Stochastic Processes and their Applications 178 (2024) 104462

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

Empirical optimal transport under estimated costs: Distributional limits and statistical applications

Shayan Hundrieser [a], Gilles Mordant [a], Christoph A. Weitkamp [a], Axel Munk [a,b,c,*]

[a] Institute for Mathematical Stochastics, Georg-August-University Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany
[b] Max Planck Institute for Multidisciplinary Sciences, Am Faßberg 11, 37077 Göttingen, Germany
[c] Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells"(MBExC), University Medical Center, Robert-Koch-Straße 40, 37075 Göttingen, Germany

ARTICLE INFO

MSC:
primary 60B12
60F05
60G15
62E20
62F40
secondary 90C08
90C31

Keywords:
Optimal transport
Central limit theorem
Stability analysis
Curse of dimensionality
Empirical process
Bootstrap

ABSTRACT

Optimal transport (OT) based data analysis is often faced with the issue that the underlying cost function is (partially) unknown. This is addressed in this paper with the derivation of distributional limits for the empirical OT value when the cost function and the measures are estimated from data. For statistical inference purposes, but also from the viewpoint of a stability analysis, understanding the fluctuation of such quantities is paramount. Our results find direct application in the problem of goodness-of-fit testing for group families, in machine learning applications where invariant transport costs arise, in the problem of estimating the distance between mixtures of distributions, and for the analysis of empirical sliced OT quantities.

The established distributional limits assume either weak convergence of the cost process in uniform norm or that the cost is determined by an optimization problem of the OT value over a fixed parameter space. For the first setting we rely on careful lower and upper bounds for the OT value in terms of the measures and the cost in conjunction with a Skorokhod representation. The second setting is based on a functional delta method for the OT value process over the parameter space. The proof techniques might be of independent interest.

1. Introduction

Statistically sound methods for data analysis relying on the optimal transport (OT) theory (see *e.g.*, Rachev and Rüschendorf [1], Santambrogio [2], Villani [3]) have won acclaim in recent years. Exemplarily, we mention fitting of generative adversarial networks [4], novel notions of multivariate quantiles [5,6] and dependence [7–9] or tools for causal inference [10].

Recall that for Polish spaces $\mathcal{X}$ and $\mathcal{Y}$ and a continuous cost function $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$, the OT value between two (Borel) probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ is defined as

$$OT(\mu, \nu, c) := \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\,\mathrm{d}\pi(x, y), \tag{1}$$

where $\Pi(\mu, \nu)$ denotes the set of couplings of μ and ν. Under mild assumptions (1) also admits a dual formulation (see, *e.g.*, Santambrogio [2]),

$$OT(\mu, \nu, c) = \sup_{f\in C(\mathcal{X})} \int_{\mathcal{X}} f^{cc}(x)\,\mathrm{d}\mu(x) + \int_{\mathcal{Y}} f^{c}(y)\,\mathrm{d}\nu(y), \tag{2}$$

* Corresponding author at: Institute for Mathematical Stochastics, Georg-August-University Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany.
E-mail addresses: s.hundrieser@math.uni-goettingen.de (S. Hundrieser), gilles.mordant@uni-goettingen.de (G. Mordant), cweitka@mathematik.uni-goettingen.de (C.A. Weitkamp), munk@math.uni-goettingen.de (A. Munk).

https://doi.org/10.1016/j.spa.2024.104462
Received 2 February 2023; Received in revised form 24 June 2024; Accepted 14 August 2024
Available online 22 August 2024
0304-4149/

where $C(\mathcal{X})$ stands for the set of real-valued, continuous functions on $\mathcal{X}$. Further, denote by $f^c(y) := \inf_{x\in\mathcal{X}} c(x,y) - f(x)$ and $f^{cc}(x) := \inf_{y\in\mathcal{Y}} c(x,y) - f^c(y)$ the cost-transformations of f and f^c under c, respectively; also often referred to as c-transformations.

If $\mathcal{X} = \mathcal{Y}$ and the cost function $c = d_{\mathcal{X}}^p$ is the pth power ($p \geq 1$) of a metric $d_{\mathcal{X}}$ on $\mathcal{X}$ the OT value gives rise to the p-Wasserstein distance

$$W_p(\mu,\nu) := (OT(\mu,\nu,d_{\mathcal{X}}^p))^{1/p},$$

which defines a metric on the space of probability measures with p-th moments [3, Chapter 6]. This metric is particularly useful for many data analysis tasks due to its potential awareness of the "inner geometry" of $\mathcal{X}$. For instance, interpreting (normalized) images, or more precisely the corresponding pixel locations and intensities, as probability measures, it has been argued that the distance induced by OT corresponds to the natural expectations of what appears close or far away for the human eye [11]. Meanwhile, there is a plenitude of real world showcases where OT based distances (and their associated transport plans) prove useful for applications *e.g.*, in cell biology [12,13], genetics [14,15], protein structure analysis [16,17] or fingerprint analysis [18], to mention but a few. In these works, the cost function is selected by the practitioner and tailored to the concrete application, *e.g.*, a tree distance on the space of phylogenetic trees as in [14] or the Euclidean distance as in [13].

However, there are various instances where the underlying cost naturally depends on the measures. Examples include, *e.g.*, Wasserstein based goodness-of-fit testing under group families [19] or Wasserstein Procrustes analysis [20], where it is central that the underlying OT problem is invariant with respect to certain transformations. Moreover, for sliced OT [21] or projection robust OT [22,23], the Wasserstein distance between multiple low-dimensional projections of measures is computed. Taking the maximum of all one-dimensional (resp. low-dimensional) projections gives rise to the max-sliced Wasserstein distance [24] (resp. projection robust OT cost [22]) which induces a measure-dependent cost since maximizing directions are determined by the underlying measures. Since the underlying measures are generally unknown and have to be estimated from data, the cost function has to be estimated as well. The interest of learning the cost function from observations has also gained interest in economics to model migration flows [25,26] and in the machine learning community for geodesic flows of probability measures [27]. Finally, let us note for completeness that the problem of identifiability of the cost function from OT values, transport plans, or potentials has only recently been analyzed [28].

Motivated by these considerations, we provide in this work a general framework for the statistical analysis for empirical OT under costs that are dependent on the underlying measures. Adopting a statistical point of view, we assume that we do not have access to the measures μ and ν but only to independent samples $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ and $\{Y_i\}_{i=1}^m \sim \nu^{\otimes m}$ with $n,m \in \mathbb{N}$. Upon defining the empirical measures $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\nu_m := \frac{1}{m}\sum_{i=1}^m \delta_{Y_i}$ and given a random cost function[1] $c_{n,m}$ such that $OT(\mu_n,\nu_m,c_{n,m})$ estimates the quantity $OT(\mu,\nu,c)$, our main focus is on characterizing for $n,m\to\infty$ with $m/(n+m)\to\lambda\in(0,1)$ the limit distribution of

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)\Big). \tag{3}$$

This is of particular interest for asymptotic tests about the relation between μ and ν for unknown c based on the OT value. Further, this enables the derivation of confidence intervals for the OT value $OT(\mu,\nu,c)$. As it is practically more relevant, we mainly focus on the scenario where both measures μ and ν are unknown. However, we stress that our theory also provides distributional limits for the one-sample case, *i.e.*, when only μ is estimated from data while ν is assumed to be known and vice versa (see Remarks 4 and 9). Moreover, although we mostly focus on empirical measures to estimate the underlying measures, our theory also enables the derivation of distributional limits for alternative measure estimators (see, *e.g.*, [29]), provided that the corresponding distributional limits for the measures can be determined.

For a fixed cost function, *i.e.*, for $c_{n,m}\equiv c$ for some $c\in C(\mathcal{X}\times\mathcal{Y})$, already various works derived limit distribution results for the empirical OT quantity in (3). A specific situation arises for probability measures on $\mathbb{R}$ with $c_p(x,y) = |x-y|^p$ for $p\geq 1$ [30–34] where the OT plan can be represented via a quantile coupling. For this setting, quantile process theory [35] in combination with integrability conditions on the underlying densities have been exploited to derive distributional limits.

Moreover, on general Euclidean spaces $\mathbb{R}^d$ with $d\geq 1$ and pth power costs $c_p(x,y) = \|x-y\|^p$ with $p>1$ it has been shown by del Barrio et al.[36,37] for probability measures μ,ν with connected support and finite $2p$-th moments for $n,m\to\infty$ with $m/(n+m)\to\lambda\in(0,1)$ that

$$\sqrt{\frac{nm}{n+m}}\left(OT(\mu_n,\nu_m,c_p) - \mathbb{E}\left[OT(\mu_n,\nu_m,c_p)\right]\right) \rightsquigarrow \mathcal{N}(0,\sigma^2_{\mu,\nu}), \tag{4}$$

where $\sigma^2_{\mu,\nu} > 0$ if and only if $\mu\neq\nu$. Here and throughout, "$\rightsquigarrow$" denotes weak convergence in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner [38, Chapter 1.3]). Their proof is based on an L^2-linearization technique of the OT value and relies on the Efron-Stein inequality. In general, the centering quantity $\mathbb{E}[OT(\mu_n,\nu_m,c_p)]$ in (4) cannot be replaced by its population quantity $OT(\mu,\nu,c_p)$ which hinders further statistical inference purposes. Indeed, for identical absolutely continuous probability measures $\mu=\nu$ on $\mathbb{R}^d$ with sufficiently many moments it follows for $d>2p$ by Fournier and Guillin [39],Weed and Bach [40] that

$$\mathbb{E}\left[OT(\mu_n,\nu_m,c_p)\right] \asymp \min(n,m)^{-p/d}.$$

[1] Here, $c_{n,m}$ is either a direct estimator for c or chosen via an OT-related optimization problem over a parameter class.

Moreover, for different measures $\mu \neq \nu$ on $\mathbb{R}^d$ which are absolutely continuous and sub-Weibull it has been shown for $d \geq 5$ by Manole and Niles-Weed [41] that

$$\mathbb{E}\left[OT(\mu_n, \nu_m, c_p)\right] - OT(\mu, \nu, c_p) \asymp \min(n, m)^{-\min(p,2)/d}.$$

These rates are also minimax optimal (up to logarithmic factors) over appropriate collections of identical measures $\mu = \nu$ [42] as well as different measures $\mu \neq \nu$ [41]. In particular, this demonstrates that estimation of the OT value suffers from the curse of dimensionality and showcases that it is in general for $d \geq 5$, due to the dominance of the bias, not possible to replace $\mathbb{E}[OT(\mu_n, \nu_m, c_p)]$ with $OT(\mu, \nu, c_p)$ in (4).

Nevertheless, according to the recently discovered *lower complexity adaptation principle* for empirical OT [43], fast convergence rates are still achieved if one of the population measures, μ or ν, is supported on a sufficiently low dimensional domain. Based on this observation, Hundrieser et al. [44] proved for compactly supported μ, ν on $\mathbb{R}^d$, with μ supported on a finite set or a smooth submanifold of dimension $\tilde{d} < 2\min(p, 2)$ using the functional delta method [45],

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_p) - OT(\mu, \nu, c_p)\Big) \rightsquigarrow \sup_{f \in S_{c_p}(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{c_p c_p}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^{c_p}), \tag{5}$$

where $S_{c_p}(\mu, \nu)$ is the set of optimizers of (2) and $\mathbb{G}^\mu, \mathbb{G}^\nu$ denote μ-, ν-Brownian bridges, *i.e.*, centered Gaussian processes with covariance structure characterized by

$$\mathrm{Cov}[\mathbb{G}^\mu(f), \mathbb{G}^\mu(g)] = \int f g \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu \int g \,\mathrm{d}\mu \quad \text{for } f, g \in C(\mathcal{X}) \tag{6}$$

and likewise for $\mathbb{G}^\nu$. The asymptotic theory laid out in (5) also provides a unified framework for distributional limits of the empirical OT value under discrete population measures [18,46] and the semi-discrete setting [47].

The central contribution of this work is to extend such distributional limits from (5) to settings where the cost function is not fixed and additionally may depend on the underlying measures. We focus on the following two special instances.

(A) The cost estimator $c_{n,m}$, centered by its population counterpart c and suitably rescaled, weakly converges in $C(\mathcal{X} \times \mathcal{Y})$ to a tight limit, *i.e.*,

$$\sqrt{\frac{nm}{n+m}}\left(c_{n,m} - c\right) \rightsquigarrow \mathbb{G}^c \quad \text{in } C(\mathcal{X} \times \mathcal{Y}).$$

(B) There exists a collection $\{c_\theta\}_{\theta\in\Theta}$ of costs such that for any $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ the corresponding cost function $c_{\mu,\nu} := c_\theta$ is selected according to an optimization problem of the OT value over Θ, *i.e.*, either

$$\theta \in \arg\max_{\theta\in\Theta} OT(\mu, \nu, c_\theta) \quad \text{or} \quad \theta \in \arg\min_{\theta\in\Theta} OT(\mu, \nu, c_\theta).$$

These two settings are natural and treat a wide spectrum of problems. Furthermore, they are strongly related. It is noteworthy that setting **(B)** could be treated in the framework of **(A)** by estimating the optimal θ. However, this approach requires the existence of a unique population cost function and weak convergence of the cost process as a random element in $C(\mathcal{X}\times\mathcal{Y})$. Since we are only interested in the empirical infimal or supremal OT value it is instead more natural to rely on an alternative approach which does not require uniqueness of the population cost function or weak convergence of the cost process.

For setting **(A)** we allow the cost function to be estimated from the given data and thus capture the asymptotic dependency between the cost estimator and the empirical measures. In particular, this enables an analysis of the empirical OT cost when the cost estimator is parametrized by a plug-in estimator, *e.g.*, a maximum likelihood procedure. Notably, setting **(A)** also allows the cost function to be estimated from independent data. Overall, this setting covers many scenarios with "extrinsically estimated costs". We refer to Sections 4.1 and 4.3 for examples. For setting **(B)** the motivation slightly differs. Here, the selected cost function depends on the OT problem itself and often brings invariance of the OT problem with respect to a class of transformation parametrized by Θ. One could describe this as OT with "intrinsically estimated costs". Examples of this setting are provided in Sections 4.2 and 4.4.

Under suitable assumptions we show in Theorem 2 for setting **(A)** that

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_{n,m}) - OT(\mu, \nu, c)\Big) \rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c),$$

where $\Pi_c^\star(\mu, \nu)$ represents the set of optimizers for (1) for μ, ν with costs c and $\pi(\mathbb{G}^c) := \int \mathbb{G}^c \mathrm{d}\pi$. For setting **(B)** we only state below the distributional limit for supremal costs; a similar distributional limit also occurs for infimal costs (Theorem 6). Upon defining the set $S_+(\Theta, \mu, \nu) = \mathrm{argmax}_{\theta\in\Theta} OT(\mu, \nu, c_\theta)$ of maximizers we show in Theorem 7 that

$$\sqrt{\frac{nm}{n+m}}\left(\sup_{\theta\in\Theta} OT(\mu_n, \nu_m, c_\theta) - \sup_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right) \rightsquigarrow \sup_{\theta\in S_+(\Theta,\mu,\nu)} \sup_{f_\theta\in S_{c_\theta}(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}).$$

In addition to these distributional limits we show for both settings **(A)** and **(B)** consistency of a bootstrap principle. This is of practical importance since quantiles of the respective distributional limits are difficult to express explicitly due to their dependency on the collection of primal and dual optimizers for population measures and cost.

Our proof technique for the distributional limit under setting **(A)** differs from previous approaches and might be of interest in its own right. More precisely, due to the estimation of the cost function, we cannot rely on any of the techniques from the references mentioned above. Instead, we derive certain lower and upper bounds on the OT value which fulfill appropriate (semi-)continuity properties. In conjunction with a Skorokhod representation for the empirical process jointly with the cost process, this enables us to prove that the law of the empirical OT value with estimated costs is asymptotically stochastically dominated from above and below by the asserted limit distribution.

For the analysis of setting **(B)** we show under suitable assumptions on the cost family $\{c_\theta\}_{\theta\in\Theta}$ and the underlying probability measures, that the empirical OT process $\sqrt{n}(OT(\mu_n, \nu_m, c_\theta) - OT(\mu, \nu, c_\theta))_{\theta\in\Theta}$ weakly converges in $C(\Theta)$ to a tight random variable. We prove this result by invoking the functional delta method in conjunction with a general result on Hadamard directional differentiability for extremal-type functionals uniformly over a compact parameter space (see Appendix A). The latter can be viewed as an extension of Fang and Santos [48, Lemma S.4.9] to processes over Θ and relies on Dini's theorem [49, Corollary 1]. Central for this differentiability result is a certain continuity condition among the sets of maximizing elements for varying parameter. For the OT process it is fulfilled, *e.g.*, if for every $\theta \in \Theta$ the set of dual optimizer $S_{c_\theta}(\mu, \nu)$ is unique (up to constant shift). A similar assumption has been imposed by Xi and Niles-Weed [50] for weak convergence of the empirical sliced OT process, which can be viewed as a special instance of our results for general OT processes, see Section 4.4. The distributional limits for the empirical infimal and supremal OT value over $\theta \in \Theta$ then follow by another application of the functional delta method.

Outline. We begin our exposition by deriving in Section 2.1 an appropriate dual formulation of the OT value which proves useful for our subsequent considerations. We then proceed with our main contributions, distributional limits for the empirical OT value under weakly converging costs in Section 2.2 as well as for the empirical OT value under extremal-type costs in Section 2.3. These asymptotic results are complemented with consistency results of bootstrap resampling schemes in Section 2.4. We discuss our assumptions for the distributional limits and the bootstrap principles in Section 3 and provide sufficient conditions for their validity. Statistical applications of our theory are provided in Section 4, where we also derive a deterministic (first-order) stability result for the OT cost under joint perturbations of measures and cost function. In Section 5 we explicitly construct functionals which enables us to "elevate" the regularity of cost estimators to that of their population counterparts. We employ them in the proofs of our main results which are stated in Section 6. All remaining proofs as well as auxiliary results and lemmata are relegated to the Appendices. For ease of reading the technical discussion of Section 3 can be skipped for a first reading and one may forward from Section 2 immediately to Section 4 on applications which provides an entry point to those readers who would like to apply our results to a particular setting of their own.

Notation and technical prerequisites. Given a set T denote by $\ell^\infty(T)$ the Banach space of bounded functionals on T equipped with uniform norm $\|\varphi\| := \sup_{t\in T} |\varphi(t)|$. Moreover, if T is equipped with a topology τ denote by $C(T)$ the Banach space of real valued, bounded, continuous functions on T equipped with uniform norm. If d_T denotes a metric on T, then we define by $C_u(T, d_T)$, or $C_u(T)$ when the metric d_T is clear from context, the space of real-valued, bounded, uniformly continuous functions on (T, d). Endowed with the uniform norm, it is a Banach space as well. A real-valued function class $\mathcal{F}$ on $\mathcal{X}$ is always be equipped with uniform norm. This specifies the Banach space $C_u(\mathcal{F})$ which is a closed subset of the Banach space $\ell^\infty(\mathcal{F})$. Moreover, for $\varepsilon > 0$ the covering number $\mathcal{N}(\varepsilon, T, d)$ denotes the minimal number of sets with diameter 2ε to cover T, and we write $x \lesssim y$ when there exists a constant $C > 0$ with $x \le Cy$. For a topological space $\mathcal{X}$ the set $\mathcal{P}(\mathcal{X})$ denotes the collection of Borel probability measures on $\mathcal{X}$. Integration $\int f \,\mathrm{d}\mu$ of a real-valued Borel measurable function $f : \mathcal{X} \to \mathbb{R}$ with respect to $\mu \in \mathcal{P}(\mathcal{X})$ is abbreviated by $\mu(f)$ or μf. Further, we denote by $f_\#\mu$ the pushforward of μ under f. We define all random variables on the same probability space $(\Omega, \mathcal{A}, \mathrm{P})$. We further assume a product structure of that space to define samples and the random weights of the bootstrap, *i.e.*, $\Omega = \Omega_0 \times \Omega_1 \times \cdots$ and $\mathrm{P} = \mathrm{P}_0 \otimes \mathrm{P}_1 \times \cdots$ so that the samples only depend on (Ω_0, P_0), the weights of the first bootstrap replicate on (Ω_1, P_1) and so on. The law of a random variable X is denoted by $\mathcal{L}(X)$. We finally assume that there exist infinite sequences of measurable maps $X_1, X_2, \dots$ from (Ω_0, P_0) to $\mathcal{X}$, respectively, and that samples of cardinality n are obtained from the infinite sequence by projection of the first n coordinates. Outer probability measures are denoted by P^* (see [38, Chapter 1.2]). Denoting by BL_1 the set of real-valued functions on a metric space (T, d_T) which are bounded by one in uniform norm and such that $|f(x) - f(y)| \le d_T(x, y)$ for any $x, y \in T$, we define the bounded Lipschitz metric between two probability measures μ, ν as $d_{BL}(\mu, \nu) := \sup_{f\in \mathrm{BL}_1} |\mu(f) - \nu(f)|$. For a set A and a function f, we write $f(\mathbf{A}) := \{f(a) \mid a \in A\}$. For two subsets A, B of a vector space, $A + B := \{a + b \mid a \in A, b \in B\}$. The set of symmetric (resp. symmetric positive definite) matrices in $\mathbb{R}^{d\times d}$ is denoted by $\mathrm{S}(d)$ (resp. $\mathrm{SPD}(d)$).

2. Main results

2.1. Preliminaries

For our theory on distributional limits for the empirical OT value under estimated cost functions we consider throughout compact Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. Given a continuous cost function $c \in C(\mathcal{X} \times \mathcal{Y})$ and probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ there always exist optimizers to both primal and dual problem [3, Theorems 4.1 and 5.10].

According to Villani [51, Remark 1.13], dual optimizers can always be selected from the function class

$$\mathcal{H}_c := \left\{ h : \mathcal{X} \to \mathbb{R} \;\middle|\; \exists g : \mathcal{Y} \to [-\|c\|_\infty, \|c\|_\infty], h(\cdot) = \inf_{y\in\mathcal{Y}} c(\cdot, y) - g(y) \right\}, \tag{7}$$

which yields for any $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ the alternative dual representation of the OT value,

$$OT(\mu, \nu, c) = \sup_{h\in\mathcal{H}_c} \mu(h^{cc}) + \nu(h^c). \tag{8}$$

The function class $\mathcal{H}_c$ is uniformly bounded, and each element exhibits the same modulus of continuity as c, hence it is compact in $C(\mathcal{X})$ by the Theorem of Arzelà-Ascoli. Formula (8) was exploited by Hundrieser et al. [44] for distributional limits of the empirical OT value under a fixed cost function.

For our purposes, we require a dual formulation over a fixed function class which holds for more than a single cost function and to circumvent potential measurability issues we seek a function class which is compact in $C(\mathcal{X})$ (cf. Lemma 45). To this end, let $B > 0$ and consider a concave modulus of continuity $w : \mathbb{R}_+ \to \mathbb{R}_+$. Then, for a continuous metric $d_{\mathcal{X}}$ on $\mathcal{X}$ we define the compact function class $\mathcal{F}(B, w) \subseteq C(\mathcal{X})$,

$$\mathcal{F}(B,w) := \left\{ f : \mathcal{X} \to \mathbb{R} \;\middle|\; \|f\|_\infty \le 2B,\ |f(x) - f(x')| \le w(d_{\mathcal{X}}(x,x')) \text{ for all } x, x' \in \mathcal{X} \right\}, \tag{9}$$

which will be utilized for a dual representation of the OT value under suitable costs.

Lemma 1 (*Dual formulation*). *Let $c \in C(\mathcal{X} \times \mathcal{Y})$ with $\|c\|_\infty \le B$ and $|c(x,y) - c(x',y)| \le w\left(d_{\mathcal{X}}(x,x')\right)$ for all $x, x' \in \mathcal{X}, y \in \mathcal{Y}$. Then, for $\mathcal{F} := \mathcal{F}(B,w)$ the following inclusions hold*

$$\mathcal{H}_c \subseteq \mathcal{F}^{cc} \subseteq \mathcal{H}_c + [-2B, 2B] \quad \text{and} \quad \mathcal{H}_c^c \subseteq \mathcal{F}^c \subseteq \mathcal{H}_c^c + [-2B, 2B].$$

Further, for arbitrary probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ it follows that

$$OT(\mu, \nu, c) = \sup_{f \in \mathcal{F}} \mu(f^{cc}) + \nu(f^c) \tag{10}$$

and the set of dual optimizers $S_c(\mu,\nu)$ of (10), referred to as Kantorovich potentials, is non-empty.

The proof of Lemma 1 is deferred to Section 6.1. Overall, Lemma 1 justifies the use of the function class $\mathcal{F} = \mathcal{F}(B,w)$ for a dual OT formulation and enables us to state conditions of distributional limits in terms of $\mathcal{F}$ instead of potentially varying collections of functions.

2.2. Distributional limits under weakly converging costs

For the distributional limits in all the statements below, we consider independent and identically distributed random variables $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ and independent $\{Y_i\}_{i=1}^m \sim \nu^{\otimes m}$ defined on the probability space put forward in the introduction. Based on these samples, we define empirical measures $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\nu_m := \frac{1}{m}\sum_{i=1}^m \delta_{Y_i}$. All the subsequent asymptotic results are to be understood for $n, m \to \infty$ with $m/(n+m) \to \lambda \in (0,1)$, which we do not recall each time for space considerations.

Our main result on the limit law for the empirical OT value under weakly converging costs is given as follows for the two-sample case. The one-sample case is discussed in Remark 4(ii).

Theorem 2 (*OT under weakly converging costs*). *Let $c \in C(\mathcal{X} \times \mathcal{Y})$ and consider an estimator $c_{n,m} \in C(\mathcal{X} \times \mathcal{Y})$ for c such that $c_{n,m}(x,y)$ is measurable for each $(x,y) \in \mathcal{X} \times \mathcal{Y}$. Let $w : \mathbb{R}_+ \to \mathbb{R}_+$ be a concave modulus of continuity for c with $w(\delta) > 0$ for $\delta > 0$ such that $|c(x,y) - c(x',y)| \le w(d_{\mathcal{X}}(x,x'))$ for all $x, x' \in \mathcal{X}, y \in \mathcal{Y}$. Assume for $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ the following.*

(JW) *For the function class $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ from (9) joint weak convergence occurs,*

$$\sqrt{\frac{nm}{n+m}} \begin{pmatrix} \mu_n - \mu \\ \nu_m - \nu \\ c_{n,m} - c \end{pmatrix} \rightsquigarrow \begin{pmatrix} \sqrt{\lambda}\, \mathbb{G}^\mu \\ \sqrt{1-\lambda}\, \mathbb{G}^\nu \\ \mathbb{G}^c \end{pmatrix} \quad \text{in } \ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X} \times \mathcal{Y}),$$

where $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ is a tight random variable and $\mathbb{G}^\mu, \mathbb{G}^\nu$ have covariance as in (6).

Further, suppose either one of the following two assumptions.

(OP) *There exists a unique OT plan $\pi \in \Pi_c^\star(\mu,\nu)$ between μ and ν for the cost function c.*

(Sup) *The empirical processes $\mathbb{G}_n^\mu := \sqrt{n}(\mu_n - \mu)$ and $\mathbb{G}_m^\nu := \sqrt{m}(\nu_m - \nu)$ fulfill the convergence $\sup_{f \in \mathcal{F}} \mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}} - f^{cc}) \xrightarrow{\mathrm{P}^*} 0$ and $\sup_{f \in \mathcal{F}} \mathbb{G}_m^\nu(f^{c_{n,m}} - f^c) \xrightarrow{\mathrm{P}^*} 0$.*

Then, it follows that

$$\sqrt{\frac{nm}{n+m}} \Big(OT(\mu_n, \nu_m, c_{n,m}) - OT(\mu, \nu, c) \Big) \rightsquigarrow \inf_{\pi \in \Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f \in S_c(\mu,\nu)} \sqrt{\lambda}\, \mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\, \mathbb{G}^\nu(f^c).$$

A key insight of Theorem 2 is that the limit distribution for the estimated OT value can be decomposed into two terms: the fluctuation of the cost estimators evaluated at the collection of OT plans and the Kantorovich potentials evaluated at the limit of the empirical process. Under uniqueness of primal and dual optimizers for the population OT problem we obtain the following.

Corollary 3 (*OT under weakly converging costs and uniqueness*). *In the setting of Theorem 2 assume (JW)and (OP), and suppose that the set of Kantorovich potentials $S_c(\mu,\nu)$ for μ, ν with cost function c is unique (up to a constant shift).*[2] *Then, for $\pi \in \Pi_c^\star(\mu,\nu)$ and*

[2] By this we mean, for any $f, g \in S_c(\mu,\nu)$ the difference $f - g$ is constant on $\mathrm{supp}(\mu)$.

$f \in S_c(\mu, \nu)$, it follows that

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_{n,m}) - OT(\mu, \nu, c)\Big) \rightsquigarrow \pi(\mathbb{G}^c) + \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c). \tag{11}$$

In particular, if $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ is a jointly centered Gaussian process in $\ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X} \times \mathcal{Y})$, the weak limit in (11) is centered normal.

The proof of Theorem 2 is deferred to Section 6.2.1 and relies on careful lower and upper bounds for the empirical OT value due to the primal (1) and dual formulation (10), as well as arguments from empirical process theory. In the course of this, a key argument is the application of Lemma 1 for $c_{n,m}$ and c. Notably, we do not demand that the cost estimator $c_{n,m}$ is suitably bounded or exhibits a similar modulus of continuity as c itself. Instead, we construct by Corollary 39 an alternative cost estimator $\overline{c}_{n,m}$ such that the conditions, $\|\overline{c}_{n,m}\|_\infty \le 2\|c\|_\infty + 1$ as well as $|\overline{c}_{n,m}(x, y) - \overline{c}_{n,m}(x', y)| \le 2w(d_{\mathcal{X}}(x, x'))$ for all $x, x' \in \mathcal{X}, y \in \mathcal{Y}$, are fulfilled deterministically and $\sqrt{nm/(n+m)}\|\overline{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0$. The latter implies by Lemma 43 that

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, \overline{c}_{n,m}) - OT(\mu_n, \nu_m, c_{n,m})\Big) \le \sqrt{\frac{nm}{n+m}}\|\overline{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0.$$

It thus suffices to show the assertion for $\overline{c}_{n,m}$ where the dual formulation from Lemma 1 involving the function class $\mathcal{F}(2\|c\|_\infty + 1, 2w)$ is available. We call $\overline{c}_{n,m}$ a *regularity elevation* of $c_{n,m}$; details on different kinds of regularity elevations are given in Section 5. The notion of regularity elevations also proves to be useful for showing the validity of (Sup) as outlined in Section 3.3.

Remark 4. We like to comment on a few aspects of the derived distributional limits.

(i) The assumptions of Theorem 2 and sufficient conditions for their validity are discussed in Sections 3.1–3.3. Effectively, (JW) delimits the theory to settings of low dimensionality. In such settings (Sup) is often also valid as long as the population cost is sufficiently regular.

(ii) Our proof technique for Theorem 2 and Corollary 3 also asserts distributional limits for the one-sample setting, *i.e.*, when μ is estimated by μ_n and ν is assumed to be known. For this setting, (JW) reduces to the condition

$$\sqrt{n}\begin{pmatrix}\mu_n - \mu\\ c_n - c\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{G}^\mu\\ \mathbb{G}^c\end{pmatrix} \quad \text{in } \ell^\infty(\mathcal{F}^{cc}) \times C(\mathcal{X} \times \mathcal{Y}).$$

Moreover, in (Sup) we only require that $\sup_{f\in\mathcal{F}} \mathbb{G}_n^\mu(f^{c_n c_n} - f^{cc}) \xrightarrow{\mathrm{P}^*} 0$. Then,

$$\sqrt{n}\Big(OT(\mu_n, \nu, c_n) - OT(\mu, \nu, c)\Big) \rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^\mu(f^{cc}).$$

In case of a *fixed* cost function, *i.e.*, when selecting $c_n = c$, List **(Sup)** is trivially met and the conditions of Theorem 2 reduce to $\mathcal{F}^{cc}$ being μ-Donsker and $\mathcal{F}^c$ being ν-Donsker. Further, by Lemma 1 this is equivalent to $\mathcal{H}_c$ and $\mathcal{H}_c^c$ being Donsker for μ and ν [38, Theorem 2.10.1 and Example 2.10.7], respectively, matching the Donsker conditions of Theorem 2.2 under bounded, continuous costs on compact domains in Hundrieser et al. [44] (in this setting Assumptions (C), (E), and (P) are all met) which imply that

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c) - OT(\mu, \nu, c)\Big) \rightsquigarrow \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c).$$

(iii) Our proof technique also yields distributional limits for the estimated OT value when instead of empirical measures μ_n and ν_m one considers measurable estimators $\tilde{\mu}_n \in \mathcal{P}(\mathcal{X})$, $\tilde{\nu}_m \in \mathcal{P}(\mathcal{Y})$, respectively, that fulfill $\tilde{\mu}_n \rightsquigarrow \mu$ and $\tilde{\nu}_m \rightsquigarrow \nu$ in probability. This would mean to replace the empirical measures μ_n and ν_m in Assumptions (JW) and (Sup) by $\tilde{\mu}_n$ and $\tilde{\nu}_m$, respectively. In addition, instead of the scaling rate $\sqrt{nm/(n+m)}$ our proof technique theory also permits a different scaling rate $a_{n,m}$ which diverges to infinity for $n, m \to \infty$.

(iv) In Proposition 34 we prove that the OT value is Gateaux differentiable in all three entries (μ, ν, c) for admissible directions $(\Delta^\mu, \Delta^\nu, \Delta^c) \in (\mathcal{P}(\mathcal{X}) - \mu) \times (\mathcal{P}(\mathcal{Y}) - \nu) \times C(\mathcal{X} \times \mathcal{Y})$ with derivative,

$$(\Delta^\mu, \Delta^\nu, \Delta^c) \mapsto \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\Delta^c) + \sup_{f\in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c).$$

Hence, the asymptotic distribution described in Theorem 2 may also be interpreted as a derivative of the OT value with respect to the triple (μ, ν, c) evaluated at the limit process. Proving Theorem 2 via an application of the functional delta method would amount to showing Hadamard directional differentiability of the OT value [45]. However, this turns out be a challenging issue without imposing additional assumptions on the measure and cost estimators, see Remark 35.

(v) In case of a centered normal limit in (11) the limit variance is given by

$$\mathrm{Var}\big(\pi(\mathbb{G}^c)\big) + \lambda\,\mathrm{Var}_{X\sim\mu}\big(f^{cc}(X)\big) + (1-\lambda)\,\mathrm{Var}_{Y\sim\mu}\big(f^c(Y)\big) + 2\sqrt{\lambda}\,\mathrm{Cov}\big(\pi(\mathbb{G}^c), \mathbb{G}^\mu(f^{cc})\big) + 2\sqrt{1-\lambda}\,\mathrm{Cov}\big(\pi(\mathbb{G}^c), \mathbb{G}^\nu(f^c)\big),$$

where we used that the random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are independent. In particular, the limit law degenerates if both Kantorovich potentials (f^{cc}, f^c) are (μ, ν)-almost surely constant and $c_{n,m}$ converges to c with a faster rate than $(nm/(n+m))^{-1/2}$, uniformly on the support of the OT plan π. For a sharp characterization of the occurrence of almost surely constant Kantorovich potentials we refer to Section 4 of Hundrieser et al. [44] where the authors showcase that for most cost functions of practical interest a.s. constancy typically does not occur if the underlying measures are different.

2.3. Distributional limits under extremal-type costs

As noted in the introduction, could the empirical infimal or supremal OT value over a fixed collection of cost functions also be analyzed using the previously described framework. However, as part of this approach, we would require the existence of a single underlying population cost function as well as weak convergence of the cost estimator. To broaden the scope of our theory, we follow in this subsection a different route to derive limiting distributions where such conditions are not required. More precisely, we first prove a uniform distributional limit for the empirical OT process indexed over the collection of cost functions before relying on a delta method to characterize the distributional limits for the respective infimal and supremal statistics.

For the subsequent assertions we again adhere to the sampling convention provided at the beginning of Section 2.2. The one-sample case is discussed in Remark 9(iii).

Theorem 5 (*OT process uniformly over compact Θ*). *Let Θ be a compact Polish space and consider a continuous map $c : \Theta \to C(\mathcal{X} \times \mathcal{Y})$, $\theta \mapsto c_\theta$. Let $w : \mathbb{R}_+ \to \mathbb{R}_+$ be a modulus of continuity such that $\sup_{\theta\in\Theta} |c_\theta(x, y) - c_\theta(x', y)| \le w(d_{\mathcal{X}}(x, x'))$ for all $x, x' \in \mathcal{X}, y \in \mathcal{Y}$. Assume for $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ the following.*

(Don) *For the function class $\mathcal{F} = \mathcal{F}(\sup_{\theta\in\Theta} \|c_\theta\|_\infty , w)$ from (9) the collection $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta c_\theta}$ is μ-Donsker and $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta}$ is ν-Donsker.*

(KP) *For any $\theta \in \Theta$, the set of Kantorovich potentials $S_{c_\theta}(\mu, \nu) \subseteq \mathcal{F}$ for the OT problem between μ and ν and cost c_θ is unique (up to a constant shift).*

Then, upon selecting $f_\theta \in S_{c_\theta}(\mu, \nu)$ for any $\theta \in \Theta$, it follows that

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_\theta) - OT(\mu, \nu, c_\theta)\Big)_{\theta\in\Theta} \rightsquigarrow \Big(\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta})\Big)_{\theta\in\Theta} \quad \text{in } C(\Theta).$$

The proof of Theorem 5 is based on Hadamard directional differentiability of the OT cost process, which follows from a general sensitivity analysis for extremal-type functions uniformly over a compact parameter space Appendix A. The assertion for the empirical OT process then follows by invoking the functional delta method [45]; the proof is deferred to Section 6.3.1.

From the above result, given any functional $\Phi : C(\Theta) \to \mathbb{R}$ that is Hadamard directionally differentiable at the function $OT(\mu, \nu, c(\cdot)) \in C(\Theta)$, Theorem 5 yields by another application of the functional delta method, the distributional limit

$$\sqrt{\frac{nm}{n+m}}\Big(\Phi\big((OT(\mu_n, \nu_m, c_\theta))_{\theta\in\Theta}\big) - \Phi\big((OT(\mu, \nu, c_\theta))_{\theta\in\Theta}\big)\Big) \rightsquigarrow D^H_{OT(\mu,\nu,c(\cdot))}\Phi\Big(\big(\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta})\big)_{\theta\in\Theta}\Big).$$

Here, $D^H_{OT(\mu,\nu,c(\cdot))}\Phi$ denotes the directional Hadamard derivative of Φ. This enables the derivation of the limit distribution for the infimal mapping using Fang and Santos [48, Lemma S.4.9] (see also Cárcamo et al. [52, Corollary 2.3]).

Theorem 6 (*OT infimum over compact Θ*). *Consider the setting of Theorem 5. Then, upon selecting $f_\theta \in S_{c_\theta}(\mu, \nu)$ for any $\theta \in \Theta$, it follows that*

$$\sqrt{\frac{nm}{n+m}}\left(\inf_{\theta\in\Theta} OT(\mu_n, \nu_m, c_\theta) - \inf_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right) \rightsquigarrow \inf_{\theta\in S_-(\Theta,\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}),$$

where $S_-(\Theta, \mu, \nu) = \operatorname{argmin}_{\theta\in\Theta} OT(\mu, \nu, c_\theta)$ denotes the set of minimizers of $OT(\mu, \nu, c_\theta)$ over Θ.

In case only (Don) holds, one can still infer the limit law for the empirical supremal OT value.

Theorem 7 (*OT supremum over compact Θ*). *Consider the setting of Theorem 5 and only assume (Don). Then, it follows that*

$$\sqrt{\frac{nm}{n+m}}\left(\sup_{\theta\in\Theta} OT(\mu_n, \nu_m, c_\theta) - \sup_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right) \rightsquigarrow \sup_{\substack{\theta\in S_+(\Theta,\mu,\nu)\\ f_\theta\in S_{c_\theta}(\mu,\nu)}} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}),$$

where $S_+(\Theta, \mu, \nu) = \operatorname{argmax}_{\theta\in\Theta} OT(\mu, \nu, c_\theta)$ denotes the set of maximizers of $OT(\mu, \nu, c_\theta)$ over Θ.

The proofs of Theorems 6 and 7 are documented in Sections 6.3.2 and 6.3.3, respectively. Moreover, in some contexts the compactness assumption on Θ might be too restrictive. The following result provides an extension to non-compact spaces Θ and focuses on the infimal statistic; an analogue statement also holds for the supremal statistic. Its proof is deferred to Section 6.3.4.

Proposition 8 (*OT infimum over general Θ*). *Let Θ be a Polish space and consider a continuous map $c : \Theta \to C(\mathcal{X} \times \mathcal{Y})$. Consider again two measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ and suppose there is a compact set $K \subseteq \Theta$ such that $S_-(\Theta, \mu, \nu) \subseteq K$, there is a sequence of minimizers $\theta_{n,m} \in S_-(\Theta, \mu_n, \nu_m)$ with $\lim_{n,m\to\infty} \mathrm{P}^*(\theta_{n,m} \notin K) = 0$, and that the assumptions of Theorem 6 hold with Θ replaced by K. Then, the assertion of Theorem 6 on the empirical infimal OT value over Θ remains valid.*

Remark 9. A few comments are in order concerning the weak limits for the empirical OT cost process as well as the respective infimal and supremal statistic.

(i) In the setting of Theorem 5 the parameter space Θ is compact and $c : \Theta \to C(\mathcal{X} \times \mathcal{Y})$ is continuous, therefore the range $c(\Theta)$ is also compact in $C(\mathcal{X} \times \mathcal{Y})$. In particular, by the Theorem of Arzelà-Ascoli, we conclude that $\sup_{\theta\in\Theta} \|c_\theta\|_\infty < \infty$ and there exists a suitable modulus of continuity for all cost functions uniformly on Θ.

(ii) Both assumptions of Theorem 5 and sufficient conditions are discussed in Sections 3.4 and 3.5. Assumption (Don) appears natural in order to control the empirical OT process uniformly over Θ, whereas (KP) is to ensure that the limit process is supported in $C(\Theta)$ and stays tight. Our proof technique suggests that (KP) can be slightly lifted, but not much. For instance, one could demand that Kantorovich potentials $S_{c_\theta}(\mu, \nu)$ which attain the supremum in the derivative can be approximated by Kantorovich potentials $S_{c_{\theta'}}(\mu, \nu)$ for θ' in the immediate vicinity of θ (as required in Lemma 53(i) in Appendix A). In particular, if $\Theta := \{\theta_1, \dots, \theta_K\}$ is a finite set equipped with discrete topology, then (KP) can be omitted.

(iii) The results also extend to the one-sample setting, *i.e.*, when μ is estimated by μ_n and ν is assumed to be known. For the one-sample version of Theorem 5 it suffices to assume in (Don) that the function class $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}$ is μ-Donsker in conjunction with (KP). Upon selecting $f_\theta \in S_{c_\theta}(\mu, \nu)$ for any $\theta \in \Theta$, the limit distribution is then given for $n \to \infty$ by

$$\sqrt{n}\Big(OT(\mu_n, \nu, c_\theta) - OT(\mu, \nu, c_\theta)\Big)_{\theta\in\Theta} \rightsquigarrow \big(\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})\big)_{\theta\in\Theta} \quad \text{in } C(\Theta).$$

Under identical assumptions, the one-sample analogue of Theorem 6 is available. For the validity of the one-sample result in Theorem 7 it suffices that $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}$ is μ-Donsker.

(iv) rem:NormalityDegeneracyhe obtained weak limits highlight an intimate dependency of limit distributions to the collection of Kantorovich potentials. In Theorem 5 the limit process is centered Gaussian due to Assumption List **(KP)**. For fixed $\theta \in \Theta$ the limiting random variables degenerates to a Dirac measure at zero if the respective Kantorovich potentials are (μ, ν)-almost surely constant. Moreover, the limit distribution in Theorem 6 is also centered normal if Kantorovich potentials $(f^{c_\theta c_\theta}, f^{c_\theta})$ for μ, ν and c_θ coincide (up to a constant shift) on $\text{supp}(\mu) \times \text{supp}(\nu)$ for any $\theta \in S_-(\Theta, \mu, \nu)$. Under analogous assumptions for $\theta \in S_+(\Theta, \mu, \nu)$ the limit distribution in Theorem 7 is centered normal. In particular, assuming (KP), this condition is fulfilled if $S_-(\Theta, \mu, \nu)$ or $S_+(\Theta, \mu, \nu)$ consist of a singleton. The resulting limit distributions degenerate if Kantorovich potentials are (μ, ν)-almost surely constant. A sharp characterization of almost surely constant potentials is detailed in Section 4 of Hundrieser et al. [44].

2.4. Bootstrap principle for optimal transport costs

Since the limit distributions in Theorems 2, 6 and 7 involve the set of Kantorovich potentials (and OT plans), under non-unique optimizers there is little hope for an explicit, closed-form description of the quantiles for these distributions, which is required for further practical purposes. To circumvent this issue we suggest the use of a k-out-of-n bootstrap procedure with $k = o(n)$ whose consistency is shown in this subsection.

For simplicity, we state the subsequent results for equal sample sizes, *i.e.*, $n = m$ as well as bootstrap samples of equal size $k = o(n)$. Under differing sample sizes $n \neq m$ one would select bootstrap samples of size $k = o(n), l = o(m)$ such that $l/(l + k) \approx m/(n + m)$. Below, we always consider the same bootstrap approach that we now introduce. For the two sequences of i.i.d. random variables $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$, $\{Y_i\}_{i=1}^n \sim \nu^{\otimes n}$, with respective empirical measures μ_n, ν_n, consider another sequence of i.i.d. bootstrap random variables $\{X_i^b\}_{i=1}^k \sim \mu_n^{\otimes k}$, $\{Y_i^b\}_{i=1}^k \sim \nu_n^{\otimes k}$ and define the bootstrap empirical measures $\mu_{n,k}^b := \frac{1}{k}\sum_{i=1}^k \delta_{X_i^b}$ and $\nu_{n,k}^b := \frac{1}{k}\sum_{i=1}^k \delta_{Y_i^b}$. Moreover, we write in the subsequent statement c_n for the cost estimator and $c_{n,k}^b$ for the bootstrap cost estimator.

Proposition 10 (*Bootstrap for OT under weakly converging costs*). *In the setting of Theorem 2, assume (JW) and either (OP) or (Sup). Let $c_{n,k}^b \in C(\mathcal{X} \times \mathcal{Y})$ be the bootstrap cost estimator such that $c_{n,k}^b(x, y)$ is measurable for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Further, assume the following.*

(JW)* *The bootstrap empirical processes are conditionally on $X_1, \dots, X_n, Y_1, \dots Y_n$ consistent in the space $\ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X} \times \mathcal{Y})$ for $n, k \to \infty$ with $k = o(n)$, i.e.,*

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\begin{pmatrix}\mu_{n,k}^b - \mu_n\\ \nu_{n,k}^b - \nu_n\\ c_{n,k}^b - c_n\end{pmatrix}\middle| X_1, \dots, X_n, Y_1, \dots Y_n\right), \mathcal{L}\left(\sqrt{n}\begin{pmatrix}\mu_n - \mu\\ \nu_n - \nu\\ c_n - c\end{pmatrix}\right)\right) \xrightarrow{\mathrm{P}^*} 0.$$

In case of setting (Sup) additionally assume the following.

(Sup)* *The unconditional bootstrap empirical processes $\mathbb{G}_{n,k}^\mu := \sqrt{k}(\mu_{n,k}^b - \mu)$ and $\mathbb{G}_{n,k}^\nu := \sqrt{k}(\nu_{n,k}^b - \nu)$ fulfill the conditions*

$$\sup_{f\in\mathcal{F}} \mathbb{G}_{n,k}^\mu(f^{c_{n,k}^b c_{n,k}^b} - f^{cc}) \xrightarrow{\mathrm{P}^*} 0 \text{ and } \sup_{f\in\mathcal{F}} \mathbb{G}_{n,k}^\nu(f^{c_{n,k}^b} - f^c) \xrightarrow{\mathrm{P}^*} 0 \text{ for } n, k \to \infty, k = o(n).$$

Then, it follows for $n, k \to \infty$ with $k = o(n)$ that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\left(OT(\mu_{n,k}^b, \nu_{n,k}^b, c_{n,k}^b) - OT(\mu_n, \nu_n, c_n)\right) \middle| X_1, \dots, X_n, Y_1, \dots, Y_n\right), \mathcal{L}\left(\sqrt{n}\left(OT(\mu_n, \nu_n, c_n) - OT(\mu, \nu, c)\right)\right)\right) \xrightarrow{\mathrm{P}^*} 0.$$

Despite not relying on the functional delta method for the derivation of the limit distribution of the empirical OT value under weakly converging costs, we obtain a similar bootstrap principle as Dümbgen [53, Proposition 2] by employing an equivalent formulation for bootstrap consistency [54] in conjunction with the use of a Skorokhod representation. The full proof is provided in Section 6.2.2.

Remark 11. When employing the functional delta method, the n-out-of-n bootstrap is not consistent if the Hadamard directional derivative is not linear [53, Proposition 1]. Although Proposition 10 does not build on a differentiability result, we show in Section 4.5 that the OT functional is Gateaux directional differentiability with a derivative that is non-linear if primal or dual optimizers are non-unique. Since Gateaux directional differentiability is implied by Hadamard directional differentiability, this suggests in the regime of non-unique optimizers the inconsistency of the naive n-out-of-n bootstrap for the empirical OT cost under weakly converging costs.

Verification of the bootstrap consistency in the settings of Theorems 5–7 is straightforward. It is a direct consequence of consistency of the k-out-of-n bootstrap empirical processes with $k = o(n)$ [38, Theorem 3.6.13] and the functional delta method for the bootstrap [53, Proposition 2]. Hence, we omit the proof of the following proposition.

Proposition 12 (*Bootstrap for OT process, supremum, and infimum*)**.** *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and Θ a compact topological space. Consider a continuous map $c : \Theta \to C(\mathcal{X} \times \mathcal{Y}), \theta \mapsto c_\theta$, let $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ and assume (Don).*

(i) (OT process in $C(\Theta)$) Then, under (KP), it follows for $n, k \to \infty$ with $k \le n$ that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\left(OT(\mu^b_{n,k}, \nu^b_{n,k}, c_\theta) - OT(\mu_n, \nu_n, c_\theta)\right)_{\theta\in\Theta} | X_1, \dots, X_n, Y_1, \dots, Y_n\right),\right.$$
$$\left.\mathcal{L}\left(\sqrt{n}\left(OT(\mu_n, \nu_n, c_\theta) - OT(\mu, \nu, c_\theta)\right)\right)_{\theta\in\Theta} \right) \xrightarrow{P^*} 0.$$

(ii) (OT infimum over Θ) Then, under (KP), it follows for $n, k \to \infty$ with $k \le o(n)$ that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\left(\inf_{\theta\in\Theta} OT(\mu^b_{n,k}, \nu^b_{n,k}, c_\theta) - \inf_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta)\right) | X_1, \dots, X_n, Y_1, \dots, Y_n\right),\right.$$
$$\left.\mathcal{L}\left(\sqrt{n}\left(\inf_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta) - \inf_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right)\right) \right) \xrightarrow{P^*} 0.$$

(iii) (OT supremum over Θ) Then, it follows for $n, k \to \infty$ with $k = o(n)$ that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\left(\sup_{\theta\in\Theta} OT(\mu^b_{n,k}, \nu^b_{n,k}, c_\theta) - \sup_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta)\right) | X_1, \dots, X_n, Y_1, \dots, Y_n\right),\right.$$
$$\left.\mathcal{L}\left(\sqrt{n}\left(\sup_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta) - \sup_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right)\right) \right) \xrightarrow{P^*} 0.$$

Notably, we also obtain consistency of the n-out-of-n bootstrap for setting (i) since (KP) implies linearity of the Hadamard directional derivative.

3. Discussion of the assumptions

In this section we discuss the assumptions on the distributional limits and the bootstrap consistency. We also provide sufficient conditions for their validity. All the proofs are deferred to Appendix B.

3.1. Assumptions (JW) and (JW)*: Joint weak convergence

For the empirical OT value under estimated costs we demand in (JW) and (JW)* weak convergence of the empirical processes in $\ell^\infty(\mathcal{F}^{cc})$ and $\ell^\infty(\mathcal{F}^c)$, where $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ is selected as in Theorem 2. This requires $\mathcal{F}^{cc}$ and $\mathcal{F}^c$ to be μ- and ν-Donsker, respectively. Moreover, we demand weak convergence of the estimated cost function in $C(\mathcal{X}\times\mathcal{Y})$ to ensure that any sequence of OT plans for μ_n, ν_m and $c_{n,m}$ tends towards an OT plan in $\Pi^\star_c(\mu, \nu)$. Finally, we stress the necessity of *joint* weak convergence in (JW) and (JW)* as the limit distribution is determined by the random variable $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ and thus characterized by their dependency.

Even though apparently unavoidable, these conditions are somewhat restrictive and delimit the theory to low dimensional settings. This is to be expected as estimation of the OT value (under population costs) suffers from the curse of dimensionality [41], leading to slow convergence rates when both population measures μ, ν exhibit high-dimensional support. However, in view of the recently discovered *lower complexity adaptation* principle [43], it suffices that one measure, μ or ν, is supported on a low dimensional space. The following proposition provides bounds on the covering numbers (see the notation section for a definition) of $\mathcal{F}^c$ and $\mathcal{F}^{cc}$ under uniform norm which leads to a universal Donsker property for both function classes.

Proposition 13 (*Universal Donsker property*)**.** *Let $c \in C(\mathcal{X} \times \mathcal{Y})$ be a continuous cost function with $\|c\|_\infty \le 1$. Assume one of the three settings.*

(i) $\mathcal{X} = \{x_1, \dots, x_N\}$ is a finite space (and no additional assumption on c).

(ii) There exists a pseudo metric[3] $\tilde{d}_{\mathcal{X}}$ *on* $\mathcal{X}$ *such that* $\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\mathcal{X}}) \lesssim \varepsilon^{-\beta}$ *for* $\varepsilon > 0$ *sufficiently small and some* $\beta \in (0,2)$ *and* $c(\cdot, y)$ *is* 1*-Lipschitz under* $\tilde{d}_{\mathcal{X}}$ *for all* $y \in \mathcal{Y}$.

(iii) $\mathcal{X} = \bigcup_{i=1}^{I} \zeta_i(\mathcal{U}_i)$ *for* $I \in \mathbb{N}$ *compact, convex subsets* $\mathcal{U}_i \subseteq \mathbb{R}^{d_i}$, $d_i \leq 3$ *with non-empty interior and maps* $\zeta_i : \mathcal{U}_i \to \mathcal{X}$ *such that for each* $i \in \{1, \dots, I\}$ *the function* $c(\zeta_i(\cdot), y)$ *is* $(\gamma_i, 1)$*-Hölder*[4] *on* $\mathcal{U}_i$ *for some* $\gamma_i \in (d_i/2, 2]$ *for all* $y \in \mathcal{Y}$.

Let $B \geq 0$ *and consider a modulus of continuity* $w : \mathbb{R}_+ \to \mathbb{R}_+$ *with respect to a metric* $d_{\mathcal{X}}$ *on* $\mathcal{X}$. *Then, for each setting there exists some* $\alpha < 2$ *such that for* $\varepsilon > 0$ *sufficiently small,*

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^c, \|\cdot\|_\infty) = \log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \text{for } \mathcal{F} = \mathcal{F}(B, w),$$

where the hidden constant depends for (i) on N, *for (ii) on* $\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\mathcal{X}})$, *and for (iii) on* $(\zeta_i, \mathcal{U}_i)_{i=1}^{I}$. *In particular, the function classes* $\mathcal{F}^c$ *and* $\mathcal{F}^{cc}$ *are universal Donsker.*

The bounds for the covering numbers stated in the above proposition are essential for the weak convergence of the empirical processes $\sqrt{n}(\mu_n - \mu)$ and $\sqrt{m}(\nu_m - \nu)$ and represent an important tool for verifying (JW). In order to clarify the assumptions of Proposition 13, we rephrase them in the example below. We additionally refer to [44, Section 5] and [43, Section 3] for more illustrative examples of similar type.

Example 14. Below we list a few cases for which Proposition 13 is applicable and asserts that the function classes $\mathcal{F}^{cc}$ and $\mathcal{F}^c$ are universal Donsker.

1. Both measures are supported on finitely many points and the cost function is real-valued. This includes measures on finite trees and a tree metric as the cost function (see, *e.g.*, [14] for an application in genetics).
2. Only one measure is supported on finitely many points and the cost function is bounded (see, *e.g.*, [55] for an application in resource allocation).
3. The cost function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is locally Lipschitz and both measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ are compactly supported. In particular, this setting is met for the cost function $c_p(x, y) = |x - y|^p$ of the p-Wasserstein distance on the real line if $p \geq 1$ (see, *e.g.*, [34] for an application in clinical trials).
4. Under $d \leq 3$, the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable and both measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are compactly supported. This setting captures the cost function $c_p(x, y) = \|x - y\|^p$ of the Wasserstein distance on Euclidean spaces for dimension d (see, *e.g.*, [13] for an application in cell biology).
5. On general Euclidean spaces, similar conditions can be stated if the minimum intrinsic dimension of both measures is sufficiently small. For locally Lipschitz costs $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ we require that one measure is concentrated on a compact polygonal path or a Lipschitz curve. In case of twice continuously differentiable costs it suffices that one measure is concentrated on a compact smooth submanifold $\mathcal{X}$ of dimension at most three.

To state sufficient conditions for (JW) and (JW)* we assume that the population cost as well as the empirical and bootstrap estimators are determined by the underlying measures via a Hadamard directionally differentiable functional. For simplicity, we consider in the subsequent proposition random variables $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}, \{Y_i\}_{i=1}^n \sim \nu^{\otimes n}$ of identical sample size n with empirical measures μ_n, ν_n, and bootstrap samples $\{X_i^b\}_{i=1}^k \sim \mu_n^{\otimes k}$, $\{Y_i^b\}_{i=1}^k \sim \nu_n^{\otimes k}$ of size $k = k(n) = \mathrm{o}(n)$ with corresponding bootstrap empirical measures $\mu_{n,k}^b, \nu_{n,k}^b$.

Proposition 15 (*Joint weak convergence*). *Let* $\mathcal{F}_{\mathcal{X}}$, $\mathcal{F}_{\mathcal{Y}}$ *be bounded function classes on* $\mathcal{X}$ *and* $\mathcal{Y}$, *respectively, and assume there is a functional* $\Phi_c : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \subseteq \ell^\infty(\mathcal{F}_{\mathcal{X}}) \times \ell^\infty(\mathcal{F}_{\mathcal{Y}}) \to C(\mathcal{X} \times \mathcal{Y})$ *such that, for all* $n, k \in \mathbb{N}$,

$$c = \Phi_c(\mu, \nu), \quad c_n = \Phi_c(\mu_n, \nu_n), \quad \text{and} \quad c_{n,k}^b = \Phi_c(\mu_{n,k}^b, \nu_{n,k}^b).$$

If Φ_c *is Hadamard directionally differentiable at* (μ, ν) *tangentially to* $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$, *and if* $\mathcal{F}_{\mathcal{X}} \cup \mathcal{F}^{cc}$ *is* μ*-Donsker while* $\mathcal{F}_{\mathcal{Y}} \cup \mathcal{F}^c$ *is* ν*-Donsker, then both (JW) and (JW)* are fulfilled.*

Remark 16. We like to point out that if the functional Φ_c is additionally continuous with respect to the topology induced by weak convergence on $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$, it follows that $c_n(x, y)$ and $c_{n,k}^b(x, y)$ are measurable for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and, due to compactness of $\mathcal{X}$ and $\mathcal{Y}$, measurable in $C(\mathcal{X} \times \mathcal{Y})$.

3.2. Assumption (OP) : Uniqueness of optimal transport plans

The subject of uniqueness of OT plans between probability measures and a given cost function is of long-standing interest and has been addressed by various authors. General conditions for continuous settings were derived by Gangbo and McCann [56] and Levin

[3] A non-negative function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+$ on a set $\mathcal{M}$ is a pseudo-metric if the three conditions $d(x, x) = 0$, $d(x, y) = d(y, x)$ and $d(x, y) \leq d(x, z) + d(z, y)$ are fulfilled for any $x, y, z \in \mathcal{M}$.

[4] A function $f : \mathcal{U} \to \mathbb{R}$ on a convex set $\mathcal{U} \subseteq \mathbb{R}^d$ with non-empty interior is (γ, Λ)-Hölder with modulus $\Lambda \geq 0$ and $\gamma \in (0, 1]$ if $\|f\|_\infty < \Lambda$ and $|f(x) - f(y)| \leq \Lambda \|x - y\|^\gamma$ for any $x, y \in \mathcal{U}$. Further, f is called (γ, Λ)-Hölder for $\gamma \in (1, 2]$ if every partial derivative of f is $(\gamma - 1, \Lambda)$-Hölder. If $\mathcal{U}$ is not open, we assume the existence of an extension $\tilde{f}$ of f onto an open convex set containing $\mathcal{U}$ such that $\tilde{f}$ is (γ, Λ)-Hölder theorem, cf. Hundrieser et al. [43].

[57], building on previous works. The subject has since been covered in depth in Chapters 9 and 10 of the reference textbook by Villani [3]; further advances have been made since.

To guarantee the uniqueness of the OT plan, many works resort to the so-called *Twist condition* which demands for differentiable costs the injectivity of the map $y \to \nabla_x c(x,y)$ for all $x \in \mathcal{X}$. The following proposition formalizes a uniqueness criterion based on this condition and should fulfill the reader's needs for many practical applications. The result can be deduced from Theorem 10.28 and Remark 10.33 in Villani [3].

Proposition 17. *Assume that $\mathcal{X}, \mathcal{Y}$ are compact Polish spaces where $\mathcal{X} \subseteq \mathbb{R}^d$ is a Euclidean subset with non-empty interior and μ is absolutely continuous with respect to the Lebesgue measure. Further, assume that c is locally Lipschitz on $\mathcal{X}\times\mathcal{Y}$, that $c(\cdot, y)$ is differentiable on $\operatorname{int}(\mathcal{X})$ for each $y \in \mathcal{Y}$ and that $y \mapsto \nabla_x c(x,y)$ is injective for each $x \in \mathcal{X}$. Then, the OT plan is unique.*

Even though in certain cases weaker conditions can yield uniqueness [58], these more general conditions are typically considerably more difficult to verify. Nevertheless, unless the cost function exhibits some kind of symmetry or is constant in some region, uniqueness of OT plans is often to be expected. Indeed, for fixed measures there is a residual set of cost functions such that for any such costs the OT plan is unique [59].

In finite discrete settings, *i.e.*, when both underlying measures are supported on finitely many points, results on the uniqueness of OT plans are mostly based on the theory of finite-dimensional linear programs, and we refer to Klatt et al. [60, Section 6] for a detailed account. Among others, they provide sufficient conditions for uniqueness of OT plans which solely depend on the cost function and the support points but are independent of the weights of the measures. For Euclidean-based costs their condition is fulfilled for Lebesgue-almost every arrangement of support points of μ and ν, it is however violated if the support points obey some regular or repetitive pattern.

3.3. Assumptions (Sup) and (Sup)*: Control of supremum over empirical processes

Our assumptions on the suprema of the empirical processes ensure that the fluctuation on the set of feasible dual potentials caused by estimation of the cost function is asymptotically negligible. Let us also point out that the suprema in (Sup) and (Sup)* are (Borel) measurable by Lemmas 45 and 46. This implies that the convergence in outer probability occurs, in fact, in probability. Indeed, following along the proof of Lemma 46 and due to measurability of $c_{n,m}$ it follows for fixed $f \in \mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ that both maps $\omega \mapsto \mathbb{G}_n^\mu(f^{cc})$ and $\omega \mapsto \mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}})$ are measurable. In conjunction with $\mathbb{G}_n^\mu((\cdot)^{cc}), \mathbb{G}_n^\mu((\cdot)^{c_{n,m}c_{n,m}}) \in C_u(\mathcal{F}, \|\cdot\|_\infty)$ by Lemma 45 and compactness of $(\mathcal{F}, \|\cdot\|_\infty)$ the measurability of the $\mathbb{G}_n^\mu((\cdot)^{cc} - (\cdot)^{c_{n,m}c_{n,m}})$ as well as its supremum follow.

In the following, we derive sufficient conditions for the validity of Assumption (Sup) (as well as Assumption (Sup)*). Based on empirical process theory, in order to suitably control the suprema

$$\sup_{f\in\mathcal{F}} \mathbb{G}_n^\mu(f^{c_{n,m},c_{n,m}} - f^{cc}) \quad \text{and} \quad \sup_{f\in\mathcal{F}} \mathbb{G}_n^\nu(f^{c_{n,m}} - f^{c})$$

a canonical route would be to impose metric entropy bounds for $\mathcal{F}^{c_{n,m},c_{n,m}} \cup \mathcal{F}^{cc}$ and $\mathcal{F}^{c_{n,m}} \cup \mathcal{F}^{c}$. Such bounds, however, would impose certain regularity requirements on the cost estimator $c_{n,m}$. Hence, in order not to narrow our scope concerning cost estimators, we employ the same ideas as in Section 2.2 and approximate the cost estimator $c_{n,m}$ by a more regular cost estimator $\tilde{c}_{n,m}$. The subsequent result formalizes these considerations for our context. Its proof relies on techniques developed by van der Vaart and Wellner [61] for empirical processes indexed over estimated function classes.

Proposition 18. *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and consider a continuous cost function c.*

(i) Assume (JW) for random elements $c_{n,m} \in C(\mathcal{X}\times\mathcal{Y})$. Take random elements $\tilde{c}_{n,m} \in C(\mathcal{X}\times\mathcal{Y})$ with $\sqrt{nm/(n+m)}\|c_{n,m} - \tilde{c}_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0$ for $n, m = m(n) \to \infty$ and $m/(n+m) \to \lambda \in (0,1)$ such that for $\varepsilon > 0$ sufficiently small,

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) + \sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \text{with } \alpha < 2. \tag{12}$$

Then Assumption (Sup) is fulfilled.

(ii) Assume (JW) and (JW) for random elements $c^b_{n,k} \in C(\mathcal{X}\times\mathcal{Y})$ and let $\tilde{c}^b_{n,k} \in C(\mathcal{X}\times\mathcal{Y})$ be random elements with $\sqrt{k}\|c^b_{n,k} - \tilde{c}^b_{n,k}\|_\infty \xrightarrow{\mathrm{P}} 0$ for $n, k = k(n) \to \infty$ and $k = \mathrm{o}(n)$ such that for $\varepsilon > 0$ sufficiently small,*

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) + \sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \text{with } \alpha < 2. \tag{13}$$

Then Assumption (Sup) is fulfilled.*

As a straightforward corollary of Proposition 18 we find that (Sup) and (Sup)* are fulfilled if the cost estimators $c_{n,m}$ and $c^b_{n,k}$ fulfill certain deterministic regularity conditions once n, m, k are sufficiently large. In the large sample regime we then choose $\tilde{c}_{n,m} := c_{n,m}$ and $\tilde{c}^b_{n,k} := c^b_{n,k}$.

Corollary 19. *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces, consider a continuous cost function c. Assume (JW) for c_n (and (JW)* for $c^b_{n,k}$) and that c, c_n (and $c^b_{n,k}$) each fulfill one of the three conditions of Proposition 13 for $n \geq N$, $k \geq K$ with random variables $N, K \in \mathbb{N}$. Then, (Sup) (and (Sup)*) hold.*

Hence, if the population cost c and the estimators $c_n, c_{n,k}$ are determined by some parameter $\theta \in \Theta$ and estimators $\theta_n, \theta_{n,k}$, such that the regularity properties of Proposition 13 are met uniformly in an open neighborhood of θ and if the estimators are consistent, then Corollary 19 asserts the validity of Assumptions (Sup) and (Sup)*.

Moreover, under mild additional assumptions on the space $\mathcal{X}$ and the cost function c, we can define $\Psi : C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$, a functional such that $\tilde{c}_n := \Psi(c_n)$ fulfills the entropy bound (12) while satisfying $\sqrt{n}\,\|\tilde{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0$ for $n \to \infty$. We call such a functional Ψ a *regularity elevation functional* since it lifts the degree of regularity of the cost estimator. Details on regularity elevations are deferred to Section 5.

Corollary 20. *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and consider a continuous cost. Assume (JW) (and (JW)*). Suppose that c fulfills one of the three conditions of Proposition 13. Under (ii) or (iii) further assume the subsequent condition (ii)' or (iii)', respectively.*

(ii)' The weak limit $\mathbb{G}^c$ is almost surely continuous with respect to $(\mathcal{X}, \tilde{d}_{\mathcal{X}}) \times \mathcal{Y}$.

(iii)' For each $i \in \{1, \dots I\}$ the set $\mathcal{U}_i \subseteq \mathbb{R}^{d_i}$ is convex and compact, the map $\zeta_i : \mathcal{U}_i \to \zeta_i(\mathcal{U}_i)$ is a homeomorphism, and the function $c_i := c(\zeta_i(\cdot), \cdot) : \mathcal{U}_i \times \mathcal{Y} \to \mathbb{R}$ is continuously differentiable in $u \in \mathcal{U}_i$ on $\mathcal{U}_i \times \mathcal{Y}$, i.e., the derivative $\nabla_u c_i : \mathrm{int}(\mathcal{U}_i) \times \mathcal{Y} \to \mathbb{R}^d$ can be continuously extended to $\mathcal{U}_i \times \mathcal{Y}$. Further, there exists a continuous partition of unity[5] $\{\eta_i\}_{i=1}^I$ on $\mathcal{X}$ with $\mathrm{supp}(\eta_i) \subseteq \zeta_i(\mathcal{U}_i)$.

Then, Assumption (Sup) (and (Sup)) is fulfilled.*

Recall that the assumptions in Proposition 13 are all met in the settings described in Example 14. The additional assumptions in Corollary 20 are fairly mild. For instance, Assumption *(ii)'* on $\mathbb{G}^c$ is always met if $d_{\mathcal{X}}$ metrizes the topology on $\mathcal{X}$. Further, Assumption *(iii)'* holds if $\mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}^d$ with $d \le 3$ are convex, compact sets and c is a twice continuously differentiable cost function on $\mathbb{R}^d \times \mathbb{R}^d$. Assumption *(iii)'* is also fulfilled for general $d \in \mathbb{N}$ if $\mathcal{X} \subseteq \mathbb{R}^d$ is a compact submanifold of dimension at most three while $\mathcal{Y} \subseteq \mathbb{R}^d$ is bounded subset and $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable. This way, thanks to compactness of $\mathcal{X}$, a partition of unity always exists and up to potentially choosing an atlas of smaller charts $(\zeta_i, \mathcal{U}_i)$ for $\mathcal{X}$, the continuous extendability of $\nabla_u c_i : \mathrm{int}(\mathcal{U}_i) \times \mathcal{Y} \to \mathbb{R}^d$ can always be ensured.

3.4. *Assumption (Don): Donsker property uniformly over Θ*

For the distributional limits by Hundrieser et al. [44] on the empirical OT value under a fixed cost function c, the authors effectively assume that the function classes $\mathcal{F}^{cc}$ and $\mathcal{F}^c$ are μ- and ν-Donsker, respectively (Remark 4(i)). Hence, for the uniform convergence result from Theorem 5 it is natural that we demand the μ- and ν-Donsker property for the unions $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}$ and $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}$ for $\mathcal{F} = \mathcal{F}(\sup_{\theta\in\Theta}\|c\|_\infty, w)$. The validity of this condition can be ensured under assumptions on the domain $\mathcal{X}$ in conjunction with regularity conditions imposed on the cost function.

Proposition 21 (*Universal Donsker property over Θ*). *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and let (Θ, d_Θ) be a metric space such that $\log \mathcal{N}(\varepsilon, \Theta, d_\Theta) \lesssim \varepsilon^{-\alpha}$ for $\alpha < 2$. Suppose that $c : (\Theta, d_\Theta) \to C(\mathcal{X}\times\mathcal{Y}), \theta \mapsto c_\theta$ is 1-Lipschitz and assume $\sup_{\theta\in\Theta}\|c_\theta\|_\infty \le 1$. Consider one of the three settings.*

(i) $\mathcal{X} = \{x_1, \dots, x_N\}$ is a finite space (and no additional assumption on c).

(ii) For any $\theta \in \Theta$ there exists a pseudo metric $\tilde{d}_{\theta,\mathcal{X}}$ on $\mathcal{X}$ such that $\sup_{\theta\in\Theta}\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\theta,\mathcal{X}}) \lesssim \varepsilon^{-\beta}$ for $\beta < 2$ and $c_\theta(\cdot, y)$ is 1-Lipschitz under $\tilde{d}_{\theta,\mathcal{X}}$ for all $y \in \mathcal{Y}$.

(iii) $\mathcal{X} = \bigcup_{i=1}^I \zeta_i(\mathcal{U}_i)$ for $I \in \mathbb{N}$ compact, convex subsets $\mathcal{U}_i \subseteq \mathbb{R}^{d_i}$, $d_i \le 3$ with non-empty interior and maps $\zeta_i : \mathcal{U}_i \to \mathcal{X}$ so that for each $i \in \{1, \dots, I\}$ the function $c_\theta(\zeta_i(\cdot), y)$ is $(\gamma_i, 1)$-Hölder on $\mathcal{U}_i$ (recall footnote (iii)) for some $\gamma_i \in (d_i/2, 2]$ for all $y \in \mathcal{Y}$, $\theta \in \Theta$.

Then, for each setting, there exists some $\alpha < 2$ such that

$$\log \mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \text{and} \quad \log \mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha}.$$

In particular, $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}$, $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}$ are universal Donsker, and Assumption (Don) is fulfilled.

The proof of Proposition 21 is a simple consequence of Proposition 13 in combination with the subsequent lemma whose proof is deferred to Appendix B.6.

Lemma 22. *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and let (Θ, d_Θ) be a metric space. Suppose $c : (\Theta, d_\Theta) \to C(\mathcal{X}\times\mathcal{Y}), \theta \mapsto c_\theta$ is 1-Lipschitz. Then, it follows for any $\varepsilon > 0$ that*

$$\max\Big(\mathcal{N}\big(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}, \|\cdot\|_\infty\big), \mathcal{N}\big(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty\big)\Big) \le \mathcal{N}\left(\frac{\varepsilon}{4}, \Theta, d_\Theta\right) \sup_{\theta\in\Theta} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty\right).$$

[5] A collection $\{\eta_i\}_{i=1}^I$ is a continuous partition of unity if $\eta_i \in C(\mathcal{X})$, $\eta_i \ge 0$ for each i and $\sum_{i=1}^I \eta_i \equiv 1$ on $\mathcal{X}$.

3.5. Assumption (KP): Uniqueness of Kantorovich potentials

The uniform weak limit of the empirical OT process from Theorem 5 demonstrates a close relation to the collection of Kantorovich potentials. In particular, for the limit to be supported on $C(\Theta)$ a certain continuity property on the Kantorovich potentials $S_{c_\theta}(\mu, \nu)$ with respect to θ is required. Assumption (KP) on the uniqueness of Kantorovich potentials represents a sufficient condition to ensure this property.

The recent work by Staudt et al. [62] thoroughly analyzes the topic of uniqueness in Kantorovich potentials and highlights that it is often expected. More precisely, for differentiable costs and assuming that one probability measure is supported on the closure of a connected open set on a smooth manifold, Kantorovich potentials are unique. As Example 3 in their work showcases, uniqueness also occurs under continuous costs if one measure is discrete while the other has connected support. In case both measures have disconnected support, then uniqueness can still be guaranteed if potentials on restricted OT sub-problems are unique and if there exists, in the language of Staudt et al., a *non-degenerate* OT plan, meaning that all connected components of both measures are linked via that OT plan. The existence of such OT plans can be guaranteed under mild conditions on the underlying measures (see (14)) and intuitively demands that the OT problem cannot be divided into distinct sub-problems.

The following statement is a direct consequence of Staudt et al. [62], which we have included for ease of reference.

Proposition 23. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a differentiable cost function. Consider probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with compact support and suppose $\operatorname{supp}(\mu) = \bigcup_{i\in I} \mathcal{X}_i$ and $\operatorname{supp}(\nu) = \bigcup_{j\in J} \mathcal{Y}_j$ for finitely many disjoint sets. Assume each set $\mathcal{X}_i$ is either (i) the closure of a connected open set, (ii) the closure of a connected open set in a smooth compact submanifold of $\mathbb{R}^d$, or (iii) a single point. Further, if $\min(|I|, |J|) \geq 2$, suppose for all non-empty, proper $I' \subset I$ and $J' \subset J$ that*

$$\sum_{i\in I'} \mu(X_i) \neq \sum_{j\in J'} \nu(Y_j), \tag{14}$$

Then, Kantorovich potentials for μ, ν and c are unique (up to a constant shift).

The proposition follows by verifying the conditions of Theorem 1 in Staudt et al. [62]. Indeed, continuity of Kantorovich potentials on $\operatorname{supp}(\nu)$ follows due to the compactness assumption and continuity of the cost function, uniqueness of Kantorovich potentials on sub-problems follows from the assumptions on the cost and the sets $\mathcal{X}_i$ [62, Corollary 2], and existence of non-degenerate plans follows via [62, Lemma 6] due to (14).

4. Applications

In this section we employ our theory from Section 2 to obtain novel insights about various OT-related concepts and applications. All proofs for this section are deferred to Appendix C.

4.1. Optimal transport-based one-sample goodness-of-fit-testing

Hallin et al. [19] proposed to use the Wasserstein distance between a sample measure and a reference measure for goodness-of-fit testing under group actions. In the following, we briefly recall the setting for compactly supported measures. Let $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$ be compactly supported, define $\mathcal{Y}$ as the convex hull of $\operatorname{supp}(\nu_0)$, and let $G_\Theta = \{g_\vartheta : \vartheta \in \Theta\}$ be a group of measurable transformations $g_\vartheta : \mathbb{R}^d \to \mathbb{R}^d$ that is parametrized by $\vartheta \in \Theta \subseteq \mathbb{R}^k$ for $k \in \mathbb{N}$. Further, assume that the map $x \mapsto g_\vartheta(x)$ is continuous for every $\vartheta \in \Theta$ and that the mappings $\vartheta \mapsto g_\vartheta$ and $g_\vartheta \mapsto (g_\vartheta)_\# \nu_0$ are bijective (this implies the identifiability of the model parameter). Hallin et al. [19] consider the subsequent testing problem:

> Let G_Θ be a group and define $\mathcal{M} = \{g_{\vartheta\#}\nu_0 : g_\vartheta \in G_\Theta\}$. Given an i.i.d. sample $\{X_i\}_{i=1}^n$ from some unknown $\mu \in \mathcal{P}(\mathcal{X})$ with $\mathcal{X} \subset \mathbb{R}^d$ compact, the aim is to test
>
> $$\mathcal{H}_0^* : \mu \in \mathcal{M} \quad \text{against} \quad \mathcal{H}_1^* : \mu \notin \mathcal{M}. \tag{15}$$

Note that the parameter ϑ^* under $\mathcal{H}_0$, such that $(g_{\vartheta^*})_\# \nu_0 = \mu$, is unknown. To construct a test for the above hypothesis, which is for instance of particular interest in the analysis of location-scale families, the authors propose to rely on the (2-)Wasserstein distance, *i.e.*, Hallin et al. [19] propose a test based on an empirical version of

$$OT\left(\mu, \nu_0, \left\|g_{\vartheta^*}^{-1}(\cdot) - \cdot\right\|^2\right) = \inf_{\pi\in\Pi(\mu,\nu_0)} \int \left\|g_{\vartheta^*}^{-1}(x) - y\right\|^2 \mathrm{d}\pi(x, y).$$

For this purpose, the unknown measure μ is replaced by μ_n and the cost function $c(x, y) = \|g_{\vartheta^*}^{-1}(x) - y\|^2$ by $c_n(x, y) = \|g_{\vartheta_n}^{-1}(x) - y\|^2$, where $\vartheta_n \in \Theta$ denotes a suitable estimator for ϑ^*. Thus, the proposed test statistic is given as

$$OT\left(\mu_n, \nu_0, \left\|g_{\vartheta_n}^{-1}(\cdot) - \cdot\right\|^2\right) = \inf_{\pi\in\Pi(\mu_n,\nu_0)} \int \left\|g_{\vartheta_n}^{-1}(x) - y\right\|^2 \mathrm{d}\pi(x, y), \tag{16}$$

which amounts to solving an OT problem with an estimated cost function. Hence, we can apply our theory to derive the limiting distribution of

$$\sqrt{n}\left(OT\left(\mu_n, \nu_0, \left\|g_{\vartheta_n}^{-1}(\cdot) - \cdot\right\|^2\right) - OT\left(\mu, \nu_0, \left\|g_{\vartheta^*}^{-1}(\cdot) - \cdot\right\|^2\right)\right) \tag{17}$$

under the null hypothesis H_0^* in (15) (see Remark 27 for a discussion). In addition, we are able to extend this to testing whether H_0^* holds approximately, which is often preferable in practice (see, *e.g.*, [34,63,64]). For this purpose, we fix an estimation procedure for ϑ^*, *i.e.*, we choose a specific estimator ϑ_n (taking values in Θ) for estimating ϑ^* and denote its population quantity by $\vartheta^o \in \Theta$ (under $\mathcal{H}_0^*$ we assume $\vartheta^* = \vartheta^o$). Then, we consider the subsequent testing problem:

> *Let G_Θ be a group. Given an i.i.d. sample $\{X_i\}_{i=1}^n$ from some unknown $\mu \in \mathcal{P}(\mathcal{X})$ with $\mathcal{X} \subset \mathbb{R}^d$ compact, the aim is to test for some prespecified $\Delta > 0$ the hypothesis*
>
> $$\mathcal{H}_0 : W_2((g_{\vartheta^o}^{-1})_{\#}\mu, \nu_0) \le \Delta \quad \textit{against} \quad \mathcal{H}_1 : W_2((g_{\vartheta^o}^{-1})_{\#}\mu, \nu_0) > \Delta. \tag{18}$$

In order to construct a test for the above problem, we have to derive the distributional limits of (17) under the assumption that $\mu \notin \mathcal{M}$. To this end, we employ the theory from Sections 2 and 3. The first step for the derivation of distributional limits of (17) is to establish Hölder regularity (cf. footnote (iii)) for costs induced by $C_\Theta : \Theta \to C(\mathcal{X} \times \mathcal{Y}), \vartheta \mapsto ((x, y) \mapsto \|g_\vartheta^{-1}(x) - y\|^2)$ near ϑ^o.

Lemma 24. *Let $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ be compact and denote by $C(\mathcal{X}, \mathbb{R}^d)$ the space of continuous functions from $\mathcal{X}$ to $\mathbb{R}^d$. Assume that $K_\Theta : \Theta \subseteq \mathbb{R}^k \to C(\mathcal{X}, \mathbb{R}^d)$, $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ is continuous near ϑ^o. Then, there is an open (w.r.t. relative topology) neighborhood $U \subseteq \Theta$ of ϑ^o and some $\Lambda \ge 0$ such that for any $x \in \mathcal{X}$ and $\vartheta \in U$ the cost function $C_\Theta(\vartheta)(x, \cdot) := \|g_\vartheta^{-1}(x) - \cdot\|^2$ is $(2, \Lambda)$-Hölder on $\mathcal{Y}$.*

Next, we verify that Hadamard differentiability of K_Θ at ϑ^o implies Hadamard differentiability of the cost parametrizing map $C_\Theta : \Theta \to C(\mathcal{X} \times \mathcal{Y}), \vartheta \mapsto ((x, y) \mapsto \|g_\vartheta^{-1}(x) - y\|^2)$ at ϑ^o. To this end, we additionally impose the following assumption.

(G) For ϑ^o there exists $m_{\vartheta^o} > 0$ such that for all $\vartheta' \in \Theta$ in some neighborhood of ϑ^o,

$$\sup_{x \in \mathbb{R}^d} \frac{\left\| g_{\vartheta'}^{-1}(x) - g_{\vartheta^o}^{-1}(x) \right\|}{1 + \left\| g_{\vartheta^o}^{-1}(x) \right\|} \le m_{\vartheta^o} \left\| \vartheta' - \vartheta^o \right\|.$$

This condition is fulfilled, *e.g.*, for location-scale families and affine transformations. A global version of the above assumption, *i.e.*, where the condition is to be fulfilled for any ϑ and not only ϑ^0, has been used by Hallin et al. [19] to ensure the consistency of their goodness-of-fit test described above.

Lemma 25. *Assume that the function $K_\Theta : \Theta \to C(\mathcal{X}, \mathbb{R}^d)$, $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ is Hadamard differentiable at ϑ^o tangentially to Θ, i.e., for any sequence $(\vartheta^o + t_n h_n)_{n \in \mathbb{N}} \subseteq \Theta$ such that $t_n \searrow 0$ and $h_n \to h \in \mathbb{R}^k$ as $n \to \infty$,*

$$\lim_{n \to \infty} \left\| \frac{K_\Theta(\vartheta^o + t_n h_n) - K_\Theta(\vartheta^o)}{t_n} - D^H_{|\vartheta^o} K_\Theta(h) \right\|_\infty = 0,$$

where $D^H_{|\vartheta^o} K_\Theta(h) : \mathcal{X} \to \mathbb{R}^d$ is a continuous function. Then, if Assumption (G) is satisfied, C_Θ is Hadamard differentiable at ϑ^o tangentially to Θ with derivative $D^H_{|\vartheta^o} C_\Theta(h) \in C(\mathcal{X} \times \mathcal{Y})$ given by

$$D^H_{|\vartheta^o} C_\Theta(h) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}, \quad (x, y) \mapsto 2\left\langle D^H_{\vartheta^o} K_\Theta(h)(x), g_{\vartheta^o}^{-1}(x) - y \right\rangle.$$

Moreover, if $\sqrt{n}(\vartheta_n - \vartheta^o) \rightsquigarrow \mathbb{G}^\vartheta$ for $n \to \infty$, we obtain for $c_n := C_\Theta(\vartheta_n)$ and $c := C_\Theta(\vartheta)$ that

$$\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c := \left(2\left\langle D^H_{\vartheta^o} K_\Theta(\mathbb{G}^\vartheta)(x), g_{\vartheta^o}^{-1}(x) - y \right\rangle \right)_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \qquad \textit{in } C(\mathcal{X} \times \mathcal{Y}). \tag{19}$$

Under the conditions in the proposition above, our main result from Theorem 2 yields a (typically) non-degenerate limiting distributions for the statistic $OT\left(\mu_n, \nu, c_n\right)$ under the assumption that $\mu \notin \mathcal{M}$. In particular, this allows us to construct an asymptotic level α test for the null hypotheses given in (18) (see [34] for the precise construction).

Proposition 26. *Let $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$ for $d \le 3$ be compactly supported and define $\mathcal{Y}$ as the convex hull of $\operatorname{supp}(\nu_0)$, and let $\mathcal{X} \subseteq \mathbb{R}^d$ be compact. Assume that (G) holds, and suppose that $K_\Theta : \Theta \to C(\mathcal{X}, \mathbb{R}^d)$, $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ is continuous near ϑ^o and Hadamard differentiable at ϑ^o. Define for $\Lambda \ge 0$ from Lemma 25 the function class*

$$\mathcal{F} := \left\{ f : \mathcal{Y} \to \mathbb{R} \;\middle|\; \|f\|_\infty \le \Lambda + 1, |f(y) - f(y')| \le 2\Lambda \left\| y - y' \right\| \textit{ for all } y, y' \in \mathcal{Y} \right\}.$$

Then, the function class $\mathcal{F}^{C_\Theta(\vartheta^o)}$ on $\mathcal{X}$ is universal Donsker. Moreover, for i.i.d. random variables $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ consider a measurable estimator ϑ_n and suppose for $n \to \infty$ joint weak convergence,

$$\sqrt{n} \begin{pmatrix} \mu_n - \mu \\ \vartheta_n - \vartheta^o \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{G}^\mu \\ \mathbb{G}^\vartheta \end{pmatrix} \qquad \textit{in } \ell^\infty(\mathcal{F}^{C_\Theta(\vartheta^o)}) \times \mathbb{R}^k. \tag{20}$$

Then, for $c_n := C_\Theta(\vartheta_n)$ and $c := C_\Theta(\vartheta)$ and by denoting the limit from (19) as $\mathbb{G}^c$, it follows that

$$\sqrt{n} \left(OT\left(\mu_n, \nu_0, c_n\right) - OT\left(\mu, \nu_0, c\right) \right) \rightsquigarrow \inf_{\pi \in \Pi_c^\star(\mu, \nu_0)} \pi(\mathbb{G}^c) + \sup_{f \in S_c(\mu, \nu_0)} \mathbb{G}^\mu(f^c),$$

where $S_c(\mu, \nu_0)$ represents the set of optimizers for $\sup_{f \in \mathcal{F}} \mu(f^c) + \nu_0(f^{cc})$.

Remark 27. A few comments on the distributional limits are in order.

(i) Given that the function class $\mathcal{F}^{C_\Theta(\vartheta^o)}$ is universal Donsker and thus μ-Donsker, and assuming that $\sqrt{n}(\vartheta_n - \vartheta^o)$ converges in distribution, the requirement of joint convergence as required in (20) is very mild. Indeed, if $\sqrt{n}(\vartheta_n - \vartheta^o)$ can be expressed asymptotically in terms of a suitable linear functional of an empirical process, *i.e.*, if it admits an asymptotic influence function $\psi \in L^2(\mu)$ (cf. van der Vaart [65, p. 58]), joint convergence follows since the union $\mathcal{F}^{cc} \cup \{\psi\}$ is μ-Donsker.

(ii) We like to point out that Proposition 26 also remains valid if $\mu \in \mathcal{M}$. However, under this assumption it follows that $(g^{-1}_{\vartheta^o})_{\#}\mu = \nu_0$ which implies that the corresponding OT plan between μ and ν_0 is given by $\pi = (\mathrm{Id}, g^{-1}_{\vartheta^o}(\cdot))_{\#}\mu$. Hence, by (19) the process $\mathbb{G}^c$ vanishes along the support of π and the first term in the limit degenerates. Further, if the support of ν_0 is connected then Kantorovich potentials are unique up to a constant shift [62, Corollary 2] and ν_0-a.s. constant [44, Corollary 4.6(i)]. Consequently, for this setting the corresponding limit distribution is degenerate. In contrast, if ν_0 has disconnected support, non-constant Kantorovich potentials exist [62, Lemma 11] which results in a non-degenerate limit distribution.

(iii) The elements presented for the one-sample case can also be generalized to the case where both empirical measures undergo a transformation, either separately or jointly. One might think of choosing the Mahalanobis distance $(x-y)^\top \Sigma^{-1}(x-y)$ as a cost function where Σ^{-1} has to be estimated and could, *e.g.*, be a diagonal matrix. As the OT value is not invariant with respect to affine transformations, rescaling the variables would ensure that no component has an overwhelming impact on the cost function compared to the other components.

Let us now exhibit the use of Proposition 26 with the following example.

Example 28 (*Location-scale family*)**.** For a compactly supported probability measure $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$, $d \le 3$, that has mean equal to zero and covariance equal to the identity matrix, consider the location scale family $\mathcal{M} = \{g_{\vartheta\#}\nu_0 : \vartheta \in \Theta := \mathbb{R}^d \times \mathrm{SPD}(d)\}$ where $g_\vartheta(x) = \Sigma^{1/2}x + m$ for $\vartheta = (m, \Sigma) \in \Theta$. In consequence, the measure $g_{\vartheta\#}\nu_0$ has mean m and covariance Σ. Given i.i.d. sample $\{X_i\}_{i=1}^n$ from some unknown $\mu \in \mathcal{P}(\mathcal{X})$ for $\mathcal{X} \subseteq \mathbb{R}^d$ compact, we pose the question of how close μ is to the family $\mathcal{M}$ or alternatively how close $(g^{-1}_{\vartheta^o})_{\#}\mu$ is to ν, where $\vartheta^o = (m_0, \Sigma_0)$ with m_0 and Σ_0 denoting the mean and covariance of μ, respectively. To this end, we employ the empirical estimators for the mean and covariance $m_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\Sigma_n := \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$, define $\vartheta_n := (m_n, \Sigma_n)$, and consider the test statistic $\sqrt{n}\left(OT\left(\mu_n, \nu_0, c_n\right) - \Delta\right)$ where $c_n(x,y) = \|g^{-1}_{\vartheta_n}(x) - y\|^2$ and $\Delta > 0$ is some positive threshold (recall testing problem (18)). To apply Proposition 26 two types of conditions are required. The first concerns the regularity and differentiability of the parametrized cost functions, and the second addresses joint weak convergence (20). For the latter we note that the function class $\mathcal{F}^{C_\Theta(\vartheta^o)} \cup \{x \mapsto x_i \mid i \in \{1,\dots,d\}\} \cup \{x \mapsto x_i \cdot x_j \mid i,j \in \{1,\dots,d\}\}$ is Donsker, being a finite union of Donsker classes, which ensures the validity of (20) according to Remark 27(i). For the first type of conditions, we note that condition (G) is fulfilled, hence it suffices to confirm continuity and Hadamard differentiability of

$$K_\Theta : \Theta \to C(\mathcal{X}, \mathbb{R}^d), \quad \vartheta = (m, \Sigma) \mapsto \left(x \mapsto g^{-1}_\vartheta(x) = \Sigma^{-1/2}(x - m)\right)$$

at $\vartheta^o = (m_0, \Sigma_0)$. Based on Fréchet differentiability of the inverse matrix root (see Lemma 29 below) and with compactness of $\mathcal{X}$ it follows that this condition is met, with derivative given by

$$D^F_{|\vartheta^o} K_\Theta : \mathbb{R}^d \times \mathrm{S}(d) \to C(\mathcal{X}, \mathbb{R}^d), \quad h = (h^m, H^\Sigma) \mapsto \left(x \mapsto -\Sigma_0^{-\frac{1}{2}}\tilde{H}^\Sigma \Sigma_0^{-\frac{1}{2}}(x - m_0) - \Sigma_0^{-\frac{1}{2}} h^m\right),$$

where $\tilde{H}^\Sigma \in \mathbb{R}^{d\times d}$ denotes the unique solution to the (continuous-time) Lyapunov equation $\Sigma_0^{1/2}X + X\Sigma_0^{1/2} = H^\Sigma$. Proposition 26 now details the asymptotic distribution of the optimal transport based test statistic. In particular, under uniqueness of OT plan and potentials, it will be centered normal with a variance that could be estimated via plug-in estimators for $\pi \in \Pi_c^\star(\mu, \nu)$ and $\mathbb{G}^c$ as well as $\mathbb{G}^\mu$ and $f \in S_c(\mu, \nu_0)$. A potential difficulty to take into account is the dependency of $\mathbb{G}^c$ and $\mathbb{G}^\mu$, which however could be sidestepped by splitting samples (one for estimation of mean and covariance, one for estimation of μ). Alternatively, since Proposition 10 is applicable here and thanks to the explicit choice of the cost estimators, the quantiles for the test statistic could be estimated via a bootstrap k-out-of-n resampling scheme with $k = o(n)$ even without imposing uniqueness of OT plans and potentials.

Lemma 29. *The inverse matrix root* $\Im\Re : \mathrm{SPD}(d) \to \mathrm{SPD}(d), \Sigma \mapsto \Sigma^{-1/2}$ *defined on the set of symmetric positive definite matrices is Fréchet differentiable and thus also Hadamard directionally differentiable. The corresponding Fréchet derivative at an element* $\Sigma_0 \in \mathrm{SPD}(d)$ *is given by*

$$D_{|\Sigma_0}\Im\Re : \mathrm{S}(d) \to \mathrm{S}(d), \quad H \mapsto -\Sigma_0^{-1/2}\tilde{H}\Sigma_0^{-1/2}, \tag{21}$$

where $\tilde{H} \in \mathbb{R}^{d\times d}$ *represents the unique solution to the (continuous-time) Lyapunov equation* $\Sigma_0^{1/2}X + X\Sigma_0^{1/2} = H$.

4.2. Optimal transport with embedded invariances

In a similar spirit as to the previous section, another strand of the literature [20,66] aims at making OT invariant to a class of transformation $\mathcal{T}$, with $\tau : \mathbb{R}^d \to \mathbb{R}^d$ continuously differentiable for each $\tau \in \mathcal{T}$, by considering

$$\inf_{\tau\in\mathcal{T}} OT\left(\mu, \nu, \|\cdot - \tau(\cdot)\|^2\right) = \inf_{\tau\in\mathcal{T}} \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x - \tau(y)\|^2 \mathrm{d}\pi(x,y). \tag{22}$$

This distance is useful in many contexts, among which the word embedding problem or protein alignment. If the class of transformations considered is the set of rotations, analyses relying on that distance is coined Wasserstein–Procrustes Analysis [20,67]. Theorem 6 provides the required tools for statistical inference for the empirical version of the optimization problem in (22). We stress that the following proposition is the first result of its kind for empirical optimal transport under embedded invariances.

Corollary 30. *Consider a set of transformations $(\mathcal{T}, d_{\mathcal{T}})$ that is a compact metric space with $\log \mathcal{N}(\epsilon, \mathcal{T}, d_{\mathcal{T}}) \lesssim \epsilon^{-\alpha}$ for $\alpha < 2$. Let $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ be compact subsets and assume that the functional $c : (\mathcal{T}, d_{\mathcal{T}}) \to C(\mathcal{X} \times \mathcal{Y}), \tau \mapsto c_\tau$, with $c_\tau(x,y) = \|x - \tau(y)\|^2$, is L-Lipschitz for some $L \geq 0$. Further, assume for $\mathcal{X}$ and $\{c_t\}_{t\in\mathcal{T}}$ any of the settings from Proposition 21 and take $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ such that the support of μ or ν is the closure of a connected open set in $\mathbb{R}^d$. Then, for $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ and $\{Y_i\}_{i=1}^m \sim \nu^{\otimes m}$, respectively, with $n, m \to \infty$ such that $m/(n+m) \to \lambda \in (0,1)$, it holds that*

$$\sqrt{\frac{nm}{n+m}}\Big(\inf_{\tau\in\mathcal{T}} OT(\mu_n, \nu_m, c_\tau) - \inf_{\tau\in\mathcal{T}} OT(\mu, \nu, c_\tau)\Big) \rightsquigarrow \inf_{\tau\in S_-(\mathcal{T},\mu,\nu)} \sqrt{\lambda}\, \mathbb{G}^\mu(f_\tau^{c_\tau c_\tau}) + \sqrt{1-\lambda}\, \mathbb{G}^\nu(f_\tau^{c_\tau}),$$

where $f_\tau \in S_{c_\tau}(\mu,\nu)$ denotes a Kantorovich potential between μ, ν and cost c_τ for $\tau \in S_-(\mathcal{T},\mu,\nu)$.

The above result details that for settings of dimension $d \leq 3$ the OT value under embedded invariances can be well-approximated by replacing population measures by empirical counterparts. Insofar, our results mathematically justifies randomized computational schemes for (22) while providing statistical guarantees. Moreover, we like to point out that the assumption on the class of transformations $\mathcal{T}$ captures many compact finite dimensional classes (*e.g.*, orthogonal or bounded linear transformations) but also permits non-parametric ones (*e.g.*, via sufficiently regular autoencoders) as long as their metric entropy does not grow to large. As previously noted, one can relax the requirement that $\mathcal{T}$ is compact to the assumption that the sequence of estimated optimal transformation τ_n is contained within a compact set with probability tending to one (Proposition 8). In the setting of $\mathcal{T}$ consisting of diffeomorphisms we have by Lemma 1 of [8]

$$\inf_{\tau\in\mathcal{T}} OT\left(\mu, \nu, \|\cdot - \tau(\cdot)\|^2\right) = \inf_{\tau\in\mathcal{T}} OT\left(\mu, \tau_\# \nu, \|\cdot - \cdot\|^2\right) = \inf_{\tau\in\mathcal{T}} W_2^2\left(\mu, \tau_\#\nu\right),$$

for which convergence of empirical minimizers τ_n can be verified for various settings using results by Bernton et al. [68].

Remark 31 (*Wasserstein–Procrustes*)**.** The above proposition can be applied under mild regularity assumptions on the measures to the special orthogonal group $\mathcal{T} := \mathrm{SO}(d)$ for $d \leq 3$. Indeed, upon choosing $\mathcal{X}, \mathcal{Y}$ as compact, convex sets of $\mathbb{R}^d$ setting (iii) of Proposition 21 is fulfilled, asserting (Don). Moreover, if the support of μ or ν is the closure of a connected open set in $\mathbb{R}^d$, then (KP) holds and the distributional limits of Theorem 6 follow.

4.3. Sketched Wasserstein distance for mixture distributions

Recently, Bing et al. [69] and Delon and Desolneux [70] investigated a distance between (Gaussian) mixture distributions. These distributions are ubiquitous in statistics and machine learning, see McLachlan et al. [71] and the references therein. One way of understanding that distance is to start from the Wasserstein distance between discrete measures but instead of using a cost function between points, one replaces the points by distributions and one must thus choose a cost between distributions. Before formally defining that concept, recall that, for a set of distributions $\mathcal{A} := (\mathrm{A}_1, \dots, \mathrm{A}_K)$ of finite cardinality K, a mixture r is a convex combination of components from $\mathcal{A}$ given by a vector $\alpha \in \Delta_K$, *i.e.*, $r = \sum_{i=1}^K \alpha_i \mathrm{A}_i$, where Δ_K is the probability simplex in $\mathbb{R}^K$. Given a distance $d : \mathcal{A}\times\mathcal{A} \to \mathbb{R}_+$ between mixture components of $\mathcal{A}$, the aforementioned authors define the *sketched Wasserstein distance* between two mixture distributions with weights α and β as

$$W(\alpha,\beta,d) := \inf_{\pi\in\Pi(\alpha,\beta)} \sum_{k,\ell=1}^K \pi_{k,\ell}\, d(\mathrm{A}_k, \mathrm{A}_\ell),$$

where the infimum is taken over elements of the set of couplings

$$\Pi(\alpha,\beta) = \left\{ \pi \in \Delta_{K\times K} \;\middle|\; \begin{array}{ll} \sum_{\ell=1}^K \pi_{k,\ell} = \alpha_k, & \text{for all } k \in \{1,\dots,K\} \\ \sum_{k=1}^K \pi_{k,\ell} = \beta_\ell, & \text{for all } \ell \in \{1,\dots,K\} \end{array} \right\}.$$

Understanding the fluctuations of a plug-in estimator for this distance can be achieved using our proof technique for Theorem 2 and is formalized in the following proposition.

Proposition 32. *Let $(\alpha_n, \beta_n, d_n) \in \Delta_K \times \Delta_K \times \mathbb{R}_+^{K^2}$ be measurable estimators for α, β, d, respectively. Further, for a positive sequence $(a_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} a_n = \infty$, assume for $n\to\infty$ that*

$$a_n \begin{pmatrix} \alpha_n - \alpha \\ \beta_n - \beta \\ d_n - d \end{pmatrix} = a_n \begin{pmatrix} (\alpha_{n,k} - \alpha_k)_{k=1}^K \\ (\beta_{n,k} - \beta_k)_{k=1}^K \\ (d_n(\mathrm{A}_k, \mathrm{A}_\ell) - d(\mathrm{A}_k, \mathrm{A}_\ell))_{l,k=1}^K \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{G}^\alpha \\ \mathbb{G}^\beta \\ \mathbb{G}^d \end{pmatrix} \quad \text{in } \mathbb{R}^{2K+K^2}, \tag{23}$$

where $(\mathbb{G}^\alpha, \mathbb{G}^\beta, \mathbb{G}^d)$ represents a tight (possibly non-Gaussian) random variable on $\mathbb{R}^{2K+K^2}$. Then,

$$a_n\Big(W(\alpha_n,\beta_n,d_n) - W(\alpha,\beta,d)\Big) \rightsquigarrow \inf_{\pi\in\Pi_d^\star(\alpha,\beta)} \langle \pi, \mathbb{G}^d\rangle + \sup_{f\in S_d(\alpha,\beta)} \langle f^{dd}, \mathbb{G}^\alpha\rangle + \langle f^d, \mathbb{G}^\beta\rangle.$$

The proof follows along the same approach as for showing Theorem 2 and is therefore omitted, see Remark 4 (iii). In this context, the requirement of weak convergence for the measure estimators $(\alpha_n, \beta_n) \rightsquigarrow (\alpha, \beta)$ in probability follows from our assumption in (23) since the population measures and its estimators are supported on finitely many points.

For comparison, Bing et al. [69, Proposition 1] obtain a distributional limit in the case where the asymptotic fluctuation of the cost is negligible compared to that of the estimated measures. Their result is captured by Proposition 32, which in addition covers the setting where the cost is estimated on the same data and converges at the same rate. Insofar, our analysis more explicitly characterizes how the additional uncertainty caused by the cost estimation affects the asymptotic fluctuation of the estimated OT value.

Finally, we stress that Proposition 32 enables potentially statistical inference in the context of parameter estimation for finite mixture models, most notably Gaussian mixtures. Mathematically, the *mixing measure* is described by $\mu = \sum_{k=1}^K \alpha_k \delta_{\theta_k}$ with $\sum_{k=1}^K \alpha_k = 1$ and $\alpha_k \geq 0$ and $\theta_1, \ldots, \theta_K \in \mathbb{R}^d$ for all $k \in \{1, \ldots, K\}$ and the mixture distribution, which describes the distribution of the data, has a density on $\mathbb{R}^d$ given by $f(x \mid \mu) := \sum_{k=1}^K \alpha_k f(x, \theta_k)$ where $f(x, \theta)$ denotes a suitable family of probability densities. Within the last decade, the convergence rates of various parameter estimators $\tilde{\mu}_n$ for μ with respect to the Wasserstein distance have been analyzed, including maximum likelihood estimators [72–75], moment-based estimators [76,77] and Bayesian estimators [78,79]. As part of Proposition 32, to obtain a distributional limit for $W(\tilde{\mu}_n, \mu)$ where $\tilde{\mu}_n$ is a finitely supported measure estimator it is necessary to derive distributional limits for the mass assigned to each point of $\tilde{\mu}_n$ as well as the respective locations. So far, such an analysis is still open and would be an interest venue for future research.

4.4. Sliced optimal transport

Our theory from Section 2.3 also enables the analysis of sliced OT quantities and complement or extend available results from the literature [50,80–82]. In the following, we formalize this statement. For two Borel probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ the *average-sliced* and *max-sliced Wasserstein distances of order* $1 \leq p < \infty$ are defined, respectively, as

$$\underline{W}_p(\mu, \nu) := \left(\int_{\mathbb{S}^{d-1}} OT(\mathfrak{p}^\theta_\# \mu, \mathfrak{p}^\theta_\# \nu, |\cdot - \cdot|^p) \, d\sigma(\theta) \right)^{\frac{1}{p}} \quad \text{and} \quad \overline{W}_p(\mu, \nu) := \max_{\theta \in \mathbb{S}^{d-1}} \left(OT(\mathfrak{p}^\theta_\# \mu, \mathfrak{p}^\theta_\# \nu, |\cdot - \cdot|^p) \right)^{\frac{1}{p}},$$

where $\mathfrak{p}^\theta : \mathbb{R}^d \to \mathbb{R}$ is the projection map $x \mapsto \theta^T x$ and σ represents the uniform distribution on the unit sphere $\mathbb{S}^{d-1}$. Note by Lemma 1 in Nies et al. [8] for any $\theta \in \mathbb{S}^{d-1}$ that

$$OT(\mathfrak{p}^\theta_\# \mu, \mathfrak{p}^\theta_\# \nu, |\cdot - \cdot|^p) = OT(\mu, \nu, |\mathfrak{p}^\theta(\cdot) - \mathfrak{p}^\theta(\cdot)|^p),$$

which enables to view the sliced Wasserstein quantities in the framework of Section 2.3 and asserts by Theorems 5–7 the following result.

Corollary 33. *Let $p \geq 1$, $d \geq 2$, and define for $\theta \in \mathbb{S}^{d-1}$ the cost $c_\theta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $(x, y) \mapsto |\mathfrak{p}^\theta(\cdot) - \mathfrak{p}^\theta(\cdot)|^p$. Further, take compactly supported probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with empirical measures μ_n, ν_m, respectively. For all assertions, we let $n, m \to \infty$ with $m/(n+m) \to \lambda \in (0, 1)$.*

(i) Assume that the set of Kantorovich potentials $S_{c_\theta}(\mu, \nu)$ is unique (up to a constant shift) for any $\theta \in \mathbb{S}^{d-1}$. Then, it follows upon selecting $f_\theta \in S_{c_\theta}(\mu, \nu)$ for any $\theta \in \mathbb{S}^{d-1}$ that

$$\sqrt{\frac{nm}{n+m}} \Big(OT(\mu_n, \nu, c_\theta) - OT(\mu, \nu, c_\theta) \Big)_{\theta \in \mathbb{S}^{d-1}} \rightsquigarrow \left(\sqrt{\lambda} \mathbb{G}^\mu (f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda} \mathbb{G}^\nu (f_\theta^{c_\theta}) \right)_{\theta \in \mathbb{S}^{d-1}} \quad \text{in } C(\mathbb{S}^{d-1})$$

(ii) Assume the same as in (i). Then, it follows that

$$\sqrt{\frac{nm}{n+m}} \left(\underline{W}_p^p(\mu_n, \nu_m) - \underline{W}_p^p(\mu, \nu) \right) \rightsquigarrow \int_{\mathbb{S}^{d-1}} \sqrt{\lambda} \mathbb{G}^\mu (f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda} \mathbb{G}^\nu (f_\theta^{c_\theta}) \, d\theta.$$

(iii) Without imposing the assumption on uniqueness of Kantorovich potentials, it follows that

$$\sqrt{\frac{nm}{n+m}} \left(\overline{W}_p^p(\mu_n, \nu_m) - \overline{W}_p^p(\mu, \nu) \right) \rightsquigarrow \sup_{\substack{\theta \in S_+(\mathbb{S}^{d-1}, \mu, \nu) \\ f_\theta \in S_{c_\theta}(\mu, \nu)}} \sqrt{\lambda} \mathbb{G}^\mu (f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda} \mathbb{G}^\nu (f_\theta^{c_\theta}).$$

Comparing Corollary 33 to the literature for $p > 1$, results in Goldfeld et al. [80] and Xi and Niles-Weed [50] are recovered under slightly weaker assumptions. For the analysis of both types of empirical sliced Wasserstein distances Goldfeld et al. [80] require the underlying measures to have compact, convex support. Moreover, for the uniform central limit theorem by Xi and Niles-Weed [50] of the sliced OT process, they assume for each $u \in \mathbb{S}^{d-1}$ that one of the projected measures has compact, connected support. These conditions are sufficient for the uniqueness of Kantorovich potentials, but it can also be guaranteed for measures with disconnected support (cf. Proposition 23 and more generally Staudt et al. [62]). Corollary 33(ii) also complements results by Manole et al. [81] on the trimmed sliced Wasserstein distance as we do not require the existence of a density but the underlying measures to be compactly supported.

For the special case $p = 1$, unlike in our results, distributional limits by Goldfeld et al. [80],Xu and Huang [82] for the average- and max-sliced Wasserstein distance do not require uniqueness of the Kantorovich potentials. Further, their theory remains valid for non-compactly supported measures by imposing suitable moment-conditions. Crucial to their approach is the special characterization of the 1-Wasserstein distance as an integral probability metric over Lipschitz functions [3, Remark 6.5], a property which we do not exploit in our general theory. Still, under uniqueness of Kantorovich potentials, which occurs, *e.g.*, if one measure is discrete while the other has connected support and is absolutely continuous [62, Example 3], Corollary 33(i) asserts weak convergence for the sliced OT process in $C(\mathbb{S}^{d-1})$.

As potential extensions we like to mention that our theory can also naturally be adapted to more general projection operations, as exemplified in the context of generalized sliced OT [83], or when considering more non-Euclidean data but which involve a projection onto a one-dimensional domain, *e.g.*, for sliced OT on the sphere [84], on hyperbolic spaces [85], or on the space of symmetric positive definite matrices [86]. In all these cases our results provide statistical guarantees for randomized computation schemes of sliced OT quantities.

4.5. Stability analysis of optimal transport

In addition to statistical applications, our theory for the empirical OT value under weakly converging costs enables a deterministic stability analysis of the OT problem (1) under joint perturbations of the costs and the measures, which may be of independent interest, *e.g.*, from the viewpoint of optimization. More precisely, we prove in the following Gateaux differentiability of the OT value in $(\mu,\nu,c) \in \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C(\mathcal{X}\times\mathcal{Y})$ for all admissible directions. This extends previous a sensitivity result for the OT value which was limited to the setting of finitely supported measures, and as based on the theory of finite-dimensional linear programs [87, Theorem 3.1]. In particular, relevant to ongoing progress on the domain of identifying and learning appropriate cost functions in a data-driven manner [27,28], the following result provides an additional insight on the robustness of the OT value.

Proposition 34. *Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $c \in C(\mathcal{X}\times\mathcal{Y})$ be fixed. Define for $t>0$ sufficiently small the quantities $\mu_t = \mu + t\Delta^\mu$ and $\nu_t = \nu + t\Delta^\nu$, where $\Delta^\mu \in (\mathcal{P}(\mathcal{X})-\mu)$ and $\Delta^\nu \in (\mathcal{P}(\mathcal{Y})-\nu)$, respectively. Further, let $c_t = c + t\Delta^c$ for some $\Delta^c \in C(\mathcal{X}\times\mathcal{X})$. Then, it follows that*

$$\lim_{t\searrow 0}\frac{1}{t}\left(OT(\mu_t,\nu_t,c_t) - OT(\mu,\nu,c)\right) = \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\Delta^c) + \sup_{f\in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c).$$

Remark 35 (*On Hadamard directional differentiability*)**.** Since the set of admissible directions $(\mathcal{P}(\mathcal{X})-\mu)\times(\mathcal{P}(\mathcal{Y})-\nu)\times C(\mathcal{X}\times\mathcal{Y})$ is not a normed vector space, we are in general unable to infer Hadamard directional differentiability by additionally proving Lipschitzianity of the OT problem with respect to the measures μ,ν and the cost function c.

Invoking the same proof strategy as in Proposition 34 would require us to show for any $t_n \searrow 0$ and any sequence of measures and cost $(\mu_n,\nu_n,c_n) = (\mu + t_n\Delta_n^\mu, \nu + t_n\Delta_n^\nu, c + t_n\Delta_n^c) \in \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C(\mathcal{X}\times\mathcal{Y})$ with $(\Delta_n^\mu,\Delta_n^\nu,\Delta_n^c) \to (\Delta^\mu,\Delta^\nu,\Delta^c)$ in $\ell^\infty(\mathcal{F}^{cc})\times\ell^\infty(\mathcal{F}^c)\times C(\mathcal{X}\times\mathcal{Y})$ that

$$\sup_{f\in\mathcal{F}}\left|\Delta_n^\mu(f^{c_nc_n} - f^{cc}) + \Delta_n^\nu(f^{c_n} - f^c)\right| \to 0. \tag{24}$$

Showing this remains a challenge and would enable us to omit conditions (Sup) and (Sup)* in the formulations of Theorem 2 and Proposition 10, respectively. Another challenge in such an attempt is that any such sequence (μ_n,ν_n) does not necessarily converge weakly for $n\to\infty$ to (μ,ν), which is relevant for our proof, since the topology induced by $\ell^\infty(\mathcal{F}^{cc})\times\ell^\infty(\mathcal{F}^c)$ may be too weak.

Though it is likely possible to show Hadamard directional differentiability of the OT problem jointly in the measures and the cost by selecting a sufficiently strong norm that metrizes weak convergence of measures, the functional delta method would inevitably require the empirical process to weakly converge in this norm and impose additional conditions. A similar trade-off for the choice of the norm is natural and known in the literature (cf. Dudley [88, p.76] and Jourdain and Tse [89]).

5. Regularity elevation functionals

In this section, we construct regularity elevation maps, *i.e.*, continuous maps $\Psi : C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$ such that for measurable estimators c_n with $\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c$ for $n\to\infty$, it follows that

$$(i)\ \sqrt{n}\big(c_n - \Psi(c_n)\big) \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad (ii)\ \Psi(c_n) \text{ fulfills certain regularity properties.} \tag{25}$$

Based on Lipschitzianity of the OT value with respect to the cost function (Lemma 43), condition (i) allows us to substitute a cost estimator with one that enjoys certain regularity properties, effectively "elevating" its level of regularity. Such maps prove useful in our work at two particular instances. For one, it enables us to assume in the proof of Theorem 2 that cost estimators are suitably bounded and exhibit the same modulus of continuity as the population cost function (cf. Corollary 39). This represents an important step to rely on Lemma 1. Moreover, the notion of regularity elevations also represents a useful tool to prove Corollary 20, for which we employ Proposition 18 and set $\tilde{c}_n := \Psi(c_n)$ for a suitable regularity elevation map. Insofar, these maps serve as an effective tool for the theoretical analysis of distributional limits.

The subsequent result provides a first set of conditions to ensure condition (i) of (25). Its proof as well as the proof of all subsequent results of this section are detailed in Appendix D.

Proposition 36. *Let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and let $c_n \in C(\mathcal{X}\times\mathcal{Y})$ be a (Borel measurable) random sequence such that $a_n(c_n - c) \rightsquigarrow \mathcal{L}$ in $C(\mathcal{X}\times\mathcal{Y})$ for some $c \in C(\mathcal{X}\times\mathcal{Y})$ and $(a_n)_{n\in\mathbb{N}}$ such that $a_n \to \infty$ for $n \to \infty$. Let $U \subseteq C(\mathcal{X}\times\mathcal{Y})$ be a linear subspace such that $\mathcal{L}$ is a.s. contained in U. Then, if $\Psi : C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$ is continuous near c, Hadamard directionally differentiable at f with a derivative such that $D^H_c\Psi|_U = \mathrm{Id}_U$ and $\Psi(c) = c$, it follows for $n \to \infty$ that*

$$a_n\big(c_n - \Psi(c_n)\big) \xrightarrow{\mathrm{P}} 0 \quad \text{for } n \to \infty.$$

Notably, in case Ψ is Hadamard differentiable with $D^H_f\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$, one may select $U = C(\mathcal{X}\times\mathcal{Y})$ and the condition on the limit $\mathcal{L}$ becomes vacuous.

To conclude various types of useful regularity properties, as required in (ii) of (25), we thus define in the following subsections various maps such that the conditions of Proposition 36 are met. Additionally, we provide suitable metric entropy bounds for $\mathcal{F}^{\Psi(\tilde{c})\Psi(\tilde{c})}$ independent of $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$.

5.1. Regularity elevation to deterministic boundedness

Consider compact Polish spaces $\mathcal{X}, \mathcal{Y}$ and let $c \in C(\mathcal{X}\times\mathcal{Y})$ be a continuous cost function such that $\|c\|_\infty \le 1$. We define the regularity elevation functional for boundedness as

$$\Psi_{\mathrm{bdd}} : C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y}), \quad \tilde{c} \mapsto \Big((x,y) \mapsto \max(\min(\tilde{c}(x,y), 2), -2)\Big).$$

Proposition 37. *For the above setting, $\Psi = \Psi_{\mathrm{bdd}}$ fulfills $\Psi(c) = c$, is continuous, and it is Hadamard differentiable at c with $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$. In particular, if $\mathcal{X}$ is a finite space, we obtain for any uniformly bounded function class $\mathcal{G}$ on $\mathcal{Y}$ that*

$$\sup_{\tilde{c}\in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim |\log(\varepsilon)|.$$

Hence, for our analysis of the empirical OT value under estimated cost functions we can assume without loss of generality that cost estimators are deterministically bounded by a constant that depends on the population cost. In the following we prove a similar insight for the modulus of continuity for cost estimators on compact (pseudo-)metric spaces.

5.2. Regularity elevation to concave modulus of continuity and lipschitzianity

Consider compact Polish spaces $\mathcal{X}, \mathcal{Y}$ and let $\tilde{d}_{\mathcal{X}}$ be a continuous (pseudo-) metric on $\mathcal{X}$. Denote by $\tilde{\mathcal{X}}$ the space $\mathcal{X}$ equipped with the topology induced by $\tilde{d}_{\mathcal{X}}$ which is also compact (Lemma 57) but potentially does not satisfy the Hausdorff property. Let $c \in C(\tilde{\mathcal{X}}\times\mathcal{Y})$ be a cost function such that $\|c\|_\infty \le 1$ and consider a concave modulus $w : \mathbb{R}_+ \to \mathbb{R}_+$ with $w(\delta) > 0$ for $\delta > 0$ such that

$$|c(x,y) - c(x',y)| \le w(\tilde{d}_{\mathcal{X}}(x,x')) \quad \text{for any } x, x' \in \tilde{\mathcal{X}}, y \in \mathcal{Y}. \tag{26}$$

If $c(\cdot,y)$ is 1-Lipschitz under $\tilde{d}_{\mathcal{X}}$, then select $w(t) := t$ and if $c(\cdot,y)$ is $(\gamma,1)$-Hölder for $\gamma \in (0,1]$ (recall footnote (iii)), choose $w(t) := t^\gamma$. The regularity elevation functional for $w\circ d_{\mathcal{X}}$ is then given by

$$\Psi^{w\circ\tilde{d}_{\mathcal{X}}}_{\mathrm{mod}} : C(\mathcal{X}\times\mathcal{Y}) \to C(\tilde{\mathcal{X}}\times\mathcal{Y}), \quad \tilde{c} \mapsto \left((x,y) \mapsto \inf_{x'\in\mathcal{X}} \tilde{c}(x',y) + 2w(\tilde{d}_{\mathcal{X}}(x,x'))\right)$$

Proposition 38. *For the above setting, $\Psi = \Psi^{w\circ\tilde{d}_{\mathcal{X}}}_{\mathrm{mod}} \circ \Psi_{\mathrm{bdd}}$ fulfills $\Psi(c) = c$, it is continuous near c, and it is Hadamard directionally differentiable at c with $D^H_{|c}\Psi|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$. Further, for any uniformly bounded function class $\mathcal{G}$ on $\mathcal{Y}$ it holds that*

$$\sup_{\tilde{c}\in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim \mathcal{N}(\varepsilon/8, \mathcal{X}, w\circ\tilde{d}_{\mathcal{X}})|\log(\varepsilon)|.$$

An appealing consequence of the above considerations is that they allow us to construct a regularity elevated estimator $\tilde{c}_{n,m}$ from $c_{n,m}$ such that $\mathcal{H}_{\tilde{c}_{n,m}} \subseteq \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}$, for $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ defined in (9), holds deterministically.

Corollary 39. *Let $c \in C(\mathcal{X}\times\mathcal{Y})$, set $B := \|c\|_\infty + 1/2$ and let $w : \mathbb{R}_+ \to \mathbb{R}_+$ be a concave modulus with $w(\delta) > 0$ for $\delta > 0$ such that (26) holds for a metric $d_{\mathcal{X}}$ on $\mathcal{X}$. Assume for a random sequence $c_n \in C(\mathcal{X}\times\mathcal{Y})$ that $a_n(c_n - c) \rightsquigarrow \mathbb{G}^c$ in $C(\mathcal{X}\times\mathcal{Y})$ with $a_n \to \infty$. Then, the random sequence*

$$\overline{c}_n := B \cdot \Psi^{w\circ d_{\mathcal{X}}/B}_{\mathrm{mod}} \circ \Psi_{\mathrm{bdd}}(c_n/B) \in C(\mathcal{X}\times\mathcal{Y})$$

satisfies $a_n\|\overline{c}_n - c_n\|_\infty \xrightarrow{\mathrm{P}} 0$ for $n \to \infty$ and deterministically fulfills $\|\overline{c}_n\|_\infty \le 2B = 2\|c\|_\infty + 1$, relation (26), and the inclusion $\mathcal{H}_{\overline{c}_n} \subseteq \mathcal{F}^{\overline{c}_n\overline{c}_n}(2\|c\|_\infty + 1, 2w)$.

5.3. Regularity elevation to Hölder functions of order $\gamma \in (1,2]$

Since we are able to leverage for convergence rates of the empirical OT value (recall Proposition 13(i)) the regularity of the underlying cost function up to Hölder degree $\gamma \le 2$, we provide in this subsection a corresponding regularity elevation map. As the setting for $\gamma \le 1$ can be treated using the theory from previous subsection, we only focus on the regime of $\gamma \in (1,2]$.

Consider a convex, compact set $\mathcal{X} \subseteq \mathbb{R}^d$ with non-empty interior. Let $c \in C(\mathcal{X} \times \mathcal{Y})$ be a cost function such that $\|c\|_\infty \leq 1$ and assume c is continuously differentiable in x, *i.e.*, suppose that $\nabla_x c : \mathrm{int}(\mathcal{X}) \times \mathcal{Y} \to \mathbb{R}^d$ can be continuously extended to $\mathcal{X} \times \mathcal{Y}$. Further, suppose that $c(\cdot, y)$ is $(\gamma, 1)$-Hölder for each $y \in \mathcal{Y}$ for $\gamma \in (1, 2]$. We define the regularity elevation map for Hölder functions of order $\gamma \in (1, 2]$ by

$$\Psi^{c,\gamma}_{\mathrm{Hol}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}), \tilde{c} \mapsto \left((x, y) \mapsto \inf_{x' \in \mathcal{X}} \tilde{c}(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + 2\sqrt{d}\, \|x - x'\|^\gamma \right)$$

Notably, it is crucial that the scalar product term involves the partial derivative of the respective (population) cost function c. Moreover, we like to point out that the image under $\Psi^{c,\gamma}_{\mathrm{Hol}}$ does not necessarily lead to $(\gamma, 1)$-Hölder functions but nonetheless ensures suitable metric entropy bounds.

Proposition 40. *For the above setting with $\mathcal{X} \subseteq \mathbb{R}^d$ convex and compact, $\Psi = \Psi^{c,\gamma}_{\mathrm{Hol}} \circ \Psi_{\mathrm{bdd}}$ fulfills $\Psi(c) = c$, it is continuous near c, and it is Hadamard differentiable at c with $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$. Further, for any uniformly bounded function class $\mathcal{G}$ on $\mathcal{Y}$ we obtain that*

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim \varepsilon^{-d/\gamma}.$$

5.4. Combination of regularity elevations

Finally, we also outline a constructive way to combine regularity elevation maps defined on different spaces. This is important since it enables to leverage regularity properties of the population cost function for different regions of the domain.

Hence, let $\mathcal{X}, \mathcal{Y}$ be compact Polish spaces and assume existence of a collection of homeomorphisms $\zeta_i : \mathcal{U}_i \to \zeta_i(\mathcal{U}_i)$ for $1 \leq i \leq I$ such that $\mathcal{X} = \bigcup_{i=1}^I \zeta_i(\mathcal{U}_i)$. Further, assume there exists a partition on unity $\{\eta_i\}_{i=1}^I$ on $\mathcal{X}$ with $\mathrm{supp}(\eta_i) \subseteq \zeta_i(\mathcal{U}_i)$. Consider a continuous cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and let $c_i : \mathcal{U}_i \times \mathcal{Y} \to \mathbb{R}, (u, y) \mapsto c(\zeta_i(u), y)$. Assume there exist maps $\Psi_i : C(\mathcal{U}_i \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y})$ such that $\Psi_i(c_i) = c_i$ and where Ψ_i is continuous near c_i and Hadamard differentiable at c_i with derivative $D^H_{|c_i}\Psi_i = \mathrm{Id}$. Using these maps we define the combination of regularity elevations as

$$\Psi_{\mathrm{com}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}), \quad \tilde{c} \mapsto \left((x, y) \mapsto \sum_{i=1}^I \eta_i(x) \Psi_i\Big(\tilde{c}(\zeta_i(\cdot), \cdot)\Big)(\zeta_i^{-1}(x), y) \right).$$

Indeed, by continuity of the partition of one η_i as well as the functionals Ψ_i and ζ_i, ζ_i^{-1} for each $i \in \{1, \dots, I\}$ it follows that the range of this functional is indeed contained in $C(\mathcal{X} \times \mathcal{Y})$.

Proposition 41. *For the above setting, $\Psi = \Psi_{\mathrm{com}}$ fulfills $\Psi(c) = c$, it is continuous near c, and it is Hadamard differentiable at c with $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$. Further, for any uniformly bounded function class $\mathcal{G}$ on $\mathcal{Y}$ we obtain*

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \leq \sum_{i=1}^I \sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c}(\zeta_i(\cdot),\cdot))}, \|\cdot\|_\infty).$$

6. Proofs of main results

In this section, we provide the full proofs of Lemma 1 for the dual representation of the OT value, Theorem 2 and Proposition 10 for the distributional limit of the empirical OT value under weakly converging costs, as well as Theorems 5–7 and Proposition 8 for empirical OT with extremal-type costs. Proofs for all auxiliary results of this section are deferred to Appendix E.

6.1. Proof of Lemma 1: Dual representation of optimal transport value

The subsequent auxiliary lemma establishes an important property of cost-transformations which is essential throughout this section.

Lemma 42 (*Lipschitz Property of Cost-Transformation*). *For arbitrary functions $f, \tilde{f} : \mathcal{X} \to \mathbb{R}$ and cost functions $c, \tilde{c} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, it follows that $\|f^c - \tilde{f}^{\tilde{c}}\|_\infty \leq \|f - \tilde{f}\|_\infty + \|c - \tilde{c}\|_\infty$. In particular, upon selecting the constant functions $\tilde{f}, \tilde{c} \equiv 0$, it follows that $\|f^c\|_\infty \leq \|f\|_\infty + \|c\|_\infty$.*

Proof of Lemma 1. For any $h \in \mathcal{H}_c$ there exists $g : \mathcal{Y} \to [-\|c\|_\infty, \|c\|_\infty]$ with $h = g^c$, and hence

$$-\|c\|_\infty - \sup_{y \in \mathcal{Y}} g(y) \leq h(x) = \inf_{y \in \mathcal{Y}} c(x, y) - g(y) \leq \|c\|_\infty - \sup_{y \in \mathcal{Y}} g(y).$$

In consequence, we find that $\|h\|_\infty \leq 2\|c\|_\infty \leq 2B$. Further, for arbitrary $x, x' \in \mathcal{X}$ and $\varepsilon > 0$, consider $y' \in \mathcal{Y}$ such that $h(x') \geq c(x', y') - g(y') - \varepsilon$. Then, it follows that

$$\begin{aligned} h(x) - h(x') &= \left[\inf_{y \in \mathcal{Y}} c(x, y) - g(y) \right] - \left[\inf_{y \in \mathcal{Y}} c(x', y) - g(y) \right] \\ &\leq c(x, y') - g(y') - c(x', y') + g(y') + \varepsilon \end{aligned}$$

$$\leq w(d_{\mathcal{X}}(x,x')) + \varepsilon.$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, we obtain that $|h(x) - h(x')| \leq w(d_{\mathcal{X}}(x,x'))$. This yields $\mathcal{H}_c \subseteq \mathcal{F}$ and thus $\mathcal{H}_c^c \subseteq \mathcal{F}^c$. Further, by Santambrogio [2, Proposition 1.34] we infer $\mathcal{H}_c = \mathcal{H}_c^{cc} \subseteq \mathcal{F}^{cc}$. To show the remaining inclusions note for $f \in \mathcal{F}$ that

$$-\|c\|_\infty - \sup_{x\in\mathcal{X}} f(x) \leq f^c \leq \|c\|_\infty - \sup_{x\in\mathcal{X}} f(x).$$

Hence, the function $g := f^c + \sup_{x\in\mathcal{X}} f(x)$ fulfills $\|g\|_\infty \leq \|c\|_\infty$, and since $\|f\|_\infty \leq 2B$, we find that

$$f^{cc}(x) = (f^c)^c = (g)^c + \sup_{x\in\mathcal{X}} f(x) \in \mathcal{H}_c + [-2B, 2B],$$

which yields $\mathcal{F}^{cc} \subseteq \mathcal{H}_c + [-2B,2B]$ as well as $\mathcal{F}^c = \mathcal{F}^{ccc} \subseteq \mathcal{H}_c^c + [-2B,2B]$. To show representation (10), we combine the inclusions $\mathcal{H}_c \subseteq \mathcal{F} \subseteq C(\mathcal{X})$ with the alternative dual representations (2) and (8). For the final claim, take a maximizing sequence $\{f_n\}_{n\in\mathbb{N}}$ for (10) which admits by compactness of $\mathcal{F}$ a converging subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ with uniform limit $f \in \mathcal{F}$. Then by Lemma 42 it follows that $\{f_{n_k}^c\}_{k\in\mathbb{N}}$ and $\{f_{n_k}^{cc}\}_{k\in\mathbb{N}}$ also uniformly converge to f^c and f^{cc}, respectively. We thus obtain that

$$\mu(f^{cc}) + \nu(f^c) = \lim_{k\to\infty} \mu(f_{n_k}^{cc}) + \nu(f_{n_k}^c) = OT(\mu,\nu,c)$$

which shows that $f \in \mathcal{F}$ is a maximizing element hence the set of optimizers $S_c(\mu,\nu)$ for (10) is non-empty. $\square$

6.2. *Proofs for distributional limits under weakly converging costs*

6.2.1. *Proof of Theorem 2*

For the proof of Theorem 2 the following auxiliary results are crucial. We start with lower and upper bound on the difference between OT values for varying costs and probability measures which are a consequence of the OT problem having a representation in terms of an infimum over feasible couplings as well as a supremum over feasible potentials.

Lemma 43 (*Lower and upper bounds*). *Define for $B > 0$ and a concave modulus of continuity $w : \mathbb{R}_+ \to \mathbb{R}_+$ the collection*

$$C(B,w) := \left\{\tilde c \in C(\mathcal{X}\times\mathcal{Y}) \mid \|\tilde c\|_\infty \leq B, |\tilde c(x,y) - \tilde c(x',y)| \leq w(d_{\mathcal{X}}(x,x')) \text{ for all } x,x' \in \mathcal{X}, y \in \mathcal{Y}\right\}.$$

Then, for costs $c, \tilde c \in C(B,w)$ and probability measures $\mu, \tilde\mu \in \mathcal{P}(\mathcal{X}), \nu, \tilde\nu \in \mathcal{P}(\mathcal{Y})$ it holds that

$$\begin{aligned}
&\inf_{\pi\in\Pi^\star_{\tilde c}(\tilde\mu,\tilde\nu)} \pi(\tilde c - c) + \sup_{f\in S_c(\mu,\nu)} (\tilde\mu-\mu)f^{cc} + (\tilde\nu-\nu)f^c\\
&\leq\; OT(\tilde\mu,\tilde\nu,\tilde c) - OT(\mu,\nu,c)\\
&\leq \min\Bigg(\inf_{\pi\in\Pi^\star_{c}(\tilde\mu,\tilde\nu)} \pi(\tilde c - c) + \sup_{f\in S_c(\tilde\mu,\tilde\nu)} (\tilde\mu-\mu)f^{cc} + (\tilde\nu-\nu)f^c,\\
&\qquad \inf_{\pi\in\Pi^\star_{c}(\mu,\nu)} \pi(\tilde c - c) + \sup_{f\in S_{\tilde c}(\tilde\mu,\tilde\nu)} (\tilde\mu-\mu)f^{cc} + (\tilde\nu-\nu)f^c + \sup_{f\in\mathcal{F}}(\tilde\mu-\mu)(f^{\tilde c\tilde c} - f^{cc}) + (\tilde\nu-\nu)(f^{\tilde c} - f^c).\Bigg)
\end{aligned}$$

In particular, for fixed measures or fixed costs it follows that

$$|OT(\mu,\nu,\tilde c) - OT(\mu,\nu,c)| \leq \|\tilde c - c\|_\infty,\quad |OT(\tilde\mu,\tilde\nu,c) - OT(\mu,\nu,c)| \leq \sup_{f\in\mathcal{F}^{cc}} \left|(\tilde\mu-\mu)f\right| + \sup_{f\in\mathcal{F}^{c}} \left|(\tilde\nu-\nu)f\right|.$$

To employ the lower and upper bounds of Lemma 43 for the proof of Theorem 2 we additionally require a number of continuity and measurability properties which are captured in the following lemma. Notably, we equip $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ with the bounded Lipschitz norm, which turns it into a Polish metric space and metrizes weak convergence of measures.

Lemma 44 (*Continuity and measurability*).*Let $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$, and $c \in C(\mathcal{X}\times\mathcal{Y})$. Take a concave modulus of continuity $w : \mathbb{R}_+ \to \mathbb{R}_+$ for c and set $C := C(2\|c\|_\infty + 1, 2w)$ (for the definition of $C(2\|c\|_\infty + 1, 2w)$ see Lemma 43). Further, recall the function class $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ introduced in (9) and define the functions*

$$\begin{aligned}
&T_1 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C \to \mathbb{R}, && (\mu',\nu',c') \mapsto OT(\mu',\nu',c'),\\
&T_2 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C\times C(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}, && (\mu',\nu',c',h_c) \mapsto \inf_{\pi\in\Pi^\star_{c'}(\mu',\nu')} \pi(h_c),\\
&T_3 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C\times C_u(\mathcal{F})^2 \to \mathbb{R}, && (\mu',\nu',c',h_\mu,h_\nu) \mapsto \sup_{f\in S_{c'}(\mu',\nu')} h_\mu(f) + h_\nu(f),\\
&T_4 : C_u(\mathcal{F})^4 \to \mathbb{R}, && (h_\mu,\tilde h_\mu,h_\nu,\tilde h_\nu) \mapsto \sup_{f\in\mathcal{F}} h_\mu(f) - \tilde h_\mu(f) + h_\nu(f) - \tilde h_\nu(f).
\end{aligned}$$

Then, T_1 and T_4 are continuous, T_2 is lower semi-continuous, and T_3 is upper semi-continuous. If $\Pi^\star_{c'}(\mu',\nu')$ is unique, T_2 is continuous at (μ',ν',c',h_c). Moreover, for fixed (μ',ν',c') the map T_2 is continuous in h_c while T_3 is continuous in (h_μ,h_ν). In particular, each function T_i for $1 \leq i \leq 4$ is Borel measurable.

The previous two assertions fully deal with *deterministic* statements on the OT functional and related terms that arise from corresponding bounds. The following two results provide the relevant tools to control the stochastic aspects. More precisely, for our proof of Theorem 2 we consider a Skorokhod representation of the random sequence detailed in (JW) which additionally fulfills the property that μ_n and ν_n weakly converge to μ and ν, respectively. For this purpose, we state the following measurability assertions and joint weak convergence statements.

Lemma 45 (*Measurability of empirical process*)**.** *For a Polish space $\mathcal{X}$ consider a totally bounded function class $\mathcal{G} \subseteq C(\mathcal{X})$ under uniform norm. Then, the following assertions hold.*

(i) Any probability measure $\mu \in \mathcal{P}(\mathcal{X})$ defines via evaluation a uniformly continuous functional on $\mathcal{G}$, i.e., $\mu \in C_u(\mathcal{G})$.

(ii) A map $\omega \mapsto \mu(\omega) \in \mathcal{P}(\mathcal{X}) \subseteq C_u(\mathcal{G})$ is Borel measurable if and only if for any $g \in \mathcal{G}$ the evaluation map $\omega \mapsto \mu(\omega)(g)$ is Borel measurable.

(iii) The empirical process $\sqrt{n}(\mu_n - \mu)$ and the bootstrap empirical process $\sqrt{k}(\mu^b_{n,k} - \mu_n)$ are both Borel measurable random variables in $C_u(\mathcal{G})$.

Lemma 46 (*Joint Weak Convergence*)**.** *For the setting of Theorem 2, assume (JW). Then, for $n, m \to \infty$, weak convergence in the Polish space $C_u(\mathcal{F})^2 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ to a tight limit occurs*

$$\left(\left(\mathbb{G}^\mu_n(f^{cc}), \mathbb{G}^\nu_m(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c_{n,m}, \mu_n, \nu_m\right) \rightsquigarrow \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu\right). \tag{27}$$

If (Sup) of Theorem 2 is also valid, then, for $n, m \to \infty$, it follows in the Polish space $C_u(\mathcal{F})^4 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ that

$$\left(\left(\mathbb{G}^\mu_n(f^{cc}), \mathbb{G}^\mu_n(f^{c_{n,m}c_{n,m}}), \mathbb{G}^\nu_m(f^c), \mathbb{G}^\nu_m(f^{c_{n,m}})\right)_{f\in\mathcal{F}}, \mathbb{G}^c_{n,m}, \mu_n, \nu_n, \right) \rightsquigarrow \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu\right). \tag{28}$$

Each sequence element for (27) and (28) as well as the weak limit are Borel measurable.

Remark 47 (*Skorokhod Representation*)**.** When dealing with weak convergence of empirical processes in non-separable spaces, special care is required due to potential measurability issues. However, since the different maps of interest are defined between Polish spaces and measurable, we circumvent such issues. In particular, since the random variables from Lemma 46 converge weakly to a tight limit with separable support, the conditions of Billingsley [90, Theorem 6.7] are met and a (measurable) Skorokhod representation exists.

With these tools at our disposal, we now proceed with the proof of Theorem 2.

Proof of Theorem 2. Invoking Corollary 39, as $\sqrt{nm/(n+m)}(c_{n,m} - c) =: \mathbb{G}^c_{n,m} \rightsquigarrow \mathbb{G}^c$ in the space $C(\mathcal{X} \times \mathcal{Y})$, there exists $\overline{c}_{n,m}$ such that the inclusion $\mathcal{H}_{\overline{c}_{n,m}} \subseteq \mathcal{F}^{\overline{c}_{n,m}\overline{c}_{n,m}}$ (recall the function classes from Section 2.1) holds deterministically for $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ and $\sqrt{n}(c_{n,m} - \overline{c}_{n,m}) \xrightarrow{\mathrm{P}} 0$. The latter implies by Lemma 43 that

$$\sqrt{\frac{nm}{n+m}} \left(OT(\mu_n, \nu_m, \overline{c}_{n,m}) - OT(\mu_n, \nu_m, c_{n,m})\right) \xrightarrow{\mathrm{P}} 0.$$

Hence, to prove the assertion it suffices by Slutzky's lemma to show that

$$\sqrt{\frac{nm}{n+m}}\left(OT(\mu_n, \nu_m, \overline{c}_{n,m}) - OT(\mu, \nu, c)\right) \rightsquigarrow \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c). \tag{29}$$

Without loss of generality, we may therefore assume $c_{n,m} = \overline{c}_{n,m}$. Further, set $\lambda_n := m/(n+m)$. Then, by Lemma 43, the subsequent lower and upper bounds follow,

$$\begin{aligned}
&\inf_{\pi\in\Pi^\star_{c_{n,m}}(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\, \mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\, \mathbb{G}^\nu_m(f^c)\\
&\leq \sqrt{\frac{nm}{n+m}}(OT(\mu_n, \nu_m, c_{n,m}) - OT(\mu, \nu, c))\\
&\leq \min\Bigg(\inf_{\pi\in\Pi^\star_c(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu_n,\nu_m)} \sqrt{\lambda_n}\, \mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\, \mathbb{G}^\nu_m(f^c),\\
&\qquad \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_{c_{n,m}}(\mu_n,\nu_m)} \sqrt{\lambda_n}\, \mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\, \mathbb{G}^\nu_m(f^c)\\
&\qquad + \sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\, \left(\mathbb{G}^\mu_n(f^{c_{n,m}c_{n,m}}) - \mathbb{G}^\mu_n(f^{cc})\right) + \sqrt{1-\lambda_n}\, \left(\mathbb{G}^\nu_m(f^{c_{n,m}}) - \mathbb{G}^\nu_m(f^c)\right)\Bigg).
\end{aligned} \tag{30}$$

For each setting (OP) and (Sup) we show that the upper and lower bounds asymptotically converge in distribution to the limit in (29), which then asserts that the empirical OT value also tends to this limit. To this end, we take for the random variables of Lemma 46 a Skorokhod representation on a probability space $(\Omega, \mathcal{A}, P)$ [90, p. 70] which is well-defined by Remark 47. More precisely, under (OP) we take the Skorokhod representation such that

$$\left(\left(\tilde{\mathbb{G}}_n^{\mu}(f^{cc}), \tilde{\mathbb{G}}_m^{\nu}(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}_{n,m}^c, \tilde{\mu}_n, \tilde{\nu}_m\right) \xrightarrow{\text{a.s.}} \left(\left(\tilde{\mathbb{G}}^{\mu}(f^{cc}), \tilde{\mathbb{G}}^{\nu}(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c, \mu, \nu\right) \tag{31}$$

in $C_u(\mathcal{F})^2 \times C(\mathcal{X}\times\mathcal{Y})\times\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$, whereas under (Sup) we choose it such that

$$\left(\left(\tilde{\mathbb{G}}_n^{\mu}(f^{cc}), \tilde{\mathbb{G}}_n^{\mu}(f^{\tilde{c}_{n,m}\tilde{c}_{n,m}}), \tilde{\mathbb{G}}_m^{\nu}(f^c), \tilde{\mathbb{G}}_m^{\nu}(f^{\tilde{c}_n})\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}_{n,m}^c, \tilde{\mu}_n, \tilde{\nu}_m\right) \xrightarrow{\text{a.s.}} \left(\left(\tilde{\mathbb{G}}^{\mu}(f^{cc}), \tilde{\mathbb{G}}^{\mu}(f^{cc}), \tilde{\mathbb{G}}^{\nu}(f^c), \tilde{\mathbb{G}}^{\nu}(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c, \mu, \nu\right) \tag{32}$$

in $C_u(\mathcal{F})^4 \times C(\mathcal{X}\times\mathcal{Y})\times\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$. We also set $\tilde{c}_{n,m} := c + \tilde{\mathbb{G}}_{n,m}^c/\sqrt{nm/(n+m)}$ which a.s. converges to c.

For the subsequent argument recall the functions T_1, T_2, T_3, T_4 from Lemma 44 and their (semi-) continuity properties. Furthermore, note that an application of Lemma 42 in combination with the arguments of the proof of Lemma 45 (i) yields that the maps $\mathcal{F}\to\mathbb{R}$,

$$f\mapsto \mathbb{G}_n^{\mu}(f^{cc}), \quad f\mapsto \mathbb{G}_n^{\mu}(f^{c_{n,m}c_{n,m}}), \quad f\mapsto \mathbb{G}_m^{\nu}(f^c), \quad f\mapsto \mathbb{G}_m^{\nu}(f^{c_{n,m}})$$

are uniformly continuous, *i.e.*, elements in $C_u(\mathcal{F})$. For both settings (OP) and (Sup) it follows by measurability of the maps T_1, T_2, T_3 for each $n, m\in\mathbb{N}$ that

$$\sqrt{\frac{nm}{n+m}}(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)) \overset{d}{=} \sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c)),$$

$$\inf_{\pi\in\Pi^\star_{c_{n,m}}(\mu_n,\nu_m)} \pi(\mathbb{G}_{n,m}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\mathbb{G}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}_m^{\nu}(f^c) \overset{d}{=} \inf_{\pi\in\Pi^\star_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c).$$

Under **(OP)** we also notice that

$$\inf_{\pi\in\Pi^\star_c(\mu_n,\nu_m)} \pi(\mathbb{G}_{n,m}^c) + \sup_{f\in S_c(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}_m^{\nu}(f^c) \overset{d}{=} \inf_{\pi\in\Pi^\star_c(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c),$$

whereas under (Sup) we additionally employ measurability of T_4 to infer for each $n\in\mathbb{N}$ that

$$\begin{aligned}
&\inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\mathbb{G}_{n,m}^c) + \sup_{f\in S_{c_{n,m}}(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}_m^{\nu}(f^c)\\
&+\sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\mathbb{G}_n^{\mu}(f^{cc}) - \mathbb{G}_n^{\mu}(f^{c_{n,m}c_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\mathbb{G}_m^{\nu}(f^c) - \mathbb{G}_m^{\nu}(f^{c_{n,m}})\right)\\
&\qquad\overset{d}{=} \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c)\\
&\qquad+\sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) - \tilde{\mathbb{G}}_n^{\mu}(f^{\tilde{c}_{n,m}\tilde{c}_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\tilde{\mathbb{G}}_m^{\nu}(f^c) - \tilde{\mathbb{G}}_m^{\nu}(f^{\tilde{c}_{n,m}})\right).
\end{aligned}$$

Hence, it suffices to work with the Skorokhod representation to obtain the weak limit for the empirical OT value. Invoking Lemma 43, identical lower and upper bounds on the quantity of interest, $\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c))$, as for (30) can be concluded.

To obtain a suitable bound on the limit inferior of $\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c))$ take for both (OP) and (Sup) a measurable set $A\in\mathcal{A}$ of full measure such that the convergence from (31) and (32) is fulfilled thereon, respectively. Then, for each $\omega\in A$ it follows by lower semi-continuity of T_2 jointly with continuity of T_3 under fixed (μ,ν,c) that

$$\begin{aligned}
&\liminf_{n,m\to\infty} \inf_{\pi\in\Pi^\star_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c)\\
&\quad\geq \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^c).
\end{aligned}$$

Under (OP), we find for each $\omega\in A$ by continuity of T_2 at $(\mu,\nu,c,\mathbb{G}_{n,m}^c)$ as a consequence of (OP) and upper semi-continuity of T_3 that

$$\begin{aligned}
&\limsup_{n,m\to\infty} \inf_{\pi\in\Pi^\star_c(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c)\\
&\quad\leq \pi^\star(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^c)\\
&\quad= \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^c).
\end{aligned}$$

Under (Sup), we note for each $\omega\in A$ by continuity of T_2 in h_c for fixed (μ,ν,c), upper semi-continuity of T_3 and continuity of T_4 that

$$\begin{aligned}
&\limsup_{n,m\to\infty} \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^{\nu}(f^c)\\
&+\sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\tilde{\mathbb{G}}_n^{\mu}(f^{cc}) - \tilde{\mathbb{G}}_n^{\mu}(f^{c_{n,m}c_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\tilde{\mathbb{G}}_m^{\nu}(f^c) - \tilde{\mathbb{G}}_m^{\nu}(f^{c_{n,m}})\right)\\
&\quad\leq \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^c)
\end{aligned}$$

$$+\sup_{f\in\mathcal{F}}\sqrt{\lambda}\,\left(\tilde{\mathbb{G}}^{\mu}(f^{cc})-\tilde{\mathbb{G}}^{\mu}(f^{cc})\right)+\sqrt{1-\lambda}\,\left(\tilde{\mathbb{G}}^{\nu}(f^{c})-\tilde{\mathbb{G}}^{\nu}(f^{c})\right)$$
$$=\inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\tilde{\mathbb{G}}^c)+\sup_{f\in S_c(\mu,\nu)}\sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc})+\sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^{c}).$$

As the lower bound and the upper bounds for $\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m})-OT(\mu,\nu,c))$ asymptotically match for all $\omega\in A$, it follows under both (OP) and (Sup) that

$$\lim_{n,m\to\infty}\sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m})-OT(\mu,\nu,c))=\inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\tilde{\mathbb{G}}^c)+\sup_{f\in S_c(\mu,\nu)}\sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc})+\sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^{c}).$$

As the set A has full measure we obtain that

$$\sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m})-OT(\mu,\nu,c))\xrightarrow{\text{a.s.}}\inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\tilde{\mathbb{G}}^c)+\sup_{f\in S_c(\mu,\nu)}\sqrt{\lambda}\,\tilde{\mathbb{G}}^{\mu}(f^{cc})+\sqrt{1-\lambda}\,\tilde{\mathbb{G}}^{\nu}(f^{c}),$$

where the limit has by measurability of T_2 and T_3 the same Borel law as the limit in the assertion, which finishes the proof. □

6.2.2. Proof of Proposition 10

Before turning to the proof of the bootstrap consistency, *i.e.*, Proposition 10, we state an important result on the convergence of the bootstrap empirical measure.

Lemma 48. *For a Polish space $\mathcal{X}$ let $\mu\in\mathcal{P}(\mathcal{X})$. Consider i.i.d. random variables $\{X_i\}_{i=1}^n\sim\mu^{\otimes n}$ and define the empirical measure $\mu_n:=n^{-1}\sum_{i=1}^n\delta_{X_i}$. Further, consider $k(n)$ i.i.d. random variables $\{X_k^b\}_{i=1}^n\sim\mu_n^{\otimes k(n)}$ to similarly define the bootstrap empirical measure $\mu_{n,k}^b:=\frac{1}{k(n)}\sum_{j=1}^{k(n)}\delta_{X_i^b}$. Then, provided that $k(n)\to\infty$ as $n\to\infty$, it follows under $n\to\infty$ that $\mu_{n,k}^b$ weakly converges to μ, in probability.*

The above lemma is a corollary of Theorem 2 in [91] and was added to ease further referencing. We can now prove Proposition 10 on bootstrap consistency under weakly converging costs.

Proof of Proposition 10. By Assumptions (JW) and (JW)*, and by measurability of the empirical and bootstrap empirical processes (Lemma 45) we infer using Lemma 2.2$(c)\Rightarrow(a)$ in Bücher and Kojadinovic [54] for two bootstrap versions $(\mu_{n,k}^{(1)},\nu_{n,k}^{(1)},c_{n,k}^{(1)})$, $(\mu_{n,k}^{(2)},\nu_{n,k}^{(2)},c_{n,k}^{(2)})$ based on independent bootstrap samples $\{X_i^{(1)}\}_{i=1}^k,\{X_i^{(2)}\}_{i=1}^k\sim\mu_n^{\otimes k}$ and $\{Y_i^{(1)}\}_{i=1}^k,\{Y_i^{(2)}\}_{i=1}^k\sim\nu_n^{\otimes k}$ for $n,k\to\infty$, $k=\mathrm{o}(n)$ that

$$\begin{pmatrix}\sqrt{n}\begin{pmatrix}\mu_n-\mu\\ \nu_n-\nu\\ c_n-c\end{pmatrix}\\ \sqrt{k}\begin{pmatrix}\mu_{n,k}^{(i)}-\mu_n\\ \nu_{n,k}^{(i)}-\nu_n\\ c_{n,k}^{(i)}-c_n\end{pmatrix}_{i=1,2}\end{pmatrix}\rightsquigarrow\begin{pmatrix}\begin{pmatrix}\mathbb{G}^{\mu}\\ \mathbb{G}^{\nu}\\ \mathbb{G}^{c}\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}^{\mu,(i)}\\ \mathbb{G}^{\nu,(i)}\\ \mathbb{G}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}$$

in $\left(C_u(\mathcal{F}^{cc})\times C_u(\mathcal{F}^{c})\times C(\mathcal{X}\times\mathcal{Y})\right)^3$. Since $k=\mathrm{o}(n)$ we also obtain by Slutzky's lemma that

$$\begin{pmatrix}\sqrt{n}\begin{pmatrix}\mu_n-\mu\\ \nu_n-\nu\\ c_n-c\end{pmatrix}\\ \sqrt{k}\begin{pmatrix}\mu_{n,k}^{(i)}-\mu\\ \nu_{n,k}^{(i)}-\nu\\ c_{n,k}^{(i)}-c\end{pmatrix}_{i=1,2}\end{pmatrix}=:\begin{pmatrix}\begin{pmatrix}\mathbb{G}_n^{\mu}\\ \mathbb{G}_n^{\nu}\\ \mathbb{G}_n^{c}\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}_{n,k}^{\mu,(i)}\\ \mathbb{G}_{n,k}^{\nu,(i)}\\ \mathbb{G}_{n,k}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}\rightsquigarrow\begin{pmatrix}\begin{pmatrix}\mathbb{G}^{\mu}\\ \mathbb{G}^{\nu}\\ \mathbb{G}^{c}\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}^{\mu,(i)}\\ \mathbb{G}^{\nu,(i)}\\ \mathbb{G}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}.$$

Herein, the triples $(\mathbb{G}^{\mu},\mathbb{G}^{\nu},\mathbb{G}^{c})$, $(\mathbb{G}^{\mu,(1)},\mathbb{G}^{\nu,(1)},\mathbb{G}^{c,(1)})$, and $(\mathbb{G}^{\mu,(2)},\mathbb{G}^{\nu,(2)},\mathbb{G}^{c,(2)})$ are independent and have identical law. Notably, invoking Corollary 39 we may assume without loss of generality that the empirical and bootstrap cost function c_n and $c_{n,k}^{(i)}$ for $i\in\{1,2\}$ deterministically satisfy the relation $\mathcal{F}_{\bar{c}}\subseteq\mathcal{F}^{\bar{c}\bar{c}}$, $\bar{c}\in\{c_n,c_{n,k}^{(i)}\}$. Moreover, by Varadarajan [92] we know that $\mu_n\rightsquigarrow\mu,\nu_n\rightsquigarrow\nu$ a.s. for $n\to\infty$, and by Lemma 48 we infer for $i\in\{1,2\}$ that $\mu_{n,k}^{(i)}\rightsquigarrow\mu,\nu_{n,k}^{(i)}\rightsquigarrow\nu$ in probability for $n,k\to\infty$, $k=\mathrm{o}(n)$. Hence, Slutzky's lemma asserts that

$$\begin{pmatrix}(\mathbb{G}_n^{\mu},\mathbb{G}_n^{\nu},\mathbb{G}_n^{c},\mu_n,\nu_n)^T\\ \left((\mathbb{G}_{n,k}^{\mu,(i)},\mathbb{G}_{n,k}^{\nu,(i)},\mathbb{G}_{n,k}^{c,(i)},\mu_{n,k}^{(i)},\nu_{n,k}^{(i)}\right)_{i=1,2}^T\end{pmatrix}\rightsquigarrow\begin{pmatrix}(\mathbb{G}^{\mu},\mathbb{G}^{\nu},\mathbb{G}^{c},\mu,\nu)^T\\ \left(\mathbb{G}^{\mu,(i)},\mathbb{G}^{\nu,(i)},\mathbb{G}^{c,(i)},\mu,\nu\right)_{i=1,2}^T\end{pmatrix}\tag{33}$$

in $\left(C_u(\mathcal{F}^{cc})\times C_u(\mathcal{F}^c)\times C(\mathcal{X}\times\mathcal{Y})\times\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\right)^3$. Moreover, using an analogous argument as for the proof of Lemma 46 we conclude that

$$\begin{pmatrix} \left(\left(\mathbb{G}_n^\mu(f^{cc}), \mathbb{G}_n^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}_n^c, \mu_n, \nu_n\right)^T \\ \left(\left(\left(\mathbb{G}_{n,k}^{\mu,(i)}(f^{cc}), \mathbb{G}_{n,k}^{\nu,(i)}(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,k}^{c,(i)}, \mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}\right)^T\right)_{i=1,2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu\right)^T \\ \left(\left(\left(\mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\nu,(i)}(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^{c,(i)}, \mu, \nu\right)^T\right)_{i=1,2} \end{pmatrix} \tag{34}$$

in the Polish space $\left(C_u(\mathcal{F})^2\times C(\mathcal{X}\times\mathcal{Y})\times\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\right)^3$, and we use under Assumption (OP) a Skorokhod representation for the process in (34).

Under Assumptions (Sup) and (Sup)*, by measurability of c_n and $c_{n,k}$ as maps to $C(\mathcal{X}\times\mathcal{Y})$, Lipschitzianity under c-transformation (Lemma 42) and Slutzky's lemma we conclude weak convergence of the random variables

$$\begin{aligned} &\begin{pmatrix} \left(\left(\mathbb{G}_n^\mu(f^{cc}), \mathbb{G}_n^\mu(f^{c_nc_n}), \mathbb{G}_n^\nu(f^c), \mathbb{G}_n^\nu(f^{c_n})\right)_{f\in\mathcal{F}}, \mathbb{G}_n^c, \mu_n, \nu_n\right)^T \\ \left(\left(\left(\mathbb{G}_{n,k}^{\mu,(i)}(f^{cc}), \mathbb{G}_{n,k}^{\mu,(i)}(f^{c_{n,k}^{(i)}c_{n,k}^{(i)}}), \mathbb{G}_{n,k}^{\nu,(i)}(f^c), \mathbb{G}_{n,k}^{\nu,(i)}(f^{c_{n,k}^{(i)}})\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,k}^{c,(i)}, \mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}\right)^T\right)_{i=1,2} \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu\right)^T \\ \left(\left(\left(\mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\nu,(i)}(f^c), \mathbb{G}^{\nu,(i)}(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^{c,(i)}, \mu, \nu\right)^T\right)_{i=1,2} \end{pmatrix} \end{aligned} \tag{35}$$

in the Polish space $\left(C_u(\mathcal{F})^4\times C(\mathcal{X}\times\mathcal{Y})\times\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\right)^3$. For the random variables from (35) we now take a Skorokhod representation.

To denote the random elements from the Skorokhod representation, we equip to the respective random variable with a tilde, *e.g.*, we write $\tilde{\mu}_n$ for the representation of μ_n. Following the same proof technique as Theorem 2 we thus conclude with Lemma 43 and Lemma 44 that

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}\left(OT(\tilde{\mu}_n, \tilde{\nu}_n, \tilde{c}_n) - OT(\mu, \nu, c)\right) \\ \left(\sqrt{k}\left(OT(\tilde{\mu}_{n,k}^{(i)}, \tilde{\nu}_{n,k}^{(i)}, \tilde{c}_{n,k}^{(i)}) - OT(\mu, \nu, c)\right)\right)_{i=1,2} \end{pmatrix} \\ &\xrightarrow{\text{a.s.}} \begin{pmatrix} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \tilde{\mathbb{G}}^\mu(f^{cc}) + \tilde{\mathbb{G}}^\nu(f^c) \\ \left(\inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^{c,(i)}) + \sup_{f\in S_c(\mu,\nu)} \tilde{\mathbb{G}}^{\mu,(i)}(f^{cc}) + \tilde{\mathbb{G}}^{\nu,(i)}(f^c)\right)_{i=1,2} \end{pmatrix}. \end{aligned}$$

Consequently, we infer for the original random variables and using that $k = o(n)$ that

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}\left(OT(\mu_n, \nu_n, c_n) - OT(\mu, \nu, c)\right) \\ \left(\sqrt{k}\left(OT(\mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}, c_{n,k}^{(i)}) - OT(\mu_n, \nu_n, c_n)\right)\right)_{i=1,2} \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^\mu(f^{cc}) + \mathbb{G}^\nu(f^c) \\ \left(\inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^{c,(i)}) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^{\mu,(i)}(f^{cc}) + \mathbb{G}^{\nu,(i)}(f^c)\right)_{i=1,2} \end{pmatrix}. \end{aligned}$$

Since the three components in the limit have identical distributions and are independent, the assertion follows at once from Bücher and Kojadinovic [54, Lemma 2.2 $(a) \Rightarrow (c)$]. □

6.3. Proofs for distributional limits under extremal-type costs

Before we proceed with the proofs for the results from Section 2.3 which rely on an application of the functional delta method, we provide a simple result on the support of the limiting processes. Its proof is deferred to Appendix E.6.

Lemma 49. *For a Polish space $\mathcal{X}$ let $\mu \in \mathcal{P}(\mathcal{X})$ and take a bounded, measurable function class $\tilde{\mathcal{F}}$ on $\mathcal{X}$. Then, the following assertions hold.*

(i) The contingent cone of $\mathcal{P}(\mathcal{X})$ at μ is given by $T_\mu\mathcal{P}(\mathcal{X}) = \mathrm{Cl}\{\frac{\mu'-\mu}{t} | t > 0, \mu' \in \mathcal{P}(\mathcal{X})\} \subseteq \ell^\infty(\tilde{\mathcal{F}})$.

(ii) For any $\Delta \in T_\mu\mathcal{P}(\mathcal{X})$ and $f, f' \in \tilde{\mathcal{F}}$ with $f - f' \equiv \kappa$ for some $\kappa \in \mathbb{R}$ it holds that $\Delta(f) = \Delta(f')$.

(iii) If $\tilde{\mathcal{F}}$ is μ-Donsker, then the tight limit $\mathbb{G}^\mu$ of the empirical process $\sqrt{n}(\mu_n - \mu)$ in $\ell^\infty(\tilde{\mathcal{F}})$ is a.s. contained in $T_\mu\mathcal{P}(\mathcal{X})$.

6.3.1. Proof of Theorem 5

The result follows by an application of the functional delta method [45]. Without loss of generality, we assume that $\mathcal{X} = \mathrm{supp}(\mu)$ and $\mathcal{Y} = \mathrm{supp}(\nu)$. This ensures that Kantorovich potentials are by (KP) unique on the full domains $\mathcal{X}$ and $\mathcal{Y}$. Assumption (Don) in conjunction with independence of the underlying random variables from μ and ν ensure by van der Vaart and Wellner [38, Example 1.4.6] that the joint process $\sqrt{nm/n+m}(\mu_n - \mu, \nu_m - \nu)$ weakly converge in $\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta})$. Further, by Lemma 49 the limit is a.s. contained in $T_\mu\mathcal{P}(\mathcal{X})\times T_\nu\mathcal{P}(\mathcal{Y})$. It remains to show that the map

$$(OT(\cdot,\cdot,c_\theta))_{\theta\in\Theta} : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}) \subseteq \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}) \to C(\Theta),$$

$$(\mu,\nu) \mapsto \left(\theta \mapsto \sup_{f\in\mathcal{F}} \mu(f^{c_\theta,c_\theta}) + \nu(f^{c_\theta})\right)$$

is Hadamard directionally differentiable at (μ,ν) tangentially to $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$. In the language of Theorem 52, take $\mathcal{F}$ and Θ as they are and set

$$V := \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}),\quad U := \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}),\quad E((\mu,\nu),f,\theta) := \mu(f^{c_\theta c_\theta})+\nu(f^{c_\theta}).$$

Then, Assumption **(EC)** follows from Lemma 42, while **(Lip)** and **(Lin)** are simple to verify by definition of V and E. Moreover, by Assumption List **(KP)** the condition of point *(ii)* in Lemma 53 holds, since the evaluations of E in f with $(\Delta^\mu,\Delta^\nu)\in T_\mu\mathcal{P}(\mathcal{X})\times T_\nu\mathcal{P}(\mathcal{Y})$ are invariant under constant shifts (Lemma 49), and since Kantorovich potentials are unique on $\mathcal{X}$ and $\mathcal{Y}$ up to a constant shift. This establishes **(DC)**, and the proof is complete. □

6.3.2. Proof of Theorem 6

Since Θ is a compact Polish space, it follows by Fang and Santos [48, Lemma S.4.9] (see also Cárcamo et al. [52, Corollary 2.3]) that the infimal mapping,

$$I : C(\Theta)\to\mathbb{R},\quad h\mapsto \inf_{\theta\in\Theta} h(\theta),$$

is Hadamard directionally differentiable at $OT(\mu,\nu,c(\cdot))\in C(\Theta)$ with derivative given by

$$D^H_{OT(\mu,\nu,c(\cdot))} I : C(\Theta)\to\mathbb{R},\quad \Delta^h\mapsto \inf_{\theta\in S_-(\Theta,\mu,\nu)}\Delta^h(\theta).$$

Hence, applying the functional delta method [45] for the infimal mapping I onto the uniform weak limit for the empirical OT process from Theorem 5 asserts the claim. □

6.3.3. Proof of Theorem 7

From the dual formulation (10) the supremal OT value over Θ is given by

$$\sup_{\theta\in\Theta} OT(\cdot,\cdot,c_\theta) : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\subseteq \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta})\to\mathbb{R},$$
$$(\mu,\nu)\mapsto \sup_{(f,\theta)\in\mathcal{F}\times\Theta}\mu(f^{c_\theta c_\theta})+\nu(f^{c_\theta}).$$

The results of Appendix A readily apply, with the choices for V, U, and E as in the proof of Theorem 5; the only difference being that the supremum is taken over $\mathcal{F}\times\Theta$ instead of $\mathcal{F}$. In particular, **(EC)**, **(Lip)**, and **(Lin)** are valid, whereas **(DC)** is now trivially fulfilled. Overall, Theorem 52 asserts that $\sup_{\theta\in\Theta} OT(\cdot,\cdot,c_\theta)$ is Hadamard directionally differentiable tangentially to $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ with derivative

$$D^H_{|(\mu,\nu)}\sup_{\theta\in\Theta} OT(\cdot,\cdot,c_\theta) : T_\mu\mathcal{P}(\mathcal{X})\times T_\nu\mathcal{P}(\mathcal{Y})\to\mathbb{R},\quad (\Delta^\mu,\Delta^\nu)\mapsto \sup_{\substack{\theta\in S_+(\Theta,\mu,\nu)\\ f_\theta\in S_{c_\theta}(\mu,\nu)}}\Delta^\mu(f_\theta^{c_\theta c_\theta})+\Delta^\nu(f_\theta^{c_\theta}).$$

Combined with weak convergence of $\sqrt{nm/n+m}(\mu_n-\mu,\nu_m-\nu)$ in $\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta})$ by (Don) in conjunction with the independence of the underlying samples [38, Example 1.4.6], and the inclusion of the limit in $T_\mu\mathcal{P}(\mathcal{X})\times T_\nu\mathcal{P}(\mathcal{Y})$ by Lemma 49, the functional delta method [45] implies the claim. □

Remark 50. In addition to the proof presented above, it is also possible to show Theorem 7 with similar arguments to those found in the proof of Fang and Santos [48, Lemma S.4.9] or Cárcamo et al. [52, Corollary 2.3]. However, their statements only provide sufficient conditions for Hadamard directional differentiability for tangentially to the space $C(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times C(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta})$, whereas the supremal OT value is defined only on the strict subset $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$.

6.3.4. Proof of Proposition 8

Define $\Delta(\mu_n,\nu_m,K) := \inf_{\theta\in\Theta} OT(\mu_n,\nu_m,c_\theta)-\inf_{\theta\in K} OT(\mu_n,\nu_m,c_\theta)$. Then,

$$\mathrm{P}^*\left(\Delta(\mu_n,\nu_m,K)\neq 0\right)\le \mathrm{P}^*\left(\left\{\Delta(\mu_n,\nu_m,K)\neq 0\right\}\cap\left\{\theta_{n,m}\in K\right\}\right)+\mathrm{P}^*(\theta_{n,m}\notin K),$$

and as the first summand in the above display is null while $\lim_{n\to\infty}\mathrm{P}^*(\theta_{n,m}\notin K)=0$, the right-hand side converges to zero. Hence, invoking Slutzky's Lemma [38, Example 1.4.7] it follows from Theorem 6 that

$$\begin{aligned}
&\sqrt{\frac{nm}{n+m}}\left(\inf_{\theta\in\Theta} OT(\mu_n,\nu_m,c_\theta)-\inf_{\theta\in\Theta} OT(\mu,\nu,c_\theta)\right)\\
&=\sqrt{\frac{nm}{n+m}}\Delta(\mu_n,\nu_m,K)+\sqrt{\frac{nm}{n+m}}\left(\inf_{\theta\in K} OT(\mu_n,\nu_m,c_\theta)-\inf_{\theta\in K} OT(\mu,\nu,c_\theta)\right)\\
&\rightsquigarrow 0+\inf_{\theta\in S_-(K,\mu,\nu)}\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})+\sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}).
\end{aligned}$$

The claim now follows at once after observing that $S_-(K,\mu,\nu)=S_-(\Theta,\mu,\nu)$. □

Funding

S. Hundrieser and C.A. Weitkamp gratefully acknowledge support from the DFG Research Training Group 2088 *Discovering structure in complex data: Statistics meets optimization and inverse problems*. G. Mordant gratefully acknowledges support from the DFG CRC 1456 *Mathematics of the Experiment A04* and A. Munk gratefully acknowledges support from the DFG CRC 1456 *A04, C06* and the Cluster of Excellence 2067 *MBExC Multiscale bioimaging—from molecular machines to networks of excitable cells*.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Uniform Hadamard differentiability of extremal-type functionals

A number of results in this work rely on the notion of Hadamard directional differentiability and the functional delta method. More precisely, both the result on the weak convergence of the empirical OT process from Section 2.3 and the formulation of regularity elevation functionals from Section 5 rely on this approach. Although, these two findings are conceptually rather unrelated, their proof techniques are based on a more general insight which we lay out in this section.

Let $(V, \|\cdot\|_V)$ be a normed vector space and consider sets $\mathcal{F}$ and Θ. Additionally, consider a real-valued function $E : V\times\mathcal{F}\times\Theta \to \mathbb{R}$ which assigns each triple (v, f, θ) to a some objective value $E(v, f, \theta)$. We are interested in sensitivity results for extremal-type functionals

$$\Psi(v) := \left(\sup_{f\in\mathcal{F}} E(v, f, \theta)\right)_{\theta\in\Theta} \qquad \text{and} \qquad \tilde{\Psi}(v) := \left(\inf_{f\in\mathcal{F}} E(v, f, \theta)\right)_{\theta\in\Theta}.$$

Herein, Θ provides the collection of feasible parameters which affect the optimization problem, while $\mathcal{F}$ represents the collection of feasible solutions. The space V denotes another set of parameters that determine the optimization problem and exhibit a vector space structure. Overall, these optimization problems characterize the general structure of processes indexed over Θ which are pointwise defined as the supremum or infimum over a collection $\mathcal{F}$ and depend on some parameter in V with an additive structure.

For our sensitivity analysis under perturbations of v it suffices to focus only on Ψ since

$$\inf_{f\in\mathcal{F}} E(v, f, \theta) = -\sup_{f\in\mathcal{F}} -E(v, f, \theta) \quad \text{for any } (v, \theta) \in V \times \Theta.$$

In the following, we first establish sufficient conditions in terms of E for the continuity properties of Ψ and the underlying sets of optimizers.

Lemma 51 (*Continuity*). *Let $(V, \|\cdot\|_V)$ be a normed vector space, consider compact topological spaces $\mathcal{F}$ and Θ whose topologies are generated by (pseudo-)metrics $d_{\mathcal{F}}$ and d_Θ, respectively, and assume that $E : V \times \mathcal{F} \times \Theta \to \mathbb{R}$ satisfies the following.*

(EC) *For any $v \in V$ the functional $E(v, \cdot, \cdot) : \mathcal{F} \times \Theta \to \mathbb{R}$ is continuous.*

(Lip) *There exists some $L \geq 0$ such that for any $(f, \theta) \in \mathcal{F} \times \Theta$ the functional $E(\cdot, f, \theta) : V \to \mathbb{R}$ is L-Lipschitz with respect to $\|\cdot\|_V$.*

Then, $\mathrm{Range}(\Psi) \subseteq C(\Theta)$ *and the functional $\Psi : V \to C(\Theta)$ is L-Lipschitz. Further, for any $(v, \theta) \in V \times \Theta$ the set of optimizers $S(v, \theta) := \{f \in \mathcal{F} \,|\, \sup_{f'\in\mathcal{F}} E(v, f', \theta) = E(v, f, \theta)\}$ is non-empty, and for fixed $v \in V$ the set-valued map*

$$(\theta, t) \in \Theta \times \mathbb{R}_+ \mapsto S(v, \theta; t) := \left\{ f \in \mathcal{F} \;\middle|\; \sup_{f'\in\mathcal{F}} E(v, f', \theta) \leq E(v, f, \theta) + t \right\}$$

is upper semi-continuous in terms of inclusion, i.e., for $\theta_n \to \theta$ and $t_n \to t$ any sequence $f_n \in S(\overline{v}, \theta_n; t_n)$ admits a converging subsequence $(f_{n_k})_{k\in\mathbb{N}}$ in $\mathcal{F}$ with limit $f \in S(\overline{v}, \theta; t)$.

Proof of Lemma 51. By Assumption **(EC)** and compactness of $\Theta\times\mathcal{F}$ it follows for any $v \in V$ that $E(v, \cdot, \cdot)$ is uniformly continuous, hence the function

$$w_{E,v} : \mathbb{R}_+ \to \mathbb{R}_+, \quad t \mapsto \sup_{\substack{d_\Theta(\theta,\theta')\leq t\\ d_{\mathcal{F}}(f,f')\leq t}} |E(v, f, \theta) - E(v, f', \theta')|$$

is finite for all $t \geq 0$ and fulfills $\lim_{t\searrow 0} w_{E,v}(t) = 0$. For $\theta, \theta' \in \Theta$ we thus find that

$$\left|\sup_{f\in\mathcal{F}} E(v, f, \theta) - \sup_{f\in\mathcal{F}} E(v, f, \theta')\right| \leq \sup_{f\in\mathcal{F}} |E(v, f, \theta) - E(v, f, \theta')| \leq w_{E,v}(d_\Theta(\theta, \theta')),$$

which implies for $v \in V$ that $\Psi(v) \in C(\Theta)$ and therefore $\mathrm{Range}(\Psi) \subseteq C(\Theta)$. For the Lipschitzianity of Ψ, note by Assumption **(Lip)** for any $v, v' \in V$ that

$$\|\Psi(v) - \Psi(v')\|_{C(\Theta)} = \sup_{\theta\in\Theta}\left|\sup_{f\in\mathcal{F}} E(v, f, \theta) - \sup_{f\in\mathcal{F}} E(v', f, \theta)\right|$$

$$\leq \sup_{\substack{\theta\in\Theta\\ f\in\mathcal{F}}} |E(v,f,\theta)-E(v',f,\theta)| \leq L\left\|v-v'\right\|_V.$$

To see that $S(v,\theta)\neq\emptyset$, note that the function $E(v,\cdot,\theta):\mathcal{F}\to\mathbb{R}$ is continuous for any $(v,\theta)\in V\times\Theta$; hence, by compactness of $\mathcal{F}$ the supremum over $\mathcal{F}$ is attained.

It remains to prove the assertion on upper semi-continuity. Consider converging sequences $t_n\to t\geq 0$ and $\theta_n\to\theta\in\Theta$ and take a sequence $f_n\in S(v,\theta_n;t_n)$. By compactness of $\mathcal{F}$ a converging subsequence $(f_{n_k})_{k\in\mathbb{N}}$ exists with limit $f\in\mathcal{F}$. Hence, by Assumption **(EC)** and since $\sup_{f\in\mathcal{F}} E(v,f,\cdot)=\Psi(v)(\cdot)\in C(\Theta)$ we obtain that $f\in S(v,\theta;t)$ since

$$E(v,f,\theta)+t=\lim_{k\to\infty} E(v,f_{n_k},\theta_{n_k})+t_{n_k}\geq \lim_{k\to\infty}\sup_{f\in\mathcal{F}} E(v,f,\theta_{n_k})=\sup_{f\in\mathcal{F}} E(v,f,\theta). \quad \square$$

With these tools at our disposal, we can state our general sensitivity result.

Theorem 52 (*Differentiability*). *Assume in the setting of Lemma 51 Assumptions* **(EC)** *and* **(Lip)**. *Let $\overline{v}\in V$ and consider a convex set $U\subset V$. Denote by $T_{\overline{v}}U:=\mathrm{Cl}\{\frac{v-\overline{v}}{t}\mid t>0, v\in U\}\subseteq V$ its contingent cone at $\overline{v}$. Further, assume the following.*

(Lin) *For any $(f,\theta)\in\mathcal{F}\times\Theta$ the function $\Delta_{|\overline{v}}E(\cdot,f,\theta):V\to\mathbb{R}, v\mapsto E(\overline{v}+v,f,\theta)-E(\overline{v},f,\theta)$ is linear.*

(DC) *For any $h\in T_{\overline{v}}U$ the function $\theta\in\Theta\mapsto\sup_{f\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f,\theta)$ is lower semi-continuous.*

Then, the functional

$$\Psi:V\to C(\Theta),\quad v\mapsto\left(\sup_{f\in\mathcal{F}} E(v,f,\theta)\right)_{\theta\in\Theta}$$

is Hadamard directionally differentiable at $\overline{v}$ tangentially to U with derivative given by

$$D^H_{|\overline{v}}\Psi:T_{\overline{v}}U\to C(\Theta),\quad h\mapsto\left(\sup_{f\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f,\theta)\right)_{\theta\in\Theta}.$$

Theorem 52 can be viewed as an extension of Fang and Santos [48, Lemma S.4.9] to a uniform perturbation result over the parameter space Θ. Additionally, our result does not require regularity properties on the full domain V but only a convex set U, an appealing property which we exploit in the context of our analysis for the OT process (where we choose $U=\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$) as well as regularity elevations (see proof of Proposition 38).

Assumptions **(EC)**, **(Lip)**, and **(Lin)** are fairly straightforward and often simple to verify. The first two conditions also appear to be necessary to infer that $\mathrm{Range}(\Psi)\subseteq C(\Theta)$ and Lipschitzianity of $\Psi:V\to C(\Theta)$. Assumption **(DC)** is more technical and requires some knowledge on the set of optimizers $S(\overline{v},\theta)$. As the proof of Theorem 52 reveals, is the functional $\theta\in\Theta\mapsto\sup_{f\in S(\overline{v},\theta)}E(h,f,\theta)$ under the assumptions of Lemma 51 always upper semi-continuous. Hence, the sole purpose of **(DC)** is to ensure $\mathrm{Range}(D^H_{|\overline{v}}\Psi)\subseteq C(\Theta)$. Sufficient conditions for its validity are stated as follows.

Lemma 53. *Assume the setting of Lemma 51 and Theorem 52. Then under either of the following conditions Assumption* **(DC)** *of Theorem 52 is fulfilled.*

(i) For any $\theta\in\Theta$ and $h\in T_{\overline{v}}U$ there exists $f\in S(\overline{v},\theta)$ with $\sup_{f'\in S(\overline{v},\theta)}\Delta_{svert\overline{v}}E(h,f',\theta)=\Delta_{|\overline{v}}E(h,f,\theta)$ such that any converging sequence $\theta_n\to\theta$ admits a sub-sequence (θ_{n_k}) and a converging sequence $f_{n_k}\in S(\overline{v},\theta_{n_k})$ with $f_{n_k}\to f$ in $\mathcal{F}$.

(ii) For any $\theta\in\Theta$ and $h\in T_{\overline{v}}U$ it holds that $\Delta_{|\overline{v}}E(h,f,\theta)=\Delta_{|\overline{v}}E(h,f',\theta)$ for any $f,f'\in S(\overline{v},\theta)$.

Proof of Lemma 53. Let $\theta_n\to\theta$ and consider an element $f\in S(\overline{v},\theta)$ such that $\Delta_{|\overline{v}}E(h,f,\theta)=\sup_{f'\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f,\theta)$. For setting (i) take an arbitrary subsequence θ_{n_k} and take another subsequence $\theta_{n_{k_l}}$ such that $f_{n_{k_l}}\in S(\overline{v},\theta_{n_{k_l}})$ converges to f for $l\to\infty$. Then, by **(EC)**,

$$\lim_{l\to\infty}\Delta_{|\overline{v}}E(h,f_{n_{k_l}},\theta_{n_{k_l}})=\Delta_{|\overline{v}}E(h,f,\theta)=\sup_{f'\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f',\theta).$$

This implies by monotonicity of the limit inferior and Lemma 56 that

$$\liminf_{n\to\infty}\sup_{f'\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f',\theta_n)\geq\liminf_{n\to\infty}\Delta_{|\overline{v}}E(h,f_n,\theta_n)\geq\sup_{f'\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f',\theta),$$

which asserts the validity of Assumption **(DC)** of Theorem 52. For setting (ii) take $f_n\in S(\overline{v},\theta_n)$ and consider by Lemma 51 a converging subsequence f_{n_k} with limit $f\in S(\overline{v},\theta)$. Hence, it holds that $\Delta_{|\overline{v}}E(h,f,\theta)=\sup_{f'\in S(\overline{v},\theta)}\Delta_{|\overline{v}}E(h,f',\theta)$ and the assertion follows from (i). $\square$

Proof of Theorem 52. The proof strategy is inspired by Römisch [45] who performs a sensitivity analysis for when Θ is a singleton. To extend the claim for a compact topological space Θ we employ the subsequent version of Dini's theorem.

Lemma 54 (*Dini's theorem, [49, Corollary 1]*). *Let Θ be a compact topological space and consider a decreasing $f_n:\Theta\to\mathbb{R}$ sequence (i.e., $f_n\geq f_{n+1}$ for all $n\in\mathbb{N}$) of upper semi-continuous functions. Further, assume that f_n pointwise converges to a (lower semi-)continuous function $f:\Theta\to\mathbb{R}$. Then, f_n converges to f uniformly on Θ.*

Take a positive null sequence $t_n \searrow 0$ with $t_n > 0$ for all $n \in \mathbb{N}$ and let $h \in T_{\overline{v}}U$. Further, take a sequence $h_n \in V$ such that $v_n := \overline{v} + t_n h_n \in U$ for all $n \in \mathbb{N}$ and $h_n \to h$ in V. For any $\theta \in \Theta$, we then observe by **(Lin)** and **(Lip)** for any $n \in \mathbb{N}$ the lower bound

$$\begin{aligned}\Psi(v_n)(\theta) - \Psi(\overline{v})(\theta) &= \sup_{f\in\mathcal{F}} E(v_n, f, \theta) - \sup_{f\in\mathcal{F}} E(\overline{v}, f, \theta)\\ &\geq \sup_{f\in S(\overline{v},\theta)} E(v_n, f, \theta) - E(\overline{v}, f, \theta)\\ &\geq \sup_{f\in S(\overline{v},\theta)} \Delta_{|\overline{v}} E(t_n h_n, f, \theta)\\ &\geq t_n \sup_{f\in S(\overline{v},\theta)} \Delta_{|\overline{v}} E(h, f, \theta) - 2t_n L \left\|h - h_n\right\|_V .\end{aligned} \tag{A.1}$$

Analogously, we obtain the upper bound

$$\begin{aligned}\Psi(v_n)(\theta) - \Psi(\overline{v})(\theta) &= \sup_{f\in\mathcal{F}} E(v_n, f, \theta) - \sup_{f\in\mathcal{F}} E(\overline{v}, f, \theta)\\ &\leq \sup_{f\in S(v_n,\theta)} E(v_n, f, \theta) - E(\overline{v}, f, \theta)\\ &\leq t_n \sup_{f\in S(v_n,\theta)} \Delta_{|\overline{v}} E(h, f, \theta) + 2t_n L \left\|h - h_n\right\|_V .\end{aligned} \tag{A.2}$$

Note that $S(v_n,\theta) \subseteq S(\overline{v},\theta; 2L\left\|v_n - v\right\|_V)$ since any $f^* \in S(v_n,\theta)$ fulfills by **(Lip)** the bound

$$\begin{aligned}E(v, f^*, \theta) &\geq E(v_n, f^*, \theta) - L\left\|v_n - v\right\|_V\\ &= \sup_{f\in\mathcal{F}} E(v_n, f, \theta) - L\left\|v_n - v\right\|_V\\ &\geq \sup_{f\in\mathcal{F}} E(v, f, \theta) - 2L\left\|v_n - v\right\|_V .\end{aligned}$$

Hence, it follows from (A.2) upon defining $\varepsilon_n := \sup_{k\geq n} 2L\left\|v_k - v\right\|_V$ that

$$\begin{aligned}\Psi(v_n)(\theta) - \Psi(\overline{v})(\theta) &\leq t_n \sup_{f\in S(\overline{v},\theta;2L\left\|v_n - v\right\|_V)} \Delta_{|\overline{v}} E(h, f, \theta) + 2t_n L\left\|h - h_n\right\|_V\\ &\leq t_n \sup_{f\in S(\overline{v},\theta;\varepsilon_n)} \Delta_{|\overline{v}} E(h, f, \theta) + 2t_n L\left\|h - h_n\right\|_V .\end{aligned} \tag{A.3}$$

Combining (A.1) and (A.3) we thus obtain for any $\theta \in \Theta$ that

$$\sup_{f\in S(\overline{v},\theta)} \Delta_{|\overline{v}} E(h, f, \theta) - 2L\left\|h - h_n\right\|_V \leq \frac{\Psi(v_n)(\theta) - \Psi(\overline{v})(\theta)}{t_n} \leq \sup_{f\in S(\overline{v},\theta;\varepsilon_n)} \Delta_{|\overline{v}} E(h, f, \theta) + 2L\left\|h - h_n\right\|_V .$$

To conclude the claim we show that the lower and upper bound uniformly converge on Θ for $n \to \infty$ to the $D^H_{|\overline{v}}\Psi$. Since $\left\|h_n - h\right\|_V \to \infty$, it suffices to prove for the functions

$$\Phi := D^H_{|\overline{v}}\Psi : \Theta \to \mathbb{R},\ \ \theta \mapsto \sup_{f\in S(\overline{v},\theta)} \Delta_{|\overline{v}} E(h, f, \theta), \qquad \Phi_n : \Theta \to \mathbb{R},\ \ \theta \mapsto \sup_{f\in S(\overline{v},\theta,\varepsilon_n)} \Delta_{|\overline{v}} E(h, f, \theta),$$

that $\lim_{n\to\infty} \left\|\Phi - \Phi_n\right\|_{C(\Theta)} = 0$. For this purpose, we employ Dini's theorem (Lemma 54).

In this context note, since $(\varepsilon_n)_{n\in\mathbb{N}}$ is a decreasing null-sequence, for all $n \in \mathbb{N}$ and any $\theta \in \Theta$ that $S(\overline{v},\theta) \subseteq S(\overline{v},\theta;\varepsilon_{n+1}) \subseteq S(\overline{v},\theta;\varepsilon_n)$ and consequently

$$\Phi(\theta) \leq \Phi_{n+1}(\theta) \leq \Phi_n(\theta) \leq 2 \sup_{\theta\in\Theta} \sup_{f\in\mathcal{F}} E(h, f, \theta) < \infty, \tag{A.4}$$

where the upper bound is finite due to Assumption **(EC)** and compactness of $\mathcal{F} \times \Theta$.

Further, let us show for any $\theta \in \Theta$ that $\lim_{n\to\infty} \Phi_n(\theta) = \Phi(\theta)$. Take a sequence $f_n \in S(\overline{v},\theta;\varepsilon_n)$ such that $\Phi_n(\theta) \leq \Delta_{|\overline{v}} E(h, f_n, \theta) + 1/n$. Consider a converging subsequence $(f_{n_k})_{k\in\mathbb{N}}$ with limit $f_\infty \in S(\overline{v},\theta)$. Then, by **(EC)** it follows that

$$\limsup_{k\to\infty} \Phi_{n_k}(\theta) \leq \lim_{k\to\infty} \Delta_{|\overline{v}} E(h, f_{n_k}, \theta) + 1/n_k = \Delta_{|\overline{v}} E(h, f_\infty, \theta) \leq \sup_{f\in S(\overline{v},\theta)} \Delta_{|\overline{v}} E(v, f, \theta) = \Phi(\theta).$$

Recalling (A.4), it thus follows that $\lim_{n\to\infty} \Phi_n(\theta) = \Phi(\theta)$.

To conclude the assertion with Dini's theorem it remains to show upper-continuity of Φ_n and of Φ; recall by Assumption **(DC)** that Φ is already lower semi-continuous. To this end, let $\varepsilon \geq 0$ and consider a converging sequence $\theta_n \to \theta$. Select $f_n \in S(\overline{v},\theta_n,\varepsilon)$ such that $\sup_{f\in S(\overline{v},\theta_n,\varepsilon)} \Delta_{|\overline{v}} E(h, f, \theta_n) \leq \Delta_{|\overline{v}} E(h, f_n, \theta_n) + 1/n$. Take a subsequence f_{n_k} and select by Lemma 51 another converging subsequence $f_{n_{k_l}}$ with limit $f_\infty \in S(\overline{v},\theta;\varepsilon)$. Using Assumption **(EC)** it thus follows that

$$\lim_{l\to\infty} \Delta_{|\overline{v}} E(h, f_{n_{k_l}}, \theta_{n_{k_l}}) + 1/n_{k_l} = \Delta_{|\overline{v}} E(h, f_\infty, \theta) \leq \sup_{f\in S(\overline{v},\theta;\varepsilon)} \Delta_{|\overline{v}} E(h, f, \theta)$$

Invoking monotonicity of the limit superior and Lemma 56 we thus obtain that

$$\limsup_{n\to\infty} \sup_{f\in S(\overline{v},\theta_n;\varepsilon)} \Delta_{|\overline{v}} E(h, f, \theta) \leq \limsup_{l\to\infty} \Delta_{|\overline{v}} E(h, f_n, \theta_n) + 1/n \leq \sup_{f\in S(\overline{v},\theta;\varepsilon)} \Delta_{|\overline{v}} E(h, f, \theta),$$

Hence, by Lemma 56 we conclude that Φ_n is upper semi-continuous and that Φ is continuous. Dini's theorem (Lemma 54) thus implies $\lim_{n\to\infty} \|\Phi - \Phi_n\|_\infty = 0$, asserting the Hadamard directional differentiability of Ψ at $\bar{v}$ tangentially to U. Finally, note that the range of $D^H_{\bar{v}}\Psi$ is indeed contained in $C(\Theta)$. $\square$

Appendix B. Proofs for Section 3: Sufficient criteria for assumptions

B.1. Proof of Proposition 13

By Lemma 1 it follows that $\mathcal{F}^c \subseteq \mathcal{H}_c^c + [-2B, 2B]$ and $\mathcal{F}^{cc} \subseteq \mathcal{H}_c + [-2B, 2B]$ with $\mathcal{H}_c$ defined in (7). Invoking Hundrieser et al. [43, Lemma 2.1] and Santambrogio [2, Proposition 1.34] we obtain for any $\varepsilon > 0$ that

$$\mathcal{N}(\varepsilon, \mathcal{F}^c, \|\cdot\|_\infty) = \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) \le \left\lceil \frac{2B}{\varepsilon} \right\rceil \mathcal{N}(\varepsilon/2, \mathcal{H}_c^c, \|\cdot\|_\infty) = \left\lceil \frac{2B}{\varepsilon} \right\rceil \mathcal{N}(\varepsilon/2, \mathcal{H}_c, \|\cdot\|_\infty).$$

For the function class $\mathcal{H}_c$, the asserted uniform metric entropy bounds are available in Section 3.1 and Appendix A of Hundrieser et al. [43]. Note by uniform boundedness of the cost function that $\mathcal{H}_c$ and $\mathcal{H}_c^c$ are uniformly bounded. The assertion on the universal Donsker property then follows from van der Vaart and Wellner [38, Theorem 2.5.6]. $\square$

B.2. Proof of Proposition 15

By assumption the functional

$$\begin{aligned} \overline{\Phi}_c : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) &\to \ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X} \times \mathcal{Y}) \\ (\mu, \nu) &\mapsto (\mu, \nu, \Phi_c(\mu, \nu)), \end{aligned}$$

where the domain is viewed as a subset of $\ell^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c)$, is Hadamard differentiable at (μ, ν) tangentially to $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. Moreover, since $\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc}$ is μ-Donsker it follows that $\sqrt{n/2}(\mu_n - \mu) \rightsquigarrow \mathbb{G}^\mu$ in $\ell^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc})$. Likewise, since $\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c$ is ν-Donsker it follows that $\sqrt{n/2}(\nu_n - \nu) \rightsquigarrow \mathbb{G}^\nu$ in $\ell^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c)$. Further, by independence of the random variables $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ it follows from [38, Theorem 1.4.6] that the joint empirical processes $\sqrt{n/2}(\mu_n - \mu, \nu_n - \nu)$ weakly converge in $\ell^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c)$ to $(\mathbb{G}^\mu, \mathbb{G}^\nu)$, contained in $T_\mu \mathcal{P}(\mathcal{X}) \times T_\nu \mathcal{P}(\mathcal{Y})$ by Lemma 49. We thus conclude by the functional delta method [45] for $\overline{\Phi}_c$ that (JW) is fulfilled.

Moreover, by the Donsker property and independence of the random variables, we also infer by van der Vaart and Wellner [38, Theorem 3.6.13] in the space $\ell^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c)$ that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\begin{pmatrix} \mu^b_{n,k} - \mu_n \\ \nu^b_{n,k} - \nu_n \end{pmatrix} | X_1, \dots, X_n, Y_1, \dots Y_n\right), \mathcal{L}\begin{pmatrix} \mathbb{G}^\mu \\ \mathbb{G}^\nu \end{pmatrix}\right) \xrightarrow{\mathrm{P}^*} 0.$$

Hence, by the functional delta method for conditionally weakly converging random variables Dümbgen [53] for $\overline{\Psi}_c$ we infer that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\begin{pmatrix} \mu^b_{n,k} - \mu_n \\ \nu^b_{n,k} - \nu_n \\ c^b_{n,k} - c_n \end{pmatrix} | X_1, \dots, X_n, Y_1, \dots Y_n\right), \mathcal{L}\left(\sqrt{n}\begin{pmatrix} \mu_n - \mu \\ \nu_n - \nu \\ c_n - c \end{pmatrix}\right)\right) \xrightarrow{\mathrm{P}^*} 0. \quad \square$$

B.3. Proof of Proposition 18

Before, we start to prove Proposition 18, we establish an auxiliary lemma.

Lemma 55. *Let $\mathcal{X}$ and $\mathcal{Y}$ be compact Polish spaces and consider $c \in C(\mathcal{X} \times \mathcal{Y})$.*

(i) For any function $g : \mathcal{X} \to \mathbb{R}$ and any constant κ, it holds that $(g + \kappa)^c = g^c - \kappa$.
(ii) Let $B > 0$. Then, for any $g : \mathcal{X} \to \mathbb{R}$ and $\Delta^c \in C(\mathcal{X} \times \mathcal{Y})$ with $\|g\|_\infty + 2\|c + \Delta^c\|_\infty \le B$ it holds that $g^{(c+\Delta^c)(c+\Delta^c)cc} \in \mathcal{H}_c + [-B, B]$.

The proof of the above lemma can be found in Appendix E.7.

Proof of Proposition 18. The proof is strongly inspired by van der Vaart and Wellner [61, Theorem 2.3] and employs standard empirical process arguments. In order to simplify the notation, we only consider the case $n = m$ and write c_n instead of $c_{n,n}$. Note that the claim for $n \neq m$ follows by the analogous arguments.

To show (i) first note by triangle inequality and using Lemma 42 that

$$\begin{aligned} \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{c_n c_n} - f^{cc})| &\le \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{c_n c_n} - f^{\tilde{c}_n \tilde{c}_n})| + \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n \tilde{c}_n} - f^{cc})| \\ &\le 4\sqrt{n}\|c_n - \tilde{c}_n\|_\infty + \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n \tilde{c}_n} - f^{cc})|. \end{aligned} \tag{B.1}$$

The first term converges by assumption for $n \to \infty$ in probability to zero. For the latter term note by (JW) and the assumption on $\tilde{c}_n$ that $\sqrt{n/2}(\tilde{c}_n - c) \rightsquigarrow \mathbb{G}^c$. By tightness of the law of $\mathbb{G}^c$ there exists for any $\varepsilon > 0$ a compact set $K \subseteq C(\mathcal{X} \times \mathcal{Y})$ such that $\mathrm{P}(\mathbb{G}^c \in K) > 1 - \varepsilon$; thus for any $\delta > 0$ the set K^δ of elements in $C(\mathcal{X} \times \mathcal{Y})$ with distance less than $\delta > 0$ to K fulfills

$$\liminf_{n\to\infty} \mathrm{P}\left(\sqrt{n/2}(\tilde{c}_n - c) \in K^\delta\right) \geq \mathrm{P}(\mathbb{G}^c \in K^\delta) > 1 - \varepsilon. \tag{B.2}$$

By compactness of K there exists a finite $\delta/2$-covering $\{h_1, \dots, h_p\}$ which implies that $K^{\delta/2} \subseteq \bigcup_{i=1}^p B(h_i, \delta)$, where $B(h, \delta)$ denotes the open ball of radius δ around h in the space $C(\mathcal{X} \times \mathcal{Y})$. We thus obtain

$$\left\{\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta/2}\right\} \subset \bigcup_{i=1}^p \left\{\tilde{c}_n \in B(c + 2n^{-1/2}h_i, \delta)\right\}.$$

Moreover, by Santambrogio [2, Proposition 1.34] it follows for any $f \in \mathcal{F}$ and $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ that $f^{\tilde{c}\tilde{c}} = f^{\tilde{c}\tilde{c}\tilde{c}\tilde{c}}$. Therefore, by triangle inequality,

$$\begin{aligned}
\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n} - f^{cc})| &= \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n\tilde{c}_n\tilde{c}_n} - f^{cccc})| \\
&\leq \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n\tilde{c}_n\tilde{c}_n} - f^{\tilde{c}_n\tilde{c}_n cc})| + \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n cc} - f^{cccc})| \\
&\leq \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n} - f^{cc})| + \sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n cc} - f^{cccc})|.
\end{aligned} \tag{B.3}$$

Assuming $\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta/2}$, it follows for the first term in (B.3) that

$$\begin{aligned}
\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n} - f^{cc})| &\leq \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} \sup_{\|h-h_i\|_\infty<\delta} |\mathbb{G}_n^\mu(f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})} - f^{cc})| \\
&\leq \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} \sup_{\|h-h_i\|_\infty<\delta} |\mathbb{G}_n^\mu(f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})} - f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})})| \\
&+ \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc})| \\
&\leq 8\delta + \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc})|.
\end{aligned} \tag{B.4}$$

Here, we used in the last inequality Lemma 42 to infer

$$\left\| f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})} - f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} \right\|_\infty \leq 4\|h_i - h\|_\infty / \sqrt{n} \leq 4\delta/\sqrt{n}$$

in conjunction with $\mathbb{G}_n^\mu(g) = \sqrt{n}(\mu_n - \mu)(g) \leq 2\sqrt{n}\|g\|_\infty$ for any measurable function g on $\mathcal{X}$. Now, define for $1 \leq i \leq p$ the function class

$$\tilde{\mathcal{G}}_n^i := \tilde{\mathcal{G}}_n^i(h_i) := \left\{ f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc} \,\middle|\, f \in \mathcal{F}^{\tilde{c}_n\tilde{c}_n} \right\}.$$

For each $1 \leq i \leq p$ and any $\varepsilon > 0$ we then observe that

$$\begin{aligned}
\log N(\varepsilon, \tilde{\mathcal{G}}_n^i, \|\cdot\|_\infty) &\leq \log N(\varepsilon, \mathcal{F}^{\tilde{c}_n\tilde{c}_n(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}, \|\cdot\|_\infty) + \log N(\varepsilon, \tilde{\mathcal{F}}^{\tilde{c}_n\tilde{c}_n cc}, \|\cdot\|_\infty) \\
&\leq 2\log N(\varepsilon, \mathcal{F}^{\tilde{c}_n\tilde{c}_n}, \|\cdot\|_\infty),
\end{aligned}$$

where the last step follows by Lemma 2.1 in [43]. In consequence, it follows by Dudley's entropy integral (see, *e.g.*,[93, Chapter 5]) that

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} \left|\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc})\right|\right] &\leq \sum_{i=1}^p \mathbb{E}\left[\sup_{g\in\tilde{\mathcal{G}}_n^i} |\mathbb{G}_n^\mu(g)|\right] \\
&\lesssim \sum_{i=1}^p \int_0^{4\|h_i\|_\infty/\sqrt{n}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_n\tilde{c}_n}, \|\cdot\|_\infty)\right)} d\varepsilon \\
&\lesssim \sum_{i=1}^p \int_0^{4\|h_i\|_\infty/\sqrt{n}} \varepsilon^{-\alpha/2} d\varepsilon \\
&\lesssim \sum_{i=1}^p (\|h_i\|_\infty / \sqrt{n})^{1-\alpha/2},
\end{aligned}$$

where by assumption the hidden constants do not depend on n. We thus infer conditionally on the event $\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta/2}$ for $n \to \infty$ that

$$\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}} \max_{i=1,\dots,p} \mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc}) \xrightarrow{\mathrm{P}} 0. \tag{B.5}$$

For the second term in (B.3) we assume $\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta/2}$ and obtain by similar arguments,

$$\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde{c}_n\tilde{c}_n cc} - f^{cccc})| \leq 8\delta + \sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc})|. \tag{B.6}$$

Upon defining the function class

$$\tilde{\mathcal{G}}_n := \tilde{\mathcal{G}}_n(h_1,\dots,h_p) := \left\{ f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc} \,\middle|\, f \in \mathcal{F} \right\} \tag{B.7}$$

we note again by Lemma 42 that any $g \in \mathcal{G}_n$ fulfills $\|g\|_\infty \le \max_{i=1,\dots,p} 4\,\|h_i\|_\infty / \sqrt{n}$. Further, for n sufficiently large there exists a constant $B > 0$ such that Lemma 55 is applicable for any $f \in \mathcal{F}$, and we obtain

$$f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} \in \mathcal{H}_c + [-B, B].$$

Hence, for sufficiently large n, it follows by Lemma 1 for any $\varepsilon > 0$ that

$$\mathcal{N}(\varepsilon, \tilde{\mathcal{G}}_n(h_1,\dots,h_p), \|\cdot\|_\infty) \le \left(\mathcal{N}(\varepsilon, \mathcal{F}^{cc} + [-B,B], \|\cdot\|_\infty)\right)^2.$$

Again invoking, Dudley's entropy integral asserts such for n that

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} \left|\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc})\right|\right] &= \mathbb{E}\left[\sup_{\tilde f\in\tilde{\mathcal{G}}_n} |\mathbb{G}_n^\mu(\tilde f)|\right] \\
&\lesssim \int_0^{\max_{i=1,\dots,p} 4\|h_i\|_\infty/\sqrt{n}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \tilde{\mathcal{G}}_n, \|\cdot\|_\infty)\right)}\, d\varepsilon \\
&\le \int_0^{\max_{i=1,\dots,p} 4\|h_i\|_\infty/\sqrt{n}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{cc}+[-B,B], \|\cdot\|_\infty)\right)}\, d\varepsilon \\
&\lesssim \int_0^{\max_{i=1,\dots,p} 4\|h_i\|_\infty/\sqrt{n}} \varepsilon^{-\alpha/2}\, d\varepsilon \\
&\lesssim \left(\max_{i=1,\dots,p} \|h_i\|_\infty / \sqrt{n}\right)^{1-\alpha/2}.
\end{aligned}$$

This implies conditionally on the event $\sqrt{n/2}(\tilde c_n - c) \in K^{\delta/2}$ for $n \to \infty$ that

$$\sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc})| \xrightarrow{\mathrm{P}} 0. \tag{B.8}$$

Concluding, for any $\varepsilon > 0$ it follows for $\delta := \varepsilon/32 > 0$ from (B.2)–(B.8) that

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde c_n \tilde c_n} - f^{cc})| > \varepsilon\right) \\
&\le \limsup_{n\to\infty}\left(\mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde c_n\tilde c_n} - f^{cc})| > \varepsilon, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) + \mathbb{P}\left(\sqrt{n/2}(\tilde c_n - c) \notin K^{\delta/2}\right)\right) \\
&\le \limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde c_n\tilde c_n} - f^{cc})| > \varepsilon, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) + \varepsilon \\
&\le \limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}^{\tilde c_n\tilde c_n}} |\mathbb{G}_n^\mu(f^{\tilde c_n\tilde c_n} - f^{cc})| > \varepsilon/2, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) \\
&\quad + \limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde c_n\tilde c_n cc} - f^{cccc})| > \varepsilon/2, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) + \varepsilon \\
&\le \limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}^{\tilde c_n\tilde c_n}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc})| > \varepsilon/4, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) \\
&\quad + \limsup_{n\to\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}_n^\mu(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc})| > \varepsilon/4, \sqrt{n/2}(\tilde c_n - c)\in K^{\delta/2}\right) + \varepsilon = \varepsilon,
\end{aligned}$$

which shows the convergence in probability of $\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\mu(f^{\tilde c_n\tilde c_n} - f^{cc})|$ to zero. We thus conclude the convergence in probability for both terms of (B.1). An analogous argument yields the convergence $\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\nu(f^{c_n} - f^c)| \xrightarrow{\mathrm{P}} 0$ for $n\to\infty$, where we apply Lemma 2.1 of [43] to obtain

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde c_n} \cup \mathcal{F}^c, \|\cdot\|_\infty) &\le \sup_{n\in\mathbb{N}} \left(\log\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde c_n}, \|\cdot\|_\infty) + \log\mathcal{N}(\varepsilon, \mathcal{F}^c, \|\cdot\|_\infty)\right) \\
&= \sup_{n\in\mathbb{N}} \left(\log\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde c_n\tilde c_n}, \|\cdot\|_\infty) + \log\mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty)\right) \\
&\le \sup_{n\in\mathbb{N}} 2\log\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde c_n\tilde c_n} \cup \mathcal{F}^{cc}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \qquad \text{for } \alpha < 2,
\end{aligned}$$

which overall verifies (Sup) of Theorem 2.

For (ii) note by Bücher and Kojadinovic [54] and since $k = k(n) = o(n)$ for $n\to\infty$ that

$$\sqrt{k}(\tilde c_{n,k}^b - c) = \sqrt{k}(\tilde c_{n,k}^b - c_{n,k}^b) + \sqrt{k}(c_{n,k}^b - c_n) + \sqrt{\frac{k}{n}}\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c.$$

Likewise, it follows for $n \to \infty$ that $\sqrt{k}(\mu^b_{n,k} - \mu) \rightsquigarrow \mathbb{G}^\mu$ in $\ell^\infty(\mathcal{F}^{cc})$, $\sqrt{k}(\nu^b_{n,k} - \nu) \rightsquigarrow \mathbb{G}^\nu$ in $\ell^\infty(\mathcal{F}^c)$. This means that we can pursue a similar proof strategy as for (i). Define $\mathbb{G}^\mu_{n,k} := \sqrt{k}(\mu^b_{n,k} - \mu)$ and $\mathbb{G}^\nu_{n,k} := \sqrt{k}(\nu^b_{n,k} - \nu)$. Then, we infer from Lemma 42 that

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{c^b_{n,k}c^b_{n,k}} - f^{cc})| &\le \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{c^b_{n,k}c^b_{n,k}} - f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}})| + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})|\\ &\le 4\sqrt{k}\left\|c^b_{n,k} - \tilde{c}^b_{n,k}\right\|_\infty + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})|,\end{aligned} \tag{B.9}$$

where the first term converges for $n \to \infty$ in probability to zero. By Santambrogio [2, Proposition 1.34] we obtain that

$$\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \le \sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}cc} - f^{cccc})|.$$

Moreover, by analogous arguments to those for (i) we obtain with probability at least $1-\varepsilon$ for n sufficiently large that

$$\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \le 8\delta + \sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})| \tag{B.10}$$

as well as

$$\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}cc} - f^{cccc})| \le 8\delta + \sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})cc} - f^{cccc})|. \tag{B.11}$$

Next, we verify that the suprema on the right-hand sides of (B.10) and (B.11) converge (unconditionally with respect to the μ_n but conditionally on the set with probability at least $1-\varepsilon$) to zero. We note by Dudley's entropy integral for the bootstrap empirical process $\sqrt{k}(\mu^b_{n,k} - \mu_n)$ and the empirical process $\sqrt{n}(\mu_n - \mu)$ as well as our previous considerations that

$$\begin{aligned}&\lesssim \mathbb{E}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})|\right]\\ &= \sum_{i=1}^p \mathbb{E}_{\mu_n}\left[\mathbb{E}_{\mu^b_{n,k}}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\sqrt{k}(\mu^b_{n,k} - \mu_n)(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})|\,\middle|\, \mu_n\right]\right]\\ &\quad + \sqrt{\frac{k}{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_n(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})|\right]\\ &\lesssim \sum_{i=1}^p \mathbb{E}_{\mu_n}\int_0^{4\|h_i\|_\infty/\sqrt{k}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}, \|\cdot\|_\infty)\right)}d\varepsilon + \sqrt{\frac{k}{n}}\int_0^{4\|h_i\|_\infty/\sqrt{k}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}, \|\cdot\|_\infty)\right)}d\varepsilon\\ &\lesssim \sum_{i=1}^p \left(1 + \sqrt{\frac{k}{n}}\right)\int_0^{4\|h_i\|_\infty/\sqrt{k}} \varepsilon^{-\alpha/2}d\varepsilon \lesssim \sum_{i=1}^p \left(1 + \sqrt{\frac{k}{n}}\right)\left(\|h_i\|_\infty/\sqrt{k}\right)^{1-\alpha/2},\end{aligned}$$

which tends to zero for $n \to \infty$ with $k = k(n) = \mathrm{o}(n)$ since the hidden constants do not depend on n, k. Recalling the definition of the function class $\tilde{\mathcal{G}}_k$ in (B.7) with n replaced by k, we obtain

$$\lesssim \mathbb{E}\left[\sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})cc} - f^{cccc})|\right] = \mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k} \left|\mathbb{G}^\mu_{n,k}(f)\right|\right].$$

Hence, Dudley's entropy integral in combination with our previous considerations yields

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k} \left|\mathbb{G}^\mu_{n,k}(f)\right|\right] &\le \mathbb{E}_{\mu_n}\left[\mathbb{E}_{\mu^b_{n,k}}\left[\sup_{f\in\tilde{\mathcal{G}}_k} |\sqrt{k}(\mu^b_{n,k} - \mu_n)(f)|\,\middle|\, \mu_n\right]\right] + \sqrt{\frac{k}{n}}\mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k} |\mathbb{G}^\mu_n(f)|\right]\\ &\lesssim \left(1 + \sqrt{\frac{k}{n}}\right)\int_0^{\max_{i=1,\dots,p} 4\|h_i\|_\infty/\sqrt{k}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{cc} + [-B, B], \|\cdot\|_\infty)\right)}d\varepsilon\\ &\lesssim \left(1 + \sqrt{\frac{k}{n}}\right)\left(\max_{i=1,\dots,p} \|h_i\|_\infty/\sqrt{k}\right)^{1-\alpha/2},\end{aligned}$$

which goes to zero for $n, k(n) \to \infty$ with $k(n) = \mathrm{o}(n)$ (the hidden constants are independent of n, k).

Using the same arguments as in (i), we conclude that

$$\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \xrightarrow{\mathrm{P}} 0.$$

Finally, analogous arguments yield that $\sup_{f\in\mathcal{F}} |\mathbb{G}^\nu_{n,k}(f^{c^b_{n,k}} - f^c)| \xrightarrow{\mathrm{P}} 0$, thus showing (ii). □

B.4. Proof of Corollary 19

Define the random variables

$$\tilde{c}_n := \begin{cases} c & \text{if } n < N, \\ c_n & \text{if } n \geq N, \end{cases} \quad \text{and} \quad \tilde{c}^b_{n,k} := \begin{cases} c & \text{if } n < N \text{ or } k < K, \\ c^b_{n,k} & \text{if } n \geq N \text{ and } k \geq K. \end{cases}$$

By Proposition 13 the cost estimators $\tilde{c}_n$ and $\tilde{c}^b_{n,k}$ satisfy the entropy bounds in (12) and (13). Tightness of N and K implies that $\sqrt{n}\|\tilde{c}_n - c_n\|_\infty \xrightarrow{\mathrm{P}} 0$ and $\sqrt{k}\|\tilde{c}^b_{n,k} - c^b_{n,k}\|_\infty \xrightarrow{\mathrm{P}} 0$ for $n, k \to \infty$, which asserts the claim by Proposition 18. □

B.5. Proof of Corollary 20

By Assumption List **(JW)** it follows that $\sqrt{nm/(n+m)}(c_{n,m} - c) \rightsquigarrow \mathbb{G}^c$, whereas under (JW)* we infer from Bücher and Kojadinovic [54] and $k = \mathrm{o}(n)$ that $\sqrt{k}(c^b_{n,k} - c) \rightsquigarrow \mathbb{G}^c$ unconditionally. In what follows we state the arguments for (Sup); for (Sup)* a similar proof strategy applies by replacing the empirical costs process by the bootstrap cost process.

First, assume without loss of generality that the population cost function fulfills $\|c\|_\infty \leq 1$. Then, for all three settings of Proposition 13 it follows that $\log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha}$ with $\alpha < 2$.

For setting (i) we set $\tilde{c}_{n,m} := \Psi_{\mathrm{bdd}}(c_{n,m})$, for Ψ_{bdd} defined in Section 5.1. Since $\|\tilde{c}_{n,m}\|_\infty \leq 2$ and $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ is uniformly bounded by 6, we obtain that $\mathcal{F}^{\tilde{c}_{n,m}}$ is uniformly bounded by 8. By Propositions 36 and 37 both conditions of Proposition 18(i) are met, asserting (Sup).

For setting (ii) we take $\tilde{c}_{n,m} := \Psi^{\tilde{d}_\mathcal{X}}_{\mathrm{mod}} \circ \Psi_{\mathrm{bdd}}(c_{n,m})$ for $\Psi^{\tilde{d}_\mathcal{X}}_{\mathrm{mod}}$ from Section 5.2. Then, $\|\tilde{c}_{n,m}\|_\infty \leq 2$ and $\mathcal{F}^{\tilde{c}_{n,m}}$ is uniformly bounded by 8. Moreover, by Assumption (ii)' it follows with Propositions 36 and 38 that $\sqrt{nm/(n+m)}\|\tilde{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0$ and that

$$\sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}, \|\cdot\|_\infty) \lesssim \mathcal{N}(\varepsilon/8, \mathcal{X}, \tilde{d}_\mathcal{X})|\log(\varepsilon)| \lesssim \varepsilon^{-\beta}|\log(\varepsilon)| \lesssim \varepsilon^{-2+(2-\beta)/2},$$

where we used the covering number assumption on $\mathcal{X}$. (Sup) then follows from Proposition 18(i).

For setting (iii) define $c_i \in C(\mathcal{U}_i \times \mathcal{Y})$ as $c_i(u, y) := c(\zeta_i(u), y)$. We consider $\tilde{c}_{n,m} := \Psi_{\mathrm{com}}(c_{n,m})$ where Ψ_{com} denotes the combination (Section 5.4) of regularity elevation functionals $\Psi_i : C(\mathcal{U}_i \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y})$ defined by $\Psi_i = \Psi^{\|\cdot\|^{\gamma_i}}_{\mathrm{mod}} \circ \Psi_{\mathrm{bdd}}$ from Section 5.2 if $\gamma_i \in (0,1]$, and $\Psi_i = \Psi^{c_i,\gamma_i}_{\mathrm{Hol}} \circ \Psi_{\mathrm{bdd}}$ from Section 5.3 if $\gamma_i \in (1,2]$, where we replace $\mathcal{X}$ by $\mathcal{U}_i$. Then, by Propositions 38, 40, and 41 the functional Ψ fulfills the assumptions of Proposition 36 and therefore $\sqrt{nm/(n+m)}\|\tilde{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{\mathrm{P}} 0$. Moreover, since for any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ it holds that $\|\Psi(\tilde{c})\|_\infty < C$ for a deterministic constant $C \geq 0$ that only depends on the functions c_i and the spaces $\mathcal{U}_i$, it follows that $\mathcal{F}^{\Psi(\tilde{c})}$ is uniformly bounded by $C + 6$ and therefore

$$\sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}, \|\cdot\|_\infty) \lesssim \sum_{i=1}^{I} \sup_{\tilde{c}_i \in C(\mathcal{U}_i \times \mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\Psi_i(\tilde{c}_i)}, \|\cdot\|_\infty) \lesssim \max_{i=1,\ldots,I} \varepsilon^{-d_i/\gamma_i},$$

where we use for the first inequality Proposition 41, and for the second we employ the bounds from Proposition 38 with $\mathcal{N}(\varepsilon, \mathcal{U}_i, \|\cdot\|^{\gamma_i}) \lesssim \varepsilon^{-d_i/\gamma_i}$ for $0 < \gamma_i \leq 1$ and Proposition 40 for $1 < \gamma_i \leq 2$. The assertion then follows by an application of Proposition 18(i). □

B.6. Proof of Lemma 22

For $\varepsilon > 0$ suppose that the right-hand side is finite since otherwise the claim is vacuous. Set $k = \mathcal{N}(\varepsilon/4, \Theta, d_\Theta)$ and let $\{\theta_1, \ldots, \theta_k\}$ be a minimal $\varepsilon/4$-covering of Θ. Further, for each $i = 1, \ldots, k$ let $\{f^i_1, \ldots, f^i_{k_i}\}$ be a minimal $\varepsilon/2$-covering of $\mathcal{F}^{c_{\theta_i}c_{\theta_i}}$, *i.e.*, $k_i = \mathcal{N}\left(\varepsilon/2, \mathcal{F}^{c_{\theta_i}c_{\theta_i}}, \|\cdot\|_\infty\right)$. Once we show that $\mathcal{F}_\mathcal{X}(\varepsilon) := \bigcup_{i=1}^k \{f^i_1, \ldots, f^i_{k_i}\}$ is an ε-covering for $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta c_\theta}$ and that $\mathcal{F}_\mathcal{Y}(\varepsilon) := \bigcup_{i=1}^k \{(f^i_1)^{c_{\theta_i}}, \ldots, (f^i_{k_i})^{c_{\theta_i}}\}$ is an ε-covering for $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta}$ the claim follows, since

$$|\mathcal{F}_\mathcal{Y}(\varepsilon)| \leq |\mathcal{F}_\mathcal{X}(\varepsilon)| = \sum_{i=1}^{k} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_{\theta_i}c_{\theta_i}}, \|\cdot\|_\infty\right) \leq \mathcal{N}\left(\frac{\varepsilon}{4}, \Theta, d_\Theta\right) \sup_{\theta\in\Theta} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty\right).$$

Hence, let $\theta \in \Theta$ and $f \in \mathcal{F}^{c_\theta c_\theta}$, and choose $\tilde{f} \in \mathcal{F}$ with $f = \tilde{f}^{c_\theta c_\theta}$. Select θ_i with $d_\Theta(\theta, \theta_i) \leq \varepsilon/4$ and choose $f^i_{l_i} \in \mathcal{F}_\mathcal{X}(\varepsilon)$ such that $\left\|f^i_{l_i} - \tilde{f}^{c_{\theta_i}c_{\theta_i}}\right\|_\infty \leq \varepsilon/2$. Now, by Lipschitzianity of the cost in θ and Lemma 42 we infer $\|\tilde{f}^{c_{\theta_i}c_{\theta_i}} - \tilde{f}^{c_\theta c_\theta}\|_\infty \leq 2d_\Theta(\theta, \theta_i) \leq \varepsilon/2$, and it follows that

$$\left\|f^i_{l_i} - f\right\|_\infty = \left\|f^i_{l_i} - \tilde{f}^{c_\theta c_\theta}\right\|_\infty \leq \left\|f^i_{l_i} - \tilde{f}^{c_{\theta_i}c_{\theta_i}}\right\|_\infty + \left\|\tilde{f}^{c_{\theta_i}c_{\theta_i}} - \tilde{f}^{c_\theta c_\theta}\right\|_\infty \leq \varepsilon,$$

which verifies that $\mathcal{F}_\mathcal{X}(\varepsilon)$ is an ε-covering of $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}$.

Moreover, for any $f \in \mathcal{F}^{c_\theta}$ there exists $\tilde{f} \in \mathcal{F}$ with $f = \tilde{f}^{c_\theta}$ and by Santambrogio [2, Proposition 1.34] it follows that $\tilde{f}^{c_\theta} = \tilde{f}^{c_\theta c_\theta c_\theta}$. Hence, upon selecting $f^i_{l_i} \in \mathcal{F}_\mathcal{X}(\varepsilon)$ as above, we find by Lemma 42 that

$$\left\|(f^i_{l_i})^{c_{\theta_i}} - \tilde{f}^{c_{\theta_i}}\right\|_\infty = \left\|(f^i_{l_i})^{c_{\theta_i}} - \tilde{f}^{c_{\theta_i}c_{\theta_i}c_{\theta_i}}\right\|_\infty \leq \left\|f^i_{l_i} - \tilde{f}^{c_{\theta_i}c_{\theta_i}}\right\|_\infty \leq \varepsilon/2.$$

Again invoking Lemma 42 yields $\|\tilde f^{c_{\theta_i}} - \tilde f^{c_\theta}\|_\infty \le d(\theta,\theta_i) \le \varepsilon/4$. Consequently, we find that

$$\left\|(f^i_{l_i})^{c_{\theta_i}} - f\right\|_\infty = \left\|(f^i_{l_i})^{c_{\theta_i}} - \tilde f^{c_\theta}\right\|_\infty \le \left\|(f^i_{l_i})^{c_{\theta_i}} - \tilde f^{c_{\theta_i}}\right\|_\infty + \|\tilde f^{c_{\theta_i}} - \tilde f^{c_\theta}\|_\infty \le \frac{3\varepsilon}{4} \le \varepsilon,$$

which proves that $\mathcal{F}_{\mathcal{Y}}(\varepsilon)$ is an ε-covering of $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}$ and finishes the proof. □

Appendix C. Proofs for Section 4: Applications

C.1. Proof of Lemma 24

Select U as the pre-image of $\{\tilde g \in C(\mathcal{X},\mathbb{R}^d) : \|\tilde g - g^{-1}_{\vartheta^o}\|_\infty < 1\}$ under K_Θ, which is open (relative) in Θ due to continuity. Hence, by compactness of $\mathcal{X}$, the collection $\{g^{-1}_\vartheta\}_{\vartheta\in U}$ is uniformly bounded on $\mathcal{X}$. Invoking the Cauchy–Schwarz inequality and, due to compactness of $\mathcal{Y}$, we infer that $\{C_\Theta(\vartheta)(x,\cdot)\}_{\vartheta\in U, x\in\mathcal{X}}$ is also uniformly bounded on $\mathcal{Y}$. Further, since $\nabla_y C_\Theta(\vartheta)(x,y) = 2\big(g^{-1}_\vartheta(x) - y\big)$ for $y \in \operatorname{int}(\mathcal{Y})$ the collection $\{\nabla_y C_\Theta(\vartheta)(x,\cdot)\}$ is bounded on $\mathcal{Y}$ uniformly over $\vartheta\in U, x\in\mathcal{X}$. Finally, note that $\operatorname{Hess}_y C_\Theta(\vartheta)(x,y) = -2\mathrm{Id}$, independent of $\vartheta \in U, x\in\mathcal{X}$. Combining these observations, we conclude the existence of $\Lambda \ge 0$ such that the $(2,\Lambda)$-Hölder regularity is met. □

C.2. Proof of Lemma 25

To establish the Hadamard differentiability of C_Θ at ϑ^o note that

$$\begin{aligned}
&\left\|\frac{C_\Theta(\vartheta^o + t_n h_n) - C_\Theta(\vartheta^o)}{t_n} - D^H_{|\vartheta^o} C_\Theta(h)\right\|_\infty\\
&= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left|\frac{1}{t_n}\left\langle g^{-1}_{\vartheta^o+t_nh_n}(x) - g^{-1}_{\vartheta^o}(x), g^{-1}_{\vartheta^o+t_nh_n}(x) + g^{-1}_{\vartheta^o}(x) - 2y\right\rangle - 2\left\langle D^H_{\vartheta^o} K_\Theta(h)(x), g^{-1}_{\vartheta^o}(x) - y\right\rangle\right|\\
&\le \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left(\left|2\left\langle \frac{1}{t_n}\left(g^{-1}_{\vartheta^o+t_nh_n}(x) - g^{-1}_{\vartheta^o}(x)\right) - D^H_{\vartheta^o} K_\Theta(h)(x), g^{-1}_{\vartheta^o}(x) - y\right\rangle\right| + \frac{1}{t_n}\left\|g^{-1}_{\vartheta^o+t_nh_n}(x) - g^{-1}_{\vartheta^o}(x)\right\|^2\right).
\end{aligned}$$

For $n\to\infty$, the first term tends to zero by Hadamard differentiability of K_Θ whereas the second term tends to zero by **(G)**. Hence, C_Θ is Hadamard differentiable at ϑ^o. The second assertion follows from the functional delta method for Hadamard differentiable functionals [45]. □

C.3. Proof of Proposition 26

First note that, in comparison to Sections 2 and 3, the roles of $\mathcal{X}$ and $\mathcal{Y}$ are interchanged. The universal Donsker property of $\mathcal{F}^{C_\Theta(\vartheta^o)}$ follows from Proposition 13(*iii*) since $d\le 3$ and $C_\Theta(\vartheta^o)(x,\cdot)$ is $(2,\Lambda)$-Hölder for some $\Lambda\ge 0$ uniformly in $x\in\mathcal{X}$ (Lemma 24). Moreover, note by measurability of ϑ_n and continuity of C_Θ near ϑ^o that c_n is also measurable. By joint weak convergence (20) we infer from Hadamard differentiability of C_Θ at ϑ^o (Lemma 25) using the functional delta method that the one-sample version of (JW) (recall Remark 4(ii)) is fulfilled. Further, since $\vartheta_n \xrightarrow{\mathrm{P}} \vartheta^o$, as n tends to infinity, we infer from Corollary 20 and Lemma 24 that the one-sample version of (Sup) is also met. The assertion now follows at once from Theorem 2. □

C.4. Proof of Lemma 29

First note that the matrix root operation $\mathfrak{R}: \mathrm{SPD}(d) \to \mathrm{SPD}(d), \Sigma \mapsto \Sigma^{1/2}$ [94, p. 134] is Fréchet differentiable with derivative given by $D_{|\Sigma_0}\mathfrak{R}: \mathrm{S}(d) \to \mathrm{S}(d), H \mapsto \tilde H$, and that the inverse operation $\mathfrak{I}: \mathrm{SPD}(d) \to \mathrm{SPD}(d), \Sigma \mapsto \Sigma^{-1}$ is also Fréchet differentiable [95, Theorem 8.3] with derivative given by $D_{|\Sigma_0}\mathfrak{I}: \mathrm{S}(d)\to\mathrm{S}(d), H\mapsto -\Sigma_0 H\Sigma_0$. The assertion now follows at once from the chain rule [95, Theorem 5.12] for compositions of Fréchet differentiable functions. □

C.5. Proof of Corollary 30

Note that by assumption, $(\mathcal{T}, d_{\mathcal{T}})$ and c fulfill the requirements of Theorem 6. Furthermore, note that Assumption (Don) can be established via Proposition 21 and Assumption List **(KP)** is implied by the assumptions on the support of μ and ν [62, Corollary 2]. Hence, the statement follows from Theorem 6. □

C.6. Proof of Corollary 33

Select $\mathcal{X}\subseteq\mathbb{R}^d$ as a compact set which contains the supports of μ and ν. Note that $\mathbb{S}^{d-1}$ is a compact Polish space and consider the Lipschitz map $c_{\mathbb{S}^{d-1}} : (\mathbb{S}^{d-1}, \|\cdot\|) \to C(\mathcal{X}\times\mathcal{X}), \theta\mapsto c_\theta|_{\mathcal{X}\times\mathcal{X}}$ whose modulus depends on $\mathcal{X}$ and p. By compactness of $\mathcal{X}$ and $\mathbb{S}^{d-1}$ it thus follows from the Theorem of Arzelà-Ascoli that $\{c_\theta|_{\mathcal{X}\times\mathcal{X}}\}_{\theta\in\mathbb{S}^{d-1}}$ is uniformly bounded and equicontinuous with a uniform modulus. Therefore, upon choosing the function class $\mathcal{F}$ as in Theorem 5, Assertion (i) follows by Theorem 5 once we verify that

Assumption (Don) is fulfilled. To this end, note that $\log N(\varepsilon, \mathbb{S}^{d-1}, \|\cdot\|) \lesssim |\log(\varepsilon)|$. Moreover, define for $\theta \in \mathbb{S}^{d-1}$ the pseudo metric $\tilde{d}_{\theta,\mathcal{X}}(x, y) = |\theta^T x - \theta^T y|$ on $\mathcal{X}$ which fulfills $\sup_{\theta\in\mathbb{S}^{d-1}} \mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\theta,\mathcal{X}}) \lesssim \varepsilon^{-1}$ and for any $x, x', y \in \mathcal{X}$,

$$\left|c_\theta(x, y) - c_\theta(x', y)\right| \le p \,\mathrm{diam}(\mathfrak{p}_\theta(\mathcal{X}))^{p-1} |\theta^T x - \theta^T x'| \le p \,\mathrm{diam}(\mathcal{X})^{p-1} \tilde{d}_{\theta,\mathcal{X}}(x, x').$$

Since the upper bound for the Lipschitz modulus does not depend on θ, Proposition 21(ii) is applicable, and we conclude that $\bigcup_{\theta\in\mathbb{S}^{d-1}} \mathcal{F}^{c_\theta c_\theta}$ and $\bigcup_{\theta\in\mathbb{S}^{d-1}} \mathcal{F}^{c_\theta}$ are universal Donsker. By applying the continuous mapping theorem [38, Theorem 1.11.1] for the integration operator over $\mathbb{S}^{d-1}$ we obtain Assertion (ii). Finally, Assertion (iii) follows from Theorem 7. $\square$

C.7. Proof of Proposition 34

Since $\mathcal{X} \times \mathcal{Y}$ is compact and by continuity of c and Δ^c there exists a common modulus of continuity w for $\{c_t(\cdot, y)\}_{y\in\mathcal{Y}, t\in[0,1]}$. Hence, for any $t \in [0, 1]$ we have $c, c_t \in C(\|c\|_\infty + \|\Delta^c\|_\infty + 1, w)$ (see Lemma 43 for the definition of $C(\cdot, \cdot)$). Consequently, we infer by Lemma 43 the inequalities,

$$\begin{aligned}
&\frac{1}{t}\left(\inf_{\pi\in\Pi^\star_{c_t}(\mu_t,\nu_t)} \pi(t\Delta^c) + \sup_{f\in S_c(\mu,\nu)} t\Delta^\mu(f^{cc}) + t\Delta^\nu(f^c)\right)\\
\le&\frac{1}{t}(OT(\mu_t, \nu_t, c_t) - OT(\mu, \nu, c))\\
\le&\frac{1}{t}\left(\inf_{\pi\in\Pi^\star_{c}(\mu,\nu)} \pi(t\Delta^c) + \sup_{f\in S_{c_t}(\mu_t,\nu_t)} t\Delta^\mu(f^{cc}) + t\Delta^\nu(f^c) + \sup_{f\in\mathcal{F}} t\Delta^\mu(f^{c_t c_t} - f^{cc}) + t\Delta^\nu(f^{c_t} - f^c)\right).
\end{aligned}$$

Next, we observe that $\Delta^\mu = \tilde{\mu} - \mu$ for some $\tilde{\mu} \in \mathcal{P}(\mathcal{X})$. This yields using Lipschitzianity under cost transformations with respect to the cost function (Lemma 42) that

$$\sup_{f\in\mathcal{F}} |\Delta^\mu(f^{c_t c_t} - f^{cc})| = \sup_{f\in\mathcal{F}} |(\tilde{\mu} - \mu)(f^{c_t c_t} - f^{cc})| \le 4\left\|c_t - c\right\|_\infty = 4t\,\|\Delta^c\|_\infty \xrightarrow{t\to 0} 0.$$

Likewise, it follows that $|\sup_{f\in\mathcal{F}} \Delta^\nu(f^{c_t} - f^c)| \to 0$ for $t \to 0$. Finally, since the pair (μ_t, ν_t) weakly converges for $t \searrow 0$ to (μ, ν) it follows by Lemma 44 that

$$\begin{aligned}
&\liminf_{t\searrow 0} \inf_{\pi\in\Pi^\star_{c_t}(\mu_t,\nu_t)} \pi(\Delta^c) + \sup_{f\in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c)\\
\ge& \inf_{\pi\in\Pi^\star_{c}(\mu,\nu)} \pi(\Delta^c) + \sup_{f\in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c)
\end{aligned}$$

as well as

$$\begin{aligned}
&\limsup_{t\searrow 0} \inf_{\pi\in\Pi^\star_{c}(\mu,\nu)} \pi(\Delta^c) + \sup_{f\in S_{c_t}(\mu_t,\nu_t)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c)\\
\le& \inf_{\pi\in\Pi^\star_{c}(\mu,\nu)} \pi(\Delta^c) + \sup_{f\in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c),
\end{aligned}$$

which yields the claim. $\square$

Appendix D. Proofs for Section 5: Regularity elevation functionals

D.1. Proof of Proposition 36

By the functional delta method [45] and the assumptions on Ψ and $\mathcal{L}$ it follows that

$$a_n\begin{pmatrix}(f_n - f)\\(\Psi(f_n) - f)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathcal{L}\\D^H_f\Psi(\mathcal{L})\end{pmatrix} \overset{d}{=} \begin{pmatrix}\mathcal{L}\\\mathcal{L}\end{pmatrix} \quad \text{for } n \to \infty.$$

The continuous mapping theorem [38, Theorem 1.11.1] in combination with measurability of the random elements f_n and $\Psi(f_n)$ (due to continuity Ψ near f) thus asserts

$$a_n\big(\Psi(f_n) - f_n\big) \xrightarrow{\mathrm{P}} 0 \quad \text{for } n \to \infty. \quad \square$$

D.2. Proof of Proposition 37

First note that $\Psi(\tilde{c}) \in C(\mathcal{X}\times\mathcal{Y})$ for any $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$ as a concatenation of continuous functions and under $\|\tilde{c}\|_\infty < 2$ that $\Psi(\tilde{c}) = \tilde{c}$, which yields $\Psi(c) = c$. In particular, this shows that $\Psi : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y})$ is continuous near c. For Hadamard differentiability at c consider a positive sequence $t_n \searrow 0$ and take a converging sequence $(h_n)_{n\in\mathbb{N}} \subseteq C(\mathcal{X} \times \mathcal{Y})$ with limit h. Since h is bounded and $\|c\|_\infty \le 1$, for n sufficiently large we have $\left\|c + t_n h_n\right\|_\infty < 2$ and therefore $\Psi(c + t_n h_n) = c + t_n h_n$. We then obtain

$$\left\|\frac{\Psi(c + t_n h_n) - \Psi(c)}{t_n} - h\right\|_\infty = \left\|h_n - h\right\|_\infty \to 0.$$

Finally, since for any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ it holds that $\|g^{\Psi(\tilde{c})}\|_\infty \le B + 2$ where $B := \sup_{g\in\mathcal{G}} \|g\|_\infty$ we find for a finite space $\mathcal{X}$ that

$$\sup_{\tilde{c}\in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\epsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \le |\mathcal{X}|(\log(B+2) + |\log(\epsilon)|) \lesssim |\log(\epsilon)|. \quad \square$$

D.3. Proof of Proposition 38

By condition (26) it follows for $x, x' \in \mathcal{X}$ with $\tilde{d}_{\mathcal{X}}(x, x') = 0$ that $c(x, y) = c(x', y)$, whereas under $\tilde{d}_{\mathcal{X}}(x, x') > 0$ we have by $w(\delta) > 0$ for $\delta > 0$ that

$$c(x, y) \le c(x', y) + w(\tilde{d}_{\mathcal{X}}(x, x')) < c(x', y) + 2w(\tilde{d}_{\mathcal{X}}(x, x')).$$

This asserts for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that

$$S(c, (x, y)) := \operatorname*{argmin}_{x'\in\mathcal{X}} c(x', y) + 2w\left(\tilde{d}_{\mathcal{X}}(x, x')\right) = \{x'' \in \mathcal{X} \mid \tilde{d}(x, x'') = 0\},$$

and overall yields by $\|c\|_\infty \le 1$ that $\Psi(c) = c$.

For the second and third claim, recall from Proposition 37 that $\Psi_{\text{bdd}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y})$ is continuous near c and Hadamard differentiable at c with derivative $\mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$. Hence, it suffices to verify that $\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}$ is continuous near c and Hadamard directionally differentiable with $D^H_{|c}\Psi|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$ for which we rely on Lemma 51 and Theorem 52. Define the spaces $V := C(\mathcal{X} \times \mathcal{Y})$, $\mathcal{F} = \mathcal{X}$, $\Theta = \tilde{\mathcal{X}} \times \mathcal{Y}$ and the functional

$$E^{w\circ\tilde{d}_{\mathcal{X}}} : V \times \mathcal{F} \times \Theta = C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{X} \times (\tilde{\mathcal{X}} \times \mathcal{Y}) \mapsto \mathbb{R}, \quad (\tilde{c}, x', (x, y)) \mapsto -\tilde{c}(x', y) - 2w(\tilde{d}_{\mathcal{X}}(x, x')).$$

For any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ the function $E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}, \cdot, \cdot) : \mathcal{X} \times (\tilde{\mathcal{X}} \times \mathcal{Y}) \to \mathbb{R}$ is continuous as a sum of continuous functions. Further, for any $(x', (x, y)) \in \mathcal{X} \times (\tilde{\mathcal{X}} \times \mathcal{Y})$ note that the function $E^{w\circ\tilde{d}_{\mathcal{X}}}(\cdot, x', (x, y)) : C(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ is 1-Lipschitz under uniform norm and that

$$\Delta_c E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}, x', (x, y)) := E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c} + c, x', (x, y)) - E^{w\circ\tilde{d}_{\mathcal{X}}}(c, x', (x, y)) = -\tilde{c}(x', y)$$

is linear in $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$. Hence, by Lemma 51 we obtain continuity of the functional

$$\begin{aligned}
&\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\tilde{\mathcal{X}} \times \mathcal{Y}),\\
&\qquad \tilde{c} \mapsto \left((x, y) \mapsto \inf_{x'\in\mathcal{X}} \tilde{c}(x', y) + 2w(\tilde{d}_{\mathcal{X}}(x, x')) = -\sup_{x'\in\mathcal{X}} E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}, x', (x, y))\right).
\end{aligned}$$

Consider the closed sub-vector space $U := C(\tilde{\mathcal{X}} \times \mathcal{Y}) \subseteq C(\mathcal{X} \times \mathcal{Y})$, cf. Lemma 57. It remains to show Assumption **(DC)** of Theorem 52. To this end, note for $h \in C(\tilde{\mathcal{X}} \times \mathcal{Y})$ that

$$h(\overline{x}, y) + 2w(\tilde{d}_{\mathcal{X}}(\overline{x}, x')) = h(\overline{x}', y) + 2w(\tilde{d}_{\mathcal{X}}(\overline{x}', x')) \quad \text{for any } \overline{x}, \overline{x}' \in S(c, (x, y))$$

since $\tilde{d}_{\mathcal{X}}(\overline{x}, \overline{x}') = 0$. This implies by Lemma 53 that **(DC)** is fulfilled. Theorem 52 thus asserts that $\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}$ is Hadamard directionally differentiable at c with derivative given by

$$\begin{aligned}
&D^H_{|c}\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\tilde{\mathcal{X}} \times \mathcal{Y}),\\
&\qquad h \mapsto \left((x, y) \mapsto \inf_{x' :\, \tilde{d}_{\mathcal{X}}(x', x)=0} h(x', y) = -\sup_{x' :\, \tilde{d}_{\mathcal{X}}(x', x)=0} -\Delta_c E^{w\circ\tilde{d}_{\mathcal{X}}}(h, x', (x, y))\right).
\end{aligned}$$

Hence, if $h \in C(\tilde{\mathcal{X}} \times \mathcal{Y})$, then $D^H_{|c}\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(h) = h$, which yields $D^H_{|c}\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$.

For the last claim note that any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ fulfills for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that

$$-\|\tilde{c}\|_\infty \le \Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x, y) \le \tilde{c}(x, y) \le \|\tilde{c}\|_\infty,$$

and hence $\|\Psi(\tilde{c})\|_\infty = \|\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}} \circ \Psi_{\text{bdd}}(\tilde{c})\|_\infty \le 2$. Further, for any $x, x' \in \mathcal{X}, y \in \mathcal{Y}$ we have

$$\begin{aligned}
\Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x, y) - \Psi_{\text{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x', y) &\le \inf_{x''\in\mathcal{X}} c(x'', y) + 2w(\tilde{d}_{\mathcal{X}}(x'', x)) - c(x'', y) - 2w(\tilde{d}_{\mathcal{X}}(x'', x'))\\
&\le 2w(\tilde{d}_{\mathcal{X}}(x, x')),
\end{aligned} \tag{D.1}$$

where we used the reverse triangle inequality since $w\circ\tilde{d}_{\mathcal{X}}$ defines a (pseudo-)metric on $\mathcal{X}$. We thus conclude for any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ and a bounded function class $\mathcal{G}$ with $B := \sup_{g\in\mathcal{G}} \|g\|_\infty < \infty$ from $\|\Psi(\tilde{c})\|_\infty \le 2$ and (D.1) that the elements of $\mathcal{G}^{\Psi(\tilde{c})}$ are bounded by $B + 2$ and 2-Lipschitz under $w\circ\tilde{d}_{\mathcal{X}}$ as an infimum over such 2-Lipschitz functions. Hence, $\mathcal{G}^{\Psi(\tilde{c})} \subseteq \mathrm{BL}_{(B+2),2}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}})$ where for the latter class uniform metric entropy bounds are available by Kolmogorov and Tikhomirov [96, Section 9], asserting for any $\epsilon > 0$

$$\begin{aligned}
\mathcal{N}(\epsilon, \mathrm{BL}_{(B+2),2}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}}), \|\cdot\|_\infty) &= \mathcal{N}(\epsilon/2, \mathrm{BL}_{(B+2)/2,1}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}}), \|\cdot\|_\infty)\\
&\lesssim \mathcal{N}(\epsilon/8, \mathcal{X}, w\circ\tilde{d}_{\mathcal{X}})|\log(\epsilon)|. \quad \square
\end{aligned}$$

D.4. Proof of Corollary 39

We infer from Proposition 38 that $\Psi := B \cdot \Psi_{\text{mod}}^{w\circ d_{\mathcal{X}}/B} \circ \Psi_{\text{bdd}}(\cdot/B)$ is continuous near c and Hadamard differentiable at c with derivative $D^H_{|c}\Psi = \text{Id}_{C(\mathcal{X}\times\mathcal{Y})}$. Hence, invoking Proposition 36 it follows that $a_n(\overline{c}_n - c_n) \xrightarrow{\text{P}} 0$ for $n \to \infty$. Moreover, by definition of Ψ_{bdd} and $\Psi_{\text{mod}}^{w/B,d_{\mathcal{X}}}$ it follows that $\|\overline{c}_n\|_\infty \leq 2B$ and that $\overline{c}_n$ fulfills (26) with w replaced by $2w$. The inclusion now follows at once from Lemma 1. □

D.5. Proof of Proposition 40

Since c is $(\gamma, 1)$-Hölder it follows for any $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$ as in Lemma A.4 of Hundrieser et al. [43] by convexity of $\mathcal{X}$ that

$$c(x, y) = c(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + R_{x'}(x) \quad \text{with} \quad |R_{x'}(x)| \leq \sqrt{d}\, \|x - x'\|^\gamma,$$

and consequently, for $x \neq x'$ we obtain

$$c(x, y) < c(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + 2\sqrt{d}\, \|x - x'\|^\gamma.$$

This asserts for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that

$$S(c, (x, y)) := \operatorname*{argmin}_{x' \in \mathcal{X}} c(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + 2\sqrt{d}\, \|x - x'\|^\gamma = \{x\}$$

and yields by $\|c\|_\infty \leq 1$ that $\Psi(c) = c$.

To show the claim on continuity and Hadamard differentiability it suffices to verify that Ψ_{Hol} is continuous near c and that is Hadamard differentiable at c with derivative $D^H_{|c}\Psi_{\text{Hol}} = \text{Id}_{C(\mathcal{X}\times\mathcal{Y})}$ for which we rely on Lemma 51 and Theorem 52. Set $V := C(\mathcal{X}\times\mathcal{Y})$, $\mathcal{F} := \mathcal{X}$ and $\Theta := \mathcal{X} \times \mathcal{Y}$ and define the functional

$$\begin{aligned} &E_{\text{Hol}} : V \times \mathcal{F} \times \Theta = C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{X} \times (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}, \\ &\qquad (\tilde{c}, x', (x, y)) \mapsto -\left(\tilde{c}(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + 2\sqrt{d}\, \|x - x'\|^\gamma\right). \end{aligned}$$

For any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ the functional $E_{\text{Hol}}(\tilde{c}, \cdot, \cdot) : \mathcal{X} \times (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ is continuous by continuity of $\nabla_x c(\cdot, \cdot)$ and for any $(x', (x, y)) \in \mathcal{X} \times (\mathcal{X} \times \mathcal{Y})$ the functional $E_{\text{Hol}}(\cdot, x', (x, y)) : C(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ is 1-Lipschitz under uniform norm while

$$\Delta_c E_{\text{Hol}}(\tilde{c}, x', (x, y)) = E_{\text{Hol}}(\tilde{c} + c, x', (x, y)) - E_{\text{Hol}}(c, x', (x, y)) = -\tilde{c}(x', y)$$

is linear in $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$. Finally, condition **(DC)** follows by Lemma 53 since $S(c, (x, y)) = \{x\}$ is a singleton. Hence, by Lemma 51 and Theorem 52 the functional

$$\Psi_{\text{Hol}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}),\ \tilde{c} \mapsto \left((x, y) \mapsto -\sup_{x' \in \mathcal{X}} E_{\text{Hol}}(\tilde{c}, x', (x, y))\right)$$

is continuous near c and Hadamard differentiable at c with derivative

$$D^H_{|c}\Psi_{\text{Hol}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\tilde{\mathcal{X}} \times \mathcal{Y}), \quad h \mapsto \Big((x, y) \mapsto h(x, y) = -\Delta_c E_{\text{Hol}}(h, x', (x, y))\Big).$$

For the claim on the uniform metric entropy bound let $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$, and assume (after application of Ψ_{bdd}) that $\|\tilde{c}\|_\infty \leq 2$. Define the collection of functions $(\tilde{E}_{x',y})_{x' \in \mathcal{X}, y \in \mathcal{Y}}$ with

$$\tilde{E}_{x',y} : \mathcal{X} \to \mathbb{R}, \quad x \mapsto \tilde{c}(x', y) + \langle \nabla_x c(x', y), x - x' \rangle + 2\sqrt{d}\, \|x - x'\|^\gamma,$$

which is $(\gamma, 2)$-Hölder on $\mathcal{X}$. Hence, by Hundrieser et al. [43, Lemma A.5] there exists another collection $(\tilde{E}^\sigma_{x',y})_{x' \in \mathcal{X}, y \in \mathcal{Y}, \sigma \in (0,1]}$ of smooth functions on $\mathcal{X}$ such that

$$\sup_{\substack{x' \in \mathcal{X} \\ y \in \mathcal{Y}}} \left\|\tilde{E}_{x',y} - \tilde{E}^\sigma_{x',y}\right\|_\infty \leq K\sigma^\gamma \quad \text{and} \quad \sup_{\substack{x' \in \mathcal{X} \\ y \in \mathcal{Y}}} \left\|\tilde{E}^\sigma_{x',y}\right\|_{C^2(\mathcal{X})} \leq K\sigma^{\gamma - 2}, \tag{D.2}$$

for all $\sigma > 0$ and some independent $K > 0$. Here, the $C^2(\mathcal{X})$-norm of a twice continuously differentiable function $g : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\|g\|_{C^2(\mathcal{X})} := \max_{|\beta| \leq 2} \left\|D^\beta g\right\|_\infty, \quad \text{where} \quad D^\beta g = \partial^{|\beta|} g / \partial x_1^{\beta_1} \cdots x_d^{\beta_d} \ \text{ for } \beta \in \mathbb{N}_0^d.$$

Note that a function with $\|g\|_{C^2(\mathcal{X})} \leq \Gamma$ for $\Gamma > 0$ is absolutely bounded by Γ, it is Γ-Lipschitz, and $d\Gamma$-semi-concave (for a formal definition see [97] or [43]), since the Eigenvalues of its Hessian are bounded by $d \cdot \Gamma$. Upon defining $\overline{c}(x, y) := \Psi(\tilde{c})(x, y) = \inf_{x' \in \mathcal{X}} \tilde{E}_{x',y}(x)$ and $\overline{c}^\sigma(x, y) := \inf_{x' \in \mathcal{X}} \tilde{E}^\sigma_{x',y}(x)$ we thus obtain from (D.2) that $\|\overline{c} - \overline{c}^\sigma\|_\infty \leq K\sigma^\gamma$ and that $\overline{c}^\sigma$ is semi-concave of order $\Gamma(\sigma) := dK\sigma^{\gamma-2}$. Hence, following along the lines the proofs of Lemma A.4 in [43] we obtain for any $\varepsilon > 0$ and $\sigma(\varepsilon) := (\varepsilon/4K)^{1/\gamma}$ that

$$\mathcal{N}(\varepsilon, \mathcal{G}^{\overline{c}}, \|\cdot\|_\infty) \leq \mathcal{N}(\varepsilon/2, \mathcal{G}^{\overline{c}^{\sigma(\varepsilon)}}, \|\cdot\|_\infty) = \mathcal{N}\left(\frac{\varepsilon}{2\Gamma(\sigma(\varepsilon))}, \frac{\mathcal{G}^{\overline{c}^{\sigma(\varepsilon)}}}{2\Gamma(\sigma(\varepsilon))}, \|\cdot\|_\infty\right) \lesssim \left(\frac{\varepsilon}{2\Gamma(\sigma(\varepsilon))}\right)^{-d/2} \lesssim \varepsilon^{-d/\gamma}.$$

Here, we used in the second inequality that $\mathcal{G}^{\bar{c}^{\sigma(\varepsilon)}}/2\Gamma(\sigma(\varepsilon))$ is contained in the collection of functions on $\mathcal{X}$ which are absolutely bounded by $B \geq 0$, Lipschitz with modulus $L \geq 0$ and 1-semi-concave, where B depends on $\mathcal{G}$ and L depends on $\mathcal{X}$, in conjunction with uniform metric entropy bounds by Bronshtein [98],Guntuboyina and Sen [99] for convex functions. In particular, since the hidden constants do not depend on $\tilde{c}$, the claim follows. □

D.6. Proof of Proposition 41

For the first claim note that $\Psi_i(c_i) = c_i$ for each $i \in \{1, \dots, I\}$ and consequently, it follows for $x \in \zeta_i(\mathcal{U}_i)$ that $\Psi_i(c_i)(\zeta_i^{-1}(x), y) = c(x, y)$. Hence, since $\sum_{i=1}^I \eta_i(x) \equiv 1$ it follows that $\Psi(c) = c$.

The claim on continuity of Ψ near c follows by continuity of the functionals $\Psi_i : C(\mathcal{U}_i \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y})$ near c_i for each $1 \leq i \leq I$.

For the claim on Hadamard differentiability of Ψ define for each $i \in \{1, \dots, I\}$ the functionals

$$\begin{aligned}
&\Psi^1_{\mathrm{com},i} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y}), && \tilde{c} \mapsto \big((u, y) \mapsto \tilde{c}(\zeta_i(u), y)\big),\\
&\Psi^2_{\mathrm{com},i} : C(\mathcal{U}_i \times \mathcal{Y}) \to C(\zeta_i(\mathcal{U}_i) \times \mathcal{Y}), && \tilde{c} \mapsto \big((x, y) \mapsto \tilde{c}(\zeta_i^{-1}(x), y)\big),
\end{aligned}$$

where both maps assign to the respective spaces of continuous functions since ζ_i^{-1} and ζ_i are both continuous. Further, note for any $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ that $\Psi(\tilde{c}) = \sum_{i=1}^I \eta_i \cdot \Psi^2_{\mathrm{com},i} \circ \Psi_i \circ \Psi^1_{\mathrm{com},i}(\tilde{c})$. Both functionals $\Psi^1_{\mathrm{com},i}$ and $\Psi^2_{\mathrm{com},i}$ are Hadamard differentiable at c_i with derivative

$$\begin{aligned}
&D^H_{|c_i}\Psi^1_{\mathrm{com},i} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y}), \quad h \mapsto \big((u, y) \mapsto h(\zeta_i(u), y)\big),\\
&D^H_{|c_i}\Psi^2_{\mathrm{com},i} : C(\mathcal{U}_i \times \mathcal{Y}) \to C(\zeta_i(\mathcal{U}_i) \times \mathcal{Y}), \quad h \mapsto \big((x, y) \mapsto h(\zeta_i^{-1}(x), y)\big).
\end{aligned}$$

By assumption on Ψ_i and chain rule we infer that Ψ is Hadamard differentiable at c with derivative

$$\begin{aligned}
&D^H_c\Psi : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}),\\
&\qquad h \mapsto \left((x, y) \mapsto \sum_{i=1}^I \eta_i(x) h(\zeta_i^{-1}(\zeta_i(x)), y) = \sum_{i=1}^I \eta_i(x) h(x, y) = h(x, y)\right)
\end{aligned}$$

and conclude that $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$.

Finally, the bound on the covering numbers is a consequence of Lemma 3.1 and Lemma A.1 in [43] as they assert for arbitrary $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ that

$$\begin{aligned}
\log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) &\leq \sum_{i=1}^I \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}|_{\zeta_i(\mathcal{U}_i)}, \|\cdot\|_\infty)\\
&\leq \sum_{i=1}^I \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c})} \circ \zeta_i, \|\cdot\|_\infty)\\
&= \sum_{i=1}^I \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c}(\zeta_i(\cdot), \cdot))}, \|\cdot\|_\infty). \quad \square
\end{aligned}$$

Appendix E. Proofs for Section 6: Lemmata of distributional limits

E.1. Proof of Lemma 42

Assume $\|f - \tilde{f}\|_\infty + \|c - \tilde{c}\|_\infty < \infty$ since otherwise the claim is vacuous. For $\tilde{f}$ and $\tilde{c}$ there exists for $y \in \mathcal{Y}$ and $\varepsilon > 0$ some $x' \in \mathcal{X}$ such that $\tilde{f}^{\tilde{c}}(y) \geq \tilde{c}(x', y) - \tilde{f}(x') - \varepsilon$. Hence,

$$\begin{aligned}
f^c(y) - \tilde{f}^{\tilde{c}}(y) &= \left[\inf_{x\in\mathcal{X}} c(x, y) - f(x)\right] - \left[\inf_{x\in\mathcal{X}} \tilde{c}(x, y) - \tilde{f}(x)\right]\\
&\leq c(x', y) - f(x') - \tilde{c}(x', y) + \tilde{f}(x') + \varepsilon\\
&\leq \|f - \tilde{f}\|_\infty + \|c - \tilde{c}\|_\infty + \varepsilon.
\end{aligned}$$

As $\varepsilon > 0$ can be chosen arbitrarily small, we obtain for any $y \in \mathcal{Y}$ the inequality

$$f^c(y) - \tilde{f}^{\tilde{c}}(y) \leq \|f - \tilde{f}\|_\infty + \|c - \tilde{c}\|_\infty.$$

Repeating the argument for f and c asserts the converse inequality and proves the claim. □

E.2. Proof of Lemma 43

Let us start by splitting the problem in two different ways,

$$\begin{aligned}
OT(\tilde{\mu}, \tilde{\nu}, \tilde{c}) - OT(\mu, \nu, c) &= (OT(\tilde{\mu}, \tilde{\nu}, \tilde{c}) - OT(\tilde{\mu}, \tilde{\nu}, c)) + (OT(\tilde{\mu}, \tilde{\nu}, c) - OT(\mu, \nu, c))\\
&= (OT(\tilde{\mu}, \tilde{\nu}, \tilde{c}) - OT(\mu, \nu, \tilde{c})) + (OT(\mu, \nu, \tilde{c}) - OT(\mu, \nu, c)).
\end{aligned}$$

Since $c, \tilde{c} \in C(2\|c\|_\infty + 1, 2w)$, we can employ the dual representation of the OT value from Lemma 1 with $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$. Hence, for each bracket in the display above, one can choose to plug-in a feasible plan in the primal formulation or a potential from $\mathcal{F}$ in the dual formulation to obtain upper and lower bounds. Doing so, we obtain

$$\begin{aligned}\inf_{\pi \in \Pi^\star_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c} - c) \le OT(\tilde{\mu}, \tilde{\nu}, \tilde{c}) - OT(\tilde{\mu}, \tilde{\nu}, c) &\le \inf_{\pi \in \Pi^\star_{c}(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c} - c),\\ \sup_{f \in S_c(\mu,\nu)} (\tilde{\mu} - \mu) f^{cc} + (\tilde{\nu} - \nu) f^c \le OT(\tilde{\mu}, \tilde{\nu}, c) - OT(\mu, \nu, c) &\le \sup_{f \in S_c(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu) f^{cc} + (\tilde{\nu} - \nu) f^c,\\ OT(\mu, \nu, \tilde{c}) - OT(\mu, \nu, c) &\le \inf_{\pi \in \Pi^\star_{c}(\mu,\nu)} \pi(\tilde{c} - c),\\ OT(\tilde{\mu}, \tilde{\nu}, \tilde{c}) - OT(\mu, \nu, \tilde{c}) &\le \sup_{f \in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu) f^{\tilde{c}\tilde{c}} + (\tilde{\nu} - \nu) f^{\tilde{c}}.\end{aligned}$$

In particular, for the last upper bound we further note that

$$\sup_{f \in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu) f^{\tilde{c}\tilde{c}} + (\tilde{\nu} - \nu) f^{\tilde{c}} \le \sup_{f \in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu) f^{cc} + (\tilde{\nu} - \nu) f^c + \sup_{f \in \mathcal{F}} (\tilde{\mu} - \mu)(f^{\tilde{c}\tilde{c}} - f^{cc}) + (\tilde{\nu} - \nu)(f^{\tilde{c}} - f^c),$$

which overall yields the lower and upper bounds for the OT cost under varying measures and costs.

Finally, the bound under fixed measures μ, ν it follows by Hölder's inequality for any $\pi \in \Pi(\mu, \nu)$ that $|\pi(\tilde{c} - c)| \le \|\tilde{c} - c\|_\infty$, whereas under a fixed cost function c we have

$$\begin{aligned}\sup_{f \in S_c(\tilde{\mu},\tilde{\nu}) \cup S_c(\mu,\nu)} \left|(\tilde{\mu} - \mu) f^{cc} + (\tilde{\nu} - \nu) f^c\right| &\le \sup_{f \in \mathcal{F}} \left|(\tilde{\mu} - \mu) f^{cc}\right| + \sup_{f \in \mathcal{F}} \left|(\tilde{\nu} - \nu) f^c\right|\\ &= \sup_{f \in \mathcal{F}^{cc}} \left|(\tilde{\mu} - \mu) f\right| + \sup_{f \in \mathcal{F}^{c}} \left|(\tilde{\nu} - \nu) f\right|. \quad \square\end{aligned}$$

E.3. Proof of Lemma 44

The continuity of T_1 is a consequence of Villani [3, Theorem 5.20]. Indeed, any converging sequence (μ_n, ν_n, c_n) with limit $(\mu_\infty, \nu_\infty, c_\infty)$ admits a sequence of OT plans $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$ which converges weakly along a subsequence, say $(\pi_{n_k})_{k\in\mathbb{N}}$, to an OT plan $\pi_\infty \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$. Hence,

$$\begin{aligned}\limsup_{k\to\infty} |T_1(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - T_1(\mu_\infty, \nu_\infty, c_\infty)| &= \limsup_{k\to\infty} |OT(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - OT(\mu_\infty, \nu_\infty, c_\infty)|\\ &= \limsup_{k\to\infty} |\pi_{n_k}(c_{n_k}) - \pi_\infty(c_\infty)|\\ &\le \limsup_{k\to\infty} |(\pi_{n_k} - \pi_\infty)(c_\infty)| + \left\|c_{n_k} - c_\infty\right\|_\infty = 0.\end{aligned}$$

Since this holds for any sequence of converging OT plans, continuity of T_1 follows from Lemma 56.

For the lower semi-continuity of T_2 take a sequence $(h_{c,n})_{n\in\mathbb{N}}$ with limit $h_{c,\infty}$ and consider OT plans $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$ such that

$$\inf_{\pi \in \Pi^\star_{c_n}(\mu_n,\nu_n)} \pi(h_{c,n}) \ge \pi_n(h_{c,n}) - 1/n.$$

Then, by Villani [3, Theorem 5.20] a converging subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ with limit $\pi_\infty \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$ exists, and it follows that

$$\begin{aligned}\liminf_{k\to\infty} T_2(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{c,n_k}) &= \liminf_{k\to\infty} \inf_{\pi \in \Pi^\star_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} \pi(h_{c,n_k})\\ &\ge \liminf_{k\to\infty} \pi_{n_k}(h_{c,n_k}) - 1/n_k\\ &\ge \liminf_{k\to\infty} \pi_{n_k}(h_{c,\infty}) - \left\|h_{c,\infty} - h_{c,n_k}\right\|_\infty - 1/n_k\\ &= \pi_\infty(h_{c,\infty}) \ge T_2(\mu_\infty, \nu_\infty, c_\infty, h_{c,\infty}).\end{aligned}$$

Consequently, by Lemma 56, lower semi-continuity of T_2 follows. To infer upper semi-continuity of T_2, and thus continuity, at $(\mu_\infty, \nu_\infty, c_\infty, h_{c,\infty})$ under the assumption of a unique OT plan $\pi^\star \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$ note by Villani [3, Theorem 5.20] that for any sequence of OT plans $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$ there exists a weakly converging subsequence π_{n_k} which tends to $\pi^\star$ for $k \to \infty$. Hence, we conclude that

$$\begin{aligned}\limsup_{k\to\infty} T_2(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{c,n_k}) &= \limsup_{k\to\infty} \inf_{\pi \in \Pi^\star_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} \pi(h_{c,n_k})\\ &\le \limsup_{k\to\infty} \pi_{n_k}(h_{c,n_k})\\ &\le \limsup_{k\to\infty} \pi_{n_k}(h_{c,\infty}) - \left\|h_{c,\infty} - h_{c,n_k}\right\|_\infty\\ &= \pi_\infty(h_{c,\infty}) = T_2(\mu_\infty, \nu_\infty, c_\infty, h_{c,\infty}).\end{aligned}$$

This implies by Lemma 56 the upper semi-continuity of T_2. Moreover, for fixed (μ', ν', c') the map T_2 is continuous in h_c since for any $\tilde{h}_c$ it holds that

$$|T_2(\mu, \nu, c, h_c) - T_2(\mu, \nu, c, \tilde{h}_c)| \le \|h_c - \tilde{h}_c\|_\infty .$$

To show upper semi-continuity of T_3 take a sequence $(h_{\mu,n}, h_{\nu,n})_{n\in\mathbb{N}}$ with limit $(h_{\mu,\infty}, h_{\nu,\infty})$. Further, by definition of C, it follows from Lemma 1 that any $c_n \in C$ fulfills $\mathcal{H}_{c_n} \subseteq \mathcal{F}^{c_n c_n} \subseteq \mathcal{F}$. Take a sequence $f_n \in S_{c_n}(\mu_n, \nu_n) \subseteq \mathcal{F}$ such that

$$T_3(\mu_n, \nu_n, c_n, h_{\mu,n}, h_{\nu,n}) \le h_\mu(f_n) + h_\nu(f_n) + 1/n.$$

By compactness of $\mathcal{F}$ there exists a uniformly converging subsequence, say $(f_{n_k})_{k\in\mathbb{N}}$, with limit $f_\infty \in \mathcal{F}$. Next, we demonstrate that $f_\infty \in S_{c_\infty}(\mu_\infty, \nu_\infty)$. To this end, we note

$$\begin{aligned}
OT(\mu_\infty, \nu_\infty, c_\infty) &\ge \mu_\infty(f_\infty^{c_\infty c_\infty}) + \nu_\infty(f_\infty^{c_\infty})\\
&= \lim_{k\to\infty} \mu_{n_k}(f_\infty^{c_\infty c_\infty}) + \nu_{n_k}(f_\infty^{c_\infty})\\
&\ge \lim_{k\to\infty} \mu_{n_k}(f_{n_k}^{c_{n_k} c_{n_k}}) + \nu_{n_k}(f_{n_k}^{c_{n_k}}) - \left\| f_\infty^{c_\infty c_\infty} - f_{n_k}^{c_{n_k} c_{n_k}} \right\|_\infty - \left\| f_\infty^{c_\infty} - f_{n_k}^{c_{n_k}} \right\|_\infty\\
&= \lim_{k\to\infty} OT(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - \left\| f_\infty^{c_\infty c_\infty} - f_{n_k}^{c_{n_k} c_{n_k}} \right\|_\infty - \left\| f_\infty^{c_\infty} - f_{n_k}^{c_{n_k}} \right\|_\infty\\
&= OT(\mu_\infty, \nu_\infty, c_\infty),
\end{aligned}$$

where the last equality follows by continuity of T_1. Hence, we get $f_\infty \in S_{c_\infty}(\mu_\infty, \nu_\infty)$.

By continuity of h_μ and h_ν on $\mathcal{F}$ and upon denoting the norm on $C_u(\mathcal{F})$ by $\|\cdot\|_{\mathcal{F}}$, we infer that

$$\begin{aligned}
\limsup_{k\to\infty} T_3(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{\mu,n_k}, h_{\nu,n_k}) &= \limsup_{k\to\infty} \sup_{f\in S_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} h_{\mu,n_k}(f) + h_{\nu,n_k}(f)\\
&\le \limsup_{k\to\infty} h_{\mu,n_k}(f_{n_k}) + h_{\nu,n_k}(f_{n_k}) + 1/n_k\\
&\le \limsup_{k\to\infty} h_{\mu,\infty}(f_{n_k}) + h_{\nu,\infty}(f_{n_k}) + \left\| h_{\mu,\infty} - h_{\mu,n_k} \right\|_{\mathcal{F}} + \left\| h_{\nu,\infty} - h_{\nu,n_k} \right\|_{\mathcal{F}} + 1/n_k\\
&= h_{\mu,\infty}(f_\infty) + h_{\nu,\infty}(f_\infty) \le T_3(\mu_\infty, \nu_\infty, c_\infty, h_{\mu,\infty}, h_{\nu,\infty})
\end{aligned}$$

and consequently, by Lemma 56, upper semi-continuity of T_3 follows. Further, for fixed (μ', ν', c') the map T_3 is continuous in (h_μ, h_ν) since for another $(\tilde{h}_\mu, \tilde{h}_\nu)$ it holds that

$$|T_3(\mu, \nu, c, h_\mu, h_\nu) - T_3(\mu, \nu, c, \tilde{h}_\mu, \tilde{h}_\nu)| \le \left\| \tilde{h}_\mu - \tilde{h}_\mu \right\|_{\mathcal{F}^{cc}} + \| \tilde{h}_\nu - \tilde{h}_\nu \|_{\mathcal{F}^c}.$$

Finally, for T_4 take $(h_{1,\mu}, \tilde{h}_{1,\mu}, h_{1,\nu}, \tilde{h}_{1,\nu})$, $(h_{2,\mu}, \tilde{h}_{2,\mu}, h_{2,\nu}, \tilde{h}_{2,\nu}) \in C_u(\mathcal{F})^4$ and note that

$$\begin{aligned}
&|T_4(h_{1,\mu}, \tilde{h}_{1,\mu}, h_{1,\nu}, \tilde{h}_{1,\nu}) - T_4(h_{2,\mu}, \tilde{h}_{2,\mu}, h_{2,\nu}, \tilde{h}_{2,\nu})|\\
&\le \left\| h_{1,\mu} - h_{2,\mu} \right\|_{\mathcal{F}} + \left\| \tilde{h}_{1,\mu} - \tilde{h}_{2,\mu} \right\|_{\mathcal{F}} + \| h_{1,\nu} - h_{2,\nu} \|_{\mathcal{F}} + \| \tilde{h}_{1,\nu} - \tilde{h}_{2,\nu} \|_{\mathcal{F}},
\end{aligned}$$

which asserts continuity. □

E.4. Proof of Lemma 45

For (i) take $f, g \in \mathcal{G}$, then $|\mu(f) - \mu(g)| \le \|f - g\|_\infty$ and hence $\mu : \mathcal{G} \to \mathbb{R}$ defines a Lipschitz map under uniform norm which asserts $\mu \in C_u(\mathcal{G})$. Assertion (ii) follows from Giné and Nickl [100, p. 17]. Finally, (iii) follows from (ii) since for any $g \in \mathcal{G}$ the evaluations $\mu_n(g) = n^{-1}\sum_{i=1}^n g(X_i)$ and $\mu_{n,k}^b(g) = k^{-1}\sum_{i=1}^k g(X_i^b)$ are Borel measurable. □

E.5. Proof of Lemma 46

We first prove that Assumption List (JW) implies for $n, m \to \infty$ with $m/(n+m) \to \lambda \in (0,1)$ that

$$\begin{pmatrix} \sqrt{n}\Big((\mu_n - \mu)(f^{cc})\Big)_{f\in\mathcal{F}} \\ \sqrt{m}\Big((\nu_m - \nu)(f^{c})\Big)_{f\in\mathcal{F}} \\ \sqrt{\frac{nm}{n+m}}(c_{n,m} - c) \end{pmatrix} = \begin{pmatrix} \Big(\mathbb{G}_n^\mu(f^{cc})\Big)_{f\in\mathcal{F}} \\ \Big(\mathbb{G}_m^\nu(f^{c})\Big)_{f\in\mathcal{F}} \\ \mathbb{G}_{n,m}^c \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Big(\mathbb{G}^\mu(f^{cc})\Big)_{f\in\mathcal{F}} \\ \Big(\mathbb{G}^\nu(f^{c})\Big)_{f\in\mathcal{F}} \\ \mathbb{G}^c \end{pmatrix} \tag{E.1}$$

in the Polish space $C_u(\mathcal{F}) \times C_u(\mathcal{F}) \times C(\mathcal{X} \times \mathcal{Y})$. To this end, consider the map

$$\begin{aligned}
\Psi : C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^c) \times C(\mathcal{X}\times\mathcal{Y}) &\to C_u(\mathcal{F}) \times C_u(\mathcal{F}) \times C(\mathcal{X}\times\mathcal{Y}),\\
(\alpha, \beta, \gamma) &\mapsto \left(\big(\alpha(f^{cc})\big)_{f\in\mathcal{F}}, \big(\beta(f^c)\big)_{f\in\mathcal{F}}, \gamma \right).
\end{aligned}$$

This map is well-defined (*i.e.*, its range is correct) since for any $(\alpha, \beta) \in C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^c)$ there exist moduli of continuity $w_\alpha, w_\beta : \mathbb{R}_+ \to \mathbb{R}_+$ such that for $f, \tilde{f} \in \mathcal{F}$ it follows by Lemma 42 that

$$\begin{aligned}
|\alpha(f^{cc}) - \alpha(\tilde{f}^{cc})| &\le w_\alpha(\|f^{cc} - \tilde{f}^{cc}\|_\infty) \le w_\alpha(\|f - \tilde{f}\|_\infty),\\
|\beta(f^{c}) - \beta(\tilde{f}^{c})| &\le w_\beta(\|f^{c} - \tilde{f}^{c}\|_\infty) \le w_\beta(\|f - \tilde{f}\|_\infty),
\end{aligned}$$

which assert that $((\alpha(f^{cc}))_{f\in\mathcal{F}}, (\beta(f^c))_{f\in\mathcal{F}}, \gamma) \in C_u(\mathcal{F})\times C_u(\mathcal{F})\times C(\mathcal{X}\times\mathcal{Y})$. Moreover, for any $(\alpha,\beta),(\tilde\alpha,\tilde\beta)\in C_u(\mathcal{F}^{cc})\times C_u(\mathcal{F}^c)$ we have

$$\sup_{f\in\mathcal{F}}|\alpha(f^{cc})-\tilde\alpha(f^{cc})| = \sup_{\tilde f\in\mathcal{F}^{cc}}|\alpha(\tilde f)-\tilde\alpha(\tilde f)| \quad\text{and}\quad \sup_{f\in\mathcal{F}}|\beta(f^{c})-\tilde\beta(f^{c})| = \sup_{\tilde f\in\mathcal{F}^{c}}|\beta(\tilde f)-\tilde\beta(\tilde f)|,$$

hence the map Ψ is continuous. Consequently, Assumption List **(JW)** and the continuous mapping theorem [38, Theorem 1.11.1] assert weak convergence (E.1).

Moreover, by Varadarajan [92] the empirical measures (μ_n,ν_n) weakly converge a.s. in $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ to (μ,ν). Note that $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ is by compactness of $\mathcal{X}$ and $\mathcal{Y}$ a separable, complete metric space [101]. Invoking Slutzky's lemma [38, Example 1.4.7] in conjunction with (E.1) we thus obtain the first claim. In particular, by measurability of $c_{n,m}$ and Lemma 45, all involved quantities are Borel measurable.

For the second claim note by Lemma 42 that any realization of μ_n,ν_m and $c_{n,m}$ leads the processes $\mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}})$ and $\mathbb{G}_m^\nu(f^{c_{n,m}})$ to be $2\sqrt{n}$-Lipschitz and $2\sqrt{m}$-Lipschitz in f, respectively. Thus, they are uniformly continuous in f. Moreover, for fixed $f\in\mathcal{F}$ we can show that the function

$$\tilde{\mathbb{G}}_n^\mu : \mathcal{P}(\mathcal{X})\times C(\mathcal{X}\times\mathcal{Y})\to\mathbb{R},\quad (\tilde\mu,\tilde c)\mapsto\sqrt{n}(\tilde\mu-\mu)(f^{\tilde c\tilde c})$$

is upper semi-continuous (*i.e.*, in particular measurable). Indeed, for $\tilde\mu_k \rightsquigarrow \tilde\mu$ in $\mathcal{P}(\mathcal{X})$ and $\tilde c_k\to\tilde c$ in $C(\mathcal{X}\times\mathcal{Y})$ it follows by Lemma 42, upper semi-continuity of $f^{\tilde c\tilde c}$ and the Portmanteau Theorem [38, Theorem 1.3.4] that

$$\limsup_{k\to\infty}\sqrt{n}(\tilde\mu_k-\mu)(f^{\tilde c_k\tilde c_k}) \le \limsup_{k\to\infty}\sqrt{n}(\tilde\mu_k-\mu)(f^{\tilde c\tilde c}) + 2\sqrt{n}\left\|f^{\tilde c_k\tilde c_k}-f^{\tilde c\tilde c}\right\|_\infty \le \sqrt{n}(\tilde\mu-\mu)(f^{\tilde c\tilde c}).$$

Hence, by Lemma 45(ii) we conclude that $(\mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}}))_{f\in\mathcal{F}}$ is Borel measurable. Likewise, we conclude $(\mathbb{G}_m^\nu(f^{c_{n,m}}))_{f\in\mathcal{F}}$ is Borel measurable.

Consequently, by (Sup) we infer, for $n,m\to\infty$, that

$$\left(\mathbb{G}_n^\mu(f^{cc})-\mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}}),\mathbb{G}_m^\nu(f^{c})-\mathbb{G}_m^\nu(f^{c_n})\right)_{f\in\mathcal{F}}\xrightarrow{\mathrm{P}}(0,0)\quad\text{in } C_u(\mathcal{F})^2.$$

The claim now follows by a combination of Slutzky's lemma and the continuous mapping theorem [38, Example 1.4.7, Theorem 1.11.1]. □

E.6. Proof of Lemma 49

The first claim follows by an observation in Römisch [45] since the set of probability measures $\mathcal{P}(\mathcal{X})$ is convex. For additional insights see Aubin and Frankowska [102, Proposition 4.2.1]

For the second claim consider a sequence $\Delta_n=(\tilde\mu_n-\mu)/t_n$ with $t_n>0$ and $\tilde\mu_n\in\mathcal{P}(\mathcal{X})$ such that $\|\Delta_n-\Delta\|_{\tilde{\mathcal{F}}}=\sup_{f\in\tilde{\mathcal{F}}}|\Delta_n(f)-\Delta(f)|\to 0$. Then, it follows from triangle inequality that

$$\begin{aligned}|\Delta(f)-\Delta(f')| &= |\Delta_n(f)-\Delta_n(f')+(\Delta-\Delta_n)(f)+(\Delta-\Delta_n)(f')|\\ &\le |\Delta_n(f-f')|+2\|\Delta-\Delta_n\|_{\tilde{\mathcal{F}}}\end{aligned}$$

Herein, the first term vanishes since $\Delta_n(f-f')=(\tilde\mu_n-\mu)(\kappa)/t_n=0$, whereas the second term converges for $n\to\infty$ to zero. Hence, $\Delta(f)=\Delta(f')$.

The third claim relies on Portemanteau's theorem [38, Lemma 1.3.4] which asserts using the notion of outer probabilities P^* that

$$\mathbb{P}\left(\mathbb{G}^\mu\in T_\mu\mathcal{P}(\mathcal{X})\right)\ge\limsup_{n\to\infty}\mathrm{P}^*\left(\sqrt{n}(\mu_n-\mu)\in T_\mu\mathcal{P}(\mathcal{X})\right)=1.\quad\square$$

E.7. Proof of Lemma 55

We start by proving (i). Note for $\kappa\in\mathbb{R}$ that

$$(f+\kappa)^c(y)=\inf_{x\in\mathcal{X}}c(x,y)-f(x)-\kappa=f^c(y)-\kappa,$$

which yields the claim. To show assertion (ii), observe by Lemma 42 that

$$\left\|g^{(c+\Delta^c)(c+\Delta^c)}\right\|_\infty\le\|g\|_\infty+2\|c+\Delta^c\|_\infty\le B. \tag{E.2}$$

Further, we find that

$$-\|c\|_\infty-\sup_{x\in\mathcal{X}}g^{(c+\Delta^c)(c+\Delta^c)}(x)\ \le g^{(c+\Delta^c)(c+\Delta^c)c}(y)\le\|c\|_\infty-\sup_{x\in\mathcal{X}}g^{(c+\Delta^c)(c+\Delta^c)}(x). \tag{E.3}$$

Using part (i) of this lemma, we obtain

$$g^{(c+\Delta^c)(c+\Delta^c)cc}=\left(\left(g^{(c+\Delta^c)(c+\Delta^c)}\right)^c+\sup_{x\in\mathcal{X}}g^{(c+\Delta^c)(c+\Delta^c)}(x)\right)^c+\sup_{x\in\mathcal{X}}g^{(c+\Delta^c)(c+\Delta^c)}(x).$$

Combining (E.2) and (E.3) with the above equation demonstrates that $g^{(c+\Delta^c)(c+\Delta^c)cc}\in\mathcal{H}_c+[-B,B]$ and hence yields the claim. □

Appendix F. Elementary analytical results

Lemma 56. *Consider a real-valued sequence $(a_n)_{n\in\mathbb{N}}$ and let $K \in \mathbb{R}$.*

(i) If for any subsequence $(a_{n_k})_{k\in\mathbb{N}}$ there exists a subsequence $(a_{n_{k_l}})_{l\in\mathbb{N}}$ with $\limsup_{l\to\infty} a_{n_{k_l}} \leq K$, then it follows that $\limsup_{n\to\infty} a_n \leq K$.

(ii) If for any subsequence $(a_{n_k})_{k\in\mathbb{N}}$ there exists a subsequence $(a_{n_{k_l}})_{l\in\mathbb{N}}$ with $\liminf_{l\to\infty} a_{n_{k_l}} \geq K$, then it follows that $\liminf_{n\to\infty} a_n \geq K$.

(iii) If for any subsequence $(a_{n_k})_{k\in\mathbb{N}}$ there exists a subsequence $(a_{n_{k_l}})_{l\in\mathbb{N}}$ with $\lim_{l\to\infty} a_{n_{k_l}} = K$, then it follows that $\lim_{n\to\infty} a_n = K$.

Proof. We only prove (i) and note that (ii) and (iii) can be shown analogously. Assume that $\limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}}(\sup_{m\geq n} a_m) \geq K+\varepsilon$ for some $\varepsilon > 0$. Since $(\sup_{m\geq n} a_m)_{n\in\mathbb{N}}$ is decreasing in n, this would imply that $\sup_{m\geq n} a_m \geq K+\varepsilon$ for all $n \in \mathbb{N}$. Hence, there would exist a subsequence of $(a_n)_{n\in\mathbb{N}}$, say $(a_{n_l})_{l\in\mathbb{N}}$, with $a_{n_l} \geq K+\varepsilon/2$ for all $l \in \mathbb{N}$. However, this would assert $\liminf_{l\to\infty} a_{n_l} \geq K+\varepsilon/2 > K$, contradicting the assumption. Thus, $\limsup_{n\to\infty} a_n \leq K$. □

Lemma 57. *Let $(\mathcal{X}, d_\mathcal{X})$ be a compact metric space and consider a continuous (pseudo-)metric $\tilde{d}_\mathcal{X}$ on $\mathcal{X}$. Then, $(\mathcal{X}, \tilde{d}_\mathcal{X})$ is a compact (pseudo-)metric space. Moreover, given a Polish space $\mathcal{Y}$ it follows that $C((\mathcal{X}, \tilde{d}_\mathcal{X}) \times \mathcal{Y}) \subseteq C((\mathcal{X}, d_\mathcal{X}) \times \mathcal{Y})$.*

Proof. The (pseudo-)metric properties are clearly fulfilled for $(\mathcal{X}, \tilde{d}_\mathcal{X})$. By continuity of $\tilde{d}_\mathcal{X}$ under $d_\mathcal{X}$ the canonical inclusion $\iota : (\mathcal{X}, d_\mathcal{X}) \to (\mathcal{X}, \tilde{d}_\mathcal{X}), x \mapsto x$ is continuous. As the image of a compactum under a continuous map is again compact the first claim follows. For the second claim, take $h \in C((\mathcal{X}, \tilde{d}_\mathcal{X}) \times \mathcal{Y})$. Then, the composition map $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}, (x, y) \mapsto h(\iota(x), y)$ is continuous and therefore the canonical embedding $h\circ(\iota, \mathrm{Id}_\mathcal{Y})$ of h is included in $C((\mathcal{X}, d_\mathcal{X}) \times \mathcal{Y})$. □

References

[1] S.T. Rachev, L. Rüschendorf, Mass Transportation Problems: Volume I: Theory, Springer Science & Business Media, 1998.

[2] F. Santambrogio, Optimal Transport for Applied Mathematicians, in: Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015.

[3] C. Villani, Optimal Transport: Old and New, Vol. 338, Springer Science & Business Media, 2008.

[4] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 214–223.

[5] V. Chernozhukov, A. Galichon, M. Hallin, M. Henry, Monge–Kantorovich depth, quantiles, ranks and signs, Ann. Statist. 45 (2017) 223–256.

[6] M. Hallin, E. del Barrio, J. Cuesta-Albertos, C. Matrán, Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach, Ann. Statist. 49 (2021) 1139–1165.

[7] G. Mordant, J. Segers, Measuring dependence between random vectors via optimal transport, J. Multivariate Anal. 189 (2022) 104912.

[8] T.G. Nies, T. Staudt, A. Munk, Transport dependency: Optimal transport based dependency measures, 2021, Preprint arXiv:2105.02073.

[9] J.C. Wiesel, Measuring association with Wasserstein distances, Bernoulli 28 (2022) 2816–2832.

[10] W. Torous, F. Gunsilius, P. Rigollet, An optimal transport approach to causal inference, 2021, Preprint arXiv:2108.05858.

[11] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vis. 40 (2000) 99–121.

[12] C. Bunne, S.G. Stark, G. Gut, J.S. Del Castillo, M. Levesque, K.V. Lehmann, L. Pelkmans, A. Krause, G. Rätsch, Learning single-cell perturbation responses using neural optimal transport, Nat. Methods 20 (2023) 1759–1768.

[13] C. Tameling, S. Stoldt, T. Stephan, J. Naas, S. Jakobs, A. Munk, Colocalization for super-resolution microscopy via optimal transport, Nat. Comput. Sci. 1 (2021) 199–211.

[14] S.N. Evans, F.A. Matsen, The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples, J. R. Stat. Soc. Ser. B Stat. Methodol. 74 (2012) 569–592.

[15] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al., Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming, Cell 176 (2019) 928–943.

[16] M. Gellert, M.F. Hossain, F.J.F. Berens, L.Z. Bruhn, C. Urbainsky, V. Liebscher, C.H. Lillig, Substrate specificity of thioredoxins and glutaredoxins–towards a functional classification, Heliyon 5 (2019) e02943.

[17] C.A. Weitkamp, K. Proksch, C. Tameling, A. Munk, Distribution of distances based object matching: asymptotic inference, J. Amer. Statist. Assoc. 119 (2024) 538–551.

[18] M. Sommerfeld, A. Munk, Inference for empirical Wasserstein distances on finite spaces, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2018) 219–238.

[19] M. Hallin, G. Mordant, J. Segers, Multivariate goodness-of-fit tests based on Wasserstein distance, Electron. J. Stat. 15 (2021) 1328–1371.

[20] E. Grave, A. Joulin, Q. Berthet, Unsupervised alignment of embeddings with Wasserstein procrustes, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 1880–1890.

[21] N. Bonneel, J. Rabin, G. Peyré, H. Pfister, Sliced and Radon Wasserstein Barycenters of measures, J. Math. Imaging Vision 51 (2015) 22–45.

[22] T. Lin, Z. Zheng, E. Chen, M. Cuturi, M.I. Jordan, On projection robust optimal transport: Sample complexity and model misspecification, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2021, pp. 262–270.

[23] J. Niles-Weed, P. Rigollet, Estimation of Wasserstein distances in the spiked transport model, Bernoulli 28 (2022) 2663–2688.

[24] I. Deshpande, Y.T. Hu, R. Sun, A. Pyrros, N. Siddiqui, S. Koyejo, Z. Zhao, D. Forsyth, A.G. Schwing, Max-sliced Wasserstein distance and its use for gans, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10648–10656.

[25] G. Carlier, A. Dupuy, A. Galichon, Y. Sun, Sista: learning optimal transport costs under sparsity constraints, Comm. Pure Appl. Math. 76 (2023) 1659–1677.

[26] A.M. Stuart, M.-T. Wolfram, Inverse optimal transport, SIAM J. Appl. Math. 80 (2020) 599–619.

[27] A.-A. Pooladian, C. Domingo-Enrich, R.T.Q. Chen, B. Amos, Neural optimal transport with Lagrangian costs, in: ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems, 2023.

[28] A. González-Sanz, M. Groppe, A. Munk, Nonlinear inverse optimal transport: Identifiability of the transport cost from its marginals and optimal values, 2023, Preprint arXiv:2312.05843.

[29] T. Manole, S. Balakrishnan, J. Niles-Weed, L. Wasserman, Plugin estimation of smooth optimal transport maps, Ann. Statist. (2024) (in press) preprint arXiv:2107.12364.

[30] E. del Barrio, E. Giné, C. Matrán, Central limit theorems for the Wasserstein distance between the empirical and the true distributions, Ann. Probab. 27 (1999) 1009–1071.

[31] E. del Barrio, E. Giné, F. Utzet, Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances, Bernoulli 11 (2005) 131–189.

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

[33] D.M. Mason, A weighted approximation approach to the study of the empirical Wasserstein distance, in: High Dimensional Probability VII, Springer, 2016, pp. 137–154.

[34] A. Munk, C. Czado, Nonparametric validation of similar distributions and assessment of goodness of fit, J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (1998) 223–241.

[35] M. Csörgö, L. Horváth, Weighted Approximations in Probability and Statistics, John Wiley & Sons, 1993.

[36] E. del Barrio, A. González-Sanz, J.M. Loubes, Central limit theorems for general transportation costs, Ann. l'Inst. Henri Poincaré 60 (2024) 847–873.

[37] E. del Barrio, J.M. Loubes, Central limit theorems for empirical transportation cost in general dimension, Ann. Probab. 47 (2019) 926–951.

[38] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes: with Applications to Statistics, in: Springer Series in Statistics, Springer, 1996.

[39] N. Fournier, A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probab. Theory Related Fields 162 (2015) 707–738.

[40] J. Weed, F. Bach, Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance, Bernoulli 25 (2019) 2620–2648.

[41] T. Manole, J. Niles-Weed, Sharp convergence rates for empirical optimal transport with smooth costs, Ann. Appl. Probab. 34 (2024) 1108–1135.

[42] S. Singh, B. Póczos, Minimax distribution estimation in Wasserstein distance, 2018, Preprint arXiv:1802.08855.

[43] S. Hundrieser, T. Staudt, A. Munk, Empirical optimal transport between different measures adapts to lower complexity, Ann. l'Inst. Henri Poincaré 60 (2024) 847–873.

[44] Shayan Hundrieser, Marcel Klatt, Axel Munk, Thomas Staudt, A unifying approach to distributional limits for empirical optimal transport, Bernoulli 30 (2024) 2846–2877.

[45] W. Römisch, Delta method, infinite dimensional, in: Encyclopedia of Statistical Sciences, vol. 16, Wiley, New York, 2006, pp. 1575–1583.

[46] C. Tameling, M. Sommerfeld, A. Munk, Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications, Ann. Appl. Probab. 29 (2019) 2744–2781.

[47] E. del Barrio, A. González-Sanz, J.M. Loubes, Central limit theorems for semi-discrete Wasserstein distances, Bernoulli 30 (2024) 554–580.

[48] Z. Fang, A. Santos, Inference on directionally differentiable functions, Rev. Econ. Stud. 86 (2019) 377–412.

[49] V. Toma, Strong convergence and Dini theorems for non-uniform spaces, Ann. Math. Blaise Pascal 4 (1997) 97–102.

[50] J. Xi, J. Niles-Weed, Distributional convergence of the sliced Wasserstein process, Adv. Neural Inf. Process. Syst. 35 (2022) 13961–13973.

[51] C. Villani, Topics in Optimal Transportation, Vol. 58, American Mathematical Society, 2003.

[52] J. Cárcamo, A. Cuevas, L.A. Rodríguez, Directional differentiability for supremum-type functionals: Statistical applications, Bernoulli 26 (2020) 2143–2175.

[53] L. Dümbgen, On nondifferentiable functions and the bootstrap, Probab. Theory Related Fields 95 (1993) 125–140.

[54] A. Bücher, I. Kojadinovic, A note on conditional versus joint unconditional weak convergence in bootstrap consistency results, J. Theoret. Probab. 32 (2019) 1145–1165.

[55] V. Hartmann, D. Schuhmacher, Semi-discrete optimal transport: a solution procedure for the unsquared euclidean distance case, Math. Methods Oper. Res. 92 (2020) 133–163.

[56] W. Gangbo, R.J. McCann, The geometry of optimal transportation, Acta Math. 177 (1996) 113–161.

[57] V. Levin, Abstract cyclical monotonicity and monge solutions for the general Monge–Kantorovich problem, Set-Valued Anal. 7 (1999) 7–32.

[58] N. Ahmad, H.K. Kim, R.J. McCann, Optimal transportation, topology and uniqueness, Bull. Math. Sci. 1 (2011) 13–32.

[59] R.J. McCann, L. Rifford, The intrinsic dynamics of optimal transport, J. l'École Polytech. Math. 3 (2016) 67–98.

[60] M. Klatt, A. Munk, Y. Zemel, Limit laws for empirical optimal solutions in stochastic linear programs, Ann. Oper. Res. 315 (2022) 251–278.

[61] A.W. van der Vaart, J.A. Wellner, Empirical processes indexed by estimated functions, in: Asymptotics: Particles, Processes and Inverse Problems, Institute of Mathematical Statistics, 2007, pp. 234–252.

[62] T. Staudt, S. Hundrieser, A. Munk, On the uniqueness of Kantorovich potentials, 2022, Preprint arXiv:2201.08316.

[63] H. Dette, A. Munk, Validation of linear regression models, Ann. Statist. 26 (1998) 778–800.

[64] H. Dette, W. Wu, Detecting relevant changes in the mean of nonstationary processes—a mass excess approach, Ann. Statist. 47 (2019) 3578–3608.

[65] A.W. van der Vaart, Asymptotic statistics, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.

[66] D. Alvarez-Melis, S. Jegelka, T.S. Jaakkola, Towards optimal transport with global invariances, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 1870–1879.

[67] K. Jin, C. Liu, C. Xia, Two-sided Wasserstein procrustes analysis, in: IJCAI, 2021, pp. 3515–3521.

[68] E. Bernton, P.E. Jacob, M. Gerber, C.P. Robert, On parameter estimation with the Wasserstein distance, Inf. Inference J. IMA 8 (2019) 657–676.

[69] X. Bing, F. Bunea, J. Niles-Weed, Estimation and inference for the Wasserstein distance between mixing measures in topic models, 2022, Preprint arXiv:2206.12768.

[70] J. Delon, A. Desolneux, A Wasserstein-type distance in the space of Gaussian mixture models, SIAM J. Imaging Sci. 13 (2020) 936–970.

[71] G.J. McLachlan, S.X. Lee, S.I. Rathnayake, Finite mixture models, Annu. Rev. Stat. Appl. 6 (2019) 355–378.

[72] P. Heinrich, J. Kahn, Strong identifiability and optimal minimax rates for finite mixture estimation, Ann. Statist. 46 (2018) 2844–2870.

[73] N. Ho, X. Nguyen, Convergence rates of parameter estimation for some weakly identifiable finite mixtures, Ann. Statist. 44 (2016a) 2726–2755.

[74] N. Ho, X. Nguyen, On strong identifiability and convergence rates of parameter estimation in finite mixtures, Electron. J. Stat. 10 (2016b) 271–307.

[75] T. Manole, N. Ho, Refined convergence rates for maximum likelihood estimation under finite mixture models, in: International Conference on Machine Learning, PMLR, 2022, pp. 14979–15006.

[76] N. Doss, Y. Wu, P. Yang, H.H. Zhou, Optimal estimation of high-dimensional gaussian location mixtures, Ann. Statist. 51 (2023) 62–95.

[77] Y. Wu, P. Yang, Optimal estimation of Gaussian mixtures via denoised method of moments, Ann. Statist. 48 (2020) 1981–2007.

[78] A. Guha, N. Ho, X. Nguyen, On posterior contraction of parameters and interpretability in Bayesian mixture modeling, Bernoulli 27 (2021) 2159–2188.

[79] I. Ohn, L. Lin, Optimal bayesian estimation of gaussian mixtures with growing number of components, Bernoulli 29 (2023) 1195–1218.

[80] Z. Goldfeld, K. Kato, G. Rioux, R. Sadhu, Statistical inference with regularized optimal transport, Inf. Inference J. IMA 5 (2024) iaad056.

[81] T. Manole, S. Balakrishnan, L. Wasserman, Minimax confidence intervals for the sliced Wasserstein distance, Electron. J. Stat. 16 (2022) 2252–2345.

[82] X. Xu, Z. Huang, Central limit theorem for the sliced 1-Wasserstein distance and the max-sliced 1-Wasserstein distance, 2022, Preprint arXiv:2205.14624.

[83] S. Kolouri, K. Nadjahi, U. Şimşekli, R. Badeau, G.K. Rohde, Generalized sliced wasserstein distances, in: H. Wallach, H. Larochelle, et al. (Eds.), Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019, pp. 261–272.

[84] M. Quellmalz, R. Beinert, G. Steidl, Sliced optimal transport on the sphere, Inverse Problems 39 (2023) 105005.

[85] C. Bonet, L. Chapel, L. Drumetz, N. Courty, Hyperbolic sliced-Wasserstein via geodesic and horospherical projections, in: Topological, Algebraic and Geometric Learning Workshops 2023, PMLR, 2023, pp. 334–370.

[86] C. Bonet, B. Malézieux, A. Rakotomamonjy, L. Drumetz, T. Moreau, M. Kowalski, N. Courty, Sliced-Wasserstein on symmetric positive definite matrices for M/EEG signals, in: International Conference on Machine Learning, PMLR, 2023, pp. 2777–2805.

[87] T. Gal, H.J. Greenberg, Advances in Sensitivity Analysis and Parametric Programming, Vol. 1, Springer Science & Business Media, 1997.
[88] R.M. Dudley, Nonlinear functionals of empirical measures and the bootstrap, in: Probability in Banach Spaces, vol. 7, Springer, 1990, pp. 63–82.
[89] B. Jourdain, A. Tse, Central limit theorem over non-linear functionals of empirical measures with applications to the mean-field fluctuation of interacting diffusions, Electron. J. Probab. 26 (2021) 1–34.
[90] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1999.
[91] R.J. Beran, L. Le Cam, P.W. Millar, Convergence of stochastic empirical measures, J. Multivariate Anal. 23 (1987) 159–168.
[92] V.S. Varadarajan, On the convergence of sample probability distributions, Sankhyā Indian J. Stat. (1933-1960) 19 (1958) 23–26.
[93] M.J. Wainwright, High-Dimensional Statistics: A Non-Asymptotic Viewpoint, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 48, Cambridge University Press, 2019.
[94] N.J. Higham, Functions of Matrices: Theory and Computation, SIAM, 2008.
[95] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley & Sons, 2019.
[96] A.N. Kolmogorov, V.M. Tikhomirov, ϵ-Entropy and ϵ-capacity of sets in functional spaces, in: S. Cernikov, N. Cernikova, A. Kolmogorov, A. Mal'cev, B. Plotkin (Eds.), Twelve Papers on Algebra and Real Functions, in: American Mathematical Society Translations–series, vol. 2, American Mathematical Society, 1961, pp. 277–364.
[97] P. Albano, Some properties of semiconcave functions with general modulus, J. Math. Anal. Appl. 271 (2002) 217–231.
[98] E.M. Bronshtein, ϵ-entropy of convex sets and functions, Sib. Math. J. 17 (1976) 393–398.
[99] A. Guntuboyina, B. Sen, Covering numbers for convex functions, IEEE Trans. Inform. Theory 59 (2013) 1957–1965.
[100] E. Giné, R. Nickl, Mathematical Foundations of Infinite-Dimensional Statistical Models, Cambridge University Press, 2016.
[101] F. Bolley, Separability and completeness for the Wasserstein distance, in: Séminaire de probabilités XLI, Springer, 2008, pp. 371–377.
[102] J. Aubin, H. Frankowska, Set-Valued Analysis, Modern Birkhäuser Classics, Springer, 1990.