SIAM J. MATH. DATA SCI.
Vol. 5, No. 3, pp. 639–669

An Improved Central Limit Theorem and Fast Convergence Rates for Entropic Transportation Costs*

Eustasio del Barrio†, Alberto González Sanz‡, Jean-Michel Loubes‡, and Jonathan Niles-Weed§

Abstract. We prove a central limit theorem for the entropic transportation cost between subgaussian probability measures, centered at the population cost. This is the first result which allows for asymptotically valid inference for entropic optimal transport between measures which are not necessarily discrete. In the compactly supported case, we complement these results with new, faster, convergence rates for the expected entropic transportation cost between empirical measures. Our proof is based on strengthening convergence results for dual solutions to the entropic optimal transport problem.

Key words. optimal transport, entropic regularization, central limit theorem, Sinkhorn divergence

MSC codes. 62H10, 60F05, 49Q22

DOI. 10.1137/22M149260X

1. Introduction.

Optimal transport has emerged as a leading methodology in many areas of data science, machine learning, and statistics [1, 3, 7, 8, 10, 11, 13, 16, 17, 22, 29, 30, 34, 38, 39, 40, 41, 44, 48, 49, 55, 56, 62, 65, 66, 70, 71, 72, 75], with applications in fields ranging from high-energy physics [47, 67] to computational biology [69, 79]. Central to its recent success in practice is the paradigm of entropic regularization, popularized by [15], which leads to a highly efficient parallelizable algorithm suitable for large-scale data analysis [60]. This regularization is defined by augmenting the standard optimal transportation problem by a penalization term based on relative entropy, defined between two probability measures α and β as $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(x))d\alpha(x)$ if α is absolutely continuous with respect to β, $\alpha \ll \beta$, and $+\infty$ otherwise. Given $P, Q \in \mathcal{P}(\mathbb{R}^d)$ and $\epsilon > 0$, the resulting problem reads

$$S_\epsilon(P,Q) = \min_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|^2 d\pi(x,y) + \epsilon H(\pi|P \times Q), \tag{1.1}$$

where $\Pi(P,Q)$ denotes the set of couplings between P and Q.

*Received by the editors April 27, 2022; accepted for publication (in revised form) April 3, 2023; published electronically July 21, 2023.
https://doi.org/10.1137/22M149260X
Funding: The research of the first and second authors was partially supported by grant PID2021-128314NB-I00 funded by MCIN/AEI/ 10.13039/501100011033/FEDER, UE. The research of the second and third authors was partially supported by the AI Interdisciplinary Institute ANITI, which was funded by the French "Investing for the Future-PIA3" program under grant agreement ANR-19-PI3A-0004. The research of the fourth author was partially funded by National Science Foundation grant DMS-2015291.
†IMUVA, Universidad de Valladolid, Spain (tasio@eio.uva.es).
‡IMT, Université de Toulouse, France (alberto.gonzalez_sanz@math.univ-toulouse.fr, loubes@math.univ-toulouse.fr).
§Courant Institute for Mathematical Sciences and Center for Data Sciences, New York University, New York, NY 10012 USA (jnw@cims.nyu.edu).

Alongside its computational virtues, entropic regularization also brings substantial statistical benefits: entropically regularized transportation costs enjoy faster convergence rates than their unregularized counterparts, even in high dimensions, making them useful for estimation tasks [12, 31, 54, 61]. Moreover, entropic regularization seems well suited to problems involving data corrupted with Gaussian noise [64]. Together, this body of results suggests the strengths of entropic optimal transport as an applied and theoretical statistical tool.

Obtaining limit theorems for the unregularized transportation costs is a long-standing question in probability theory and statistics. (See the recent work [43] for references and an account of the history of this problem.) Under relatively stringent assumptions on the measures, it is known that the empirical unregularized transport cost possesses asymptotically Gaussian fluctuations around its expectation [20, 23]; stronger results can be obtained when one or both of the measures are discrete [21, 73, 74], when the measures are smooth [53], and in one dimension [18, 19].

The strict convexity and differentiability of the regularized optimal transportation problem make it possible to prove significantly more general results. A central limit theorem (CLT) for entropy regularized transportation costs, centered at the expectation of the empirical cost, was first obtained by [54] (see (1.2) below). Generalizations and extensions for discrete measures have been proved by [6, 46]. A growing body of work investigates the properties of the entropy regularized optimal transport problem from the perspective of probability and analysis, including its asymptotic properties as $\epsilon \to 0$ [2, 5, 9, 25, 26, 33, 42, 58, 59], opening the door to further statistical applications of entropy regularized transport.

A crucial question in statistical applications of entropic optimal transport costs is the construction of asymptotic confidence intervals, to permit asymptotically valid inference. The most general results known in this direction are due to [54], which showed that if P and Q are subgaussian probabilities on $\mathbb{R}^d$, then

$$\sqrt{n}(S_\epsilon(P_n,Q) - \mathbb{E}S_\epsilon(P_n,Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)), \tag{1.2}$$

with $S_\epsilon(\cdot,\cdot)$ as in (1.1). (See section 2 for further background and definitions.) A limitation of this result in practical inference problems is the centering at $\mathbb{E}S_\epsilon(P_n,Q)$ rather than at the population quantity $S_\epsilon(P,Q)$. This result parallels known results for the unregularized transport cost: [20, 23] show that, under suitable technical conditions on P and Q, there exists $\sigma \geq 0$ such that

$$\sqrt{n}(W_p^p(P_n,Q) - \mathbb{E}W_p^p(P_n,Q)) \xrightarrow{w} N(0,\sigma^2)\,, \tag{1.3}$$

where W_p denotes the unregularized p-transportation cost, $p > 1$. See also [36] for its generalization to the flat torus. In this case, it is known that the centering at $\mathbb{E}W_p^p(P_n,Q)$ is unavoidable, and that it is not possible in general to replace $\mathbb{E}W_p^p(P_n,Q)$ by $W_p^p(P,Q)$, in view of the fact that known lower bounds on convergence rates of the Wasserstein distance imply that $\sqrt{n}(W_p^p(P_n,Q) - W_p^p(P,Q))$ is typically not stochastically bounded when $d > 2p$.

However, this limitation does *not* apply to the entropically regularized transport costs. Indeed, the results of [31] imply that, for compactly supported measures,

$$|S_\epsilon(P,Q) - \mathbb{E}S_\epsilon(P_n,Q)| \leq C_{P,Q} n^{-1/2} \tag{1.4}$$

for a positive constant $C_{P,Q}$ depending exponentially on ϵ. A further refinement due to [54] showed that the same bound holds for subgaussian distributions, and that $C_{P,Q}$ can be taken to have only polynomial dependence on ϵ. As a consequence, prior work does not rule out the possibility that $\sqrt{n}(S_\epsilon(P_n,Q) - S_\epsilon(P,Q))$ enjoys a CLT, but neither does it provide a proof that such a theorem holds. In this paper, we close this gap. We show a CLT of the form

$$\sqrt{n}(S_\epsilon(P_n,Q) - S_\epsilon(P,Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)), \tag{1.5}$$

valid for any subgaussian probabilities P and Q in any dimension. Prior to our work, such a bound was known only when P and Q were supported on a finite or countable set [6, 46]. Our results represent a significant generalization of these results and imply that, under sufficiently strong moment conditions, asymptotically valid inference is always possible for the entropic transportation cost.

Our proof of (1.5) is based on an important strengthening of (1.4). Specifically, we show that, for subgaussian probability measures,

$$|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| = o(n^{-1/2}) \quad \text{as } n \to \infty. \tag{1.6}$$

Combining this result with (1.2) yields (1.5).

When P and Q are supported on a bounded set Ω, we are able to obtain substantially more precise results, which are of independent interest. Our techniques imply that for compactly supported P and Q,

$$|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| \le C_{P,Q} n^{-1} .$$

(See Remark 3.5.) This result implies that the bias of $S_\epsilon(P_n,Q)$ decays at the fast n^{-1} rate, thereby recovering the rate typically obtained for *parametric* estimation problems.[1] Our proof also yields new sample complexity results for the Sinkhorn divergence, defined as $D_\epsilon(P,Q) = S_\epsilon(P,Q) - \frac{1}{2}(S_\epsilon(P,P) + S_\epsilon(Q,Q))$. For probability measures on compact sets, convergence in Sinkhorn divergence is equivalent to weak convergence [27], implying that $D_\epsilon(P_n,P) \to 0$ a.s. In Theorem 5.1, we show the quantitative bound

$$\mathbb{E}D_\epsilon(P_n,P) \le C_P n^{-1},$$

valid for all compactly supported P. This convergence rate could have been anticipated from known distributional limits for Sinkhorn divergences between finitely supported measures [6, 46], but was unknown prior to our work.

In the bounded case, these results are all derived as corollaries of new convergence results for the *optimal dual potentials* in the entropic transport problem. In Theorem 4.5, we prove that, when P and Q are bounded, the entropic potentials converge fast in Hölder norm:

$$\mathbb{E}\|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} \le C_{P,Q} n^{-1} , \tag{1.7}$$

[1] A similar result was obtained in independent concurrent work [63].

where $s = [d/2]+1$. We prove this result, as well as its two-sample analogue, in section 4. To our knowledge, these bounds for the derivatives are new, even for finitely supported probability measures. A similar bound showing convergence of the potentials in L^∞ has been proved using completely different techniques in [51], though the bounds obtained do not seem to be strong enough to obtain Theorem 5.1.

When P and Q are not necessarily bounded but have subgaussian tails, we prove a nonquantitative analogue of (1.7), showing that f_n and g_n converge to f^* and g^* a.s. in a suitably strong topology. This result is a strengthening of a similar convergence result obtained by [54].

The results we develop here have multiple applications in statistics; here we underline some of them. Since the first version of this work appeared, the bounds on the potentials have been used to define and analyze a new Gaussian process on the space of distributions with a.s. continuous sample paths [4], with applications to uncertainty quantification and Bayesian modeling based on regularized optimal transport. These potential bounds have also been used to obtain weak limits for estimators of the Sinkhorn divergence [37]. Theorem 5.1 also implies improvements to tests already existing in the literature. Prior work has suggested the use of the Sinkhorn divergence for independence testing [50], but that paper's theoretical results use a bound similar to that of [31] which implies that the critical value for valid independence testing with the Sinkhorn divergence should be chosen to be of order $n^{-1/2}$. The results of this paper show that this bound is unnecessarily conservative at the null and can be improved to n^{-1}, which, of course, increases the power of the proposed tests.

The remaining sections of this paper are organized as follows. Section 2 provides some background results on entropic transportation costs. The CLT (1.5) and the faster rate (1.6) are given in section 3. Section 4 contains the announced results about the convergence rates of the potentials. The bounds for Sinkhorn divergences are proved in section 5. Finally we include a section with some numerical illustration of our limit theorems.

2. Preliminaries on entropic transportation costs.

This selection collects several background results on the entropic transportation problem (1.1).

We say that a distribution ν is the pushforward by a map T of a distribution μ if $\nu = \mu \circ T^{-1}$. A simple computation shows that if P^ε and Q^ε denote the pushforwards of P and Q under the map $x \mapsto \varepsilon^{-\frac{1}{2}} x$, then $S_\epsilon(P,Q) = \varepsilon S_1(P^\varepsilon, Q^\varepsilon)$. Hence, we focus on the case $\varepsilon = 1$ and write simply $S(P,Q)$ instead of $S_1(P,Q)$. The minimization problem (1.1) admits a dual formulation. In fact, if $\pi \in \Pi(P,Q)$ and $r = \frac{d\pi}{d(P\times Q)}$, then, for any $f \in L_1(P)$, $g \in L_1(Q)$,

$$\int \left[\frac{1}{2}\|x-y\|^2 + \log r(x,y)\right] r(x,y) dP(x) dQ(y) \geq \int f(x) dP(x) + \int g(y) dQ(y) \\ - \int e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x) dQ(y) + 1,$$

with equality if and only if $r(x,y) = e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2}$ $P \times Q$-a.s. (This follows from the elementary fact that $s \log s \geq s-1$, $s>0$, with equality if and only if $s=1$.) This inequality implies the following version of weak duality:

$$S(P,Q) \geq \sup_{f\in L_1(P), g\in L_1(Q)} \left\{ \int_{\mathbb{R}^d} f(x)dP(x) + \int_{\mathbb{R}^d} g(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) + 1 \right\}.$$

It shows also that if $\frac{d\pi}{d(P\times Q)} = e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2}$ for some $f \in L_1(P_1)$ and $g \in L_1(Q)$, then π is a minimizer for the entropic transportation problem (indeed, by the strict convexity of H, it is the unique minimizer). The theory of entropic optimal transportation (see [14, 57]) shows that the last inequality is, in fact, an equality, namely,

$$S(P,Q) = \sup_{f\in L_1(P), g\in L_1(Q)} \left\{ \int_{\mathbb{R}^d} f(x)dP(x) + \int_{\mathbb{R}^d} g(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) + 1 \right\}. \tag{2.1}$$

Maximizing pairs in (2.1) are called optimal potentials. These optimal potentials exist and satisfy some regularity conditions under integrability assumptions on P and Q.

Following the framework in [54], we say that a probability P is σ^2-subgaussian if $\mathbb{E}(e^{\frac{\|X\|^2}{2d\sigma^2}}) \leq 2$ when $X \sim P$. When P and Q are subgaussian there exist optimal potentials, denoted by f^*, g^*, satisfying the *optimality conditions,* i.e.,

$$\begin{aligned} &\int e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y) = 1 \qquad \forall x \in \mathbb{R}^d, \\ &\int e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dP(x) = 1 \qquad \forall y \in \mathbb{R}^d; \end{aligned} \tag{2.2}$$

see Proposition 6 in [54]. Moreover, the pair (f^*, g^*) satisfying (2.2) is unique up to constant shifts, and is uniquely specified by adopting the normalization convention

$$\int f^*(x)dP(x) = \int g^*(y)dQ(y). \tag{2.3}$$

In what follows, we tacitly assume that (2.3) holds unless we explicitly specify an alternate convention.

The above considerations imply that the minimizer in the primal formulation is

$$d\pi^* = e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y)dP(x),$$

where f^* and g^* satisfy

$$\begin{aligned} f^*(x) &= -\log\left(\int e^{g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y)\right), \\ g^*(y) &= -\log\left(\int e^{f^*(y)-\frac{1}{2}\|x-y\|^2} dP(y)\right). \end{aligned} \tag{2.4}$$

Let $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ be a multi-index. If $P, Q \in \mathcal{P}(\mathbb{R}^d)$ are σ^2-subgaussian, then (see Proposition 1 in [54]), the optimal potential f^* specified above is such that

$$\left| D^\alpha \left(f^* - \frac{1}{2}\|\cdot\|^2 \right)(x) \right| \leq C_{k,d} \begin{cases} 1+\sigma^4 & \text{if } k=0 \\ \sigma^k(\sigma+\sigma^2)^k & \text{otherwise} \end{cases} \qquad \text{if } \|x\| \leq \sqrt{d}\sigma, \tag{2.5}$$

$$\left| D^\alpha \left(f^* - \frac{1}{2}\|\cdot\|^2 \right)(x) \right| \leq C_{k,d} \begin{cases} 1 + (1+\sigma^2)\|x\|^2 & \text{if } k=0 \\ \sigma^k(\sqrt{\sigma\|x\|} + \sigma\|x\|)^k & \text{otherwise} \end{cases} \quad \text{if } \|x\| \geq \sqrt{d}\sigma, \tag{2.6}$$

and likewise for g^*, where in both cases $k := |\alpha|$, and the constant $C_{k,d}$ depends only on d and k.

Throughout the paper, we will assume that $P, Q \in \mathcal{P}(\mathbb{R}^d)$ are σ^2-subgaussian probabilities and $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are independent samples of independent and identically distributed (i.i.d.) random variables with laws P and Q, respectively. We will denote by P_n and Q_m the associated empirical measures. We will require that the measures P_n and Q_n are also subgaussian, which is guaranteed by the following result, which summarizes Lemmas 2 and 4 in [54].

Lemma 2.1. *Let $X_1, \dots, X_n$ be i.i.d. random variables with σ^2-subgaussian law $P \in \mathcal{P}(\mathbb{R}^d)$, and let P_n be the associated empirical measure. Then, there exists a random variable $\tilde{\sigma}$, such that*

1. *for every $n \in \mathbb{N}$, the probabilities P and P_n are uniformly $\tilde{\sigma}^2$-subgaussian a.s.,*
2. *for any $k \in \mathbb{N}$, we have $\mathbb{E}(\tilde{\sigma}^{2k}) \leq 2k^k\sigma^{2k}$.*

3. An improved central limit theorem for subgaussian probability measures. This section shows that, for subgaussian probability measures, the expected empirical entropic transportation cost converges to its population counterpart with rate $o(n^{-1/2})$. This is an improvement over the bound (1.4) derived in [54] and has, as a main consequence, a CLT for the empirical entropic transportation cost with the natural centering constants (see Theorem 3.6), which, in turn, yields an asymptotically valid confidence interval for $S_\varepsilon(P, Q)$ regardless of the dimension, d.

Let s be a nonnegative integer. To prove the main result in this section, we introduce the class $\mathcal{G}^s(C)$, consisting of all $f \in \mathcal{C}^s(\mathbb{R}^d)$ such that

$$\begin{aligned} &|f(x)| \leq C(1 + \|x\|^3), \\ &|D^\alpha f(x)| \leq C(1 + \|x\|^{s+1}), \quad |\alpha| \leq s. \end{aligned} \tag{3.1}$$

Our next results gives an estimate of the complexity of this class, in terms of covering numbers[2] with respect to the random metric $L^2(dP_n)$. The proof can be easily adapted from the proof of Proposition 3 in [54]. We omit further details.

Lemma 3.1. *Assume $\mathcal{G}^s(C)$ is as above. If $X_1, \dots, X_n$ are i.i.d. random variables with σ^2-subgaussian law $P \in \mathcal{P}(\mathbb{R}^d)$, P_n is the associated empirical measure, and $L = \frac{1}{n}\sum_{i=1}^n e^{-\frac{\|X_i\|^2}{4d\sigma^2}}$, then, for a constant $C_{s,d}$ depending only on s and d,*

$$\log \mathcal{N}(\epsilon, \mathcal{G}^s(C), L^2(dP_n)) \leq C_{s,d} L^{\frac{d}{2s}} \epsilon^{\frac{-d}{s}} (1+\sigma^d)(1+\sigma^s)^{\frac{d}{s}}. \tag{3.2}$$

[2] Here $\mathcal{N}(\epsilon, \mathcal{G}^s(C), L^2(dP_n))$ denotes the covering numbers of the class $\mathcal{G}^s(C)$ with respect to the metric $L^2(P_n)$, i.e., the minimal number of balls of radius ϵ needed to cover that class (see [77, Definition 2.1.5] for the definition).

Finally, we introduce the space $\mathcal{G}^s = \bigcup_{C\geq 0}\mathcal{G}^s(C)$ endowed with the norm

$$\|f\|_s = \left\|\frac{f}{1+\|\cdot\|^3}\right\|_\infty + \sum_{i=1}^s \sum_{|k|=i} \left\|\frac{D^k f}{1+\|\cdot\|^{s+1}}\right\|_\infty .$$

Let $(\mathcal{G}^s)'$ denote the dual space of $\mathcal{G}^s$, endowed with the dual norm

$$\|G\|'_s = \sup_{f\in\mathcal{G}^s,\ \|f\|_s\leq 1} |G(f)|.$$

With these ingredients we are ready to prove the main technical result of this section, from which we obtain the CLT for the entropic transportation cost with natural centering constants (Theorem 3.6 below).

Lemma 3.2. *If $P,Q\in\mathcal{P}(\mathbb{R}^d)$ are σ^2-subgaussian probabilities, then*

$$\sqrt{n}\,|\mathbb{E}S(P_n,Q) - S(P,Q)| \to 0. \tag{3.3}$$

Moreover, if $m=m(n)$ and $\lambda := \lim_{n\to\infty}\frac{n}{n+m}\in(0,1)$, then

$$\sqrt{\frac{nm}{n+m}}\,|\mathbb{E}S(P_n,Q_m) - S(P,Q)| \to 0. \tag{3.4}$$

Proof. Let $(f_n,g_n)\in L_1(P_n)\times L_1(Q)$ be the unique pair of optimal potentials satisfying (2.2) and (2.3) for P_n,Q. As noted above, by Proposition 6 in [54], this pair satisfies (2.5) and (2.6). We observe that, by optimality of the potentials,

$$S(P,Q) \geq \int_{\mathbb{R}^d} f_n(x)dP(x) + \int_{\mathbb{R}^d} g_n(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) + 1,$$

which yields

$$\begin{aligned} 0 \leq \sqrt{n}\,(\mathbb{E}S(P_n,Q) - S(P,Q)) \leq & \mathbb{E}\int_{\mathbb{R}^d} f_n(x)\sqrt{n}(dP_n - dP)(x) \\ & - \mathbb{E}\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2}\sqrt{n}(dP_n-dP)(x)dQ(y). \end{aligned}$$

Now the optimality condition

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2} dQ(y) = 1 \quad \forall x\in\mathbb{R}^d$$

implies that

$$0 \leq \sqrt{n}\,(\mathbb{E}S_1(P_n,Q) - S_1(P,Q)) \leq \mathbb{E}\int_{\mathbb{R}^d} f_n(x)\sqrt{n}(dP_n - dP)(x).$$

Set $s=[d/2]+1$, and let $(f^*,g^*)\in L_1(P)\times L_1(Q)$ be the unique pair of optimal potentials satisfying (2.2) and (2.3) for P,Q. Since $\mathbb{E}\int_{\mathbb{R}^d} f^*(x)\sqrt{n}(dP_n-dP)(x)=0$,

$$0 \leq \sqrt{n}\,(\mathbb{E}S_1(P_n,Q) - S_1(P,Q)) \leq \mathbb{E}\int_{\mathbb{R}^d}\{f_n(x)-f^*(x)\}\sqrt{n}(dP_n-dP)(x).$$

We write now $\mathbb{G}_n$ for the empirical process indexed by $\mathcal{G}^s$, that is, $\mathbb{G}_n(f) = \sqrt{n}(P_n(f) - P(f))$, $f \in \mathcal{G}^s$, and note that

$$|\mathbb{G}_n(f)| \le \sqrt{n}(P_n + P)(|f|) \le \left\| \frac{f}{1 + \|\cdot\|^3} \right\|_\infty \sqrt{n}(P_n + P)(1 + \|\cdot\|^3)$$
$$\le \sqrt{n}(P_n + P)(1 + \|\cdot\|^3)\|f\|_s \le \sqrt{n}(2 + 48(d\tilde{\sigma}^2)^3)\|f\|_s,$$

where the last inequality comes from Lemma 1 in [54]. Consequently, we deduce that $\mathbb{G}_n$ belongs to the dual space $(\mathcal{G}^s)'$ for all $n \in \mathbb{N}$, and we get the bound

$$\sqrt{n}\left(\mathbb{E}S_1(P_n, Q) - S_1(P, Q)\right) \le \mathbb{E}\left\{\|\mathbb{G}_n\|_s' \|f^* - f_n\|_s\right\}.$$

Using the Cauchy–Schwarz inequality we see that

$$\sqrt{n}\left(\mathbb{E}S_1(P_n, Q) - S_1(P, Q)\right) \le \sqrt{\mathbb{E}\|\mathbb{G}_n\|_s'^2 \mathbb{E}\|f^* - f_n\|_s^2}. \tag{3.5}$$

Note that $\|\mathbb{G}_n\|_s'$ is the sup taken on the unit ball of $\mathcal{G}^s$, which is contained in $\mathcal{G}^s(1)$. We can conclude, by using (3.2) and Theorem 3.5.1 and Exercise 2.3.1 in [35], that there exists a constant $C_{s,d} > 0$ such that

$$\mathbb{E}\|\mathbb{G}_n\|_s'^2 \le C_{s,d}\mathbb{E}\left(\int_0^{\max_{\|f\|_s \le 1}\|f\|_{L^2(dP_n)}} \sqrt{L^{\frac{d}{2s}}\epsilon^{\frac{-d}{s}}(1 + \sigma^d)(1 + \sigma^s)^{\frac{d}{s}}}\, d\epsilon\right)^2$$
$$\le (1 + \sigma^d)(1 + \sigma^s)^{\frac{d}{s}} C_{s,d}\mathbb{E}\left(\int_0^{1 + 4\sqrt{3}d^{3/2}\tilde{\sigma}^3} L^{\frac{d}{4s}}\epsilon^{\frac{-d}{2s}}\, d\epsilon\right)^2$$
$$\le C'_{s,d}(1 + \sigma^{2d})\mathbb{E}L^{\frac{d}{2s}}(1 + \tilde{\sigma}^3)^{\frac{2s-d}{s}},$$

where we have used first Lemma 1 in [54] to bound

$$\max_{\|f\|_s \le 1}\|f\|_{L^2(dP_n)} \le 1 + \left(\int \|x\|^6 dP_n(x)\right)^{1/2} \le 1 + 4\sqrt{3}d^{3/2}\tilde{\sigma}^3$$

and then the fact that $s = [d/2] + 1$. Using the Cauchy–Schwarz inequality we see that

$$\mathbb{E}L^{\frac{d}{2s}}(1 + \tilde{\sigma}^3)^{\frac{2s-d}{s}} \le \sqrt{\mathbb{E}L^{\frac{d}{s}}\mathbb{E}(1 + \tilde{\sigma}^3)^{\frac{2(2s-d)}{s}}},$$

where we can use the fact that $\mathbb{E}L < C$ for a positive constant C independent of n and Lemma 2.1 to conclude that $\limsup \mathbb{E}\|\mathbb{G}_n\|_s'^2 < \infty$.

To deal with the second term in (3.5) we denote $\Delta_n = f^* - f_n$. We prove next that $\|\Delta_n\|_s \to 0$ a.s., and then that it is dominated by a random variable with finite second moment. Together, these facts imply that $\mathbb{E}\|f^* - f_n\|_s^2 \to 0$ and conclude the proof. The first claim is given by the following result.

Lemma 3.3. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be σ-subgaussian probabilities, and P_n, Q_n associated empirical measures. Then, the optimal transport potentials (f_n, g_n) for P_n, Q_n satisfy $\|f_n - f^*\|_s \to 0$ and $\|g_n - g^*\|_s \to 0$ a.s.*

Proof. We prove the result for f_n, with the same conclusion following for g_n by symmetry. First, we use induction to prove convergence of the derivatives up to order s. We follow classical arguments in real analysis (see [68]):

(1) *For $J=0$,* Proposition 4 in [54] shows that, a.s., $\Delta_n := f^* - f_n \to 0$ uniformly in compact sets.

(2) *Assume that for every k with $|k| \le J-1$, we have $D^k\Delta_n \to 0$, uniformly in compact sets.* Let $k=(k_1,\dots,k_d)$ be such that $|k|=J$, and let $B_R \subset \mathbb{R}^d$ be the ball of radius R centered at 0. Using the fact that all the derivatives of $D^k\Delta_n$ are bounded and $D^k\Delta_n$ is itself pointwise bounded (see Proposition 1 and Lemma 2 in [54]), we derive that the sequence $D^k\Delta_n$ is equicontinuous and bounded for all points. We can then apply the Arzelà–Ascoli theorem on B_R to deduce that, up to subsequences, $D^k\Delta_n \to \Delta^k$ uniformly on B_R. Suppose, without loss of generality, that $k_1 \ge 1$, set $k'=(k_1-1,\dots,k_d)$, and note that

$$D^{k'}\Delta_n(x) = \int_0^{x_1} D^k\Delta_n(t,,x_2,\dots,x_d)dt + D^{k'}\Delta_n(0,x_2,\dots,x_d),$$

which implies that

$$\begin{aligned}&\left|D^{k'}\Delta_n(x) - \int_0^{x_1}\Delta^k(t,,x_2,\dots,x_d)dt\right| \\ &\quad \le \int_0^{x_1} |D^k\Delta_n(t,,x_2,\dots,x_d) - \Delta^k(t,,x_2,\dots,x_d)|dt + |D^{k'}\Delta_n(0,x_2,\dots,x_d)|.\end{aligned}$$

As a consequence,

$$\begin{aligned}&\sup_{x\in B_R}\left|D^{k'}\Delta_n(x) - \int_0^{x_1}\Delta^k(t,,x_2,\dots,x_d)dt\right| \\ &\quad \le R\sup_{x\in B_R}|D^k\Delta_n(x) - \Delta^k(x)| + |D^{k'}\Delta_n(0,x_2,\dots,x_d)| \to 0,\end{aligned}$$

where the limit follows from the induction hypothesis (recall that $\sup_{x\in B_R}|D^{k'}\Delta_n(x)| \to 0$). By uniqueness of the limit we conclude that $0=\int_0^x \Delta^k dx_1$, which implies that $\Delta^k=0$. By taking $R\to\infty$ we conclude that $D^k\Delta_n \to 0$ uniformly on the compact sets of $\mathbb{R}^d$.

To show convergence in the norm $\|\cdot\|_s$, it suffices to show that for any $\epsilon>0$, there exists an n_0 such that $\|\Delta_n\|_s \le \epsilon$ for all $n\ge n_0$. Recall that by Lemma 2.1(1) and Proposition 1 in [54], there exists an a.s. finite random variable $\tilde\sigma$ and a constant $K_{s,d}$ such that for all $n\in\mathbb{N}$ and $x\in\mathbb{R}^d$,

$$\begin{aligned}\frac{|\Delta_n(x)|}{1+\|x\|^3} &\le K_{s,d}\frac{1+\tilde\sigma^4}{1+\|x\|}, \\ \frac{|D^k\Delta_n(x)|}{1+\|x\|^{s+1}} &\le K_{s,d}\frac{1+\tilde\sigma^{3s}}{1+\|x\|} \qquad \forall |k|\le s.\end{aligned} \tag{3.6}$$

We obtain that there exists a finite random variable $\tilde K$ such that

$$\frac{|\Delta_n(x)|}{1+\|x\|^3} + \sum_{i=1}^s\sum_{|k|=i}\frac{|D^k\Delta_n(x)|}{1+\|x\|^{s+1}} \le \epsilon/2 \qquad \forall \|x\| > \tilde K\epsilon^{-1}, n\ge 0\,.$$

Since Δ_n and $D^k\Delta_n$ converge uniformly to zero on the compact set $\{x\in\mathbb{R}^d : \|x\|\le c_{d,s}\tilde{K}\epsilon^{-1}\}$, there exists an n_0 for which

$$\frac{|\Delta_n(x)|}{1+\|x\|^3}+\sum_{i=1}^{s}\sum_{|k|=i}\frac{|D^k\Delta_n(x)|}{1+\|x\|^{s+1}}\le \epsilon/2 \qquad \forall \|x\|\le \tilde{K}\epsilon^{-1}, n\ge n_0\,.$$

Combining these claims, we obtain that $\|\Delta_n\|_s\le\epsilon$ for all $n\ge n_0$, as desired. ■

To complete the proof of Lemma 3.2 it only remains to prove that $\|f_n-f^*\|_s$ can be dominated by a random variable with finite second moment. We have from (3.6) above that $\|\Delta_n\|_s^2\le K'_{s,d}(1+\tilde{\sigma}^{3s})^2$ for some constant $K'_{s,d}$. It only remains to show that $\mathbb{E}\left(1+\tilde{\sigma}^{3s}\right)^2<\infty$. But Lemma 2.1 implies that all moments of $\tilde{\sigma}$ are finite, which completes the proof.

To deal with the two-sample case, we split the difference as follows:

$$\begin{aligned}&\sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P_n,Q_m)-S_1(P,Q)|\\ &\quad\le\sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P_n,Q_m)-S_1(P,Q_m)|+\sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P,Q_m)-S_1(P,Q)|.\end{aligned}$$

The second term tends to 0 by using (3.3). For the first one we denote $g_{n,m}$ a potential of $S_1(P_n,Q_m)$ and g_m a potential of $S_1(P,Q_m)$. Applying (3.5) we derive

$$\begin{aligned}\sqrt{m}|\left(\mathbb{E}S(P_n,Q_m)-S(P,Q_m)\right)|&\le\sqrt{\mathbb{E}\|F_m\|_s'^2\,\mathbb{E}\|g_{n,m}-g_m\|_s^2}\\ &\le 2\sqrt{\mathbb{E}\|F_m\|_s'^2\left(\mathbb{E}\|g_{n,m}-g^*\|_s^2+\mathbb{E}\|g_m-g^*\|^2\right)}.\end{aligned}$$

We conclude using Lemma 3.3 (which can be trivially adapted to this setup) and the subsequent argument. ■

As a consequence of Lemma 3.2, by simply considering the change of variables $x\mapsto x\epsilon^{-\frac{1}{2}}$ (recall the comments at the beginning of this section) we obtain the generalization to any $\epsilon>0$.

Corollary 3.4. *Let $P,Q\in\mathcal{P}(\mathbb{R}^d)$ be σ-subgaussian probabilities and P_n, Q_m associated empirical measures. Then*

$$\sqrt{n}\,|\mathbb{E}S_\epsilon(P_n,Q)-S_\epsilon(P,Q)|\to 0$$

and

$$\sqrt{\frac{nm}{n+m}}\,|\mathbb{E}S_\epsilon(P_n,Q_m)-S_\epsilon(P,Q)|\to 0.$$

As announced, Corollary 3.4 improves over Corollary 1 in [54], which implied $|\mathbb{E}S_\epsilon(P_n,Q)-S_\epsilon(P,Q)|=O(n^{-1/2})$ rather than $|\mathbb{E}S_\epsilon(P_n,Q)-S_\epsilon(P,Q)|=o(n^{-1/2})$.

Remark 3.5. In some cases we can go much further in this direction. In fact, if P and Q are compactly supported, then (see Theorem 4.5 below)

$$\mathbb{E}\|f_n-f^*\|^2_{\mathcal{C}^s(\Omega)}\le\frac{c_\Omega}{n}$$

for some constant c_Ω. Plugging this into (3.5) and using again the fact that $\limsup_n \mathbb{E}\|\mathbb{G}_n\|_s'^2 < \infty$, we conclude that

$$|\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P, Q)| \le \frac{C_\Omega}{n} \tag{3.7}$$

for some constant $C_\Omega > 0$, which depends exponentially on the diameter of Ω. A similar conclusion holds for the two-sample problem. Whether this improved rate remains valid for general subgaussian probabilities is an open question. The exponential dependence of C_Ω on the diameter of the support makes (3.7) tighter than [54]—where the rate is $\frac{K_\Omega}{\sqrt{n}}$, with K_Ω polynomially dependent on the diameter of Ω—for big sample sizes. However, since the constant goes from polynomial to exponential in the diameter, the bound in [54] may be more appropriate for small data sets.

The following central limit becomes a direct consequence of Theorem 3 in [54], which shows that the fluctuations around the mean are asymptotically Gaussian, i.e.,

$$\sqrt{n}(S_\epsilon(P_n, Q) - \mathbb{E}S_\epsilon(P_n, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)). \tag{3.8}$$

Here $(f_\epsilon^*, g_\epsilon^*)$ are optimal potentials for $S_\epsilon(P, Q)$. We observe that, while the pair of optimal potentials is not uniquely defined, it follows from the uniqueness of the minimizer in (1.1) that if $(\tilde{f}_\epsilon, \tilde{g}_\epsilon)$ is another pair of optimal potentials, then $\tilde{f}_\epsilon = f_\epsilon^* + K$ P-a.s. for some constant K. Hence, $\mathrm{Var}_P(f_\epsilon^*)$ is well defined in the sense that it does not depend on the choice of optimal potential.

Theorem 3.6. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be σ-subgaussian probabilities; then*

$$\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)),$$

where $(f_\epsilon^*, g_\epsilon^*)$ are optimal potentials for $S_\epsilon(P, Q)$. Moreover, if $\lambda := \lim_{n,m\to\infty} \frac{n}{n+m} \in (0,1)$,

$$\sqrt{\frac{nm}{n+m}}(S_\epsilon(P_n, Q_m) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, (1-\lambda)\,\mathrm{Var}_P(f_\epsilon^*) + \lambda\,\mathrm{Var}_Q(g_\epsilon^*)).$$

The limits appearing in Theorem 3.6 are nondegenerate as long as the potentials have positive variance; that is, degeneracy can only occur when f_ϵ^* is P-a.s. constant in the one-sample case, or when f_ϵ^* and g_ϵ^* are P-a.s. and Q-a.s. constant, respectively, in the two-sample case. It is easy to see that this situation can occur, for instance, when both P and Q are concentrated on a single point. In the case of the unregularized optimal transport problem, it is known that degenerate limits can arise in other situations as well [43, section 4], including when the distributions are absolutely continuous. However, the following result shows that, in the regularized case, a degenerate limit cannot arise when at least one of the two distributions is mutually absolutely continuous with the Lebesgue measure. We believe that the hypothesis can be relaxed, but leave a further investigation of the degeneracy of the limit to future work.

Proposition 3.7. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$, and assume that one of the two measures is equivalent to the Lebesgue measure. Then f_ϵ^* and g_ϵ^* cannot both be constant. In particular, the two-sample limit in Theorem 3.6 is nondegenerate.*

Proof. We prove the claim for $\epsilon = 1$ and proceed by contradiction. Assume without loss of generality that Q is mutually absolutely continuous with the Lebesgue measure. Suppose that $f^*(x) + g^*(y) = c$ for $P \times Q$-a.e. (x,y). Then by the optimality condition (2.2),

$$\int e^{c - \frac{1}{2}\|x-y\|^2} dP(x) = 1 \quad Q\text{-a.e.}$$

Since Q is equivalent to the Lebesgue measure, this equality holds Lebesgue-a.e. In particular, integrating both sides with respect to the Lebesgue measure, we obtain

$$\int\int e^{c - \frac{1}{2}\|x-y\|^2} dP(x) dy = +\infty \, .$$

On the other hand, since the integrand is nonnegative, we can switch the order of integration and explicitly evaluate the inner integral, yielding

$$\int\int e^{c - \frac{1}{2}\|x-y\|^2} dy dP(x) = \int e^{c} (2\pi)^{d/2} dP(x) < +\infty \, ,$$

since P is a probability measure. This results in a contradiction, so f^* and g^* cannot both be constant. ■

One important advantage of Theorem 3.6 over (3.8) is that it can be exploited for inferential purposes. For instance, it enables us to build confidence intervals for $S_\epsilon(P,Q)$ as follows. Note that we can estimate the asymptotic variance in the one-sample CLT by

$$\hat{\sigma}_n^2 := \mathrm{Var}_{P_n}(f_{n,\epsilon}) = \frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) - \Big(\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)\Big)^2, \tag{3.9}$$

where $(f_{n,\epsilon}, g_{n,\epsilon})$ is a pair of optimal potentials for $S_\epsilon(P_n,Q)$. It follows from the proof of Lemma 3.2 that $\mathbb{E}\|f_{n,\epsilon} - f_\epsilon^*\|_s^2 \to 0$. Hence, $\|f_{n,\epsilon} - f_\epsilon^*\|_s \to 0$ in probability. Using the elementary bound $|a^2 - b^2| \le |a-b|^2 + 2|b||a-b|$, we see that

$$\left|\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) - \frac{1}{n}\sum_{i=1}^n f_\epsilon^{*2}(X_i)\right| \le (\|f_{n,\epsilon} - f_\epsilon^*\|_s^2 + 2\|f_\epsilon^*\|_s \|f_{n,\epsilon} - f_\epsilon^*\|_s)\frac{1}{n}\sum_{i=1}^n (1 + \|X_i\|^3).$$

Since $\frac{1}{n}\sum_{i=1}^n f_\epsilon^{*2}(X_i) \to \mathbb{E}_P(f_\epsilon^{*2})$ a.s., we conclude that $\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) \to \mathbb{E}_P(f_\epsilon^{*2})$ in probability and, arguing similarly for $\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)$, that

$$\hat{\sigma}_n^2 \to \mathrm{Var}_P(f_\epsilon^*) \quad \text{in probability.}$$

We conclude that

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(S_\epsilon(P_n,Q) - S_\epsilon(P,Q)) \xrightarrow{w} N(0,1)$$

and, as a consequence, that, writing z_β for the β quantile of the standard normal distribution,

$$\left[S_\epsilon(P_n,Q) \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\right] \tag{3.10}$$

is a confidence interval for $S_\epsilon(P,Q)$ of asymptotic level $1-\alpha$. A similar confidence interval can be constructed from the two-sample statistic. Such results will be illustrated in the simulations in section 6.

4. Convergence rates for optimal potentials. The goal of this section is to prove a bound on the difference between empirical potentials and their population counterparts. In this section we assume that both measures, P, Q, are supported on a compact set $\Omega \subset \mathbb{R}^d$. By translation invariance of the optimal transport problem, we may assume without loss of generality that $0 \in \Omega$. We write D_Ω for the diameter of Ω and let (f^*, g^*) be a pair of optimal potentials (maximizers of (2.1) for P and Q) and (f_n, g_n) their empirical counterpart (maximizers of (2.1) for P_n and Q). As noted above, these optimal potentials are unique up to an additive constant. In this section, we adopt the following normalization convention:

$$\int g^*(y)dQ(y) = \int g_n(y)dQ(y) = 0, \tag{4.1}$$

We show below that derivatives of the optimal potentials are uniformly bounded (see Lemma 4.1). Additionally, the choice of optimal potentials in (4.1) allows us to control uniformly the optimal potentials, as we show in Lemma 4.4. These are key ingredients for the aim of the section, namely, showing that the convergence rate of f_n (resp., g_n) towards f^* (resp., g^*) is $O_p(\frac{1}{\sqrt{n}})$.

The optimal potentials belong to the space $\mathcal{C}^s(\Omega)$ for $s = \lceil \frac{d}{2} \rceil + 1$, in which we consider the norm

$$\|f\|_{\mathcal{C}^s(\Omega)} = \sum_{i=0}^{s} \sum_{|\alpha|=i} \|D^\alpha f\|_\infty.$$

In this section, we use the notation $c_{d,s}, c'_{d,s}, \dots$ to indicate unspecified positive constants depending on d and s whose value may change from line to line. The optimality conditions (2.4) imply the following bounds (see Proposition 1 in [31]).

Lemma 4.1. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$. Then the optimal potentials (f^*, g^*) satisfy*

(i) $\min_{y\in\Omega}\{\frac{1}{2}\|x-y\|^2 - g^*(y)\} \le f^*(x) \le \max_{y\in\Omega}\{\frac{1}{2}\|x-y\|^2 - g^*(y)\}$,

(ii) *$f^*(x)$ is D_Ω-Lipschitz,*

(iii) *$f^* \in \mathcal{C}^\infty(\Omega)$ and $\|D^\alpha f^*\|_\infty \le C_{d,\alpha}(1 + D_\Omega^{|\alpha|})$ for all multi-indices α with $|\alpha| \ge 1$, for some constant $C_{d,\alpha}$ depending only on d and α.*

Proof. The first two claims are proven in Proposition 1 of [31], so it suffices to consider the last claim. We prove it for f^*, the case of g^* being similar. Define $\bar{f}^*(x) = f^*(x) - \frac{1}{2}\|x\|^2$. As in the proof of Proposition 1 in [54], the Faà di Bruno formula yields

$$-D^\alpha \overline{f}^*(x) = \sum_{\beta_1+\cdots+\beta_k=\alpha} \lambda_{\alpha,\beta_1,\dots,\beta_k} \prod_{j=1}^{k} \mu_{\beta_j,g} \quad \forall x \in \Omega,$$

where $\lambda_{\alpha,\beta_1,\dots,\beta_k}$ are combinatorial quantities and for a multi-index β we define

$$\begin{aligned}\mu_{\beta,g} &= \frac{\int y^\beta e^{g^*(y)-\frac{1}{2}\|y\|^2+\langle x,y\rangle}dQ(y)}{\int e^{g^*(y)-\frac{1}{2}\|y\|^2+\langle x,y\rangle}dQ(y)}\\ &= \int y^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dQ(y)\\ &= \int \prod_{i=1}^d y_i^{\beta_i} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dQ(y).\end{aligned}$$

By the optimality condition (2.4), $\int e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dQ(y)=1$. As a consequence, there exists $C'_{d,\alpha}$ such that $\|D^\alpha \overline{f}^*\|_\infty \le C'_{d,\alpha} D_\Omega^{|\alpha|}$. Since $\|D^\alpha \frac{1}{2}\|x\|^2\|_\infty \le 1+D_\Omega$ for $|\alpha|\ge 1$, we obtain $\|D^a f^*\| \le C'_{d,\alpha} D_\Omega^\alpha + 1 + D_\Omega \le C_{d,\alpha}(1+D_\Omega^{|\alpha|})$. ∎

Remark 4.2. Since the probabilities P_n and Q_n are also supported on the same compact set Ω, Lemma 4.1 holds also for f_n and g_n.

We also obtain bounds on the derivatives of $e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$.

Lemma 4.3. *For any multi-index β, the function $x \mapsto x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$ has $\mathcal{C}^s(\Omega)$ norm at most $c_{d,s} e^{D_\Omega^2}(1+D_\Omega^{s+|\beta|})$.*

Proof. By Lemma 4.1, $\|x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}\|_\infty \le D_\Omega^{|\beta|} e^{D_\Omega^2}$. For any $1\le|\alpha|\le s$, the Faà di Bruno formula implies

$$D^\alpha e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} = e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} \sum_{\lambda_1+\cdots+\lambda_s=\alpha} \gamma_{\alpha,\lambda_1,\ldots,\lambda_s} \prod_{j=1}^s D^{\lambda_j}\left(h^*(x,y)-\frac{1}{2}\|x-y\|^2\right)$$

for some combinatorial coefficients $\gamma_{\alpha,\lambda_1,\ldots,\lambda_s}$, where the derivative operators are taken with respect to the x variable. By Lemma 4.1, this quantity is bounded in magnitude by $c_{d,s}e^{D_\Omega^2}(1+D_\Omega^{|\alpha|})$ for some constant $c'_{d,s}$. This implies

$$|D^\alpha x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}| \le c''_{d,s} e^{D_\Omega^2}(1+D_\Omega^{s+|\beta|}) \qquad \forall 1\le|\alpha|\le s.$$

Therefore, choosing $c_{d,s}$ to be a sufficiently large constant depending on d and s yields the claim. ∎

For our particular choice of optimal potentials we can also control the uniform norm, as follows.

Lemma 4.4. *Under (4.1), we have*

$$\|f^*\|_\infty, \|f_n\|_\infty, \|g^*\|_\infty, \|g_n\|_\infty \le \frac{1}{2}D_\Omega^2.$$

Proof. Since

$$S_\epsilon(P,Q) = \int_{\mathbb{R}^d} f^*(x)dP(x) + \int_{\mathbb{R}^d} g^*(y)dQ(y) \ge 0,$$

(4.1) implies $\int_{\mathbb{R}^d} f^*(x)dP(x) \ge 0$. Therefore, using first the optimality conditions, then Jensen's inequality, and finally (4.1), we obtain

$$g^*(y) = -\log\left(\int e^{f^*(y)-\frac{1}{2}\|x-y\|^2}dP(y)\right) \le \int \left\{\frac{1}{2}\|x-y\|^2 - f^*(y)\right\}dP(y) \le \frac{1}{2}D_\Omega^2$$

for all $y \in \Omega$. By the same argument, $f^*(x) \le \frac{1}{2} D_\Omega^2$ for all $x \in \Omega$. Set $x \in \Omega$; then by Lemma 4.1,

$$f^*(x) \ge \min_{y\in\Omega} \left\{ \frac{1}{2}\|x-y\|^2 - g^*(y) \right\} \ge -\max_{y\in\Omega} g^*(y) \ge -\frac{1}{2} D_\Omega^2,$$

and the same for g^*. ■

For any $a, b \in \mathcal{C}^s(\Omega)$ denote

$$L(a,b) = \int_{\mathbb{R}^d} a(x)dP(x) + \int_{\mathbb{R}^d} b(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{a(x)+b(y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) + 1$$

and its semi-empirical counterpart

$$L_n(a,b) = \int_{\mathbb{R}^d} a(x)dP_n(x) + \int_{\mathbb{R}^d} b(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{a(x)+b(y)-\frac{1}{2}\|x-y\|^2} dP_n(x)dQ(y) + 1.$$

Let us denote by h^* and h_n the functions, belonging to $\mathcal{C}^s(\Omega\times\Omega)$, defined by

$$h^*(x,y) = f^*(x) + g^*(y), \qquad h_n(x,y) = f_n(x) + g_n(y), \tag{4.2}$$

and by $\pi^* \in \mathcal{P}(\Omega\times\Omega)$ the optimal coupling defined by $d\pi^* = e^{h^*(x,y)-\|x-y\|^2} dP(x)dQ(y)$. Abusing notation, we write $L(f_n, g_n) = L(h_n)$ and $L(f^*, g^*) = L(h^*)$. As a consequence of Lemma 4.4 we obtain the following useful bound:

$$\|h^*\|_\infty,\ \|h_n\|_\infty \le D_\Omega^2 \quad \forall n \in \mathbb{N}. \tag{4.3}$$

At this point, we can state the main theorem of this section.

Theorem 4.5. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$. Assume (f^*, g^*) are optimal potentials for P, Q and (f_n, g_n) for P_n, Q satisfying (4.1). Then there exists a constant C_d, depending only on d, such that*

$$\mathbb{E}\|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} \le \frac{C_d}{n} D_\Omega^{5(d+1)} e^{15 D_\Omega^2}.$$

Moreover, if $(f_{n,m}, g_{n,m})$ are optimal potentials for P_n, Q_m satisfying (4.1), then

$$\mathbb{E}\|g_{n,m} - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_{n,m} - f^*\|^2_{\mathcal{C}^s(\Omega)} \le \frac{C_d}{\min\{n,m\}} D_\Omega^{5(d+1)} e^{15 D_\Omega^2}.$$

The proof is divided in a sequence of technical lemmas, of some independent interest. We show first (Lemma 4.6) that the functional L is well-behaved in the sense of being strongly concave near its maximum. Then (in Lemma 4.7) we show that the functional $L_n - L$ is Lipschitz. Typically, these two results are enough to prove convergence at the fast n^{-1} rate (see, e.g., Theorem 3.2.5 of [77]). Unfortunately, in our case, the norms appearing in Lemmas 4.6 and 4.7 are different. This technical issue can be handled thanks to Lemmas 4.8 and 4.9.

Lemma 4.6. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$; then*

$$L(h_n) - L(h^*) \le -\frac{1}{2}\|h_n - h^*\|^2_{L^2(d\pi^*)} e^{-\|h_n - h^*\|_\infty}, \tag{4.4}$$

where h^*, h_n, and π^* are defined in (4.2).

Proof. The inequality $e^x \geq 1 + x + \frac{e^{-|x|}}{2} x^2$, which can be checked by elementary means, implies that

$$\begin{aligned}
&\int e^{h_n(x,y)-\frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&= \int e^{h_n(x,y)-h^*(x,y)} e^{h^*(x,y)-\frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&\geq \int \left\{ 1 + (h_n(x,y) - h^*(x,y)) + \frac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-|h_n(x,y)-h^*(x,y)|} \right\} d\pi^*(x,y) \\
&\geq \int \left\{ 1 + (h_n(x,y) - h^*(x,y)) + \frac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-||h_n-h^*||_\infty} \right\} d\pi^*(x,y).
\end{aligned}$$

The optimality conditions yield $L(h^*) = \int h^*(x,y)dP(x)dQ(y)$. Hence,

$$\begin{aligned}
&\int e^{h_n(x,y)-\frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&\quad \geq -L(h^*) + \int \left\{ 1 + h_n(x,y) + \frac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-||h_n-h^*||_\infty} \right\} d\pi^*(x,y).
\end{aligned}$$

We conclude by using the relation $\int h_n(x,y)d\pi^* = \int h_n(x,y)dP(x)dQ(y)$, which follows from the optimality conditions. ■

Lemma 4.7. *Under the assumptions of Lemma 4.6, we have*

$$L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \leq \|P - P_n\|_{\mathcal{C}_1^s(\Omega)} \|h_n - h^*\|_{\mathcal{C}^s(\Omega^2)}, \quad a.s., \tag{4.5}$$

where

$$\|P - P_n\|_{\mathcal{C}_1^s(\Omega)} := \sup_{\|f\|_{\mathcal{C}^s(\Omega)} \leq 1} \int f(x)(dP_n(x) - dP(x)). \tag{4.6}$$

Proof. As noted above, the optimality conditions imply that

$$\begin{aligned}
L_n(h_n) &= \int h_n(x,y)dP_n(x)dQ(y), \quad L(h_n) = \int h_n(x,y)dP(x)dQ(y), \\
L_n(h^*) &= \int h^*(x,y)dP_n(x)dQ(y), \quad L(h^*) = \int h^*(x,y)dP(x)dQ(y).
\end{aligned}$$

Therefore we have

$$\begin{aligned}
&L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \\
&\quad = \int h_n(x,y)dQ(y)(dP_n(x) - dP(x)) - \int h^*(x,y)dQ(y)(dP_n(x) - dP(x)) \\
&\quad = \int (h_n(x,y) - h^*(x,y))dQ(y)(dP_n(x) - dP(x)) \\
&\quad \leq \|h_n - h^*\|_{\mathcal{C}^s(\Omega)} \sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)} \leq 1 \\ h(x,y)=f(x)+g(y)}} \int h(x,y)dQ(y)(dP_n(x) - dP(x)).
\end{aligned}$$

Note that

$$\begin{aligned}&\sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)}\le 1\\ h(x,y)=f(x)+g(y)}}\int h(x,y)dQ(y)(dP_n(x)-dP(x))\\ &\quad=\sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)}\le 1\\ h(x,y)=f(x)+g(y)}}\int g(y)dQ(y)(dP_n(x)-dP(x))+\int f(x)dQ(y)(dP_n(x)-dP(x)).\end{aligned}$$

Since the first term is 0 and the second is not affected by adding a constant to f, we see that it equals

$$\sup_{\|f\|_{\mathcal{C}^s(\Omega)}\le 1}\int f(x)(dP_n(x)-dP(x)).$$

■

As anticipated, Lemma 4.7 works with the norm $\|\cdot\|^2_{\mathcal{C}^s(\Omega^2)}$ and Lemma 4.6 with $\|\cdot\|_{L^2(d\pi^*)}$. Both norms are different, but the next technical results show how these norms are related in the present setup.

Lemma 4.8. *Under the assumptions of Lemma 4.6,*

$$\begin{aligned}&\|D^\alpha f^*-D^\alpha f_n\|^2_\infty\le c_{d,s}D_\Omega^{2|\alpha|}\|h_n-h^*\|^2_\infty,\\ &\|D^\alpha g^*-D^\alpha g_n\|^2_\infty\le c_{d,s}D_\Omega^{2|\alpha|}e^{2D_\Omega^2}\left(\|h_n-h^*\|^2_\infty+D_\Omega^{2s}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}\right)\end{aligned}$$

for every multi-index α, with $1\le|\alpha|\le s$.

Proof. We let $c_{d,s}$ denote a positive constant depending on d and s whose value may change from line to line. We note first that $f^*(x)-f_n(x)=\bar f^*(x)-\bar f_n(x)$, where

$$\bar f^*(x)=f^*(x)-\frac{1}{2}\|x\|^2\quad\text{and}\quad \bar f_n(x)=f_n(x)-\frac{1}{2}\|x\|^2. \tag{4.7}$$

As in the proof of Lemma 4.1, the Faá di Bruno formula implies

$$\begin{aligned}&D^\alpha f_n(x)-D^\alpha f^*(x)=D^\alpha\overline{f}_n(x)-D^\alpha\overline{f}^*(x)\\ &=\sum_{\beta_1+\dots+\beta_s=\alpha}\lambda_{\alpha,\beta_1,\dots,\beta_s}\left(\prod_{j=1}^s\int y^{\beta_j}e^{h_n(x,y)-\frac12\|x-y\|^2}dQ(y)-\prod_{j=1}^s\int y^{\beta_j}e^{h^*(x,y)-\frac12\|x-y\||^2}dQ(y)\right).\end{aligned}$$

Splitting the product, this last term equals

$$\begin{aligned}&\sum_{\beta_1+\dots+\beta_s=\alpha}\lambda_{\alpha,\beta_1,\dots,\beta_s}\sum_{i=1}^s\prod_{j<i}\int y^{\beta_j}e^{h_n(x,y)-\frac12\|x-y\|^2}dQ(y)\\ &\qquad\prod_{j>i}\int y^{\beta_j}e^{h^*(x,y)-\frac12\|x-y\|^2}dQ(y)\int y^{\beta_i}e^{-\frac12\|x-y\|^2}\left\{e^{h_n(x,y)}-e^{h^*(x,y)}\right\}dQ(y).\end{aligned}$$

Since $0\in\Omega$, it follows that $|y^{\beta_j}|\le D_\Omega^{|\beta_j|}$. Using that $|e^x-e^y|\le(e^y+e^x)|x-y|$ we upper bound $|D^\alpha f_n(x)-D^\alpha f^*(x)|$ by

$$D_\Omega^{|\alpha|}\sum_{\beta_1+\cdots+\beta_s=\alpha}|\lambda_{\alpha,\beta_1,\dots,\beta_s}|\sum_{i=1}^s\int(e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}+e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2})|h_n(x,y)-h^*(x,y)|dQ(y)$$
$$\le 2sD_\Omega^{|\alpha|}\sum_{\beta_1+\cdots+\beta_s=\alpha}|\lambda_{\alpha,\beta_1,\dots,\beta_s}|\|h_n-h^*\|_\infty,$$

where we have used (2.4) to bound the integral. We conclude that $\|D^af_n(x)-D^af^*(x)\|_\infty^2\le c_{d,s}D_\Omega^{2|\alpha|}\|h_n-h^*\|_\infty^2$.

Turning to g_n and g_n^*, we can argue similarly to obtain

$$\begin{aligned}
&|D^\alpha g_n(y)-D^\alpha g^*(y)|=|D^\alpha\overline{g}_n(y)-D^\alpha\overline{g}^*(y)|\\
&\quad\le\sum_{\beta_1+\cdots+\beta_s=\alpha}|\lambda_{\alpha,\beta_1,\dots,\beta_s}|\left(\prod_{j=1}^s\int x^{\beta_j}e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2}dP_n(x)-\prod_{j=1}^s\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dP(x)\right)\\
&\quad=\sum_{\beta_1+\cdots+\beta_s=\alpha}|\lambda_{\alpha,\beta_1,\dots,\beta_s}|\sum_{i=1}^s\prod_{j<i}\int x^{\beta_j}e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2}dP_n(x)\prod_{j>i}\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dP(x)\cdot\\
&\qquad\left(\int x^{\beta_j}e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2}dP_n(x)-\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dP(x)\right).
\end{aligned}$$

Note that

$$\begin{aligned}
&\left|\int x^{\beta_j}e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2}dP_n(x)-\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}dP(x)\right|\\
&\quad\le\left|\int x^{\beta_j}(e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2}-e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2})dP_n(x)\right|+\left|\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}(dP(x)-P_n(x))\right|.
\end{aligned}$$

Since $\|h_n\|_\infty,\|h^*\|_\infty\le D_\Omega^2$ by (4.3), the first term can be bounded by $2D_\Omega^{|\beta_j|}e^{D_\Omega^2}\|h_n-h^*\|_\infty$. For the other term observe that by Lemma 4.3, the function $x\mapsto x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$ belongs to $\mathcal{C}^s(\Omega)$, with norm at most $c_{d,s}e^{D_\Omega^2}(1+D_\Omega^{s+|\beta_j|})$. We conclude that there exists some constant $c_{d,s}$ such that

$$\left|\int x^{\beta_j}e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}(dP_n(x)-dP(x))\right|\le c_{d,s}D_\Omega^{|\beta_j|+s}e^{D_\Omega^2}\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}. \tag{4.8}$$

Combining the last two estimates we finally have

$$\|D^ag^*-D^ag_n\|_\infty^2\le c_{d,s}D_\Omega^{2|\alpha|}e^{2D_\Omega^2}\left(\|h_n-h^*\|_\infty^2+D_\Omega^{2s}\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}^2\right),$$

which allows us to conclude the proof. ∎

Now we need to compare the norms $\|\cdot\|_\infty$ and $\|\cdot\|_{L^2(d\pi^*)}$. We set $C=e^{-\frac{3}{2}D_\Omega^2}$ and note that (4.1) implies

$$\begin{aligned}\int (h_n(x,y) - h^*(x,y))^2 e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) \\ \geq C \int (h_n(x,y) - h^*(x,y))^2 dP(x)dQ(y) \\ = C\Big\{ \int (f_n(x) - f^*(x))^2 dP(x) + \int (g_n(y) - g^*(y))^2 dQ(y) \\ +2 \int (f_n(x) - f^*(x))(g_n(y) - g^*(y)) dP(x)dQ(y) \Big\}.\end{aligned}$$

Since the last term equals 0, we obtain the bound

$$\|h_n - h^*\|^2_{L^2(d\pi^*)} \geq e^{-\frac{3}{2}D_\Omega^2} \left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)} \right). \tag{4.9}$$

Finally, we prove the last technical result, which relates the L^2 and L^∞ norms for the difference of the potentials.

Lemma 4.9. *Under the assumptions of Lemma 4.6, we have*

$$\begin{aligned}\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)} \\ \geq \frac{1}{2} e^{-2D_\Omega^2} \left(\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty \right) - c_{d,s} D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}.\end{aligned}$$

Proof. We will work separately with f^* and g^*. Fixing $x \in \Omega$, Jensen's inequality yields

$$|e^{-f^*(x)} - e^{-f_n(x)}|^2 \leq \int \left(|e^{g^*(y)-\frac{1}{2}\|x-y\|^2} - e^{g_n(y)-\frac{1}{2}\|x-y\|^2}| \right)^2 dQ(y).$$

Now, the mean value theorem implies

$$|x-y| e^{\min\{x,y\}} \leq |e^x - e^y| \leq e^{\max\{x,y\}} |x-y|, \quad x,y \in \mathbb{R},$$

yielding

$$\begin{aligned}e^{-2\max\{\|f^*\|_\infty, \|f_n\|_\infty\}} |f_n(x) - f^*(x)|^2 \leq |e^{-f^*(x)} - e^{-f_n(x)}|^2 \\ \leq e^{2\max\{\|g^*\|_\infty, \|g_n\|_\infty\}} \|g_n - g^*\|^2_{L^2(dQ)}.\end{aligned}$$

Consequently, using Lemma 4.4, we have proved that

$$\|g_n - g^*\|^2_{L^2(dQ)} \geq e^{-2D_\Omega^2} \|f_n - f^*\|^2_\infty.$$

Now we deal with $\|g_n - g^*\|^2_\infty$. We fix $y \in \Omega$. By the triangle inequality we have

$$\begin{aligned}|e^{-g^*(y)} - e^{-g_n(y)}| \\ \leq \int \left| e^{f^*(x)+\frac{1}{2}\|x-y\|^2} - e^{f_n(y)+\frac{1}{2}\|x-y\|^2} \right| dP(y) + \left| \int e^{f_n(x)+\frac{1}{2}\|x-y\|^2} (dP(x) - dP_n) \right|.\end{aligned}$$

Squaring both sides we see that

$$\begin{aligned}|e^{-g^*(y)} - e^{-g_n(y)}|^2 \\ \leq 2 \int \left| e^{f^*(x)+\frac{1}{2}\|x-y\|\|^2} - e^{f_n(y)+\frac{1}{2}\|x-y\|^2} \right|^2 dP(y) + 2 \left| \int e^{f_n(x)+\frac{1}{2}\|x-y\|^2} (dP(x) - dP_n) \right|^2.\end{aligned}$$

The first term is bounded by $2e^{D_\Omega^2}\|f_n - f^*\|^2_{L^2(dP)}$ as in the previous case. Repeating the arguments which led to the bound (4.8), the second term is at most $c_{d,s}e^{D_\Omega^2}(1 + D_\Omega^{2s})\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}$. Together, these estimates yield

$$e^{-D_\Omega^2}\|g_n - g^*\|^2_\infty \leq 2e^{D_\Omega^2}\|f_n - f^*\|^2_{L^2(dP)} + c_{d,s}D_\Omega^{2s}e^{D_\Omega^2}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}.$$

We conclude by rearranging this inequality and combining it with the bound on $\|g_n - g^*\|^2_{L^2(dQ)}$ derived above. ■

We are ready now for the proof of the main result in this section.

Proof of Theorem 4.5. Combining Lemma 4.6, (4.9), and Lemma 4.9, we see that

$$\begin{aligned} L(h^*) - L(h_n) &\geq \frac{1}{2}e^{-\|h_n - h^*\|_\infty}e^{-\frac{3}{2}D_\Omega^2}\left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)}\right) \\ &\geq \frac{1}{2}e^{-\frac{7}{2}D_\Omega^2}\left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)}\right) \\ &\geq \frac{1}{2}e^{-\frac{7}{2}D_\Omega^2}\left(\frac{1}{2}e^{-2D_\Omega^2}\left(\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty\right) - c_{d,s}D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}\right). \end{aligned}$$

Moreover, since $\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty \geq \frac{1}{2}\|h_n - h^*\|^2_\infty$, we obtain

$$L(h^*) - L(h_n) \geq \frac{1}{2}e^{-\frac{7}{2}D_\Omega^2}\left(\frac{1}{4}e^{-2D_\Omega^2}\|h_n - h^*\|^2_\infty - D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}\right).$$

Lemma 4.8 implies the existence of some constant $c_{d,s}$ such that

$$\|h_n - h^*\|^2_\infty \geq \frac{1}{c_{d,s}D_\Omega^{2s}e^{2D_\Omega^2}}\left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)}\right) - D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)},$$

which yields

$$\begin{aligned} L(h^*) - L(h_n) \geq c_{d,s}e^{-\frac{15}{2}D_\Omega^2}D_\Omega^{-2s}&\left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)}\right) \\ &- c'_{d,s}e^{-\frac{7}{2}D_\Omega^2}D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}. \end{aligned} \tag{4.10}$$

On the other hand, Lemma 4.7 yields

$$\begin{aligned} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)}\left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)}\right) &\geq L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \\ &\geq L_n(h^*) - L(h_n) - L_n(h^*) + L(h^*) \\ &= L(h^*) - L(h_n). \end{aligned}$$

The previous bound and (4.10) yield

$$\begin{aligned} &\|P - P_n\|_{\mathcal{C}_1^s(\Omega)}\left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)}\right) \\ &\quad\geq c_{d,s}e^{-\frac{15}{2}D_\Omega^2}D_\Omega^{-2s}\left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)}\right) - c'_{d,s}e^{-\frac{7}{2}D_\Omega^2}D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}, \end{aligned}$$

which, by using the inequality $(a+b)^2 \le 2(a^2+b^2)$, implies

$$\begin{aligned}\sqrt{2}\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}&\left(\|f_n-f^*\|^2_{\mathcal{C}^s(\Omega)}+\|g_n-g^*\|^2_{\mathcal{C}^s(\Omega)}\right)^{1/2}\\ &\ge c_{d,s}e^{-\frac{15}{2}D_\Omega^2}D_\Omega^{-2s}\left(\|f_n-f^*\|^2_{\mathcal{C}^s(\Omega)}+\|g_n-g^*\|^2_{\mathcal{C}^s(\Omega)}\right)-c'_{d,s}e^{-\frac{7}{2}D_\Omega^2}D_\Omega^{2s}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}.\end{aligned}$$

Denoting $\Delta_n = \left(\|f_n-f^*\|^2_{\mathcal{C}^s(\Omega)}+\|g_n-g^*\|^2_{\mathcal{C}^s(\Omega)}\right)^{\frac{1}{2}}$, we get

$$\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}\Delta_n \ge c_{d,s}e^{-\frac{15}{2}D_\Omega^2}D_\Omega^{-2s}\Delta_n^2 - c'_{d,s}e^{-\frac{7}{2}D_\Omega^2}D_\Omega^{2s}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}. \tag{4.11}$$

From this we obtain

$$\begin{aligned}\Delta_n &\le c_{d,s}D_\Omega^{2s}e^{\frac{15}{2}D_\Omega^2}\left(\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}+\sqrt{\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}+e^{-11D_\Omega^2}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}}\right)\\ &\le c_{d,s}D_\Omega^{2s}e^{\frac{15}{2}D_\Omega^2}\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}.\end{aligned} \tag{4.12}$$

Next, we analyze $\|P-P_n\|_{\mathcal{C}_1^s(\Omega)}$. Theorem 3.5.1 and Exercise 2.3.1 in [35] imply that there exists a numerical constant C such that

$$n\mathbb{E}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)} \le C\mathbb{E}\left(\int_0^{\max_{\|f\|_{\mathcal{C}^s(\Omega)}\le 1}\|f\|_{L^2(dP_n)}}\sqrt{\log\left(2N(\epsilon,\mathcal{C}_1^s(\Omega),\|\cdot\|_\infty)\right)}d\epsilon\right)^2.$$

By Proposition 1.1. in [76],

$$\log\left(2N(\epsilon,\mathcal{C}_1^s(\Omega),\|\cdot\|_\infty)\right) \le c_{s,d}D_\Omega^d\left(\frac{1}{\epsilon}\right)^{\frac{d}{s}}.$$

Since $\|f\|_{L^2(dP_n)} \le \|f\|_{\mathcal{C}^s(\Omega)}$ a.s., the choice $s=[d/2]+1$ yields the bound

$$n\mathbb{E}\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)} \le c_{s,d}D_\Omega^d\left(\int_0^1\left(\frac{1}{\epsilon}\right)^{\frac{d}{d+1}}d\epsilon\right)^2 = c_{s,d}D_\Omega^d, \tag{4.13}$$

which completes the proof for the one-sample case.

The two-sample case can be handled with the same argument plus some minor modifications, as follows. Let $f_{n,m}$ be the optimal potential for P_n and Q_m. Then,

$$\|f_{n,m}-f^*\|^2_{\mathcal{C}^s(\Omega)} \le 2\|f_n-f^*\|^2_{\mathcal{C}^s(\Omega)} + 2\|f_{n,m}-f_n\|^2_{\mathcal{C}^s(\Omega)}.$$

The first term can be controlled by (4.12). Moreover, observe that the derivation of (4.12) did not use any facts about the measure Q apart from the fact that it is supported on Ω. Since P_n is also supported on Ω, this implies that $\|f_{n,m}-f_n\|_{\mathcal{C}^s(\Omega)}$ can also be controlled by (4.12), so that the bound

$$\|f_{n,m}-f^*\|^2_{\mathcal{C}^s(\Omega)} \le c_{d,s}D_\Omega^{2s}e^{\frac{15}{2}D_\Omega^2}\left(\|P-P_n\|^2_{\mathcal{C}_1^s(\Omega)}+\|Q-Q_m\|^2_{\mathcal{C}_1^s(\Omega)}\right) \tag{4.14}$$

holds. This and (4.13) complete the proof. ■

5. Convergence rates for Sinkhorn divergences. In this section, we develop faster convergence rates for the *Sinkhorn divergence.* The entropic transportation cost, $S_\epsilon(P,Q)$, is not symmetric in P,Q and does not satisfy $S_\epsilon(P,P)=0$. These observations motivated the introduction of Sinkhorn divergences [32]: For probabilities $P,Q\in\mathcal{P}(\mathbb{R}^d)$ the quadratic Sinkhorn divergence is defined as

$$D_\epsilon(P,Q)=S_\epsilon(P,Q)-\frac{1}{2}\left(S_\epsilon(P,P)+S_\epsilon(Q,Q)\right).$$

Clearly, $D_\epsilon(P,Q)$ is symmetric in P,Q and $D_\epsilon(P,P)=0$. In fact (see Theorem 1 in [27]), $D_\epsilon(P,Q)\geq 0$, with $D_\epsilon(P,Q)=0$ if and only if $P=Q$, and for measures supported on a compact set, convergence in Sinkhorn distance is equivalent to weak convergence. This makes the Sinkhorn divergence a suitable measure of dissimilarity in applications.

In this section we obtain rates of convergence for empirical Sinkhorn divergences. More precisely, we consider independent samples $X_1,\dots,X_n$, $Y_1,\dots,Y_m$ of i.i.d. random variables with law $P\in\mathcal{P}(\Omega)$ and associated empirical measures P_n and P'_m, respectively. Since P_n and P'_m converge weakly to P, the Sinkhorn divergence satisifes $D_\epsilon(P_n,P'_m)\to 0$ a.s. The main result of this section gives a rate for this convergence.

Theorem 5.1. *Assume $\Omega\subset\mathbb{R}^d$ is compact, $P\in\mathcal{P}(\Omega)$, and P_n and P'_m are empirical measures as above. Then there exist constants c_d and c'_d, depending only on d, such that*

(i) *(one-sample case)*

$$\mathbb{E}D_1(P_n,P)\leq\frac{c_d}{n}D_\Omega^{\frac{3d}{2}+1}\frac{32}{(d+1)^2}e^{\frac{19}{2}D_\Omega^2};$$

(ii) *(two-sample case)*

$$\mathbb{E}D_1(P_n,P'_m)\leq\frac{c'_d}{\min\{n,m\}}D_\Omega^{\frac{3d}{2}+1}\frac{32}{(d+1)^2}e^{\frac{19}{2}D_\Omega^2}.$$

Proof. We deal first with the one-sample case. We denote by $(f_{n,n},g_{n,n})$ the optimal potentials for $S_1(P_n,P_n)$, set $h_{n,n}(x,y)=f_{n,n}(x)+g_{n,n}(y)$, and write $d\pi_{n,n}(x,y)=e^{h_{n,n}(x,y)-\frac{1}{2}\|x-y\|^2}$ for the optimal measure. Further, as in (4.2), we write h^*, π^* for the corresponding objects in the case of $S_1(P,P)$ and h_n, π_n in the case of $S_1(P_n,P)$. Then we can write

$$D_1(P_n,P)=\int h_n(x,y)d\pi_n(x,y)-\frac{1}{2}\left(\int h_{n,n}(x,y)d\pi_{n,n}(x,y)+\int h^*(x,y)d\pi^*(x,y)\right).\tag{5.1}$$

Moreover, using the optimality conditions, we have

$$\int\left(h_n(x,y)-h^*(x,y)\right)d\pi^*(x,y)=L(h_n)-L(h^*)\leq 0$$

and

$$\int\left(h_n(x,y)-h_{n,n}(x,y)\right)d\pi_{n,n}(x,y)=L_{n,n}(h_n)-L_{n,n}(h_{n,n})\leq 0,$$

where L is defined as in the previous section and

$$L_{n,n}(h) = \int \left\{h(x,y) - e^{h(x,y)-\frac{1}{2}||x-y||^2} + 1\right\} dP_n(x)dP_n(x).$$

Therefore, from (5.1) we obtain

$$D_1(P_n,P) \le \int h_n(x,y)d\pi_n(x,y) - \frac{1}{2}\left(\int h_n(x,y)d\pi_{n,n}(x,y) + \int h_n(x,y)d\pi^*(x,y)\right). \tag{5.2}$$

Note, moreover, that the upper bound in (5.2) can be rewritten as

$$\begin{aligned}
&\int f_n(x)dP_n(x) + \int g_n(y)dP(y)\\
&\quad - \frac{1}{2}\left(\int f_n(x)dP_n(x) + \int g_n(y)dP_n(y) + \int f_n(x)dP(x) + \int g_n(y)dP(y)\right)\\
&\quad = \frac{1}{2}\int (f_n(x) - g_n(x))\, dP_n(x) + \frac{1}{2}\int (g_n(x) - f_n(x))\, dP(x)\\
&\quad = \frac{1}{2}\int (f_n(x) - g_n(x))\,(dP_n(x) - dP(x))\\
&\quad = \frac{1}{2}\int (f_n(x) - g^*(x))\,(dP_n(x) - dP(x))\\
&\quad + \frac{1}{2}\int (g^*(x) - g_n(x))\,(dP_n(x) - dP(x)),
\end{aligned} \tag{5.3}$$

where (f_n, g_n) are optimal entropic potentials for P_n, P and (f^*, g^*) are optimal transport potentials for (P,P), where we adopt the normalization convention $\int g^*(y)dP(y) = \int g_n(y) dP(y) = 0$. The symmetry of $S_1(P,P)$ and the uniqueness of the entropic potentials up to additive constants imply that $f^* = g^* + a$ for some constant $a \in \mathbb{R}$. Plugging this into (5.3) we obtain from (5.2) that

$$D_1(P_n,P) \le \frac{1}{2}\|P - P_n\|_{\mathcal{C}_1^s(\Omega)}\left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)}\right). \tag{5.4}$$

From (4.12), we obtain, for some constant $c_{d,s}$, the bound

$$\begin{aligned}
\left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)}\right) &\le 2\left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)}\right)^{\frac{1}{2}}\\
&\le c_{d,s} D_\Omega^{2s} e^{\frac{15}{2}D_\Omega^2}\|P - P_n\|_{\mathcal{C}_1^s(\Omega)}.
\end{aligned}$$

We conclude as in the proof of Theorem 4.5.

For the two-sample case we can adapt the argument above without much effort. Indeed, observe that we can write

$$\begin{aligned}
D_1(P_n, P'_m) = &\int h_{n,m}(x,y)d\pi_{n,m}(x,y)\\
&- \frac{1}{2}\left(\int h_{n,n}(x,y)d\pi_{n,n}(x,y) + \int h_{m,m}(x,y)d\pi_{m,m}(x,y)\right)
\end{aligned} \tag{5.5}$$

and argue as in (5.2) to get

$$D_1(P_n, P'_m) \leq \int h_{n,m}(x,y)d\pi_{n,m}(x,y) - \frac{1}{2}\left(\int h_{n,m}(x,y)d\pi_{n,n}(x,y) + \int h_{n,m}(x,y)d\pi_{m,m}(x,y)\right). \tag{5.6}$$

Now we can reuse the same arguments leading to (5.2)—just replacing P by P'_m—to upper bound $D_1(P_n, P'_m)$ by

$$\frac{1}{2}\int (f_{n,m}(x) - g^*(x))\left(dP_n(x) - dP'_m(x)\right) + \frac{1}{2}\int (g^*(x) - g_{n,m}(x))\left(dP_n(x) - dP'_m(x)\right). \tag{5.7}$$

Once again, since (f^*, g^*) agree up to an additive constant, (5.7) is equivalent to

$$\frac{1}{2}\int (f_{n,m}(x) - f^*(x))\left(dP_n(x) - dP'_m(x)\right) + \frac{1}{2}\int (g^*(x) - g_{n,m}(x))\left(dP_n(x) - dP'_m(x)\right).$$

Finally, the two-sample case can be deduced directly from the inequality

$$\begin{aligned} D_1(P_n, P) &\leq \frac{1}{2}\|P_n - P'_m\|_{\mathcal{C}_1^s(\Omega)}\left(\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)}\right) \\ &\leq \frac{1}{2}\left(\|P'_m - P\|_{\mathcal{C}_1^s(\Omega)} + \|P_n - P\|_{\mathcal{C}_1^s(\Omega)}\right)\left(\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)}\right) \end{aligned} \tag{5.8}$$

and (4.14), which yields

$$\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)} \leq c_{d,s} D_\Omega^{2s} e^{\frac{15}{2}D_\Omega^2}\left(\|P - P_n\|_{\mathcal{C}_1^s(\Omega)} + \|P - P'_m\|_{\mathcal{C}_1^s(\Omega)}\right)$$

for a constant $c_{d,s}$ depending on d and s. We conclude as above. ■

6. Implementation issues and empirical results.

In this section we provide details about the practical implementation and statistical performance of the two-sample analogue of the confidence interval (3.10).

Recall from Theorem 3.6 that

$$\sqrt{\frac{nm}{n+m}}\frac{1}{\sigma_{\epsilon,\lambda}(P,Q)}(S_\epsilon(P_n, Q_m) - S_\epsilon(P,Q)) \xrightarrow{w} N(0,1), \tag{6.1}$$

where $\sigma^2_{\epsilon,\lambda}(P,Q)$ is the asymptotic variance of the two-sample case. This variance can be consistently estimated by

$$\begin{aligned} \hat{\sigma}^2_{n,m} := &\frac{m}{n+m}\left(\frac{1}{n}\sum_{i=1}^n f^2_{n,\epsilon}(X_i) - \left(\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)\right)^2\right) \\ &+ \frac{n}{n+m}\left(\frac{1}{m}\sum_{i=1}^m g^2_{n,\epsilon}(Y_i) - \left(\frac{1}{m}\sum_{i=1}^m g_{n,\epsilon}(Y_i)\right)^2\right), \end{aligned} \tag{6.2}$$

where $(f_{n,\epsilon}, g_{n,\epsilon})$ is a pair of empirical potentials. Since the optimal transport potentials can be computed directly from the Sinkhorn iterates, the computation of $(f_{n,\epsilon}, g_{n,\epsilon})$ does not cost more than the one of $S_\epsilon(P_n, Q_m)$; see [15]. Hence, writing z_β for the β quantile of the standard normal distribution and arguing as in section 3, we can conclude that the interval

$$CI_\alpha^{n,m} = \left[S_\epsilon(P_n, Q_m) \pm \hat{\sigma}_{n,m} \sqrt{\frac{n+m}{nm}} z_{1-\alpha/2} \right]$$

is an asymptotic confidence interval of level $1-\alpha$.

We investigate here the finite sample performance of this confidence interval. We consider the scenario where $P \sim N(0, I_d/2)$ and $Q \sim N((1,\ldots,1)^t, I_d/2)$. The population entropy regularized cost has a closed form for Gaussian measures (see [24], [45], or [52]), which, for our case, is

$$S_\epsilon(P,Q) = 2\,d - \frac{\epsilon}{2}\left(d\sqrt{1+\frac{4}{\epsilon^2}} - d\log\left(1+\sqrt{1+\frac{4}{\epsilon^2}}\right) + d\,\log(2) - d\right).$$

We focus on the case $n = m$ for several choices of $n = 50, 100, 250, 500, 1000, 5000$ and study the influence of the parameters d and ϵ on the rate of convergence of the true confidence level to the nominal level $1-\alpha$ for $\alpha = 0.05$. To approximate this true confidence level we use Monte Carlo simulation, with 1000 replicates of the interval. The results are reported in Table 1. In particular, we compute $CI_{0.05}^{n,n}$ for different values of $\epsilon \in [0.5, 2, 5, 10]$ and $d \in [2, 10, 15]$. To calculate $S_\epsilon(P_n, Q_n)$ and the empirical potentials—which allows us to compute $CI_{0.05}^{n,n}$—we use the Python library POT; see [28].

We observe that both d and ϵ affect the estimation of the asymptotic confidence interval $CI_{0.05}^{n,n}$. A large sample size is required to achieve the nominal confidence interval for small values of ϵ and large dimension. This is more or less expected: in view of Remark 3.5, the value $n\,|\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P,Q)|$ can be upper bounded by a constant C_Ω, which depends

Table 1

Evolution of the Monte Carlo estimation (number of iterations equals 1000*) of* $\mathbb{P}(S_\epsilon(P,Q) \in CI_{0.05}^{n,n})$ *for different values of the dimension* d *and regularization factor* ϵ.

		$\mathbb{P}(S_\epsilon(P,Q) \in CI_{0.05}^{n,n})$			
	n	$\epsilon = 0.5$	$\epsilon = 2$	$\epsilon = 5$	$\epsilon = 10$
$d = 2$	50	0.935	0.936	0.932	0.941
	100	0.937	0.952	0.929	0.941
	250	0.95	0.94	0.935	0.949
	500	0.954	0.947	0.95	0.958
	1000	0.944	0.954	0.947	0.96
	5000	0.939	0.957	0.947	0.955
$d = 10$	50	0.781	0.945	0.958	0.932
	100	0.787	0.937	0.951	0.945
	250	0.775	0.941	0.948	0.943
	500	0.785	0.955	0.953	0.947
	1000	0.803	0.94	0.945	0.954
	5000	0.862	0.944	0.946	0.951
$d = 15$	50	0.487	0.944	0.933	0.944
	100	0.396	0.944	0.957	0.944
	250	0.271	0.938	0.943	0.953
	500	0.194	0.94	0.941	0.947
	1000	0.173	0.938	0.945	0.955
	5000	0.134	0.942	0.943	0.943

exponentially on the support's diameter, and—extrapolating this argument to the case where the probabilities are not supported in a compact set—this provides a possible explanation of the inaccuracy produced by the choice of small values ϵ or large d. (Note that this exponential dependency on the diameter is translated directly to an exponential dependence on ϵ^{-1} by the change of variables $x \mapsto \epsilon^{-\frac{1}{2}}x$.) Moreover, as $\epsilon \to 0$, the entropic regularized potentials approach the unregularized optimal transport potentials (see [59]), whose $n \to \infty$ convergence suffers from the curse of dimensionality (see [78]). These observations help explain the results of Table 1.

The most counterintuitive aspect we find in these simulations is the fact that in dimension 15 with $\epsilon = 0.5$, the performance worsens as sample size increases. It is reasonable to conjecture that this failure is related to the fact that the bias $|\mathbb{E}S_\epsilon(P_n, Q_n) - S_\epsilon(P, Q)|$ is still nonnegligible for moderate sample sizes. To explore this phenomenon, Figure 1 computes $|\mathbb{E}S_\epsilon(P_n, Q_n) - S_\epsilon(P, Q)|$ for $n = 10, 20, 30, 50, 100, 200, 500$ and $d = 2, 5, 10, 15$.

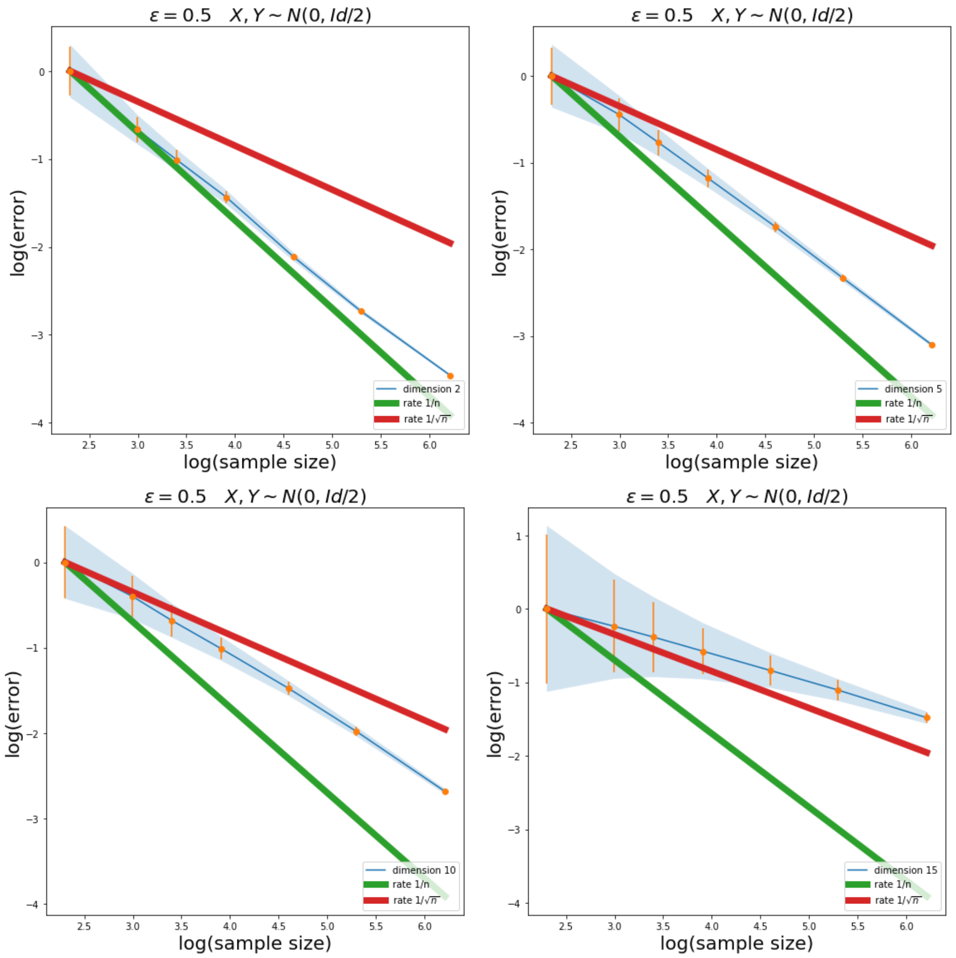

Figure 1. *Estimation, via Monte Carlo with* 1000 *repetitions, of* $|\mathbb{E}S_\epsilon(P_n, Q_n) - S_\epsilon(P, Q)|$ *for* $n = 10, 20, 30, 50, 100, 200, 500$ *and* $d = 2, 5, 10, 15$. *The results are plotted in double logarithm scale. The orange line represents the Monte Carlo standard deviation and the shadow the one given by* (6.2).

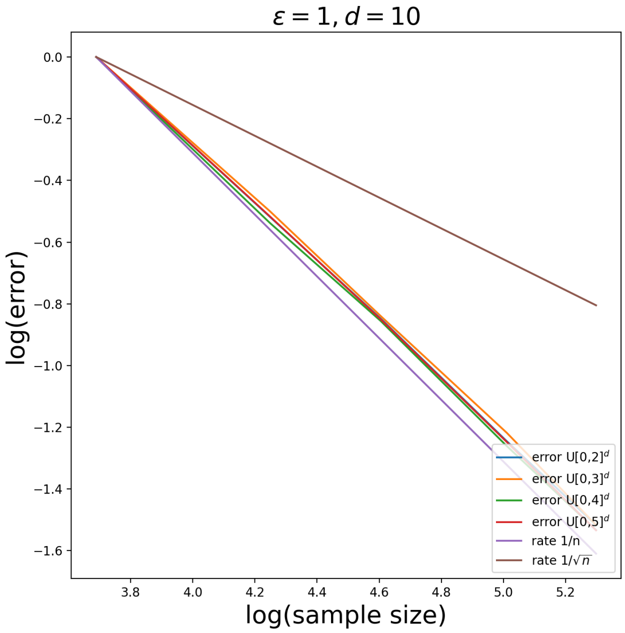

Figure 2. *Estimation, via Monte Carlo with* 1000 *repetitions, of the Sinkhorn divergence between* $X_1, \ldots, X_n$ *and* $Y_1, \ldots, Y_n$ *uniformly distributed in the square* $[0, k]^10$ *for* $k = 2, 3, 4, 5$ *and* $n = 40, 70, 100, 150, 200$. *The results are plotted in double logarithm scale.*

The empirical conclusions could not be more clear: the error lies in the bias. Though our bounds imply that the bias is asymptotically of order n^{-1}, at small sample sizes the rate of decay is slower and gets worse as the dimension increases. As Figure 1 indicates, the bias is still negligible for $d = 2, 5, 10$ but becomes dominant for $d = 15$. Though our confidence intervals are asymptotically valid, the large bias causes the coverage to worsen as n increases when the overall number of samples is small.

To illustrate the rate given in Theorem 5.1, we sample $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ uniformly distributed in the square $[0, k]^{10}$ for $k = 2, 3, 4, 5$ and $n = 40, 70, 100, 150, 200$. We estimate the expectation of the divergence by Monte Carlo with 1000 independent samples. The results are reported in Figure 2. This simulation agrees closely with the theoretical bounds.

REFERENCES

[1] A. Abadie and G. W. Imbens, *Large sample properties of matching estimators for average treatment effects*, Econometrica, 74 (2006), pp. 235–267.

[2] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme, *Asymptotics for semidiscrete entropic optimal transport*, SIAM J. Math. Anal., 54 (2022), pp. 1718–1741, https://doi.org/10.1137/21M1440165.

[3] D. Alvarez-Melis and T. S. Jaakkola, *Gromov-Wasserstein alignment of word embedding spaces*, in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018, pp. 1881–1890, https://doi.org/10.18653/v1/D18-1214.

[4] F. Bachoc, L. Béthune, A. González-Sanz, and J.-M. Loubes, *Gaussian processes on distributions based on regularized optimal transport*, in Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, PMLR 206:4986-5010, 2023.

[5] R. J. Berman, *The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations*, Numer. Math., 145 (2020), pp. 771–836, https://doi.org/10.1007/s00211-020-01127-x.

[6] J. Bigot, E. Cazelles, and N. Papadakis, *Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications*, Electron. J. Stat., 13 (2019), pp. 5120–5150, https://doi.org/10.1214/19-EJS1637.

[7] E. Black, S. Yeom, and M. Fredrikson, *Fliptest: Fairness testing via optimal transport,* in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 111–121.

[8] J. Blanchet, Y. Kang, and K. Murthy, *Robust Wasserstein profile inference and applications to machine learning*, J. Appl. Probab., 56 (2019), pp. 830–857, https://doi.org/10.1017/jpr.2019.49.

[9] V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry, *Monge-Kantorovich depth, quantiles, ranks and signs*, Ann. Statist., 45 (2017), pp. 223–256, https://doi.org/10.1214/16-AOS1450.

[10] S. Chiappa, R. Jiang, T. Stepleton, A. Pacchiano, H. Jiang, and J. Aslanides, *A general approach to fairness with optimal transport*, in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 3633–3640.

[11] S. Chiappa and A. Pacchiano, *Fairness with Continuous Optimal Transport*, preprint, arXiv:2101.02084, 2021.

[12] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard, and G. Peyré, *Faster Wasserstein distance estimation with the Sinkhorn divergence*, in Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020, 190.

[13] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, *Optimal transport for domain adaptation*, IEEE Trans. Pattern Anal. Mach. Intell., 39 (2017), pp. 1853–1865.

[14] I. Csiszar, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab., 3 (1975), pp. 146–158, https://doi.org/10.1214/aop/1176996454.

[15] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport,* in Advances in Neural Information Processing Systems 26 (NIPS 2013), 2013, pp. 2292–2300, http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

[16] L. de Lara, A. González-Sanz, N. Asher, and J.-M. Loubes, *Transport-Based Counterfactual Models*, preprint, https://arxiv.org/abs/2108.13025, 2021.

[17] N. Deb and B. Sen, *Multivariate rank-based distribution-free nonparametric testing using measure transportation*, J. Amer. Statist. Assoc., 118 (2023), pp. 192–207.

[18] E. del Barrio, E. Giné, and C. Matrán, *Central limit theorems for the Wasserstein distance between the empirical and the true distributions*, Ann. Probab., 27 (1999), pp. 1009–1071, https://doi.org/10.1214/aop/1022677394.

[19] E. del Barrio, E. Giné, and F. Utzet, *Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances*, Bernoulli, 11 (2005), pp. 131–189, https://doi.org/10.3150/bj/1110228245.

[20] E. del Barrio, A. González-Sanz, and J.-M. Loubes, *Central limit theorems for general transportation costs*, Ann. Inst. H. Poincaré, to appear; preprint, https://arxiv.org/abs/2102.06379, 2021.

[21] E. del Barrio, A. González-Sanz, and J.-M. Loubes, *Central limit theorems for semidiscrete Wasserstein distances*, Bernoulli, to appear; preprint, https://arxiv.org/abs/2105.11721, 2021.

[22] E. del Barrio, P. Gordaliza, and J.-M. Loubes, *A central limit theorem for L_p transportation cost on the real line with application to fairness assessment in machine learning*, Inf. Inference, 8 (2019), pp. 817–849.

[23] E. del Barrio and J.-M. Loubes, *Central limit theorems for empirical transportation cost in general dimension*, Ann. Probab., 47 (2019), pp. 926–951, https://doi.org/10.1214/18-AOP1275.

[24] E. del Barrio and J.-M. Loubes, *The Statistical Effect of Entropic Regularization in Optimal Transportation*, preprint, https://arxiv.org/abs/2006.05199, 2020.

[25] A. Delalande, *Nearly tight convergence bounds for semi-discrete entropic optimal transport,* in International Conference on Artificial Intelligence and Statistics, 2021.

[26] S. Eckstein and M. Nutz, *Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm*, preprint, https://arxiv.org/pdf/2110.06798.pdf, 2021.

[27] J. Feydy, T. Séjourné, F.-X. Vialard, S. Amari, A. Trouvé, and G. Peyré, *Interpolating between optimal transport and MMD using Sinkhorn divergences*, in Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, 2019, pp. 2681–2690.

[28] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer, *POT: Python Optimal Transport*, J. Mach. Learn. Res., 22 (2021), pp. 1–8, http://jmlr.org/papers/v22/20-451.html.

[29] R. Flamary, M. Cuturi, N. Courty, and A. Rakotomamonjy, *Wasserstein discriminant analysis*, Mach. Learn., 107 (2018), pp. 1923–1945, https://doi.org/10.1007/s10994-018-5717-1.

[30] R. Gao and A. J. Kleywegt, *Distributionally Robust Stochastic Optimization with Wasserstein Distance*, preprint, arXiv:1604.02199, 2016.

[31] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré, *Sample complexity of Sinkhorn divergences,* in Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, K. Chaudhuri and M. Sugiyama, eds., PMLR 89, 2019, pp. 1574–1583, http://proceedings.mlr.press/v89/genevay19a.html.

[32] A. Genevay, G. Peyré, and M. Cuturi, *Learning generative models with Sinkhorn divergences*, in Proceedings of the International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain, 2018, pp. 1608–1617, http://proceedings.mlr.press/v84/genevay18a.html.

[33] P. Ghosal, M. Nutz, and E. Bernton, *Stability of Entropic Optimal Transport and Schrödinger Bridges*, preprint, https://arxiv.org/pdf/2106.03670.pdf, 2022.

[34] P. Ghosal and B. Sen, *Multivariate Ranks and Quantiles Using Optimal Transport: Consistency, Rates, and Nonparametric Testing*, preprint, arXiv:1905.05340, 2019.

[35] E. Giné and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Camb. Ser. Stat. Probab. Math. 40, Cambridge University Press, New York, 2016.

[36] J. González-Delgado, A. González-Sanz, J. Cortés, and P. Neuvial, *Two-Sample Goodness-of-Fit Tests on the Flat Torus Based on Wasserstein Distance and Their Relevance to Structural Biology*, preprint, https://arxiv.org/abs/2108.00165, 2021,

[37] A. González-Sanz, J.-M. Loubes, and J. Niles-Weed, *Weak Limits of Entropy Regularized Optimal Transport; Potentials, Plans and Divergences*, preprint, https://arxiv.org/abs/2207.07427, 2022.

[38] P. Gordaliza, E. Del Barrio, G. Fabrice, and J.-M. Loubes, *Obtaining fairness using optimal transport theory*, in Proceedings of the 36th International Conference on Machine Learning, PMLR 97, 2019, pp. 2357–2365.

[39] E. Grave, A. Joulin, and Q. Berthet, *Unsupervised alignment of embeddings with Wasserstein Procrustes*, in Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, PMLR 89, 2019, pp. 1880–1890.

[40] M. Hallin, E. del Barrio, J. Cuesta-Albertos, and C. Matrán, *Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach*, Ann. Statist., 49 (2021), pp. 1139–1165, https://doi.org/10.1214/20-AOS1996.

[41] M. Hallin, G. Mordant, and J. Segers, *Multivariate goodness-of-fit tests based on Wasserstein distance*, Electron. J. Stat., 15 (2021), pp. 1328–1371, https://doi.org/10.1214/21-ejs1816.

[42] Z. Harchaoui, L. Liu, and S. Pal, *Asymptotics of Entropy-Regularized Optimal Transport via Chaos Decomposition*, 2020, https://doi.org/10.48550/arXiv.2011.08963.

[43] S. Hundrieser, M. Klatt, T. Staudt, and A. Munk, *A Unifying Approach to Distributional Limits for Empirical Optimal Transport*, preprint, https://arxiv.org/pdf/2202.12790.pdf, 2022.

[44] J.-C. Hütter and P. Rigollet, *Minimax estimation of smooth optimal transport maps*, Ann. Statist., 49 (2021), pp. 1166–1194, https://doi.org/10.1214/20-aos1997.

[45] H. Janati, B. Muzellec, G. Peyré, and M. Cuturi, *Entropic optimal transport between unbalanced gaussian measures has a closed form*, in Advances in Neural Information Processing Systems 33, 2020, pp. 10468–10479.

[46] M. Klatt, C. Tameling, and A. Munk, *Empirical regularized optimal transport: Statistical theory and applications*, SIAM J. Math. Data Sci., 2 (2020), pp. 419–443, https://doi.org/10.1137/19M1278788.

[47] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Metric space of collider events*, Phys. Rev. Lett., 123 (2019), 041801.
[48] D. Kuhn, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, *Wasserstein distributionally robust optimization: Theory and applications in machine learning*, INFORMS TutORials in Operations Research, 2019 (2019), pp. 130–166.
[49] M. Li, Y.-M. Zhai, Y.-W. Luo, P.-F. Ge, and C.-X. Ren, *Enhanced transport distance for unsupervised domain adaptation*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13936–13944.
[50] L. Liu, S. Pal, and Z. Harchaoui, *Entropy regularized optimal transport independence criterion*, in Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, PMLR 151, 2022.
[51] G. Luise, S. Salzo, M. Pontil, and C. Ciliberto, *Sinkhorn barycenters with free support via Frank-Wolfe algorithm*, in Advances in Neural Information processing Systems 32, 2019.
[52] A. Mallasto, A. Gerolin, and H. Q. Minh, *Entropy-regularized* 2*-Wasserstein distance between Gaussian measures*, Inf. Geom., 5 (2022), pp. 289–323.
[53] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman, *Plugin Estimation of Smooth Optimal Transport Maps*, preprint, https://arxiv.org/pdf/2107.12364.pdf, 2022.
[54] G. Mena and J. Niles-Weed, *Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem*, in Advances in Neural Information Processing Systems 32, Curran Associates, 2019, https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.
[55] S. L. Morgan and D. J. Harding, *Matching estimators of causal effects: Prospects and pitfalls in theory and practice*, Sociol. Method Res., 35 (2006), pp. 3–60.
[56] T. F. Móri and G. J. Székely, *The earth mover's correlation*, Ann. Univ. Sci. Budapest. Sect. Comput., 50 (2020), pp. 249–268, http://ac.inf.elte.hu/Vol_050_2020/249_50.pdf.
[57] M. Nutz, *Introduction to Entropic Optimal Transport*, Ph.D., Ecole Polytechnique, Columbia in Paris, 2021.
[58] M. Nutz and J. Wiesel, *Entropic optimal transport: Convergence of potentials*, Probab. Theory Related Fields, 184 (2022), pp. 401–424.
[59] M. Nutz and J. Wiesel, *Stability of Schrödinger Potentials and Convergence of Sinkhorn's Algorithm*, preprint, https://arxiv.org/pdf/2201.10059.pdf, 2022.
[60] G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Found. Trends Mach. Learn., 11 (2019), pp. 355–607, https://doi.org/10.1561/2200000073.
[61] A.-A. Pooladian and J. Niles-Weed, *Entropic Estimation of Optimal Transport Maps*, preprint, https://arxiv.org/pdf/2109.12004.pdf, 2022.
[62] I. Redko, A. Habrard, and M. Sebban, *Theoretical analysis of domain adaptation with optimal transport*, in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2017, pp. 737–753.
[63] P. Rigollet and A. Stromme, *On the Sample Complexity of Entropic Optimal Transport*, preprint, https://arxiv.org/abs/2206.13472, 2022.
[64] P. Rigollet and J. Weed, *Entropic optimal transport is maximum-likelihood deconvolution*, C. R. Math. Acad. Sci. Paris, 356 (2018), pp. 1228–1235, https://doi.org/10.1016/j.crma.2018.10.010.
[65] P. Rigollet and J. Weed, *Uncoupled isotonic regression via minimum Wasserstein deconvolution*, Inf. Inference, 8 (2019), pp. 691–717.
[66] L. Risser, A. González Sanz, Q. Vincenot, and J.-M. Loubes, *Tackling algorithmic bias in neural-network classifiers using Wasserstein-*2 *regularization*, J. Math. Imaging Vision, 64 (2022), pp. 672–689.
[67] M. C. Romão, N. Castro, J. Milhano, R. Pedro, and T. Vale, *Use of a Generalized Energy Mover's Distance in the Search for Rare Phenomena at Colliders*, preprint, arXiv:2004.09360, 2020.
[68] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.
[69] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al., *Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming*, Cell, 176 (2019), pp. 928–943.

[70] H. Shi, M. Drton, M. Hallin, and F. Han, *Center-Outward Sign- and Rank-Based Quadrant, Spearman, and Kendall Tests for Multivariate Independence*, https://arxiv.org/abs/2111.15567v1, 2021.
[71] H. Shi, M. Drton, and F. Han, *Distribution-free consistent independence tests via center-outward ranks and signs*, J. Amer. Statist. Assoc., 117 (2022), pp. 395–410, https://doi.org/10.1080/01621459.2020.1782223.
[72] H. Shi, M. Hallin, M. Drton, and F. Han, *On universally consistent and fully distribution-free rank tests of vector independence*, Ann. Statist., 50 (2022), pp. 1933–1959.
[73] M. Sommerfeld and A. Munk, *Inference for empirical Wasserstein distances on finite spaces*, J. R. Stat. Soc. Ser. B Stat. Methodol., 80 (2018), pp. 219–238.
[74] C. Tameling, M. Sommerfeld, and A. Munk, *Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications*, Ann. Appl. Probab., 29 (2019), pp. 2744–2781, https://doi.org/10.1214/19-AAP1463.
[75] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag, New York, 1996.
[76] P. E. Todd, *Matching estimators*, in Microeconometrics, Springer, 2010, pp. 108–121.
[77] A. van der Vaart, *Bracketing smooth functions*, Stochastic Process. Appl., 52 (1994), pp. 93–105, https://doi.org/10.1016/0304-4149(94)90102-3.
[78] J. Weed and F. Bach, *Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance*, Bernoulli, 25 (2019), pp. 2620–2648, https://doi.org/10.3150/18-BEJ1065.
[79] K. D. Yang, K. Damodaran, S. Venkatachalapathy, A. C. Soylemezoglu, G. Shivashankar, and C. Uhler, *Predicting cell lineages using autoencoders and optimal transport*, PLoS Comput. Biol., 16 (2020), e1007828.