Electron. Commun. Probab. **30** (2025), article no. 36, 1–13.
https://doi.org/10.1214/25-ECP682
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Approximation rates of entropic maps in semidiscrete optimal transport*

Ritwik Sadhu† Ziv Goldfeld‡ Kengo Kato§

Abstract

Entropic optimal transport offers a computationally tractable approximation to the classical problem. We study the approximation rate of the entropic optimal transport map (in approaching the Brenier map) when the regularization parameter ε tends to zero in the semidiscrete setting, where the input measure is absolutely continuous while the output is finitely discrete. Previous work shows that the approximation rate is $O(\sqrt{\varepsilon})$ under the L^2-norm with respect to the input measure. In this work, we establish faster, $O(\varepsilon^2)$ rates up to polylogarithmic factors, under the dual Lipschitz norm, which is weaker than the L^2-norm. For the said dual norm, the $O(\varepsilon^2)$ rate is sharp. As a corollary, we derive a central limit theorem for the entropic estimator for the Brenier map in the dual Lipschitz space when the regularization parameter tends to zero as the sample size increases.

Keywords: Brenier map; entropic map; entropic optimal transport; semidiscrete optimal transport.
MSC2020 subject classifications: 49Q22; 60F05; 49N15.
Submitted to ECP on November 24, 2024, final version accepted on March 30, 2025.

1 Introduction

1.1 Overview

For an absolutely continuous input distribution P and a generic output distribution Q, both on $\mathbb{R}^d$ with finite second moments, the *Brenier map* [7] sending P to Q induces the optimal coupling for the optimal transport problem with quadratic cost:

$$\inf_{\pi \in \Pi(P,Q)} \int \|x-y\|^2 \, d\pi(x,y), \tag{1.1}$$

where $\Pi(P,Q)$ denotes the collection of couplings of P and Q. The Brenier map can be characterized as a P-a.e. unique transport map given by the gradient of a convex function. This celebrated result has seen numerous applications in statistics and machine learning, ranging from transfer learning and domain adaptation to vector quantile regression and causal inference; see [12] for a review of the recent development in statistical optimal

*Z. Goldfeld is partially supported by NSF grants CCF-2046018, and DMS-2210368, and the IBM Academic Award. K. Kato is partially supported by NSF grants DMS-2210368 and DMS-2413405.
†University of Washington, United States of America. E-mail: rsadhu@uw.edu
‡Cornell University, United States of America. E-mail: goldfeld@cornell.edu
§Cornell University, United States of America. E-mail: kk976@cornell.edu

transport. From a mathematical standpoint, the Brenier map provides a powerful tool to derive functional inequalities [17] and suggests natural extensions of the quantile function to the multivariate setting [11], among others.

In practice, however, directly solving the optimal transport problem (1.1) and computing the Brenier map is challenging, especially when d is large. A popular remedy for this computational difficulty is entropic regularization, whereby (1.1) is replaced with

$$\inf_{\pi\in\Pi(P,Q)}\int \|x-y\|^2\,d\pi(x,y)+\varepsilon D_{\mathsf{KL}}(\pi\|P\otimes Q), \tag{1.2}$$

where $\varepsilon>0$ is the regularization parameter and D_{KL} is the Kullback-Leibler divergence defined by $D_{\mathsf{KL}}(\alpha\|\beta):=\int\log\frac{d\alpha}{d\beta}\,d\alpha$ if $\alpha\ll\beta$ and $:=\infty$ otherwise. Entropic optimal transport is amenable to efficient computation via Sinkhorn's algorithm, for which rigorous convergence guarantees have been developed under different settings [25, 18, 1, 40, 3, 8, 23, 26, 36, 16, 14]. As ε shrinks, various objects from entropic optimal transport converge to those for unregularized optimal transport—a topic that has seen extensive research activities in recent years; see the literature review below.

Denoting by π^ε the (unique) optimal coupling for the entropic problem (2.3), an entropic surrogate of the Brenier map is given by $T^\varepsilon(x)=\mathbb{E}_{(X,Y)\sim\pi^\varepsilon}[Y\mid X=x]$, which we shall call the *entropic map*[42]. To understand the quality of this computationally tractable approximation, the rate at which the entropic map approaches the Brenier map as $\varepsilon\downarrow 0$ has received recent attention. [9] showed that if P and Q are compactly supported and the Brenier map T^0 is M-Lipschitz (which precludes Q being discrete), then $\|T^\varepsilon-T^0\|^2_{L^2(P)}\le M(d\varepsilon\log(1/\varepsilon)+O(\varepsilon))$. In the continuous-to-continuous setting, imposing stronger smoothness conditions on the densities of P and Q and the dual potentials, [42] established faster $O(\varepsilon^2)$ rates for $\|T^\varepsilon-T^0\|^2_{L^2(P)}$. In the semidiscrete setting (i.e., when P is absolutely continuous and Q is finitely discrete), [43] showed that

$$\|T^\varepsilon-T^0\|^2_{L^2(P)}=O(\varepsilon), \tag{1.3}$$

and their Example 3.5 demonstrates that this rate is sharp under $L^2(P)$. The follow-up work by the same authors [22] derived quantitative upper bounds on the $L^2(P)$ error.

The goal of this paper is to explore quantitative upper bounds on the bias of T^ε for small ε in the semidiscrete setting, but from a different angle. Instead of the L^2-norm, we shall look at the linear functional $\langle\varphi,T^\varepsilon\rangle_{L^2(P)}$ for a suitable Borel vector field φ and derive quantitative upper bounds on $\langle\varphi,T^\varepsilon-T^0\rangle_{L^2(P)}$. There are several applications where a linear functional of the Brenier map is an object of interest (cf. [5, 49, 30][1]). In particular, while pointwise inference is essentially infeasible for the semidiscrete Brenier map [45], a local average of the Brenier map via a suitable kernel can provide a meaningful object for inference. The preceding bound (1.3) directly implies that, for any bounded Borel vector field φ,

$$|\langle\varphi,T^\varepsilon-T^0\rangle_{L^2(P)}|\le\|\varphi\|_\infty\|T^\varepsilon-T^0\|_{L^2(P)}=O(\sqrt{\varepsilon}). \tag{1.4}$$

Perhaps somewhat surprisingly, we show that this rate can be *much* faster for smooth test functions. Indeed, our main result shows that, if P is supported on a compact convex set and has a positive Lipschitz density on the support, then for any α-Hölder vector field φ with $\alpha\in(0,1]$,

$$|\langle\varphi,T^\varepsilon-T^0\rangle_{L^2(P)}|=O(\varepsilon^{1+\alpha}\vee\varepsilon^2\log^3(1/\varepsilon)).$$

In particular, this implies near $O(\varepsilon^2)$ approximation rates for Lipschitz test functions. The hidden constant depends on φ only through its α-Hölder norm, so by taking the supremum

[1] Those references concern estimation of a linear functional of the univariate quantile function, which naturally extends to the multivariate case by viewing the Brenier map as a multivariate quantile function.

over φ whose α-Hölder norm is at most 1, the same rate holds for $\|T^\varepsilon - T^0\|_{(\mathcal{C}^\alpha)^*}$, where $\|\cdot\|_{(\mathcal{C}^\alpha)^*}$ is the dual norm. This fast convergence rate under the dual norm is in line with the (sharp) approximation rate of ε^2 for the semidiscrete optimal transportation cost itself [2]. Finally, building on our recent work [45], we derive a central limit theorem in the dual space $(\mathcal{C}^\alpha)^*$ for the empirical entropic map with vanishing regularization parameters.

1.2 Literature review

There is now a large literature on convergence and approximation rates of entropic optimal transport costs, potentials, couplings, and maps when the regularization parameter tends to zero [33, 34, 31, 10, 13, 42, 15, 2, 38, 4, 20, 9, 43, 39, 22]. Among others, [2] derived an asymptotic expansion of the entropic cost in the semidiscrete case when the regularization parameter tends to zero, showing faster convergence at the rate ε^2 than the continuous-to-continuous case. See also the follow-up work by [20].

There is also a growing interest in estimation and inference for the Brenier map and its entropic variant [11, 29, 27, 42, 26, 41, 21, 45, 43, 32, 19, 44, 28]. Among them, [42] proposed using the entropic map with vanishing regularization parameters to estimate the Brenier map, and established convergence rates under the $L^2(P)$-norm in the continuous-to-continuous setting. However, these rate are suboptimal from a minimax point of view [29]. For the semidiscrete setting, [43] established the $O(n^{-1/2})$ rate for the entropic estimator with vanishing regularization levels $\varepsilon = \varepsilon_n = O(n^{-1/2})$ under the squared $L^2(P)$-norm. Our recent work [45] derived various limiting distribution results for certain functionals of the empirical (unregularized) Brenier map, when the input P is known but the discrete output Q is unknown.

1.3 Organization and notation

The rest of the note is organized as follows. Section 2 contains background material on the optimal transport problem and its entropic counterpart. Section 3 presents our main results. All the proofs are gathered in Section 4.

For $a, b \in \mathbb{R}$, we use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ to denote the Euclidean norm and inner product, respectively. Let $\mathbf{1}_N \in \mathbb{R}^N$ denote the vector of ones. For $d \in \mathbb{N}$ and $0 \le r \le d$, $\mathcal{H}^r$ denotes the r-dimensional Hausdorff measure on $\mathbb{R}^d$; cf. [24].

2 Background

2.1 Optimal transport

Let P and Q be Borel probability measures on $\mathbb{R}^d$ with finite second moments, and write $\mathcal{X}$ and $\mathcal{Y}$ for their respective supports. Recall the quadratic optimal transport problem (1.1), which, upon expanding the square, is equivalent to

$$\sup_{\pi \in \Pi(P,Q)} \int \langle x, y \rangle \, d\pi(x, y). \tag{2.1}$$

The Brenier theorem [7] yields that whenever P is absolutely continuous, the problem (2.1) admits a unique optimal solution π^0, which is induced by a P-a.e. unique map $T^0 : \mathcal{X} \to \mathbb{R}^d$, in the sense that $\pi^0 = P \circ (\mathrm{id}, T^0)^{-1}$ with id denoting the identity map. We call T^0 the *Brenier map*. See, e.g., [48, 46] for background of optimal transport.

We focus herein on the semidiscrete setting, where P is absolutely continuous while Q is finitely discrete with support $\mathcal{Y} = \{y_1, \dots, y_N\}$. Let $q = (q_1, \dots, q_N)^\intercal$ be the vector of masses with $q_i = Q(\{y_i\})$ for $i \in [N] := \{1, \dots, N\}$. In this case, the (semi)dual problem for (2.1) reads as

$$\inf_{z\in\mathbb{R}^N}\int \max_{1\le i\le N}(\langle x,y_i\rangle - z_i)\,dP(x) + \langle z,q\rangle. \tag{2.2}$$

Given any $z^0=(z^0_1,\dots,z^0_N)^\intercal$ optimal solution to (2.2), the Brenier map is then given by

$$T^0(x)=\nabla_x\Big(\max_{1\le i\le N}(\langle x,y_i\rangle - z^0_i)\Big),\quad P\text{-a.e. } x.$$

To simplify its description, for $z\in\mathbb{R}^N$, define the *Laguerre cells* $\{C_i(z)\}_{i=1}^N$,

$$C_i(z):=\bigcap_{j\neq i;1\le j\le N}\big\{z\in\mathcal{X}:\langle y_i-y_j,x\rangle\ge z_i-z_j\big\},$$

using which the Brenier map is given by $T^0(x)=y_i$ for $x\in C_i(z^0)$ and $i\in[N]$. The Laguerre cells form a partition of $\mathcal{X}$ up to Lebesgue negligible sets, so the preceding description specifies a P-a.e. defined map with values in $\mathcal{Y}$. Furthermore, as T^0 is a transport map, we have $P(C_i(z^0))=Q(\{y_i\})=q_i>0$ for $i\in[N]$.

The dual vector z^0 is not unique as adding the same constant to all z_i does not change the value of the objective in (2.2). So, we always normalize z^0 in such a way that $\langle z^0,\mathbf{1}_N\rangle=0$, which, together with mild conditions on P, guarantees uniqueness of z^0.

2.2 Entropic optimal transport

The entropic optimal transport problem corresponding to (2.1) is

$$\sup_{\pi\in\Pi(P,Q)}\int\langle x,y\rangle\,d\pi(x,y)-\varepsilon D_{\mathsf{KL}}(\pi\|P\otimes Q), \tag{2.3}$$

where $\varepsilon>0$ is the regularization parameter. For any P and Q with finite second moments (i.e., beyond the semidiscrete setting), the problem (2.3) admits a unique optimal solution π^ε, which is of the form

$$\frac{d\pi^\varepsilon}{d(P\otimes Q)}(x,y)=e^{\frac{\langle x,y\rangle-\phi^\varepsilon(x)-\psi^\varepsilon(y)}{\varepsilon}},$$

where $(\phi^\varepsilon,\psi^\varepsilon)$ is any optimal solution to the dual problem[2]

$$\inf_{(\phi,\psi)\in L^1(P)\times L^1(Q)}\int\phi\,dP+\int\psi\,dQ+\varepsilon\iint e^{\frac{\langle x,y\rangle-\phi(x)-\psi(y)}{\varepsilon}}\,dP(x)dQ(y).$$

Here, since π^ε is a coupling, one has $\int e^{\frac{\langle x,y\rangle-\phi^\varepsilon(x)-\psi^\varepsilon(y)}{\varepsilon}}dQ(y)=1$, that is, $\phi^\varepsilon(x)=\varepsilon\log\int e^{(\langle x,y\rangle-\psi^\varepsilon(y))/\varepsilon}\,dQ(y)$ for P-a.e. x. Substituting this expression leads to the semidual problem. See [37] for a comprehensive overview of entropic optimal transport. An entropic counterpart of the Brenier map was proposed in [42] by observing that $T^0(x)=\mathbb{E}_{(X,Y)\sim\pi^0}[Y\mid X=x]$, i.e., the Brenier map agrees with the conditional expectation of the second coordinate given the first under π^0. Replacing π^0 with π^ε leads to the *entropic map* $T^\varepsilon(x)=\mathbb{E}_{(X,Y)\sim\pi^\varepsilon}[Y\mid X=x]$ for $x\in\mathcal{X}$.

Specializing to the semidiscrete setting where Q has support $\mathcal{Y}=\{y_1,\dots,y_N\}$, one may reduce the semidiscrete problem to

$$\inf_{z\in\mathbb{R}^N}\int\left\{\varepsilon\log\sum_{i=1}^N q_ie^{(\langle\cdot,y_i\rangle-z_i)/\varepsilon}\right\}dP+\langle z,q\rangle.$$

[2] Pairs of optimal potentials are a.e. unique up to additive constants, i.e., if $(\tilde\phi,\tilde\psi)$ is another optimal pair then $\tilde\phi=\phi+c$ P-a.e. and $\tilde\psi=\psi-c$ Q-a.e., for some $c\in\mathbb{R}$.

Replacing z_i with $z_i + \varepsilon \log q_i$, the above semidual problem is equivalent to

$$\inf_{z\in\mathbb{R}^N} \int \left\{ \varepsilon \log \sum_{i=1}^N e^{(\langle \cdot, y_i\rangle - z_i)/\varepsilon} \right\} dP + \langle z, q\rangle. \tag{2.4}$$

The latter semidual problem (2.4) admits a unique optimal solution z^ε subject to the normalization $\langle z^\varepsilon, \mathbf{1}_N\rangle = 0$. In this case, the entropic map simplifies to

$$T^\varepsilon(x) = \sum_{i=1}^N y_i \frac{e^{(\langle x, y_i\rangle - z_i^\varepsilon)/\varepsilon}}{\sum_{j=1}^N e^{(\langle x, y_j\rangle - z_j^\varepsilon)/\varepsilon}}, \quad x \in \mathcal{X}.$$

3 Main results

We derive approximation rates of the entropic map T^ε towards the Brenier map T^0 as $\varepsilon \downarrow 0$. In contrast to [43, 22] that focus on the (squared) $L^2(P)$-norm $\|T^\varepsilon - T^0\|^2_{L^2(P)}$, we consider the linear functional $\langle \varphi, T^\varepsilon - T^0\rangle_{L^2(P)} = \int \langle \varphi(x), T^\varepsilon(x) - T^0(x)\rangle \, dP(x)$, for a suitable Borel vector field $\varphi : \mathcal{X} \to \mathbb{R}^d$, and establish the rates. Taking the supremum over a certain function class leads to the convergence rates under the corresponding dual norm. We start from the assumption under which the results hold.

Assumption 3.1 (Conditions on marginals)**.** *(i) The input measure P is supported on a compact convex set $\mathcal{X} \subset \mathbb{R}^d$ with nonempty interior and has a Lebesgue density ρ that is Lipschitz continuous and strictly positive on $\mathcal{X}$. (ii) The output measure Q is finitely discrete with support $\mathcal{Y} = \{y_1, \dots, y_N\} \subset \mathbb{R}^d$. For $q = (q_1, \dots, q_N)^\intercal$ with $q_i = Q(\{y_i\})$, we assume that $\min_{1\le i\le N} q_i \ge c_0$ for some (sufficiently small) constant $c_0 \in (0,1)$.*

Condition (i) guarantees uniqueness of the dual vector z^0 (subject to the normalization $\langle z^0, \mathbf{1}_N\rangle = 0$); cf. Theorem 7.18 in [46]. For a vector-valued mapping $\varphi : \mathcal{X} \to \mathbb{R}^d$ and $\alpha \in (0,1]$, the α-Hölder norm $\|\varphi\|_{\mathcal{C}^\alpha}$ (Lipschitz norm when $\alpha = 1$) is defined by

$$\|\varphi\|_{\mathcal{C}^\alpha} := \|\varphi\|_\infty + \sup_{x,y\in\mathcal{X}; x\ne y} \|\varphi(x) - \varphi(y)\| / \|x - y\|^\alpha,$$

where $\|\varphi\|_\infty = \sup_{x\in\mathcal{X}} \|\varphi(x)\|$. The following is our main result.

Theorem 3.2 (Convergence rates for Hölder test functions)**.** *Fix $\alpha \in (0,1]$. Under Assumption 3.1, for every α-Hölder vector field $\varphi : \mathcal{X} \to \mathbb{R}^d$,*

$$|\langle \varphi, T^\varepsilon - T^0\rangle_{L^2(P)}| \lesssim \|\varphi\|_\infty \varepsilon^2 \log^3(1/\varepsilon) + \|\varphi\|_{\mathcal{C}^\alpha} \varepsilon^{1+\alpha}, \quad \forall \varepsilon \in (0, e^{-1}),$$

where the inequality $\lesssim$ holds up to a constant that depends only on $\alpha, \mathcal{X}, \rho, \mathcal{Y}$, and c_0.

Remark 3.3 (Bounded test functions)**.** Inspection of the proof shows that if the test function φ is only (measurable and) bounded, then $|\langle \varphi, T^\varepsilon - T^0\rangle_{L^2(P)}| \lesssim \|\varphi\|_\infty \varepsilon$ for $\varepsilon \in (0,1)$, where the hidden constant depends only on $\mathcal{X}, \rho, \mathcal{Y}$, and c_0.

Theorem 3.2 implies that the (right) derivative of the mapping $\varepsilon \mapsto \langle \varphi, T^\varepsilon\rangle_{L^2(P)}$ at $\varepsilon = 0$ vanishes for any Hölder vector field φ. Indeed, the proof of the theorem shows that the right derivative at $\varepsilon = 0$ agrees with

$$\sum_{i\ne j} \frac{\log 2}{\|y_i - y_j\|} \int_{C_i(z^0)\cap C_j(z^0)} \langle y_j - y_i, \varphi(x)\rangle \rho(x) \, d\mathcal{H}^{d-1}(x) = 0.$$

Hence, we need to look at a higher-order expansion of the mapping $\varepsilon \mapsto \langle \varphi, T^\varepsilon\rangle_{L^2(P)}$ around $\varepsilon = 0$, which requires careful analysis of the facial structures of the Laguerre cells. In particular, special care is needed when $y_i - y_j$ and $y_i - y_k$ for some distinct indices i, j, k are linearly dependent; see, e.g., the proof of Lemma 4.1 ahead. The proof of Theorem 3.2 is inspired by the proofs in [2, 20] for the asymptotic expansions of the entropic cost, but differs from them in some important ways, as detailed in Remark 4.3.

Remark 3.4 (Sharpness of $O(\varepsilon^2)$ rate when $\alpha = 1$)**.** Theorem 1.1 in [2] establihses the asymptotic expansion of $\mathbb{E}_{(X,Y)\sim\pi^\varepsilon}[\|X-Y\|^2]$, which, after rearranging terms, implies

$$\langle \mathrm{id}, T^\varepsilon - T^0\rangle_{L^2(P)} = -\frac{\varepsilon^2\pi^2}{24}\sum_{i<j}\frac{1}{\|y_i-y_j\|}\int_{C_i(z^0)\cap C_j(z^0)}\rho(x)\,d\mathcal{H}^{d-1}(x) + o(\varepsilon^2).$$

Since the identity mapping id is Lipschitz, the rate in Theorem 3.2 is sharp up to the $\log^3(1/\varepsilon)$ factor. The question of whether the polylogarithmic factor can be dropped for a generic Lipschitz vector is left for future research.

Remark 3.5 (Sharpness of $O(\varepsilon^{\alpha+1})$ rate in $d=1$)**.** As in [2, 43], consider $d=1, P = \mathrm{Unif}([-1,1])$, and $Q = \frac{1}{2}(\delta_{-1}+\delta_1)$, for which the entropic map is $T^\varepsilon(x) = \tanh(2x/\varepsilon)$ and the Brenier map is $T^0(x) = \mathrm{sign}(x)$. For $\varphi(x) = \mathrm{sign}(x)|x|^\alpha$ with $\alpha \in (0,1]$, which is α-Hölder on $[-1,1]$, one can verify from the dominated convergence theorem that $\lim_{\varepsilon\downarrow 0}\varepsilon^{-1-\alpha}\langle\varphi, T^\varepsilon - T^0\rangle_{L^2(P)} = \int_0^\infty x^\alpha(\tanh(2x)-1)\,dx$, where the integral on the right-hand side is absolutely convergent. Hence, the $O(\varepsilon^{1+\alpha})$ rate in Theorem 3.2 is in general sharp for $\alpha \in (0,1)$.

Let $\mathcal{C}^\alpha = \mathcal{C}^\alpha(\mathcal{X};\mathbb{R}^d)$ be the Banach space of α-Hölder mappings $\mathcal{X}\to\mathbb{R}^d$ endowed with the norm $\|\cdot\|_{\mathcal{C}^\alpha}$. The topological dual $(\mathcal{C}^\alpha)^*$ is the Banach space of continuous linear functionals on $\mathcal{C}^\alpha$ endowed with the dual norm, $\|\ell\|_{(\mathcal{C}^\alpha)^*} = \sup_{\varphi:\|\varphi\|_{\mathcal{C}^\alpha}\le 1}\ell(\varphi)$. One may think of any bounded measurable mapping $T:\mathcal{X}\to\mathbb{R}^d$ as an element of the dual space $(\mathcal{C}^\alpha)^*$ by identifying T with the linear functional $\varphi\mapsto\langle\varphi,T\rangle_{L^2(P)}$. With this identification, the preceding theorem yields rates of convergence of the entropic map under $\|\cdot\|_{(\mathcal{C}^\alpha)^*}$.

Corollary 3.6 (Convergence rates under dual Hölder norm)**.** *Fix* $\alpha \in (0,1]$. *Under Assumption 3.1,* $\|T^\varepsilon - T^0\|_{(\mathcal{C}^\alpha)^*} \lesssim \varepsilon^{1+\alpha}\vee\varepsilon^2\log^3(1/\varepsilon)$ *for all* $\varepsilon\in(0,e^{-1})$*, where the inequality* $\lesssim$ *holds up to a constant that depends only on* $\alpha,\mathcal{X},\rho,\mathcal{Y}$*, and* c_0*.*

We discuss a statistical application of the preceding result. Suppose the input measure P is known but the output Q is unknown, and we have access to an i.i.d. sample $Y_1,\dots,Y_n$ from Q. Such a setting is natural when we think of the Brenier map as a multivariate quantile function, where P serves as a reference measure (cf. [11]). Let $\hat{Q}_n = n^{-1}\sum_{i=1}^n\delta_{Y_i}$ denote the empirical distribution, which is supported in $\mathcal{Y}$. In addition, let $\hat{T}_n^0$ and $\hat{T}_n^\varepsilon$ with $\varepsilon>0$ be the Brenier and entropic maps, respectively, for the pair $(P,\hat{Q}_n)$. Our recent work [45] established a central limit theorem for $\hat{T}_n^0$ in $(\mathcal{C}^\alpha)^*$,

$$\sqrt{n}(\hat{T}_n^0 - T^0)\xrightarrow{d}\mathbb{G}\quad\text{in } (\mathcal{C}^\alpha)^*,\quad\text{as } n\to\infty, \tag{3.1}$$

where $\xrightarrow{d}$ signifies convergence in distribution and $\mathbb{G}$ is a centered Gaussian variable in $(\mathcal{C}^\alpha)^*$ (the exact form of $\mathbb{G}$ can be found in Theorem 4 in [45]). The next result shows that the same weak limit holds for the entropic estimator with $\varepsilon=\varepsilon_n\downarrow 0$ sufficiently fast.

Corollary 3.7 (Central limit theorem under dual Hölder space)**.** *Suppose Assumption 3.1 holds and in addition that one of the following holds for* $\mathcal{X}$*: (a)* $\mathcal{X}$ *is a polytope, or (b)* $\mathcal{H}^{d-1}(\partial\mathcal{X}\cap H)=0$ *for every hyperplane* H *in* $\mathbb{R}^d$*. Then,* $\sqrt{n}(\hat{T}_n^{\varepsilon_n}-T^0)\xrightarrow{d}\mathbb{G}$ *in* $(\mathcal{C}^\alpha)^*$*, provided that* $\varepsilon_n = o\big(n^{-\frac{1}{2(1+\alpha)}}\wedge n^{-1/4}/\log^{3/2}n\big)$*, where* $\mathbb{G}$ *is the same centered Gaussian variable in* $(\mathcal{C}^\alpha)^*$ *as that in* (3.1)*.*

Remark 3.8 (Comparison with [43])**.** [43] showed that $\mathbb{E}[\|\hat{T}_n^{\varepsilon_n}-T^{\varepsilon_n}\|^2_{L^2(P)}] = O(\varepsilon_n^{-1}n^{-1})$. Combining the bias estimate in (1.3), they established $\mathbb{E}[\|\hat{T}_n^{\varepsilon_n}-T^0\|^2_{L^2(P)}] = O(n^{-1/2})$ by choosing ε_n decaying at the rate $n^{-1/2}$. It is interesting to observe that, under the dual norm $\|\cdot\|_{(\mathcal{C}^\alpha)^*}$, the empirical entropic map enjoys the parametric rate with ε_n decaying substantially slower than $n^{-1/2}$.

4 Proofs

4.1 Preliminaries

Define

$$\Delta_{ij}^{\varepsilon}(x) \coloneqq \langle y_i - y_j, x\rangle - z_i^{\varepsilon} + z_j^{\varepsilon}, \quad \varepsilon \geq 0.$$

Observe that $C_i(z^0) = \{x \in \mathcal{X} : \Delta_{ij}^0 \geq 0, \forall j \neq i\}$ and $T^{\varepsilon}(x) = \sum_{j=1}^N y_j \frac{e^{-\Delta_{ij}^{\varepsilon}(x)/\varepsilon}}{\sum_{k=1}^N e^{-\Delta_{ik}^{\varepsilon}(x)/\varepsilon}}$ for any $i \in [N]$ and $x \in \mathcal{X}$. Furthermore, define

$$H_{ij}(t) \coloneqq \{x \in C_i(z^0) : \Delta_{ij}^0(x) = t\}.$$

Observe that $H_{ij}(0) = H_{ji}(0) = C_i(z^0) \cap C_j(z^0)$. For notational convenience, set $M_\rho = \sup_{x\in\mathcal{X}} \rho(x) < \infty$ and $\delta = \min_{i\neq j} \|y_i - y_j\| > 0$. In what follows, the notation $\lesssim$ means that the left-hand side is upper bounded by the right-hand side up to a constant that depends only on $\alpha, \mathcal{X}, \rho, \mathcal{Y}$, and c_0. We first establish the following preliminary estimates.

Lemma 4.1. *Under Assumption 3.1, the following hold. (i). For any distinct indices i, j, one has $\int_{C_i(z^0)} e^{-\Delta_{ij}^0(x)/\varepsilon}\rho(x)\,dx \leq \frac{\varepsilon M_\rho (\operatorname{diam}\mathcal{X})^{d-1}}{\|y_i - y_j\|}$. (ii). For any distinct indices i, j, k, one has $\int_{C_i(z^0)} e^{-\Delta_{ij}^0(x)/\varepsilon} e^{-\Delta_{ik}^0(x)/\varepsilon}\,\rho(x)\,dx \lesssim \varepsilon^2 \log^2(1/\varepsilon)$ for $\varepsilon > 0$.*

Proof of Lemma 4.1. (i). By the coarea formula [24, Theorem 3.11],

$$\int_{C_i(z^0)} e^{-\Delta_{ij}^0(x)/\varepsilon}\rho(x)\,dx = \frac{1}{\|y_i - y_j\|}\int_0^\infty \left(\int_{H_{ij}(t)} \rho(x)\,d\mathcal{H}^{d-1}(x)\right) e^{-t/\varepsilon} dt. \tag{4.1}$$

The inner integral can be bounded by $M_\rho \mathcal{H}^{d-1}\big(H_{ij}(t)\big) \leq M_\rho (\operatorname{diam}\mathcal{X})^{d-1}$, as $H_{ij}(t)$ is a hyperplane section of $\mathcal{X}$, which implies that the right-hand side on (4.1) can be bounded by $\varepsilon\|y_i - y_j\|^{-1} M_\rho (\operatorname{diam}\mathcal{X})^{d-1}$.

(ii). Fix $\eta > 0$. Set $A_{i\ell}(\eta) \coloneqq \{x \in C_i(z^0) : \Delta_{i\ell}^0(x) \geq \eta\}$ and $B_{i\ell}(\eta) \coloneqq \{x \in \mathcal{X} : 0 \leq \Delta_{i\ell}^0(x) < \eta\}$, for $\ell = j, k$. Then, applying the coarea formula, the integral $\int_{C_i(z^0)} e^{-\Delta_{ij}^0(x)/\varepsilon} e^{-\Delta_{ik}^0(x)/\varepsilon}\rho(x)\,dx$ can be bounded by

$$\begin{aligned}
&\left(\int_{A_{ij}(\eta)} + \int_{A_{ik}(\eta)} + \int_{C_i(z^0)\cap A_{ij}(\eta)^c \cap A_{ik}(\eta)^c}\right) e^{-\Delta_{ij}^0(x)/\varepsilon} e^{-\Delta_{ik}^0(x)/\varepsilon}\rho(x)\,dx \\
&\leq \delta^{-1} e^{-\eta/\varepsilon}\int_0^\infty \left\{\left(\int_{H_{ij}(t)} + \int_{H_{ik}(t)}\right)\rho(x)\,d\mathcal{H}^{d-1}(x)\right\} e^{-t/\varepsilon}\,dt + M_\rho \mathcal{H}^d(B_{ij}(\eta)\cap B_{ik}(\eta)) \\
&\leq 2\delta^{-1}\varepsilon e^{-\eta/\varepsilon}(\operatorname{diam}\mathcal{X})^{d-1} M_\rho + M_\rho \mathcal{H}^d(B_{ij}(\eta)\cap B_{ik}(\eta)).
\end{aligned}$$

For the second term on the right-hand side, we separately consider the following two cases.

Case (a). Suppose that $y_i - y_j$ and $y_i - y_k$ are linearly independent. In this case,

$$\mathcal{H}^d(B_{ij}(\eta)\cap B_{ik}(\eta)) \leq (\operatorname{diam}\mathcal{X})^{d-2}\frac{\eta^2}{\sqrt{\|y_i - y_j\|^2\|y_i - y_k\|^2 - \langle y_i - y_j, y_i - y_k\rangle^2}}.$$

Case (b). Suppose that $y_i - y_j$ and $y_i - y_k$ are linearly dependent, so that $y_i - y_k = c(y_i - y_j)$ for some $c \neq 0$. We will show that there exists $\eta_0 > 0$ that depends only on $\mathcal{X}, \rho, \mathcal{Y}$, and c_0 such that $B_{ij}(\eta)\cap B_{ik}(\eta) = \varnothing$ for all $\eta \in (0, \eta_0)$. We only consider the $c < 0$ case. The $c > 0$ case is similar (see Step 1 of the proof of Theorem 1 (i) in [45] for a similar argument). Suppose $B_{ij}(\eta)\cap B_{ik}(\eta) \neq \varnothing$, which entails that there exists some $x \in \mathcal{X}$ such that

$$0 \leq \langle y_i - y_j, x\rangle - b_{ij} < \eta \quad \text{and} \quad 0 \leq \langle y_i - y_k, x\rangle - b_{ik} < \eta, \tag{4.2}$$

where $b_{ij} = z_i^0 - z_j^0$. Let L_1 and L_2 be the hyperplanes defined by $L_1 = \{x : \langle y_i - y_j, x\rangle = b_{ij}\}$ and $L_2 = \{x : \langle y_i - y_k, x\rangle = b_{ik}\}$, which are parallel as $y_i - y_j$ and $y_j - y_k$ are linearly dependent. As such, $\operatorname{dist}(L_1, L_2) = \frac{|b_{ij} - c^{-1} b_{ik}|}{\|y_i - y_j\|}$. On the other hand, by our choice of x from (4.2),

$$\operatorname{dist}(L_1, L_2) \le \operatorname{dist}(x, L_1) + \operatorname{dist}(x, L_2) \le \frac{\eta}{\|y_i - y_j\|} + \frac{\eta}{\|y_i - y_k\|} = \frac{\eta(1 + |c|^{-1})}{\|y_i - y_j\|},$$

so that $|b_{ij} - c^{-1} b_{ik}| \le \eta(1 + |c|^{-1})$. Observe that

$$\begin{aligned} C_i(z^0) &\subset \{x : \langle y_i - y_j, x\rangle \ge b_{ij}\} \cap \{x : \langle y_i - y_k, x\rangle \ge b_{ik}\} \\ &\subset \{x : \langle y_i - y_j, x\rangle \ge b_{ij}\} \cap \{x : \langle y_i - y_j, x\rangle \le b_{ij} + \eta(1 + |c|^{-1})\}, \end{aligned}$$

which implies that

$$q_i = P(C_i(z^0)) \le M_\rho (\operatorname{diam} \mathcal{X})^{d-1} \frac{\eta(1 + |c|^{-1})}{\|y_i - y_j\|}.$$

Hence, if we choose

$$\eta_0 = \frac{\delta c_0}{2(1 + |c|^{-1}) M_\rho (\operatorname{diam} \mathcal{X})^{d-1}},$$

then $q_i < c_0 \le \min_\ell q_\ell$ for $\eta < \eta_0$, which is a contradiction. Conclude that $B_{ij}(\eta) \cap B_{ik}(\eta) = \varnothing$ for $\eta < \eta_0$.

Finally, by choosing $\eta = \varepsilon \log(1/\varepsilon)$, we see that the desired estimate holds for all $\varepsilon \in (0, \varepsilon_0)$ for some $\varepsilon_0 > 0$ that depends only on $\mathcal{X}, \rho, \mathcal{Y}$, and c_0. For $\varepsilon \ge \varepsilon_0$, one may use the crude upper bound $\int_{C_i(z^0)} e^{-\Delta_{ij}^0(x)/\varepsilon} e^{-\Delta_{ik}^0(x)/\varepsilon} \rho(x)\, dx \le \int_{C_i(z^0)} \rho(x)\, dx \le 1$, and adjust the constant hidden in $\lesssim$. □

4.2 Proof of Theorem 3.2

The proof is divided into two steps.

Step 1. We first establish that

$$|\langle \varphi, T^\varepsilon - T^0\rangle_{L^2(P)}| \lesssim \|\varphi\|_\infty \big(\|z^\varepsilon - z^0\|_\infty e^{2\|z^\varepsilon - z^0\|_\infty/\varepsilon} + \varepsilon^2 \log^2(1/\varepsilon)\big) + \|\varphi\|_{\mathcal{C}^\alpha} \varepsilon^{1+\alpha}. \tag{4.3}$$

Since $\{C_i(z^0)\}_{i=1}^N$ forms a partition of $\mathcal{X}$ up to Lebesgue negligible sets, one has

$$\langle \varphi, T^\varepsilon\rangle_{L^2(P)} = \sum_{i=1}^N \sum_{j=1}^N \int_{C_i(z^0)} \langle y_j, \varphi(x)\rangle \frac{e^{-\Delta_{ij}^\varepsilon(x)/\varepsilon}}{\sum_{k=1}^N e^{-\Delta_{ik}^\varepsilon(x)/\varepsilon}} \rho(x)\, dx.$$

On the other hand, $\langle \varphi, T^0\rangle_{L^2(P)} = \sum_{i=1}^N \sum_{j=1}^N \int_{C_i(z^0)} \langle y_i, \varphi(x)\rangle \frac{e^{-\Delta_{ij}^\varepsilon(x)/\varepsilon}}{\sum_{k=1}^N e^{-\Delta_{ik}^\varepsilon(x)/\varepsilon}} \rho(x)\, dx$. Subtracting these expressions leads to

$$\langle \varphi, T^\varepsilon - T^0\rangle_{L^2(P)} = \sum_{i \ne j} \int_{C_i(z^0)} \langle y_j - y_i, \varphi(x)\rangle \frac{e^{-\Delta_{ij}^\varepsilon(x)/\varepsilon}}{1 + \sum_{k \ne j} e^{-\Delta_{ik}^\varepsilon(x)/\varepsilon}} \rho(x)\, dx. \tag{4.4}$$

We will replace Δ_{ij}^ε with Δ_{ij}^0 on the right-hand side.

Noting that $e^{-\Delta_{ij}^\varepsilon/\varepsilon} = e^{(z_i^\varepsilon - z_i^0 - z_j^\varepsilon + z_j^0)/\varepsilon} e^{-\Delta_{ij}^0/\varepsilon}$ and $\Delta_{ik}^0 \ge 0$ for $k \ne i$ on $C_i(z^0)$ and using the elementary inequality $|e^t - 1| \le e^{|t|}|t|$, one has, for $x \in C_i(z^0)$,

$$\begin{aligned} &\left| \frac{e^{-\Delta_{ij}^\varepsilon(x)/\varepsilon}}{1 + \sum_{k \ne j} e^{-\Delta_{ik}^\varepsilon(x)/\varepsilon}} - \frac{e^{-\Delta_{ij}^0(x)/\varepsilon}}{1 + \sum_{k \ne j} e^{-\Delta_{ik}^0(x)/\varepsilon}} \right| \\ &\le \left| e^{-\Delta_{ij}^\varepsilon(x)/\varepsilon} \Big(1 + \sum_{k \ne i} e^{-\Delta_{ik}^0(x)/\varepsilon}\Big) - e^{-\Delta_{ij}^0(x)/\varepsilon} \Big(1 + \sum_{k \ne i} e^{-\Delta_{ik}^\varepsilon(x)/\varepsilon}\Big) \right| \\ &\le 4\varepsilon^{-1} N \|z^\varepsilon - z^0\|_\infty e^{-\Delta_{ij}^0(x)/\varepsilon} e^{2\|z^\varepsilon - z^0\|_\infty/\varepsilon}. \end{aligned}$$

Lemma 4.1 (i) then yields

$$\left|\langle\varphi, T^\varepsilon - T^0\rangle_{L^2(P)} - \sum_{i\neq j}\int_{C_i(z^0)}\langle y_j - y_i, \varphi(x)\rangle \frac{e^{-\Delta^0_{ij}(x)/\varepsilon}}{1+\sum_{k\neq j} e^{-\Delta^0_{ik}(x)/\varepsilon}}\rho(x)\,dx\right|$$
$$\leq 4N^3\|\varphi\|_\infty M_\rho(\operatorname{diam}\mathcal{X})^{d-1}\|z^\varepsilon - z^0\|_\infty e^{2\|z^\varepsilon - z^0\|_\infty/\varepsilon}.$$

Furthermore,

$$\left|\frac{e^{-\Delta^0_{ij}(x)/\varepsilon}}{1+\sum_{k\neq j} e^{-\Delta^0_{ik}(x)/\varepsilon}} - \frac{e^{-\Delta^0_{ij}(x)/\varepsilon}}{1+e^{-\Delta^0_{ij}(x)/\varepsilon}}\right| \leq e^{-\Delta^0_{ij}(x)/\varepsilon}\sum_{k\neq i,j} e^{-\Delta^0_{ik}(x)/\varepsilon}.$$

Hence, by Lemma 4.1 (ii), we conclude that

$$\left|\langle\varphi, T^\varepsilon - T^0\rangle_{L^2(P)} - \sum_{i\neq j}\int_{C_i(z^0)}\langle y_j - y_i, \varphi(x)\rangle \frac{e^{-\Delta^0_{ij}(x)/\varepsilon}}{1+e^{-\Delta^0_{ij}(x)/\varepsilon}}\rho(x)\,dx\right|$$
$$\lesssim \|\varphi\|_\infty\big(\|z^\varepsilon - z^0\|_\infty e^{2\|z^\varepsilon - z^0\|_\infty/\varepsilon} + \varepsilon^2\log^2(1/\varepsilon)\big).$$

Setting $h^\varphi_{ij}(t) = \int_{H_{ij}(t)}\langle y_j - y_i, \varphi(x)\rangle\rho(x)\,d\mathcal{H}^{d-1}(x)$, an application of the coarea formula yields

$$\int_{C_i(z^0)}\langle y_j - y_i, \varphi(x)\rangle \frac{e^{-\Delta^0_{ij}(x)/\varepsilon}}{1+e^{-\Delta^0_{ij}(x)/\varepsilon}}\rho(x)\,dx = \frac{\varepsilon}{\|y_i - y_j\|}\int_0^\infty h^\varphi_{ij}(\varepsilon t)\frac{e^{-t}}{1+e^{-t}}\,dt.$$

We will replace $h^\varphi_{ij}(\varepsilon t)$ with $h^\varphi_{ij}(0)$. To this end, we need the following estimate, whose proof will be given after the proof of this theorem.

Lemma 4.2. *For any distinct indices i,j, $\mathcal{H}^{d-1}\big(H_{ij}(t)\Delta[H_{ij}(0)+tv_{ij}]\big)\lesssim t$ for all $t>0$ with $v_{ij} = (y_i - y_j)/\|y_i - y_j\|^2$. Here $[H_{ij}(0)+tv_{ij}] = \{x + tv_{ij} : x\in H_{ij}(0)\}$.*

The above lemma yields

$$|h^\varphi_{ij}(t) - h_{ij}(0)| \lesssim \|\varphi\|_\infty t + \int_{H_{ij}(0)}\|\varphi(x+tv_{ij})\rho(x+tv_{ij}) - \varphi(x)\rho(x)\|\,d\mathcal{H}^{d-1}(x) \lesssim \|\varphi\|_{\mathcal{C}^\alpha}(t\vee t^\alpha),$$

where we used the fact that ρ is Lipschitz and $\mathcal{X}$ is bounded. This implies

$$\left|\int_0^\infty (h^\varphi_{ij}(\varepsilon t) - h^\varphi_{ij}(0))\frac{e^{-t}}{1+e^{-t}}\,dt\right| \lesssim \|\varphi\|_{\mathcal{C}^\alpha}\varepsilon^\alpha, \quad \varepsilon\in(0,1),$$

Furthermore, since $h^\varphi_{ij}(0) = -h^\varphi_{ji}(0)$ (as $H_{ij}(0) = H_{ji}(0) = C_i(z^0)\cap C_j(z^0)$), we have $\sum_{i\neq j} h^\varphi_{ij}(0)/\|y_i - y_j\| = 0$. Putting everything together, we obtain the estimate in (4.3).

Step 2. In this step, we establish that $\|z^\varepsilon - z^0\|\lesssim \varepsilon^2\log^3(1/\varepsilon)$ for $\varepsilon\in(0,e^{-1})$, which, combined with Step 1, leads to the result of the theorem. This is a slight improvement on Corollary 2.2 in [20], but follows from the arguments there with a minor modification. We provide an outline below.

Set

$$G_i(\varepsilon, z) = \int \frac{e^{(\langle x, y_i\rangle - z_i)/\varepsilon}}{\sum_{j=1}^N e^{(\langle x,y_j\rangle - z_j)/\varepsilon}}\rho(x)\,dx - q_i,\ i\in[N],$$

and $G(\varepsilon,z) = (G_1(\varepsilon,z),\dots,G_N(\varepsilon,z))^\intercal$. By the first-order condition for the semidual problem (2.4), z^ε for $\varepsilon>0$ satisfies $G(\varepsilon,z^\varepsilon)=0$. By Theorem 3.2 in [20], $\nabla_z G(\varepsilon,z^\varepsilon)$ is invertible on $(\mathbf{1}_N)^\perp$ (the vector subspace of $\mathbb{R}^N$ orthogonal to $\mathbf{1}_N$), so the implicit function theorem yields that the mapping $\varepsilon\mapsto z^\varepsilon$ is $\mathcal{C}^1$ on $(0,\infty)$ with $\dot z^\varepsilon = -\big[\nabla_z G(\varepsilon,z^\varepsilon)\big]^{-1}\dot G(\varepsilon,z^\varepsilon)$, where $\dot{z}^\varepsilon = dz^\varepsilon/d\varepsilon$ and $\dot{G}(\varepsilon, z) = \partial G(\varepsilon, z)/\partial\varepsilon$ (note here that $\dot{G}(\varepsilon, z) \in (\mathbf{1}_N)^\perp$). Again, using Theorem 3.2 in [20], one obtains $\|\dot{z}^\varepsilon\| \lesssim \|\dot{G}(\varepsilon, z^\varepsilon)\|/\lambda_2$, where λ_2 denotes the second smallest eigenvalue of the matrix $\operatorname{diag}\{q_1, \dots, q_N\} - qq^\intercal$. By [47], $\lambda_2 \gtrsim 1$. Finally, the proof of Theorem 3.3 in [20] yields that for any $\eta > 0$,

$$|\dot{G}_i(\varepsilon, z^\varepsilon)| \lesssim \varepsilon^{-2}\eta^3 + \varepsilon^{-2}e^{-\eta/\varepsilon}\big(1 + \eta^2 + \varepsilon\eta + (\eta + \varepsilon^2)e^{-\eta/\varepsilon}\big),\ i \in [N].$$

Choosing $\eta = 3\varepsilon\log(1/\varepsilon)$ leads to $\|\dot{z}^\varepsilon\| \lesssim \varepsilon\log^3(1/\varepsilon)$, so that $\|z^\varepsilon - z^0\| \le \int_0^\varepsilon \|\dot{z}^t\|\,dt \lesssim \int_0^\varepsilon t\log^3(1/t)\,dt \lesssim \varepsilon^2\big(1 + \log(1/\varepsilon) + \log^2(1/\varepsilon) + \log^3(1/\varepsilon)\big) \lesssim \varepsilon^2\log^3(1/\varepsilon)$ for $\varepsilon \in (0, e^{-1})$. This completes the proof. □

Proof of Lemma 4.2. Set $b_{ij} = z_i^0 - z_j^0$ for notational convenience. Since $x \in [H_{ij}(0)+tv_{ij}]$ for $t > 0$ satisfies $\langle y_i - y_j, x\rangle - b_{ij} = t$, one sees that $[H_{ij}(0) + tv_{ij}] \setminus H_{ij}(t) \subset C_i(z^0)^c \cap C_j(z^0)^c$. Set $H_{ijk}(t) = \{x : x \in H_{ij}(0), x + tv_{ij} \in C_k(z^0)\}$, then $[H_{ij}(0) + tv_{ij}] \setminus H_{ij}(t) \subset \bigcup_{k\neq i,j}\big[H_{ijk}(t) + tv_{ij}\big]$. For $x \in H_{ijk}(t)$, the translation of x by tv_{ij} alters the sign of $\langle y_i - y_k, x\rangle - b_{ik}$, which can happen only when $0 \le \langle y_i - y_k, x\rangle - b_{ik} \le t\|y_i - y_k\|/\|y_i - y_j\|$. This implies $H_{ijk}(t) \subset \big\{x \in \mathcal{X} : \langle y_i - y_j, x\rangle = b_{ij}, b_{ik} \le \langle y_i - y_k, x\rangle \le b_{ik} + R_{\mathcal{Y}}t\big\} =: A_{ijk}(t)$ with $R_{\mathcal{Y}} = \max_{i,\,j,\,k\text{ distinct}} \frac{\|y_i - y_k\|}{\|y_i - y_j\|}$. We separately consider the following two cases.

Case (i). Suppose that $y_i - y_j$ and $y_i - y_k$ are linearly independent. In this case $\mathcal{H}^{d-1}(A_{ijk}(t)) \lesssim t$.

Case (ii). Suppose that $y_i - y_j$ and $y_i - y_k$ are linearly dependent, i.e., $y_i - y_k = c(y_i - y_j)$ for some $c \neq 0$. Set $L_1 = \{x : \langle y_i - y_j, x\rangle = b_{ij}\}$ and $L_2 = \{x : \langle y_i - y_k, x\rangle = b_{ik}\} = \{x : \langle y_i - y_j, x\rangle = c^{-1}b_{ik}\}$. Since L_1 and L_2 are parallel, we have $\operatorname{dist}(L_1, L_2) = \frac{|b_{ij} - c^{-1}b_{ik}|}{\|y_i - y_j\|}$. In addition, if $x \in A_{ijk}(t)$, then $\operatorname{dist}(x, L_1) = 0$ and $\operatorname{dist}(x, L_2) \le \frac{R_{\mathcal{Y}}t}{\|y_i - y_k\|}$. Arguing as in the proof of Lemma 4.1 (ii), one can show that there exists a sufficiently small t_0 that depends only on $\mathcal{X}, \rho, \mathcal{Y}$, and c_0 such that $A_{ijk}(t) = \varnothing$ for all $t \in (0, t_0)$.

Now, since the Hausdorff measure is translation invariant, we have

$$\mathcal{H}^{d-1}\big([H_{ij}(0) + tv_{ij}] \setminus H_{ij}(t)\big) \le \sum_{k\neq i,j} \mathcal{H}^{d-1}(A_{ijk}(t)) \lesssim t, \quad t \in (0, t_0). \tag{4.5}$$

For $t \ge t_0$, one may use the crude estimate $\mathcal{H}^{d-1}\big([H_{ij}(0)+tv_{ij}]\setminus H_{ij}(t)\big) \le \mathcal{H}^{d-1}(H_{ij}(0)) \le (\operatorname{diam}\mathcal{X})^{d-1}$ and adjust the constant in $\lesssim$ to see that the estimate (4.5) holds for all $t > 0$.

Next, consider the set $H_{ij}(t) \setminus [H_{ij}(0) + tv_{ij}]$. Each $x \in H_{ij}(t) \setminus [H_{ij}(0) + tv_{ij}]$ satisfies $\langle y_i - y_j, x - tv_{ij}\rangle = b_{ij}$, so one must have $x - tv_{ij} \in C_i(z^0)^c \cap C_j(z^0)^c$. This implies that $H_{ij}(t) \setminus [H_{ij}(0) + tv_{ij}] \subset \bigcup_{k\neq i,j}[\tilde{H}_{ijk}(t) + tv_{ij}]$, where $\tilde{H}_{ijk}(t) = \big\{x \in C_k(z^0) : x + tv_{ij} \in C_i(z^0), \langle y_i - y_j, x\rangle = b_{ij}\big\}$. In this case, each $x \in \tilde{H}_{ijk}(t)$ satisfies $-R_{\mathcal{Y}}t \le \langle y_i - y_k, x\rangle - b_{ik} \le 0$, so that $\tilde{H}_{ijk}(t) \subset \big\{x \in \mathcal{X} : b_{ik} - R_{\mathcal{Y}}t \le \langle y_i - y_k, x\rangle \le b_{ik}, \langle y_i - y_j, x\rangle = b_{ij}\big\} =: B_{ijk}(t)$. Arguing as in the previous case, we have $\mathcal{H}^{d-1}(B_{ijk}(t)) \lesssim t$. This completes the proof. □

Remark 4.3 (Comparison with [2, 20])**.** A key estimate in the proofs of Theorem 1.1 in [2] and Theorem 2.3 in [20] that concern the asymptotic expansions of the entropic cost is on the integral $\int_{C_i(z^0)} \Delta_{ij}^0(x)\frac{e^{-\Delta_{ij}^0(x)/\varepsilon}}{\sum_{k=1}^N e^{-\Delta_{ik}^0(x)/\varepsilon}}\rho(x)\,dx$. Crucial to their derivations is to use the fact that $\Delta_{ij}(x) \ge 0$ on $C_i(z^0)$ to upper and lower bound the integral. Then, applying the coarea formula and change of variables $t/\varepsilon \to t$ leads to the $O(\varepsilon^2)$ rate. In our case, the integrand in (4.4) need not be nonnegative nor a function of $\Delta_{ij}(x)$, so different arguments are needed.

4.3 Proof of Corollary 3.7

Let $\hat{q}_{n,i} = \hat{Q}_n(\{y_i\})$, then $\min_i \hat{q}_{n,i} \ge c_0/2$ with probability approaching one. Hence, Corollary 3.6 yields $\|\hat{T}_n^{\varepsilon_n} - \hat{T}_n^0\|_{(\mathcal{C}^\alpha)^*} \lesssim \varepsilon_n^{1+\alpha} \vee \varepsilon_n^2\log^3(1/\varepsilon_n)$. It remains to verify that the central limit theorem (3.1) for $\hat{T}_n^0$ holds under our assumption. To this end, it suffices to verify Assumptions 1 and 2 in [45]. Assumption 1 in [45] holds under the current Assumption 3.1 and the additional assumption made in the statement of the corollary. To verify Assumption 2 in [45] (L^1-Poincaré inequality for P), we first note that it suffices to verify the L^1-Poincaré inequality with the expectation replaced by the median; cf. Lemma 2.1 in [35]. Recall that the median minimizes the expected absolute deviation. Since $\mathcal{X}$ is convex, the uniform distribution over $\mathcal{X}$ satisfies (the median version of) the L^1-Poincaré inequality with constant K, say; cf. [6]. For any smooth function f on $\mathbb{R}^d$,

$$\min_c \int |f-c|\, dP \le M_\rho \min_c \int_{\mathcal{X}} |f-c|\, dx \le \frac{KM_\rho}{\inf_{x\in\mathcal{X}} \rho(x)} \int \|\nabla f\|\, dP.$$

This implies that P satisfies Assumption 2 in [45]. □

References

[1] Altschuler, J., Niles-Weed, J., and Rigollet, P., *Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration*, Advances in Neural Information Processing Systems, 2017.

[2] Altschuler, J. M., Niles-Weed, J., and Stromme, A. J., *Asymptotics for semidiscrete entropic optimal transport*, SIAM Journal on Mathematical Analysis, 54: 1718–1741, 2022. MR4393198

[3] Berman, R. J., *The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations*, Numerische Mathematik, 145: 771–836, 2020. MR4125978

[4] Bernton, E., Ghosal, P., and Nutz, M., *Entropic optimal transport: Geometry and large deviations*, Duke Mathematical Journal, 171: 3363–3400, 2022. MR4505361

[5] Bickel, P. J., *On some robust estimates of location*, The Annals of Mathematical Statistics: 847–858, 1965. MR0177484

[6] Bobkov, S. G., *Isoperimetric and analytic inequalities for log-concave probability measures*, The Annals of Probability, 27: 1903–1921, 1999. MR1742893

[7] Brenier, Y., *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on Pure and Applied Mathematics, 44: 375–417, 1991. MR1100809

[8] Carlier, G., *On the linear convergence of the multimarginal Sinkhorn algorithm*, SIAM Journal on Optimization, 32: 786–794, 2022. MR4418040

[9] Carlier, G., Pegon, P., and Tamanini, L., *Convergence rate of general entropic optimal transport costs*, Calculus of Variations and Partial Differential Equations, 62: 116, 2023. MR4565039

[10] Carlier, G., Duval, V., and Peyré, G., *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM Journal on Mathematical Analysis, 49: 1385–1418, 2017. MR3635459

[11] Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M., *Monge–Kantorovich depth, quantiles, ranks and signs*, The Annals of Statistics, 45: 223–256, 2017. MR3611491

[12] Chewi, S., Niles-Weed, J., and Rigollet, P., *Statistical Optimal Transport*, Lecture Notes in Mathematics, vol. 2377, Springer, 2025.

[13] Chizat, L., Roussillon, P., Léger, F., Vialard, François-Xavier, and Peyré, G., *Faster Wasserstein distance estimation with the Sinkhorn divergence*, Advances in Neural Information Processing Systems, 2020.

[14] Chizat, L., Delalande, A., and Vaškevičius, T., *Sharper Exponential Convergence Rates for Sinkhorn's Algorithm in Continuous Settings*, arXiv preprint arXiv:2407.01202, 2024.

[15] Conforti, G., and Tamanini, L., *A formula for the time derivative of the entropic cost and applications*, Journal of Functional Analysis, 280: 108964, 2021. MR4232667

[16] Conforti, G., Durmus, A., and Greco, G., *Quantitative contraction rates for Sinkhorn algorithm: beyond bounded costs and compact marginals*, arXiv preprint arXiv:2304.04451, 2023.

[17] Cordero-Erausquin, D., *Some applications of mass transport to Gaussian-type inequalities*, Archive for Rational Mechanics and Analysis, 161: 257–269, 2002. MR1894593

[18] Cuturi, M., *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, Advances in Neural Information Processing Systems, 2013.

[19] del Barrio, E., González-Sanz, A., Loubes, J.-M., and Niles-Weed, J., *An improved central limit theorem and fast convergence rates for entropic transportation costs*, SIAM Journal on Mathematics of Data Science, 5: 639–669, 2023. MR4616887

[20] Delalande, A., *Nearly tight convergence bounds for semi-discrete entropic optimal transport*, International Conference on Artificial Intelligence and Statistics, 2022.

[21] Divol, V., Niles-Weed, J., and Pooladian, A.-A., *Optimal transport map estimation in general function spaces*, arXiv preprint arXiv:2212.03722, 2022.

[22] Divol, V., Niles-Weed, J., and Pooladian, A.-A., *Tight stability bounds for entropic Brenier maps*, arXiv preprint arXiv:2404.02855, 2024.

[23] Eckstein, S., and Nutz, M., *Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm*, SIAM Journal on Mathematical Analysis, 54: 5922–5948, 2022. MR4506579

[24] Evans, L. C., and Gariepy, R. F., *Measure Theory and Fine Properties of Functions*, 1991. MR1158660

[25] Franklin, J., and Lorenz, J., *On the scaling of multidimensional matrices*, Linear Algebra and its Applications, 114: 717–735, 1989. MR0986904

[26] Ghosal, P., and Nutz, M., *On the convergence rate of Sinkhorn's algorithm*, arXiv preprint arXiv:2212.06000, 2022.

[27] Ghosal, P., and Sen, B., *Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing*, The Annals of Statistics, 50: 1012–1037, 2022. MR4404927

[28] Goldfeld, Z., Kato, K., Rioux, G., and Sadhu, R., *Limit theorems for entropic optimal transport maps and Sinkhorn divergence*, Electronic Journal of Statistics, 18: 980–1041, 2024. MR4718466

[29] Hütter, J.-C., and Philippe R., *Minimax estimation of smooth optimal transport maps*, The Annals of Statistics, 49: 1166–1194, 2021. MR4255123

[30] Koenker, R., and Portnoy, S., *L-estimation for linear models*, Journal of the American Statistical Association, 82: 851–857, 1987. MR0909992

[31] Léonard, C., *From the Schrödinger problem to the Monge–Kantorovich problem*, Journal of Functional Analysis, 262: 1879–1920, 2012. MR2873864

[32] Manole, T., Balakrishnan, S., Niles-Weed, J., and Wasserman, L., *Plugin estimation of smooth optimal transport maps*, The Annals of Statistics, 52: 966–998, 2024. MR4784066

[33] Mikami, T., *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probability theory and related fields, 129: 245–260, 2004. MR2063377

[34] Mikami, T., and Thieullen, M., *Optimal transportation problem by stochastic optimal control*, SIAM Journal on Control and Optimization, 47: 1127–1139, 2008. MR2407010

[35] Milman, E., *On the role of convexity in isoperimetry, spectral gap and concentration*, Inventiones Mathematicae, 177: 1–43, 2009. MR2507637

[36] Nutz, M., and Wiesel, J., *Stability of Schrödinger potentials and convergence of Sinkhorn's algorithm*, The Annals of Probability, 51: 699–722, 2023. MR4546630

[37] Nutz, M., *Introduction to entropic optimal transport*, Lecture notes, Columbia University, 2021.

[38] Nutz, M., and Wiesel, J., *Entropic optimal transport: Convergence of potentials*, Probability Theory and Related Fields, 184: 401–424, 2022. MR4498514

[39] Pal, S., *On the difference between entropic cost and the optimal transport cost*, The Annals of Applied Probability, 34: 1003–1028, 2024. MR4700251

[40] Peyré, G., and Cuturi, M., *Computational optimal transport: With applications to data science*, Foundations and Trends® in Machine Learning, 11: 355–607, 2019.

[41] Pooladian, A.-A., Cuturi, M., and Niles-Weed, J., *Debiaser Beware: Pitfalls of Centering Regularized Transport Maps*, International Conference on Machine Learning, pp. 17830–17847. PMLR, 2022.

[42] Pooladian, A.-A., and Niles-Weed, J., *Entropic estimation of optimal transport maps*, arXiv preprint arXiv:2109.12004, 2021.

[43] Pooladian, A.-A., Divol, V., and Niles-Weed, J., *Minimax estimation of discontinuous optimal transport maps: The semi-discrete case*, International Conference on Machine Learning, 2023.

[44] Rigollet, P., and Stromme, A.J., *On the sample complexity of entropic optimal transport*, The Annals of Statistics, 53: 61–90, 2025. MR4865008

[45] Sadhu, R., Goldfeld, Z., and Kato, K., *Stability and statistical inference for semidiscrete optimal transport maps*, The Annals of Applied Probability, 34: 5694–5736, 2024. MR4839636

[46] Santambrogio, F., *Optimal Transport for Applied Mathematicians*, Birkäuser, 2015. MR3409718

[47] Tanabe, K., and Sagae, M., *An exact Cholesky decomposition and the generalized inverse of the variance–covariance matrix of the multinomial distribution, with applications*, Journal of the Royal Statistical Society: Series B (Methodological), 54: 211–219, 1992. MR1157720

[48] Villani, C., *Optimal Transport: Old and New*, 2008. MR2459454

[49] Welsh, A. H., and Morrison, H. L., *Robust L estimation of scale with an application in astronomy*, Journal of the American Statistical Association, 85: 729–743, 1990. MR1138354

Acknowledgments. The authors would like to thank anonymous referees for their constructive comments.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: https://imstat.org/shop/donation/