

# Limit theorems for entropic optimal transport maps and Sinkhorn divergence

Ziv Goldfeld

*School of Electrical and Computer Engineering, Cornell University*  
*e-mail: [goldfeld@cornell.edu](mailto:goldfeld@cornell.edu)*

Kengo Kato

*Department of Statistics and Data Science, Cornell University*  
*e-mail: [kk976@cornell.edu](mailto:kk976@cornell.edu)*

Gabriel Rioux

*Center for Applied Mathematics, Cornell University*  
*e-mail: [ger84@cornell.edu](mailto:ger84@cornell.edu)*

Ritwik Sadhu

*Department of Statistics and Data Science, Cornell University*  
*e-mail: [rs2526@cornell.edu](mailto:rs2526@cornell.edu)*

**Abstract:** We study limit theorems for entropic optimal transport (EOT) maps, dual potentials, and the Sinkhorn divergence. The key technical tool we use is a first and second-order Hadamard differentiability analysis of EOT potentials with respect to the marginals, which may be of independent interest. Given the differentiability results, the functional delta method is used to obtain central limit theorems for empirical EOT potentials and maps. The second-order functional delta method is leveraged to establish the limit distribution of the empirical Sinkhorn divergence under the null. Building on the latter result, we further derive the null limit distribution of the Sinkhorn independence test statistic and characterize the correct order. Since our limit theorems follow from Hadamard differentiability of the relevant maps, as a byproduct, we also obtain bootstrap consistency and asymptotic efficiency of the empirical EOT map, potentials, and Sinkhorn divergence.

**MSC2020 subject classifications:** Primary 62G20, 60F05; secondary 60F17.

**Keywords and phrases:** Entropic map, entropic optimal transport, functional delta method, Hadamard differentiability, Sinkhorn divergence.

Received July 2022.

## Contents

1	Introduction . . . . .	982
1.1	Overview . . . . .	982
1.2	Contributions . . . . .	983
1.3	Literature review . . . . .	984

1.4	Concurrent work . . . . .	985
1.5	Organization . . . . .	985
1.6	Notation . . . . .	986
2	Background and preliminaries . . . . .	986
2.1	EOT problem . . . . .	986
2.2	EOT potentials . . . . .	987
2.3	EOT map . . . . .	988
2.4	Sinkhorn divergence . . . . .	989
3	Main results . . . . .	989
3.1	Limit theorems for EOT potentials and maps . . . . .	990
3.2	Limit theorems for Sinkhorn divergence . . . . .	991
4	Differentiability of EOT potentials and Sinkhorn divergence . . . . .	994
4.1	Hadamard differentiability of EOT potentials . . . . .	995
4.2	Hadamard differentiability of Sinkhorn divergence . . . . .	997
5	Bootstrap consistency and asymptotic efficiency . . . . .	997
5.1	Bootstrap consistency . . . . .	998
5.2	Asymptotic efficiency . . . . .	999
6	Proofs for Sections 3 and 4 . . . . .	1000
6.1	Proof of Theorem 3 . . . . .	1000
6.2	Proof of Theorem 4 . . . . .	1007
6.3	Proof of Lemma 3 . . . . .	1010
6.4	Proof of Theorem 5 . . . . .	1011
6.5	Proofs for Section 3 (except Proposition 2) . . . . .	1012
6.6	Proof of Proposition 2 . . . . .	1014
6.7	Proof of Proposition 3 . . . . .	1016
7	Discussions . . . . .	1017
7.1	Unbounded supports . . . . .	1017
7.2	Multimarginal EOT . . . . .	1018
A	$m$ -out-of- $n$ bootstrap for Sinkhorn null limit . . . . .	1019
A.1	Numerical experiments . . . . .	1020
A.2	Implementation details and theoretical guarantees . . . . .	1022
B	Auxiliary proofs . . . . .	1025
B.1	Proof of Lemma 1 . . . . .	1025
B.2	Proof of Lemma 6 . . . . .	1027
C	Technical tools . . . . .	1028
C.1	Hadamard differentiability and functional delta method . . . . .	1028
C.2	Hadamard differentiability of $Z$ -functional . . . . .	1030
D	Other auxiliary results . . . . .	1032
	Acknowledgments . . . . .	1034
	Funding . . . . .	1034
	References . . . . .	1035

## 1. Introduction

### 1.1. Overview

Optimal Transport (OT) [55] quantifies the discrepancy between Borel probability measures  $\mu^1, \mu^2$  on  $\mathbb{R}^d$  as

$$\text{OT}_c(\mu^1, \mu^2) := \inf_{\pi \in \Pi(\mu^1, \mu^2)} \int c \, d\pi, \quad (1)$$

where  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+ = [0, \infty)$  is the cost function and  $\Pi(\mu^1, \mu^2)$  is the set of couplings (or plans) between  $\mu^1$  and  $\mu^2$ . Under certain conditions on the marginals and the cost function, the OT plan  $\pi$  that achieves the infimum in (1) concentrates on the graph of a deterministic map  $T$ , called the OT map or the Brenier map when  $c$  is quadratic [10, 43]. OT tools have been successfully employed for various applications, encompassing machine learning, statistics, and applied mathematics; see [78, 77, 8, 88] and references therein. We refer the reader to [93, 94, 87] as standard references on OT theory.

Statistical OT seeks to estimate and carry out inference for OT and related objects thereof based on data. Two central objects of interest are the OT cost, which has natural applications to minimum distance estimation and testing, and the OT map, which is useful for transfer learning and domain adaptation tasks. Alas, despite its widespread applicability, the OT problem suffers from computational and statistical scalability issues. In general, the OT cost is difficult to compute and its plug-in empirical estimator converges towards the ground truth at the rate  $n^{-1/d}$  [34, 41, 62, 72], which is known to be minimax optimal without further assumptions [73]. Estimation of the Brenier map encounters similar difficulties, as the results of [54] suggest that the minimax rate would be  $n^{-1/d}$  in general (though formally a conjecture). Imposing smoothness on the marginals or the Brenier map can speed up minimax rates, but verification of such assumptions is nontrivial and computations of the minimax optimal estimators tend to be burdensome [30, 54, 63].

Entropic OT (EOT) has emerged as an appealing alternative to the classic Kantorovich formulation that circumvents these statistical and computational difficulties. EOT regularizes the transportation cost by the Kullback-Leibler (KL) divergence as [60]

$$S_{c,\varepsilon}(\mu^1, \mu^2) := \inf_{\pi \in \Pi(\mu^1, \mu^2)} \int c \, d\pi + \varepsilon \text{D}_{\text{KL}}(\pi \| \mu^1 \otimes \mu^2), \quad (2)$$

where  $\varepsilon > 0$  is a regularization parameter. As  $\varepsilon \rightarrow 0$ , the EOT problem converges towards OT, not only in terms of the transportation cost but also in optimal plans and dual potentials [3, 6, 15, 18, 22, 24, 59, 66, 69, 74, 76]. For fixed  $\varepsilon > 0$ , EOT alleviates the computational and statistical challenges associated with classic OT. Indeed, EOT between discrete distributions (e.g., empirical distributions) can be efficiently solved via the Sinkhorn algorithm [27, 4], whose time complexity scales quadratically in the number of support points. Regarding

empirical estimation, EOT and its centered version  $\bar{S}_{c,\varepsilon}(\mu^1, \mu^2) = S_{c,\varepsilon}(\mu^1, \mu^2) - (S_{c,\varepsilon}(\mu^1, \mu^1) + S_{c,\varepsilon}(\mu^2, \mu^2))/2$  (known as the Sinkhorn divergence) enjoy the parametric  $n^{-1/2}$  convergence rate under several settings [32, 44, 65, 83]. Recent work has further established central limit theorems (CLTs) for the EOT cost in certain cases [9, 32, 48, 57, 65] as well as estimation rate results for EOT maps and plans [80, 82, 83]; see a literature review below. Still, much is left to be desired on deeper understanding of limit behaviors of EOT and related objects, such as EOT potentials, maps, and Sinkhorn divergences, whose analysis poses a significant challenge from a probabilistic perspective.

## 1.2. Contributions

The present paper contributes to the growing literature on statistical OT by establishing limit theorems of the aforementioned objects: EOT potentials, maps, and the Sinkhorn divergence. The key ingredient of our derivations is a first and second-order Hadamard differentiability analysis of the EOT potentials with respect to (w.r.t.) the marginal distributions. Importantly, we establish Hadamard differentiability of the EOT potentials as maps into Hölder function spaces. The derivation first establishes Hadamard differentiability as maps into the space of continuous functions ( $\mathcal{C}$ -space) using the Schrödinger system, which characterizes EOT potentials, and a version of the implicit function theorem. We then lift the Hadamard differentiability to Hölder spaces by showing that derivatives of the EOT potentials are again Hadamard differentiable as maps into the  $\mathcal{C}$ -space. Having this result, the functional delta method [85] yields a CLT for the empirical EOT potentials, which, in turn, implies a CLT for the EOT map under the quadratic cost via the continuous mapping theorem. Both limit variables are characterized as Gaussians with values in the appropriate function spaces.

Hadamard differentiability results for the EOT potentials further enable developing a limit distribution theory for the Sinkhorn divergence. While asymptotic normality under the alternative (when  $\mu^1 \neq \mu^2$ ) is a straightforward consequence of existing EOT limit theorems [32, 48], the null case (when  $\mu^1 = \mu^2$ ) for general distributions is significantly more challenging and is a subject of this paper. The difficulty originates from the first-order Hadamard derivative of the Sinkhorn divergence nullifying when  $\mu^1 = \mu^2$  (as the functional achieves its global minimum there), which implies that the variance of the empirical Sinkhorn divergence vanishes under the null and the limit degenerates. To overcome this, we employ the second-order functional delta method [85], which requires second-order Hadamard derivatives of the EOT potentials and the Sinkhorn divergence. The Hadamard differentiability result of EOT potentials in Hölder spaces is key to finding such higher-order derivatives. Application of the second-order functional delta method then yields a distributional limit for the empirical Sinkhorn divergence at the rate of  $n^{-1}$ . To the best of our knowledge, the null limit distribution for the Sinkhorn divergence beyond the discrete case has been an open problem, and our result closes this gap.

The Sinkhorn divergence was applied to independence testing in [61], although the null limit distribution of the test statistic was not obtained in that

work. [61] proposed critical values of order  $n^{-1/2}$  derived from concentration inequalities, which is at odds with the  $n^{-1}$  order fluctuations under the null implied by our limit theorem. Building on the second-order Hadamard differentiability result for the Sinkhorn divergence and techniques for analysis of  $U$ -processes [19, 29], we establish the null limit distribution of the Sinkhorn independence test statistic and the correct  $n^{-1}$  order. Finally, the Hadamard differentiability results automatically yield bootstrap consistency and asymptotic efficiency of the empirical EOT potentials, map, and Sinkhorn divergence (under the alternative), which is another virtue of our approach.

### 1.3. Literature review

Statistical and probabilistic analyses of EOT and related objects have seen active research in the past couple of years. Regarding limit distribution theory, [9, 57] derived CLTs for the EOT cost with  $c(x_1, y_2) = \|x_1 - x_2\|^p$  ( $p \in [1, \infty)$ ) for finitely discrete distributions. [9] also derived the null limit of the Sinkhorn divergence in the discrete case by parameterizing it by finite-dimensional simplex vectors and directly finding the Hessian w.r.t. the simplex vectors. This approach does not directly extend to the general case. For general sub-Gaussian distributions, [65] showed asymptotic normality of  $\sqrt{n}(\mathbb{S}_{\|\cdot\|^2, \varepsilon}(\hat{\mu}_n^1, \mu^2) - \mathbb{E}[\mathbb{S}_{\|\cdot\|^2, \varepsilon}(\hat{\mu}_n^1, \mu^2)])$  and its two-sample analog. The main limitation of this result is that the centering term is the expected empirical EOT cost, which precludes performing inference for  $\mathbb{S}_{\|\cdot\|^2, \varepsilon}(\mu^1, \mu^2)$  itself. This limitation was addressed in [32], who derived a limit theorem for the population centering by combining the CLT from [65] with a bias bound of the form  $\mathbb{E}[\mathbb{S}_{\|\cdot\|^2, \varepsilon}(\hat{\mu}_n^1, \mu^2)] - \mathbb{S}_{\|\cdot\|^2, \varepsilon}(\mu^1, \mu^2) = o(n^{-1/2})$ . The recent work by the present authors [48] generalized this result to dependent data and further complemented it with asymptotic efficiency of the empirical EOT cost and consistency of the bootstrap estimate. It is worth noting that [32] derived the  $n^{-1}$  rate for the empirical Sinkhorn divergence under the null but did not derive its limit distribution.

Estimation of the EOT plan and map were also studied as a means to obtain computationally efficient proxies of the OT plan and the Brenier map, respectively. [82] considered estimation of the Brenier map under the quadratic cost via an entropic approximation; see also [80]. They analyzed the empirical EOT map and established a rate toward the Brenier map by taking  $\varepsilon = \varepsilon_n \rightarrow 0$ , which is however sub-optimal. [83] established the parametric rate toward the EOT map with  $\varepsilon > 0$  fixed. CLTs for the empirical EOT plan were studied in [57, 53] for discrete distributions and [51] for more general cases. The latter work also obtained a limit theorem for the EOT potentials in the  $\mathcal{C}$ -space, which is derived via a significantly different proof technique than ours (not relying on Hadamard differentiability) and is weaker than the convergence in Hölder spaces established herein. Hadamard differentiability in Hölder spaces is crucial for obtaining the null limit distribution of the empirical Sinkhorn divergence, which is one of our main contributions. Furthermore, our argument based on Hadamard differentiability of EOT potentials yields not only limit distributions but also

asymptotic efficiency and consistency of the bootstrap. Finally, [52] derived limit theorems for a *different* entropic regularization that makes the optimal solution explicit and thus relies on analysis techniques that significantly differ from ours.

Our Hadamard differentiability results also contribute to the study of stability of EOT, which has attracted growing interest in the mathematics literature [14, 17, 33, 37, 45, 67, 68, 75], and hence would be of independent interest beyond statistical applications. Those references study stability of EOT-related objects w.r.t. varying marginals (e.g., weak convergence of marginals) under general settings, but do not contain differentiability results like ours. The recent preprint [84] studies stability of the cost and dual potentials of the quadratic Gromov-Wasserstein distance with entropic penalty, leveraging the variational form from [96] that represents it as an infimum of a sequence of EOT problems.

#### 1.4. Concurrent work

The concurrent and independent work [50] establishes similar results concerning limit distributions of EOT potentials and the Sinkhorn divergence, but via a markedly different proof technique that does not involve Hadamard derivatives or the functional delta method. They also do not discuss bootstrapping or asymptotic efficiency. Our overlapping results—a mean-zero Gaussian limit for the EOT potentials in the Hölder space and a non-Gaussian limit for the Sinkhorn divergence under the null (with a scaling factor of  $n$ )—are consistent with each other, although they derive explicit forms of the limit distributions. Compared with [50], our contribution is to formulate and derive Hadamard differentiability for the EOT potentials and the Sinkhorn divergence, including higher-order ones, from which the limit distributions, consistency of resampling methods, and asymptotic efficiency of the empirical estimators automatically follow. The resampling methods enable performing inference without knowing the explicit limits, and our semiparametric efficiency result implies that, though the limit is not explicit, it is the best one can hope for. Furthermore, the Hadamard differentiability results enable deriving limit distributions beyond the empirical estimators, such as for the Sinkhorn independence statistic. As such, we view the contributions of our work and that of [50] as complementary to each other.

Another related work is [49], which appeared after the initial version of this paper was posted on arXiv. That work derived several limit theorems for EOT-related objects for nonsmooth costs, combining the approaches of [50] and [83], but their scope and proof techniques are substantially different.

#### 1.5. Organization

The rest of the paper is organized as follows. In Section 2, we collect background material on the EOT problem, potentials, map, and Sinkhorn divergence. In Section 3, we derive limit distributions of these objectives. We also derive the null limit distribution of the Sinkhorn independence test statistic in [61]. In Section 4, we collect Hadamard differentiability results for the relevant maps,

including higher-order ones. In Section 5, we discuss bootstrap consistency and asymptotic efficiency of the empirical estimators. Section 6 contains proofs for Sections 3 and 4. Section 7 leaves some concluding remarks and discussions on the extensions to the unbounded support case and multimarginal EOT. The Appendix contains additional results concerning the  $m$ -out-of- $n$  bootstrap for the Sinkhorn null limit (including small-scale numerical experiments), proofs that are omitted from the main text, technical tools used in the proofs, and other auxiliary results.

### 1.6. Notation

For a subset  $A$  of a topological space  $S$ , let  $\overline{A}^S$  denote the closure of  $A$ . We use  $\mathcal{P}(S)$  to denote the set of Borel probability measures on  $S$ . For  $\mu \in \mathcal{P}(S)$ ,  $\text{spt}(\mu)$  denotes its support. For a nonempty set  $S$ , let  $\ell^\infty(S)$  denote the Banach space of bounded real functions on  $S$  equipped with the sup-norm  $\|\cdot\|_{\infty,S} = \sup_{x \in S} |\cdot|$ . For every compact set  $\mathcal{X} \subset \mathbb{R}^d$ , let  $\mathcal{C}(\mathcal{X})$  denote the Banach space of continuous functions on  $\mathcal{X}$  equipped with the sup-norm  $\|\cdot\|_{\infty,\mathcal{X}}$ . For every multi-index  $k = (k_1, \dots, k_d) \in \mathbb{N}_0^d$  with  $|k| = \sum_{j=1}^d k_j$  ( $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ ), define the differential operator  $D^k$  by  $D^k = \frac{\partial^{|k|}}{\partial x_1^{k_1} \dots \partial x_d^{k_d}}$  with  $D^0 f = f$ . For every  $s \in \mathbb{N}_0$  and nonempty compact set  $\mathcal{X} \subset \mathbb{R}^d$  that agrees with the closure of its interior,  $\mathcal{C}^s(\mathcal{X})$  denotes the set of functions  $f$  on  $\mathcal{X}$  such that  $f$  has continuous derivatives of all orders  $\leq s$  on  $\text{int}(\mathcal{X})$  and the derivatives have continuous extensions to  $\mathcal{X}$  ( $\mathcal{C}^0(\mathcal{X}) = \mathcal{C}(\mathcal{X})$ ). Define the norm  $\|f\|_{\mathcal{C}^s(\mathcal{X})} = \sum_{j=0}^s \max_{|k|=j} \|D^k f\|_{\infty,\text{int}(\mathcal{X})}$ ;  $(\mathcal{C}^s(\mathcal{X}), \|\cdot\|_{\mathcal{C}^s(\mathcal{X})})$  is a separable Banach space [see Problem 5.1 in [47]; separability follows by noting that  $f \mapsto (D^k f)_{|k| \leq s}$  is isomorphic from  $\mathcal{C}^s(\mathcal{X})$  onto a closed subspace of  $\prod_{k:|k| \leq s} \mathcal{C}(\mathcal{X})$ ]. We often identify a finite signed Borel measure  $\gamma$  on  $\mathbb{R}^d$  with the linear functional  $f \mapsto \gamma(f) := \int f d\gamma$  defined on the bounded Borel functions on  $\mathbb{R}^d$ . For two real numbers  $a$  and  $b$ , let  $a \vee b = \max\{a, b\}$ .

## 2. Background and preliminaries

### 2.1. EOT problem

In this paper, we study EOT problems with smooth cost functions for compactly supported distributions on  $\mathbb{R}^d$ . We briefly review basic definitions and results concerning EOT problems. Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a smooth (i.e., infinitely differentiable) cost function. To simplify exposition concerning the Sinkhorn divergence, we will assume that  $c$  is symmetric, i.e.,  $c(x_1, x_2) = c(x_2, x_1)$  for all  $x_1, x_2 \in \mathbb{R}^d$ . A canonical example is the quadratic cost  $c(x_1, x_2) = \|x_1 - x_2\|^2/2$ . The corresponding EOT problem for compactly supported distributions  $\mu^1, \mu^2$  on  $\mathbb{R}^d$  is defined as

$$S_{c,\varepsilon}(\mu^1, \mu^2) = \inf_{\pi \in \Pi(\mu^1, \mu^2)} \int c d\pi + \varepsilon D_{\text{KL}}(\pi \| \mu^1 \otimes \mu^2), \quad (3)$$

where  $\varepsilon > 0$  is a regularization parameter,  $\Pi(\mu^1, \mu^2)$  is the set of couplings of  $(\mu^1, \mu^2)$ , and  $D_{\text{KL}}$  is the Kullback-Leibler divergence defined by

$$D_{\text{KL}}(\alpha \parallel \beta) := \begin{cases} \int \log(d\alpha/d\beta) d\alpha & \text{if } \alpha \ll \beta \\ +\infty & \text{otherwise} \end{cases}.$$

The EOT problem admits a unique solution  $\pi^*$ , which we call the *EOT plan*. Throughout this paper, we assume that the regularization parameter  $\varepsilon > 0$  is fixed, so we often omit the dependence on  $\varepsilon$ .

The EOT problem admits strong duality, which reads as

$$S_{c,\varepsilon}(\mu^1, \mu^2) = \sup_{\varphi=(\varphi_1, \varphi_2)} \int \varphi_1 d\mu^1 + \int \varphi_2 d\mu^2 - \varepsilon \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\mu^1 \otimes \mu^2 + \varepsilon, \quad (4)$$

where  $(\varphi_1 \oplus \varphi_2)(x_1, x_2) = \varphi_1(x_1) + \varphi_2(x_2)$  and the supremum is taken over all  $\varphi = (\varphi_1, \varphi_2) \in L^1(\mu^1) \times L^1(\mu^2)$ . There exist functions  $\varphi = (\varphi_1, \varphi_2) \in L^1(\mu^1) \times L^1(\mu^2)$  that achieve the supremum in the dual problem (4), which we call *EOT potentials*. EOT potentials are a.e. unique up to additive constants in the sense that if  $(\tilde{\varphi}_1, \tilde{\varphi}_2)$  is another pair of EOT potentials, then there exists a constant  $a \in \mathbb{R}$  such that  $\tilde{\varphi}_1 = \varphi_1 + a$   $\mu^1$ -a.e. and  $\tilde{\varphi}_2 = \varphi_2 - a$   $\mu^2$ -a.e. A pair of functions  $\varphi \in L^1(\mu^1) \times L^1(\mu^2)$  are EOT potentials if and only if they satisfy the so-called *Schrödinger system*

$$\int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\mu^j - 1 = 0 \quad \mu^i\text{-a.e.}, \quad i \neq j,$$

where  $\mu^j$  acts on the  $j$ -th coordinate, i.e.,

$$\int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\mu^2 = \int e^{\frac{\varphi_1(\cdot) + \varphi_2(x_2) - c(\cdot, x_2)}{\varepsilon}} d\mu^2(x_2).$$

Given EOT potentials  $(\varphi_1, \varphi_2)$ , the (unique) EOT plan  $\pi^*$  can be expressed as

$$d\pi^* = e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d(\mu^1 \otimes \mu^2). \quad (5)$$

See Section 1 in [74] and the references therein for the above results.

## 2.2. EOT potentials

In what follows, we deal with distributions supported in a compact set  $\mathcal{X} \subset \mathbb{R}^d$ . We will maintain the following assumption throughout the paper:

the set  $\mathcal{X} \subset \mathbb{R}^d$  is a bounded closed ball,

where we implicitly assume that the radius of  $\mathcal{X}$  is sufficiently large to contain the supports of the population distributions.

Our limit theorems rely on regularity properties of EOT potentials. These properties are summarized in the following lemma (proved in Appendix B), where the notation  $\equiv$  is used to represent equality that holds everywhere on the domain.



**Lemma 1** (Regularity of EOT potentials). *Pick and fix an arbitrary reference point  $(x_1^\circ, x_2^\circ) \in \mathcal{X} \times \mathcal{X}$ . The following hold.*

(i) *For every  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , there exist EOT potentials  $(\varphi_1, \varphi_2) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  such that*

$$\int e^{\frac{\varphi_1(\cdot) + \varphi_2(x_2) - c(\cdot, x_2)}{\varepsilon}} d\mu^2(x_2) - 1 \equiv 0, \int e^{\frac{\varphi_1(x_1) + \varphi_2(\cdot) - c(x_1, \cdot)}{\varepsilon}} d\mu^1(x_1) - 1 \equiv 0. \tag{6}$$

*Furthermore, if  $(\tilde{\varphi}_1, \tilde{\varphi}_2)$  is another pair of EOT potentials satisfying (6), then there exists  $a \in \mathbb{R}$  such that  $(\tilde{\varphi}_1, \tilde{\varphi}_2) \equiv (\varphi_1 + a, \varphi_2 - a)$ . Hence, there exists a unique pair of functions  $\varphi^\mu = (\varphi_1^\mu, \varphi_2^\mu) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  that satisfies (6) and  $\varphi_1^\mu(x_1^\circ) = \varphi_2^\mu(x_2^\circ)$ .*

(ii) *For every  $s \in \mathbb{N}$ , there exists  $R_s > 0$  that may depend on  $c, \varepsilon, d, \mathcal{X}$ , such that  $\|\varphi_1^\mu\|_{\mathcal{C}^s(\mathcal{X})} \vee \|\varphi_2^\mu\|_{\mathcal{C}^s(\mathcal{X})} \leq R_s$  for all  $\mu \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ .*

(iii) *Fix arbitrary  $s \in \mathbb{N}$  and equip  $\mathcal{P}(\mathcal{X})$  with the topology of weak convergence. Then, the map  $\mu \mapsto \varphi^\mu, \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  is continuous, i.e., if each  $\mu_n^i$  converges weakly to  $\mu^i$  ( $i = 1, 2$ ), then  $\varphi^{\mu_n} \rightarrow \varphi^\mu$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ .*

### 2.3. EOT map

The EOT map is an efficiently computable surrogate of the Brenier map. Recall that the (vanilla) OT problem with the quadratic cost between  $(\mu^1, \mu^2)$  with absolutely continuous  $\mu^1$  admits a ( $\mu^1$ -a.e.) unique OT map (called the Brenier map)  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and the (unique) OT plan concentrates on the graph of  $T$ ,  $\{(x_1, T(x_1)) : x_1 \in \text{spt}(\mu^1)\}$ . Hence, the Brenier map agrees with the conditional expectation of the second coordinate given the first under the OT plan (also called the barycenter projection). Motivated by this observation, [82] considered the following EOT analog of the Brenier map.

**Definition 1** (EOT map). Consider the quadratic cost  $c(x_1, x_2) = \|x_1 - x_2\|^2/2$ . For  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , the *EOT map* is defined by

$$T^\mu = \mathbb{E}_{\pi^\star}[X_2 \mid X_1 = \cdot], (X_1, X_2) \sim \pi^\star,$$

where  $\pi^\star$  is the unique EOT plan for  $(\mu^1, \mu^2)$ .

The EOT map is *a priori* defined only  $\mu^1$ -a.e. However, as noted in [82], using the expression (5) of the EOT plan, we may define a version of the conditional expectation for all  $x_1 \in \mathcal{X}$  (and indeed  $x_1 \in \mathbb{R}^d$ ) as

$$T^\mu(x_1) = \frac{\int_{\mathcal{X}} x_2 e^{\frac{\varphi_2^\mu(x_2) - \|x_1 - x_2\|^2/2}{\varepsilon}} d\mu^2(x_2)}{\int_{\mathcal{X}} e^{\frac{\varphi_2^\mu(x_2) - \|x_1 - x_2\|^2/2}{\varepsilon}} d\mu^2(x_2)}, \quad x_1 \in \mathcal{X}. \tag{7}$$

We always choose this version throughout the paper. Just as the Brenier map, the EOT map can be characterized in terms of the gradient of the EOT potential.

**Lemma 2** (Proposition 2 in [82]). *Under the setting of Definition 1, we have  $T^\mu(x_1) = x_1 - \nabla \varphi_1^\mu(x_1)$  for  $x_1 \in \mathcal{X}$ .*

[82] used the EOT map as a means to estimate the (original) Brenier map by taking  $\varepsilon = \varepsilon_n \rightarrow 0$ . Here, we view the EOT map  $T^\mu$  as an object of interest on its own right, rather than approximating the Brenier map.

**Example 1** (Vector quantile function). The Brenier map can be interpreted as a vector version of the quantile function when  $\mu^1$  is taken as a known reference measure, such as  $\mu^1 = \text{Unif}[0, 1]^d$  [13, 20, 46]. Indeed, for  $d = 1$ , the Brenier map sending  $\mu^1 = \text{Unif}[0, 1]$  to  $\mu^2$  agrees with the quantile function of  $\mu^2$  [87]. The EOT map  $T^\mu$  can thus be viewed as an efficiently computable surrogate of the vector quantile function [12].

### 2.4. Sinkhorn divergence

One drawback of EOT in (3) is that it is not a metric even for distance-like costs, such as  $c(x_1, x_2) = \|x_1 - x_2\|^p$ ,  $p \geq 1$ . In fact, EOT is not even a divergence since  $\mathcal{S}_{c,\varepsilon}(\mu^1, \mu^2) \neq 0$  when  $\mu^1 = \mu^2$ , which renders it incompatible for applications to homogeneity and independence testing.<sup>1</sup> To remedy this issue, a popular approach is center EOT to obtain the Sinkhorn divergence:

$$\bar{\mathcal{S}}_{c,\varepsilon}(\mu^1, \mu^2) = \mathcal{S}_{c,\varepsilon}(\mu^1, \mu^2) - \frac{1}{2}(\mathcal{S}_{c,\varepsilon}(\mu^1, \mu^1) + \mathcal{S}_{c,\varepsilon}(\mu^2, \mu^2)). \quad (8)$$

Under certain regularity conditions on the cost function (satisfied by the quadratic cost), the Sinkhorn divergence satisfies  $\bar{\mathcal{S}}_{c,\varepsilon}(\mu^1, \mu^2) \geq 0$  and  $\bar{\mathcal{S}}_{c,\varepsilon}(\mu^1, \mu^2) = 0$  if and only if  $\mu^1 = \mu^2$  [39].

Assuming  $x_1^o = x_2^o$ , by duality (4) and  $\varphi_1^{(\mu^1, \mu^1)} = \varphi_2^{(\mu^1, \mu^1)}$  (which follows by symmetry of the cost function), the Sinkhorn divergence admits the following expression [80]:

$$\bar{\mathcal{S}}_{c,\varepsilon}(\mu^1, \mu^2) = \int (\varphi_1^{(\mu^1, \mu^2)} - \varphi_1^{(\mu^1, \mu^1)}) d\mu^1 + \int (\varphi_2^{(\mu^1, \mu^2)} - \varphi_2^{(\mu^2, \mu^2)}) d\mu^2. \quad (9)$$

We will use this expression when studying the null limit distribution of the empirical Sinkhorn divergence.

### 3. Main results

In this section, we derive limit distributions for the empirical EOT potentials, map, and Sinkhorn divergence. For  $\mu^i \in \mathcal{P}(\mathcal{X})$ ,  $i = 1, 2$ , let  $\hat{\mu}_n^i$  denote the empirical distribution of a sample of  $n$  independent observations from  $\mu^i$ , i.e.,  $\hat{\mu}_n^i = n^{-1} \sum_{j=1}^n \delta_{X_j^i}$ , where  $X_1^i, \dots, X_n^i$  are i.i.d. according to  $\mu^i$ . The samples from  $\mu^1$  and  $\mu^2$  are assumed to be independent, and we set  $\hat{\mu}_n = (\hat{\mu}_n^1, \hat{\mu}_n^2)$ .

---

<sup>1</sup>A divergence on the space of probability distributions is a mapping to the extended reals that is nonnegative and nullifies if and only if the distributions are the same.

**3.1. Limit theorems for EOT potentials and maps**

Lemma 1 (ii) implies that EOT potentials lie in the Hölder space  $\mathcal{C}^s(\mathcal{X})$  for arbitrary  $s \in \mathbb{N}$ . The first main result concerns a limit distribution for the empirical EOT potentials in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ , which, in view of Lemma 2, automatically yields a limit distribution for the empirical EOT map in  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$ . Recall that a random variable  $G$  with values in a (real) Banach space  $\mathbb{B}$  is called Gaussian if for every  $b^* \in \mathbb{B}^*$  (the topological dual of  $\mathbb{B}$ ),  $b^*G$  is a real-valued Gaussian random variable. We say that  $G$  has mean zero if  $b^*G$  does so for every  $b^* \in \mathbb{B}^*$ . Let  $\xrightarrow{d}$  denote convergence in distribution. When necessary, convergence in distribution is understood in the sense of Hoffmann-Jørgensen (cf. Chapter 1 in [91]). For the product of two metric (or normed) spaces, we always consider a product metric (or norm).

**Theorem 1** (Limit theorem for EOT potentials). *Let  $s \in \mathbb{N}$  and  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  be arbitrary. Then, for  $\hat{\varphi}_n = \varphi^{\mu_n}$ , we have as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\varphi}_n - \varphi^\mu) \xrightarrow{d} \mathbf{G}^\mu \quad \text{in } \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X}),$$

where  $\mathbf{G}^\mu = (G_1^\mu, G_2^\mu)$  is a zero-mean Gaussian random variable with values in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ .

Theorem 1 immediately implies the following corollary concerning the limit distribution of the empirical EOT map. For  $s \in \mathbb{N}_0$ , let  $\mathcal{C}^s(\mathcal{X}; \mathbb{R}^d)$  denote the space of vector-valued functions  $f = (f_1, \dots, f_d) : \mathcal{X} \rightarrow \mathbb{R}^d$  whose coordinate functions belong to  $\mathcal{C}^s(\mathcal{X})$ , equipped with the norm  $\|f\|_{\mathcal{C}^s(\mathcal{X}; \mathbb{R}^d)} = \sum_{j=1}^d \|f_j\|_{\mathcal{C}^s(\mathcal{X})}$ .

**Corollary 1** (Limit theorem for EOT map). *Let  $s \in \mathbb{N}$  and  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  be arbitrary. Consider the quadratic cost  $c(x_1, x_2) = \|x_1 - x_2\|^2/2$  and the EOT map  $T^\mu$  given in (7). Then, for  $\hat{T}_n = T^{\mu_n}$ , we have as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{T}_n - T^\mu) \xrightarrow{d} -\nabla G_1^\mu \quad \text{in } \mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d).$$

The limit  $-\nabla G_1^\mu$  is a zero-mean Gaussian random variable in  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$ .

The recent work of [32] shows that  $\mathbb{E}[\|\hat{\varphi}_1^n - \varphi_1^\mu\|_{\mathcal{C}^s(\mathcal{X})}^2] \vee \mathbb{E}[\|\hat{\varphi}_2^n - \varphi_2^\mu\|_{\mathcal{C}^s(\mathcal{X})}^2] = O(n^{-1})$ . Theorem 1 complements their result by further showing distributional convergence of  $\sqrt{n}(\hat{\varphi}_n - \varphi^\mu)$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ .

The proof of Theorem 1 employs Hadamard differentiability of the map  $\mu \mapsto \varphi^\mu$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  (stated in Theorem 3 ahead) and the functional delta method; see Appendix C.1 for a review of Hadamard differentiability and the functional delta method. To this effect, we embed  $\mathcal{P}(\mathcal{X})$  into  $\ell^\infty(B^s)$ , where  $B^s$  is the unit ball in  $\mathcal{C}^s(\mathcal{X})$ . Since  $B^s$  is  $\mu^i$ -Donsker for  $i = 1, 2$  when  $s > d/2$  (cf. Theorem 2.7.1 in [91]), the conclusion of Theorem 1 follows from the functional delta method. The case of  $s \leq d/2$  follows by noting that the inclusion map  $f \mapsto f, \mathcal{C}^s(\mathcal{X}) \rightarrow \mathcal{C}^{s'}(\mathcal{X})$ , with  $s' < s$  is continuous. Having Theorem 1,

Corollary 1 follows by Lemma 2 and the fact that the gradient  $f \mapsto \nabla f$  is continuous from  $\mathcal{C}^s(\mathcal{X})$  into  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$ . The corollary can also be deduced directly from Hadamard differentiability of the map  $\mu \mapsto \nabla \varphi_1^\mu$  in  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$ . The covariance structure of the limiting Gaussian random variables in Theorem 1 and Corollary 1 can be inferred from their respective proofs.

*Remark 1* (One-sample case). We have only presented the two-sample limit distribution results for the EOT potentials and map, but as evident from the proof strategy, analogous conclusions continue to hold for the one-sample case where either  $\mu^1$  or  $\mu^2$  is known (cf. Example 1).

*Remark 2* (Measurability). Since  $(x_1, \dots, x_n) \mapsto n^{-1} \sum_{j=1}^n \delta_{x_j}, \mathcal{X}^n \rightarrow \mathcal{P}(\mathcal{X})$  is weakly continuous, in view of Lemma 1 (iii),  $\hat{\varphi}_n$  is a proper,  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ -valued random variable. Likewise, the empirical EOT map  $\hat{T}_n$  is a proper,  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$ -valued random variable.

*Remark 3* (Higher-order fluctuations). More can be said about higher-order fluctuations of the empirical EOT potentials. In Theorem 4, we will establish second-order Hadamard differentiability of the EOT potentials, which implies that

$$n(\hat{\varphi}_n - \varphi^\mu - [\varphi^\mu]'(\hat{\mu}_n - \mu))$$

converges in distribution in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ , where  $[\varphi^\mu]'$  is the first Hadamard derivative at  $\mu$  (cf. Theorem 3).

*Remark 4* (Comparison with [51]). Theorem 1 in the latest update of [51] (updated on July 9, 2022 on arXiv)<sup>2</sup> states a limit distribution result for the EOT potentials in  $\mathcal{C}(S_1) \times \mathcal{C}(S_2)$ , where  $S_i := \text{spt}(\mu^i)$  is a compact convex set (in fact [51] consider the multimarginal setting, but we focus our discussion on the two-marginal case). Their proof differs from ours in that they do not derive Hadamard differentiability of EOT potentials (nor does the proof contain Hadamard differentiability results). Importantly, the Hadamard differentiability result is stronger than just deriving a limit distribution, as it automatically yields bootstrap consistency and asymptotic efficiency of the empirical estimators; see Section 5 for further discussion. The question of asymptotic efficiency is not accounted for in [51]. Furthermore, Hadamard differentiability of the EOT potentials in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  plays a crucial role in deriving the null limit distribution of the empirical Sinkhorn divergence that involves the *second-order* Hadamard derivative of the EOT potentials in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ .

### 3.2. Limit theorems for Sinkhorn divergence

The second main result concerns limit distributions for the empirical Sinkhorn divergence. We first state an asymptotic normality result for the empirical Sinkhorn divergence.

---

<sup>2</sup>The initial version stated a weak convergence result in  $L^\infty(\mu^1) \times L^\infty(\mu^2)$ .

**Proposition 1** (Limit theorem for Sinkhorn divergence). *For every  $\mu=(\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , we have*

$$\sqrt{n}(\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \bar{S}_{c,\varepsilon}(\mu^1, \mu^2)) \xrightarrow{d} N(0, \sigma_\mu^2), \quad \text{as } n \rightarrow \infty,$$

where  $\sigma_\mu^2 = \text{Var}_{\mu^1}(\varphi_1^\mu - \varphi_1^{(\mu^1, \mu^1)}) + \text{Var}_{\mu^2}(\varphi_2^\mu - \varphi_2^{(\mu^2, \mu^2)})$ . Furthermore, the asymptotic variance  $\sigma_\mu^2$  is strictly positive whenever  $\bar{S}_{c,\varepsilon}(\mu^1, \mu^2) \neq 0$  and  $\text{spt}(\mu^1) \cap \text{spt}(\mu^2) \neq \emptyset$ .

*Remark 5.* Since  $(\mu^1, \mu^2) \mapsto \bar{S}_{c,\varepsilon}(\mu^1, \mu^2)$  is weakly continuous (this follows by Lemma 1 and the duality formula),  $\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  is a proper random variable.

The first claim of this proposition follows from relatively minor modifications to the proof of Theorem 7 in [48] that establishes asymptotic normality of the empirical EOT for the quadratic cost (or [32]). The (two-sample) asymptotic variance for the (uncentered) EOT cost  $S_{c,\varepsilon}$  is  $\text{Var}_{\mu^1}(\varphi_1^\mu) + \text{Var}_{\mu^2}(\varphi_2^\mu)$ , while that for the Sinkhorn divergence  $\bar{S}_{c,\varepsilon}$  involves debiased EOT potentials. The second claim provides conditions under which the limiting Gaussian distribution is nondegenerate. If  $\bar{S}_{c,\varepsilon}$  is a valid divergence (e.g., when the cost is quadratic), then  $\bar{S}_{c,\varepsilon}(\mu^1, \mu^2) \neq 0$  if and only if  $\mu^1 \neq \mu^2$ . The assumption that  $\text{spt}(\mu^1) \cap \text{spt}(\mu^2) \neq \emptyset$  can not be dropped in general. Indeed, if  $\mu^1$  and  $\mu^2$  are point masses at distinct points, we have  $\sigma_\mu^2 = 0$  but  $\mu^1 \neq \mu^2$ .

In Proposition 1, when  $\mu^1 = \mu^2$ , we have  $\sigma_\mu^2 = 0$ , which entails that  $\sqrt{n}\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) \rightarrow 0$  in probability. Indeed, [32] show that  $\mathbb{E}[\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)] = O(n^{-1})$  under  $\mu^1 = \mu^2$  for the quadratic cost, which implies that  $n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  is uniformly tight. The next theorem shows that, when  $\mu^1 = \mu^2$ ,  $n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  in fact has a weak limit, thereby determining a more precise random fluctuation of the empirical Sinkhorn divergence under the null.

**Theorem 2** (Limit theorem for Sinkhorn divergence under null). *Suppose that  $\mu^1 = \mu^2 = \mu \in \mathcal{P}(\mathcal{X})$ . Then,  $n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) \xrightarrow{d} \chi_\mu$  as  $n \rightarrow \infty$  for some random variable  $\chi_\mu$ . Furthermore, assuming nonnegativity of  $\bar{S}_{c,\varepsilon}$ , the support of  $\chi_\mu$  agrees with  $[0, \infty)$ , unless  $\chi_\mu = 0$  a.s.*

The proof of Theorem 2 is significantly more involved than that of Proposition 1 and relies on the second-order functional delta method. To this end, we establish second-order Hadamard differentiability of the map  $(\nu^1, \nu^2) \mapsto \bar{S}_{c,\varepsilon}(\nu^1, \nu^2)$  at  $(\nu^1, \nu^2) = (\mu, \mu)$ , which in turn involves the second-order Hadamard derivative of the EOT potentials. The limit variable is given by a nonlinear functional of a certain Gaussian process, but it seems highly nontrivial to derive an explicit expression of the limit distribution from our derivation; an implicit expression is provided in the proof.<sup>3</sup> Still, the limit distribution can be consistently estimated by the (two-sample version of) subsampling or the  $m$ -out-of- $n$  bootstrap [7, 79]. See Appendix A for details. The proof of the consistency of the  $m$ -out-of- $n$  bootstrap again relies on second-order Hadamard differentiability of the Sinkhorn divergence w.r.t. the marginals.

<sup>3</sup>The concurrent work [50] derives an explicit expression of the Sinkhorn null limit, albeit with a different technique.

The second claim of the theorem shows that the limit variable  $\chi_\mu$  is nondegenerate in most cases. Indeed, assuming that either  $\mu^1$  or  $\mu^2$  is not a point mass, in view of Lemma 10 below,  $\chi_\mu = 0$  a.s. if and only if the second Hadamard derivative of  $\bar{S}_{c,\varepsilon}$  at  $(\mu, \mu)$  is identically zero.<sup>4</sup> Otherwise, the support of  $\chi_\mu$  agrees with  $[0, \infty)$ .

*Remark 6* (One-sample case). Analogous results hold for the one-sample case. For example, when  $\mu^2$  is known,  $\sqrt{n}(\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \mu^2) - \bar{S}_{c,\varepsilon}(\mu^1, \mu^2)) \xrightarrow{d} N(0, \text{Var}_{\mu^1}(\varphi_1^\mu - \varphi_1^{(\mu^1, \mu^1)}))$  under the setting of Proposition 1, while  $n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \mu)$  has a weak limit under the setting of Theorem 2. The alternative limit in the one-sample case may degenerate even when the supports of the two marginals overlap. Indeed, when  $c$  is quadratic,  $\mu^1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ , and  $\mu^2 = \text{Unif}[-1, 1]$ , by symmetry considerations,  $\varphi_1^{(\mu^1, \mu^2)} - \varphi_1^{(\mu^1, \mu^1)}$  is  $\mu^1$ -a.s. constant, so the limit distribution for  $\sqrt{n}(\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \mu^2) - \bar{S}_{c,\varepsilon}(\mu^1, \mu^2))$  degenerates to 0.<sup>5</sup>

*Remark 7* (Higher-order fluctuations). The proof of Theorem 2 reveals that in general (i.e.,  $\mu^1 \neq \mu^2$  is allowed), the following stochastic expansion holds:

$$\sqrt{n}(\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \bar{S}_{c,\varepsilon}(\mu^1, \mu^2)) = \sum_{i=1}^2 \sqrt{n}(\hat{\mu}_n^i - \mu^i)(\varphi_i^\mu - \varphi_i^{(\mu^i, \mu^i)}) + n^{-1/2}r_n,$$

where  $r_n$  converges in distribution as  $n \rightarrow \infty$ . This expansion characterizes more precise random fluctuations of  $\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$ ; similar expansions hold for the (uncentered) EOT cost  $S_{c,\varepsilon}$ .

*Remark 8* (Comparison with [9]). A version of Theorem 2 was derived in [9] when the (common) distribution  $\mu$  is finitely discrete, where the limit is given by a weighted sum of independent  $\chi_1^2$ -random variables. When  $\mu$  is finitely discrete, it may be parameterized by a finite-dimensional simplex vector. Using this parameterization, [9] directly computed the Hessian matrix of the Sinkhorn divergence w.r.t. the simplex vectors and applied the second-order delta method. Clearly, this proof technique is restricted to the finitely discrete case and does not directly extend to the general case of Theorem 2. Indeed, the major challenge in the proof of Theorem 2 stems from the fact that in the general case, the problem is inherently infinite-dimensional and requires delicate functional analytic arguments. This is accounted for by the second-order Hadamard differentiability result of the Sinkhorn divergence, stated in Theorem 5.

**3.2.1. Independence testing with Sinkhorn divergence.** Let  $(V_i, W_i) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ ,  $i = 1, \dots, n$  be i.i.d. with common distribution  $\pi$ . Set  $d = d_1 + d_2$  and  $X_i = (V_i, W_i) \in \mathbb{R}^d$ . Let  $\pi^V$  and  $\pi^W$  denote the marginal distributions of  $V_i$  and  $W_i$ , respectively. Assume that  $\pi^V$  and  $\pi^W$  are compactly supported and let

<sup>4</sup>For finitely discrete marginals, [9] show that the Sinkhorn divergence, as a function of the simplex vectors representing discrete distributions, is  $\varepsilon$ -strongly convex in the first argument. This implies nondegeneracy of the Sinkhorn null limit, unless  $\mu$  is a point mass. Extending their argument to general compactly supported marginals appears to be nontrivial and is not pursued here.

<sup>5</sup>We thank an anonymous referee for this counterexample.

$\mathcal{X} \subset \mathbb{R}^d$  be a closed ball containing  $\text{spt}(\pi^V) \times \text{spt}(\pi^W)$ . Consider independence testing

$$H_0 : \pi = \pi^V \otimes \pi^W \quad \text{vs.} \quad H_1 : \pi \neq \pi^V \otimes \pi^W.$$

Motivated by computational considerations, [61] proposed a test based on the Sinkhorn divergence that compares the empirical distribution of  $X_i = (V_i, W_i)$  with the product of the marginal empirical distributions of  $V_i$  and  $W_i$ . [61] focused on the quadratic cost, but we allow here a general smooth cost. Specifically, [61] proposed a test that rejects the null for large values of the statistic

$$D_n = \bar{S}_{c,\varepsilon}(\hat{\pi}_n, \hat{\pi}_n^V \otimes \hat{\pi}_n^W),$$

where  $\hat{\pi}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ ,  $\hat{\pi}_n^V = n^{-1} \sum_{i=1}^n \delta_{V_i}$ , and  $\hat{\pi}_n^W = n^{-1} \sum_{i=1}^n \delta_{W_i}$ . The rationale behind this test is as follows. By Varadarajan's theorem, it holds that  $\hat{\pi}_n \rightarrow \pi$  and  $\hat{\pi}_n^V \otimes \hat{\pi}_n^W \rightarrow \pi^V \otimes \pi^W$  weakly a.s., so by Lemma 1 (iii) and duality, we have  $D_n \rightarrow \bar{S}_{c,\varepsilon}(\pi, \pi^V \otimes \pi^W)$  a.s. At least for the quadratic cost,  $\bar{S}_{c,\varepsilon}(\pi, \pi^V \otimes \pi^W) = 0$  if and only if  $\pi = \pi^V \otimes \pi^W$ , so it is reasonable to reject  $H_0$  when  $D_n$  is large.

[61] suggested a critical value of order  $n^{-1/2}$  derived from a finite sample deviation inequality for  $D_n$ . However, under  $H_0$ , both  $\hat{\pi}_n$  and  $\hat{\pi}_n^V \otimes \hat{\pi}_n^W$  converge to the same limit, so Theorem 2 suggests that the correct order of  $D_n$  under the null should be  $n^{-1}$ . The next proposition confirms this, thereby determining the precise rate for  $D_n$  under the null.

**Proposition 2** (Null limit of Sinkhorn independence test). *Consider the setting as stated above. Then, under the null  $H_0$ , we have  $nD_n \xrightarrow{d} \aleph_\pi$  as  $n \rightarrow \infty$  for some random variable  $\aleph_\pi$ . Furthermore, assuming nonnegativity of  $\bar{S}_{c,\varepsilon}$ , the support of  $\aleph_\pi$  agrees with  $[0, \infty)$ , unless  $\aleph_\pi = 0$  a.s.*

Note that Proposition 2 does not immediately follow from Theorem 2 since  $\hat{\pi}_n^V \otimes \hat{\pi}_n^W$  is not an empirical process but a two-sample  $V$ -process [29], and  $\hat{\pi}_n$  and  $\hat{\pi}_n^V \otimes \hat{\pi}_n^W$  are dependent. The proof first finds a joint limit distribution of  $\sqrt{n}(\hat{\pi}_n - \pi)$  and  $\sqrt{n}(\hat{\pi}_n^V \otimes \hat{\pi}_n^W - \pi)$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  with  $s > 2d$  using techniques from  $U$ -processes [19], and then applies the second-order functional delta method. Additionally, as in Theorem 2, unless  $\pi$  degenerates to a point mass or the second derivative of  $\bar{S}_{c,\varepsilon}$  at  $(\pi, \pi)$  with  $\pi = \pi^V \otimes \pi^W$  is identically zero, the support of the limit variable  $\aleph_\pi$  agrees with  $[0, \infty)$ , yielding nondegeneracy of the limit law. An implicit expression for the derived limiting random variable is provided in the proof.

#### 4. Differentiability of EOT potentials and Sinkhorn divergence

As already stated, the main ingredients of the proofs of the results in the preceding section are the first and second-order Hadamard differentiability results for the EOT potentials and Sinkhorn divergence. The present section collects those results.

### 4.1. Hadamard differentiability of EOT potentials

Our goal is to establish Hadamard differentiability of the map  $\mu \mapsto \varphi^\mu$ . Since  $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  is *a priori* not a vector space, we embed the preceding map into a normed space as follows. Note that  $\varphi^\mu$  is fully characterized as the solution of (6) (subject to the normalization  $\varphi_1(x_1^\circ) = \varphi_2(x_2^\circ)$ ), and whenever  $\varphi_i \in \mathcal{C}^s(\mathcal{X})$  for  $i = 1, 2$ , we have

$$\left( e^{\frac{\varphi_1(\cdot) + \varphi_2(x_2) - c(\cdot, x_2)}{\varepsilon}}, e^{\frac{\varphi_1(x_1) + \varphi_2(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right) \in \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X}), \quad \forall (x_1, x_2) \in \mathcal{X} \times \mathcal{X},$$

for  $s \in \mathbb{N}$  arbitrary. With this in mind, it is natural to think of  $\mu^i$  as a functional on  $\mathcal{C}^s(\mathcal{X})$ , and we identify  $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  as a subset of  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ , where

$$B^s = \{f \in \mathcal{C}^s(\mathcal{X}) : \|f\|_{\mathcal{C}^s(\mathcal{X})} \leq 1\}.$$

Formally, defining  $\tau : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \rightarrow \ell^\infty(B^s) \times \ell^\infty(B^s)$  by  $(\tau\mu)_i = (f \mapsto \int f d\mu^i)_{f \in B^s}$ ,  $i = 1, 2$ , we identify the map  $\nu \mapsto \varphi^\nu$  with  $\tau\nu \mapsto \varphi^\nu$ . Since  $\tau$  is one-to-one (cf. Lemma 16), the latter map is well-defined. Equip  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  with the norm  $\|\gamma^1\|_{\infty, B^s} \vee \|\gamma^2\|_{\infty, B^s}$  for  $\gamma = (\gamma^1, \gamma^2) \in \ell^\infty(B^s) \times \ell^\infty(B^s)$ . Finally, for  $\eta \in \mathcal{P}(\mathcal{X})$ , set

$$\mathcal{P}_\eta = \{\nu \in \mathcal{P}(\mathcal{X}) : \text{spt}(\nu) \subset \text{spt}(\eta)\} \quad \text{and} \quad \mathcal{M}_\eta = \{t(\nu - \eta) : \nu \in \mathcal{P}_\eta, t > 0\}.$$

Elements of  $\mathcal{M}_\eta$  are signed measures with total mass zero supported in  $\text{spt}(\eta)$ . Observe that  $\mathcal{M}_\eta \subset \ell^\infty(B^s)$ . These definitions are an artifact of our proof technique.

**Theorem 3** (Hadamard differentiability of EOT potentials). *For every  $s \in \mathbb{N}$  and  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , the map  $\nu \mapsto \varphi^\nu, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  is Hadamard differentiable at  $\mu$  tangentially to  $\frac{\ell^\infty(B^s)}{\mathcal{M}_{\mu^1}} \times \frac{\ell^\infty(B^s)}{\mathcal{M}_{\mu^2}}$ .*

The first step of the proof of Theorem 3 is to establish Hadamard differentiability in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  instead of  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ . To this effect, we regard  $\varphi^\mu$  as a solution to the system of functional equations (6) and use a version of the implicit function theorem to prove Hadamard differentiability of the map  $\mu \mapsto \varphi^\mu$  in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ . Precisely, Hadamard differentiability of the EOT potentials is first established in  $\mathcal{C}(\text{spt}(\mu^1)) \times \mathcal{C}(\text{spt}(\mu^2))$ , then the EOT potentials are extended to  $\mathcal{X} \times \mathcal{X}$  via (6) and differentiability in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  follows readily.

To lift the Hadamard differentiability to  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ , we again use the expression from (6),

$$e^{-\varphi_i^\mu/\varepsilon}(x_i) = \int e^{(\varphi_j^\mu(x_j) - c(x_1, x_2))/\varepsilon} d\mu^j(x_j), \quad i \neq j,$$

and show that derivatives of  $\varphi_i^\mu$  are Hadamard differentiable in  $\mathcal{C}(\mathcal{X})$ . This argument is partly inspired by the proof of Theorem 4.5 in [32]. Completeness of  $\mathcal{C}^s(\mathcal{X})$  then yields that the map  $\mu \mapsto \varphi_i^\mu$  is Hadamard differentiable in  $\mathcal{C}^s(\mathcal{X})$ . See the proof in Section 6.1 for the full details.



*Remark 9* (Tangent cone). Since  $\mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  is convex, the set  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  agrees with the tangent cone to  $\mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  at  $\boldsymbol{\mu}$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  (cf. Chapter 4 in [5]; see also Appendix C.1). Each element of  $\overline{\mathcal{M}_{\mu^i}^{\ell^\infty(B^s)}}$  extends uniquely to a bounded linear functional on  $\mathcal{C}^s(\mathcal{X})$  (cf. Lemma 1 in [70]). Hence, by the Riesz-Markov-Kakutani theorem, each  $\gamma^i \in \overline{\mathcal{M}_{\mu^i}^{\ell^\infty(B^s)}}$  corresponds to finite signed Borel measures  $(\gamma_k^i)_{k \in \mathbb{N}_0^d, |k| \leq s}$  on  $\mathcal{X}$  supported in  $\text{spt}(\mu^i)$  such that  $\gamma^i(f) = \sum_{|k| \leq s} \int D^k f d\gamma_k^i$  for  $f \in \mathcal{C}^s(\mathcal{X})$ . We remark that this action also makes sense for a function  $f \in \mathcal{C}(\text{spt}(\mu^i))$  that merely admits some extension  $\bar{f} \in \mathcal{C}^s(\mathcal{X})$ . Indeed, let  $\gamma_n^i \in \mathcal{M}_{\mu^i}$  converge to  $\gamma^i$  in  $\ell^\infty(B^s)$ , then  $\gamma_n^i(f) = \gamma_n^i(\bar{f}) \rightarrow \gamma^i(\bar{f})$  as  $\text{spt}(\gamma_n^i) \subset \text{spt}(\mu^i)$ ; the first equality also shows that the limit is independent of the choice of extension, hence we may define  $\gamma^i(f) := \gamma^i(\bar{f})$ . Abusing notation, we often denote the action of  $\gamma^i$  on  $f \in \mathcal{C}(\text{spt}(\mu^i))$  admitting a  $\mathcal{C}^s$  extension to  $\mathcal{X}$  or  $f \in \mathcal{C}^s(\mathcal{X})$  as  $\gamma^i(f) = \int f d\gamma^i$ .

*Remark 10* (Functional delta method). Recalling that  $s$  in Lemma 1 (ii) is arbitrary, if we choose  $s > d/2$ , then  $B^s$  is  $\mu^i$ -Donsker for each  $i = 1, 2$  (cf. Theorem 2.7.1 in [91]). We use this fact in the proof of the limit theorem for EOT potentials from Theorem 1. Since the support of a tight  $\mu^i$ -Brownian bridge in  $\ell^\infty(B^s)$  is contained in  $\overline{\mathcal{M}_{\mu^i}^{\ell^\infty(B^s)}}$  (see Lemma 17), the functional delta method (see Lemma 11) immediately applies to the map  $\boldsymbol{\mu} \mapsto \boldsymbol{\varphi}^\boldsymbol{\mu}$ . Similar comments apply to other Hadamard differentiability results.

The derivation of the null limit distribution of the empirical Sinkhorn divergence involves the second-order Hadamard derivative of EOT potentials, which is given next.

**Theorem 4** (Second-order Hadamard differentiability of EOT potentials). *For every  $s \in \mathbb{N}$  and  $\boldsymbol{\mu} = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , there exists a continuous map  $[\boldsymbol{\varphi}^\boldsymbol{\mu}]'' : \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}} \rightarrow \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  such that for every sequence  $(\boldsymbol{\mu}_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  with  $\boldsymbol{\gamma}_t := t^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}) \rightarrow \boldsymbol{\gamma}$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  as  $t \downarrow 0$ , we have*

$$\frac{\boldsymbol{\varphi}^{\boldsymbol{\mu}_t} - \boldsymbol{\varphi}^\boldsymbol{\mu} - t[\boldsymbol{\varphi}^\boldsymbol{\mu}]'(\boldsymbol{\gamma}_t)}{t^2/2} \rightarrow [\boldsymbol{\varphi}^\boldsymbol{\mu}]''(\boldsymbol{\gamma}) \quad \text{in } \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X}).$$

*The map  $[\boldsymbol{\varphi}^\boldsymbol{\mu}]''$  is positively homogeneous of degree 2, i.e.,  $[\boldsymbol{\varphi}^\boldsymbol{\mu}]''(t\boldsymbol{\gamma}) = t^2[\boldsymbol{\varphi}^\boldsymbol{\mu}]''(\boldsymbol{\gamma})$  for every  $t > 0$  and  $\boldsymbol{\gamma} \in \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ .*

To prove this theorem, we extend Lemma 3.9.34 in [91] to the second-order Hadamard derivative, which is presented in Lemma 15. It is worth noting that, while the second-order Hadamard differentiability result is stated in terms of  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ , its proof requires the (first-order) Hadamard derivative in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  to verify Condition (iv) in Lemma 15. See the proof in Section 6.2 for details.

#### 4.2. Hadamard differentiability of Sinkhorn divergence

In this section, we study Hadamard derivatives of the Sinkhorn divergence. The following lemma follows relatively easily from the proof of Theorem 7 in [48].

**Lemma 3** (Hadamard derivative of Sinkhorn divergence). *For every  $s \in \mathbb{N}$  and  $\boldsymbol{\mu} = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , the map  $\boldsymbol{\nu} = (\nu^1, \nu^2) \mapsto \bar{S}_{c,\varepsilon}(\nu^1, \nu^2), \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathbb{R}$  is Hadamard differentiable at  $\boldsymbol{\mu}$  tangentially to  $\overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$  with derivative*

$$[\bar{S}_{c,\varepsilon}^\boldsymbol{\mu}]'(\boldsymbol{\gamma}) = \int (\varphi_1^\boldsymbol{\mu} - \varphi_1^{\boldsymbol{\mu}^{(\mu^1, \mu^1)}}) d\gamma^1 + \int (\varphi_2^\boldsymbol{\mu} - \varphi_2^{\boldsymbol{\mu}^{(\mu^2, \mu^2)}}) d\gamma^2$$

for  $\boldsymbol{\gamma} = (\gamma^1, \gamma^2) \in \overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$ .

When  $\mu^1 = \mu^2$ , we have  $[\bar{S}_{c,\varepsilon}^\boldsymbol{\mu}]'(\boldsymbol{\gamma}) = 0$ . Therefore, to explore the limit distribution of the empirical Sinkhorn divergence under the null  $\mu^1 = \mu^2$ , we need to look at the second-order Hadamard derivative of the Sinkhorn divergence, which is given next.

**Theorem 5** (Second-order Hadamard derivative of Sinkhorn divergence). *For every  $s \in \mathbb{N}$  and  $\boldsymbol{\mu} = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , there exists a continuous functional  $\Delta_\boldsymbol{\mu} : \overline{\mathcal{M}_\mu}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_\mu}^{\ell^\infty(B^s)} \rightarrow \mathbb{R}$  such that for every sequence  $(\mu_t^1, \mu_t^2) \in \mathcal{P}_\mu \times \mathcal{P}_\mu$  with  $t^{-1}(\mu_t^i - \mu) \rightarrow \gamma^i$  in  $\ell^\infty(B^s)$  as  $t \downarrow 0$  for  $i = 1, 2$ , we have*

$$\frac{\bar{S}_{c,\varepsilon}(\mu_t^1, \mu_t^2)}{t^2/2} \rightarrow \Delta_\boldsymbol{\mu}(\boldsymbol{\gamma})$$

with  $\boldsymbol{\gamma} = (\gamma^1, \gamma^2)$ . The functional  $\Delta_\boldsymbol{\mu}$  is positively homogeneous of degree 2.

Given the first and second-order Hadamard differentiability results for the EOT potentials, the proof of Theorem 5 is reasonably straightforward. Indeed, the proof consists of expanding  $\varphi_i^{\boldsymbol{\mu}^{(\mu_t^1, \mu_t^2)}}$  and  $\varphi_i^{\boldsymbol{\mu}^{(\mu_t^i, \mu_t^i)}}$  up to the second-order and plugging these expansions into the dual expression (9) of  $\bar{S}_{c,\varepsilon}(\mu_t^1, \mu_t^2)$ .

### 5. Bootstrap consistency and asymptotic efficiency

As discussed before, our limit theorems in Theorem 1, Corollary 1, and Proposition 1 follow by establishing Hadamard differentiability of the relevant maps. Importantly, Hadamard differentiability results automatically also yield bootstrap consistency and asymptotic efficiency (cf. Chapter 3.11 in [91] and Chapter 25 in [90]) of the empirical estimators of the EOT potentials, map, and Sinkhorn divergence (with  $\mu^1 \neq \mu^2$ ). To simplify our discussion, we focus here on estimating the EOT map for the quadratic cost with known  $\mu^1$  (i.e., the one-sample case); other cases are similar. The setting of known  $\mu^1$  is motivated by the connection of the EOT map to the vector quantile function (Example 1). Consider the setting of Corollary 1.

**5.1. Bootstrap consistency**

Let  $\tilde{T}_n = T^{(\mu^1, \hat{\mu}_n^2)}$  be the empirical EOT map, and  $\tilde{T}_n^B$  be its bootstrap analog, i.e.,  $\tilde{T}_n^B = T^{(\mu^1, \hat{\mu}_n^{2,B})}$ , where  $\hat{\mu}_n^{2,B}$  is the empirical distribution of a bootstrap sample from  $\hat{\mu}_n^2$  of size  $n$  (cf. Chapter 3.6 in [91] and Chapter 23 in [90]).

Pick any  $s, s' \in \mathbb{N}_0$  with  $s' < s$  and  $s > d/2$ . Lemma 2 and Theorem 3 yield that the map  $\delta : \nu^2 \mapsto T^{(\mu^1, \nu^2)}, \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \rightarrow \mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d)$  is Hadamard differentiable at  $\nu^2 = \mu^2$  tangentially to  $\overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  with derivative  $\delta'_{\mu^2}(\gamma^2) = -[\nabla \varphi_1^\mu]'(0, \gamma^2)$ . The tangent cone  $\overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  contains a vector subspace on which a tight  $\mu^2$ -Brownian bridge  $\mathbb{G}_2^{\mu^2}$  in  $\ell^\infty(B^s)$  concentrates (cf. Lemma 17; note that  $B^s$  with  $s > d/2$  is  $\mu^2$ -Donsker). Hence, by the functional delta method, we have

$$\sqrt{n}(\tilde{T}_n - T^\mu) \xrightarrow{d} \delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) \quad \text{in } \mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d).$$

Furthermore, by Theorems 3.6.1 and 3.9.11 in [91], the following bootstrap consistency holds:

$$\sup_{h \in \text{BL}_1(\mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d))} |\mathbb{E}^B [h(\sqrt{n}(\tilde{T}_n^B - \tilde{T}_n))] - \mathbb{E}[h(\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}))]| \rightarrow 0 \quad (10)$$

in outer probability, where  $\mathbb{E}^B$  denotes the conditional expectation given the sample and  $\text{BL}_1(\mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d))$  is the class of 1-Lipschitz functions  $h : \mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d) \rightarrow [-1, 1]$ .

**Example 2** (Confidence bands for EOT map). Consider constructing confidence bands for  $T_j^\mu$  with  $T^\mu = (T_1^\mu, \dots, T_d^\mu)$ . The continuous mapping theorem yields that  $\|\sqrt{n}(\tilde{T}_{n,j} - T_j^\mu)\|_{\infty, \mathcal{X}} \xrightarrow{d} \|\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})\|_{j, \infty, \mathcal{X}}$ . For a given  $\alpha \in (0, 1)$ , let  $\hat{q}_{1-\alpha}$  denote the conditional  $(1 - \alpha)$ -quantile of  $\|\sqrt{n}(\tilde{T}_{n,j}^B - \tilde{T}_{n,j})\|_{\infty, \mathcal{X}}$  given the data. The bootstrap consistency result (10) yields that  $\{[\tilde{T}_j^n(x_1) \pm \hat{q}_{1-\alpha}/\sqrt{n}] : x_1 \in \mathcal{X}\}$  is a valid confidence band for  $T_j^\mu$ , i.e.,  $\mathbb{P}(T_j^\mu(x_1) \in [\tilde{T}_j^n(x_1) \pm \hat{q}_{1-\alpha}/\sqrt{n}]) \rightarrow 1 - \alpha$ .

*Remark 11* (Statistical inference for unregularized Brenier maps). When the source measure  $\mu^1$  is absolutely continuous, Brenier’s theorem [10] guarantees the existence of the  $\mu^1$ -a.e. unique map (called the Brenier map) given by the gradient of a convex function transporting  $\mu^1$  onto  $\mu^2$ . The EOT map approximates the (unregularized) Brenier map as  $\varepsilon \rightarrow 0$  (cf. [18, 66]). However, all the limit theorems in the present paper crucially rely on the fact that the regularization parameter  $\varepsilon > 0$  is fixed and do not directly extend to the case where  $\varepsilon = \varepsilon_n \rightarrow 0$ . Indeed, estimation of the unregularized Brenier map suffers from the curse of dimensionality [54], so  $\sqrt{n}$ -consistency toward the unregularized Brenier map does not hold in general for  $\varepsilon = \varepsilon_n \rightarrow 0$ . One exception is the semidiscrete case where  $\mu^1$  is absolutely continuous and  $\mu^2$  is finitely discrete, for which [81] establish the parametric convergence rate of empirical EOT

maps with vanishing regularization parameters under the squared  $L^2$ -loss. For the semidiscrete case, the recent preprint [86] by a subset of the authors studies statistical inference for the empirical (unregularized) Brenier map. Beyond the semidiscrete case, however, limiting distributional results for the (unregularized) Brenier are still rather limited. The recent preprint [64], which appeared on arXiv after the present work, derives pointwise limiting distributions for kernel-based estimates of the Brenier map for densities supported on the flat torus. For finitely discrete marginals, [56] derives limit laws for the unregularized OT plan, which in particular, specialize to the Brenier map, provided that the latter exists.

### 5.2. Asymptotic efficiency

Regarding asymptotic efficiency, we follow Chapter 3.11 in [91]. Consider the setting of Corollary 1 and fix  $(\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ . As before, let  $s, s' \in \mathbb{N}_0$  be such that  $s' < s$  and  $s > d/2$ . For notational convenience, set  $\mathbb{B} = \mathcal{C}^{s'}(\mathcal{X}; \mathbb{R}^d)$ . By linearity of the derivative, the limit variable  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})$  is zero-mean Gaussian in  $\mathbb{B}$ .

To apply the results of Chapter 3.11 in [91], we need to specify statistical experiments  $(\mathcal{X}_n, \mathcal{A}_n, P_{n,h} : h \in H)$  indexed by a vector subspace  $H$  of a Hilbert space and local parameters  $\kappa_n(h)$ . Choose  $H$  to be the set of bounded measurable functions on  $\mathcal{X}$  with  $\mu^2$ -mean zero equipped with the  $L^2(\mu^2)$ -inner product, and set  $\mathcal{X}_n = \mathcal{X}^n$ ,  $\mathcal{A}_n = (\text{Borel } \sigma\text{-field on } \mathcal{X}^n)$ ,  $P_{n,h} = (\mu_{n,h}^2)^{\otimes n}$  with  $d\mu_{n,h}^2 = (1 + h/\sqrt{n})d\mu^2$ , and  $\kappa_n(h) = \delta(\mu_{n,h}^2) = T^{(\mu^1, \mu_{n,h}^2)} : H \rightarrow \mathbb{B}$ . Note that for each  $h \in H$ ,  $\mu_{n,h}^2$  is a valid probability measure for sufficiently large  $n$ . By Lemma 3.10.11 in [91], the sequence of experiments  $(\mathcal{X}_n, \mathcal{A}_n, P_{n,h} : h \in H)$  is asymptotically normal in the sense of [91, p 412]. Under this setup, the following proposition regarding asymptotic efficiency of the empirical EOT map holds. We say that the parameter sequence  $\kappa_n(h)$  is *regular* if there exists a continuous linear operator  $\dot{\kappa} : H \rightarrow \mathbb{B}$  such that  $\sqrt{n}(\kappa_n(h) - \kappa_n(0)) \rightarrow \dot{\kappa}(h)$  for every  $h \in H$ ; a sequence of (Borel measurable) estimators  $\mathcal{T}_n$  is called *regular* if the limit law of  $\sqrt{n}(\mathcal{T}_n - \kappa_n(h))$  under  $P_{n,h}$  exists for every  $h \in H$  and is independent of  $h$ . Additionally, a function  $\ell : \mathbb{B} \rightarrow \mathbb{R}_+$  is called *subconvex* if for every  $c \in \mathbb{R}_+$ , the level set  $\ell^{-1}([0, c])$  is closed, convex, and symmetric.

**Proposition 3** (Asymptotic efficiency of empirical EOT map). *Consider the above setting. Then the following hold.*

- (i) (Convolution) *The sequences of parameters  $\kappa_n(h)$  and estimators  $\tilde{\mathcal{T}}_n$  are regular. For every regular sequence of Borel measurable estimators  $\mathcal{T}_n$  based on  $X_1^2, \dots, X_n^2$ , the limit law of  $\sqrt{n}(\mathcal{T}_n - T^\mu)$  under  $P_{n,0}$  equals the distribution of the sum  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) + W$  for some  $\mathbb{B}$ -valued random variable  $W$  independent of  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})$ .*
- (ii) (Local asymptotic minimaxity) *For every sequence of Borel measurable esti-*

mators  $\mathcal{T}_n$  based on  $X_1^2, \dots, X_n^2$  and every subconvex function  $\ell : \mathbb{B} \rightarrow \mathbb{R}_+$ ,

$$\sup_{I \subset H: \text{finite}} \liminf_{n \rightarrow \infty} \sup_{h \in I} \mathbb{E}_h \left[ \ell \left( \sqrt{n}(\mathcal{T}_n - \kappa_n(h)) \right) \right] \geq \mathbb{E} \left[ \ell(\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})) \right],$$

where  $\mathbb{E}_h$  denotes the expectation under  $P_{n,h}$ .

The regularity of the parameter sequence  $\kappa_n(h)$  follows from Hadamard differentiability of the map  $\nu^2 \mapsto T^{(\mu^1, \nu^2)}$ . The regularity of the empirical EOT map follows from the Hadamard differentiability result and Le Cam’s third lemma. The second claim of Proposition 3 (i) follows from applying Theorem 3.11.2 in [91]. To this effect, we need to verify that the law of the Gaussian variable appearing in the cited theorem agrees with that of  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})$  in our setting, which follows by adapting the argument in the proof of Proposition 2 in [48]. Given (i), Proposition 3 (ii) directly follows from Theorem 3.11.5 in [91].

Proposition 3 (i) shows that the limit law  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})$  of the empirical EOT map is the most concentrated around zero among all regular estimators for  $\kappa_n(h)$ . Furthermore, by regularity of the empirical EOT map, for every bounded continuous function  $\ell : \mathbb{B} \rightarrow \mathbb{R}_+$  and every finite set  $I \subset H$ ,

$$\lim_{n \rightarrow \infty} \sup_{h \in I} \mathbb{E}_h \left[ \ell \left( \sqrt{n}(\tilde{T}_n - \kappa_n(h)) \right) \right] = \mathbb{E} \left[ \ell(\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})) \right],$$

showing that  $\tilde{T}_n$  is asymptotically minimax in a local sense.

## 6. Proofs for Sections 3 and 4

### 6.1. Proof of Theorem 3

As noted in Section 4.1, we will first establish Hadamard differentiability of the map  $\nu \mapsto \varphi^\nu$  in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ .

**Lemma 4.** *Consider the setting of Theorem 3. Then, the map  $\nu \mapsto \varphi^\nu, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  is Hadamard differentiable at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ .*

The proof of Lemma 4 proceeds as follows. Fix  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  with  $S_i := \text{spt}(\mu^i)$  for  $i = 1, 2$ . For notational convenience, set  $S := S_1 \times S_2$  and  $\mathbb{D} := \mathcal{C}(S_1) \times \mathcal{C}(S_2)$ . We equip  $\mathbb{D}$  with a product norm,  $\|(\varphi_1, \varphi_2)\|_{\mathbb{D}} = \|\varphi_1\|_{\infty, S_1} \vee \|\varphi_2\|_{\infty, S_2}$ . Choose an arbitrary fixed reference point  $(x_1^\circ, x_2^\circ) \in S$ . With Lemma 1 in mind, consider

$$\Theta^s = \left\{ (\varphi_1|_{S_1}, \varphi_2|_{S_2}) : (\varphi_1, \varphi_2) \in \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X}), \right. \\ \left. \|\varphi_1\|_{\mathcal{C}^s(\mathcal{X})} \vee \|\varphi_2\|_{\mathcal{C}^s(\mathcal{X})} \leq R_s, \varphi_1(x_1^\circ) = \varphi_2(x_2^\circ) \right\} \subset \mathbb{D}.$$

Define the map  $\Psi : (\mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}) \times \Theta^s \rightarrow \mathbb{D}$  by

$$\Psi(\nu, \varphi) = \left( \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\nu^2 - 1, \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\nu^1 - 1 \right) \tag{11}$$

for  $\nu = (\nu^1, \nu^2) \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  and  $\varphi = (\varphi_1, \varphi_2) \in \Theta^s$ . Given  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$ , the corresponding EOT potentials  $\varphi^\nu$  are fully characterized by its restriction to  $S$ ,  $\varphi^\nu|_S := (\varphi_1^\nu|_{S_1}, \varphi_2^\nu|_{S_2}) \in \Theta^s$ , being the unique solution to  $\Psi(\nu, \varphi) \equiv 0$  on  $S$ . We will decompose the map  $\nu \mapsto \varphi^\nu$  into the composition of the maps  $\nu \mapsto \Psi(\nu, \cdot)$ ,  $\Psi(\nu, \cdot) \mapsto \varphi^\nu|_S$ , and  $\varphi^\nu|_S \mapsto \varphi^\nu$ , and separately show their Hadamard differentiability.

To formulate Hadamard differentiability of the map  $\Psi(\nu, \cdot) \mapsto \varphi^\nu|_S$ , we consider the following setting (cf. Section 3.9.4.7 in [91]; see also Appendix C.2). Let  $\ell^\infty(\Theta^s, \mathbb{D})$  be the Banach space of all uniformly norm-bounded maps  $z : \Theta^s \rightarrow \mathbb{D}$  equipped with the norm  $\|z\|_{\ell^\infty(\Theta^s, \mathbb{D})} = \sup_{\varphi \in \Theta^s} \|z(\varphi)\|_{\mathbb{D}}$ . Also, let  $Z(\Theta^s, \mathbb{D})$  be the subset of  $\ell^\infty(\Theta^s, \mathbb{D})$  consisting of all maps with at least one zero. Let  $\phi : Z(\Theta^s, \mathbb{D}) \rightarrow \Theta^s$  be a map that assigns each  $z \in Z(\Theta^s, \mathbb{D})$  to one of its zeros  $\phi(z)$ , i.e.,  $z(\phi(z)) = 0$ .

For a given  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$ ,  $\Psi(\nu, \cdot) : \varphi \mapsto \Psi(\nu, \varphi)$  is a uniformly norm-bounded map from  $\Theta^s$  into  $\mathbb{D}$  with a zero at  $\varphi^\nu|_S$ , guaranteeing that  $\Psi(\nu, \cdot) \in Z(\Theta^s, \mathbb{D})$ . Uniqueness of the EOT potentials then yields that

$$\varphi^\nu|_S = \phi \circ \Psi(\nu, \cdot).$$

Furthermore, in light of (6), we have the following representation of the EOT potentials  $\varphi^\nu$  for  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$ ,

$$\varphi^\nu = -\varepsilon \left( \log \left( \int e^{\frac{\varphi_2^\nu|_{S_2}(x_2) - c(\cdot, x_2)}{\varepsilon}} d\nu^2(x_2) \right), \log \left( \int e^{\frac{\varphi_1^\nu|_{S_1}(x_1) - c(x_1, \cdot)}{\varepsilon}} d\nu^1(x_1) \right) \right) \quad (12)$$

which depends on the potentials only through their restrictions to  $S_1$  and  $S_2$ .

We will establish Hadamard differentiability of  $\phi$  at  $\Psi(\mu, \cdot)$  by invoking Lemma 3.9.34 in [91] (see also Lemma 14). Hadamard differentiability of the map  $\nu \mapsto \Psi(\nu, \cdot)$  is straightforward. The chain rule for Hadamard differentiable maps (cf. Lemma 3.9.3 in [91]) then yields Hadamard differentiability of the map  $\nu \mapsto \varphi^\nu|_S$  in  $\mathbb{D}$ . Hadamard differentiability of  $\nu \mapsto \varphi^\nu$  in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  follows readily.

We shall first verify Hadamard differentiability of the map  $\nu \mapsto \Psi(\nu, \cdot)$ ,  $\mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow Z(\Theta^s, \mathbb{D}) \subset \ell^\infty(\Theta^s, \mathbb{D})$ . For notational convenience, define  $\psi : \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \rightarrow Z(\Theta^s, \mathbb{D})$  by  $\psi(\nu) = \Psi(\nu, \cdot)$ .

**Lemma 5.** *The map  $\psi : \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow Z(\Theta^s, \mathbb{D}) \subset \ell^\infty(\Theta^s, \mathbb{D})$  is Hadamard differentiable at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$  with derivative  $\psi' : \overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)} \rightarrow \ell^\infty(\Theta^s, \mathbb{D})$  given by*

$$\begin{aligned} \psi'(\gamma)(\varphi) &= \left( \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\gamma^2, \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\gamma^1 \right), \\ \varphi &= (\varphi_1, \varphi_2) \in \Theta^s, \gamma = (\gamma^1, \gamma^2) \in \overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}. \end{aligned} \quad (13)$$

As before,  $\gamma^j$  acts on the  $j$ -th coordinate, i.e.,

$$\int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\gamma^2 = \int e^{\frac{\varphi_1(\cdot) + \varphi_2(x_2) - c(\cdot, x_2)}{\varepsilon}} d\gamma^2(x_2).$$

*Proof.* Let  $\mathfrak{M}_i$  denote the space of finite signed Borel measures on  $\mathcal{X}$  supported in  $S_i = \text{spt}(\mu^i)$ . The definition (13) makes sense for  $\gamma \in \mathfrak{M}_1 \times \mathfrak{M}_2$ , and the extended map  $\psi' : \mathfrak{M}_1 \times \mathfrak{M}_2 \rightarrow \ell^\infty(\Theta^s, \mathbb{D})$  is linear. For every  $\varphi = (\varphi_1, \varphi_2) \in \Theta^s$ , let  $\bar{\varphi} = (\bar{\varphi}_1, \bar{\varphi}_2) \in \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$  denote an arbitrary  $\mathcal{C}^s$ -extension with  $\|\bar{\varphi}_1\|_{\mathcal{C}^s(\mathcal{X})} \vee \|\bar{\varphi}_2\|_{\mathcal{C}^s(\mathcal{X})} \leq R_s$ . Observe that

$$R' = \sup_{(x_1, x_2) \in \mathcal{X} \times \mathcal{X}} \sup_{(\varphi_1, \varphi_2) \in \Theta^s} \left\| e^{\frac{\bar{\varphi}_1(x_1) + \bar{\varphi}_2(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} \vee \left\| e^{\frac{\bar{\varphi}_1(\cdot) + \bar{\varphi}_2(x_2) - c(\cdot, x_2)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} < \infty. \tag{14}$$

Since  $\psi'(\gamma)(\varphi) = \psi'(\gamma)(\bar{\varphi})|_S$  for every  $\gamma \in \mathfrak{M}_1 \times \mathfrak{M}_2$ , we have

$$\|\psi'(\gamma)\|_{\ell^\infty(\Theta^s, \mathbb{D})} \leq R' (\|\gamma^1\|_{\infty, B^s} \vee \|\gamma^2\|_{\infty, B^s}). \tag{15}$$

Hence,  $\psi'$  extends uniquely to a continuous linear operator from  $\overline{\mathfrak{M}_1}^{\ell^\infty(B^s)} \times \overline{\mathfrak{M}_2}^{\ell^\infty(B^s)}$  into  $\ell^\infty(\Theta^s, \mathbb{D})$ .

Second, for every  $\gamma \in \overline{\mathfrak{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathfrak{M}_{\mu^2}}^{\ell^\infty(B^s)}$ , pick a sequence of pairs of signed Borel measures  $(\gamma_t)_{t>0}$  with total mass zero such that  $\mu + t\gamma_t \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  for sufficiently small  $t$  and  $\gamma_t \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  as  $t \downarrow 0$ . Then, as  $t \downarrow 0$ ,

$$\frac{\psi(\mu + t\gamma_t) - \psi(\mu)}{t} = \psi'(\gamma_t) \rightarrow \psi'(\gamma) \quad \text{in } \ell^\infty(\Theta^s, \mathbb{D}).$$

This completes the proof. □

Next, we shall establish Hadamard differentiability of the map  $\phi : Z(\Theta^s, \mathbb{D}) \subset \ell^\infty(\Theta^s, \mathbb{D}) \rightarrow \mathbb{D}$  at  $\Psi_\mu := \Psi(\mu, \cdot)$ . To this end, we apply Lemma 3.9.34 in [91]. The following lemma verifies the required conditions to apply the lemma. The proof, which we defer to Appendix B, relies on the results from [17].

**Lemma 6.** *The following hold.*

- (i) *The map  $\Psi_\mu : \Theta^s \ni \varphi \mapsto \Psi(\mu, \varphi) \in \mathbb{D}$  is injective and its inverse (defined on  $\Psi_\mu(\Theta^s)$ ) is continuous at 0.*
- (ii) *The map  $\Psi_\mu$  is Fréchet differentiable at  $\varphi = \varphi^\mu|_S$  with derivative  $\dot{\Psi}_\mu : \text{lin}(\Theta^s) \rightarrow \mathbb{D}$  given by*

$$\dot{\Psi}_\mu(\mathbf{h}) = \varepsilon^{-1} \left( \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2) d\mu^2, \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2) d\mu^1 \right)$$

for  $\mathbf{h} = (h_1, h_2) \in \text{lin}(\Theta^s)$ , where  $\text{lin}(\Theta^s)$  is the linear hull of  $\Theta^s$ . Furthermore,  $\dot{\Psi}_\mu : (\text{lin}(\Theta^s), \|\cdot\|_{\mathbb{D}}) \rightarrow \mathbb{D}$  is injective and its inverse is continuous.

Extend  $\dot{\Psi}_\mu^{-1}$  continuously to  $\overline{\dot{\Psi}_\mu(\text{lin}(\Theta^s))}^{\mathbb{D}}$ . Now, Lemma 3.9.34 in [91] implies the following. See also Appendix C.2.

**Lemma 7.** *The map  $\phi : Z(\Theta^s, \mathbb{D}) \subset \ell^\infty(\Theta^s, \mathbb{D}) \rightarrow \mathbb{D}$  is Hadamard differentiable at  $\Psi_\mu := \Psi(\mu, \cdot)$  tangentially to the set*

$$\mathcal{Z}_\mu = \left\{ z \in \ell^\infty(\Theta^s, \mathbb{D}) : z = \lim_{t \downarrow 0} \frac{z_t - \Psi_\mu}{t} \text{ for some } z_t \rightarrow \Psi_\mu \text{ in } Z(\Theta^s, \mathbb{D}), t \downarrow 0 \right\} \\ \cap \left\{ z \in \ell^\infty(\Theta^s, \mathbb{D}) : z \text{ is continuous at } \varphi^\mu|_S \right\}.$$

The derivative is given by  $\phi'_{\Psi_\mu}(z) = -\dot{\Psi}_\mu^{-1}(z(\varphi^\mu|_S))$ .

Hadamard differentiability of the map  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \varphi^\nu|_S \in \mathbb{D}$  at  $\mu$  now follows directly from Lemmas 5 and 7.

**Lemma 8.** *The map  $\nu \mapsto \varphi^\nu|_S, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathbb{D}$  is Hadamard differentiable at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ .*

*Proof.* Given Lemmas 5 and 7, the lemma follows from the chain rule for Hadamard differentiable maps; see Lemma 3.9.3 in [91]. The only thing we need to verify is that, for every  $\gamma = (\gamma^1, \gamma^2) \in \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ , it holds that  $\psi'(\gamma) \in \mathcal{Z}_\mu$ . Let  $(\mu_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  be a sequence such that  $\gamma_t := t^{-1}(\mu_t - \mu) \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ . Then  $\Psi_{\mu_t} \in Z(\Theta^s, \mathbb{D})$  and  $t^{-1}(\Psi_{\mu_t} - \Psi_\mu) = \psi'(\gamma_t) \rightarrow \psi'(\gamma)$  in  $\ell^\infty(\Theta^s, \mathbb{D})$ . It remains to show that  $\psi'(\gamma)$  is continuous at  $\varphi^\mu|_S$ . By construction, for every  $\eta > 0$ , there exists  $\tilde{\gamma} = (\tilde{\gamma}^1, \tilde{\gamma}^2) \in \mathcal{M}_{\mu^1} \times \mathcal{M}_{\mu^2}$  such that  $\|\gamma^1 - \tilde{\gamma}^1\|_{\ell^\infty(B^s)} \vee \|\gamma^2 - \tilde{\gamma}^2\|_{\ell^\infty(B^s)} \leq \eta$ . By (15), we have  $\|\psi'(\gamma) - \psi'(\tilde{\gamma})\|_{\ell^\infty(\Theta^s, \mathbb{D})} \leq R'\eta$  with  $R'$  given by (14). Since  $\tilde{\gamma}$  is a pair of signed measures, for every sequence  $\varphi_n \in \Theta^s$  with  $\|\varphi_n - \varphi^\mu|_S\|_{\mathbb{D}} \rightarrow 0$ , we see that  $\|\psi'(\tilde{\gamma})(\varphi_n) - \psi'(\tilde{\gamma})(\varphi^\mu|_S)\|_{\mathbb{D}} \rightarrow 0$ . We thus conclude that  $\limsup_{n \rightarrow \infty} \|\psi'(\gamma)(\varphi_n) - \psi'(\gamma)(\varphi^\mu|_S)\|_{\mathbb{D}} \leq 2R'\eta$ , and as  $\eta > 0$  is arbitrary, we have shown that  $\psi'(\gamma)$  is continuous at  $\varphi^\mu|_S$ . Hence, the chain rule applies, and the Hadamard derivative  $[\varphi^\mu|_S]' : \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}} \rightarrow \mathbb{D}$  is given by  $[\varphi^\mu|_S]'(\gamma) = -\dot{\Psi}_\mu^{-1}(\psi'(\gamma)(\varphi^\mu|_S))$ .  $\square$

Now, Lemma 8 and the formula for the potentials (12) together yield Lemma 4.

*Proof of Lemma 4.* We show Hadamard differentiability of the map  $\nu \mapsto \varphi_1^\nu$  at  $\mu$ ; differentiability of the second potential follows analogously. To simplify notation, define

$$\xi^\nu : (x_1, x_2) \in \mathcal{X} \times S_2 \mapsto e^{\frac{\varphi^\nu|_{S_2}(x_2) - c(x_1, x_2)}{\varepsilon}}.$$

By the formula (12) and the chain rule, it suffices to show that the map  $\Upsilon : \nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \int \xi^\nu(\cdot, x_2) d\nu^2(x_2) \in \mathcal{C}(\mathcal{X})$  is Hadamard differentiable at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ . To this effect, let  $(\mu_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  be such that  $\gamma_t := t^{-1}(\mu_t - \mu) \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ , and consider

$$t^{-1}(\Upsilon(\mu_t) - \Upsilon(\mu)) = \int t^{-1}(\xi^{\mu_t}(\cdot, x_2) - \xi^\mu(\cdot, x_2)) d\mu^2(x_2)$$



$$+ \int \xi^{\mu_t}(\cdot, x_2) d\gamma_t^2(x_2).$$

As  $t \downarrow 0$ , the first term on the right-hand side converges to the Hadamard derivative of the map  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \int \xi^\nu(\cdot, x_2) d\mu^2(x_2) \in \mathcal{C}(\mathcal{X})$  at  $\mu$ , which agrees with  $\varepsilon^{-1} \int \xi^\mu(\cdot, x_2) [\varphi_2^\mu|_{S_2}]'(\gamma)(x_2) d\mu^2(x_2)$  by Lemma 8 and the chain rule.

Pertaining to the second term, since  $\text{spt}(\gamma_t^2) \subset S_2$ , we have

$$\begin{aligned} & \sup_{x_1 \in \mathcal{X}} \left| \int \xi^{\mu_t}(x_1, x_2) d\gamma_t^2(x_2) - \int \xi^\mu(x_1, x_2) d\gamma^2(x_2) \right| \\ & \leq \sup_{x_1 \in \mathcal{X}} \left\| e^{\frac{\varphi_2^{\mu_t}(\cdot) - c(x_1, \cdot)}{\varepsilon}} - e^{\frac{\varphi_2^\mu(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2\|_{\infty, B^s} \\ & \quad + \sup_{x_1 \in \mathcal{X}} \left\| e^{\frac{\varphi_2^\mu(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2 - \gamma^2\|_{\infty, B^s}. \end{aligned}$$

The right-hand side converges to 0 as  $t \downarrow 0$  since  $\|\gamma_t^2\|_{\infty, B^s} = O(1)$ ,  $\|\gamma_t^2 - \gamma^2\|_{\infty, B^s} = o(1)$ , and  $\varphi_2^{\mu_t} \rightarrow \varphi_2^\mu$  in  $\mathcal{C}^s(\mathcal{X})$  by Lemma 1 (iii) (note that  $\mu_t \rightarrow \mu$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  and hence  $\mu_t^i \rightarrow \mu^i$  weakly for  $i = 1, 2$  by Lemma 16). Conclude that, for  $\gamma \in \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ ,

$$\Upsilon'_\mu(\gamma) = \varepsilon^{-1} \int \xi^\mu(\cdot, x_2) [\varphi_2^\mu|_{S_2}]'(\gamma)(x_2) d\mu^2(x_2) + \int \xi^\mu(\cdot, x_2) d\gamma^2(x_2), \quad (16)$$

and hence  $[\varphi_1^\mu]'(\gamma) = -\varepsilon e^{\frac{\varphi_1^\mu}{\varepsilon}} \Upsilon'_\mu(\gamma)$  by the chain rule. □

*Remark 12* (Compatibility of derivatives of EOT potentials). As Lemmas 8 and 4 establish, respectively, the Hadamard derivatives of the maps  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \varphi^\nu|_S \in \mathbb{D}$  and  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \varphi^\nu \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$ , the latter derivative should extend the former. Indeed, for every  $\gamma \in \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  and  $(\mu_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  for which  $t^{-1}(\mu_t - \mu) \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ ,  $t^{-1}(\varphi^{\mu_t} - \varphi^\mu) \rightarrow [\varphi^\mu]'(\gamma)$  in  $\mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  and hence also in  $\mathbb{D}$ . So,  $[\varphi^\mu]'(\gamma)|_S = [\varphi^\mu|_S]'(\gamma)$ , as desired.

*Remark 13* (Choice of reference point). We have chosen a reference point  $(x_1^\circ, x_2^\circ)$  from  $S$  in the proof of Lemma 4, but this is immaterial. Indeed, for a different choice of reference point  $(\tilde{x}_1^\circ, \tilde{x}_2^\circ) \in \mathcal{X} \times \mathcal{X}$ , the functions  $\tilde{\varphi}_1^\mu = \varphi_1^\mu - \frac{1}{2}(\varphi_1(\tilde{x}_1^\circ) - \varphi_2(\tilde{x}_1^\circ))$  and  $\tilde{\varphi}_2^\mu = \varphi_2^\mu + \frac{1}{2}(\varphi_1(\tilde{x}_1^\circ) - \varphi_2(\tilde{x}_1^\circ))$  are EOT potentials satisfying the constraint  $\tilde{\varphi}_1^\mu(\tilde{x}_1^\circ) = \tilde{\varphi}_2^\mu(\tilde{x}_2^\circ)$ . Clearly, by construction, the map  $\nu \mapsto \tilde{\varphi}^\nu, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$  is Hadamard differentiable at  $\mu$ . Hence the conclusion of Lemma 4 holds for an arbitrary reference point  $(x_1^\circ, x_2^\circ) \in \mathcal{X} \times \mathcal{X}$ .

We are now ready to prove Theorem 3.

*Proof of Theorem 3.* We divide the proof into two steps.

Step 1. Pick any multi-index  $k = (k_1, \dots, k_d) \in \mathbb{N}_0^d$  with  $0 < |k| \leq s$ . In what follows,  $i \in \{1, 2\}$  is arbitrary. Also,  $\mu^{-1} = \mu^2$  and  $\mu^{-2} = \mu^1$ . Similar

conventions apply to  $x_{-i}$  etc. We will show that the map

$$\nu \mapsto D^k \varphi_i^\nu, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}(\mathcal{X})$$

is Hadamard differentiable at  $\mu$ . Observe that

$$e^{-\varphi_i^\nu(x_i)/\varepsilon} = \int e^{(\varphi_{-i}^\nu(x_{-i}) - c(x_1, x_2))/\varepsilon} d\nu^{-i}(x_{-i}),$$

so that by interchanging differentiation and integration,  $D_{x_i}^k (e^{-\varphi_i^\nu(x_i)/\varepsilon})$  can be expressed as a linear combination of functions of the form

$$\int \prod_{j=1}^J [D_{x_i}^{\ell_j} c(x_1, x_2)]^{m_j} \times e^{(\varphi_{-i}^\nu(x_{-i}) - c(x_1, x_2))/\varepsilon} d\nu^{-i}(x_{-i}),$$

where  $1 \leq J \leq |k|$ , and  $m_1, \dots, m_J \in \mathbb{N}$  and  $\ell_1, \dots, \ell_J \in \mathbb{N}_0^d \setminus \{0\}$  are such that  $m_1 \ell_1 + \dots + m_J \ell_J = k$ . Combining the fact that  $(\log y)^{(n)} = (-1)^{n+1} (n-1)! y^{-n}$  and the multivariate Faà di Bruno formula (cf. Theorem 2.1 in [25]), we see that  $D^k \varphi_i^\nu(x_i) = -\varepsilon D^k \log(e^{-\varphi_i^\nu(x_i)/\varepsilon})$  can be expressed as a linear combination of products of functions of the form

$$\begin{aligned} & \frac{\int \zeta(x_1, x_2) e^{(\varphi_{-i}^\nu(x_{-i}) - c(x_1, x_2))/\varepsilon} d\nu^{-i}(x_{-i})}{\int e^{(\varphi_{-i}^\nu(x_{-i}) - c(x_1, x_2))/\varepsilon} d\nu^{-i}(x_{-i})} \\ &= \int \zeta(x_1, x_2) e^{\frac{\varphi_1^\nu(x_1) + \varphi_2^\nu(x_2) - c(x_1, x_2)}{\varepsilon}} d\nu^{-i}(x_{-i}), \end{aligned}$$

where  $\zeta$  is a smooth function on  $\mathbb{R}^d \times \mathbb{R}^d$  that depends only on the cost  $c$  and multi-index  $k$ . For example, for  $y = (y_1, \dots, y_d)$ ,

$$\begin{aligned} \frac{\partial^3}{\partial y_{j_1} \partial y_{j_2} \partial y_{j_3}} \log(f(y)) &= \frac{\frac{\partial^3}{\partial y_{j_1} \partial y_{j_2} \partial y_{j_3}} f(y)}{f(y)} - \frac{\frac{\partial^2}{\partial y_{j_1} \partial y_{j_3}} f(y)}{f(y)} \frac{\frac{\partial}{\partial y_{j_2}} f(y)}{f(y)} \\ &\quad - \frac{\frac{\partial}{\partial y_{j_1}} f(y)}{f(y)} \frac{\frac{\partial^2}{\partial y_{j_2} \partial y_{j_3}} f(y)}{f(y)} + 2 \frac{\frac{\partial}{\partial y_{j_1}} f(y)}{f(y)} \frac{\frac{\partial}{\partial y_{j_2}} f(y)}{f(y)} \frac{\frac{\partial}{\partial y_{j_3}} f(y)}{f(y)}. \end{aligned}$$

Hence, it suffices to show that, for every smooth function  $\zeta$  on  $\mathbb{R}^d \times \mathbb{R}^d$ , the map

$$\begin{aligned} \nu \mapsto & \int \zeta(x_1, x_2) e^{\frac{\varphi_1^\nu(x_1) + \varphi_2^\nu(x_2) - c(x_1, x_2)}{\varepsilon}} d\nu^{-i}(x_{-i}), \\ & \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}(\mathcal{X}) \end{aligned}$$

is Hadamard differentiable at  $\mu$ .

Let  $(\mu_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  be a sequence such that  $\gamma_t := t^{-1}(\mu_t - \mu) \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  as  $t \downarrow 0$  with  $\gamma = (\gamma_1, \gamma_2)$ . Define

$$g_t(x_1, x_2) = \zeta(x_1, x_2) e^{\frac{\varphi_1^{\mu_t}(x_1) + \varphi_2^{\mu_t}(x_2) - c(x_1, x_2)}{\varepsilon}}, \quad t \geq 0$$

with  $\mu_0 = \mu$ . We have shown in Lemma 4 that  $t^{-1}(\varphi_i^{\mu_t} - \varphi_i^\mu) \rightarrow [\varphi_i^\mu]'(\gamma)$  in  $\mathcal{C}(\mathcal{X})$  as  $t \downarrow 0$ , so that

$$t^{-1}(g_t - g_0) \rightarrow \varepsilon^{-1}\{[\varphi_1^\mu]'(\gamma) \oplus [\varphi_2^\mu]'(\gamma)\}g_0 =: h(\gamma) \quad \text{in } \mathcal{C}(\mathcal{X} \times \mathcal{X}).$$

Observe that

$$\int g_t(x_1, x_2) d\mu_t^{-i}(x_{-i}) = \int g_t(x_1, x_2) d\mu^{-i}(x_{-i}) + t \int g_t(x_1, x_2) d\gamma_t^{-i}(x_{-i}).$$

As  $t \downarrow 0$ , we have

$$\begin{aligned} & t^{-1} \left\{ \int g_t(x_1, x_2) d\mu^{-i}(x_{-i}) - \int g_0(x_1, x_2) d\mu^{-i}(x_{-i}) \right\} \\ & \rightarrow \int h(\gamma)(x_1, x_2) d\mu^{-i}(x_{-i}) \quad \text{in } \mathcal{C}(\mathcal{X}). \end{aligned}$$

To control  $\int g_t(x_1, x_2) d\gamma_t^{-i}(x_{-i})$ , observe that

$$\begin{aligned} & \left| \int g_t(x_1, x_2) d\gamma_t^2(x_2) - \int g_0(x_1, x_2) d\gamma^2(x_2) \right| \\ & \leq \left| \int (g_t - g_0)(x_1, x_2) d\gamma_t^2(x_2) \right| + \left| \int g_0(x_1, x_2) d(\gamma_t^2 - \gamma^2)(x_2) \right| \\ & \leq \|g_t(x_1, \cdot) - g_0(x_1, \cdot)\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2\|_{\infty, B^s} + \|g_0(x_1, \cdot)\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2 - \gamma^2\|_{\infty, B^s}. \end{aligned}$$

We have  $\|\gamma_t^2\|_{\infty, B^s} = O(1)$  and  $\|\gamma_t^2 - \gamma^2\|_{\infty, B^s} = o(1)$  as  $t \downarrow 0$  by construction, and  $\sup_{x_1 \in \mathcal{X}} \|g_0(x_1, \cdot)\|_{\mathcal{C}^s(\mathcal{X})} < \infty$  by Lemma 1 (ii). Since  $\mu_t^i$  converges weakly to  $\mu^i$  as  $t \downarrow 0$  (as convergence in  $\ell^\infty(B^s)$  implies weak convergence; cf. Lemma 16), Lemma 1 (iii) implies that  $\varphi^{\mu_t} \rightarrow \varphi^\mu$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ , which in turn implies that  $\sup_{x_1 \in \mathcal{X}} \|g_t(x_1, \cdot) - g_0(x_1, \cdot)\|_{\mathcal{C}^s(\mathcal{X})} = o(1)$  as  $t \downarrow 0$ . Hence, we have

$$\int g_t(\cdot, x_2) d\gamma_t^2(x_2) \rightarrow \int g_0(\cdot, x_2) d\gamma^2(x_2) \quad \text{in } \mathcal{C}(\mathcal{X}).$$

Likewise, we have  $\int g_t(x_1, \cdot) d\gamma_t^1(x_1) \rightarrow \int g_0(x_1, \cdot) d\gamma^1(x_1)$  in  $\mathcal{C}(\mathcal{X})$ . Conclude that

$$\begin{aligned} & t^{-1} \left\{ \int g_t(x_1, x_2) d\mu_t^{-i}(x_{-i}) - \int g_0(x_1, x_2) d\mu^{-i}(x_{-i}) \right\} \\ & \rightarrow \int h(\gamma)(x_1, x_2) d\mu^{-i}(x_{-i}) + \int g_0(x_1, x_2) d\gamma^{-i}(x_{-i}) \quad \text{in } \mathcal{C}(\mathcal{X}). \end{aligned}$$

The limit is linear and continuous from  $\overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$  into  $\mathcal{C}(\mathcal{X})$ .

Step 2. As in Step 1, let  $(\mu_t)_{t>0} \subset \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}$  be a sequence such that  $\gamma_t := t^{-1}(\mu_t - \mu) \rightarrow \gamma$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  as  $t \downarrow 0$ . By Lemma 4 and Step 1, for every multi-index  $k \in \mathbb{N}_0^d$  with  $|k| \leq s$ , the map  $\nu \mapsto D^k \varphi_i^\nu, \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset$

$\ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathcal{C}(\mathcal{X})$  is Hadamard differentiable at  $\boldsymbol{\mu}$ . Denote its derivative by  $[D^k \varphi_i^\mu]'$ , so that

$$t^{-1}(D^k \varphi_i^{\mu_t} - D^k \varphi_i^\mu) \rightarrow [D^k \varphi_i^\mu]'(\boldsymbol{\gamma}) \quad \text{in } \mathcal{C}(\mathcal{X}).$$

Pick any sequence  $t_n \downarrow 0$ . Then,  $t_n^{-1}(\varphi_i^{\mu_{t_n}} - \varphi_i^\mu)$  is Cauchy in  $\mathcal{C}^s(\mathcal{X})$ , so by completeness of  $\mathcal{C}^s(\mathcal{X})$ , the limit in  $\mathcal{C}^s(\mathcal{X})$  exists, i.e.,  $t_n^{-1}(\varphi_i^{\mu_{t_n}} - \varphi_i^\mu) \rightarrow \bar{\varphi}_i$  in  $\mathcal{C}^s(\mathcal{X})$ . The limit  $\bar{\varphi}_i$  satisfies that  $D^k \bar{\varphi}_i = [D^k \varphi_i^\mu]'(\boldsymbol{\gamma})$  for every multi-index  $k \in \mathbb{N}_0^d$  with  $|k| \leq s$ , which shows that  $[\varphi_i^\mu]'(\boldsymbol{\gamma}) \in \mathcal{C}^s(\mathcal{X})$  with  $D^k [\varphi_i^\mu]'(\boldsymbol{\gamma}) = [D^k \varphi_i^\mu]'(\boldsymbol{\gamma})$  for every multi-index  $k \in \mathbb{N}_0^d$  with  $|k| \leq s$ . Since the map  $\boldsymbol{\gamma} \mapsto [\varphi_i^\mu]'(\boldsymbol{\gamma})$  is linear and continuous from  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  into  $\mathcal{C}^s(\mathcal{X})$ , we obtain the desired result.  $\square$

*Remark 14.* Another possible approach would be to employ the implicit function theorem for Banach spaces (see, e.g., Theorem I.5.9 in [58]), which asks Fréchet differentiability of the map  $\boldsymbol{\mu} \mapsto \boldsymbol{\varphi}^\mu$ . However, in our problem, it seems highly nontrivial to verify the required conditions to directly apply the implicit function theorem. For statistical purposes, Hadamard differentiability is sufficient in most cases; cf. Chapter 3.9 in [91].

## 6.2. Proof of Theorem 4

We divide the proof into two steps.

**Step 1.** We first show twice Hadamard differentiability of the mapping  $\boldsymbol{\nu} \mapsto \boldsymbol{\varphi}^\nu$  in  $\mathbb{D}$ . Recall that  $\boldsymbol{\varphi}^{\mu_t}|_S = \phi \circ \Psi(\boldsymbol{\mu}_t, \cdot)$ . Observe that  $\Psi(\boldsymbol{\mu}_t, \cdot) = \Psi(\boldsymbol{\mu}, \cdot) + t\psi'(\boldsymbol{\gamma}_t)$ . We will apply Lemma 15 below with  $\Theta = \Theta^s, \mathbb{D} = \mathbb{D}, \mathbb{L} = \mathbb{D}, \Psi = \Psi(\boldsymbol{\mu}, \cdot), \theta_0 = \boldsymbol{\varphi}^\mu|_S, \theta_t = \boldsymbol{\varphi}^{\mu_t}|_S$ , and  $z_t = \psi'(\boldsymbol{\gamma}_t)$ . To this end, we shall verify the conditions in Lemma 15.

Twice Fréchet differentiability of  $\Psi(\boldsymbol{\mu}, \cdot)$  is straightforward to verify, with second derivative given by

$$\ddot{\Psi}_\mu(h_1, h_2) = \varepsilon^{-2} \left( \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2)^2 d\mu^2, \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2)^2 d\mu^1 \right).$$

We have already verified Conditions (i)–(ii) of Lemma 15 in the proof of Lemma 8. Regarding Condition (iii), since  $\psi'(\boldsymbol{\gamma}_t) \in \mathcal{Z}_\mu$  (cf. the proof of Lemma 8), in view of Remark 17 after Lemma 14, it holds that  $\psi'(\boldsymbol{\gamma}_t)(\boldsymbol{\varphi}^\mu|_S) \in \overline{\ddot{\Psi}_\mu(\text{lin}(\Theta^s))}^{\mathbb{D}}$ . It remains to verify Condition (iv) in Lemma 15. Observe that, as  $\text{spt}(\boldsymbol{\gamma}_t^i) \subset S_i$ ,

$$\begin{aligned} & t^{-1} \{ \psi'(\boldsymbol{\gamma}_t)(\boldsymbol{\varphi}^{\mu_t}|_S) - \psi'(\boldsymbol{\gamma}_t)(\boldsymbol{\varphi}^\mu|_S) \} \\ &= t^{-1} \left( \int \left( e^{\frac{\varphi_1^{\mu_t} \oplus \varphi_2^{\mu_t} - c}{\varepsilon}} - e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} \right) d\boldsymbol{\gamma}_t^2, \int \left( e^{\frac{\varphi_1^{\mu_t} \oplus \varphi_2^{\mu_t} - c}{\varepsilon}} - e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} \right) d\boldsymbol{\gamma}_t^1 \right). \end{aligned}$$

By symmetry, it suffices to show convergence of the first coordinate. Recall from Theorem 3 that  $t^{-1}(\varphi_i^{\mu_t} - \varphi_i^\mu) \rightarrow [\varphi_i^\mu]'(\boldsymbol{\gamma})$  in  $\mathcal{C}^s(\mathcal{X})$  for  $i = 1, 2$ . We shall show

that

$$\sup_{x_1 \in \mathcal{X}} \left\| t^{-1} \left\{ e^{\frac{\varphi_1^{\mu t}(x_1) + \varphi_2^{\mu t}(\cdot) - c(x_1, \cdot)}{\varepsilon}} - e^{\frac{\varphi_1^{\mu}(x_1) + \varphi_2^{\mu}(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\} - \varepsilon^{-1} \{ [\varphi_1^{\mu}]'(\gamma)(x_1) + [\varphi_2^{\mu}]'(\gamma)(\cdot) \} e^{\frac{\varphi_1^{\mu}(x_1) + \varphi_2^{\mu}(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} \rightarrow 0. \tag{17}$$

Indeed, since  $\gamma_t^2 \rightarrow \gamma^2$  in  $\ell^\infty(B^s)$ , (17) implies that

$$\int t^{-1} \left( e^{\frac{\varphi_1^{\mu t} \oplus \varphi_2^{\mu t} - c}{\varepsilon}} - e^{\frac{\varphi_1^{\mu} \oplus \varphi_2^{\mu} - c}{\varepsilon}} \right) d\gamma_t^2 \rightarrow \varepsilon^{-1} \int \{ [\varphi_1^{\mu}]'(\gamma) \oplus [\varphi_2^{\mu}]'(\gamma) \} e^{\frac{\varphi_1^{\mu} \oplus \varphi_2^{\mu} - c}{\varepsilon}} d\gamma^2$$

in  $\mathcal{C}(\mathcal{X})$ , yielding the desired result in view of Remark 9.

Observe that, in general, for  $f, g \in \mathcal{C}^s(\mathcal{X})$ , it holds that

$$\|fg\|_{\mathcal{C}^s(\mathcal{X})} \leq K \|f\|_{\mathcal{C}^s(\mathcal{X})} \|g\|_{\mathcal{C}^s(\mathcal{X})}$$

for some constant  $K$  that depends only on  $s, d$ . Now, since

$$\sup_{x_1 \in \mathcal{X}} \left\| e^{\frac{\varphi_1^{\mu}(x_1) + \varphi_2^{\mu}(\cdot) - c(x_1, \cdot)}{\varepsilon}} \right\|_{\mathcal{C}^s(\mathcal{X})} < \infty$$

by Lemma 1 (ii), the left-hand side of (17) is bounded by

$$\sup_{x_1 \in \mathcal{X}} \left\| t^{-1} \left\{ e^{\frac{\varphi_1^{\mu t}(x_1) - \varphi_1^{\mu}(x_1) + \varphi_2^{\mu t}(\cdot) - \varphi_2^{\mu}(\cdot)}{\varepsilon}} - 1 \right\} - \varepsilon^{-1} \{ [\varphi_1^{\mu}]'(\gamma)(x_1) + [\varphi_2^{\mu}]'(\gamma)(\cdot) \} \right\|_{\mathcal{C}^s(\mathcal{X})} \tag{18}$$

up to a constant independent of  $t$ . It is not difficult to see that (18) converges to zero when  $\|\cdot\|_{\mathcal{C}^s(\mathcal{X})}$  is replaced by  $\|\cdot\|_{\infty, \mathcal{X}}$ . We shall show that for every multi-index  $k \in \mathbb{N}_0^d$  with  $0 < |k| \leq s$ ,

$$\sup_{x_1 \in \mathcal{X}} \left\| t^{-1} D_{x_2}^k \left( e^{\frac{\varphi_1^{\mu t}(x_1) - \varphi_1^{\mu}(x_1) + \varphi_2^{\mu t}(\cdot) - \varphi_2^{\mu}(\cdot)}{\varepsilon}} \right) - \varepsilon^{-1} D_{x_2}^k [\varphi_2^{\mu}]'(\gamma)(\cdot) \right\|_{\infty, \mathcal{X}} \rightarrow 0. \tag{19}$$

By the multivariate Faà di Bruno formula (cf. Theorem 2.1 in [25]),

$$t^{-1} D_{x_2}^k \left( e^{\frac{\varphi_1^{\mu t}(x_1) - \varphi_1^{\mu}(x_1) + \varphi_2^{\mu t}(x_2) - \varphi_2^{\mu}(x_2)}{\varepsilon}} \right)$$

can be expressed as a linear combination of functions of the form

$$e^{\frac{\varphi_1^{\mu t}(x_1) - \varphi_1^{\mu}(x_1) + \varphi_2^{\mu t}(x_2) - \varphi_2^{\mu}(x_2)}{\varepsilon}} \times t^{-1} \prod_{j=1}^J \left[ D_{x_2}^{\ell_j} (\varphi_2^{\mu t}(x_2) - \varphi_2^{\mu}(x_2)) \right]^{m_j}, \tag{20}$$

where  $1 \leq J \leq |k|$ , and  $m_1, \dots, m_J \in \mathbb{N}$  and  $\ell_1, \dots, \ell_J \in \mathbb{N}_0^d \setminus \{0\}$  are such that  $m_1 \ell_1 + \dots + m_J \ell_J = k$ . The coefficient for the leading term with  $J = 1$  and  $m_1 = 1$  is  $\varepsilon^{-1}$ . Except for the leading term, the order of (20) is  $O(t)$  as  $t \downarrow 0$  in  $\mathcal{C}(\mathcal{X} \times \mathcal{X})$ , so we arrive at the following expansion in  $\mathcal{C}(\mathcal{X} \times \mathcal{X})$ :

$$\begin{aligned} & t^{-1} D_{x_2}^k \left( e^{\frac{\varphi_1^{\mu_t}(x_1) - \varphi_1^\mu(x_1) + \varphi_2^{\mu_t}(x_2) - \varphi_2^\mu(x_2)}{\varepsilon}} \right) \\ &= \varepsilon^{-1} e^{\frac{\varphi_1^{\mu_t}(x_1) - \varphi_1^\mu(x_1) + \varphi_2^{\mu_t}(x_2) - \varphi_2^\mu(x_2)}{\varepsilon}} \times t^{-1} D_{x_2}^k (\varphi_2^{\mu_t}(x_2) - \varphi_2^\mu(x_2)) + o(1). \end{aligned}$$

The right-hand side converges to  $\varepsilon^{-1} D_{x_2}^k [\varphi_2^\mu]'(\gamma)(x_2)$  in  $\mathcal{C}(\mathcal{X} \times \mathcal{X})$  as  $t^{-1}(\varphi_i^{\mu_t} - \varphi_i^\mu) \rightarrow [\varphi_i^\mu]'(\gamma)$  in  $\mathcal{C}^s(\mathcal{X})$  for  $i = 1, 2$ , which leads to (19) as desired.

Therefore, Lemma 15 guarantees that the limit

$$\lim_{t \downarrow 0} \frac{\varphi^{\mu_t}|_S - \varphi^\mu|_S - t[\varphi^\mu|_S]'(\gamma_t)}{t^2/2}$$

exists in  $\mathbb{D}$ . By our construction, the limit depends on  $\gamma$  but not on the choice of sequence  $\mu_t$ , so denote the limit by  $[\varphi^\mu|_S]''(\gamma)$ .

Step 2. Next, we leverage twice differentiability of  $\nu \mapsto \varphi^\nu|_S$  to establish the second derivative of  $\nu \mapsto \varphi^\nu$  using (12). As in the proof of Lemma 4, we only deal with the first potential. Recall the notation  $\xi^\nu$  and  $\Upsilon$  that appeared in the proof of Lemma 4 and observe that  $\varphi_1^\nu = -\varepsilon \log(\Upsilon(\nu))$ . Precisely,  $\varphi_1^\nu$  agrees with the composition of the maps  $f \in \mathcal{C}_+(\mathcal{X}) \mapsto -\varepsilon \log(f) \in \mathcal{C}(\mathcal{X})$  with  $\mathcal{C}_+(\mathcal{X}) = \{f \in \mathcal{C}(\mathcal{X}) : f > 0\}$  and  $\Upsilon : \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \rightarrow \mathcal{C}_+(\mathcal{X})$ . The former map is twice Hadamard (indeed Fréchet) differentiable and its domain  $\mathcal{C}_+(\mathcal{X})$  is open in  $\mathcal{C}(\mathcal{X})$ , so the tangent cone  $\mathfrak{T}_{\mathcal{C}_+(\mathcal{X})}(f)$  agrees with  $\mathcal{C}(\mathcal{X})$  for every  $f \in \mathcal{C}_+(\mathcal{X})$ . From the second-order chain rule for Hadamard differentiable maps (see Lemma 12), it suffices to establish second-order Hadamard differentiability of  $\Upsilon$  at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$ . As such, with  $\mu_t$  and  $\gamma_t$  as above, recalling the expression of the derivative of  $\Upsilon$  from (16) and that  $[\varphi_2^\mu|_{S_2}]'(\gamma_t) = [\varphi_2^{\mu_t}]'(\gamma_t)$  on  $S_2$  (see Remark 12), we have

$$\begin{aligned} & \frac{\Upsilon(\mu_t) - \Upsilon(\mu) - t\Upsilon'_\mu(\gamma_t)}{t^2/2} \\ &= (2/t^2) \int \{ \xi^{\mu_t}(\cdot, x_2) - \xi^\mu(\cdot, x_2) - t\varepsilon^{-1} \xi^\mu(\cdot, x_2) [\varphi_2^{\mu_t}]'(\gamma_t)(x_2) \} d\mu^2(x_2) \\ & \quad + (2/t) \int \{ \xi^{\mu_t}(\cdot, x_2) - \xi^\mu(\cdot, x_2) \} d\gamma_t^2(x_2), \end{aligned}$$

As  $t \downarrow 0$ , the first term on the right-hand side converges to the second-order Hadamard derivative of the map  $\nu \in \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \mapsto \int \xi^\nu(\cdot, x_2) d\mu^2(x_2) \in \mathcal{C}(\mathcal{X})$  at  $\mu$ , which agrees with

$$\int \left\{ \varepsilon^{-2} \xi^\mu(\cdot, x_2) ([\varphi_2^\mu]'(\gamma)(x_2))^2 + \varepsilon^{-1} \xi^\mu(\cdot, x_2) [\varphi_2^\mu|_{S_2}]''(\gamma)(x_2) \right\} d\mu^2(x_2)$$

by the chain rule and twice differentiability of the restricted potentials.

As  $(x_1, x_2) \in \mathcal{X} \times \mathcal{X} \mapsto e^{\frac{\varphi_2^\mu(x_2) - c(x_1, x_2)}{\varepsilon}}$  extends  $\xi^\nu$ , the second term satisfies

$$\begin{aligned} & \sup_{x_1 \in \mathcal{X}} \left| t^{-1} \int \{ \xi^{\mu^t}(x_1, \cdot) - \xi^\mu(x_1, \cdot) \} d\gamma_t^2 - \varepsilon^{-1} \int \xi^\mu(x_1, \cdot) [\varphi_2^\mu]'(\gamma)(\cdot) d\gamma^2 \right| \\ & \leq \sup_{x_1 \in \mathcal{X}} \left\| e^{\frac{\varphi_2^\mu - c(x_1, \cdot)}{\varepsilon}} \left( t^{-1} \left( e^{\frac{\varphi_2^\mu t - \varphi_2^\mu}{\varepsilon}} - 1 \right) - \varepsilon^{-1} [\varphi_2^\mu]'(\gamma)(\cdot) \right) \right\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2\|_{\infty, B^s} \\ & \quad + \sup_{x_1 \in \mathcal{X}} \left\| \varepsilon^{-1} e^{\frac{\varphi_2^\mu - c(x_1, \cdot)}{\varepsilon}} [\varphi_2^\mu]'(\gamma)(\cdot) \right\|_{\mathcal{C}^s(\mathcal{X})} \|\gamma_t^2 - \gamma^2\|_{\infty, B^s}. \end{aligned} \tag{21}$$

As in Step 1, one can show that the right-hand side of (21) converges to 0 as  $t \downarrow 0$ .

Consequently,  $\Upsilon$  is twice Hadamard differentiable at  $\mu$  with derivative

$$\begin{aligned} \Upsilon''_\mu(\gamma) &= \int \varepsilon^{-2} \xi^\mu(\cdot, x_2) \left( [\varphi_2|_{S_2}^\mu]'(\gamma)(x_2) \right)^2 d\mu^2(x_2) \\ & \quad + \int \varepsilon^{-1} \xi^\mu(\cdot, x_2) [\varphi_2|_{S_2}^\mu]''(\gamma)(x_2) d\mu^2(x_2) \\ & \quad + 2\varepsilon^{-1} \int \xi^\mu(x_1, \cdot) [\varphi_2|_{S_2}^\mu]'(\gamma)(\cdot) d\gamma^2, \quad \gamma \in \overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}, \end{aligned}$$

and hence  $[\varphi_1^\mu]''(\gamma) = \varepsilon e^{\frac{2\varphi_1^\mu}{\varepsilon}} (\Upsilon'_\mu(\gamma))^2 - \varepsilon e^{\frac{\varphi_1^\mu}{\varepsilon}} \Upsilon''_\mu(\gamma)$  by the chain rule. Finally, continuity and positive homogeneity (of degree 2) of  $[\varphi^\mu]''$  follow from the construction.  $\square$

### 6.3. Proof of Lemma 3

We divide the proof into two steps. Since the Sinkhorn divergence is invariant w.r.t. the choice of reference points and  $\gamma^i(a) = 0$  for every constant  $a \in \mathbb{R}$ , we may assume without loss of generality that  $x_1^\circ = x_2^\circ$ .

Step 1. The map  $\nu = (\nu^1, \nu^2) \mapsto S_{c, \varepsilon}(\nu^1, \nu^2), \mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2} \subset \ell^\infty(B^s) \times \ell^\infty(B^s) \rightarrow \mathbb{R}$  is Hadamard differentiable at  $\mu$  tangentially to  $\overline{\mathcal{M}_{\mu^1}}^{\ell^\infty(B^s)} \times \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$  with derivative

$$[S_{c, \varepsilon}^\mu]'(\gamma) = \int \varphi_1^\mu d\gamma^1 + \int \varphi_2^\mu d\gamma^2.$$

Given regularity of EOT potentials (Lemma 1), the proof follows similarly to the proof (or its argument) of Theorem 7 in [48], with small modifications. To avoid repetitions, we omit the details.

Step 2. The conclusion of the lemma follows by Step 1 and noting that  $\varphi_1^{(\mu^i, \mu^i)} = \varphi_2^{(\mu^i, \mu^i)}$  by symmetry of the cost function.  $\square$

#### 6.4. Proof of Theorem 5

As before, we may assume without loss of generality that  $x_1^\circ = x_2^\circ$ . Using (9), we can expand  $\bar{S}_{c,\varepsilon}(\mu_t^1, \mu_t^2)$  as

$$\begin{aligned}\bar{S}_{c,\varepsilon}(\mu_t^1, \mu_t^2) &= \int (\varphi_1^{(\mu_t^1, \mu_t^2)} - \varphi_1^{(\mu_t^1, \mu_t^1)}) d\mu_t^1 + \int (\varphi_2^{(\mu_t^1, \mu_t^2)} - \varphi_2^{(\mu_t^2, \mu_t^2)}) d\mu_t^2 \\ &= \int (\varphi_1^{(\mu_t^1, \mu_t^2)} - \varphi_1^{(\mu_t^1, \mu_t^1)}) d\mu + \int (\varphi_2^{(\mu_t^1, \mu_t^2)} - \varphi_2^{(\mu_t^2, \mu_t^2)}) d\mu \\ &\quad + t \int (\varphi_1^{(\mu_t^1, \mu_t^2)} - \varphi_1^{(\mu_t^1, \mu_t^1)}) d\gamma_t^1 + t \int (\varphi_2^{(\mu_t^1, \mu_t^2)} - \varphi_2^{(\mu_t^2, \mu_t^2)}) d\gamma_t^2.\end{aligned}$$

By Theorem 3, we have

$$\frac{\varphi_i^{(\mu_t^1, \mu_t^2)} - \varphi_i^{(\mu, \mu)}}{t} \rightarrow [\varphi_i^{(\mu, \mu)}]'(\gamma^1, \gamma^2) \quad \text{in } \mathcal{C}^s(\mathcal{X})$$

as  $t \downarrow 0$ , while

$$\frac{\varphi_i^{(\mu_t^i, \mu_t^i)} - \varphi_i^{(\mu, \mu)}}{t} \rightarrow [\varphi_i^{(\mu, \mu)}]'(\gamma^i, \gamma^i) \quad \text{in } \mathcal{C}^s(\mathcal{X}).$$

Since  $\gamma_t^i \rightarrow \gamma^i$  in  $\ell^\infty(B^s)$  as  $t \downarrow 0$ , we have

$$\int (\varphi_i^{(\mu_t^1, \mu_t^2)} - \varphi_i^{(\mu_t^i, \mu_t^i)}) d\gamma_t^i = t \int [\varphi_i^{(\mu, \mu)}]'(\gamma^1 - \gamma^i, \gamma^2 - \gamma^i) d\gamma^i + o(t).$$

On the other hand, by Theorem 4, we have

$$\begin{aligned}\frac{\varphi_1^{(\mu_t^1, \mu_t^2)} - \varphi_1^{(\mu_t^1, \mu_t^1)} - t[\varphi_1^{(\mu, \mu)}]'(0, \gamma_t^2 - \gamma_t^1)}{t^2/2} \\ \rightarrow [\varphi_1^{(\mu, \mu)}]''(\gamma^1, \gamma^2) - [\varphi_1^{(\mu, \mu)}]''(\gamma^1, \gamma^1) \quad \text{in } \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X}).\end{aligned}$$

A similar expansion holds for  $\varphi_2^{(\mu_t^1, \mu_t^2)} - \varphi_2^{(\mu_t^2, \mu_t^2)}$ . Hence, we have

$$\begin{aligned}\int (\varphi_1^{(\mu_t^1, \mu_t^2)} - \varphi_1^{(\mu_t^1, \mu_t^1)}) d\mu + \int (\varphi_2^{(\mu_t^1, \mu_t^2)} - \varphi_2^{(\mu_t^2, \mu_t^2)}) d\mu \\ = t \int ([\varphi_1^{(\mu, \mu)}]'(0, \gamma_t^2 - \gamma_t^1) + [\varphi_2^{(\mu, \mu)}]'(\gamma_t^1 - \gamma_t^2, 0)) d\mu \\ + \frac{t^2}{2} \int \left( \sum_{i=1}^2 ([\varphi_i^{(\mu, \mu)}]''(\gamma^1, \gamma^2) - [\varphi_i^{(\mu, \mu)}]''(\gamma^i, \gamma^i)) \right) d\mu + o(t^2).\end{aligned}\tag{22}$$

As the cost function is symmetric, we have

$$[\varphi_2^{(\mu, \mu)}]'(\gamma_t^1 - \gamma_t^2, 0) = [\varphi_1^{(\mu, \mu)}]'(0, \gamma_t^1 - \gamma_t^2) = -[\varphi_1^{(\mu, \mu)}]'(0, \gamma_t^2 - \gamma_t^1),$$



so that the first term on the right-hand side of (22) vanishes. Conclude that

$$\begin{aligned} \lim_{t \downarrow 0} \frac{\bar{S}_{c,\varepsilon}(\mu_t^1, \mu_t^2)}{t^2/2} &= \int \left( \sum_{i=1}^2 ([\varphi_i^{(\mu,\mu)}]''(\gamma^1, \gamma^2) - [\varphi_i^{(\mu,\mu)}]''(\gamma^i, \gamma^i)) \right) d\mu \\ &\quad + 2 \sum_{i=1}^2 \int [\varphi_i^{(\mu,\mu)}]'(\gamma^1 - \gamma^i, \gamma^2 - \gamma^i) d\gamma^i =: \Delta_\mu(\gamma). \end{aligned}$$

Continuity and positive homogeneity (of degree 2) of  $\Delta_\mu$  follows immediately (or from its construction).  $\square$

**6.5. Proofs for Section 3 (except Proposition 2)**

In what follows, let

$$\mathfrak{M}_{\mu^i} = \left\{ g d\mu^i : g : \mathcal{X} \rightarrow \mathbb{R} \text{ is bounded and measurable with } \mu^i\text{-mean zero} \right\},$$

where  $g d\mu^i$  should be understood as a signed measure  $A \mapsto \int_A g d\mu^i$ .

For a (generic) probability measure  $\mu$  and a function class  $\mathcal{F} \subset L^2(\mu)$ , a stochastic process  $G = (G(f))_{f \in \mathcal{F}}$  is called a  $\mu$ -Brownian bridge if it is a Gaussian process with mean zero and covariance function  $\mathbb{E}[G(f)G(g)] = \text{Cov}_\mu(f, g)$ ; furthermore, if  $G$  is a tight measurable map into  $\ell^\infty(\mathcal{F})$ , then we call  $G$  a tight  $\mu$ -Brownian bridge in  $\ell^\infty(\mathcal{F})$ . Recall that a (zero-mean) Gaussian process that is a tight measurable map into  $\ell^\infty(\mathcal{F})$  is an  $\ell^\infty(\mathcal{F})$ -valued Gaussian random variable (with mean zero) in the Banach space sense; see Lemma 3.9.8 in [91].

**Lemma 9.** *Let  $s$  be a positive integer with  $s > d/2$ . Then, for every  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ , we have*

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} \mathbb{G}^\mu = (\mathbb{G}_1^{\mu^1}, \mathbb{G}_2^{\mu^2}) \quad \text{in } \ell^\infty(B^s) \times \ell^\infty(B^s), \quad (23)$$

where  $\mathbb{G}_1^{\mu^1}$  and  $\mathbb{G}_2^{\mu^2}$  are independent, tight  $\mu^1$ - and  $\mu^2$ -Brownian bridges in  $\ell^\infty(B^s)$  and  $\ell^\infty(B^s)$ , respectively. Furthermore,  $\text{spt}(\mathbb{G}^\mu) \subset \overline{\mathfrak{M}_{\mu^1}^{\ell^\infty(B^s)}} \times \overline{\mathfrak{M}_{\mu^2}^{\ell^\infty(B^s)}}$ .

*Proof of Lemma 9.* The set  $B^s$  is  $\mu^i$ -Donsker by Theorem 2.7.1 in [91]. Since the samples are independent, by Example 1.4.6 in [91] (combined with Lemma 3.2.4 in [35] concerning measurable covers), we obtain the weak convergence result (23). The second claim follows by Lemma 17.  $\square$

The following lemma will be used to prove the second claim of Theorem 2.

**Lemma 10.** *Consider the setting of Theorem 5 and assume that  $\bar{S}_{c,\varepsilon}$  is non-negative. Let  $W$  be a tight random variable with values in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  whose support is a cone  $\neq \{0\}$ . Then, unless  $\Delta_\mu$  is identically zero, the support of  $\Delta_\mu(W)$  agrees with  $[0, \infty)$ .*

*Proof of Lemma 10.* Consider the restriction of  $\Delta_\mu$  on  $\text{spt}(W)$ , which we denote by the same symbol. The restriction is still continuous and positively homogeneous of degree 2 (the latter follows as  $\text{spt}(W)$  is a cone). The functional  $\Delta_\mu$  is nonnegative by construction. By positive homogeneity,  $\Delta_\mu$  is either identically zero or onto  $[0, \infty)$ . In our setting,  $\Delta_\mu$  is onto  $[0, \infty)$ . For every open interval  $(a, b) \subset [0, \infty)$ , the inverse image  $\Delta_\mu^{-1}((a, b))$  is nonempty and open in  $\text{spt}(W)$ , and hence by the definition of support, we have  $\mathbb{P}(\Delta_\mu(W) \in (a, b)) > 0$ , which implies that  $\text{spt}(\Delta_\mu(W)) = [0, \infty)$ .  $\square$

We are now ready to prove Theorem 1, Corollary 1, Proposition 1, and Theorem 2.

*Proof of Theorem 1.* The theorem follows from the Hadamard differentiability result (Theorem 3) combined with the functional delta method. Let  $s > d/2$ , so that the weak convergence (23) holds. Since  $\sqrt{n}(\hat{\mu}_n - \mu) \in \overline{\mathcal{P}_{\mu^1} \times \mathcal{P}_{\mu^2}}^{\ell^\infty(B^s)}$  a.s. (as  $\text{spt}(\hat{\mu}_n^i) \subset \text{spt}(\mu^i)$  a.s.) and  $\text{spt}(\mathbb{G}^\mu) \subset \overline{\mathfrak{M}_{\mu^1}^{\ell^\infty(B^s)} \times \mathfrak{M}_{\mu^2}^{\ell^\infty(B^s)}} \subset \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)} \times \mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  (or the inclusion  $\text{spt}(\mathbb{G}^\mu) \subset \overline{\mathcal{M}_{\mu^1}^{\ell^\infty(B^s)} \times \mathcal{M}_{\mu^2}^{\ell^\infty(B^s)}}$  follows from the portmanteau theorem), we may apply the functional delta method (Lemma 11) to conclude that

$$\sqrt{n}(\hat{\varphi}_n - \varphi^\mu) \xrightarrow{d} [\varphi^\mu]'(\mathbb{G}^\mu) \quad \text{in } \mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X}).$$

Also, since  $\overline{\mathfrak{M}_{\mu^1}^{\ell^\infty(B^s)} \times \mathfrak{M}_{\mu^2}^{\ell^\infty(B^s)}}$  is a closed subspace of  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  and the restriction of  $[\varphi^\mu]'$  to  $\overline{\mathfrak{M}_{\mu^1}^{\ell^\infty(B^s)} \times \mathfrak{M}_{\mu^2}^{\ell^\infty(B^s)}}$  is a continuous linear operator, we see that  $[\varphi^\mu]'(\mathbb{G}^\mu)$  is a zero-mean Gaussian random variable with values in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ . The result for the  $s \leq d/2$  case follows by the fact the inclusion map  $f \mapsto f, \mathcal{C}^s(\mathcal{X}) \rightarrow \mathcal{C}^{s'}(\mathcal{X})$  with  $s' < s$  is continuous.  $\square$

*Proof of Corollary 1.* Since the map  $f \mapsto \nabla f, \mathcal{C}^s(\mathcal{X}) \rightarrow \mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$  is continuous and linear, we have  $\sqrt{n}(\hat{T}_n - T^\mu) = -\nabla \sqrt{n}(\hat{\varphi}_n^1 - \varphi_1^\mu) \xrightarrow{d} -\nabla G_1^\mu$  in  $\mathcal{C}^{s-1}(\mathcal{X}; \mathbb{R}^d)$  and the limit  $-\nabla G_1^\mu$  is again a zero-mean Gaussian variable.  $\square$

*Proof of Proposition 1.* Given the weak convergence and Hadamard differentiability results (Lemmas 9 and 3), the first claim follows by applying the functional delta method.

For the second claim, assume  $\bar{S}_{c,\varepsilon}(\mu^1, \mu^2) \neq 0$  and  $\text{spt}(\mu^1) \cap \text{spt}(\mu^2) \neq \emptyset$ . For simplicity of notation, assume without loss of generality that  $\varepsilon = 1$  and  $x_1^\circ = x_2^\circ$ , and define the shorthands  $\varphi_i^\mu = \varphi_i$  and  $\varphi_i^{(\mu^1, \mu^2)} = \psi_i$ , for  $i = 1, 2$ . Suppose on the contrary that  $\sigma_\mu^2 = 0$ , which entails  $\text{Var}_{\mu^i}(\varphi_i - \psi_i) = 0$ , so  $\psi_i = \varphi_i + a_i \mu^i$ -a.e. for some  $a_i \in \mathbb{R}$  for  $i = 1, 2$ . The Schrödinger system (6) implies that

$$1 = \int e^{\varphi_1 \oplus \varphi_2 - c} d\mu^2 = e^{-a_1 - a_2} \int e^{\psi_1 \oplus \psi_2 - c} d\mu^2 = e^{-a_1 - a_2 + \psi_1 - \psi_2}$$

where we used the fact that  $\varphi_1^{(\mu^1, \mu^2)} = \varphi_2^{(\mu^1, \mu^2)} = \psi_2$ . The equality above holds  $\mu^1$ -a.e., so  $\psi_1 - \psi_2 = a_1 + a_2 \mu^1$ -a.e. By symmetry, we also have  $\psi_2 - \psi_1 = a_1 + a_2$

$\mu^2$ -a.e. Since  $\psi_1$  and  $\psi_2$  are continuous, we have  $\psi_1 - \psi_2 = a_1 + a_2$  on  $\text{spt}(\mu^1)$  and  $\psi_2 - \psi_1 = a_1 + a_2$  on  $\text{spt}(\mu^2)$ , and as  $\text{spt}(\mu^1) \cap \text{spt}(\mu^2) \neq \emptyset$ , we must have  $a_1 + a_2 = 0$ . However, the duality formula (9) then entails  $\bar{S}_{c,\varepsilon}(\mu^1, \mu^2) = 0$ , which is a contradiction.  $\square$

*Proof of Theorem 2.* Given the second-order Hadamard differentiability result (Theorem 5), the first claim of the theorem follows by applying the second-order functional delta method (Lemma 13),  $n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) \xrightarrow{d} \Delta_\mu(\mathbb{G}^{(\mu,\mu)})/2$ . The second claim follows by Lemma 10, upon noting that  $\text{spt}(\mathbb{G}^{(\mu,\mu)})$  is a vector subspace of  $\ell^\infty(B^s) \times \ell^\infty(B^s)$  (cf. the proof of Lemma 17 or Lemma 5.1 in [92]).  $\square$

**6.6. Proof of Proposition 2**

The proof uses techniques from  $U$ -processes. We refer to [29] as an excellent reference on  $U$ -processes. Assume  $\pi = \pi^V \otimes \pi^W$ . For notational convenience, set  $\hat{\pi}_n^\circ = \hat{\pi}_n^V \otimes \hat{\pi}_n^W$ . Let  $B^s$  be the unit ball in  $\mathcal{C}^s(\mathcal{X})$  with  $s > 2d$  (recall that  $d = d_1 + d_2$  and  $\mathcal{X} = \mathcal{V} \times \mathcal{W}$ ). We will derive a joint limit distribution for  $\sqrt{n}(\hat{\pi}_n - \pi)$  and  $\sqrt{n}(\hat{\pi}_n^\circ - \pi)$  in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ . As in [2], define for  $f \in B^s$ ,

$$\begin{aligned} h_f(x_1, x_2) &= h_f((v_1, w_1), (v_2, w_2)) \\ &:= f(v_1, w_1) + f(v_2, w_2) - f(v_1, w_2) - f(v_2, w_1). \end{aligned} \tag{24}$$

The function  $h_f$  is symmetric with  $\|h_f\|_{\mathcal{C}^s(\mathcal{X} \times \mathcal{X})} \leq 4$ . As  $V_i$  and  $W_i$  are independent, the mean of  $h_f(X_1, X_2)$  is zero:  $\mathbb{E}[h_f(X_1, X_2)] = 0$ . Consider the  $U$ -statistic with kernel  $h_f$ :

$$U_n(h_f) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_f(X_i, X_j).$$

Then, keeping in mind  $X_i = (V_i, W_i)$ , we can expand  $U_n(h_f)$  as

$$\begin{aligned} U_n(h_f) &= \frac{2}{n} \sum_{i=1}^n f(X_i) - \frac{2}{n(n-1)} \sum_{1 \leq i \neq j \leq n} f(V_i, W_j) \\ &= \frac{2}{n} \sum_{i=1}^n f(X_i) - \frac{2}{n^2} \sum_{i,j=1}^n f(V_i, W_j) \\ &\quad + \underbrace{\frac{2}{n^3 - n^2} \sum_{1 \leq i \neq j \leq n} f(V_i, W_j) + \frac{2}{n^2} \sum_{i=1}^n f(X_i)}_{=: A_n(f)} \\ &= 2(\hat{\pi}_n(f) - \hat{\pi}_n^\circ(f)) + A_n(f), \end{aligned}$$

that is,

$$\hat{\pi}_n^\circ(f) = \hat{\pi}_n(f) - \frac{1}{2}U_n(h_f) + \frac{1}{2}A_n(f).$$

Since  $|f| \leq 1$ , we have  $|A_n(f)| \leq 4/n$ . We will approximate  $U_n(h_f)$  by the Hajék process. Apply the Hoeffding decomposition to  $h_f$ :

$$h_f^{(1)}(x) = \int h_f(x, x') d\pi(x'), \quad h_f^{(2)}(x, x') = h_f(x, x') - h_f^{(1)}(x) - h_f^{(1)}(x').$$

Then, we can decompose  $U_n(h_f)$  as  $U_n(h_f) = 2\hat{\pi}_n(h_f^{(1)}) + U_n(h_f^{(2)})$ .

Now, since  $\{h_f : f \in B^s\} \subset \{h \in \mathcal{C}^s(\mathcal{X} \times \mathcal{X}) : \|h\|_{\mathcal{C}^s(\mathcal{X} \times \mathcal{X})} \leq 4\}$  and the  $t$ -entropy number of the latter function class w.r.t.  $\|\cdot\|_{\infty, \mathcal{X} \times \mathcal{X}}$  is of order  $t^{-2d/s}$  as  $t \downarrow 0$  by Theorem 2.7.1 in [91], applying Corollary 5.6 in [19] with  $k = 2$ , we have

$$\mathbb{E}[\|U_n(h_f^{(2)})\|_{\infty, B^s}] \leq O(n^{-1}) \times \int_0^1 t^{-2d/s} dt = O(n^{-1}).$$

Summarizing, we have  $\mathbb{E}[\|\sqrt{n}(\hat{\pi}_n^\circ(f) - \hat{\pi}_n(f - h_f^{(1)}))\|_{\infty, B^s}] \rightarrow 0$ .

For notational convenience, set  $\mathbb{G}_n = \sqrt{n}(\hat{\pi}_n - \pi)$  and  $\tilde{h}_f = f - h_f^{(1)}$ . We will show that  $((\mathbb{G}_n(f))_{f \in B^s}, (\mathbb{G}_n(\tilde{h}_f))_{f \in B^s})$  converges in distribution a tight limit in  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ . To this end, we will apply Lemma 18. Again, by Theorem 2.7.1 in [91],  $B^s$  and  $\tilde{\mathcal{H}} := \{\tilde{h}_f : f \in B^s\}$  are both  $\pi$ -Donsker. Finite-dimensional convergence follows trivially, so by Lemma 18,

$$((\mathbb{G}_n(f))_{f \in B^s}, (\mathbb{G}_n(h))_{h \in \tilde{\mathcal{H}}}) \xrightarrow{d} \mathbf{G}^\pi \quad \text{in } \ell^\infty(B^s) \times \ell^\infty(\tilde{\mathcal{H}})$$

for some tight limit  $\mathbf{G}^\pi$ . Finally, since the map  $\iota : \ell^\infty(B^s) \times \ell^\infty(\tilde{\mathcal{H}}) \rightarrow \ell^\infty(B^s) \times \ell^\infty(B^s)$  defined by

$$(\iota \mathbf{z})_1(f) = z_1(f) \quad \text{and} \quad (\iota \mathbf{z})_2(f) = z_2(\tilde{h}_f), \quad \mathbf{z} = (z_1, z_2) \in \ell^\infty(B^s) \times \ell^\infty(\tilde{\mathcal{H}})$$

is continuous, we conclude that

$$((\mathbb{G}_n(f))_{f \in B^s}, (\mathbb{G}_n(\tilde{h}_f))_{f \in B^s}) \xrightarrow{d} \iota \mathbf{G}^\pi =: \tilde{\mathbf{G}}^\pi \quad \text{in } \ell^\infty(B^s) \times \ell^\infty(B^s).$$

The limit variable  $\tilde{\mathbf{G}}^\pi = (\tilde{G}_1^\pi, \tilde{G}_2^\pi)$  is a two-dimensional Gaussian process with mean zero and covariance structure

$$\begin{aligned} \text{Cov}(\tilde{G}_1^\pi(f), \tilde{G}_1^\pi(g)) &= \text{Cov}_\pi(f, g), \quad \text{Cov}(\tilde{G}_2^\pi(f), \tilde{G}_2^\pi(g)) = \text{Cov}_\pi(\tilde{h}_f, \tilde{h}_g), \\ \text{and} \quad \text{Cov}(\tilde{G}_1^\pi(f), \tilde{G}_2^\pi(g)) &= \text{Cov}_\pi(f, \tilde{h}_g). \end{aligned}$$

In view of the discussion above, the same limit holds for  $(\sqrt{n}(\hat{\pi}_n - \pi), \sqrt{n}(\hat{\pi}_n^\circ - \pi))$ . Now, combining Theorem 4 and the second-order functional delta method (Lemma 13), we conclude that

$$nD_n = n\bar{S}_{c, \varepsilon}(\hat{\pi}_n, \hat{\pi}_n^\circ) \xrightarrow{d} \frac{1}{2} \Delta_\pi(\tilde{\mathbf{G}}^\pi).$$

Finally, the second claim of the proposition follows by Lemma 10, upon noting that the support of  $\tilde{\mathbf{G}}^\pi$  is a vector subspace of  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ ; cf. Lemma 5.1 in [92]. This completes the proof.  $\square$

### 6.7. Proof of Proposition 3

(i). We first note that, as  $hd\mu^2 \in \overline{\mathcal{M}_{\mu^2}}^{\ell^\infty(B^s)}$ , by Hadamard differentiability,  $\kappa_n$  satisfies

$$\sqrt{n}(\kappa_n(h) - \kappa_n(0)) \rightarrow \delta'_{\mu^2}(hd\mu^2) =: \dot{\kappa}(h).$$

Since  $h \mapsto hd\mu^2$  is linear and continuous from  $H$  into  $\ell^\infty(B^s)$  (cf. Example 1.5.10 in [91]),  $\dot{\kappa} : H \rightarrow \mathbb{B}$  is a continuous linear map, establishing regularity of the parameter sequence  $\kappa_n(h)$ .

Next, we shall verify regularity of the empirical EOT map  $\tilde{T}_n$ . Hadamard differentiability enables finding limit distributions under local alternatives. By the second claim of the functional delta method (Lemma 11), we have

$$\sqrt{n}(\tilde{T}_n - T^\mu) - \delta'_{\mu^2}(\sqrt{n}(\hat{\mu}_n^2 - \mu^2)) \rightarrow 0$$

in probability under  $P_{n,0}$ . From the proof of Theorem 3.10.12 in [91],

$$\left( \sqrt{n}(\hat{\mu}_n^2 - \mu^2), \frac{dP_{n,h}}{dP_{n,0}} \right) \xrightarrow{d} (\mathbb{G}_2^{\mu^2}, \Lambda) \quad \text{in } \ell^\infty(B^s) \times \mathbb{R}$$

under  $P_{n,0}$ , and the law  $L$  on  $\ell^\infty(B^s)$  defined by  $L(A) = \mathbb{E}[\mathbb{1}_A(\mathbb{G}_2^{\mu^2})\Lambda]$  agrees with the law of  $\mathbb{G}_2^{\mu^2} + hd\mu^2$  (note:  $hd\mu^2$  should be understood as an element of  $\ell^\infty(B^s)$ ). Combining these two displays above yields that

$$\left( \sqrt{n}(\tilde{T}_n - T^\mu), \frac{dP_{n,h}}{dP_{n,0}} \right) \xrightarrow{d} (\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}), \Lambda) \quad \text{in } \mathbb{B} \times \mathbb{R}$$

under  $P_{n,0}$ . Hence, by Le Cam's third lemma (Theorem 3.10.7 in [91]),  $\sqrt{n}(\tilde{T}_n - T^\mu) \xrightarrow{d} \delta'_{\mu^2}(\mathbb{G}_2^{\mu^2} + hd\mu^2)$  under  $P_{n,h}$ . Note that  $\mathbb{G}_2^{\mu^2} + hd\mu^2 \in \overline{\mathfrak{M}_{\mu^2}}^{\ell^\infty(B^s)}$ . By linearity of the derivative, we have  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2} + hd\mu^2) = \delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) + \delta'_{\mu^2}(hd\mu^2)$ . Conclude that

$$\sqrt{n}(\tilde{T}_n - T^\mu) \xrightarrow{d} \delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) + \delta'_{\mu^2}(hd\mu^2) \quad \text{in } \mathbb{B}$$

under  $P_{n,h}$ . This immediately implies that  $\tilde{T}_n$  is regular, that is,  $\sqrt{n}(\tilde{T}_n - T^{(\mu^1, \mu_n^2, h)}) \xrightarrow{d} \delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})$  under  $P_{n,h}$ .

Now, in view of Theorem 3.11.2 in [91], the last claim follows by verifying that  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) \stackrel{d}{=} G$ , where  $G$  is a Gaussian random variable in  $\mathbb{B}$  such that for every  $b^* \in \mathbb{B}^*$ ,  $b^*G$  is Gaussian with mean zero and variance

$$\sup_{h \in H: \|h\|_{L^2(\mu^2)}=1} (b^* \dot{\kappa}(h))^2 = \sup_{h \in H: \|h\|_{L^2(\mu^2)}=1} (b^* \delta'_{\mu^2}(hd\mu^2))^2.$$

By construction  $\delta'_{\mu^2, b^*} := b^* \circ \delta'_{\mu^2}$  is a continuous linear functional on  $\overline{\mathfrak{M}_{\mu^2}}^{\ell^\infty(B^s)} \supset \text{spt}(\mathbb{G}_2^{\mu^2})$  with  $\mathfrak{M}_{\mu^2} = \{hd\mu^2 : h \in H\}$  (Lemma 17). Note that  $\overline{\mathfrak{M}_{\mu^2}}^{\ell^\infty(B^s)}$  is a

closed subspace of  $\ell^\infty(B^s)$ . Arguing as in the proof of Proposition 2 in [48], we obtain

$$\text{Var}(b^*G) = \sup_{\substack{h \in H \\ \|h\|_{L^2(\mu^2)}=1}} (\delta'_{\mu^2, b^*}(hd\mu^2))^2 = \text{Var}(\delta'_{\mu^2, b^*}(\mathbb{G}_2^{\mu^2})) = \text{Var}(b^*\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2})).$$

Conclude that  $\delta'_{\mu^2}(\mathbb{G}_2^{\mu^2}) \stackrel{d}{=} G$ .

(ii). Given the result of Part (i), Part (ii) follows directly from Theorem 3.11.5 in [91].  $\square$

### 7. Discussions

We have established limit distributions for the EOT potentials, map, and Sinkhorn divergence. The main ingredient of our proofs was Hadamard differentiability of the relevant maps. Importantly, the Hadamard differentiability results yield not only limit distributions but also bootstrap consistency and asymptotic efficiency. Regarding the Sinkhorn divergence, our main contribution is the derivation of the null limit distribution for compactly supported distributions. When the population measures agree, the first Hadamard derivative of the Sinkhorn divergence vanishes, which necessitates looking into higher-order Hadamard derivatives. For this, Hadamard differentiability of the EOT map in (sufficiently regular) Hölder spaces plays an important role. As another contribution, we have derived the null limit distribution of the Sinkhorn independence test statistic and determined the precise order of the test statistic under the null, which was not available before. We end this paper with discussions of two possible extensions.

#### 7.1. Unbounded supports

In the present paper, we have assumed that the marginals  $\mu^1$  and  $\mu^2$  are compactly supported, which excludes, for instance, Gaussian distributions (one exception is Proposition 1, where the compactness assumption can be relaxed to a sub-Gaussian condition when  $c$  is quadratic; see [31, 48]). Indeed, the compactness assumption is essential to formulate the Hadamard differentiability results, where we first regard the dual potentials as a mapping into the Hölder space with arbitrary smoothness level (cf. Lemma 1 (ii)), and then embed each marginal into the topological dual of the Hölder space (cf. Remark 9). The second step is natural in view of the Schrödinger system, since whenever  $\varphi_1, \varphi_2 \in \mathcal{C}^s(\mathcal{X})$ , the exponentiated functions

$$e^{\frac{\varphi_1(x_1) + \varphi_2(\cdot) - c(x_1, \cdot)}{\varepsilon}} \quad \text{and} \quad e^{\frac{\varphi_1(\cdot) + \varphi_2(x_2) - c(\cdot, x_2)}{\varepsilon}} \tag{25}$$

both lie in  $\mathcal{C}^s(\mathcal{X})$  for every  $x_1, x_2 \in \mathcal{X}$ .

For certain smooth costs, even when the marginals are not compactly supported, as long as they have sufficiently light tails, we may uniformly upper bound derivatives of EOT potentials by a polynomial of  $1 + \|\cdot\|$  (cf. [65]), so a natural idea would be to replace the Hölder space with a certain weighted Hölder space (cf. [71]). However, even when the dual potentials lie in a weighted Hölder space, the exponentiated functions in (25) need not lie in the same weighted Hölder space, so it is nontrivial to find a suitable normed space into which the marginals are embedded when they are not compactly supported. In particular, the desired space should enable extending Lemma 6 (ii) to the unboundedly supported setting to ensure that the Fréchet derivative  $\dot{\Psi}_\mu$  is injective with a continuous inverse. While these properties are crucial for our argument, they seem challenging to establish for marginals with unbounded supports. We therefore leave this exploration for future work.

It is worth mentioning that [53] derive a Hadamard derivative for the EOT plan w.r.t. the marginals when the supports are *countable*. They use a curvature-type condition on the cost to allow for unbounded supports (see Remark 5.9 in [53]; see also [52] for a similar condition), but their proof technique is restricted to the countable support case where we may naturally embed the marginals into a sequence space, and does not directly extend to a more general case.

## 7.2. Multimarginal EOT

The results of the present paper extend to the multimarginal case, as long as we deal with compactly supported marginals. Multimarginal optimal transport finds applications spanning from economics [16, 21, 38], through density functional theory in quantum chemistry [23, 26], to generative modeling in machine learning [11]. The multimarginal problem is also intimately related to Wasserstein barycenters; cf., e.g., [1], where entropic regularization is often applied to attain computational tractability [28].

Given a smooth cost function  $c : (\mathbb{R}^d)^N \rightarrow \mathbb{R}_+$  and marginals  $\mu^1, \dots, \mu^N \in \mathcal{P}(\mathcal{X})$  with  $N \geq 2$  arbitrary, the multimarginal EOT problem reads as

$$\inf_{\pi \in \Pi(\mu^1, \dots, \mu^N)} \int c d\pi + \varepsilon \text{D}_{\text{KL}}(\pi \| m), \quad m = \mu^1 \otimes \dots \otimes \mu^N, \quad (26)$$

where  $\Pi(\mu^1, \dots, \mu^N)$  is the set of Borel probability measures on  $\mathcal{X}^N$  with marginals  $\mu^1, \dots, \mu^N$ . To simplify notation, we set  $\varepsilon = 1$ . The corresponding dual problem is

$$\sup_{(\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^N L^\infty(\mu^i)} \sum_{i=1}^N \int \varphi_i d\mu^i - \int e^{\sum_{i=1}^N \varphi_i(x_i) - c(x_1, \dots, x_N)} dm(x_1, \dots, x_N) + 1.$$

Bounded functions  $\varphi = (\varphi_1, \dots, \varphi_N)$  solve the dual problem if and only if they satisfy the Schrödinger system, i.e.,

$$\int e^{\sum_{j=1}^N \varphi_j(x_j) - c(x_1, \dots, x_N)} dm^{-i}(x_{-i}) - 1 = 0 \quad \text{for } \mu^i\text{-a.e. } x_i, \quad i = 1, \dots, N, \quad (27)$$

where  $m^{-i} = \otimes_{j \neq i} \mu^j$  and  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$ . See [17]. Theorem 4.3 in [17] shows that the Schrödinger system (27) admits a solution  $\varphi = (\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^N L^\infty(\mu^i)$ , which is unique in the sense that if  $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_N) \in \prod_{i=1}^N L^\infty(\mu^i)$  is another solution to the Schrödinger system (27), then  $\sum_{i=1}^N \varphi(x_i) = \sum_{i=1}^N \tilde{\varphi}_i(x_i)$  for  $m$ -a.e.  $(x_1, \dots, x_N)$ , i.e., there exist constants  $a_1, \dots, a_N$  that sum to zero such that  $(\tilde{\varphi}_1, \dots, \tilde{\varphi}_N) = (\varphi_1 + a_1, \dots, \varphi_N + a_N)$  for  $m$ -a.e.  $(x_1, \dots, x_N)$ . We call  $\varphi$  EOT potentials. Then, the (unique) optimal solution to the multimarginal EOT problem (26) is given by

$$d\pi^*(x_1, \dots, x_N) = e^{\sum_{i=1}^N \varphi_i(x_i) - c(x_1, \dots, x_N)} dm(x_1, \dots, x_N).$$

Given these preparations, it is not difficult to see that Lemma 1 naturally extends to the multimarginal setting. For simplicity, we choose a common reference point  $x^\circ \in \mathcal{X}$ . Then, for every  $\mu = (\mu^1, \dots, \mu^N) \in \prod_{i=1}^N \mathcal{P}(\mathcal{X})$ , there exists a unique set of functions  $\varphi^\mu = (\varphi_1^\mu, \dots, \varphi_N^\mu) \in \prod_{i=1}^N \mathcal{C}(\mathcal{X})$  satisfying the Schrödinger system (27) for every  $x_i \in \mathcal{X}$  for each  $i = 1, \dots, N$ , and such that  $\varphi_1^\mu(x^\circ) = \dots = \varphi_N^\mu(x^\circ)$ . Furthermore, for every  $s \in \mathbb{N}$ , there exists a constant  $R_s$  such that  $\max_{1 \leq i \leq N} \|\varphi_i^\mu\|_{\mathcal{C}^s(\mathcal{X})} \leq R_s$  for all  $\mu \in \prod_{i=1}^N \mathcal{P}(\mathcal{X})$ . Then, Theorem 3 extends to the multimarginal setting as follows.

**Theorem 6.** *For every  $s \in \mathbb{N}$  and  $\mu = (\mu^1, \dots, \mu^N) \in \prod_{i=1}^N \mathcal{P}(\mathcal{X})$ , the map  $\nu \mapsto \varphi^\nu, \prod_{i=1}^N \mathcal{P}_{\mu^i} \subset \prod_{i=1}^N \ell^\infty(B^s) \rightarrow \prod_{i=1}^N \mathcal{C}^s(\mathcal{X})$  is Hadamard differentiable tangentially to  $\prod_{i=1}^N \overline{\mathcal{M}_{\mu^i}^{\ell^\infty(B^s)}}$ .*

The proof is similar to the two-marginal case, so omitted for brevity. Note that the proof of Lemma 6 relies on the results of [17], but their results cover the multimarginal case. Other results extend similarly. For example, let  $\hat{\mu}_n^i$  denote the empirical distribution of  $n$  i.i.d. data from  $\mu^i$  and assume the samples from different marginal distributions are independent. Then, for  $\hat{\varphi}_n = \varphi^{\hat{\mu}_n}$ , we have that  $\sqrt{n}(\hat{\varphi}_n - \varphi^\mu)$  converges in distribution to a zero-mean Gaussian random variable in  $\prod_{i=1}^N \mathcal{C}^s(\mathcal{X})$ .

## Appendix A: $m$ -out-of- $n$ bootstrap for Sinkhorn null limit

We consider estimating the Sinkhorn null limit distribution in Theorem 2 by the two-sample  $m$ -out-of- $n$  bootstrap. Let  $\mu = (\mu^1, \mu^2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  be arbitrary. For each  $i = 1, 2$ , let  $X_1^i, \dots, X_n^i$  be i.i.d. data from  $\mu^i$  with  $\hat{\mu}_n^i = n^{-1} \sum_{j=1}^n \delta_{X_j^i}$ . Consider the pooled empirical distribution  $\hat{\rho}_n = (2n)^{-1} \sum_{j=1}^n (\delta_{X_j^1} + \delta_{X_j^2})$ , and let  $Z_1^B, \dots, Z_{2m}^B$  be an independent sample from  $\hat{\rho}_n$ , where  $m = m_n \rightarrow \infty$ . Set

$$\hat{\rho}_{m,n}^{1,B} = \frac{1}{m} \sum_{j=1}^m \delta_{Z_j^B} \quad \text{and} \quad \hat{\rho}_{m,n}^{2,B} = \frac{1}{m} \sum_{j=m+1}^{2m} \delta_{Z_j^B}.$$

The following proposition shows that the  $m$ -out-of- $n$  bootstrap can consistently estimate the null limit law in Theorem 2 when  $\mu^1 = \mu^2 = \mu$ . Recall that  $\text{BL}_1(\mathbb{R})$  denotes the collection of 1-Lipschitz functions  $g: \mathbb{R} \rightarrow [-1, 1]$ .



**Proposition 4.** *Consider the above setting and set  $\rho = (\mu^1 + \mu^2)/2$ . Assume  $m = o(n)$ . Then, we have*

$$\sup_{g \in \text{BL}_1(\mathbb{R})} \left| \mathbb{E}^B \left[ g(m\bar{S}_{c,\varepsilon}(\hat{\rho}_{m,n}^{1,B}, \hat{\rho}_{m,n}^{2,B})) \right] - \mathbb{E}[g(\chi_\rho)] \right| \rightarrow 0 \quad (28)$$

in probability, where  $\mathbb{E}^B$  denotes the conditional expectation given the sample and  $\chi_\rho$  follows the limit law in Theorem 2 with  $\mu$  replaced by  $\rho$ .

*Remark 15.* Since  $\text{BL}_1(\mathbb{R})$  is compact w.r.t. the topology of locally uniform convergence by the Ascoli-Arzelà theorem, the left-hand side of (28) reduces to the supremum over a countable subcollection of  $\text{BL}_1(\mathbb{R})$  and hence is a proper random variable.

*Proof.* Fix  $s > d/2$ . Set  $\mathbb{G}_{m,n}^{i,B} = \sqrt{m}(\hat{\rho}_{m,n}^{i,B} - \hat{\rho}_n)$  for  $i = 1, 2$ . Given the sample,  $(\mathbb{G}_{m,n}^{1,B}, \mathbb{G}_{m,n}^{2,B})$  has only finitely many possible values, so induces the conditional distribution defined on the Borel  $\sigma$ -field on  $\ell^\infty(B^s) \times \ell^\infty(B^s)$ . Then, since  $B^s$  is uniformly bounded and Donsker for  $\mu^1$  and  $\mu^2$ , arguing as in the proof of Theorem 3.7.1 in [91], we have

$$(\mathbb{G}_{m,n}^{1,B}, \mathbb{G}_{m,n}^{2,B}) \xrightarrow{d} (\mathbb{G}_1^\rho, \mathbb{G}_2^\rho) \quad \text{in } \ell^\infty(B^s) \times \ell^\infty(B^s)$$

given almost every sequence  $X_1^1, X_2^1, \dots, X_1^2, X_2^2, \dots$ , where  $\mathbb{G}_1^\rho$  and  $\mathbb{G}_2^\rho$  are independent tight  $\rho$ -Brownian bridges in  $\ell^\infty(B^s)$ . Furthermore,  $\mathbb{E}[\|\sqrt{m}(\hat{\rho}_n - \rho)\|_{\infty, B^s}] = O(\sqrt{m/n}) = o(1)$  by Theorem 2.14.1 in [91] and the assumption that  $m = o(n)$ , so  $\|\sqrt{m}(\hat{\rho}_n - \rho)\|_{\infty, B^s} \rightarrow 0$  in probability by Markov's inequality. Pick any subsequence  $n'$  of  $n$  and choose a further subsequence  $n''$  for which  $\|\sqrt{m''}(\hat{\rho}_{n''} - \rho)\|_{\infty, B^s} \rightarrow 0$  a.s. with  $m'' = m_{n''}$ . Now, since  $\sqrt{m}(\hat{\rho}_{m,n}^{i,B} - \rho) = \mathbb{G}_{m,n}^{i,B} + \sqrt{m}(\hat{\rho}_n - \rho)$ , we have

$$(\sqrt{m''}(\hat{\rho}_{m'',n''}^{1,B} - \rho), \sqrt{m''}(\hat{\rho}_{m'',n''}^{2,B} - \rho)) \xrightarrow{d} (\mathbb{G}_1^\rho, \mathbb{G}_2^\rho) \quad \text{in } \ell^\infty(B^s) \times \ell^\infty(B^s) \quad (29)$$

given almost every sequence  $X_1^1, X_2^1, \dots, X_1^2, X_2^2, \dots$ . Then, by the second-order Hadamard differentiability of  $\bar{S}_{c,\varepsilon}$  (Theorem 5) and the second-order functional delta method (Lemma 13)<sup>6</sup>, we have

$$m''\bar{S}_{c,\varepsilon}(\hat{\rho}_{m'',n''}^{1,B}, \hat{\rho}_{m'',n''}^{2,B}) \xrightarrow{d} \frac{1}{2}\Delta_\rho(\mathbb{G}_1^\rho, \mathbb{G}_2^\rho) \stackrel{d}{=} \chi_\rho$$

given almost every sequence  $X_1^1, X_2^1, \dots, X_1^2, X_2^2, \dots$ . Since the limit is independent of the choice of subsequence  $n'$ , we obtain the result.  $\square$

### A.1. Numerical experiments

We present numerical experiments to assess the scaling for the Sinkhorn null limit in Theorem 2 as well as the finite sample performance of the two-sample  $m$ -out-of- $n$  bootstrap procedure.

<sup>6</sup>We have chosen a suitable subsequence for which the weak convergence (29) holds for almost every sequence  $X_1^1, X_2^1, \dots, X_1^2, X_2^2, \dots$ . Therefore, for every fixed such sequence, we may apply the second-order functional delta method.

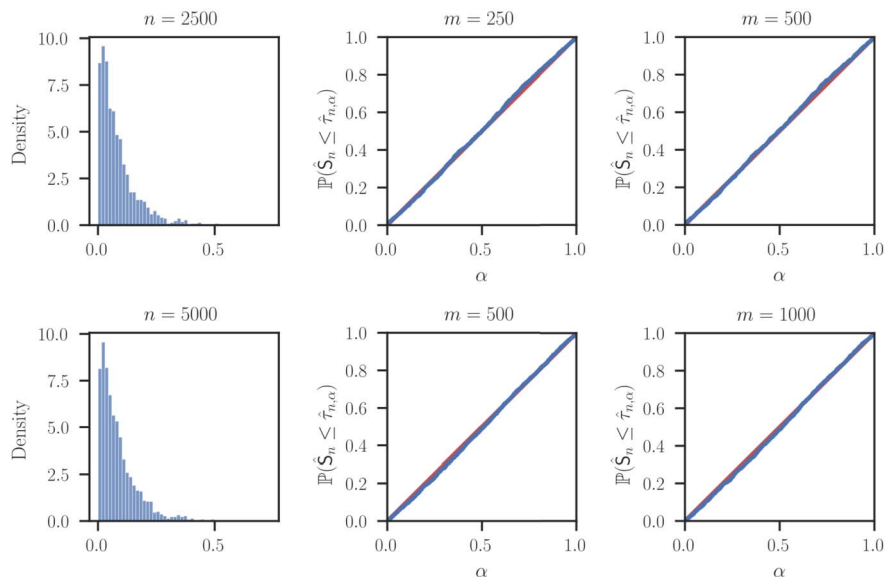


FIG 1. The top row consists (from left to right) of a histogram of  $\hat{S}_n = n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  for  $n = 2500$ , and P-P plots of  $\hat{S}_n$  generated using the methodology described in Appendix A.1 with  $m = 0.1n$  and  $m = 0.2n$  respectively along with a red  $45^\circ$  reference line. The bottom row compiles analogous plots for  $n = 5000$ . See Appendix A.2 for additional implementation details.

Throughout, the cost function  $c$  is the squared Euclidean distance,  $\varepsilon = 1$ , and  $\mu^1 = \mu^2$  is taken to be the uniform distribution on  $(0, 1/2)^2$ . Figure 1 consists of two experiments. The first (top row) involves plotting a histogram of  $\hat{S}_n = n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  for  $n = 2500$  based on 1500 repetitions along with corresponding P-P plots for  $\hat{S}_n$ . As the limit distribution  $\chi_\mu$  in Theorem 2 is not defined explicitly, we compute the coverage probabilities,  $\mathbb{P}(\hat{S}_n \leq \hat{\tau}_{n,\alpha})$  for  $\alpha \in (0, 1)$ , where  $\hat{\tau}_{n,\alpha}$  is the  $\alpha$ -quantile of the subsampled distribution. Precisely, for each of the 1500 repetitions, we construct the subsampled distribution function  $\hat{F}_{m,n}^B$  of  $m\bar{S}_{c,\varepsilon}(\hat{\rho}_{m,n}^{1,B}, \hat{\rho}_{m,n}^{2,B})$  for  $m = 0.1n$  and  $m = 0.2n$  based on 1000 repetitions; for each such repetition, we compute the rank of  $\hat{S}_n$  w.r.t.  $\hat{F}_{m,n}^B$  (i.e.,  $\hat{F}_{m,n}^B(\hat{S}_n)$ ), and then approximate  $\mathbb{P}(\hat{S}_n \leq \hat{\tau}_{n,\alpha})$  by the number of ranks with value less than  $\alpha$ . The second experiment (bottom row) is performed similarly, but with  $n = 5000$  rather than 2500 to see how these results vary with increasing  $n$ .

From these experiments, one can see that the empirical distribution of  $\hat{S}_n$  is observed to be reasonably stable in the finite sample regime, which is consistent with Theorem 2, and the coverage probabilities  $\mathbb{P}(\hat{S}_n \leq \hat{\tau}_{n,\alpha})$  are close to the  $45^\circ$  line uniformly over  $\alpha \in (0, 1)$ , so the  $m$ -out-of- $n$  bootstrap approximates well the sampling distribution of  $\hat{S}_n$ .

**A.2. Implementation details and theoretical guarantees**

To compute the relevant EOT quantities, we use the standard implementation of Sinkhorn’s algorithm from the Python Optimal Transport package [40] (see also Algorithm 3 in [84]). To estimate the EOT plan  $\mathbf{\Pi}^*$   $\in \mathbb{R}^{N \times N}$  between measures  $\mu^1, \mu^2$  supported on  $N$  points  $(x^{1,i})_{i=1}^N, (x^{2,i})_{i=1}^N \subset \mathcal{X}$  with weights  $w^1 = (\mu^1(x^{1,1}), \dots, \mu^1(x^{1,N}))$ ,  $w^2 = (\mu^2(x^{2,1}), \dots, \mu^2(x^{2,N}))$ , Sinkhorn’s algorithm leverages the representation (5) to iteratively construct matrices  $\mathbf{\Pi}^k \in \mathbb{R}^{N \times N}$  satisfying  $\mathbf{\Pi}^k \mathbf{1}_N = w^1$ , where  $\mathbf{1}_N \in \mathbb{R}^N$  is the vector of 1’s. Convergence of this implementation of Sinkhorn’s algorithm to  $\mathbf{\Pi}^*$  follows from the note on p. 731 in [42] with only minor modifications. Two stopping conditions are implemented in this version of Sinkhorn’s algorithm. The first limits the total number of iterations to 1000, while the second sets a desired threshold on the violation of the marginal constraint  $\|(\mathbf{\Pi}^k)^\top \mathbf{1}_N - w^2\| \leq 1e^{-9}$ .

Given the linear scaling from Theorem 2, to obtain statistically meaningful simulation results we must compute  $\hat{S}_n = n\bar{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$ , for a fixed  $\varepsilon > 0$ , to within  $o_P(n^{-1})$  precision. To the best of our knowledge, past works concerning computational complexity of Sinkhorn’s algorithm do not provide guarantees for this setting. Rather, they focus on approximating the unregularized OT cost within a desired precision (cf., e.g., [4, Theorem 1] and [36, Theorem 2]), which requires scaling  $\varepsilon \searrow 0$  and treating the entropy term as a type of bias.

To bridge this gap, we analyze the total number of Sinkhorn iterations required to achieve accuracy  $o_P(n^{-1})$  when computing  $\hat{S}_n$  with fixed  $\varepsilon > 0$ . Corollary 2 ahead states that  $o_P(\log(n \log(n)))$  iterations are sufficient, and we recall that each iteration requires  $O(n^2)$  arithmetic operations [27]. As this setting has not been treated before, we believe that this result may be of independent interest.

To arrive at the iteration complexity bound, we start by setting some notation. For positive vectors  $a, b \in \mathbb{R}^n$ , define

$$d_H(a, b) = \log \left( \max_{1 \leq i, j \leq n} \frac{a_i b_j}{b_i a_j} \right) = \max_{1 \leq i \leq n} \log \left( \frac{a_i}{b_i} \right) - \min_{1 \leq i \leq n} \log \left( \frac{a_i}{b_i} \right),$$

and, fixing  $\mathbf{K} \in \mathbb{R}^{n \times n}$  with positive entries, let

$$E_{\mathbf{K}} = \{ \mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} = \text{diag}(u)\mathbf{K}\text{diag}(v), \mathbf{A}_{ij} > 0 \}.$$

Define a metric  $d$  on  $E_{\mathbf{K}} \times E_{\mathbf{K}}$  via

$$d(\mathbf{A}, \mathbf{B}) = d_H(x, \mathbf{1}_n) + d_H(y, \mathbf{1}_n), \text{ for } x, y \text{ satisfying } \mathbf{A} = \text{diag}(x)\mathbf{B}\text{diag}(y).$$

For a positive matrix  $\mathbf{K} \in \mathbb{R}^{n \times n}$ , let

$$\lambda(\mathbf{K}) := \frac{\sqrt{\eta(\mathbf{K})} - 1}{\sqrt{\eta(\mathbf{K})} + 1} < 1, \quad \eta(\mathbf{K}) := \max_{\substack{1 \leq i, j \leq n \\ 1 \leq k, l \leq n}} \frac{\mathbf{K}_{ik}\mathbf{K}_{jl}}{\mathbf{K}_{jk}\mathbf{K}_{il}}.$$

In the sequel we let  $\mathbf{C} \in \mathbb{R}^{n \times n}$  be the matrix of pairwise costs  $\mathbf{C}_{ij} = c(x^{1,i}, x^{2,j})$  and  $\mathbf{K} \in \mathbb{R}^{n \times n}$  be given by  $\mathbf{K}_{ij} = e^{-\mathbf{C}_{ij}/\varepsilon}$  for  $i, j \in \{1, \dots, n\}$ .

**Proposition 5.** *Fix  $\varepsilon > 0$ ,  $\delta > 0$ , and assume that  $\hat{\mu}_n^1, \hat{\mu}_n^2$  are supported on  $(x^{1,i})_{i=1}^n, (x^{2,i})_{i=1}^n$  respectively. Then, Sinkhorn’s algorithm produces a matrix  $\tilde{\Pi} \in \mathbb{R}^{n \times n}$  that achieves a cost of  $\tilde{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  for the discrete EOT primal functional (3) satisfying*

$$|\mathbf{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)| \leq (6\|c\|_{\infty, \mathcal{X} \times \mathcal{X}} + 2\varepsilon \log(n)) (e^\delta - 1) + \varepsilon\delta$$

in  $O(\log(1/\delta))$  iterations. This cost estimate is also satisfied by  $\tilde{\Pi} = \Pi^k$  when the termination condition  $\|(\Pi^k)^\top \mathbb{1}_n - \mathbb{1}_n/n\| \leq \frac{1}{n} \frac{\delta(1-\lambda(\mathbf{K})) + 2 - \sqrt{\delta^2(1-\lambda(\mathbf{K}))^2 + 4}}{2}$  is met, which requires  $O(\frac{n}{\delta} \log n)$  iterations.

*Proof.* It follows from equation (28) in [84] and the surrounding discussion that the iterates  $\Pi^k$  of Sinkhorn’s algorithm achieve  $d(\Pi^*, \Pi^k) \leq \delta$  in at most

$$1 + \frac{1}{2 \log((\lambda(\mathbf{K}))^{-1})} \log \left( \frac{d_H((\Pi^1)^\top \mathbb{1}_n, \frac{\mathbb{1}_n}{n})}{\delta(1-\lambda(\mathbf{K}))} \right),$$

iterations, recalling that  $\lambda(\mathbf{K}) < 1$ . Further, the 2-norm termination condition is met in at most

$$1 + \frac{4n}{\delta(1-\lambda(\mathbf{K})) + 2 - \sqrt{\delta^2(1-\lambda(\mathbf{K}))^2 + 4}} \log \left( e^{\|\mathbf{C}\|_\infty/\varepsilon n} \right)$$

iterations; the value for the termination condition is set as to guarantee that if  $\tilde{\Pi}$  satisfies this termination condition, then  $d(\Pi^*, \tilde{\Pi}) \leq \delta$  (see Proposition 31 in [84]). Observing that  $\lambda(\mathbf{K})$  is independent of  $n$ , we proceed with showing that  $d_H((\Pi^1)^\top \mathbb{1}_n, \frac{\mathbb{1}_n}{n})$  is independent of  $n$ , such that the claimed dependencies on  $\delta$  and  $n$  hold.

As  $\Pi^1 = \text{diag}(u^1) \mathbf{K} \text{diag}(v^1)$ , for  $u^1 = \frac{1}{n} \mathbb{1}_n / (\mathbf{K}(\mathbb{1}_n / (\mathbf{K}^\top \mathbb{1}_n)))$  and  $v^1 = \mathbb{1}_n / (\mathbf{K}^\top \mathbb{1}_n)$ , where division is understood componentwise (cf. e.g. Algorithm 3 in [84]), we have

$$\begin{aligned} d_H \left( (\Pi^1)^\top \mathbb{1}_n, \frac{\mathbb{1}_n}{n} \right) &= \max_{1 \leq i \leq n} \log(n((\Pi^1)^\top \mathbb{1}_n)_i) - \min_{1 \leq i \leq n} \log(n((\Pi^1)^\top \mathbb{1}_n)_i) \\ &= \max_{1 \leq i \leq n} \log(((\Pi^1)^\top \mathbb{1}_n)_i) - \min_{1 \leq i \leq n} \log(((\Pi^1)^\top \mathbb{1}_n)_i), \end{aligned}$$

and we note that

$$\begin{aligned} n \min_{i,j=1}^n K_{ji} &\leq (\mathbf{K}^\top \mathbb{1}_n)_i = \sum_{j=1}^n K_{ji} \leq n \max_{i,j=1}^n K_{ji}, \\ n \min_{i,j=1}^n K_{ij} &\leq (\mathbf{K} \mathbb{1}_n)_i = \sum_{j=1}^n K_{ij} \leq n \max_{i,j=1}^n K_{ij}, \end{aligned}$$

so

$$\frac{1}{n} \frac{(\min_{i,j=1}^n K_{ij})^2}{(\max_{i,j=1}^n K_{ij})^2} \leq ((\mathbf{\Pi}^1)^\top \mathbb{1}_n)_i \leq \frac{1}{n} \frac{(\max_{i,j=1}^n K_{ij})^2}{(\min_{i,j=1}^n K_{ij})^2},$$

such that

$$2 \log \left( \frac{\min_{i,j=1}^n K_{ij}}{\max_{i,j=1}^n K_{ij}} \right) \leq d_H \left( (\mathbf{\Pi}^1)^\top \mathbb{1}_n, \frac{\mathbb{1}_n}{n} \right) \leq 2 \log \left( \frac{\max_{i,j=1}^n K_{ij}}{\min_{i,j=1}^n K_{ij}} \right).$$

Conclude that  $d_H((\mathbf{\Pi}^1)^\top \mathbb{1}_n, \frac{\mathbb{1}_n}{n})$  can be bounded independently of  $n$ .

We now lift the bound  $d(\mathbf{\Pi}^*, \tilde{\mathbf{\Pi}}) \leq \delta$  to a bound on  $|\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)|$ , where  $\tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  is the discrete EOT primal functional (3) evaluated at  $\tilde{\mathbf{\Pi}}$  (note that  $\tilde{\mathbf{\Pi}}$  is not necessarily a coupling, but is positive with entries summing to 1). By Lemma 3.9 in [42], it holds that, for any  $i, j \in \{1, \dots, n\}$ ,

$$e^{-\delta} - 1 \leq \frac{\mathbf{\Pi}_{ij}^*}{\tilde{\mathbf{\Pi}}_{ij}} - 1 \leq e^\delta - 1, \text{ so } |\mathbf{\Pi}_{ij}^* - \tilde{\mathbf{\Pi}}_{ij}| \leq \tilde{\mathbf{\Pi}}_{ij}(e^\delta - 1). \tag{30}$$

It follows that  $\|\mathbf{\Pi}^* - \tilde{\mathbf{\Pi}}\|_1 \leq \sum_{i,j=1}^n \tilde{\mathbf{\Pi}}_{ij}(e^\delta - 1) = e^\delta - 1$ , where the 1-norm is taken over all entries. From (30), we also obtain  $-\delta \leq \log(\mathbf{\Pi}_{ij}^*) - \log(\tilde{\mathbf{\Pi}}_{ij}) \leq \delta$ , and so

$$\|\log(\mathbf{\Pi}^*) - \log(\tilde{\mathbf{\Pi}})\|_\infty \leq \delta,$$

where the  $\infty$ -norm is taken over entries. Now, note that

$$\begin{aligned} & |\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)| \\ &= \left| \sum_{i,j=1}^n \mathcal{C}_{ij}(\mathbf{\Pi}_{ij}^* - \tilde{\mathbf{\Pi}}_{ij}) + \varepsilon \sum_{i,j=1}^n (\mathbf{\Pi}_{ij}^* \log(n^2 \mathbf{\Pi}_{ij}^*) - \tilde{\mathbf{\Pi}}_{ij} \log(n^2 \tilde{\mathbf{\Pi}}_{ij})) \right|. \end{aligned} \tag{31}$$

We control each term as follows,

$$\left| \sum_{i,j=1}^n \mathcal{C}_{ij}(\mathbf{\Pi}_{ij}^* - \tilde{\mathbf{\Pi}}_{ij}) \right| \leq \|\mathcal{C}\|_\infty \|\mathbf{\Pi}^* - \tilde{\mathbf{\Pi}}\|_1 \leq \|\mathcal{C}\|_\infty (e^\delta - 1), \tag{32}$$

where the first inequality is due to Hölder’s inequality. Similarly,

$$\begin{aligned} & \left| \sum_{i,j=1}^n (\mathbf{\Pi}_{ij}^* \log(\mathbf{\Pi}_{ij}^*) - \tilde{\mathbf{\Pi}}_{ij} \log(\tilde{\mathbf{\Pi}}_{ij})) \right| \\ &= \left| \sum_{i,j=1}^n (\mathbf{\Pi}_{ij}^* \log(\mathbf{\Pi}_{ij}^*) - \tilde{\mathbf{\Pi}}_{ij} \log(\mathbf{\Pi}_{ij}^*) + \tilde{\mathbf{\Pi}}_{ij} \log(\mathbf{\Pi}_{ij}^*) - \tilde{\mathbf{\Pi}}_{ij} \log(\tilde{\mathbf{\Pi}}_{ij})) \right| \\ &\leq \|\mathbf{\Pi}^* - \tilde{\mathbf{\Pi}}\|_1 \|\log(\mathbf{\Pi}^*)\|_\infty + \|\tilde{\mathbf{\Pi}}\|_1 \|\log(\mathbf{\Pi}^*) - \log(\tilde{\mathbf{\Pi}})\|_\infty \\ &\leq (e^\delta - 1) \|\log(\mathbf{\Pi}^*)\|_\infty + \delta. \end{aligned} \tag{33}$$

It suffices, therefore, to bound  $\|\log(\mathbf{\Pi}^*)\|_\infty$ . From (5), we have, for  $i, j \in \{1, \dots, n\}$ ,

$$n^2 \mathbf{\Pi}_{ij}^* = e^{\frac{\varphi_1(x^{1,i}) + \varphi_2(x^{2,j}) - c(x^{1,i}, x^{2,j})}{\varepsilon}},$$

such that  $\log(\mathbf{\Pi}_{ij}^*) = \varepsilon^{-1} (\varphi_1(x^{1,i}) + \varphi_2(x^{2,j}) - c(x^{1,i}, x^{2,j})) - 2 \log(n)$ . Consequently,  $\|\log(\mathbf{\Pi}^*)\|_\infty \leq 5\varepsilon^{-1} \|c\|_{\infty, \mathcal{X} \times \mathcal{X}} + 2 \log(n)$ , choosing a version of the EOT potentials  $(\varphi_1, \varphi_2)$  satisfying (6) with  $\|\varphi_1\|_{\infty, \mathcal{X}} \vee \|\varphi_2\|_{\infty, \mathcal{X}} \leq 2\|c\|_{\infty, \mathcal{X} \times \mathcal{X}}$  (see Lemma 2.1 in [74]). Combining (31)-(33) with this bound proves the claimed result,

$$|\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)| \leq (6\|c\|_{\infty, \mathcal{X} \times \mathcal{X}} + 2\varepsilon \log(n)) (e^\delta - 1) + \varepsilon \delta.$$

□

The iteration complexity required to achieve the desired precision of  $o_P(n^{-1})$  for approximating  $\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  is provided next.

**Corollary 2.** *In the setting of Proposition 5, Sinkhorn's algorithm produces an approximation  $\tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  of  $\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$  with  $|\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)| = o_P(n^{-1})$  in  $o_P(\log(n \log(n)))$  iterations. The  $o_P(n^{-1})$  cost approximation also holds when using  $\mathbf{\Pi} = \mathbf{\Pi}^k$ , which satisfies the termination condition  $\|(\mathbf{\Pi}^k)^\top \mathbb{1}_n - \mathbb{1}_n/n\| = o_P\left(\frac{1}{n^2 \log(n)}\right)$ , which, in turn, requires  $o_P(n^2 (\log n)^2)$  iterations.*

*Proof.* The result follows from Proposition 5 by setting  $\delta(n) = o_P\left(\frac{1}{n \log(n)}\right)$ , as

$$e^{\delta(n)} - 1 = \delta(n) + O((\delta(n))^2) \text{ as } n \rightarrow \infty,$$

such that

$$(6\|c\|_{\infty, \mathcal{X} \times \mathcal{X}} + 2\varepsilon \log(n)) (e^{\delta(n)} - 1) + \varepsilon \delta(n) = o_P(n^{-1}),$$

whereby  $|\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2) - \tilde{\mathcal{S}}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)| = o_P(n^{-1})$ . □

In light of (8), the implications of Corollary 2 also hold for the Sinkhorn divergence by approximating  $\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^2)$ ,  $\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^1, \hat{\mu}_n^1)$ , and  $\mathcal{S}_{c,\varepsilon}(\hat{\mu}_n^2, \hat{\mu}_n^2)$  using Sinkhorn's algorithm. In our experiments,  $n = 5000$  and the termination condition  $\|(\mathbf{\Pi}^k)^\top \mathbb{1}_n - \mathbb{1}_n/n\| \leq 1e^{-9}$  is always met prior to the completion of 1000 iterations. As  $\frac{1}{5000^2 \log(5000)} \sim 4.7e^{-9}$ , the chosen threshold,  $1e^{-9}$ , is of a reasonable magnitude.

## Appendix B: Auxiliary proofs

### B.1. Proof of Lemma 1

The results are standard (except possibly (iii)), but we include the proof for completeness.

(i). The argument is similar to Proposition 6 in [65]. Pick any pair of EOT potentials  $(\varphi_1^\circ, \varphi_2^\circ)$ . Update  $(\varphi_1^\circ, \varphi_2^\circ)$  as

$$\begin{aligned} \varphi_1(x_1) &= -\varepsilon \log \int e^{\frac{\varphi_2^\circ(x_2) - c(x_1, x_2)}{\varepsilon}} d\mu^2(x_2), \quad x_1 \in \mathcal{X}, \\ \varphi_2(x_2) &= -\varepsilon \log \int e^{\frac{\varphi_1(x_1) - c(x_1, x_2)}{\varepsilon}} d\mu^1(x_1), \quad x_2 \in \mathcal{X}. \end{aligned}$$

The functions  $(\varphi_1, \varphi_2)$  are well-defined pointwise by Jensen’s inequality. By construction,  $\int e^{\frac{\varphi_1(x'_1) + \varphi_2(x_2) - c(x'_1, x_2)}{\varepsilon}} d\mu^1(x'_1) - 1 = 0$  for all  $x_2 \in \mathcal{X}$ . Also, by Jensen’s inequality,

$$\begin{aligned} &\int (\varphi_1 - \varphi_1^\circ) d\mu^1 + \int (\varphi_2 - \varphi_2^\circ) d\mu^2 \\ &= -\varepsilon \int \log e^{\frac{\varphi_1^\circ - \varphi_1}{\varepsilon}} d\mu^1 - \varepsilon \int \log e^{\frac{\varphi_2^\circ - \varphi_2}{\varepsilon}} d\mu^2 \\ &\geq -\varepsilon \log \int e^{\frac{\varphi_1^\circ - \varphi_1}{\varepsilon}} d\mu^1 - \varepsilon \log \int e^{\frac{\varphi_2^\circ - \varphi_2}{\varepsilon}} d\mu^2 \\ &= -\varepsilon \log \int e^{\frac{\varphi_1^\circ \oplus \varphi_2^\circ - c}{\varepsilon}} d(\mu^1 \otimes \mu^2) - \varepsilon \log \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d(\mu^1 \otimes \mu^2) \\ &= 0, \end{aligned}$$

so that  $(\varphi_1, \varphi_2)$  is a pair of EOT potentials and the inequality above is an equality. In particular,  $\int \log e^{\frac{\varphi_2^\circ - \varphi_2}{\varepsilon}} d\mu^2 = \log \int e^{\frac{\varphi_2^\circ - \varphi_2}{\varepsilon}} d\mu^2$ , and by strict concavity of the logarithm, we have  $e^{\frac{\varphi_2^\circ - \varphi_2}{\varepsilon}} = 1$ , i.e.,  $\varphi_2^\circ = \varphi_2$   $\mu^2$ -a.e. Thus,

$$\int e^{\frac{\varphi_1(x_1) + \varphi_2(x'_2) - c(x_1, x'_2)}{\varepsilon}} d\mu^2(x'_2) = \int e^{\frac{\varphi_1(x_1) + \varphi_2^\circ(x'_2) - c(x_1, x'_2)}{\varepsilon}} d\mu^2(x'_2) = 1$$

for all  $x_1 \in \mathcal{X}$ . The other claims are straightforward.

(ii). This follows from the expressions

$$\begin{aligned} \varphi_1^\mu(x_1) &= -\varepsilon \log \int e^{\frac{\varphi_2^\mu(x_2) - c(x_1, x_2)}{\varepsilon}} d\mu^2(x_2), \quad x_1 \in \mathcal{X}, \\ \varphi_2^\mu(x_2) &= -\varepsilon \log \int e^{\frac{\varphi_1^\mu(x_1) - c(x_1, x_2)}{\varepsilon}} d\mu^1(x_1), \quad x_2 \in \mathcal{X}. \end{aligned}$$

First, by Lemma 2.1 in [74], there exists a version of EOT potentials  $(\varphi_1, \varphi_2)$  satisfying (6) with  $\|\varphi_1\|_{\infty, \mathcal{X}} \vee \|\varphi_2\|_{\infty, \mathcal{X}} \leq 2\|c\|_{\infty, \mathcal{X} \times \mathcal{X}}$ . By uniqueness,  $\varphi_1^\mu = \varphi_1 - \frac{1}{2}(\varphi_1(x_1^\circ) - \varphi_2(x_2^\circ))$  and  $\varphi_2^\mu = \varphi_2 + \frac{1}{2}(\varphi_1(x_1^\circ) - \varphi_2(x_2^\circ))$ , so  $\|\varphi_1^\mu\|_{\infty, \mathcal{X}} \vee \|\varphi_2^\mu\|_{\infty, \mathcal{X}} \leq 3\|c\|_{\infty, \mathcal{X} \times \mathcal{X}}$ . Derivatives of  $\varphi_1^\mu$  and  $\varphi_2^\mu$  can be evaluated by interchanging differentiation and integration, which is guaranteed under the current assumption. See [44, 65] for similar arguments.

(iii). For notational convenience, let  $\varphi_n = \varphi^{\mu_n}$  and  $\varphi = \varphi^\mu$ . By Part (ii), the Ascoli-Arzelà theorem, and the diagonal argument, for every subsequence  $n'$ , there exists a further subsequence  $n''$  along which the derivatives  $D^k \varphi_{n, i}$

converge in  $\mathcal{C}(\mathcal{X})$  for all  $i = 1, 2$  and  $k \in \mathbb{N}_0^d$  with  $|k| \leq s$ , which implies that  $\varphi_{n'',i}$  is Cauchy in  $\mathcal{C}^s(\mathcal{X})$  for  $i = 1, 2$ . By completeness of  $\mathcal{C}^s(\mathcal{X})$ , we have  $\varphi_{n'',i} \rightarrow \bar{\varphi}_i$  in  $\mathcal{C}^s(\mathcal{X})$  for  $i = 1, 2$ . By assumption,  $\mu_n^i$  converges weakly to  $\mu^i$ , which implies that  $\sup_{f \in \text{BL}_1(\mathcal{X})} |\int f d(\mu_n^i - \mu^i)| \rightarrow 0$ , where  $\text{BL}_1(\mathcal{X})$  is the class of 1-Lipschitz functions  $f : \mathcal{X} \rightarrow [-1, 1]$  (cf. Chapter 1.12 in [91]). Again, by Part (ii), we see that

$$\sup_{n \in \mathbb{N}} \sup_{(x_1, x_2) \in \mathcal{X} \times \mathcal{X}} \left\| \nabla_{x_i} \left( e^{\frac{\varphi_{n,1}(x_1) + \varphi_{n,2}(x_2) - c(x_1, x_2)}{\varepsilon}} \right) \right\| < \infty.$$

Hence, for each fixed  $x_1 \in \mathcal{X}$ ,

$$\begin{aligned} \int e^{\frac{\varphi_{n'',1}(x_1) + \varphi_{n'',2}(x_2) - c(x_1, x_2)}{\varepsilon}} d\mu_{n''}^2(x_2) \\ &= \int e^{\frac{\varphi_{n'',1}(x_1) + \varphi_{n'',2}(x_2) - c(x_1, x_2)}{\varepsilon}} d\mu^2(x_2) + o(1) \\ &= \int e^{\frac{\bar{\varphi}_1(x_1) + \bar{\varphi}_2(x_2) - c(x_1, x_2)}{\varepsilon}} d\mu^2(x_2) + o(1), \end{aligned}$$

where the second equality follows as  $\|\varphi_{n'',i} - \bar{\varphi}_i\|_{\infty, \mathcal{X}} \rightarrow 0$ . Since the left-hand side is  $\equiv 1$ , we conclude that  $\int e^{\frac{\bar{\varphi}_1 \oplus \bar{\varphi}_2 - c}{\varepsilon}} d\mu^2 \equiv 1$ . By symmetry, we also have  $\int e^{\frac{\bar{\varphi}_1 \oplus \bar{\varphi}_2 - c}{\varepsilon}} d\mu^1 \equiv 1$ . By construction,  $\bar{\varphi}_1(x_1^\circ) = \bar{\varphi}_2(x_2^\circ)$ , so that  $\bar{\varphi} = \varphi$ , i.e.,  $\varphi_{n''} \rightarrow \varphi$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ . Since the limit does not depend on the choice of subsequence, we have  $\varphi_n \rightarrow \varphi$  in  $\mathcal{C}^s(\mathcal{X}) \times \mathcal{C}^s(\mathcal{X})$ .  $\square$

## B.2. Proof of Lemma 6

The proof relies on the results from [17]. We will use the following observation throughout the proof: for any continuous functions  $f, g$  on  $S_i = \text{spt}(\mu^i)$ , if  $f = g$   $\mu^i$ -a.e., then  $f \equiv g$ . Indeed, for every  $A \subset S_i$  with  $\mu^i$ -measure 1, its closure  $\bar{A}$  agrees with  $S_i$ , since otherwise  $S_i \setminus \bar{A}$  is a nonempty open set in  $S_i$ , so  $\mu^i(S_i \setminus \bar{A}) > 0$ , which contradicts the assumption that  $A$  has  $\mu^i$ -measure 1. Also, for any continuous function  $f$  on  $S_i$ , we have  $\|f\|_{L^\infty(\mu^i)} = \|f\|_{\infty, S_i}$ .

(i). Consider the map  $\mathcal{T} : L^\infty(\mu^1) \times L^\infty(\mu^2) \rightarrow L^\infty(\mu^1) \times L^\infty(\mu^2)$  defined by

$$\mathcal{T}(\varphi) = \left( \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\mu^2, \int e^{\frac{\varphi_1 \oplus \varphi_2 - c}{\varepsilon}} d\mu^1 \right).$$

By Theorem 4.3 in [17],  $\mathcal{T}$  is injective in the sense that, if  $\mathcal{T}(\varphi) = \mathcal{T}(\tilde{\varphi})$  ( $\mu^1 \otimes \mu^2$ )-a.e., then there exists a constant  $a \in \mathbb{R}$  such that  $(\tilde{\varphi}_1, \tilde{\varphi}_2) = (\varphi_1 + a, \varphi_2 - a)$  ( $\mu^1 \otimes \mu^2$ )-a.e. If  $\varphi, \tilde{\varphi} \in \Theta^s$ , then  $(\tilde{\varphi}_1, \tilde{\varphi}_2) \equiv (\varphi_1 + a, \varphi_2 - a)$  on  $S$ , but because of the normalization  $\varphi_1(x_1^\circ) = \varphi_2(x_2^\circ)$  and  $\tilde{\varphi}_1(x_1^\circ) = \tilde{\varphi}_2(x_2^\circ)$ , we have  $a = 0$ , i.e.,  $\tilde{\varphi} \equiv \varphi$ . This shows injectivity of  $\Psi_\mu$ .

To show that the inverse of  $\Psi_\mu$  is continuous at 0, it suffices to show that

$$\varphi_n \in \Theta^s, \|\Psi_\mu(\varphi_n)\|_{\mathbb{D}} \rightarrow 0 \Rightarrow \|\varphi_n - \varphi^\mu\|_S \|_{\mathbb{D}} \rightarrow 0.$$



Recall  $\Psi_\mu(\varphi^\mu|_S) \equiv 0$  on  $S$ . Observe that the map  $\Psi_\mu$  makes sense on  $\mathbb{D}$  and is continuous from  $\mathbb{D}$  into  $\mathbb{D}$ . Since  $\varphi_n \in \Theta^s$ , by the Ascoli-Arzelà theorem, for any subsequence  $n'$ , there exists a further subsequence  $n''$  along which  $\varphi_{n''} \rightarrow \bar{\varphi}$  in  $\mathbb{D}$  for some  $\bar{\varphi} \in \mathbb{D}$ . Since  $\Psi_\mu$  is continuous from  $\mathbb{D}$  into  $\mathbb{D}$ , we have  $\Psi_\mu(\bar{\varphi}) \equiv 0$ . By construction,  $\bar{\varphi}_1(x_1^\circ) = \bar{\varphi}_2(x_2^\circ)$ , so by Lemma 1 (i), we have  $\bar{\varphi} \equiv \varphi^\mu|_S$ . Since the limit  $\varphi^\mu|_S$  is independent of the choice of subsequence, we have  $\varphi_n \rightarrow \varphi^\mu|_S$  in  $\mathbb{D}$ .

(ii). The first claim is straightforward. To show the second claim, it suffices to show that

$$\inf_{\mathbf{h} \in \text{lin}(\Theta^s), \|\mathbf{h}\|_{\mathbb{D}}=1} \|\dot{\Psi}_\mu(\mathbf{h})\|_{\mathbb{D}} > 0.$$

Consider the map  $\dot{\mathcal{T}}_\mu : L^\infty(\mu^1) \times L^\infty(\mu^2) \rightarrow L^\infty(\mu^1) \times L^\infty(\mu^2)$  defined by

$$\dot{\mathcal{T}}_\mu(\mathbf{h}) = \left( \varepsilon^{-1} \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2) d\mu^2, \varepsilon^{-1} \int e^{\frac{\varphi_1^\mu \oplus \varphi_2^\mu - c}{\varepsilon}} (h_1 \oplus h_2) d\mu^1 \right).$$

Equip  $L^\infty(\mu^1) \times L^\infty(\mu^2)$  with a product norm  $\|\mathbf{h}\|_{L^\infty(\mu)} := \|h_1\|_{L^\infty(\mu^1)} \vee \|h_2\|_{L^\infty(\mu^2)}$ . By Proposition 3.1 in [17],

$$\inf_{\|\mathbf{h}\|_{L^\infty(\mu)}=1, \int h_1 d\mu^1=0} \|\dot{\mathcal{T}}_\mu(\mathbf{h})\|_{L^\infty(\mu)} > 0.$$

For  $\mathbf{h} \in \text{lin}(\Theta^s)$ , define  $\bar{\mathbf{h}} = (h_1 - \int h_1 d\mu^1, h_2 + \int h_1 d\mu^1)$ . Then  $\|\dot{\mathcal{T}}_\mu(\bar{\mathbf{h}})\|_{L^\infty(\mu)} = \|\dot{\mathcal{T}}_\mu(\mathbf{h})\|_{L^\infty(\mu)} = \|\dot{\Psi}_\mu(\mathbf{h})\|_{\mathbb{D}}$ . It remains to show that

$$\inf_{\mathbf{h} \in \text{lin}(\Theta^s), \|\mathbf{h}\|_{\mathbb{D}}=1} \|\bar{\mathbf{h}}\|_{\mathbb{D}} > 0.$$

Suppose on the contrary that  $\inf_{\mathbf{h} \in \text{lin}(\Theta^s), \|\mathbf{h}\|_{\mathbb{D}}=1} \|\bar{\mathbf{h}}\|_{\mathbb{D}} = 0$ . Then, there exists a sequence  $\mathbf{h}_n \in \text{lin}(\Theta^s)$  with  $\|\mathbf{h}_n\|_{\mathbb{D}} = 1$  such that  $\|\bar{\mathbf{h}}_n\|_{\mathbb{D}} \rightarrow 0$ . Since  $\int h_{n,1} d\mu^1$  is bounded, there exists a subsequence along which  $\int h_{n,1} d\mu^1 \rightarrow a$  for some  $a \in \mathbb{R}$ . Along the subsequence,  $\|\mathbf{h}_n - (a, -a)\|_{\mathbb{D}} \rightarrow 0$ . However, since  $h_{n,1}(x_1^\circ) = h_{n,2}(x_2^\circ)$  by construction, we must have  $a = 0$ , i.e.,  $\|\mathbf{h}_n\|_{\mathbb{D}} \rightarrow 0$ , which contradicts the assumption that  $\|\mathbf{h}_n\|_{\mathbb{D}} = 1$ .  $\square$

## Appendix C: Technical tools

### C.1. Hadamard differentiability and functional delta method

In this appendix, we review concepts of Hadamard differentiability and the functional delta methods. Our exposition mostly follows [85]. Other standard references are [91, 90].

Let  $\mathfrak{D}, \mathfrak{E}$  be normed spaces and  $\phi : \Theta \subset \mathfrak{D} \rightarrow \mathfrak{E}$  be a map. We say that  $\phi$  is *Hadamard directionally differentiable* at  $\theta \in \Theta$  if there exists a map  $\phi'_\theta : \mathfrak{X}_\Theta(\theta) \rightarrow \mathfrak{E}$  such that

$$\lim_{t \downarrow 0} \frac{\phi(\theta_t) - \phi(\theta)}{t} = \phi'_\theta(h) \quad (34)$$

for any sequence  $(\theta_t)_{t>0} \subset \Theta$  with  $t^{-1}(\theta_t - \theta) \rightarrow h$  as  $t \downarrow 0$ , where  $\mathfrak{T}_\Theta(\theta)$  is the tangent (or adjacent) cone to  $\Theta$  at  $\theta$ ,

$$\mathfrak{T}_\Theta(\theta) = \left\{ h \in \mathfrak{D} : h = \lim_{t \downarrow 0} \frac{\theta_t - \theta}{t} \text{ for some } \theta_t \rightarrow \theta \text{ in } \Theta, t \downarrow 0 \right\}.$$

The derivative  $\phi'_\theta$  is continuous (cf. Proposition 3.1 in [89]) and positively homogeneous (but need not be linear). Furthermore, the tangent cone  $\mathfrak{T}_\Theta(\theta)$  is closed, and: (i) if  $\Theta$  is open, then  $\mathfrak{T}_\Theta(\theta)$  agrees with  $\mathfrak{D}$ ; and (ii) if  $\Theta$  is convex, then  $\mathfrak{T}_\Theta(\theta)$  agrees with  $\overline{\{t(\vartheta - \theta) : \vartheta \in \Theta, t > 0\}}^\mathfrak{D}$  (cf. Chapter 4 in [5]).<sup>7</sup>

If (34) only holds for  $h \in \mathfrak{D}_0$  for a subset  $\mathfrak{D}_0 \subset \mathfrak{T}_\Theta(\theta)$ , then we say that  $\phi$  is Hadamard directionally differentiable at  $\theta \in \Theta$  *tangentially* to  $\mathfrak{D}_0$ . In that case, the derivative  $\phi'_\theta$  is defined only on  $\mathfrak{D}_0$ . Finally, if the derivative  $\phi'_\theta$  is linear, then we say that  $\phi$  is *Hadamard differentiable* at  $\theta$  (tangentially to  $\mathfrak{D}_0$  if  $\phi'_\theta$  is defined only on  $\mathfrak{D}_0$ ). The tangent set  $\mathfrak{D}_0$  need not be a vector subspace of  $\mathfrak{D}$ , so by linearity, we mean that, for any  $\alpha_1, \dots, \alpha_J \in \mathbb{R}$  and  $h_1, \dots, h_J \in \mathfrak{D}_0$ , whenever  $\sum_{j=1}^J \alpha_j h_j = 0$ , it holds that  $\sum_{j=1}^J \alpha_j \phi'_{\theta_0}(h_j) = 0$ , which is equivalent to  $\phi'_{\theta_0}$  admitting a linear extension to the linear hull of  $\mathfrak{D}_0$  by Lemma 2.5.3 in [35].

**Lemma 11** (Functional delta method; [85]). *Let  $\mathfrak{D}, \mathfrak{E}$  be normed spaces and  $\phi : \Theta \subset \mathfrak{D} \rightarrow \mathfrak{E}$  be a map that is Hadamard directionally differentiable at  $\theta \in \Theta$  tangentially to a set  $\mathfrak{D}_0 \subset \mathfrak{T}_\Theta(\theta)$ . Let  $T_n : \Omega \rightarrow \Theta$  be maps such that  $r_n(T_n - \theta) \xrightarrow{d} T$  for some  $r_n \rightarrow \infty$  and Borel measurable map  $T : \Omega \rightarrow \mathfrak{D}$  with values in a separable subset of  $\mathfrak{D}_0$ . Then the following hold: (i)  $r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'_\theta(T)$ ; (ii) If in addition  $\Theta$  is convex and  $\mathfrak{D}_0 = \mathfrak{T}_\Theta(\theta)$ , then  $r_n(\phi(T_n) - \phi(\theta)) - \phi'_\theta(r_n(T_n - \theta)) \rightarrow 0$  in outer probability.*

*Remark 16.* Our definition of Hadamard differentiability is slightly different from [91, 90], in that those references do not require the tangent set  $\mathfrak{D}_0$  to be a subset of  $\mathfrak{T}_\Theta(\theta)$ . Our modification is made to be consistent with the definition of Hadamard directional differentiability in [85]. However, this modification is innocuous since, for any sequence  $(\theta_t)_{t>0} \subset \Theta$  with  $h := \lim_{t \downarrow 0} t^{-1}(\theta_t - \theta)$  (if exists), we must have  $h \in \mathfrak{T}_\Theta(\theta)$ .

The proof of Theorem 2 relies on the second-order functional delta method, which we describe next. We say that a map  $\phi : \Theta \subset \mathfrak{D} \rightarrow \mathfrak{E}$  (with  $\Theta$  being convex) is *second-order Hadamard directionally differentiable* at  $\theta \in \Theta$  if it is (first-order) Hadamard directionally differentiable at  $\theta$  and there exists a map  $\phi''_\theta : \mathfrak{T}_\Theta(\theta) \rightarrow \mathfrak{E}$  such that

$$\lim_{t \downarrow 0} \frac{\phi(\theta_t) - \phi(\theta) - t\phi'_\theta(h_t)}{t^2/2} = \phi''_\theta(h)$$

<sup>7</sup>Indeed, [85] defines the Hadamard derivative on the (Bouligand) contingent cone, which is in general slightly bigger than the adjacent cone. For our purpose, this difference is immaterial. Note that both concepts agree when  $\Theta$  is convex.

for any sequence  $(\theta_t)_{t>0} \subset \Theta$  with  $h_t := t^{-1}(\theta_t - \theta) \rightarrow h$  as  $t \downarrow 0$  (note: as  $\Theta$  is convex,  $h_t \in \mathfrak{X}_\Theta(\theta)$ , so that  $\phi'_\theta(h_t)$  is well-defined). The map  $\phi''_\theta$  is continuous and positively homogeneous of degree 2.

**Lemma 12** (Second-order chain rule). *Let  $\mathfrak{D}, \mathfrak{E}, \mathfrak{F}$  be normed spaces,  $\Theta \subset \mathfrak{D}$ ,  $\Xi \subset \mathfrak{E}$  be convex,  $\phi : \Theta \rightarrow \Xi$  be a map that is twice Hadamard directionally differentiable at  $\theta \in \Theta$ , and  $\psi : \Xi \rightarrow \mathfrak{F}$  be twice Hadamard directionally differentiable at  $\phi(\theta)$  with linear first-order derivative at  $\phi(\theta)$  defined on the linear hull of  $\mathfrak{X}_\Xi(\phi(\theta))$ . Then, the composition  $\psi \circ \phi : \Theta \rightarrow \mathfrak{F}$  is twice Hadamard directionally differentiable at  $\theta$  with*

$$[\psi \circ \phi]''_\theta = \psi''_{\phi(\theta)} \circ \phi'_\theta + \psi'_{\phi(\theta)} \circ \phi''_\theta.$$

*Proof.* Let  $(\theta_t)_{t>0} \subset \Theta$  be such that  $h_t := t^{-1}(\theta_t - \theta) \rightarrow h$  in  $\mathfrak{D}$  as  $t \downarrow 0$ . Observe that  $(\phi(\theta_t))_{t>0} \subset \Xi$  with  $t^{-1}(\phi(\theta_t) - \phi(\theta)) \rightarrow \phi'_\theta(h)$  as  $t \downarrow 0$ , so that

$$\psi(\phi(\theta_t)) - \psi(\phi(\theta)) = t\psi'_{\phi(\theta)}(t^{-1}(\phi(\theta_t) - \phi(\theta))) + \frac{t^2}{2}\psi''_{\phi(\theta)}(\phi'_\theta(h)) + o(t^2).$$

By linearity of  $\psi'_{\phi(\theta)}$ ,

$$t\psi'_{\phi(\theta)}(t^{-1}(\phi(\theta_t) - \phi(\theta))) - t\psi'_{\phi(\theta)}(\phi'_\theta(h_t)) = \psi'_{\phi(\theta)}(\phi(\theta_t) - \phi(\theta) - t\phi'_\theta(h_t)).$$

It follows from continuity and positive homogeneity of  $\psi'_{\phi(\theta)}$  that

$$\lim_{t \downarrow 0} \frac{\psi(\phi(\theta_t)) - \psi(\phi(\theta)) - t\psi'_{\phi(\theta)}(\phi'_\theta(h_t))}{t^2/2} = \psi''_{\phi(\theta)}(\phi'_\theta(h)) + \psi'_{\phi(\theta)}(\phi''_\theta(h)).$$

□

**Lemma 13** (Second-order functional delta method; [85]). *Let  $\mathfrak{D}, \mathfrak{E}$  be normed spaces,  $\Theta \subset \mathfrak{D}$  be convex, and  $\phi : \Theta \rightarrow \mathfrak{E}$  be a map that is second-order Hadamard directionally differentiable at  $\theta \in \Theta$ . Let  $T_n : \Omega \rightarrow \Theta$  be maps such that  $r_n(T_n - \theta) \xrightarrow{d} T$  for some  $r_n \rightarrow \infty$  and Borel measurable map  $T : \Omega \rightarrow \mathfrak{D}$  with values in a separable subset of  $\mathfrak{X}_\Theta(\theta)$ . Then,  $r_n^2(\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta)) \xrightarrow{d} \frac{1}{2}\phi''_\theta(T)$  and  $r_n^2(\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta) - \frac{1}{2}\phi''_\theta(T_n - \theta)) \rightarrow 0$  in outer probability.*

### C.2. Hadamard differentiability of Z-functional

Let  $\mathbb{D}$  be a Banach space,  $\Theta$  be an arbitrary nonempty subset of  $\mathbb{D}$ , and  $\mathbb{L}$  be another Banach space. Let  $Z(\Theta, \mathbb{L})$  be the subset of  $\ell^\infty(\Theta, \mathbb{L})$  consisting of all maps with at least one zero, and let  $\phi : Z(\Theta, \mathbb{L}) \rightarrow \Theta$  be a map that assigns one of its zeros of to each  $z \in Z(\Theta, \mathbb{L})$ . Following [91], we call  $\phi$  the Z-functional. We say that a map  $\Psi : \Theta \subset \mathbb{D} \rightarrow \mathbb{L}$  is Fréchet differentiable at  $\theta_0 \in \Theta$  if there exists a bounded linear operator  $\dot{\Psi}_{\theta_0} : \text{lin}(\Theta) \rightarrow \mathbb{L}$  such that

$$\lim_{\substack{\|h\|_{\mathbb{D}} \rightarrow 0 \\ \theta_0 + h \in \Theta}} \frac{\|\Psi(\theta_0 + h) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(h)\|_{\mathbb{L}}}{\|h\|_{\mathbb{D}}} = 0.$$

The following is taken from Lemma 3.9.34 in [91], with a minor modification to adapt to our definition of Hadamard differentiability; cf. Remark 16.

**Lemma 14** (Hadamard derivative of  $Z$ -functional). *Suppose that (i)  $\Psi : \Theta \rightarrow \mathbb{L}$  is uniformly norm-bounded, one-to-one, possesses a zero at  $\theta_0 \in \Theta$ , and has an inverse that is continuous at 0, and (ii)  $\Psi$  is Fréchet differentiable at  $\theta_0 \in \Theta$  with derivative  $\dot{\Psi}_{\theta_0}$  that is one-to-one and such that its inverse is continuous on  $\dot{\Psi}_{\theta_0}(\text{lin}(\Theta))$ . Then,  $\phi$  is Hadamard differentiable at  $\Psi$  tangentially to the set*

$$\mathcal{Z}_{\Psi} = \left\{ z \in \ell^{\infty}(\Theta, \mathbb{L}) : z = \lim_{t \downarrow 0} \frac{z_t - \Psi}{t} \text{ for some } z_t \rightarrow \Psi \text{ in } Z(\Theta, \mathbb{L}), t \downarrow 0 \right\} \\ \cap \left\{ z \in \ell^{\infty}(\Theta, \mathbb{L}) : z \text{ is continuous at } \theta_0 \right\}.$$

The derivative is given by  $\phi'_{\Psi}(z) = -\dot{\Psi}_{\theta_0}^{-1}(z(\theta_0))$ .

*Remark 17.* It is *a priori* not clear whether  $z(\theta_0)$  is in the domain of  $\dot{\Psi}_{\theta_0}^{-1}$ , but this follows from the proof of Lemma 3.9.34 in [91]. Indeed, let  $(z_t)_{t>0} \subset \ell^{\infty}(\Theta, \mathbb{L})$  be such that  $\Psi + tz_t \in Z(\Theta, \mathbb{L})$  for  $t$  sufficiently small,  $z_t \rightarrow z$  as  $t \downarrow 0$ , and  $z$  is continuous at  $\theta_0$ . Set  $\theta_t = \phi(\Psi + tz_t)$ . Arguing as in the proof of Lemma 3.9.34 in [91], we have  $\Psi(\theta_t) = \dot{\Psi}_{\theta_0}(\theta_t - \theta_0) + R(t)$  with  $\|R(t)\|_{\mathbb{L}} = o(t)$  as  $t \rightarrow 0$ , so that  $t^{-1}\{\Psi(\theta_t) - R(t)\} = \dot{\Psi}_{\theta_0}(t^{-1}(\theta_t - \theta_0)) \in \dot{\Psi}_{\theta_0}(\text{lin}(\Theta))$  while  $t^{-1}\{\Psi(\theta_t) - R(t)\} = -z_t(\theta_t) - t^{-1}R(t) \rightarrow -z(\theta_0)$  in  $\mathbb{L}$ . Hence  $z(\theta_0) \in \overline{\dot{\Psi}_{\theta_0}(\text{lin}(\Theta))}^{\mathbb{L}}$ .

The following lemma concerns the second-order Hadamard derivative for the  $Z$ -functional, which is used in the proof of Theorem 4.

**Lemma 15** (Second-order Hadamard derivative of  $Z$ -functional). *Consider the assumption of the preceding lemma. Assume further that  $\Psi$  is twice Fréchet differentiable at  $\theta_0$ , in the sense that there exists a continuous operator  $\ddot{\Psi}_{\theta_0} : \overline{\text{lin}(\Theta)}^{\mathbb{D}} \rightarrow \mathbb{L}$  positively homogeneous of degree 2 such that*

$$\lim_{\substack{\|h\|_{\mathbb{D}} \rightarrow 0 \\ \theta_0 + h \in \Theta}} \frac{\|\Psi(\theta_0 + h) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(h) - \frac{1}{2}\ddot{\Psi}_{\theta_0}(h)\|_{\mathbb{L}}}{\|h\|_{\mathbb{D}}^2} = 0.$$

Let  $(z_t)_{t>0} \in \ell^{\infty}(\Theta, \mathbb{L})$  be a sequence of maps such that (i)  $\Psi + tz_t \in Z(\Theta, \mathbb{L})$  for  $t$  sufficiently small, (ii)  $z_t \rightarrow z$  as  $t \downarrow 0$  for some  $z$  that is continuous at  $\theta_0$ , (iii)  $z_t(\theta_0) \in \overline{\dot{\Psi}_{\theta_0}(\text{lin}(\Theta))}^{\mathbb{L}}$  for  $t$  sufficiently small, and (iv) there exists a limit  $t^{-1}(z_t(\theta_t) - z_t(\theta_0)) \rightarrow \dot{z}_{\theta_0}$  in  $\mathbb{L}$  as  $t \downarrow 0$  with  $\theta_t = \phi(\Psi + tz_t)$ . Then, we have

$$\frac{\phi(\Psi + tz_t) - \phi(\Psi) - t\phi'_{\Psi}(z_t)}{t^2/2} \rightarrow -\dot{\Psi}_{\theta_0}^{-1}(2\dot{z}_{\theta_0} + \ddot{\Psi}_{\theta_0}(\phi'_{\Psi}(z))).$$

The above definition of second Fréchet derivative differs from the standard one (cf. [95]), but suffices for our purpose.

*Proof.* Recall  $\theta_t = \phi(\Psi + tz_t)$ , i.e.,  $\Psi(\theta_t) + tz_t(\theta_t) = 0$ . Arguing as in the proof

of Lemma 3.9.34 in [91], we have  $\|\theta_t - \theta_0\| = O(t)$ . Thus,

$$\underbrace{\Psi(\theta_t)}_{-tz_t(\theta_t)} - \underbrace{\Psi(\theta_0)}_{=0} = \dot{\Psi}_{\theta_0}(\theta_t - \theta_0) + \frac{1}{2}\ddot{\Psi}_{\theta_0}(\theta_t - \theta_0) + o(t^2).$$

Subtracting  $-tz_t(\theta_0)$  from both sides, we have

$$-t(z_t(\theta_t) - z_t(\theta_0)) = \dot{\Psi}_{\theta_0}(\theta_t - \theta_0 - t\phi'_{\Psi}(z_t)) + \frac{1}{2}\ddot{\Psi}_{\theta_0}(\theta_t - \theta_0) + o(t^2).$$

Since  $t^{-1}(\theta_t - \theta_0) \rightarrow \phi'_{\Psi}(z)$  by the preceding lemma, we have

$$\ddot{\Psi}_{\theta_0}(\theta_t - \theta_0) = t^2\ddot{\Psi}_{\theta_0}(\phi'_{\Psi}(z)) + o(t^2).$$

Also, by assumption,

$$-t(z_t(\theta_t) - z_t(\theta_0)) = -t^2\dot{z}_{\theta_0} + o(t^2).$$

Conclude that

$$\frac{\theta_t - \theta_0 - t\phi'_{\Psi}(z_t)}{t^2/2} \rightarrow -\dot{\Psi}_{\theta_0}^{-1}(2\dot{z}_{\theta_0} + \ddot{\Psi}_{\theta_0}(\phi'_{\Psi}(z))). \quad \square$$

### Appendix D: Other auxiliary results

**Lemma 16** (Convergence in  $\ell^\infty(B^s)$  implies weak convergence). *Let  $\mathcal{X} \subset \mathbb{R}^d$  be a compact set that agrees with the closure of its interior. Pick any  $s \in \mathbb{N}$  and set  $B^s$  to be the unit ball in  $C^s(\mathcal{X})$ . For  $\mu_n, \mu \in \mathcal{P}(\mathcal{X})$ , if  $\mu_n \rightarrow \mu$  in  $\ell^\infty(B^s)$ , then  $\mu_n \rightarrow \mu$  weakly.*

*Proof.* Pick any bounded 1-Lipschitz function  $f$  on  $\mathcal{X}$ . By the Kirszbraun-McShane theorem, we may extend  $f$  to a 1-Lipschitz function on  $\mathbb{R}^d$ , which we denote by the same symbol  $f$ . Let  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  be a compactly supported smooth density function and approximate  $f$  by  $f_t = t^{-d} \int_{\mathbb{R}^d} f(y)K((y-\cdot)/t)dy = \int_{\mathbb{R}^d} f(\cdot + tz)K(z)dz$  for  $t > 0$ . As  $f$  is 1-Lipschitz,  $f_t$  is smooth and 1-Lipshitz with  $\|f - f_t\|_{\infty, \mathbb{R}^d} \leq t$ . The restriction of  $f_t$  to  $\mathcal{X}$  belongs to  $C^s(\mathcal{X})$ , so that for any  $t > 0$ ,

$$\int_{\mathcal{X}} f d\mu_n \leq \int_{\mathcal{X}} f_t d\mu_n + t \leq \int_{\mathcal{X}} f d\mu + 2t + o(1), \quad n \rightarrow \infty.$$

The reverse inequality follows similarly, so that we have  $\limsup_{n \rightarrow \infty} |\int_{\mathcal{X}} f d(\mu_n - \mu)| \leq 2t$ . Sending  $t \downarrow 0$ , we have  $\int_{\mathcal{X}} f d\mu_n \rightarrow \int_{\mathcal{X}} f d\mu$ , implying  $\mu_n \rightarrow \mu$  weakly.  $\square$

**Lemma 17** (Support of Brownian bridge). *Let  $\mu$  be a probability measure on a measurable space  $S$  and  $\mathcal{F} \subset L^2(\mu)$  be a  $\mu$ -pre-Gaussian class, i.e., there exists a tight  $\mu$ -Brownian bridge  $G_\mu$  in  $\ell^\infty(\mathcal{F})$ . Let*

$$\mathfrak{M}_\mu = \left\{ g d\mu : g \text{ is a bounded measurable function on } S \text{ with } \mu\text{-mean zero} \right\}.$$

*Then  $\text{spt}(G_\mu) \subset \overline{\mathfrak{M}_\mu}^{\ell^\infty(\mathcal{F})}$ .*

*Proof.* Let  $\mathcal{C}_u(\mathcal{F})$  denote the space of uniformly continuous functions on  $\mathcal{F}$  relative to the pseudometric  $d_\mu(f, g) = \sqrt{\text{Var}_\mu(f - g)}$ . Since  $\mathcal{F}$  is  $\mu$ -pre-Gaussian,  $\mathcal{F}$  is totally bounded for  $d_\mu$  (so that  $\mathcal{C}_u(\mathcal{F})$  is a closed subspace of  $\ell^\infty(\mathcal{F})$ ) and  $G_\mu \in \mathcal{C}_u(\mathcal{F})$  a.s. (cf. Example 1.5.10 in [91]). For every  $f \in \mathcal{F}$  and  $g_1, g_2 \in L^2(\mu)$  with  $\mu$ -mean zero,

$$\int f(g_1 - g_2) d\mu = \int (f - \int f d\mu)(g_1 - g_2) d\mu \leq \sqrt{\text{Var}_\mu(f)} d_\mu(g_1 - g_2)$$

by the Cauchy-Schwarz inequality. Since  $\sup_{f \in \mathcal{F}} \text{Var}_\mu(f) < \infty$  by total boundedness of  $\mathcal{F}$  w.r.t.  $d_\mu$ , we have that  $g d\mu \in \overline{\mathfrak{M}_\mu}^{\ell^\infty(\mathcal{F})}$  for every  $g \in L^2(\mu)$  with mean zero.

By Lemma 5.1 in [92], the support  $\text{spt}(G_\mu)$  agrees with the  $\|\cdot\|_{\infty, \mathcal{F}}$ -closure of the reproducing kernel Hilbert space (RKHS) for  $G_\mu$  (think of  $G_\mu$  as a zero-mean Gaussian random variable in  $\mathcal{C}_u(\mathcal{F})$ , which is a separable Banach space). There are two ways to define the RKHS for  $G_\mu$ ; by viewing  $G_\mu$  as a stochastic process or as a random variable with values in the Banach space  $\mathcal{C}_u(\mathcal{F})$ . In this case, however, they both agree; see Theorem 2.1 in [92]. With this in mind, any element of the RKHS for  $G_\mu$  is of the form

$$f \mapsto \mathbb{E}[G_\mu(f)X], \quad X \in \overline{\text{lin}\{G_\mu(g) : g \in \mathcal{F}\}}^{L^2(\mathbb{P})}.$$

For  $X = \sum_{i=1}^s \alpha_i G_\mu(g_i)$  with  $\alpha_i \in \mathbb{R}$  and  $g_i \in \mathcal{F}$ , we have

$$\begin{aligned} \mathbb{E}[G_\mu(f)X] &= \sum_{i=1}^s \alpha_i \mathbb{E}[G_\mu(f)G_\mu(g_i)] = \sum_{i=1}^s \alpha_i \text{Cov}_\mu(f, g_i) \\ &= \int f \left( \sum_{i=1}^s \alpha_i (g_i - \int g_i d\mu) \right) d\mu, \end{aligned}$$

so that  $\mathbb{E}[G_\mu(\cdot)X] \in \overline{\mathfrak{M}_\mu}^{\ell^\infty(\mathcal{F})}$ . Further, for any  $X \in \overline{\text{lin}\{G_\mu(g) : g \in \mathcal{F}\}}^{L^2(\mathbb{P})}$ , choose  $X_n \in \text{lin}\{G_\mu(g) : g \in \mathcal{F}\}$  such that  $\mathbb{E}[|X_n - X|^2] \rightarrow 0$ . Then,

$$\|\mathbb{E}[G_\mu(\cdot)X_n] - \mathbb{E}[G_\mu(\cdot)X]\|_{\infty, \mathcal{F}} \leq \sqrt{\sup_{f \in \mathcal{F}} \text{Var}_\mu(f)} \sqrt{\mathbb{E}[|X_n - X|^2]} \rightarrow 0,$$

which shows that  $\mathbb{E}[G_\mu(\cdot)X] \in \overline{\mathfrak{M}_\mu}^{\ell^\infty(\mathcal{F})}$ . Conclude that  $\text{spt}(G_\mu) \subset \overline{\mathfrak{M}_\mu}^{\ell^\infty(\mathcal{F})}$ .  $\square$

**Lemma 18** (Weak convergence in product space). *Let  $S, T$  be nonempty sets and let  $X_n = (X_n(s))_{s \in S}, Y_n = (Y_n(t))_{t \in T}$  be sequences of stochastic processes with bounded paths. Suppose that, marginally,  $X_n$  and  $Y_n$  converge in distribution to tight random variables in  $\ell^\infty(S)$  and  $\ell^\infty(T)$ , respectively. Then, if the finite-dimensional distributions of  $(X_n, Y_n)$  converge weakly, i.e., for every  $s_1, \dots, s_m \in S, t_1, \dots, t_\ell \in T$ ,  $(X_n(s_1), \dots, X_n(s_m), Y_n(t_1), \dots, Y_n(t_\ell))$  jointly converges in distribution (in  $\mathbb{R}^{m+\ell}$ ), then  $(X_n, Y_n) \xrightarrow{d} (X, Y)$  in  $\ell^\infty(S) \times \ell^\infty(T)$  for some tight limit  $(X, Y)$ .*

*Proof.* The lemma follows from Prohorov's theorem, upon observing that a tight limit is uniquely determined by the finite-dimensional convergence. One way to show the latter is to apply Lemma 1.3.12 in [91]. We present another more direct proof. Let  $(X, Y) : \Omega \rightarrow \ell^\infty(S) \times \ell^\infty(T)$  be a tight (Borel measurable) random variable. Then,  $X$  is tight in  $\ell^\infty(S)$ , so there exists a pseudometric  $\rho_S$  on  $S$  that makes  $S$  totally bounded and such that  $X \in \mathcal{C}_u(S)$  a.s., where  $\mathcal{C}_u(S)$  is the space of  $\rho_S$ -uniformly continuous functions on  $S$  equipped with the sup-norm  $\|\cdot\|_{\infty, S}$ ; see Chapter 1.5 in [91]. Consider the Borel  $\sigma$ -field on  $\mathcal{C}_u(S)$ . Define  $\mathcal{C}_u(T)$  analogously. Since  $\mathcal{C}_u(S)$  and  $\mathcal{C}_u(T)$  are separable, the Borel  $\sigma$ -field on  $\mathcal{C}_u(S) \times \mathcal{C}_u(T)$  (defined w.r.t. the product topology) agrees with the product  $\sigma$ -field. In turn, the Borel  $\sigma$ -field on  $\mathcal{C}_u(S)$  agrees with the cylinder  $\sigma$ -field (i.e., the smallest  $\sigma$ -field that makes every coordinate projection  $f \mapsto f(s)$  measurable). For  $s_1, \dots, s_m \in S$ , let  $\pi_{s_1, \dots, s_m}^S : \mathcal{C}_u(S) \rightarrow \mathbb{R}^m$  be the projection onto  $s_1, \dots, s_m$ , i.e.,  $\pi_{s_1, \dots, s_m}^S(f) = (f(s_1), \dots, f(s_m))$ . Define  $\pi_{t_1, \dots, t_\ell}^T$  analogously. Then, the collection of sets of the form

$$\begin{aligned} & [\pi_{s_1, \dots, s_m}^S]^{-1}(A) \times [\pi_{t_1, \dots, t_\ell}^T]^{-1}(B) \subset \mathcal{C}_u(S) \times \mathcal{C}_u(T), \\ & s_i \in S, t_j \in T, A \subset \mathbb{R}^m, B \subset \mathbb{R}^\ell : \text{Borel sets} \end{aligned}$$

is a  $\pi$ -system that generates the Borel  $\sigma$ -field on  $\mathcal{C}_u(S) \times \mathcal{C}_u(T)$ . Hence, the joint law of  $(X, Y)$  is uniquely determined by the collection of the joint laws of random vectors of the form  $(X(s_1), \dots, X(s_m), Y(t_1), \dots, Y(t_\ell))$ .

The rest of the proof is standard. Since  $X_n$  and  $Y_n$  are marginally asymptotically tight and asymptotically measurable (Lemma 1.3.8 in [91]),  $(X_n, Y_n)$  is jointly asymptotically tight and asymptotically measurable (Lemmas 1.4.3 and 1.4.4 in [91]). By Prohorov's theorem (Theorem 1.3.9 in [91]), every subsequence has a further subsequence weakly convergent to a tight law. By finite-dimensional convergence, the weak limit is unique. Hence  $(X_n, Y_n) \xrightarrow{d} (X, Y)$  in  $\ell^\infty(S) \times \ell^\infty(T)$  for some tight limit  $(X, Y)$ .  $\square$

## Acknowledgments

The authors would like to thank Editor Grace Y. Yi, an associate editor, and two anonymous referees for their constructive comments that help improve the quality of the manuscript.

## Funding

Ziv Goldfeld is supported by NSF grants CCF-1947801, CCF-2046018, and DMS-2210368, and the 2020 IBM Academic Award. Kengo Kato is supported by NSF grants DMS-1952306, DMS-2014636, and DMS-2210368. Gabriel Rioux is supported by the NSERC postgraduate fellowship PGSD-567921-2022.

## References

- [1] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011. [MR2801182](#)
- [2] Mélisande Albert, Yann Bouret, Magalie Fromont, and Patricia Reynaud-Bouret. Bootstrap and permutation tests of independence for point processes. *The Annals of Statistics*, 43(6):2537–2564, 2015. [MR3405603](#)
- [3] Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme. Asymptotics for semidiscrete entropic optimal transport. *SIAM Journal on Mathematical Analysis*, 54(2):1718–1741, 2022. [MR4393198](#)
- [4] Jason M. Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1961–1971, 2017.
- [5] Jean-Pierre Aubin and Hélène Frankowska. *Set-valued Analysis*. Springer Science & Business Media, 2009. [MR2458436](#)
- [6] Espen Bernton, Promit Ghosal, and Marcel Nutz. Entropic optimal transport: Geometry and large deviations. *Duke Mathematical Journal*, 171(16):3363–3400, 2022. [MR4505361](#)
- [7] P. J. Bickel, F. Götze, and W. R. van Zwet. Resampling fewer than  $n$  observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31, 1997. [MR1441142](#)
- [8] Jérémie Bigot. Statistical data analysis in the Wasserstein space. *ESAIM: Proceedings and Surveys*, 68:1–19, 2020. [MR4142495](#)
- [9] Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electronic Journal of Statistics*, 13(2):5120–5150, 2019. [MR4041704](#)
- [10] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991. [MR1100809](#)
- [11] Jiezhong Cao, Langyuan Mo, Yifan Zhang, Kui Jia, Chunhua Shen, and Mingkui Tan. Multi-marginal Wasserstein GAN. *Advances in Neural Information Processing Systems*, 32, 2019.
- [12] Guillaume Carlier, Victor Chernozhukov, Gwendoline De Bie, and Alfred Galichon. Vector quantile regression and optimal transport, from theory to numerics. *Empirical Economics*, pages 1–28, 2020.
- [13] Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression: an optimal transport approach. *The Annals of Statistics*, 44(3):1165–1192, 2016. [MR3485957](#)
- [14] Guillaume Carlier, Lénaïc Chizat, and Maxime Laborde. Lipschitz continuity of the Schrödinger map in entropic optimal transport. *arXiv preprint arXiv:2210.00225*, 2022.
- [15] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient



- flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017. [MR3635459](#)
- [16] Guillaume Carlier and Ivar Ekeland. Matching for teams. *Economic theory*, 42:397–418, 2010. [MR2564442](#)
- [17] Guillaume Carlier and Maxime Laborde. A differential approach to the multi-marginal Schrödinger system. *SIAM Journal on Mathematical Analysis*, 52(1):709–717, 2020. [MR4062805](#)
- [18] Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs. *arXiv preprint [arXiv:2206.03347](#)*, 2022. [MR4565039](#)
- [19] Xiaohui Chen and Kengo Kato. Jackknife multiplier bootstrap: finite sample approximations to the  $U$ -process supremum with applications. *Probability Theory and Related Fields*, 176(3):1097–1163, 2020. [MR4087490](#)
- [20] Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256, 2017. [MR3611491](#)
- [21] Pierre-André Chiappori, Robert J McCann, and Lars P Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, 42:317–354, 2010. [MR2564439](#)
- [22] Lenaïc Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster Wasserstein distance estimation with the Sinkhorn divergence. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [23] Maria Colombo, Luigi De Pascale, and Simone Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canadian Journal of Mathematics*, 67(2):350–368, 2015. [MR3314838](#)
- [24] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):108964, 2021. [MR4232667](#)
- [25] G. H. Constantine and T. H. Savits. A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996. [MR1325915](#)
- [26] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Density functional theory and optimal transportation with coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013. [MR3020313](#)
- [27] Marco Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pages 2292–2300, 2013.
- [28] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *International conference on machine learning*, pages 685–693. PMLR, 2014.
- [29] Victor de la Peña and Evarist Giné. *Decoupling: From Dependence to Independence*. Springer, 1999. [MR1666908](#)
- [30] Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Advances in Neural Information Processing Systems*, 34:29736–29753,

- 2021.
- [31] Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for semidiscrete Wasserstein distances. *arXiv preprint arXiv:2202.06380*, 2022. [MR4665589](#)
  - [32] Eustasio del Barrio, Alberto González-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *arXiv preprint arXiv:2204.09105*, 2022. [MR4616887](#)
  - [33] George Deligiannidis, Valentin De Bortoli, and Arnaud Doucet. Quantitative uniform stability of the iterative proportional fitting procedure. *arXiv preprint arXiv:2108.08129*, 2021. [MR4696283](#)
  - [34] Richard M. Dudley. The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969. [MR0236977](#)
  - [35] Richard M. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, 1999. [MR1720712](#)
  - [36] Pavel Dvurechensky, Alexander Gasnikov, and Alexey Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn’s algorithm. In *International conference on machine learning*, pages 1367–1376. PMLR, 2018.
  - [37] Stephan Eckstein and Marcel Nutz. Quantitative stability of regularized optimal transport and convergence of Sinkhorn’s algorithm. *SIAM Journal on Mathematical Analysis*, 54(6):5922–5948, 2022. [MR4506579](#)
  - [38] Ivar Ekeland. An optimal matching problem. *ESAIM: Control, Optimisation and Calculus of Variations*, 11(1):57–71, 2005. [MR2110613](#)
  - [39] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trounev, and Gabriel Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. PMLR, 2019.
  - [40] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
  - [41] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015. [MR3383341](#)
  - [42] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its applications*, 114:717–735, 1989. [MR0986904](#)
  - [43] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996. [MR1440931](#)
  - [44] Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of Sinkhorn divergences. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*,

- pages 1574–1583, 2019.
- [45] Promit Ghosal, Marcel Nutz, and Espen Bernton. Stability of entropic optimal transport and Schrödinger bridges. *Journal of Functional Analysis*, 283(9):109622, 2022. [MR4460341](#)
  - [46] Promit Ghosal and Bodhisattva Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *The Annals of Statistics*, 50(2):1012–1037, 2022. [MR4404927](#)
  - [47] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*, volume 224. Springer, 2015. [MR0737190](#)
  - [48] Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Statistical inference with regularized optimal transport. *arXiv preprint [arXiv:2205.04283](#)*, 2022. [MR4701828](#)
  - [49] Alberto González-Sanz and Shayan Hundrieser. Weak limits for empirical entropic optimal transport: Beyond smooth costs. *arXiv preprint [arXiv:2305.09745](#)*, 2023.
  - [50] Alberto González-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. Weak limits of entropy regularized optimal transport; potentials, plans and divergences. *arXiv preprint: [arXiv:2207.07427](#)*, 2022.
  - [51] Florian Gunsilius and Yuliang Xu. Matching for causal effects via multi-marginal unbalanced optimal transport. *arXiv preprint [arXiv:2112.04398](#)*, 2021. Updated on July 9, 2022.
  - [52] Zaid Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv preprint [arXiv:2011.08963](#)*, 2020.
  - [53] Shayan Hundrieser, Marcel Klatt, and Axel Munk. Limit distributions and sensitivity analysis for entropic optimal transport on countable spaces. *arXiv preprint [arXiv:2105.00049](#)*, 2021. [MR4700262](#)
  - [54] Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, 2021. [MR4255123](#)
  - [55] Leonid V. Kantorovich. On the translocation of masses. In *Doklady Akademii Nauk USSR*, volume 37, pages 199–201, 1942. [MR0009619](#)
  - [56] Marcel Klatt, Axel Munk, and Yoav Zemel. Limit laws for empirical optimal solutions in random linear programs. *Annals of Operations Research*, 315(1):251–278, 2022. [MR4458612](#)
  - [57] Marcel Klatt, Carla Taming, and Axel Munk. Empirical regularized optimal transport: Statistical theory and applications. *SIAM Journal on Mathematics of Data Science*, 2(2):419–443, 2020. [MR4105566](#)
  - [58] Serge Lang. *Fundamentals of differential geometry*, volume 191. Springer Science & Business Media, 2012. [MR1666820](#)
  - [59] Christian Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012. [MR2873864](#)
  - [60] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete & Continuous Dynamical Systems*, 34(4):1533, 2014. [MR3121631](#)

- [61] Lang Liu, Soumik Pal, and Zaid Harchaoui. Entropy regularized optimal transport independence criterion. In *International Conference on Artificial Intelligence and Statistics*, pages 11247–11279. PMLR, 2022.
- [62] T. Manole and J. Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *arXiv preprint [arXiv:2106.13181](https://arxiv.org/abs/2106.13181)*, 2021. [MR4700254](https://doi.org/10.1.1.1.1)
- [63] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *arXiv preprint [arXiv:2107.12364](https://arxiv.org/abs/2107.12364)*, 2021.
- [64] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Central limit theorems for smooth optimal transport maps. *arXiv preprint [arXiv:2312.12407](https://arxiv.org/abs/2312.12407)*, 2023.
- [65] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32, 2019.
- [66] Toshio Mikami. Monge’s problem with a quadratic cost by the zero-noise limit of  $h$ -path processes. *Probability theory and related fields*, 129(2):245–260, 2004. [MR2063377](https://doi.org/10.1007/s00332-004-0000-0)
- [67] Toshio Mikami. Regularity of Schrödinger’s functional equation and mean field pdes for  $h$ -path processes. *Osaka Journal of Mathematics*, 56(4):831–842, 2019. [MR4020639](https://doi.org/10.1007/s12015-019-0000-0)
- [68] Toshio Mikami. Regularity of Schrödinger’s functional equation in the weak topology and moment measures. *Journal of the Mathematical Society of Japan*, 73(1):99–123, 2021. [MR4203323](https://doi.org/10.1007/s12015-021-0000-0)
- [69] Toshio Mikami and Michele Thieullen. Optimal transportation problem by stochastic optimal control. *SIAM Journal on Control and Optimization*, 47(3):1127–1139, 2008. [MR2407010](https://doi.org/10.1137/070407010)
- [70] Richard Nickl. On convergence and convolutions of random signed measures. *Journal of Theoretical Probability*, 22(1):38–56, 2009. [MR2472004](https://doi.org/10.1007/s12015-009-0000-0)
- [71] Richard Nickl and Benedikt M. Pötscher. Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *Journal of Theoretical Probability*, 20:177–199, 2007. [MR2324525](https://doi.org/10.1007/s12015-007-0000-0)
- [72] Jonathan Niles-Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, Nov. 2019. [MR4003560](https://doi.org/10.1017/S1446788719000000)
- [73] Jonathan Niles-Weed and Philippe Rigollet. Estimation of Wasserstein distances in the Spiked Transport Model. *Bernoulli*, 28(4):2663–2688, 2022. [MR4474558](https://doi.org/10.1017/S1446788722000000)
- [74] Marcel Nutz and Johannes Wiesel. Entropic optimal transport: Convergence of potentials. *Probability Theory and Related Fields*, 184(1-2):401–424, 2022. [MR4498514](https://doi.org/10.1007/s00332-022-0000-0)
- [75] Marcel Nutz and Johannes Wiesel. Stability of Schrödinger potentials and convergence of Sinkhorn’s algorithm. *The Annals of Probability*, 51(2):699–722, 2023. [MR4546630](https://doi.org/10.1214/22-AOP1500)
- [76] Soumik Pal. On the difference between entropic cost and the optimal trans-

- port cost. *The Annals of Applied Probability*, to appear, 2019. [MR4700251](#)
- [77] Victor M. Panaretos and Yoav Zemel. *An invitation to statistics in Wasserstein space*. Springer Nature, 2020. [MR4350694](#)
- [78] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [79] Dimitris N. Politis and Joseph P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, pages 2031–2050, 1994. [MR1329181](#)
- [80] Aram-Alexandre Pooladian, Marco Cuturi, and Jonathan Niles-Weed. De-biaser beware: Pitfalls of centering regularized transport maps. *arXiv preprint arXiv:2202.08919*, 2022.
- [81] Aram-Alexandre Pooladian, Vincent Divol, and Jonathan Niles-Weed. Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. *arXiv preprint arXiv:2301.11302*, 2023.
- [82] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.
- [83] Philippe Rigollet and Austin J. Stromme. On the sample complexity of entropic optimal transport. *arXiv e-prints*, pages arXiv–2206, 2022.
- [84] Gabriel Rioux, Ziv Goldfeld, and Kengo Kato. Entropic Gromov-Wasserstein distances: Stability and algorithms. *arXiv preprint arXiv:2306.00182*, 2023.
- [85] Werner Römisch. Delta method, infinite dimensional. In *Encyclopedia of Statistical Sciences*. Wiley, 2004.
- [86] Ritwik Sadhu, Ziv Goldfeld, and Kengo Kato. Limit theorems for semidiscrete optimal transport maps. *arXiv preprint arXiv:2303.10155*, 2023.
- [87] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, 55(58-63):94, 2015. [MR3409718](#)
- [88] Filippo Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: An overview. *Bulletin of Mathematical Sciences*, 7(1):87–154, 2017. [MR3625852](#)
- [89] Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66:477–487, 1990. [MR1080259](#)
- [90] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK, 1998. [MR1652247](#)
- [91] Aad W. van der Vaart and Jon Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media, 1996. [MR1385671](#)
- [92] J. H. van Zanten and A. W. van der Vaart. Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 200–222. Institute of Mathematical Statistics, 2008. [MR2459226](#)
- [93] Cédric Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics. American Mathematical Society, 2003. [MR1964483](#)
- [94] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2008. [MR2459454](#)

- [95] Eberhard Zeidler. *Applied functional analysis: main principles and their applications*, volume 109. Springer Science & Business Media, 2012. [MR1347692](#)
- [96] Zhengxin Zhang, Ziv Goldfeld, Youssef Mroueh, and Bharath K. Sriperumbudur. Gromov-Wasserstein distances: entropic regularization, duality, and sample complexity. *arXiv preprint [arXiv:2212.12848](#)*, 2022.