimes Q)$ and

$$U(\cdot, \cdot, X, Y) := \oplus \left(\mathbb{L}^{-1} \left[\begin{pmatrix} \frac{\psi'(\xi_*(\cdot, Y))}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \\ \frac{\psi'(\xi_*(X, \cdot))}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \end{pmatrix} \right]_{\oplus} \right).$$

REMARK 3.7 (Regularization parameter). Let us detail how the main results, stated above for $\varepsilon = 1$, read in the case of a general regularization parameter $\varepsilon > 0$. The following can be deduced by the scaling argument in Remark 2.5. First, (f_*, g_*) are replaced by the rescaled potentials $(f_\varepsilon, g_\varepsilon)$ as defined in Remark 2.5, and ξ_* is replaced by

$$\xi_\varepsilon(x, y) = \frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x, y)}{\varepsilon}.$$

Apart from those changes, the formula for \mathbb{L} remains unchanged. In the statement of Theorem 3.4, we further add a multiplicative factor ε in front of \mathbb{L}^{-1} . While in Theorem 3.5, we replace ROT by ROT_ε on the left-hand side and add a multiplicative factor ε in front of the bracket in (16). Finally, the statement of Theorem 3.6 requires not further changes, with the understanding that π, π_n are now the optimal couplings for ROT_ε .

3.1. *Proof strategy.* Our arguments center on the potentials (f_*, g_*) and their empirical counterparts (f_n, g_n) . The first-order condition of optimality (13) can be written in operator form as $\Gamma(f_*, g_*) = (1, 1)$ and $\Gamma_n(f_n, g_n) = (1, 1)$. In fact, for reasons discussed later, we work with closely related operators $\tilde{\Gamma}$ and $\tilde{\Gamma}_n$. Linearizing the first-order condition yields an identity of the form

$$(17) \quad \tilde{\Gamma}(f_n, g_n) - \tilde{\Gamma}_n(f_n, g_n) = \mathbb{L}(f_n - f_*, g_n - g_*) + (\dots)$$

where \mathbb{L} is a linear operator and (\dots) are terms that do not cause difficulties.

The first step is to show the invertibility of \mathbb{L} (Proposition 4.4). To this end, we show that $\mathbb{L} = \text{Id} + \mathbb{A}$ where \mathbb{A} is a compact operator. Thus by Fredholm's alternative, \mathbb{L} is invertible if (and only if) -1 is not an eigenvalue of \mathbb{A} . To prove the latter, suppose that (f, g) is an eigenvector corresponding to the eigenvalue -1 of \mathbb{A} . We first show that

$$(18) \quad \|(f, g)\|_{\oplus} = \|\mathbb{A}(f, g)\|_{\oplus},$$

which implies that f is constant on the set $\cup_{y \in S_Q(x)} S_P(y)$ for all $x \in \text{argmax} |f|$, where $S_Q(x)$ and $S_P(y)$ are the sections of the set $S = \{(x, y) \in \Omega \times \Omega' : \psi''(\xi_*(x, y)) > 0\}$. In the case of EOT, S is the full space and one immediately concludes that f is constant, completing the proof. In the general case, S can be smaller (even sparse, see Remark 4.3) and then $\cup_{y \in S_Q(x)} S_P(y)$ does not cover the whole space Ω in general. Instead, we prove that (18) implies that f is constant in a neighborhood of fixed radius α around the set $\text{argmax} |f|$. We then iterate this reasoning to show that f is constant on the whole space Ω .

Once the invertibility of \mathbb{L} is shown, we focus on the central limit theorem. A high-level perspective on the proof was already given in Section 1.2 based on Z-estimation. Here, we give a more pedestrian sketch. The basic idea is that if we had a central limit theorem for the left hand side of (17), then we could apply \mathbb{L}^{-1} on both sides and deduce the desired result for $(f_n - f_*, g_n - g_*)$. To follow this strategy, one would like to replace the left hand side $\tilde{\Gamma}(f_n, g_n) - \tilde{\Gamma}_n(f_n, g_n)$ by the more tractable expression $\tilde{\Gamma}(f_*, g_*) - \tilde{\Gamma}_n(f_*, g_*)$. Indeed, general arguments show that the latter expression satisfies a central limit theorem (Lemma 5.2). Let us proceed with that replacement, by defining the difference

$$\Delta_n := [\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) - [\tilde{\Gamma} - \tilde{\Gamma}_n](f_n, g_n)$$

and hence

$$\tilde{\Gamma}(f_*, g_*) - \tilde{\Gamma}_n(f_*, g_*) = \mathbb{L}(f_n - f_*, g_n - g_*) + \Delta_n + (\dots).$$

At this point we would like an estimate along the lines of $\|\Delta_n\|_{\oplus} = o(\|(f_* - f_n, g_* - g_n)\|_{\oplus})$. A first attempt may be to bound $\|\Delta_n\|_{\oplus}$ in terms of

$$(19) \quad \|(f_* - f_n, g_* - g_n)\|_{\oplus} \sup_{\|(f, g)\|_{\oplus} \leq 1} \|[\tilde{\Gamma}_n - \tilde{\Gamma}](f, g)\|_{\infty}.$$

However, this expression does not behave like $o(\|(f_* - f_n, g_* - g_n)\|_{\oplus})$. Indeed, the empirical measures would have to converge in total variation for the supremum over continuous functions to be of order $o(1)$. Loosely speaking, this failure arises because the unit ball of \mathcal{C}_{\oplus} is not compact. We resolve this with a finer analysis decomposing Δ_n into two sequences, one that is valued in a compact subset of \mathcal{C}^0 and one whose uniform norm converges to zero almost surely (Lemma 5.4).

This leads to a key decomposition of $(f_* - f_n, g_* - g_n)$ into three sequences (Proposition 5.3): two as just described, and a third arising from the aforementioned central limit theorem in Lemma 5.2. A crucial detail is that we establish this central limit theorem not only in \mathcal{C}_{\oplus} but for a stronger norm whose unit ball is compactly embedded in the uniform topology

and hence satisfies the universal Glivenko–Cantelli property, avoiding the roadblock (19). On the strength of the decomposition, we derive in Corollary 5.7 that

$$\|\Delta_n\|_{\oplus} = o_{\mathbb{P}} \left(\|(f_* - f_n, g_* - g_n)\|_{\oplus} + n^{-\frac{1}{2}} \right).$$

Using this, the three-sequence decomposition, and the central limit theorem for two-sample U -statistics, Section 5.3 completes the proofs of our main results.

4. Linearization of the Optimality Condition. Throughout this section, the transport cost $c \in \mathcal{C}(\Omega \times \Omega')$ is fixed and (P, Q) are fixed marginals satisfying Assumption 3.1. (Differentiability of c is not assumed.) Note that Proposition 2.3 applies not only to those population marginals (P, Q) but also to the empirical measures (P_n, Q_n) of their i.i.d. samples. We choose continuous versions of the potentials as detailed in Proposition 2.3. While at this point we do not yet know the uniqueness of the population potentials, we may choose and fix one such pair (f_*, g_*) . By (13), the potentials $(f_*, g_*) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ are a solution of the equation $\Gamma(f_*, g_*) = (1, 1)$, where

$$\begin{aligned} \Gamma : \mathcal{C}_{\oplus} &\rightarrow \mathcal{C}(\Omega) \times \mathcal{C}(\Omega'), \\ (f, g) &\mapsto \begin{pmatrix} \Gamma^{(1)}(f, g) \\ \Gamma^{(2)}(f, g) \end{pmatrix} = \begin{pmatrix} \int \psi'(f(\cdot) + g(y) - c(\cdot, y)) dQ(y) \\ \int \psi'(f(x) + g(\cdot) - c(x, \cdot)) dP(x) \end{pmatrix}. \end{aligned}$$

Similarly, the empirical potentials (f_n, g_n) are a solution of $\Gamma_n(f_n, g_n) = (1, 1)$, where

$$\begin{aligned} \Gamma_n : \mathcal{C}_{\oplus} &\rightarrow \mathcal{C}(\Omega) \times \mathcal{C}(\Omega'), \\ (f, g) &\mapsto \begin{pmatrix} \Gamma_n^{(1)}(f, g) \\ \Gamma_n^{(2)}(f, g) \end{pmatrix} = \begin{pmatrix} \int \psi'(f(\cdot) + g(y) - c(\cdot, y)) dQ_n(y) \\ \int \psi'(f(x) + g(\cdot) - c(x, \cdot)) dP_n(x) \end{pmatrix}. \end{aligned}$$

As mentioned in the preceding section, we choose and fix empirical potentials (f_n, g_n) that are continuous on $\Omega \times \Omega'$ and depend measurably on $\omega \in \Omega$.

Recall the shorthand $\xi_*(x, y) = f_*(x) + g_*(y) - c(x, y)$ from (15). The following lemma details the first-order development of the operator

$$\tilde{\Gamma} : \mathcal{C}_{\oplus} \rightarrow \mathcal{C}(\Omega) \times \mathcal{C}(\Omega'), \quad (f, g) \mapsto \tilde{\Gamma}(f, g) := \begin{pmatrix} \frac{\Gamma^{(1)}}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \\ \frac{\Gamma^{(2)}}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \end{pmatrix}.$$

Here the denominators are bounded away from zero by Proposition 2.3. We introduce those denominators so that the derivative \mathbb{L} (see (20) below) becomes an operator $\mathcal{C}_{\oplus} \rightarrow \mathcal{C}_{\oplus}$ that is a perturbation of the identity. (A priori, these operators could depend on the chosen and fixed potentials (f_*, g_*) , which causes no harm. Once the proof of uniqueness is complete, it will be clear that there was in fact no ambiguity.)

LEMMA 4.1. *The operator $\tilde{\Gamma} : \mathcal{C}_{\oplus} \rightarrow \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ is continuously Fréchet differentiable; we denote its derivative at the point $(f, g) \in \mathcal{C}_{\oplus}$ by $\mathbb{L}_{(f, g)}$. That is, we have*

$$\lim_{\|(u, v)\|_{\oplus} \rightarrow 0} \frac{\|\tilde{\Gamma}(f + u, g + v) - \tilde{\Gamma}(f, g) - \mathbb{L}_{(f, g)}(u, v)\|_{\mathcal{C}(\Omega) \times \mathcal{C}(\Omega')}}{\|(u, v)\|_{\oplus}} = 0$$

and the function $\mathcal{C}_{\oplus} \ni (f, g) \mapsto \mathbb{L}_{(f, g)} \in \mathcal{L}(\mathcal{C}_{\oplus}, \mathcal{C}(\Omega) \times \mathcal{C}(\Omega'))$ is continuous. The derivative at (f_*, g_*) is given by

$$\mathbb{L}_{(f_*, g_*)} : \mathcal{C}_{\oplus} \rightarrow \mathcal{C}(\Omega) \times \mathcal{C}(\Omega'), \quad (u, v) \mapsto \begin{pmatrix} u(\cdot) + \frac{\int \psi''(f_*(\cdot) + g_*(y) - c(\cdot, y)) v(y) dQ(y)}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \\ v(\cdot) + \frac{\int \psi''((f_*(x) + g_*(\cdot) - c(x, \cdot)) u(x) dP(x))}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \end{pmatrix}.$$

PROOF. Using that ψ' is \mathcal{C}^1 , the proof is a direct calculation. \square

Composing with the quotient map, we define the operator $\mathbb{L} \in \mathcal{L}(\mathcal{C}_\oplus, \mathcal{C}_\oplus)$ by

$$(20) \quad \mathbb{L} : (u, v) \mapsto [\mathbb{L}_{(f_*, g_*)}(u, v)]_\oplus.$$

The main goal of this section is to show that \mathbb{L} is invertible. We first state an auxiliary result on the sections of the set $S := \{(x, y) \in \Omega \times \Omega' : \psi''(\xi_*(x, y)) > 0\}$ whose interpretation is discussed in Remark 4.3 below. The proof is reported in Appendix C.2.

LEMMA 4.2. *Consider the sections of $S = \{(x, y) \in \Omega \times \Omega' : \psi''(\xi_*(x, y)) > 0\}$,*

$$S_Q(x) := \{y \in \Omega' : \psi''(\xi_*(x, y)) > 0\}, \quad S_P(y) := \{x \in \Omega : \psi''(\xi_*(x, y)) > 0\}.$$

There exists $\alpha > 0$ such that

$$(x + \alpha \overline{\mathbb{B}}) \cap \Omega \subset \cup_{y \in S_Q(x)} S_P(y) \quad \text{for all } x \in \Omega.$$

REMARK 4.3. The set $S = \{(x, y) \in \Omega \times \Omega' : \psi''(\xi_*(x, y)) > 0\}$ is closely related to the support of the optimal coupling π . Indeed, $\psi' \circ \xi_*$ is the density of π by Proposition 2.3. For the KL and polynomial divergences detailed in Example 2.2, $\{t : \psi'(t) > 0\} = \{t : \psi''(t) > 0\}$, hence S is exactly the set where the density is positive, and its closure is the support of π . For the KL divergence of EOT, $\psi'' > 0$ on all of \mathbb{R} and thus $S = \Omega \times \Omega'$, so that Lemma 4.2 is trivial. For the polynomial divergences, however, it is known that the support is often sparse (see Section 1.1), and then Lemma 4.2 is relevant.

The formula in Lemma 4.1 suggests to see $\mathbb{L} = \text{Id} + \mathbb{A}$ as a perturbation of the identity $\text{Id} \in \mathcal{L}(\mathcal{C}_\oplus, \mathcal{C}_\oplus)$. To prove the invertibility of \mathbb{L} , we will show that the operator

$$(21) \quad \mathbb{A} = (\mathbb{A}_1, \mathbb{A}_2) : \mathcal{C}_\oplus \rightarrow \mathcal{C}_\oplus, \quad (f, g) \mapsto \left(\begin{array}{c} \frac{\int \psi''(\xi_*(\cdot, y))g(y)dQ(y)}{\int \psi''(\xi_*(\cdot, y))dQ(y)} \\ \frac{\int \psi''(\xi_*(x, \cdot))f(x)dP(x)}{\int \psi''(\xi_*(x, \cdot))dP(x)} \end{array} \right)$$

is compact and -1 is not an eigenvalue. Then, we conclude by the Fredholm alternative. While similar results are known for EOT [3, 23], we require a substantially different proof because our optimal coupling need not have full support.

PROPOSITION 4.4. *The operator $\mathbb{L} \in \mathcal{L}(\mathcal{C}_\oplus, \mathcal{C}_\oplus)$ admits an inverse $\mathbb{L}^{-1} \in \mathcal{L}(\mathcal{C}_\oplus, \mathcal{C}_\oplus)$.*

PROOF. We first show that the operator \mathbb{A} of (21) is compact. We need to show that if $\{(u_n, v_n)\}_{n \in \mathbb{N}}$ is bounded in \mathcal{C}_\oplus then $\{(\mathbb{A}_1 u_n, \mathbb{A}_2 v_n)\}_{n \in \mathbb{N}}$ is relatively compact in \mathcal{C}_\oplus . Since $\{(u_n, v_n)\}$ is bounded in \mathcal{C}_\oplus , there exists a sequence $\{a_n\}$ of real numbers such that $\|u_n - a_n\|_\infty$ and $\|v_n + a_n\|_\infty$ are bounded. A sufficient condition for the relative compactness of $\{(\mathbb{A}_1 v_n, \mathbb{A}_2 u_n)\}$ in \mathcal{C}_\oplus is that $\{\mathbb{A}_1(v_n + a_n)\}$ and $\{\mathbb{A}_2(u_n - a_n)\}$ are relatively compact in $\mathcal{C}(\Omega)$ and $\mathcal{C}(\Omega')$, respectively. We only prove that $\{\mathbb{A}_2(u_n - a_n)\}$ is relatively compact, the former is analogous. By the Arzelà–Ascoli theorem, it suffices to show equicontinuity. As the function $(\int \psi''(\xi_*(x, \cdot))dP(x))^{-1}$ is continuous by Proposition 2.3, and since the product of an equicontinuous sequence with a continuous function (on the compact space Ω') remains equicontinuous, it moreover suffices to show that

$$(22) \quad \int \psi''(\xi_*(x, \cdot))(u_n(x) - a_n)dP(x), \quad n \in \mathbb{N}$$

is equicontinuous. Indeed, let ρ be a modulus of continuity for $\int \psi''(\xi_*(x, \cdot))dP(x)$ and $C = \sup_n \|u_n - a_n\|_\infty$, then $C\rho$ is a modulus of continuity for the functions in (22). This completes the proof that \mathbb{A} is compact.

Next, we show that -1 is not an eigenvalue of \mathbb{A} . Suppose that

$$(23) \quad f + a = -\mathbb{A}_1(g) \quad \text{and} \quad g - a = -\mathbb{A}_2(f)$$

for some $(f, g) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ and $a \in \mathbb{R}$; we need to prove that $[(f, g)]_{\oplus} = [(0, 0)]_{\oplus}$. As a first step, we show that $a = 0$ must hold in (23). Using (23) together with the definition of \mathbb{A}_2 , we have

$$(24) \quad \int \psi''(\xi_*(x, y))(f(x) + g(y))dQ(y) = -a \int \psi''(\xi_*(x, y))dQ(y), \quad \text{for all } x \in \Omega$$

and similarly by the definition of \mathbb{A}_1 ,

$$(25) \quad \int \psi''(\xi_*(x, y))(f(x) + g(y))dP(x) = a \int \psi''(\xi_*(x, y))dP(x), \quad \text{for all } y \in \Omega'.$$

Integrating in (24) and (25) w.r.t. P and Q , respectively, we get

$$a \int \psi''(\xi_*(x, y))d(P \otimes Q)(y) = 0,$$

which implies $a = 0$. Hence, (23) simplifies to $f = -\mathbb{A}_1(g)$ and $g = -\mathbb{A}_2(f)$, which we can concatenate to deduce

$$(26) \quad f = \mathbb{A}_1\mathbb{A}_2(f).$$

Next, we claim that for any $x \in \operatorname{argmax} |f|$, f is constant on the set $\cup_{y \in S_Q(x)} S_P(y) \subset \Omega$. Indeed, taking norms on both sides of (26) yields

$$\|f\|_{\infty} = \|\mathbb{A}_1\mathbb{A}_2(f)\|_{\infty}.$$

It then suffices to show that, for any $h \in \mathcal{C}(\Omega)$,

$$\|\mathbb{A}_1\mathbb{A}_2(h)\|_{\infty} < \|h\|_{\infty}$$

unless h is constant on the set $\cup_{y \in S_Q(x)} S_P(y)$ for all $x \in \operatorname{argmax} |h|$. To see the latter, fix $x_* \in \operatorname{argmax} |h|$ and define the probability measure

$$A \mapsto \mu_{x_*}(A) = \frac{\int \psi''(\xi_*(x_*, y)) \frac{\int_A \psi''(\xi_*(x, y))dP(x)}{\int \psi''(\xi_*(x, y))dP(x)} dQ(y)}{\int \psi''(\xi_*(x_*, y))dQ(y)}$$

on Ω . Then $\int h(x)d\mu_{x_*}(x) = \mathbb{A}_1\mathbb{A}_2(h)(x_*)$ for any $h \in \mathcal{C}(\Omega)$. In particular, using also (26),

$$\|f\|_{\infty} = |f(x_*)| = |\mathbb{A}_1\mathbb{A}_2(f)(x_*)| = \left| \int f(x)d\mu_{x_*}(x) \right|.$$

As f is continuous, this implies that $f(x) = \|f\|_{\infty}$ for all $x \in \operatorname{supp}(\mu_{x_*})$, the support of μ_{x_*} . Observing that $\operatorname{supp}(\mu_{x_*})$ contains the set $\cup_{y \in S_Q(x_*)} S_P(y)$ completes the proof of the claim that f is constant on the set $\cup_{y \in S_Q(x)} S_P(y)$ for all $x \in \operatorname{argmax} |f|$.

Next, we show that $f = \|f\|_{\infty}$ on Ω . By Lemma 4.2 there is $\alpha > 0$ such that

$$(x + \alpha\overline{\mathbb{B}}) \cap \Omega \subset \cup_{y \in S_Q(x)} S_P(y) \quad \text{for all } x \in \Omega.$$

Fix $x_* \in \operatorname{argmax} |f|$. The above then implies $f = \|f\|_{\infty}$ on $(x_* + \alpha\overline{\mathbb{B}}) \cap \Omega$. We iterate this argument to prove that $f = \|f\|_{\infty}$ on Ω . Indeed, let $\bar{x} \in \Omega$ be arbitrary; we show that $f(\bar{x}) = \|f\|_{\infty}$. As Ω is compact and connected, there exist $N \in \mathbb{N}$ and $x_1, x_2, \dots, x_N \in \Omega$ with $x_1 = x_*$, $x_N = \bar{x}$ and $|x_i - x_{i+1}| \leq \alpha/2$ (e.g., such x_i arise from an open cover of Ω by balls of radius $\alpha/2$). We know $f(x_1) = \|f\|_{\infty}$ and the above argument shows that if $f(x_i) = \|f\|_{\infty}$ then also $f = \|f\|_{\infty}$ on $(x_i + \alpha\overline{\mathbb{B}}) \cap \Omega$, which in particular yields $f(x_{i+1}) = \|f\|_{\infty}$.

Inductively, we obtain $f(\bar{x}) = f(x_N) = f(x_1) = \|f\|_\infty$. This completes the proof that $f = \|f\|_\infty$ on Ω .

Using the fact that $g = -\mathbb{A}_2(f)$, we derive that $\|g\|_\infty = -\|f\|_\infty$, implying that $(f, g) \sim_\oplus (0, 0)$. In summary, we have shown that -1 is not an eigenvalue of \mathbb{A} , and hence that the linear operator $\mathbb{L} = \text{Id} + \mathbb{A}$ is injective. By Fredholm's alternative, it follows that $\mathbb{L} : \mathcal{C}_\oplus \rightarrow \mathcal{C}_\oplus$ is also surjective. As \mathcal{C}_\oplus is a Banach space, the inverse is necessarily continuous, completing the proof. \square

As a by-product of the invertibility of \mathbb{L} , we obtain the uniqueness of the potentials.

PROOF OF THEOREM 3.2. Let (f_*, g_*) be as above and let $(\tilde{f}_*, \tilde{g}_*)$ be another solution of the first-order condition (13). We first note that any convex combination $(\lambda\tilde{f}_* + (1-\lambda)f_*, \lambda\tilde{g}_* + (1-\lambda)g_*)$, for $\lambda \in (0, 1)$, is again a solution of (13). Indeed, by Proposition 2.3, the set of solutions of (13) is the set of (continuous) optimizers of the dual problem (11), and the latter set is convex since (11) is a concave maximization problem. We can now use Lemma 4.1 to derive

$$0 = \frac{d}{d\lambda} \Big|_{\lambda=0} \tilde{\Gamma}(\lambda\tilde{f}_* + (1-\lambda)f_*, \lambda\tilde{g}_* + (1-\lambda)g_*) = \mathbb{L}(\tilde{f}_* - f_*, \tilde{g}_* - g_*).$$

As \mathbb{L} is invertible by Proposition 4.4, it follows that $(\tilde{f}_* - f_*, \tilde{g}_* - g_*) = 0$ in \mathcal{C}_\oplus . \square

5. Proofs of the Central Limit Theorems.

5.1. Consistency of the Potentials. We first state the consistency of the empirical potentials towards the population counterpart.

LEMMA 5.1. *For each $n \in \mathbb{N}$, let $(f_n, g_n) \in \mathcal{C}_\oplus$ be any solution of $\Gamma_n(f_n, g_n) = (1, 1)$. Then*

$$\|(f_n, g_n) - (f_*, g_*)\|_\oplus \xrightarrow{\text{a.s.}} 0.$$

PROOF. Recall that empirical quantities such as P_n, f_n, \dots implicitly depend on the realization $X_1(\omega), \dots, X_n(\omega), Y_1(\omega), \dots, Y_n(\omega)$ of the sample. To make this dependence explicit, we add a superscript ω . As $X_1, \dots, X_n, Y_1, \dots, Y_n$ are i.i.d., the Glivenko–Cantelli theorem implies that there is a set $\Omega_1 \subset \Omega$ with $\mathbb{P}(\Omega_1) = 1$ such that for all $\omega \in \Omega_1$,

$$\sup_{\|f\|_{0,1} \leq 1} \left| \int f d(P_n^\omega - P) \right| \rightarrow 0 \quad \text{and} \quad \sup_{\|g\|_{0,1} \leq 1} \left| \int g d(Q_n^\omega - Q) \right| \rightarrow 0.$$

By Proposition 2.3 there is a constant $K > 0$ such that $\|(f_n^\omega \oplus g_n^\omega)\|_{0,1} \leq K$ for all $n \geq 1$ and $\omega \in \Omega$. Recalling the notation $\xi_n(x, y) := f_n(x) + g_n(y) - c(x, y)$, it follows that $\|\psi'(\xi_n^\omega)\|_{0,1} \leq K'$ for a constant K' . We conclude that for all $(x_0, y_0) \in \Omega \times \Omega'$ and all $\omega \in \Omega_1$,

$$\begin{aligned} (1, 1) - \Gamma(f_n^\omega, g_n^\omega)(x_0, y_0) &= [\Gamma_n^\omega(f_n^\omega, g_n^\omega) - \Gamma(f_n^\omega, g_n^\omega)](x_0, y_0) \\ (27) \quad &= \left(\int \psi'(\xi_n^\omega(x_0, y)) d(Q_n^\omega - Q)(y) \right. \\ &\quad \left. - \int \psi'(\xi_n^\omega(x, y_0)) d(P_n^\omega - P)(x) \right) \rightarrow 0. \end{aligned}$$

Fix $\omega \in \Omega_1$. After passing to a subsequence, (f_n^ω, g_n^ω) converges in \mathcal{C}_\oplus to a limit (f_0^ω, g_0^ω) . We have $\Gamma(f_n^\omega, g_n^\omega) \rightarrow \Gamma(f_0^\omega, g_0^\omega)$ by the continuity of Γ , and thus (27) shows that $\Gamma(f_0^\omega, g_0^\omega) = (1, 1)$. The uniqueness in Theorem 3.2 now implies that $(f_0^\omega, g_0^\omega) = (f_*, g_*)$ as elements of \mathcal{C}_\oplus . More precisely, this argument shows that any convergent subsequence converges to (f_*, g_*) , and hence that the whole sequence (f_n^ω, g_n^ω) converges to (f_*, g_*) in \mathcal{C}_\oplus . As $\mathbb{P}(\Omega_1) = 1$, this amounts to the claimed a.s. convergence. \square

5.2. *Key Decomposition.* To facilitate the reading we will assume from now on that (f_n, g_n) are appropriately shifted so that

$$\|(f_n - f_*, g_n - g_*)\|_{\oplus} = \|f_n - f_*\|_{\infty} + \|g_n - g_*\|_{\infty}.$$

We also fix a Hölder exponent $\beta \in (0, 1)$. As $\int \psi''(\xi_*(\cdot, y))dQ(y)$ is bounded away from zero and continuous (Proposition 2.3), the linear operator

$$\mathbb{F} : \mathcal{C}(\Omega) \times \mathcal{C}(\Omega') \ni (f, g) \mapsto \left(\frac{\int \psi''(\xi_*(\cdot, y))dQ(y)}{\int \psi''(\xi_*(x, \cdot))dP(x)} \right) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$$

is bounded and bijective. We define the Banach space $\mathcal{B}_{\mathbb{F}, \beta}$ of all functions $(f, g) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ with finite norm

$$\|(f, g)\|_{\mathbb{F}, \beta} := \|\mathbb{F}^{-1}(f, g)\|_{0, \beta}.$$

This norm is stronger than the uniform norm. Indeed, a crucial fact for our subsequent arguments is that the unit ball of $\mathcal{B}_{\mathbb{F}, \beta}$ is compactly embedded in $\mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ (by the same reasoning as in [19, Lemma 6.33]).

It follows from Theorem B.4¹ of Appendix B that the central limit theorem

$$\sqrt{n}(\Gamma(f_*, g_*) - \Gamma_n(f_*, g_*)) = \sqrt{n} \left(\int \psi'(\xi_*(\cdot, y))d(Q - Q_n)(y) \right) \xrightarrow{w} \begin{pmatrix} \mathbf{G}_Q \\ \mathbf{G}_P \end{pmatrix}$$

holds in $\mathcal{C}^{0, \beta}(\Omega) \times \mathcal{C}^{0, \beta}(\Omega')$, where $\mathbf{G}_Q, \mathbf{G}_P$ are Gaussian processes as detailed in Theorem 3.4. As a consequence, the isometric image $\tilde{\Gamma}_n = \mathbb{F}(\Gamma_n)$ satisfies the following central limit theorem in the space $\mathcal{B}_{\mathbb{F}, \beta}$.

LEMMA 5.2. *The following weak limit holds in $\mathcal{B}_{\mathbb{F}, \beta}$,*

$$\sqrt{n}[\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) := \sqrt{n} \left(\tilde{\Gamma}(f_*, g_*) - \tilde{\Gamma}_n(f_*, g_*) \right) \xrightarrow{w} \begin{pmatrix} \mathbf{G}_Q \\ \mathbf{G}_P \end{pmatrix}.$$

In particular, the sequence $\sqrt{n}[\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*)$ is tight in $\mathcal{B}_{\mathbb{F}, \beta}$.

The goal of this subsection, and indeed the key technical result, is the following decomposition of $(f_n - f_*, g_n - g_*)$ into three terms.

PROPOSITION 5.3. *There exist random functions $U_n, V_n, W_n \in \mathcal{C}(\Omega)$ and $U'_n, V'_n, W'_n \in \mathcal{C}(\Omega')$, and compact sets $\mathcal{K} \subset \mathcal{C}(\Omega)$ and $\mathcal{K}' \subset \mathcal{C}(\Omega')$, such that*

- (i) (U_n, U'_n) takes values in $\mathcal{K} \times \mathcal{K}'$ for all n ,
- (ii) $\|(V_n, V'_n)\|_{\infty} \xrightarrow{a.s.} 0$,
- (iii) $\mathbb{L}(W_n, W'_n)$ is tight in $\mathcal{B}_{\mathbb{F}, \beta}$,
- (iv) for all n ,

$$(28) \quad \begin{pmatrix} f_n - f_* \\ g_n - g_* \end{pmatrix} = \|f_n - f_*, g_n - g_*\|_{\oplus} \begin{pmatrix} U_n + V_n \\ U'_n + V'_n \end{pmatrix} + n^{-\frac{1}{2}} \begin{pmatrix} W_n \\ W'_n \end{pmatrix}.$$

Next, we collect the higher-level steps of the proof, while outsourcing the more technical parts to the subsequent lemmas.

¹To see that its condition is satisfied, note that $\psi'(\xi_*(\cdot, Y_1)) - \int \psi'(\xi_*(\cdot, y))dQ(y)$ is tight in $\mathcal{C}^{0,1}(\Omega)$ as the composition of the measurable function $Y_1 : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}^d$ and the continuous function $v \mapsto \psi'(\xi_*(\cdot, v)) - \int \psi'(\xi_*(\cdot, y))dQ(y) \in \mathcal{C}^{0,1}(\Omega)$.

PROOF OF PROPOSITION 5.3. By Lemma 4.1 there are random functions $(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}}) \in \mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ with $\|(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}})\|_\infty \xrightarrow{\text{a.s.}} 0$ such that

$$\tilde{\Gamma}(f_n, g_n) - \tilde{\Gamma}(f_*, g_*) = \mathbb{L}(f_n - f_*, g_n - g_*) - \|(f_n - f_*, g_n - g_*)\|_{\oplus}(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}}).$$

Using the fact that $\Gamma(f_*, g_*) = (1, 1) = \Gamma_n(f_n, g_n)$ and setting

$$\tilde{\Gamma}_n := \left(\frac{\Gamma_n^{(1)}}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \right),$$

this yields

$$(29) \quad \tilde{\Gamma}(f_n, g_n) - \tilde{\Gamma}_n(f_n, g_n) = \mathbb{L}(f_n - f_*, g_n - g_*) - \|(f_n - f_*, g_n - g_*)\|_{\oplus}(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}}).$$

Recall from Proposition 4.4 that \mathbb{L} is a bounded bijection. Moreover, $\|\tilde{\Gamma}(f_* - f_n, g_* - g_n)\|_{\mathcal{C}_{\oplus}} \rightarrow 0$ a.s. by Lemma 5.1 and the continuity of $\tilde{\Gamma}$. Setting also

$$(30) \quad \Delta_n := [\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) - [\tilde{\Gamma} - \tilde{\Gamma}_n](f_n, g_n),$$

we deduce

$$\begin{aligned} & (f_n - f_*, g_n - g_*) \\ &= \mathbb{L}^{-1}([\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*)) - \mathbb{L}^{-1}(\Delta_n) + \|(f_n - f_*, g_n - g_*)\|_{\oplus} \mathbb{L}^{-1}(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}}). \end{aligned}$$

Set $(W_n, W'_n) = n^{\frac{1}{2}} \mathbb{L}^{-1}([\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*))$. Then $\mathbb{L}(W_n, W'_n)$ is tight by Lemma 5.2 and $\mathbb{L}^{-1}([\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*)) = n^{-\frac{1}{2}}(W_n, W'_n)$. Therefore,

$$(f_n - f_*, g_n - g_*) = n^{-\frac{1}{2}}(W_n, W'_n) - \mathbb{L}^{-1}(\Delta_n) + \|(f_n - f_*, g_n - g_*)\|_{\oplus} \mathbb{L}^{-1}(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}}).$$

Lemma 5.4 below derives a decomposition for the remaining term Δ_n which completes the proof of Proposition 5.3 after recalling that \mathbb{L}^{-1} is continuous. Specifically, the function (V_n, V'_n) in the assertion of Proposition 5.3 arises from combining the function $\mathbb{L}^{-1}(V_{n,\mathbb{L}}, V'_{n,\mathbb{L}})$ above with the function $\mathbb{L}^{-1}(\tilde{V}_n, \tilde{V}'_n)$ from Lemma 5.4. \square

The following lemma encapsulates the main step of the preceding proof of Proposition 5.3 and uses the notation Δ_n from (30).

LEMMA 5.4. *There exist compact sets $\tilde{\mathcal{K}} \subset \mathcal{C}(\Omega)$ and $\tilde{\mathcal{K}}' \subset \mathcal{C}(\Omega')$, and random functions $\tilde{U}_n, \tilde{V}_n \in \mathcal{C}(\Omega)$ and $\tilde{U}'_n, \tilde{V}'_n \in \mathcal{C}(\Omega')$, such that*

- (i) $(\tilde{U}_n, \tilde{U}'_n)$ takes values in $\tilde{\mathcal{K}} \times \tilde{\mathcal{K}}'$ for all n ,
- (ii) $\|\tilde{V}_n, \tilde{V}'_n\|_\infty \xrightarrow{\text{a.s.}} 0$,
- (iii) for all n ,

$$\Delta_n = \left(\frac{\|f_n - f_*\|_\infty \tilde{V}_n}{\|g_n - g_*\|_\infty \tilde{V}'_n} \right) + \left(\frac{\|g_n - g_*\|_\infty \tilde{U}_n}{\|f_n - f_*\|_\infty \tilde{U}'_n} \right).$$

PROOF. We prove the claims concerning the functions on Ω ; the functions on Ω' are obtained analogously. We start by recalling that

$$\Gamma^{(1)}(f_n, g_n) - \Gamma_n^{(1)}(f_n, g_n) = \int \psi'(\xi_n(\cdot, y)) d(Q - Q_n)(y)$$

and

$$\Gamma^{(1)}(f_*, g_*) - \Gamma_n^{(1)}(f_*, g_*) = \int \psi'(\xi_*(\cdot, y)) d(Q - Q_n)(y).$$

Set

$$\Delta_n^{(1)} := \Gamma^{(1)}(f_*, g_*) - \Gamma_n^{(1)}(f_*, g_*) - (\Gamma^{(1)}(f_n, g_n) - \Gamma_n^{(1)}(f_n, g_n)).$$

Note that writing $\Delta_n^{(1)}$ is a slight abuse of notation: defining analogously $\Delta_n^{(2)}$, we have

$$(31) \quad \Delta_n = \left(\frac{\Delta_n^{(1)}}{\int \psi''(\xi_*(\cdot, y)) dQ(y)}, \frac{\Delta_n^{(2)}}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \right).$$

The fundamental theorem of calculus implies that

$$\begin{aligned} \Delta_n^{(1)} &= \int \{\psi'(\xi_*(\cdot, y)) - \psi'(\xi_n(\cdot, y))\} d(Q - Q_n)(y) \\ &= \int \int_0^1 \psi''((1 - \lambda)\xi_n(\cdot, y) + \lambda\xi_*(\cdot, y)) (\xi_*(\cdot, y) - \xi_n(\cdot, y)) d\lambda d(Q - Q_n)(y). \end{aligned}$$

We set $\psi_n(\cdot, y) := \int_0^1 \psi''((1 - \lambda)\xi_n(\cdot, y) + \lambda\xi_*(\cdot, y)) d\lambda$ and use the definitions of ξ_* , ξ_n to deduce

$$(32) \quad \begin{aligned} \Delta_n^{(1)} &= \int \psi_n(\cdot, y) (\xi_*(\cdot, y) - \xi_n(\cdot, y)) d(Q - Q_n)(y) \\ &= (f_* - f_n) \int \psi_n(\cdot, y) d(Q - Q_n)(y) + \int \psi_n(\cdot, y) (g_*(y) - g_n(y)) d(Q - Q_n)(y). \end{aligned}$$

Considering the first term in (32) and adding back the denominator from (31), set

$$\tilde{V}_n := \frac{1}{\int \psi''(\xi_*(\cdot, y)) dQ(y)} \int \psi_n(\cdot, y) d(Q - Q_n)(y).$$

Then \tilde{V}_n is a random variable with values in $\mathcal{C}(\Omega)$. As ξ_n, ξ_* have uniformly bounded Lipschitz norms by Proposition 2.3, it follows from Lemma 5.5 below that $\|\tilde{V}_n\|_\infty \rightarrow 0$ a.s. Regarding the second term in (32), set (using the convention $0/0 := 0$ if necessary)

$$\tilde{U}_n := \frac{1}{\int \psi''(\xi_*(x, \cdot)) dP(x)} \int \psi_n(\cdot, y) \frac{g_*(y) - g_n(y)}{\|g_n - g_*\|_\infty} d(Q - Q_n)(y).$$

Then Lemma 5.6 below, applied with $d\mu(y) := \frac{g_*(y) - g_n(y)}{\|g_n - g_*\|_\infty} d(Q - Q_n)(y)$, shows that \tilde{U}_n is a random variable with values in a compact set $\tilde{\mathcal{K}} \subset \mathcal{C}(\Omega)$. \square

The next two technical lemmas were used in the preceding proof of Lemma 5.4.

LEMMA 5.5. *For any $C \in \mathbb{R}$, the class of functions*

$$\mathcal{F} := \left\{ \Omega' \ni y \mapsto \int_0^1 \psi''(h(x, y, \lambda)) d\lambda : x \in \Omega, \|h\|_{0,1} \leq C \right\}$$

is relatively compact in $\mathcal{C}(\Omega)$. As a consequence,

$$\sup_{f \in \mathcal{F}} \left| \int f d(Q_n - Q) \right| \xrightarrow{a.s.} 0.$$

PROOF. Let ρ be a modulus of continuity of ψ'' on $[-C, C]$, i.e., $\psi''(t_1) - \psi''(t_2) \leq \rho(|t_1 - t_2|)$ for all $t_1, t_2 \in [-C, C]$, where ρ is monotone increasing and $\rho(s) \rightarrow 0$ as $s \rightarrow 0$. Then the modulus of continuity of $\int_0^1 \psi''(h(x, \cdot, \lambda))d\lambda$ is bounded as follows,

$$\begin{aligned} & \left| \int_0^1 \psi''(h(x, y_1, \lambda))d\lambda - \int_0^1 \psi''(h(x, y_2, \lambda))d\lambda \right| \\ & \leq \int_0^1 |\psi''(h(x, y_1, \lambda)) - \psi''(h(x, y_2, \lambda))|d\lambda \\ & \leq \int_0^1 \rho(|h(x, y_1, \lambda) - h(x, y_2, \lambda)|)d\lambda \\ & \leq \rho(C\|y_1 - y_2\|). \end{aligned}$$

In view of the Arzelà–Ascoli theorem, this implies that \mathcal{F} is relatively compact in $\mathcal{C}(\Omega)$. The second claim follows from this compactness; see [42, Theorem 2.4.1]. \square

LEMMA 5.6. Set $\psi_n(\cdot, y) := \int_0^1 \psi''((1 - \lambda)\xi_n(\cdot, y) + \lambda\xi_*(\cdot, y))d\lambda$. The family of functions

$$\int \psi_n(\cdot, y)d\mu(y), \quad n \geq 1$$

where $\mu \in \mathcal{C}(\Omega)'$ is any signed measure with $\|\mu\|_{\text{TV}} \leq 2$, is relatively compact in $\mathcal{C}(\Omega')$.

PROOF. Similarly as in the proof of Lemma 5.5, these functions are uniformly bounded and admit a common modulus of continuity. \square

We conclude this subsection with a corollary of Proposition 5.3 that will be used to prove the central limit theorem, Theorem 3.4.

COROLLARY 5.7. The sequence $\Delta_n := [\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) - [\tilde{\Gamma} - \tilde{\Gamma}_n](f_n, g_n)$ satisfies

$$\|\Delta_n\|_{\oplus} = o_{\mathbb{P}}\left(\|(f_* - f_n, g_* - g_n)\|_{\oplus} + n^{-\frac{1}{2}}\right).$$

PROOF. We write Δ_n as in (31) and detail the proof for the first component only. As in (32),

$$(33) \quad \Delta_n^{(1)} = \int \psi_n(\cdot, y)(f_* - f_n + g_*(y) - g_n(y))d(Q - Q_n)(y).$$

Recall the notation from Proposition 5.3, in particular $U_n, V_n, W_n \in \mathcal{C}(\Omega)$ and $U'_n, V'_n, W'_n \in \mathcal{C}(\Omega')$, and the compact sets \mathcal{K} and \mathcal{K}' . Inserting (28) into (33), we have

$$\begin{aligned} \Delta_n^{(1)} &= \underbrace{\|(f_* - f_n, g_* - g_n)\|_{\oplus} \int \psi_n(\cdot, y)(U_n + U'_n(y))d(Q - Q_n)(y)}_{=:A_n} \\ &+ \underbrace{\|(f_* - f_n, g_* - g_n)\|_{\oplus} \int \psi_n(\cdot, y)(V_n + V'_n(y))d(Q - Q_n)(y)}_{=:B_n} \\ &+ n^{-\frac{1}{2}} \underbrace{\int \psi_n(\cdot, y)(W_n + W'_n(y))d(Q - Q_n)(y)}_{=:C_n}. \end{aligned}$$

As $\psi_n(x, \cdot)$ is an equicontinuous and bounded family,

$$\mathcal{F} := \{y \mapsto \psi_n(x, y)(f(x) + g(y)) : (f, g) \in \mathcal{K} \times \mathcal{K}', x \in \Omega, n \geq 1\}$$

is relatively compact in $\mathcal{C}(\Omega')$. Thus \mathcal{F} is uniformly Glivenko–Cantelli and we get

$$(34) \quad \begin{aligned} \|A_n\|_\infty &\leq \|(f_* - f_n, g_* - g_n)\|_\oplus \sup_{g \in \mathcal{F}} \left| \int g(y) d(Q - Q_n)(y) \right| \\ &= o_{\mathbb{P}}(\|(f_* - f_n, g_* - g_n)\|_\oplus). \end{aligned}$$

Next, recalling $\|(V_n, V'_n)\|_\infty \xrightarrow{\text{a.s.}} 0$, the bound

$$\begin{aligned} \|B_n\|_\infty &\leq \|(f_* - f_n, g_* - g_n)\|_\oplus \|\psi_n\|_\infty \|(V_n, V'_n)\|_\infty \sup_{\|g\|_\infty \leq 1} \left| \int g(y) d(Q - Q_n)(y) \right| \\ &\leq 2\|(f_* - f_n, g_* - g_n)\|_\oplus \|\psi_n\|_\infty \|(V_n, V'_n)\|_\infty \end{aligned}$$

yields

$$(35) \quad \|B_n\|_\infty = o_{\mathbb{P}}(\|(f_* - f_n, g_* - g_n)\|_\oplus).$$

Finally, as the unit ball wrt. $\|\cdot\|_{\mathbb{F}, \beta}$ is compact in $\mathcal{C}(\Omega) \times \mathcal{C}(\Omega')$ and \mathbb{L} is a continuous bijection, the class

$$\mathcal{F}' := \{y \mapsto \psi_n(x, y)(f(x) + g(y)) : \|\mathbb{L}(f, g)\|_{\mathbb{F}, \beta} \leq 1, x \in \Omega, n \geq 1\}$$

is uniformly Glivenko–Cantelli. Since $\mathbb{L}(W_n, W'_n)$ is tight in $\mathcal{B}_{\mathbb{F}, \beta}$, $\|\mathbb{L}(W_n, W'_n)\|_{\mathbb{F}, \beta}$ is stochastically bounded and we deduce

$$(36) \quad \|C_n\|_\infty \leq n^{-\frac{1}{2}} \|\mathbb{L}(W_n, W'_n)\|_{\mathbb{F}, \beta} \sup_{g \in \mathcal{F}'} \left| \int g(y) d(Q - Q_n)(y) \right| \leq o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

Combining the bounds (34), (35) and (36), we conclude

$$\|\Delta_n^{(1)}\|_\infty = o_{\mathbb{P}}\left(\|(f_* - f_n, g_* - g_n)\|_\oplus + n^{-\frac{1}{2}}\right).$$

The result follows since the denominator in (31) is continuous and bounded away from zero. \square

5.3. Deriving the central limit theorems. We can now establish the central limit theorems. We begin with the one for the potentials.

PROOF OF THEOREM 3.4. Armed with the results of Sections 5.1 and 5.2, the central limit theorem for the potentials follows by the standard reasoning in Z -estimation; see [42, Theorem 3.3.1]. We give the details for the sake of completeness. We can write (29) as

$$\begin{aligned} \mathbb{L}(f_n - f_*, g_n - g_*) &= [\tilde{\Gamma} - \tilde{\Gamma}_n](f_n, g_n) + \|(f_n - f_*, g_n - g_*)\|_\oplus (V_{n, \mathbb{L}}, V'_{n, \mathbb{L}}) \\ &= [\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) - \Delta_n + \|(f_n - f_*, g_n - g_*)\|_\oplus (V_{n, \mathbb{L}}, V'_{n, \mathbb{L}}). \end{aligned}$$

In view of Lemma 5.2 and Corollary 5.7, we deduce that

$$\|\mathbb{L}(f_n - f_*, g_n - g_*)\|_\oplus = o_{\mathbb{P}}(\|(f_* - f_n, g_* - g_n)\|_\oplus) + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

Recalling that \mathbb{L} has a bounded inverse (Proposition 4.4), it follows that

$$\|(f_n - f_*, g_n - g_*)\|_\oplus \leq o_{\mathbb{P}}(\|(f_* - f_n, g_* - g_n)\|_\oplus) + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

As we also have $\|(f_n - f_*, g_n - g_*)\|_{\oplus} = o_{\mathbb{P}}(1)$ by Lemma 5.1, this implies the rate $\|(f_n - f_*, g_n - g_*)\|_{\oplus} = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. Substituting this into (29), we obtain

$$\left\| \tilde{\Gamma}(f_n, g_n) - \tilde{\Gamma}_n(f_n, g_n) - \mathbb{L}(f_n - f_*, g_n - g_*) \right\|_{\oplus} = o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$$

which, using again Corollary 5.7 and $\|(f_n - f_*, g_n - g_*)\|_{\oplus} = o_{\mathbb{P}}(1)$, leads to

$$(37) \quad \left\| \tilde{\Gamma}(f_*, g_*) - \tilde{\Gamma}_n(f_*, g_*) - \mathbb{L}(f_n - f_*, g_n - g_*) \right\|_{\oplus} = o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

Denote by $\tilde{\mathbf{G}}$ the right-hand side in Lemma 5.2. As $n^{1/2}(\tilde{\Gamma}(f_*, g_*) - \tilde{\Gamma}_n(f_*, g_*)) \xrightarrow{w} \tilde{\mathbf{G}}$ in \mathcal{C}_{\oplus} by Lemma 5.2, (37) implies that

$$n^{1/2} \mathbb{L}(f_n - f_*, g_n - g_*) \xrightarrow{w} \tilde{\mathbf{G}}$$

in \mathcal{C}_{\oplus} as well. Using that \mathbb{L}^{-1} is continuous (Proposition 4.4) and applying the continuous mapping theorem, we conclude that

$$n^{1/2} (f_n - f_*, g_n - g_*) \xrightarrow{w} \mathbb{L}^{-1} \tilde{\mathbf{G}},$$

which was the claim. \square

We continue with the central limit theorem for the optimal cost.

PROOF OF THEOREM 3.5. From the fact that (f_*, g_*) and (f_n, g_n) are optimizers of the dual problem (11) for (P, Q) and (P_n, Q_n) , respectively, we deduce the two inequalities

$$\begin{aligned} & \text{ROT}(P_n, Q_n) - \text{ROT}(P, Q) \\ & \geq \int \{f_*(x) + g_*(y) - \psi(f_*(x) + g_*(y) - c(x, y))\} d((P_n \otimes Q_n) - (P \otimes Q))(x, y) \end{aligned}$$

and

$$\begin{aligned} & \text{ROT}(P_n, Q_n) - \text{ROT}(P, Q) \\ & \leq \int \{f_n(x) + g_n(y) - \psi(f_n(x) + g_n(y) - c(x, y))\} d((P_n \otimes Q_n) - (P \otimes Q))(x, y). \end{aligned}$$

Using our shorthands (15) and dropping the integration variables from the notation, the difference between the upper and the lower bound is

$$\begin{aligned} (38) \quad & \int \{f_n(x) + g_n(y) - \psi(f_n(x) + g_n(y) - c(x, y)) \\ & - (f_*(x) + g_*(y) - \psi(f_*(x) + g_*(y) - c(x, y)))\} d((P_n \otimes Q_n) - (P \otimes Q))(x, y) \\ & = \int \{(\xi_n - \xi_*) - (\psi(\xi_n) - \psi(\xi_*))\} d((P_n \otimes Q_n) - (P \otimes Q)). \end{aligned}$$

Using the fundamental theorem of calculus in the same way as below (31), and setting $\tilde{\psi}_n := 1 + \int_0^1 \psi'((1 - \lambda)\xi_n + \lambda\xi_*) d\lambda$, the right-hand side of (38) can be written as

$$\int \tilde{\psi}_n \cdot (\xi_n - \xi_*) d((P_n \otimes Q_n) - (P \otimes Q))$$

which has the same structure as (33), except that now the integration is over both variables. We proceed as below (33): insert the decomposition (28), split the integral into three terms, and treat them with the same arguments as in Eqs. (34) to (36). The result is that

$$(38) = o_{\mathbb{P}} \left(\|(f_* - f_n, g_* - g_n)\|_{\oplus} + n^{-\frac{1}{2}} \right).$$

From Theorem 3.4 we further know that $\|(f_* - f_n, g_* - g_n)\|_{\oplus} = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$, whence we conclude that (38) $= o_{\mathbb{P}}(n^{-\frac{1}{2}})$. Recall that (38) is the difference of the upper and lower bounds for $\text{ROT}(P_n, Q_n) - \text{ROT}(P, Q)$. As a consequence, $\text{ROT}(P_n, Q_n) - \text{ROT}(P, Q)$ equals the lower bound up to $o_{\mathbb{P}}(n^{-\frac{1}{2}})$; to wit,

$$\begin{aligned} & \text{ROT}(P_n, Q_n) - \text{ROT}(P, Q) + o_{\mathbb{P}} \left(n^{-\frac{1}{2}} \right) \\ &= \int \left\{ (f_*(x) + g_*(y) - \psi(f_*(x) + g_*(y) - c(x, y))) \right\} d((P_n \otimes Q_n) - (P \otimes Q))(x, y). \end{aligned}$$

We can now apply the central limit theorem for U -statistics [41, Theorem 12.6] to derive the result. (In fact, to obtain the formula for the variance σ^2 , it is easier to directly use the formula of the Hájek projection stated just before [41, Theorem 12.6].) \square

Finally, we show the central limit theorem for the couplings.

PROOF OF THEOREM 3.6. Note that

$$\int \eta d(\pi_n - \pi) = \int \bar{\eta} d(\pi_n - \pi).$$

Recalling from Proposition 2.3 that $\frac{d\pi}{dP \otimes Q} = \psi'(\xi_*)$ and $\frac{d\pi_n}{dP_n \otimes Q_n} = \psi'(\xi_n)$, we obtain

$$\begin{aligned} \int \eta d(\pi_n - \pi) &= \underbrace{\int \bar{\eta} \{ \psi'(\xi_n) - \psi'(\xi_*) \} d((P_n \otimes Q_n) - (P \otimes Q))}_{=: A_n} \\ &+ \underbrace{\int \bar{\eta} \{ \psi'(\xi_n) - \psi'(\xi_*) \} d(P \otimes Q)}_{=: B_n} \\ &+ \underbrace{\int \bar{\eta} \psi'(\xi_*) d((P_n \otimes Q_n) - (P \otimes Q))}_{=: C_n}. \end{aligned}$$

Arguing as in (34) and then using Theorem 3.4, we have

$$A_n = o_{\mathbb{P}} \left(\|(f_* - f_n, g_* - g_n)\|_{\oplus} \right) = o_{\mathbb{P}} \left(n^{-\frac{1}{2}} \right).$$

Turning to B_n , a first-order Taylor development of ψ' yields

$$\begin{aligned} B_n &= \int (\xi_n - \xi_*) \bar{\eta} \psi''(\xi_*) d(P \otimes Q) + o_{\mathbb{P}} \left(\|(f_n - f_*, g_n - g_*)\|_{\oplus} \right) \\ &= \int (\xi_n - \xi_*) \bar{\eta} \psi''(\xi_*) d(P \otimes Q) + o_{\mathbb{P}} \left(n^{-\frac{1}{2}} \right). \end{aligned}$$

Moreover, we see from (37) that

$$\left\| (f_n - f_*, g_n - g_*) - \mathbb{L}^{-1} \left([\tilde{\Gamma} - \tilde{\Gamma}_n](f_*, g_*) \right) \right\|_{\oplus} = o_{\mathbb{P}} \left(n^{-\frac{1}{2}} \right).$$

After inserting the definitions of $\tilde{\Gamma}$ and $\tilde{\Gamma}_n$, this yields the third equality in

$$\begin{aligned}
B_n &= \int (\xi_n - \xi_*) \bar{\eta} \psi''(\xi_*) d(P \otimes Q) + o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) \\
&= \int (\oplus(f_n - f_*, g_n - g_*)) \bar{\eta} \psi''(\xi_*) d(P \otimes Q) + o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) \\
&= - \int \oplus \left(\mathbb{L}^{-1} \left[\begin{array}{c} \int \psi'(\xi_*(\cdot, y)) d(Q_n - Q)(y) \\ \int \psi''(\xi_*(\cdot, y)) dQ(y) \\ \int \psi'(\xi_*(x, \cdot)) d(P_n - P)(x) \\ \int \psi''(\xi_*(x, \cdot)) dP(x) \end{array} \right] \right) \bar{\eta} \psi''(\xi_*) d(P \otimes Q) + o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) \\
&= - \frac{1}{n^2} \sum_{i,j=1}^n \int \left\{ \oplus \left(\mathbb{L}^{-1} \left[\begin{array}{c} \psi'(\xi_*(\cdot, Y_j)) - \int \psi'(\xi_*(\cdot, y)) dQ(y) \\ \int \psi''(\xi_*(\cdot, y)) dQ(y) \\ \psi'(\xi_*(X_i, \cdot)) - \int \psi'(\xi_*(x, \cdot)) dP(x) \\ \int \psi''(\xi_*(x, \cdot)) dP(x) \end{array} \right] \right) \bar{\eta} \psi''(\xi_*) \right\} d(P \otimes Q) \\
&\quad + o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) =: \bar{B}_n + o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

Note also that \bar{B}_n is centered (w.r.t. \mathbb{P}) because the precomposition of a centered random variable with a bounded linear operator is again centered. Turning to C_n , we can write

$$C_n = \frac{1}{n^2} \sum_{i,j=1}^n \bar{\eta}(X_i, Y_j) \psi'(\xi_*(X_i, Y_j)) - \int \bar{\eta} \psi'(\xi_*) d(P \otimes Q)$$

which is also centered. Moreover, the variances of \bar{B}_n and C_n are bounded by a constant depending only on $\|\xi_*\|_{\infty}$, $\|\eta\|_{\infty}$,

$$\left\| \left(\int \psi''(\xi_*(\cdot, y)) dQ(y) \right)^{-1} \right\|_{\infty} \quad \text{and} \quad \left\| \left(\int \psi''(\xi_*(x, \cdot)) dP(x) \right)^{-1} \right\|_{\infty}.$$

In summary, up to negligible terms, $\int \eta d(\pi_n - \pi)$ is a U -statistic with finite variance. We can then apply the central limit theorem for U -statistics (see [41, Theorem 12.6]) to conclude the result. (Once again, to obtain the formula for the variance $\sigma^2(\eta)$, it is easier to directly use the formula of the Hájek projection stated just before [41, Theorem 12.6].) \square

Funding. SE is grateful for support by the German Research Foundation through Project 553088969 as well as the Cluster of Excellence ‘‘Machine Learning — New Perspectives for Science’’ (EXC 2064/1 number 390727645). MN acknowledges funding by NSF Grants DMS-2407074 and DMS-2106056.

REFERENCES

- [1] BAYRAKTAR, E., ECKSTEIN, S. and ZHANG, X. (2025). Stability and sample complexity of divergence regularized optimal transport. *Bernoulli* **31** 213–239.
- [2] BLONDEL, M., SEGUY, V. and ROLET, A. (2018). Smooth and Sparse Optimal Transport. *Proceedings of Machine Learning Research* **84** 880–889.
- [3] CARLIER, G. and LABORDE, M. (2020). A Differential Approach to the Multi-Marginal Schrödinger System. *SIAM Journal on Mathematical Analysis* **52** 709–717. <https://doi.org/10.1137/19M1253800>
- [4] CHEWI, S., NILES-WEED, J. and RIGOLLET, P. (2025). *Statistical Optimal Transport*. Springer.
- [5] CHOW, Y. S. and TEICHER, H. (1997). *Probability Theory*. Springer New York. <https://doi.org/10.1007/978-1-4612-1950-7>
- [6] CÁRCAMO, J., CUEVAS, A. and RODRÍGUEZ, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* **26**. <https://doi.org/10.3150/19-bej1188>
- [7] DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability* **47** 926 – 951. <https://doi.org/10.1214/18-AOP1275>

- [8] DEL BARRIO, E. and LOUBES, J.-M. (2020). The statistical effect of entropic regularization in optimal transportation. *arXiv:2006.05199*.
- [9] DEL BARRIO, E., SANZ, A. G., LOUBES, J.-M. and NILES-WEED, J. (2023). An Improved Central Limit Theorem and Fast Convergence Rates for Entropic Transportation Costs. *SIAM Journal on Mathematics of Data Science* **5** 639–669. <https://doi.org/10.1137/22M149260X>
- [10] DESSEIN, A., PAPADAKIS, N. and ROUAS, J.-L. (2018). Regularized Optimal Transport and the rot mover’s Distance. *J. Mach. Learn. Res.* **19** 1–53.
- [11] DI MARINO, S. and GEROLIN, A. (2020). Optimal Transport losses and Sinkhorn algorithm with general convex regularization. *Preprint arXiv:2007.00976v1*.
- [12] DUDLEY, R. M. (1968). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.* **40** 40–50. [MR236977](https://doi.org/10.2307/236977)
- [13] ECKSTEIN, S. and KUPPER, M. (2021). Computation of optimal transport and related hedging problems via penalization and neural networks. *Appl. Math. Optim.* **83** 639–667. [MR4239795](https://doi.org/10.1007/s00440-014-0583-7)
- [14] ECKSTEIN, S. and NUTZ, M. (2024). Convergence Rates for Regularized Optimal Transport via Quantization. *Math. Oper. Res.* **49** 1223–1240. <https://doi.org/10.1287/moor.2022.0245>
- [15] ESSID, M. and SOLOMON, J. (2018). Quadratically regularized optimal transport on graphs. *SIAM J. Sci. Comput.* **40** A1961–A1986. [MR3820384](https://doi.org/10.1137/17S10915228)
- [16] FOURNIER, N. and GUILLIN, A. (2014). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162** 707–738. <https://doi.org/10.1007/s00440-014-0583-7>
- [17] GARRIZ-MOLINA, A., GONZÁLEZ-SANZ, A. and MORDANT, G. (2024). Infinitesimal behavior of Quadratically Regularized Optimal Transport and its relation with the Porous Medium Equation. *arXiv:2407.21528*.
- [18] GENEVAY, A., CHIZAT, L., BACH, F., CUTURI, M. and PEYRÉ, G. (2019). Sample Complexity of Sinkhorn Divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (K. CHAUDHURI and M. SUGIYAMA, eds.). *Proceedings of Machine Learning Research* **89** 1574–1583. PMLR.
- [19] GILBARG, D. and TRUDINGER, N. S. (1983). *Elliptic partial differential equations of second order*, second ed. *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **224**. Springer-Verlag, Berlin. <https://doi.org/10.1007/978-3-642-61798-0> [MR737190](https://doi.org/10.1007/978-3-642-61798-0)
- [20] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Statistical inference with regularized optimal transport. *Inf. Inference* **13** Paper No. 13, 68. <https://doi.org/10.1093/imaiai/iaad056>
- [21] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Limit theorems for entropic optimal transport maps and Sinkhorn divergence. *Electronic Journal of Statistics* **18**. <https://doi.org/10.1214/24-ejs2217>
- [22] GONZÁLEZ-SANZ, A. and HUNDRIESER, S. (2023). Weak Limits for Empirical Entropic Optimal Transport: Beyond Smooth Costs. *arXiv:2305.09745*.
- [23] GONZÁLEZ-SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2024). Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences. *arXiv:2207.07427*.
- [24] GONZÁLEZ-SANZ, A. and NUTZ, M. (2024). Quantitative Convergence of Quadratically Regularized Linear Programs. *arXiv:2408.04088*.
- [25] GONZÁLEZ-SANZ, A. and NUTZ, M. (2024). Sparsity of Quadratically Regularized Optimal Transport: Scalar Case. *arXiv:2410.03353*.
- [26] GULRAJANI, I., AHMED, F., ARJOVSKY, M., DUMOULIN, V. and COURVILLE, A. (2017). Improved Training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 5769–5779.
- [27] JAIN, N. C. (1976). Central limit theorem in a Banach space. In *Probability in Banach Spaces* (A. BECK, ed.) 113–130. Springer, Berlin, Heidelberg.
- [28] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-20212-4>
- [29] LI, L., GENEVAY, A., YUROCHKIN, M. and SOLOMON, J. M. (2020). Continuous Regularized Wasserstein Barycenters. In *Advances in Neural Information Processing Systems* (H. LAROCHELLE, M. RANZATO, R. HADSELL, M. F. BALCAN and H. LIN, eds.) **33** 17755–17765. Curran Associates, Inc.
- [30] LORENZ, D. and MAHLER, H. (2022). Orlicz space regularization of continuous optimal transport problems. *Appl. Math. Optim.* **85** Paper No. 14, 33. [MR4409806](https://doi.org/10.1007/s00440-014-0583-7)
- [31] LORENZ, D., MANNS, P. and MEYER, C. (2021). Quadratically regularized optimal transport. *Appl. Math. Optim.* **83** 1919–1949. [MR4261277](https://doi.org/10.1007/s00440-014-0583-7)
- [32] MALLASTO, A., GEROLIN, A. and MINH, H. Q. (2021). Entropy-regularized 2-Wasserstein distance between Gaussian measures. *Information Geometry* **5** 289–323. <https://doi.org/10.1007/s41884-021-00052-8>
- [33] MANOLE, T. and NILES-WEED, J. (2024). Sharp convergence rates for empirical optimal transport with smooth costs. *Ann. Appl. Probab.* **34** 1108–1135. <https://doi.org/10.1214/23-aap1986> [MR4700254](https://doi.org/10.1214/23-aap1986)

- [34] MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems* (H. WALLACH, H. LAROCHELLE, A. BEYGELZIMER, F. D'ALCHÉ-BUC, E. FOX and R. GARNETT, eds.) **32**. Curran Associates, Inc.
- [35] MUZELLEC, B., NOCK, R., PATRINI, G. and NIELSEN, F. (2017). Tsallis Regularized Optimal Transport and Ecological Inference. *Proceedings of the AAAI Conference on Artificial Intelligence* **31**. <https://doi.org/10.1609/aaai.v31i1.10854>
- [36] NUTZ, M. (2024). Quadratically Regularized Optimal Transport: Existence and Multiplicity of Potentials. *SIAM J. Math. Anal.*, to appear.
- [37] PEYRÉ, G. and CUTURI, M. (2019). Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning* **11** 355–607. <https://doi.org/10.1561/22000000073>
- [38] RIGOLLET, P. and STROMME, A. J. (2025). On the sample complexity of entropic optimal transport. *Ann. Statist.* **53** 61–90. <https://doi.org/10.1214/24-aos2455> MR4865008
- [39] SCHMITZER, B. (2019). Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J. Sci. Comput.* **41** A1443–A1481. <https://doi.org/10.1137/16M1106018> MR3947294
- [40] SEGUY, V., DAMODARAN, B. B., FLAMARY, R., COURTY, N., ROLET, A. and BLONDEL, M. (2018). Large Scale Optimal Transport and Mapping Estimation. In *International Conference on Learning Representations*.
- [41] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511802256>
- [42] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer. <https://doi.org/10.1007/978-1-4757-2545-2>
- [43] VILLANI, C. (2009). *Optimal Transport: Old and New*. Springer-Verlag, Berlin. <https://doi.org/10.1007/978-3-540-71050-9> MR2459454
- [44] WIESEL, J. and XU, X. (2024). Sparsity of Quadratically Regularized Optimal Transport: Bounds on concentration and bias. *Preprint arXiv:2410.03425v1*.
- [45] ZHANG, S., MORDANT, G., MATSUMOTO, T. and SCHIEBINGER, G. (2023). Manifold Learning with Sparse Regularised Optimal Transport. *arXiv:2307.09816*.

APPENDIX A: SIMULATIONS

In this section we provide numerical examples to illustrate our results, and circumstantial evidence that the obtained results may extend to quadratic regularization. We recall from Example 2.2 that quadratic regularization is a boundary case of Assumption 2.1.

For our experiments we need to solve the population problem very accurately. In EOT, one typically uses quadratic cost with Gaussian marginals in such situations, as that example allows for a closed-form solution (e.g., [8, 32]). The first subsection below details the only example with semi-closed form solution for more general divergences that we are aware of. The second subsection uses that example to numerically study the convergence rate for two divergences.

A.1. Example with semi-explicit population potentials. Let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ be the flat torus and let μ be the volume measure on \mathbb{T}^d . Denote by $[x]$ the equivalence class of a vector $x \in \mathbb{R}^d$. The distance on \mathbb{T}^d is given by

$$d([x], [y]) = \inf_{z \in \mathbb{Z}^d} \|x - y - z\|.$$

We consider the marginals $P = Q = \mu$ and the transport cost $c([x], [y]) = d^2([x], [y])$. In this symmetric “self-transport” problem, the population potentials (f_*, g_*) can be chosen such that $g_* = f_*$; this choice also removes the ambiguity about an additive constant. Indeed, f_* is uniquely determined by the equation

$$1 = \int \psi' \left(\frac{f_*([x]) + f_*([y]) - d^2([x], [y])}{\varepsilon} \right) d\mu([y]), \quad x \in \mathbb{R}^d.$$

Define the constant $C_{\varepsilon, d, \psi} \in \mathbb{R}$ via

$$1 = \int \psi' \left(\frac{2C_{\varepsilon, d, \psi} - d^2([0], [y])}{\varepsilon} \right) d\mu([y]).$$

For fixed $x \in \mathbb{R}^d$, consider the translation $y \mapsto T_x(y) := x + y$ on \mathbb{R}^d , which induces the map

$$[y] \mapsto S_x([y]) := [T_x(y)] = [x + y]$$

on the quotient space $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$. Note that S_x is an isometry and that the “translation invariance” of the volume measure implies $(S_x)_\# \mu = \mu$. Hence

$$\begin{aligned} 1 &= \int \psi' \left(\frac{2C_{\varepsilon, d, \psi} - d^2([0], [y])}{\varepsilon} \right) d\mu(y) = \int \psi' \left(\frac{2C_{\varepsilon, d, \psi} - d^2([x], S_x([y]))}{\varepsilon} \right) d\mu(y) \\ &= \int \psi' \left(\frac{2C_{\varepsilon, d, \psi} - d^2([x], [y])}{\varepsilon} \right) d(S_x)_\# \mu(y) = \int \psi' \left(\frac{2C_{\varepsilon, d, \psi} - d^2([x], [y])}{\varepsilon} \right) d\mu(y) \end{aligned}$$

for all $x \in \mathbb{R}^d$, showing that $f_* \equiv C_{\varepsilon, d, \psi}$.

A.2. Numerical illustrations. While the above example allows for explicit calculation, we emphasize that it actually leads to a degenerate regime, since the variance of the limits is zero. This is most easily seen in Theorem 3.5, where the variance is governed by the terms $f_*(X)$ and $\int \psi \left(\frac{f_*(X) + g_*(y) - c(X, y)}{\varepsilon} \right) d\mu(y)$ for $X \sim \mu$ (as well as the symmetric terms for Y). We see that both terms are constant due to $f_* = g_* \equiv C_{\varepsilon, d, \psi}$ and the shift invariance of the cost $c = d^2$ and μ . Hence, we expect to see rates which are even faster than $n^{-1/2}$, irrespective of dimension. While ideally we would like to also illustrate the results in non-degenerate examples, we are not aware of further examples where explicit solutions exist, and the latter are required to correctly visualize the asymptotic regime.

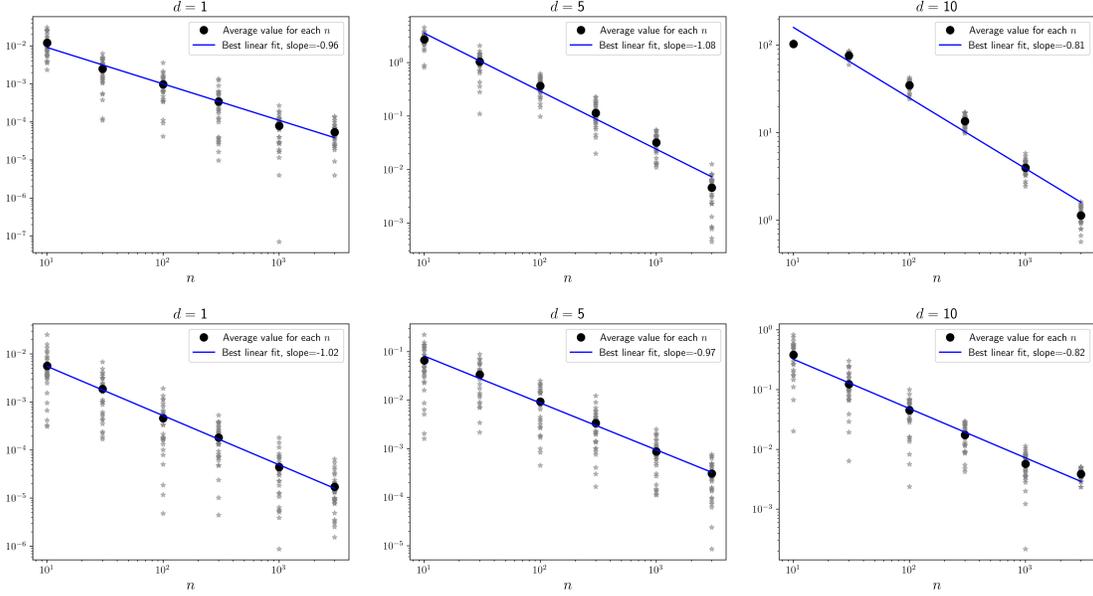


FIG 1. Log-log plots of $|\text{OT}_\varepsilon(P_n, Q_n) - \text{OT}_\varepsilon(P, Q)|$ for dimensions $d \in \{1, 5, 10\}$ (left, middle, right) and two different divergences, namely $\varphi(x) = \frac{2}{3}(x^{3/2} - 1)$ leading to $\psi(x) = \frac{1}{3}x_+^3 + \frac{2}{3}$ (top) and $\varphi(x) = \frac{1}{2}(x^2 - 1)$ leading to $\psi(x) = \frac{1}{2}x_+^2 + \frac{1}{2}$ (bottom). We use $\varepsilon = 0.5$. We numerically observe approximate rates n^α with $\alpha \approx -1$ for all cases, irrespective of dimension. The approximation is based on thirty independent evaluations of $|\text{OT}_\varepsilon(P_n, Q_n) - \text{OT}_\varepsilon(P, Q)|$ illustrated by the gray dots, while the black dots are the respective averages for each $n \in \{10, 30, 100, 300, 1000, 3000\}$.

To obtain approximate solutions for the empirical marginals P_n and Q_n , we use the Sinkhorn-like iterations described for instance in [30]. The results are illustrated in Figure A.2. As predicted by Theorem 3.5, we observe rates even faster than $n^{-1/2}$, and the slopes are similar for each dimension. (This is in contrast to upper bounds in [1], which would only yield $\max\{n^{-1/2}, n^{-2/d}\}$ in this case.) We observe similar slopes for $\varphi(x) = \frac{2}{3}(x^{3/2} - 1)$ and $\varphi(x) = \frac{1}{2}(x^2 - 1)$. Note that the former divergence satisfies Assumption 2.1 as the conjugate is cubic, whereas the latter (quadratic) divergence does not.

APPENDIX B: CENTRAL LIMIT THEOREM IN HÖLDER SPACE

The aim of this section is to provide a central limit theorem in the Hölder space $\mathcal{C}^{0,\beta}(\Omega)$ with $0 \leq \beta < 1$ for an i.i.d. sequence of random variables with values in $\mathcal{C}^{0,1}(\Omega)$ and finite second order moments. While it is possible that this result is known, we have not found a suitable reference in the literature. We first recall some preliminaries on type II spaces, following the excellent survey [27].

DEFINITION B.1. Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be separable Banach spaces such that $E \subset F$. The pair (E, F) is said to be of type II if there exists $C \geq 0$ such that

$$\mathbb{E} \left[\left\| \sum_{i=1}^n X_i \right\|_F^2 \right] \leq C \cdot \sum_{i=1}^n \mathbb{E} \left[\|X_i\|_E^2 \right]$$

for any $n \in \mathbb{N}$ and any sequence X_1, \dots, X_n of centered and independent E -valued random variables with $\mathbb{E}[\|X_i\|_E] < \infty$ for all $i = 1, \dots, n$.

The type II property is fundamental to derive central limit theorems in Banach spaces (see [27] and the references therein). To state the central limit theorem, recall that the expectation $\mathbb{E}[X]$ of a random variable X with $\mathbb{E}[\|X\|_E] < \infty$ in a Banach space E is defined as the following element of the bidual E'' :

$$E' \ni g \mapsto \mathbb{E}[g(X)] \in \mathbb{R},$$

where E' denotes the dual of E . A measure on E is called Gaussian if its finite-dimensional distributions are Gaussian; i.e., given $g_1, \dots, g_n \in E'$, the distribution on \mathbb{R}^n induced by (g_1, \dots, g_n) is Gaussian. A random variable $\mathbf{G} \in E$ is called Gaussian if its distribution is Gaussian.

THEOREM B.2 ([27, Theorem 3.1]). *Let (E, F) be of type II. Then every sequence $\{X_n\}_n \subset E$ of i.i.d. random vectors with $\mathbb{E}[\|X_i\|_E^2] < \infty$ and $\mathbb{E}[X_i] = 0$ satisfies the central limit theorem in E ; i.e., there exists a centered Gaussian random variable $\mathbf{G} \in E$ such that*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \xrightarrow{w} \mathbf{G} \quad \text{in } E.$$

To establish the desired central limit theorem, we need to show that the Banach pair $(\mathcal{C}^{0,1}(\Omega), \mathcal{C}^{0,\beta}(\Omega))$ is of type II. To that end, we first state an alternative characterization and a sufficient condition for being of type II. Recall that a Rademacher sequence $\{\epsilon_n\}_n$ is a sequence of i.i.d. (scalar) random variables with $\mathbb{P}(\epsilon_n = 1) = \mathbb{P}(\epsilon_n = -1) = \frac{1}{2}$.

PROPOSITION B.3 ([27, Theorems 3.1 and 2.3]). *Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be separable Banach spaces such that $E \subset F$. The following are equivalent:*

- (i) *The pair (E, F) is of type II,*
- (ii) *for any sequence $\{x_n\}_n \subset E$ such that $\sum_{n=1}^{\infty} \|x_n\|_E^2 < \infty$, and any Rademacher sequence $\{\epsilon_n\}_n$, the sum $\sum_{n=1}^{\infty} x_n \epsilon_n$ converges a.s. in F .*

Moreover, the following is a sufficient condition for (i) and (ii):

- (iii) *for any sequence $\{x_n\}_n \subset E$ such that $\sum_{n=1}^{\infty} \|x_n\|_E^2 < \infty$, and any i.i.d. sequence $\{\eta_n\}_n$ of standard Gaussian random variables, the sum $\sum_{n=1}^{\infty} x_n \eta_n$ converges a.s. in F .*

We can now state and prove the desired central limit theorem in $\mathcal{C}^{0,\beta}(\Omega)$. We recall that a random variable X in a Banach space $(E, \|\cdot\|_E)$ is tight if for every $\varepsilon > 0$ there exists a compact set \mathcal{K}_ε such that

$$\mathbb{P}(X \in \mathcal{K}_\varepsilon) \geq 1 - \varepsilon.$$

THEOREM B.4. *Let $\beta \in [0, 1)$ and let Ω be a compact subset of \mathbb{R}^d . Let $\{X_n\}_n$ be an i.i.d. sequence of random functions in $\mathcal{C}^{0,1}(\Omega)$ where X_1 is tight, $\mathbb{E}[\|X_1\|_{0,1}^2] < \infty$ and $\mathbb{E}[X_1] = 0$. Then $\{X_n\}_n$ satisfies the central limit theorem in $\mathcal{C}^{0,\beta}(\Omega)$; i.e., there exists a tight centered Gaussian random variable $\mathbf{G} \in \mathcal{C}^{0,\beta}(\Omega)$ such that*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \xrightarrow{w} \mathbf{G} \quad \text{in } \mathcal{C}^{0,\beta}(\Omega).$$

Note that, necessarily, $\mathbb{E}[\mathbf{G}(x)\mathbf{G}(x')] = \mathbb{E}[X_1(x)X_1(x')]$ for all $x, x' \in \Omega$.

PROOF. Proposition B.3 is not directly applicable as $\mathcal{C}^{0,1}(\Omega)$ and $\mathcal{C}^{0,\beta}(\Omega)$ are not separable.² However, we can restrict to separable subspaces as follows. Since X_1 is tight, there exists a sequence $\{\mathcal{K}_m\}_{m \in \mathbb{N}}$ of compact subsets of $\mathcal{C}^{0,1}(\Omega)$ such that X_1 (and then also X_n for any n) is a.s. valued in $\cup_{m \in \mathbb{N}} \mathcal{K}_m$. Note that $\cup_{m \in \mathbb{N}} \mathcal{K}_m$ is separable. Let E be the closed linear hull of $\cup_{m \in \mathbb{N}} \mathcal{K}_m$ in $\mathcal{C}^{0,1}(\Omega)$, and let F be the closed linear hull of $\cup_{m \in \mathbb{N}} \mathcal{K}_m$ in $\mathcal{C}^{0,\beta}(\Omega)$. Then E and F are separable Banach spaces, hence Proposition B.3 applies.

Let $\{f_n\}_n \subset E$ be a sequence of functions with $\sum_{n=1}^{\infty} \|f_n\|_{0,1}^2 < \infty$ and let $\{\eta_n\}_{n \in \mathbb{N}}$ be an i.i.d. sequence of (scalar) standard Gaussian random variables. In view of Theorem B.2 and Proposition B.3, our goal is to show that the partial sum $\sum_{n=1}^N f_n \eta_n$ converges a.s. in F (equivalently in $\mathcal{C}^{0,\beta}(\Omega)$) to some limit \mathbb{G} .

Note that for every $x \in \Omega$, the partial sum $\mathbb{G}_N(x) := \sum_{n=1}^N \eta_n f_n(x)$ converges a.s. by Kolmogorov's Three Series Theorem, hence we can define $\mathbb{G}(x)$ as the a.s. limit $\sum_{n=1}^{\infty} \eta_n f_n(x)$. Thus \mathbb{G} is a scalar, centered Gaussian process indexed by $x \in \Omega$. To show that the a.s. convergence $\mathbb{G}_N \rightarrow \mathbb{G}$ also holds in the norm of $\mathcal{C}^{0,\beta}(\Omega)$, it suffices to show weak convergence in $\mathcal{C}^{0,\beta}(\Omega)$, because by the Lévy–Itô–Nisio theorem [28, Theorem 2.4], these two convergences are equivalent for the partial sum $\mathbb{G}_N = \sum_{n=1}^N \eta_n f_n$ of the independent and symmetric random variables $\{\eta_n f_n\}_n$. In view of the pointwise convergence, weak convergence is implied by relative compactness (for the weak convergence topology), hence by Prokhorov's theorem [28, Theorem 2.1] it suffices to check that given $\delta \in (0, 1)$ there exists a compact $\mathcal{K} \subset \mathcal{C}^{0,\beta}(\Omega)$ such that

$$(39) \quad \mathbb{P}(\mathbb{G}_N \in \mathcal{K}) \geq 1 - \delta \quad \text{for all } N \geq 1.$$

Since $\mathcal{C}^{0,\beta'}(\Omega)$ is compactly embedded in $\mathcal{C}^{0,\beta}(\Omega)$ for $\beta' > \beta$ (e.g., [19, Lemma 6.33]), it is enough to show that

$$\sup_{N \geq 1} \mathbb{E}[\|\mathbb{G}_N\|_{0,\beta'}^\gamma] < \infty$$

for some $\beta' > \beta$ and $\gamma \geq 1$.

To that end, we shall use Kolmogorov's regularity theorem. Recall that two processes $X(x)$ and $\tilde{X}(x)$ indexed by $x \in \Omega$ are modifications of one another if $\mathbb{P}\{\tilde{X}(x) = X(x)\} = 1$ for all $x \in \Omega$, whereas they are called indistinguishable if

$$\mathbb{P}\{\tilde{X}(x) = X(x) \text{ for all } x \in \Omega\} = 1.$$

Kolmogorov's regularity theorem asserts that if $\{G(x)\}_{x \in \Omega}$ is a real-valued stochastic process satisfying

$$(40) \quad \mathbb{E}[|G(x) - G(y)|^\gamma] \leq C \|x - y\|^{d+\epsilon}$$

for some $\gamma \geq 1$ and $\epsilon > 0$, then G admits a modification \tilde{G} with

$$(41) \quad \mathbb{E} \left[\|\tilde{G}\|_{0,\frac{\epsilon}{\gamma}}^\gamma \right] \leq C_{d,\gamma,\epsilon,C}$$

where the constant $C_{d,\gamma,\epsilon,C}$ depends only on d , γ , ϵ and C . In particular, establishing (40) for $G = \mathbb{G}_N$ with constants independent of N implies (41) for a version $\tilde{\mathbb{G}}_N$ with a right-hand side $C_{d,\gamma,\epsilon,C}$ independent of N . Moreover, as we already know that $x \mapsto \mathbb{G}_N(x)$ is continuous, the modification $\tilde{\mathbb{G}}_N$ is necessarily indistinguishable from \mathbb{G}_N . In summary, establishing (40) for $G = \mathbb{G}_N$ with constants independent of N implies

$$\sup_{N \geq 1} \mathbb{E} \left[\|\mathbb{G}_N\|_{0,\frac{\epsilon}{\gamma}}^\gamma \right] < \infty,$$

²We thank Kengo Kato for pointing out this glitch in an earlier version of this manuscript.

which is the desired bound provided that $\beta' := \frac{\epsilon}{\gamma} \in (\beta, 1]$.

Fix $\alpha \geq 1$ large enough such that setting $\epsilon := \alpha$ and $\gamma := d + \alpha$, we have $\beta' := \frac{\epsilon}{\gamma} = \frac{\alpha}{d+\alpha} > \beta$. As explained above, it suffices to show

$$(42) \quad \mathbb{E}[|\mathbb{G}_N(x) - \mathbb{G}_N(y)|^\gamma] \leq C\|x - y\|^\gamma$$

for some $C > 0$ independent of N , as that will imply (39) and thus the claim.

By the Marcinkiewicz–Zygmund inequality (e.g., [5, Theorem 10.3.2]) there exists a constant $C > 0$ independent of N such that

$$\begin{aligned} \mathbb{E}[|\mathbb{G}_N(x) - \mathbb{G}_N(y)|^\gamma] &= \mathbb{E}\left[\left|\sum_{n=1}^N \eta_n(f_n(x) - f_n(y))\right|^\gamma\right] \\ &\leq C\mathbb{E}\left[\left|\sum_{n=1}^N (\eta_n(f_n(x) - f_n(y)))^2\right|^{\frac{\gamma}{2}}\right] \\ &\leq C\|x - y\|^\gamma \mathbb{E}\left[\left|\sum_{n=1}^N (\eta_n)^2 \|f_n\|_{0,1}^2\right|^{\frac{\gamma}{2}}\right]. \end{aligned}$$

We may assume without loss of generality that $s_N := \sum_{n=1}^N \|f_n\|_{0,1}^2 > 0$. Noting that $s_N \leq s_\infty < \infty$ by the choice of $\{f_n\}_n$, we deduce

$$\mathbb{E}[|\mathbb{G}_N(x) - \mathbb{G}_N(y)|^\gamma] \leq C(s_\infty)^{\frac{\gamma}{2}}\|x - y\|^\gamma \mathbb{E}\left[\left(\frac{1}{s_N} \sum_{n=1}^N (\eta_n)^2 \|f_n\|_{0,1}^2\right)^{\frac{\gamma}{2}}\right].$$

Now Jensen's inequality yields

$$\begin{aligned} \mathbb{E}[|\mathbb{G}_N(x) - \mathbb{G}_N(y)|^\gamma] &\leq C(s_\infty)^{\frac{\gamma}{2}}\|x - y\|^\gamma \mathbb{E}\left[\frac{1}{s_N} \sum_{n=1}^N (\eta_n)^\gamma \|f_n\|_{0,1}^2\right] \\ &= C(s_\infty)^{\frac{\gamma}{2}}\|x - y\|^\gamma \frac{1}{s_N} \sum_{n=1}^N \mathbb{E}[(\eta_n)^\gamma] \|f_n\|_{0,1}^2 \\ &= C\mathbb{E}[|\eta_1|^\gamma] (s_\infty)^{\frac{\gamma}{2}}\|x - y\|^\gamma \frac{1}{s_N} \sum_{n=1}^N \|f_n\|_{0,1}^2 \\ &= C\mathbb{E}[|\eta_1|^\gamma] (s_\infty)^{\frac{\gamma}{2}}\|x - y\|^\gamma, \end{aligned}$$

which is the desired estimate (42). \square

APPENDIX C: OMITTED PROOFS

C.1. Proofs for Section 2. In this subsection we prove Proposition 2.3 summarizing background results that were used throughout the text. We first state a lemma recalling the analogue of the c -conjugate that is standard in optimal transport. For the setting of regularization by f -divergence, this notion was introduced by [11]. (Note, however, that some of the results in [11] are flawed because the conjugate of φ was taken over \mathbb{R} instead of \mathbb{R}_+ , leading to signed measures instead of couplings.)

LEMMA C.1. *Let P, Q be probability measures on \mathbb{R}^d with supports Ω, Ω' . Let $c \in \mathcal{C}(\Omega \times \Omega')$ be bounded and have modulus of continuity ρ . Given any bounded measurable function $g : \Omega' \rightarrow \mathbb{R}$, there exists a unique function $f : \Omega \rightarrow \mathbb{R}$ such that*

$$(43) \quad \int \psi'(f(x) + g(y) - c(x, y)) dQ(y) = 1 \quad \text{for all } x \in \Omega.$$

Moreover, f is uniformly continuous with modulus ρ and its oscillation is bounded as

$$(44) \quad \sup_{x \in \Omega} f(x) - \inf_{x \in \Omega} f(x) \leq 2\|c\|_\infty,$$

while $\inf_{x, y} \{f(x) + g(y)\} \leq t_0 + \|c\|_\infty$ and $\sup_{x, y} \{f(x) + g(y)\} \geq t_0 - \|c\|_\infty$, where t_0 is defined in Assumption 2.1. Finally, f solves the concave optimization

$$\sup_{f \in L^\infty(P)} \int (f \oplus g - \psi(f \oplus g - c)) d(P \otimes Q).$$

PROOF OF LEMMA C.1. As g and c are bounded, $\lim_{s \rightarrow \infty} \psi'(s + g(y) - c(x, y)) = \infty$ and $\lim_{s \rightarrow 0} \psi'(s + g(y) - c(x, y)) < 1$ by the properties of ψ , where the limits are uniform in (x, y) . As ψ' is continuous, the intermediate value theorem yields the existence of f solving (43). Let $x, \tilde{x} \in \Omega$ and assume without loss of generality that $f(\tilde{x}) \leq f(x)$. As ψ' is nondecreasing, (43) yields

$$\begin{aligned} \int \psi'(f(x) + g(y) - c(x, y)) dQ(y) &= 1 = \int \psi'(f(\tilde{x}) + g(y) - c(\tilde{x}, y)) dQ(y) \\ &\leq \int \psi'(f(\tilde{x}) + g(y) - c(x, y) + \rho(\|x - \tilde{x}\|)) dQ(y). \end{aligned}$$

As ψ' is, in addition, strictly increasing on $[t_0 - \delta, \infty)$, this implies $f(x) \leq f(\tilde{x}) + \rho(\|x - \tilde{x}\|)$, showing that f has modulus of continuity ρ . The same argument, applied with $x = \tilde{x}$, also shows that $f(x)$ is uniquely determined by (43), and also that the oscillation of f is bounded by the one of c :

$$(45) \quad \sup_x f(x) - \inf_x f(x) \leq \sup_{x, y} c(x, y) - \inf_{x, y} c(x, y) \leq 2\|c\|_\infty.$$

Furthermore, $\psi'(t_0) = 1$ by Assumption 2.1. Hence (43) implies that

$$\inf_{x, y} \{f(x) + g(y) - c(x, y)\} \leq t_0 \leq \sup_{x, y} \{f(x) + g(y) - c(x, y)\}.$$

Thus $\inf_{x, y} \{f(x) + g(y)\} \leq t_0 + \|c\|_\infty$ and $\sup_{x, y} \{f(x) + g(y)\} \geq t_0 - \|c\|_\infty$. \square

PROOF OF PROPOSITION 2.3. (i) We first show $\text{ROT}(P, Q) \geq \text{DUAL}(P, Q)$, the so-called weak duality. The definition of ψ implies that

$$\psi(f \oplus g - c) \geq \frac{d\pi}{d(P \otimes Q)} \cdot (f \oplus g - c) - \varphi\left(\frac{d\pi}{d(P \otimes Q)}\right)$$

for any $(f, g) \in L^\infty(P) \times L^\infty(Q)$ and any $\pi \in \Pi(P, Q)$ with $\pi \ll P \otimes Q$. Combining this with $\int f \oplus g d\pi = \int f \oplus g d(P \otimes Q)$ yields the claimed weak duality. The converse inequality will be shown in (v) below.

(ii) Primal existence follows directly from the weak compactness of $\Pi(\mu, \nu)$ and the weak lower semi-continuity of the objective.

(iii) To show dual existence, let $(\tilde{f}^n, \tilde{g}^n)_n$ be a maximizing sequence for $\text{DUAL}(P, Q)$. Define f^n as the ‘‘conjugate’’ of \tilde{g}^n as provided by Lemma C.1, and let similarly g^n be the

conjugate of f^n (as provided by the symmetric analogue of Lemma C.1 with interchanged roles of the marginals). By Lemma C.1, we see that (f^n, g^n) improves upon (f^n, \tilde{g}^n) , so that it is again a maximizing sequence. Moreover, f_n and g_n are ρ -continuous. And as (44) holds for both f and g , the inequalities below (44) yield the uniform bound $\|f_n \oplus g_n\|_\infty \leq t_0 + 5\|c\|_\infty$. We can shift (f_n, g_n) by a constant such that $f_n(x_0) = 0$ at some reference point $x_0 \in \Omega$. It follows from Lemma C.1 that f_n and g_n are uniformly bounded. The Arzelà–Ascoli theorem then yields a subsequential uniform limit (f_*, g_*) . Using the uniform convergence, it is easy to see that (f_*, g_*) maximizes $\text{DUAL}(P, Q)$.

(iv) If (f_*, g_*) maximizes $\text{DUAL}(P, Q)$, taking the directional derivative in an arbitrary direction $(f, g) \in L^\infty(P) \times L^\infty(Q)$, as well as its negative $(-f, -g)$, yields the first-order condition (12). Conversely, let (f_*, g_*) solve (12). Note that for any fixed (x, y) , the function $\mathbb{R} \ni s \mapsto s - \psi(s - c(x, y))$ is concave with gradient $1 - \psi'(s - c(x, y))$. Given any functions $(f, g) \in L^\infty(P) \times L^\infty(Q)$, it follows that for $(x, y) \in \Omega \times \Omega'$,

$$(46) \quad \begin{aligned} & f(x) + g(y) - \psi(f(x) + g(y) - c(x, y)) \\ & \leq f_*(x) + g_*(y) - \psi(f_*(x) + g_*(y) - c(x, y)) \\ & \quad + \{1 - \psi'(f_*(x) + g_*(y) - c(x, y))(f(x) - f_*(x))\} \\ & \quad + \{1 - \psi'(f_*(x) + g_*(y) - c(x, y))(g(y) - g_*(y))\}. \end{aligned}$$

Integrating (46) with respect to $P \otimes Q$, the final two lines vanish due to (12), which shows that (f_*, g_*) is an optimizer of (11).

(v) We have $\psi' \geq 0$ as ψ is increasing, hence $d\pi := \psi'(f_* \oplus g_* - c)d(P \otimes Q)$ is a non-negative measure. The first equation of (12) shows that its first marginal is P and the second equation of (12) shows that its second marginal is Q . Hence $\pi \in \Pi(P, Q)$. Using the equality $\varphi(\psi'(z)) = \psi'(z)z - \psi(z)$ with $z := f_* \oplus g_* - c$, integrating, and rearranging, we find

$$\int (f_* \oplus g_* - \psi(f_* \oplus g_* - c))d(P \otimes Q) = \int c d\pi + \int \varphi\left(\frac{d\pi}{d(P \otimes Q)}\right) d(P \otimes Q).$$

In view of the weak duality $\text{ROT}(P, Q) \geq \text{DUAL}(P, Q)$ shown at the beginning of the proof, this shows that π is an optimizer of $\text{ROT}(P, Q)$ and also completes the proof of the strong duality $\text{ROT}(P, Q) = \text{DUAL}(P, Q)$ stated in (i).

(vi) Given any solution (f_*, g_*) of (12), applying Lemma C.1 twice yields versions that solve (13).

If f_*, g_* solve (13), then they are conjugates of one another, hence Lemma C.1 yields the modulus of continuity. As in the proof of (iii), the uniform bound follows from the inequalities below (44) using that (44) holds for both f and g .

(vii) Let C be such that $f_* \oplus g_* - c \leq C$. By Assumption 2.1 there are $t_0 > 0$ and $\delta, \alpha > 0$ such that $\psi'(t_0) = 1$, and $\psi'(t) < 1$ for $t \leq t_0 - \delta$, and $\psi''(t) \geq \alpha$ for $t \in [t_0 - \delta, C]$. Let $x \in \Omega$. Consider the sets

$$\begin{aligned} A &= \{y \in \Omega' : f_*(x) + g_*(y) - c(x, y) < t_0 - \delta\}, \\ B &= \{y \in \Omega' : t_0 - \delta \leq f_*(x) + g_*(y) - c(x, y) \leq C\} \end{aligned}$$

and note that $B = \Omega' \setminus A$. Set $p := Q(A)$. As ψ' is nondecreasing, (13) implies

$$1 = \int \psi'(f_*(\cdot) + g_*(y) - c(\cdot, y))dQ(y) \leq p\psi'(t_0 - \delta) + (1 - p)\psi'(C).$$

This yields the upper bound $p \leq \frac{\psi'(C) - 1}{\psi'(C) - \psi'(t_0 - \delta)} < 1$ which is uniform in x . As $\psi'' \geq 0$, we deduce the uniform lower bound

$$\int \psi''(f_*(\cdot) + g_*(y) - c(\cdot, y))dQ(y) \geq \alpha Q(B) = (1 - p)\alpha.$$

The second claim is shown analogously. \square

C.2. Proofs for Section 4. In this subsection we prove the auxiliary Lemma 4.2 on the sections of the set $S = \{(x, y) \in \Omega \times \Omega' : \psi''(\xi_*(x, y)) > 0\}$.

PROOF OF LEMMA 4.2. Recall from Proposition 2.3 (vii) that $S_Q(x) \neq \emptyset$ and $S_P(y) \neq \emptyset$ for all $(x, y) \in \Omega \times \Omega'$. This implies that $x \in O(x) := \cup_{y \in S_Q(x)} S_P(y)$ for all $x \in \Omega$. By continuity of $\zeta := \psi'' \circ \xi_*$, each set $S_P(y)$, and then also the union $O(x)$, is relatively open in Ω . That is,

$$(47) \quad \text{for each } x \in \Omega \text{ there exists } r > 0 \text{ such that } (x + r\mathbb{B}) \cap \Omega \subset O(x).$$

Consider the set

$$A = \{(x, z) \in \Omega \times \Omega : z \in \Omega \setminus O(x)\}.$$

If $z \in O(x)$, then there exists y such that $\zeta(x, y) > 0$ and $\zeta(z, y) > 0$. By continuity of ζ , for \tilde{x}, \tilde{z} sufficiently close to (x, z) , we still have $\zeta(\tilde{x}, y) > 0$ and $\zeta(\tilde{z}, y) > 0$, showing that $\tilde{z} \in O(\tilde{x})$. This proves that $\{(x, z) : z \in O(x)\}$ is open and hence that A is closed. Consider also, for each $r > 0$, the closed subset

$$A_r = A \cap \{(x, z) \in \Omega \times \Omega : z \in (x + r\overline{\mathbb{B}})\}.$$

In view of the definition of A , the fact (47) translates to $\cap_{r>0} A_r = \emptyset$. Now the finite intersection property of the compact set $\Omega \times \Omega$ yields that $A_r = \emptyset$ for some $r > 0$, which was the claim. \square