Journal of Convex Analysis
Volume 29 (2022), No. 1, 1–12

A Simple Relaxation Approach to Duality for Optimal Transport Problems in Completely Regular Spaces

Giuseppe Savaré
Department of Decision Sciences, Bocconi University, Milan, Italy
giuseppe.savare@unibocconi.it

Giacomo E. Sodini
Fakultät für Mathematik, Technische Universität München, Garching bei München, Germany
sodini@ma.tum.de

Received: June 13, 2020
Accepted: November 7, 2020

We present a simple and direct approach to duality for Optimal Transport for lower semicontinuous cost functionals in arbitrary completely regular topological spaces, showing that the Optimal Transport functional can be interpreted as the largest sublinear and weakly lower semicontinuous functional extending the cost between pairs of Dirac masses.

Keywords: Optimal transport, convex duality, Legendre transform, Fenchel-Moreau theorem.

2020 Mathematics Subject Classification: 49Q22, 2020, 49N15, 28A33.

1. Introduction

Let X_1, X_2 be Hausdorff topological spaces and let $\mathsf{c} : X_1 \times X_2 \to [0, +\infty]$ be a lower semicontinuous cost function. Given $\mu_i \in \mathcal{P}(X_i)$, the space of Radon probability measures on X_i, $i = 1, 2$, the Kantorovich formulation of the Optimal Transport problem amounts to finding the optimal transference plan $\boldsymbol{\gamma} \in \mathcal{P}(X_1 \times X_2)$ minimizing the linear functional

$$\mathscr{C}(\boldsymbol{\gamma}) := \int_{X_1 \times X_2} \mathsf{c}(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma}(x_1, x_2) \tag{1}$$

in the set $\Gamma(\mu_1, \mu_2)$ of all the plans whose marginals are μ_1 and μ_2 respectively.

One of the fundamental results concerning such a problem is its dual formulation in terms of the maximization of the linear cost

$$\mathscr{D}(\varphi_1, \varphi_2; \mu_1, \mu_2) := \int_{X_1} \varphi_1\,\mathrm{d}\mu_1 + \int_{X_2} \varphi_2\,\mathrm{d}\mu_2 \quad \varphi_i \in \mathrm{C}_b(X_i),$$

on the convex set of continuous and bounded functions

$$\mathrm{K}(\mathsf{c}) := \left\{ \begin{array}{l} (\varphi_1, \varphi_2) \in \mathrm{C}_b(X_1) \times \mathrm{C}_b(X_2) \mid \\ \varphi_1(x_1) + \varphi_2(x_2) \le \mathsf{c}(x_1, x_2)\ \forall\,(x_1, x_2) \in X_1 \times X_2 \end{array} \right\}. \tag{2}$$

ISSN 0944-6532 / $ 2.50

In fact, if X_1 and X_2 are completely regular spaces, so that $\mathrm{C}_b(X_i)$ contains enough continuous functions to generate the topology of X_i, it is possible to prove that

$$\min\Big\{\mathscr{C}(\boldsymbol{\gamma}) \mid \boldsymbol{\gamma}\in\Gamma(\mu_1,\mu_2)\Big\} = \sup\Big\{\mathscr{D}(\varphi_1,\varphi_2;\mu_1,\mu_2) \mid (\varphi_1,\varphi_2)\in\mathrm{K}(\mathsf{c})\Big\}. \tag{3}$$

We refer to [23, Section 5] for a detailed bibliographical discussion. When c is a metric in a compact space $X = X_1 = X_2$, (3) was proven by Kantorovich in his celebrated paper [12]; the result was then extended to separable metric spaces by Dudley (see [7, §11.8] and also the comments in [1, Appendix B], [9]). The result for general costs in compact spaces was derived by [11] and then considerably extended by [13] (see also [18]). In order to deal with general Borel cost functions c, [13] deeply studied the dependence of the primal and dual problem from the cost function c, once μ_1, μ_2 are fixed.

The standard approach to (3) is to initially derive the duality in a simplified setting (finite dimension, discrete measures, compact spaces) by using some general principle for convex constrained problems (as Fenchel-Rockafellar Theorem in [22, Chap. 1] or minimax for saddle points of a Lagrangian formulation in [14] or linear programming in finite dimensional spaces in [13]), ultimately relying on Hahn-Banach theorem. In this respect, the compactness of X_i plays a crucial role thanks to the Riesz-Markov-Kakutani theorem, which in particular shows that the spaces of Radon probability measures $\mathcal{P}(X_1\times X_2)$ is a weakly*-compact convex subset of the dual of the Banach space $\mathrm{C}_b(X_1\times X_2)$. The duality result is then extended to more general cases by using the particular structure of the problem (as the tightness of the measures in [22]) or by a more subtle construction involving the c-transform trick (as in [2, 23]). It is worth mentioning the dual approach by [19, §1.6.3] (see also [4]).

The aim of this short note is to present a simple and direct proof of the general duality result (3) exploiting a slightly different point of view, which in some sense is complementary with respect to the approach of [13]: instead of studying the dependence of the Optimal Transport functional

$$\mathscr{T}_{\mathsf{c}}(\mu_1,\mu_2) := \min\Big\{\int_{X_1\times X_2}\mathsf{c}\,\mathrm{d}\boldsymbol{\gamma} \mid \boldsymbol{\gamma}\in\Gamma(\mu_1,\mu_2)\Big\} \tag{4}$$

on the cost c, we keep c fixed and study the dependence of (a suitable extension of) $\mathscr{T}_{\mathsf{c}}$ with respect to the pair (μ_1,μ_2) in $\mathcal{M}_+(X_1)\times\mathcal{M}_+(X_2)$, where $\mathcal{M}_+(X)$ denotes the cone of positive measures in the space $\mathcal{M}(X)$ of real valued Radon measures on X.

Our approach is motivated by two simple remarks: first of all, the dual functional induced by the right-hand side of (3) is the supremum of a family of linear functionals on the pair (μ_1,μ_2) and is therefore a sublinear (i.e. convex and positively 1-homogeneous) functional on $\mathcal{M}(X_1)\times\mathcal{M}(X_2)$ (possibly taking the value $+\infty$), lower semicontinuous with respect to the weak topology in duality with continuous and bounded functions. Therefore, it is natural to extend the Optimal Transport functional (4) to the whole space $\mathcal{M}(X_1)\times\mathcal{M}(X_2)$, by homogeneity if $\mu_1(X_1) = \mu_2(X_2)\ge 0$ and setting it equal to $+\infty$ if one of the measures μ_i does not belong to $\mathcal{M}_+(X_i)$ or $\mu_1(X_1)\neq\mu_2(X_2)$ (see also formula (10)). A second natural property relies on the intimate relation between $\mathscr{T}_{\mathsf{c}}$ and the cost c, which relies on the obvious formula

$$\mathscr{T}_{\mathsf{c}}(m\delta_{x_1}, m\delta_{x_2}) = m\mathsf{c}(x_1,x_2)\quad\text{for every } m\ge 0,\quad (x_1,x_2)\in X_1\times X_2. \tag{5}$$

The duality formula can be used to characterize $\mathscr{T}_{\mathsf{c}}$ as the largest convex and lower semicontinuous functional on $\mathcal{M}(X_1) \times \mathcal{M}(X_2)$ which coincides with (5) on the set of pair of nonnegative discrete measures.

Our argument is based on two steps: we first consider the extension of (4) to $\mathcal{M}(X_1) \times \mathcal{M}(X_2)$ as in formula (10). It is immediate to check that $\mathscr{T}_{\mathsf{c}}$ is a sublinear functional; in Section 3 we will also prove that $\mathscr{T}_{\mathsf{c}}$ is lower semicontinuous with respect to weak convergence. The technique relies on a simple compactification argument which shows the robust dependence of the sets $\Gamma(\mu_1, \mu_2)$ w.r.t. (μ_1, μ_2) in $\mathcal{M}_+(X_1) \times \mathcal{M}_+(X_2)$, a result which is of independent interest in such a general framework (see Theorem 3.1).

In the second step (Section 4) the duality formula (3) emerges as a direct application of Fenchel-Moreau theorem $\mathscr{T}_{\mathsf{c}} = \mathscr{T}_{\mathsf{c}}^{**}$ just by the computation of the polar of $\mathscr{T}_{\mathsf{c}}$ with respect to the duality pairing between measures and continuous functions.

As a byproduct, we obtain a reinforcement of (3), showing that the spaces of bounded continuous functions $\mathrm{C}_b(X_i)$ can be replaced by smaller unital subalgebras A_i which are sufficiently rich to generate the topology of X_i. As a simple example of applications of this fact, we can obtain duality formulas with smooth C^∞ functions in finite dimensional Euclidean spaces, Lipschitz functions in metric spaces, smooth cylindrical functions in topological vector spaces.

A second advantage of this approach concerns the analysis of more general transport functionals $\mathscr{U}_{\mathsf{c}}$ for unbalanced pair of measures, generated by functions of the form

$$\mathscr{U}_{\mathsf{c}}(m_1\delta_{x_1}, m_2\delta_{x_2}) := \mathsf{c}(x_1, m_1; x_2, m_2) \quad \text{for every } (x_1, x_2) \in X_1 \times X_2,\ m_1, m_2 \geq 0,$$

which are finite even if $m_1 \neq m_2$. This class of transport problems has been recently introduced by [15, 5, 16], and we plan to study its duality properties in a forthcoming paper, by using the same technique introduced here.

Acknowledgements. We thank the anonymous reviewer for her/his constructive and detailed comments and we gratefully acknowledge the support of the Institute of Advanced Study of the Technical University of Munich. G.S. was also supported by MIUR-PRIN 2017 project *Gradient flows, Optimal Transport and Metric Measure Structures* and by IMATI-CNR, Pavia.

2. Measure theoretic and topological preliminaries

Let X be a Hausdorff topological space. We denote by $\mathscr{B}(X)$ the Borel σ-algebra on X and by $\mathcal{M}(X)$ (resp. $\mathcal{M}_+(X)$) the vector space of real valued (resp. the cone of nonnegative) Radon measures on X, i.e. the countably additive set functions $\mu : \mathscr{B}(X) \to \mathbb{R}$ (resp. $\mu : \mathscr{B}(X) \to [0, +\infty)$) s.t.

$$\forall\, B \in \mathscr{B}(X)\ \forall\, \epsilon > 0 \quad \exists\, K \subset B \text{ compact s.t. } |\mu|(B \setminus K) < \epsilon, \tag{6}$$

where $|\mu|$ denotes the total variation of μ. It is worth noticing that if X is a Polish (or even Souslin) space then every finite Borel measure satisfies (6) and therefore is Radon [20, Theorem 10, page 122].

If $f : X \to Y$ is a continuous map between two Hausdorff spaces and $\mu \in \mathcal{M}(X)$, we will denote by $f_\sharp\mu$ the push-forward measure in $\mathcal{M}(Y)$ defined by

$$f_\sharp\mu(B) := \mu(f^{-1}(B)) \quad \text{for every } B \in \mathscr{B}(Y).$$

When X is a *completely regular* (or Tychonoff) space (i.e. it is Hausdorff and for every closed set C and point $x \in X \setminus C$ there exists a continuous function $f : X \to [0,1]$ s.t. $f(x) = 0$ and $f(C) = \{1\}$) there is a natural duality pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{M}(X)$ and the space $\mathrm{C}_b(X)$ of continuous and bounded real functions defined on X

$$\langle \mu, \varphi \rangle := \int_X \varphi \,\mathrm{d}\mu \quad \text{for every } \mu \in \mathcal{M}(X),\ \varphi \in \mathrm{C}_b(X). \tag{7}$$

(7) defines a real nondegenerate bilinear form in $\mathcal{M}(X) \times \mathrm{C}_b(X)$, for if a Radon measure $\mu \in \mathcal{M}(X)$ satisfies $\int_X \varphi \,\mathrm{d}\mu = 0$ for every $\varphi \in \mathrm{C}_b(X)$, then $|\mu|(B) = 0$ for every $B \in \mathscr{B}(X)$ (e.g. by the approximation result [3, Lemma 7.2.8]) so that μ is the null measure. Hence we can endow $\mathcal{M}(X)$ with the weak (or narrow) Hausdorff topology $\sigma(\mathcal{M}(X), \mathrm{C}_b(X))$: the coarsest topology on $\mathcal{M}(X)$ for which the maps $\mu \mapsto \int_X \varphi \,\mathrm{d}\mu$ are continuous for every $\varphi \in \mathrm{C}_b(X)$.

Since in general $\mathcal{M}(X)$ is not first-countable, we will mostly deal with general nets $(\mu_\lambda)_{\lambda \in \mathbb{L}}$, i.e. maps $\lambda \to \mu_\lambda$ defined in a directed set $\mathbb{L}$ with values in $\mathcal{M}(X)$, see e.g. [10, §4.3]. By definition, a net $(\mu_\lambda)_{\lambda \in \mathbb{L}} \subset \mathcal{M}(X)$ converges to $\mu \in \mathcal{M}(X)$ in the weak topology if

$$\lim_{\lambda \in \mathbb{L}} \int_X \varphi \,\mathrm{d}\mu_\lambda = \int_X \varphi \,\mathrm{d}\mu \quad \forall \varphi \in \mathrm{C}_b(X).$$

Recall that a measure $\mu \in \mathcal{M}(X)$ is concentrated (or is carried by) $X' \subset X$ if $X \setminus X'$ is μ-negligible (thus there exists a Borel set $X'' \subset X'$ such that $\mu(X \setminus X'') = 0$); in particular

$$\text{a Radon measure } \mu \in \mathcal{M}(X) \text{ is always concentrated on a } \sigma\text{-compact subset } D. \tag{8}$$

It is enough to take $D := \cup_n K_n$, where $K_n \subset X$ are compact sets s.t. $|\mu|(X \setminus K_n) < \frac{1}{n}$, obtained by equation (6) with $\epsilon = 1/n$ and $B = X$.

We collect a list of useful properties, see [6, 54, 58, 59 Chap. III] for the proofs of the last three claims.

Lemma 2.1. *Let X, Y be completely regular spaces.*

(1) *If $f : X \to Y$ is continuous then the map $f_\sharp : \mathcal{M}(X) \to \mathcal{M}(Y)$ is continuous.*

(2) *If $\varphi : X \to (-\infty, +\infty]$ is lower semicontinuous and bounded from below and $(\mu_\lambda)_{\lambda \in \mathbb{L}}$ is a net weakly converging to μ in $\mathcal{M}_+(X)$ then*

$$\liminf_{\lambda \in \mathbb{L}} \int_X \varphi \,\mathrm{d}\mu_\lambda \ge \int_X \varphi \,\mathrm{d}\mu.$$

(3) *If $\iota : X \to Y$ is a topological embedding (i.e. a continuous map providing a homeomorphism between X and $\iota(X)$ with the topology induced by the inclusion in Y), then $\iota_\sharp : \mathcal{M}(X) \to \mathcal{M}(Y)$ is a topological embedding as well, with*

$$\iota_\sharp\big(\mathcal{M}(X)\big) = \mathcal{M}(\iota(X), Y) := \Big\{\mu \in \mathcal{M}(Y) : \mu \text{ is concentrated on } \iota(X)\Big\}.$$

(4) *If X is compact then for every $M \ge 0$ the set $\Big\{\mu \in \mathcal{M}(X) : |\mu|(X) \le M\Big\}$ is compact.*

The last statement concerns sub-algebras $\mathrm{A} \subset \mathrm{C}_b(X)$ which are rich enough to characterize weak convergence. We first state the relevant definition.

Definition 2.2. (Adapted algebra of continuous functions)
Let X be a completely regular space. We say that a unital subalgebra $\mathrm{A} \subset \mathrm{C}_b(X)$ is *adapted* if the topology of X coincides with the initial topology induced by A. Equivalently, for every net $(x_\lambda)_{\lambda\in\mathbb{L}}$ in X

$$\lim_{\lambda\in\mathbb{L}} x_\lambda = x \quad \Leftrightarrow \quad \lim_{\lambda\in\mathbb{L}} f(x_\lambda) = f(x) \quad \text{for every } f \in \mathrm{A}.$$

Since X is Hausdorff, it is immediate to check that an adapted algebra A separates the points of X. It is interesting that the above condition is also sufficient to recover the weak topology of $\mathcal{M}(X)$.

Lemma 2.3. *Let X be a completely regular space and let $\mathrm{A} \subset \mathrm{C}_b(X)$ be an adapted algebra. Then a net $(\mu_\lambda)_{\lambda\in\mathbb{L}}$ in $\mathcal{M}(X)$ weakly converges to μ if and only if*

$$\lim_{\lambda\in\mathbb{L}} \int_X f \,\mathrm{d}\mu_\lambda = \int_X f \,\mathrm{d}\mu \quad \textit{for every } f \in \mathrm{A}. \tag{9}$$

Equivalently, the weak topology of $\mathcal{M}(X)$ coincides with $\sigma(\mathcal{M}(X), \mathrm{A})$.

Proof. We consider only the nontrivial implication and we will show that a net $(\mu_\lambda)_{\lambda\in\mathbb{L}}$ satisfying (9) weakly converges in $\mathcal{M}(X)$.

Let us set $I_f := [\inf_X f, \sup_X f] \subset \mathbb{R}$ and let us consider the product space $Y = \prod_{f\in\mathrm{A}} I_f$ endowed with the product topology; the component of a point in $y \in Y$ will be denoted as y_f with $f \in \mathrm{A}$. Y is compact by Tychonoff's Theorem. Since A is adapted, the map

$$\iota : X \to Y \quad \text{defined by} \quad \iota(x)_f := f(x) \quad \text{for every } x \in X$$

is a topological embedding. By Lemma 2.1(3) it is then sufficient to show that the net $\hat\mu_\lambda := \iota_\sharp\mu_\lambda$ weakly converges to $\hat\mu := \iota_\sharp\mu$ in $\mathcal{M}(Y)$. Let $\mathcal{B}$ be the unital algebra obtained by functions of the form

$$\varphi_{F,P}(y) = P(y_{f_1}, y_{f_2}, \dots, y_{f_k}),\ y\in Y,\ F = \{f_1, f_2, \cdots, f_k\} \subset \mathrm{A},\ P \text{ polynomial in } \mathbb{R}^k.$$

Since $\mathcal{B}$ contains the unit and separates the points of Y, by Stone-Weierstrass theorem $\mathcal{B}$ is uniformly dense in $\mathrm{C}_b(Y)$, so that in order to check the convergence of $\hat\mu_\lambda$ is sufficient to test them against functions of $\mathcal{B}$. We have

$$\begin{aligned}
\lim_{\lambda\in\mathbb{L}} \int_Y \varphi_{F,P}(y) \,\mathrm{d}\hat\mu_\lambda(y) &= \lim_{\lambda\in\mathbb{L}} \int_X P(\iota(x)_{f_1}, \iota(x)_{f_2}, \cdots, \iota(x)_{f_k}) \,\mathrm{d}\mu_\lambda(x) \\
&= \lim_{\lambda\in\mathbb{L}} \int_X P(f_1(x), f_2(x), \cdots, f_k(x)) \,\mathrm{d}\mu_\lambda(x) \\
&= \int_X P(f_1(x), f_2(x), \cdots f_k(x)) \,\mathrm{d}\mu(x) = \int_Y \varphi_{F,P}(y) \,\mathrm{d}\hat\mu(y),
\end{aligned}$$

where we used (9) and the fact that the function $x \mapsto P(f_1(x), f_2(x), \cdots f_k(x))$ belongs to the algebra A as well. $\square$

3. The optimal transport functional and its lower semi continuity

In all this section we will suppose that X_1, X_2 are completely regular spaces. The marginals μ_i of a Radon measure $\boldsymbol{\mu} \in \mathcal{M}_+(X_1 \times X_2)$ in the product space are defined in terms of the projection maps $\pi^i : X_1 \times X_2 \to X_i$, $\pi^i(x_1, x_2) := x_i$, $i = 1, 2$, by $\mu_i := \pi^i_\sharp \boldsymbol{\mu}$. Conversely, given $\mu_i \in \mathcal{M}_+(X_i)$ we will set

$$\Gamma(\mu_1, \mu_2) := \Big\{ \boldsymbol{\mu} \in \mathcal{M}_+(X_1 \times X_2) \mid \pi^i_\# \boldsymbol{\mu} = \mu_i \text{ for } i = 1, 2 \Big\}.$$

It is immediate to check that $\Gamma(\mu_1, \mu_2)$ is not empty if and only if $\mu_1(X_1) = \mu_2(X_2)$. In this case, $\Gamma(\mu_1, \mu_2)$ contains the unique Radon extension to $\mathscr{B}(X_1 \times X_2)$ of the product measure $M^{-1}\mu_1 \otimes \mu_2$ defined on $\mathscr{B}(X_1) \otimes \mathscr{B}(X_2)$ (see [3, Theorem 7.6.2] and [20, Chap. I, §9]) when $M := \mu_i(X_i) > 0$ (or the null measure if $\mu_i(X_i) = 0$).

The next result provides a crucial tool to study the lower semicontinuity of the Optimal Transport functional (4).

Theorem 3.1. *(Compactness from converging marginals)*
Let $(\boldsymbol{\gamma}_\lambda)_{\lambda \in \mathbb{L}}$ be a net in $\mathcal{M}_+(X_1 \times X_2)$ with $\mu_{i,\lambda} := \pi^i_\sharp \boldsymbol{\gamma}_\lambda \in \mathcal{M}_+(X_i)$, $i = 1, 2$, $\lambda \in \mathbb{L}$. If $(\mu_{i,\lambda})_{\lambda \in \mathbb{L}}$ weakly converge to some μ_i in $\mathcal{M}(X_i)$, then there exists a subnet $(\boldsymbol{\gamma}'_\alpha)_{\alpha \in \mathbb{A}}$ weakly convergent to some $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$ in $\mathcal{M}(X_1 \times X_2)$.

Proof. We recall that every completely regular space can be topologically embedded in a compact Hausdorff space (e.g. by the construction we used in the proof of Lemma 2.3: this property, in fact, characterizes completely regular spaces). Up to an identification of X_i with its homeomorphic image, we can thus assume that X_i is a subset of a compact Hausdorff spaces $\hat{X}_i$; thanks to Lemma 2.1(3), we can also identify the measures $\mu_{i,\lambda}$, μ_i in $\mathcal{M}_+(X_i)$ with corresponding measures $\hat{\mu}_{i,\lambda}$, $\hat{\mu}_i$ in $\mathcal{M}_+(X_i, \hat{X}_i)$ concentrated on X_i s.t. $\hat{\mu}_{i,\lambda} \to \hat{\mu}_i$ weakly in $\mathcal{M}(\hat{X}_i)$. Similarly, we can identify each $\boldsymbol{\gamma}_\lambda$ with a measure $\hat{\gamma}_\lambda$ in $\mathcal{M}_+(\hat{X}_1 \times \hat{X}_2)$ concentrated on $X_1 \times X_2$. Since $\hat{X}_1 \times \hat{X}_2$ is compact and the total mass $\hat{\boldsymbol{\gamma}}_\lambda(\hat{X}_1 \times X_2)$ is converging and thus is eventually bounded, by Lemma 2.1(4) there exists some $\tilde{\boldsymbol{\gamma}} \in \mathcal{M}_+(\hat{X}_1 \times \hat{X}_2)$ and a subnet $\boldsymbol{\gamma}'_\alpha = \boldsymbol{\gamma}_{\lambda(\alpha)}$ (with corresponding subnet $\hat{\boldsymbol{\gamma}}'_\alpha = \hat{\boldsymbol{\gamma}}_{\lambda(\alpha)}$) induced by a map $\alpha \mapsto \lambda(\alpha)$, $\alpha \in \mathbb{A}$ (see [10, Theorem 4.29]), such that

$$\hat{\boldsymbol{\gamma}}'_\alpha \to \tilde{\boldsymbol{\gamma}} \quad \text{weakly in } \mathcal{M}(\hat{X}_1 \times \hat{X}_2).$$

On the other hand, since the marginals of $\hat{\boldsymbol{\gamma}}_\lambda$ are $\hat{\mu}_{i,\lambda}$ and $\hat{\mu}_{i,\lambda} \to \hat{\mu}_i$ weakly in $\mathcal{M}(\hat{X}_i)$, we deduce that the marginals of $\tilde{\boldsymbol{\gamma}}$ on $\hat{X}_i$ are $\hat{\mu}_i$. Since $\hat{\mu}_i$ are Radon measures concentrated on two sigma compact subsets $D_i \subset X_i$ (see (8)), we have

$$\tilde{\boldsymbol{\gamma}}((\hat{X}_1 \times \hat{X}_2) \setminus (D_1 \times D_2)) \le \mu_1(X_1 \setminus D_1) + \mu_2(X_2 \setminus D_2) = 0.$$

It follows that $\tilde{\boldsymbol{\gamma}}$ is concentrated on $X_1 \times X_2$, and therefore can be written as $\tilde{\boldsymbol{\gamma}} = \iota_\sharp \boldsymbol{\gamma}$ for a measure $\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)$. A further application of Lemma 2.1(3) yields that $\boldsymbol{\gamma}'_\alpha$ weakly converges to $\boldsymbol{\gamma}$ in $\mathcal{M}(X_1 \times X_2)$ and Lemma 2.1(1) shows that $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$. □

Let us state an immediate consequence of the previous result.

Corollary 3.2. *(Compactness from compact marginals) Let $\mathcal{K}_i \subset \mathcal{M}_+(X_i)$ be compact in the weak topology, $i = 1, 2$. Then the set $\mathcal{K} := \{\boldsymbol{\gamma} \in \mathcal{M}_+(X_1 \times X_2) \mid \pi^i_\# \boldsymbol{\gamma} \in \mathcal{K}_i\}$ is compact in the weak topology of $\mathcal{M}(X_1 \times X_2)$.*

Proof. Since $\mathcal{K}$ is closed in $\mathcal{M}(X_1 \times X_2)$ thanks to Claim 1 of Lemma 2.1, it is sufficient to prove that every net $(\boldsymbol{\gamma}_\lambda)_{\lambda\in\mathbb{L}}$ in $\mathcal{K}$ has a converging subnet.

Setting $\mu_{i,\lambda} := \pi^i_\sharp \boldsymbol{\gamma}_\lambda$, thanks to the compactness of $\mathcal{K}_1 \times \mathcal{K}_2$ we can find a subnet $(\mu'_{1,\alpha}, \mu'_{2,\alpha})_{\alpha\in\mathbb{A}}$, $\mu'_{i,\alpha} = \mu_{i,\alpha(\lambda)}$, converging to $(\mu_1, \mu_2) \in \mathcal{K}_1 \times \mathcal{K}_2$ in $\mathcal{M}(X_1) \times \mathcal{M}(X_2)$.

Applying Theorem 3.1 we can find a further subnet $(\boldsymbol{\gamma}''_\beta)_{\beta\in\mathbb{B}}$ of $(\boldsymbol{\gamma}'_\alpha)_{\alpha\in\mathbb{A}}$ converging to a measure $\boldsymbol{\gamma}$ in $\mathcal{M}(X_1 \times X_2)$. □

Remark 3.3. In locally compact or Polish spaces Corollary 3.2 could also be proven by using Prokhorov's characterization of compact subsets of $\mathcal{M}_+(X_i)$ in terms of uniform tightness [6, 59, Chap. III]. The argument we are presenting here is more direct (once Radon measures are involved) and works in completely regular spaces as well. In the case of arbitrary topological spaces, one has to deal with a more refined definition of the weak topology and Corollary 3.2 can also be extended to this general setting. Since we think that this result is of independent interest, we added its proof in the Appendix.

Definition 3.4. (Optimal Transport functional)
Let X_1, X_2 be completely regular topological spaces and let us set

$$\mathcal{O}(X_1, X_2) := \Big\{(\mu_1, \mu_2) \in \mathcal{M}_+(X_1) \times \mathcal{M}_+(X_2) : \mu_1(X_1) = \mu_2(X_2)\Big\}.$$

For every pair $(\mu_1, \mu_2) \in \mathcal{M}(X_1)\times\mathcal{M}(X_2)$ the *Optimal Transport functional* induced by a lower semicontinuous cost function $\mathsf{c} : X_1 \times X_2 \to [0, +\infty]$ is defined by

$$\mathscr{T}_\mathsf{c}(\mu_1, \mu_2) := \min\left\{\int_{X_1\times X_2} \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma} \mid \boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)\right\} \quad \text{if } (\mu_1, \mu_2) \in \mathcal{O}(X_1, X_2); \tag{10}$$

we set $\mathscr{T}_\mathsf{c}(\mu_1, \mu_2) := +\infty$ if $(\mu_1, \mu_2) \notin \mathcal{O}(X_1, X_2)$.

Theorem 3.5. *(Convexity and lower semicontinuity of $\mathscr{T}_\mathsf{c}$) If $\mathsf{c}: X_1\times X_2 \to [0, +\infty]$ is a lower semi continuous function then $\mathscr{T}_c : \mathcal{M}(X_1) \times \mathcal{M}(X_2) \to [0, +\infty]$ is a well defined lower semicontinuous (w.r.t. the product weak topology) and sublinear function subject to*

$$\mathscr{T}_c(m\delta_{x_1}, m\delta_{x_2}) = mc(x_1, x_2) \quad \textit{for every } m \geq 0\,,\ (x_1, x_2) \in X_1 \times X_2. \tag{11}$$

Proof. Whenever $(\mu_1, \mu_2) \in \mathcal{O}$, the compactness of $\Gamma(\mu_1, \mu_2)$ (following from Corollary 3.2) and the lower semicontinuity of the cost $\mathscr{C}$ of (1) (following by Lemma 2.1(2)) give the existence of a minimum (possibly assuming the value $+\infty$) by the direct method in Calculus of Variations. We denote then by $\Gamma_o(\mu_1, \mu_2)$ the set of optimal plans, i.e. the set of $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$ realizing the minimum in (10).

Convexity of $\mathscr{T}_\mathsf{c}$ has to be checked only between pairs $(\mu_1, \mu_2), (\mu'_1, \mu'_2) \in \mathcal{O}$. In this case it is a simple consequence of the fact that, if $\boldsymbol{\gamma}\in\Gamma_o(\mu_1, \mu_2)$ and $\boldsymbol{\gamma}'\in\Gamma_o(\mu'_1, \mu'_2)$, then $t\boldsymbol{\gamma} + (1 - t)\boldsymbol{\gamma}' \in \Gamma(t\mu_1 + (1 - t)\mu_2, t\mu'_1 + (1 - t)\mu'_2)$ for every $t \in [0, 1]$, so that

$$\begin{aligned}\mathscr{T}_c(t\mu_1 + (1 - t)\mu_2, t\mu'_1 + (1 - t)\mu'_2) &\leq \int_{X_1\times X_2} \mathsf{c}\,\mathrm{d}(t\boldsymbol{\gamma} + (1 - t)\boldsymbol{\gamma}') \\ &= t\mathscr{T}_c(\mu_1, \mu_2) + (1 - t)\mathscr{T}_c(\mu'_1, \mu'_2).\end{aligned}$$

Lower semicontinuity is a consequence of Theorem 3.1; first of all notice that it is not restrictive to check it on nets $\{(\mu_{1,\lambda}, \mu_{2,\lambda})\}_{\lambda\in\mathbb{L}} \subset \mathcal{O}$ convergent to some $(\mu_1, \mu_2) \in \mathcal{O}$ such that $\mathscr{T}_{\mathsf{c}}(\mu_{1,\lambda}, \mu_{2,\lambda}) \le a$ for some $a \in [0, +\infty)$. We select a net $(\boldsymbol{\gamma}_\lambda)_{\lambda\in\mathbb{L}} \subset \mathcal{M}_+(X_1 \times X_2)$ s.t. $\boldsymbol{\gamma}_\lambda \in \Gamma_o(\mu_{1,\lambda}, \mu_{2,\lambda})$ for every $\lambda \in \mathbb{L}$. Thanks to Theorem 3.1, we can extract a further subnet $\boldsymbol{\gamma}'_\alpha = \boldsymbol{\gamma}_{\lambda(\alpha)}$, $\alpha \in \mathbb{A}$, converging in the weak topology to some $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$. Then

$$a \ge \liminf_{\alpha\in\mathbb{A}} \mathscr{T}_{\mathsf{c}}(\mu_{1,\alpha}, \mu_{2,\alpha}) = \liminf_{\alpha\in\mathbb{A}} \int_{X_1\times X_2} \mathsf{c}\, \mathrm{d}\boldsymbol{\gamma}'_\alpha \ge \int_{X_1\times X_2} \mathsf{c}\, \mathrm{d}\boldsymbol{\gamma} \ge \mathscr{T}_c(\mu_1, \mu_2).$$

Finally, (11) is a consequence of the fact that the set $\Gamma(m\delta_{x_1}, m\delta_{x_2})$ coincides with the singleton $\{m\delta_{x_1} \otimes \delta_{x_2}\}$. □

4. Kantorovich duality

Before stating the main duality result, let us briefly recall the Fenchel-Moreau Theorem in the framework of a pair of vector spaces E, F placed in duality by a non-degenerate bilinear map $\langle \cdot, \cdot \rangle$, see e.g. [8]. We endow E with the weak topology $\sigma(E, F)$, the coarsest topology for which all the functions $e \mapsto \langle e, f\rangle$, $f \in F$, are continuous.

Definition 4.1. (Lower semicontinuous convex envelope)
Let $\mathscr{F} : E \to (-\infty, +\infty]$ be a function satisfying

$$\mathscr{F}(e) \ge \langle e, f\rangle - c \quad \text{for some } f \in F,\ c \in \mathbb{R} \text{ and every } e \in E. \tag{12}$$

The *lower semicontinuous convex envelope* of $\mathscr{F}$ is the largest minorant of $\mathscr{F}$ among the lower semicontinuous and convex functions defined in E and taking values in $(-\infty, +\infty]$.

Definition 4.2. (Polar function)
Let $\mathscr{F} : E \to (-\infty, +\infty]$ be satisfying (12) and not identically $+\infty$. The *polar* (or conjugate) function of $\mathscr{F}$ is the function $\mathscr{F}^* : F \to (-\infty, +\infty]$ defined by

$$\mathscr{F}^*(f) := \sup_{e\in E} \langle e, f\rangle - \mathscr{F}(e) \quad \text{for every } f \in F.$$

Theorem 4.3. (Fenchel-Moreau)
Let E and F be vector spaces placed in duality and let $\mathscr{F} : E \to (-\infty, +\infty]$ be satisfying (12) *and not identically $+\infty$. Then the lower semicontinuous convex envelope of $\mathscr{F}$ is given by the dual formula*

$$\mathscr{F}^{**}(e) := \sup_{f\in F} \langle e, f\rangle - \mathscr{F}^*(f) \quad \textit{for every } e \in E.$$

In particular, if $\mathscr{F}$ is convex and lower semicontinuous then $\mathscr{F} = \mathscr{F}^{**}$.

Theorem 4.4. (Duality for Optimal Transport)
Let X_1, X_2 be completely regular spaces, let $\mathsf{c} : X_1 \times X_2 \to [0, +\infty]$ be a lower semicontinuous cost function and let $\mathrm{A}_i \subset \mathrm{C}_b(X_i)$ be adapted unital algebras of continuous functions, according to Definition 2.2.

Let us consider the functional $\mathscr{F}_{\mathsf{c}} : \mathcal{M}(X_1) \times \mathcal{M}(X_2) \to [0, +\infty]$

$$\mathscr{F}_{\mathsf{c}}(\mu_1, \mu_2) := \begin{cases} mc(x_1, x_2) & \textit{if } (\mu_1, \mu_2) = (m\delta_{x_1}, m\delta_{x_2}) \textit{ for } m \geq 0\,,\ (x_1, x_2) \in X_1 \times X_2 \\ +\infty & \textit{elsewhere.} \end{cases}$$

Then the functional $\mathscr{T}_{\mathsf{c}}$ *defined by* (10) *is the convex lower semicontinuous envelope of* $\mathscr{F}_{\mathsf{c}}$ *in* $\mathcal{M}(X_1) \times \mathcal{M}(X_2)$ *and for every* $\mu_i \in \mathcal{M}(X_i)$ *it holds*

$$\mathscr{T}_c(\mu_1, \mu_2) = \sup \left\{ \begin{array}{l} \displaystyle\int_{X_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{X_2} \varphi_2 \, \mathrm{d}\mu_2 \mid (\varphi_1, \varphi_2) \in \mathrm{A}_1 \times \mathrm{A}_2, \\ \varphi_1(x_1) + \varphi_2(x_2) \leq \mathsf{c}(x_1, x_2) \textit{ for every } (x_1, x_2) \in X_1 \times X_2 \end{array} \right\}.$$

Proof. Let us set $E := \mathcal{M}(X_1) \times \mathcal{M}(X_2)$ and $F := \mathrm{A}_1 \times \mathrm{A}_2$ with the obvious bilinear form

$$\langle \cdot, \cdot \rangle : E \times F \to \mathbb{R}, \qquad ((\mu_1, \mu_2), (\varphi_1, \varphi_2)) \mapsto \int_{X_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{X_2} \varphi_2 \, \mathrm{d}\mu_2.$$

Then for every $(\varphi_1, \varphi_2) \in \mathrm{A}_1 \times \mathrm{A}_2$

$$\begin{aligned} \mathscr{F}_{\mathsf{c}}^*(\varphi_1, \varphi_2) &= \sup_{(\mu_1, \mu_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)} \big\langle (\mu_1, \mu_2), (\varphi_1, \varphi_2) \big\rangle - \mathscr{F}_c(\mu_1, \mu_2) \\ &= \sup_{(x_1, x_2) \in X_1 \times X_2,\, m \geq 0} m\varphi_1(x_1) + m\varphi_2(x_2) - m\,\mathsf{c}(x_1, x_2) \\ &= \begin{cases} 0 & \text{if } \varphi_1(x_1) + \varphi_2(x_2) \leq \mathsf{c}(x_1, x_2) \text{ for every } x_i \in X_i, \\ +\infty & \text{otherwise.} \end{cases}, \end{aligned}$$

so that, recalling (2), $\mathscr{F}_{\mathsf{c}}^*$ is the indicator function of the convex set $\mathrm{K}(\mathsf{c}) \cap (\mathrm{A}_1 \times \mathrm{A}_2)$. From Theorem 4.3, we have that

$$\mathscr{F}_c^{**}(\mu_1, \mu_2) = \sup \left\{ \int_{X_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{X_2} \varphi_2 \, \mathrm{d}\mu_2 \mid (\varphi_1, \varphi_2) \in \mathrm{K}(\mathsf{c}) \cap (\mathrm{A}_1 \times \mathrm{A}_2) \right\}$$

must coincide with the lower semicontinuous (w.r.t. $\sigma(\mathcal{M}(X_1) \times \mathcal{M}(X_2), \mathrm{A}_1 \times \mathrm{A}_2)$) convex envelope of $\mathscr{F}_c$. Since, by Theorem 3.5, $\mathscr{T}_c$ is convex and weakly lower semicontinuous w.r.t. the product weak topology (coinciding by Lemma 2.3 with $\sigma(\mathcal{M}(X_1) \times \mathcal{M}(X_2), \mathrm{A}_1 \times \mathrm{A}_2)$) and stays below $\mathscr{F}_{\mathsf{c}}$, we have $\mathscr{T}_{\mathsf{c}} \leq \mathscr{F}_{\mathsf{c}}^{**}$.

The converse inequality is immediate since for every $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$ and every $(\varphi_1, \varphi_2) \in \mathrm{K}(\mathsf{c}) \cap (\mathrm{A}_1 \times \mathrm{A}_2)$ we have

$$\begin{aligned} \int_{X_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{X_2} \varphi_2 \, \mathrm{d}\mu_2 &= \int_{X_1 \times X_2} (\varphi_1(x_1) + \varphi_2(x_2)) \, \mathrm{d}\boldsymbol{\gamma}(x_1, x_2) \\ &\leq \int_{X_1 \times X_2} \mathsf{c}(x_1, x_2) \, \mathrm{d}\boldsymbol{\gamma}(x_1, x_2). \end{aligned}$$

Passing to the supremum in (φ_1, φ_2) and to the infimum in $\boldsymbol{\gamma}$, we conclude that $\mathscr{T}_c \geq \mathscr{F}_{\mathsf{c}}^{**}$. $\square$

A. Compactness for transport plans in topological spaces

In this last section we show how to generalize Theorem 3.1 and Corollary 3.2 to arbitrary Hausdorff topological spaces.

Since duality with continuous and bounded functions cannot be used to define a Hausdorff topology in $\mathcal{M}_+(X)$, a natural topology (called *narrow topology*) can be introduced following Topsoe [20, Appendix].

Definition A.1. (Narrow topology)
Let X be a Hausdorff topological space. The narrow topology on $\mathcal{M}_+(X)$ is the coarsest topology which makes all the maps $\mu \mapsto \int_X \varphi \, \mathrm{d}\mu$ lower semicontinuous for every bounded and lower semicontinuous function $\varphi : X \to \mathbb{R}$.

In order to state a useful criterium for compactness in $\mathcal{M}_+(X)$ we give the following definition.

Definition A.2. (Domination of compact sets)
Let X be a Hausdorff topological space and let $\mathcal{K}(X)$ (respectively $\mathcal{G}(X)$) be the collection of the compact (resp. open) subsets of X. We say that a collection $\mathcal{G} \subset \mathcal{G}(X)$ *dominates* the compact subsets of X, and we write $\mathcal{G} \succ \mathcal{K}(X)$, if

$$\forall\, K \in \mathcal{K}(X) \quad \exists\, G \in \mathcal{G} : \quad K \subset G.$$

Theorem A.3. (Topsoe [21]) *Let X be a Hausdorff topological space.*

(1) *A net $(\mu_\lambda)_{\lambda\in\mathbb{L}} \subset \mathcal{M}_+(X)$ is compact (i.e. from every subnet it is possible to extract a narrowly convergent sub-subnet) if and only if $\limsup_{\lambda\in\mathbb{L}} \mu_\lambda(X) < +\infty$ and for every $\mathcal{G} \succ \mathcal{K}(X)$ and for every $\epsilon > 0$ there exists a finite subset $\mathcal{G}' \subset \mathcal{G}$ such that*
$$\limsup_{\lambda\in\mathbb{L}} \min_{G\in\mathcal{G}'} \mu_\lambda(X \setminus G) \le \epsilon. \tag{13}$$

(2) *A narrowly closed set $\mathcal{F} \subset \mathcal{M}_+(X)$ is narrowly compact if and only if it is bounded and for every $\mathcal{G} \succ \mathcal{K}(X)$ and for every $\epsilon > 0$ there exists a finite subset $\mathcal{G}' \subset \mathcal{G}$ such that*
$$\sup_{\mu\in\mathcal{F}} \min_{G\in\mathcal{G}'} \mu(X \setminus G) \le \epsilon. \tag{14}$$

Remark A.4. Condition (14) is really a relaxation of the usual uniform tightness condition: in fact, the latter guarantees the existence of a singleton $\mathcal{G}'$ satisfying (14).

We are now able to state and prove the analogous of Theorem 3.1 and Corollary 3.2.

Theorem A.5. (Compactness from converging marginals) *Let X_i, $i = 1, 2$ be Hausdorff topological spaces and let $(\boldsymbol{\gamma}_\lambda)_{\lambda\in\mathbb{L}}$ be a net in $\mathcal{M}_+(X_1 \times X_2)$ satisfying $\mu_{i,\lambda} := \pi^i_\sharp \boldsymbol{\gamma}_\lambda \in \mathcal{M}_+(X_i)$, $i = 1, 2$, $\lambda \in \mathbb{L}$. If $(\mu_{i,\lambda})_{\lambda\in\mathbb{L}}$ narrowly converge to some μ_i in $\mathcal{M}(X_i)$, then there exists a subnet $(\boldsymbol{\gamma}'_\alpha)_{\alpha\in\mathbb{A}}$ narrowly convergent to some $\boldsymbol{\gamma} \in \Gamma(\mu_1, \mu_2)$ in $\mathcal{M}(X_1 \times X_2)$.*

Proof. Let us first recall (see e.g. [17, §26, Exercise 9]) that whenever $G \subset X_1 \times X_2$ is an open set containing the product $K_1 \times K_2$ of two compact subsets $K_i \subset X_i$, $i = 1, 2$, then there exist open sets $G_i \subset X_i$ such that

$$K_1 \times K_2 \subset G_1 \times G_2 \subset G. \tag{15}$$

Let $\mathcal{G} \succ \mathcal{K}(X_1 \times X_2)$ and let $\epsilon > 0$ be fixed. Thanks to (15), in order to check (13) it is not restrictive to replace $\mathcal{G}$ with the collection of cartesian open sets

$$\mathcal{G}_c := \Big\{ G_1 \times G_2 \mid \exists\, G \in \mathcal{G} \text{ such that } G_1 \times G_2 \subset G \Big\}.$$

Let us now introduce the disjoint union $X := X_1 \sqcup X_2$ endowed with the finest topology for which the canonical injections $\iota_i : X_i \to X$ are continuous; we can thus identify X_i with $\iota_i(X_i)$ as (open and closed) subsets of X. Since a set $A \subset X$ is open (resp. compact) in X if and only if $A \cap X_i$ is open (resp. compact) in X_i for $i = 1, 2$, it is not difficult to check that the family of open sets in X

$$\hat{\mathcal{G}}_c := \Big\{ G_1 \sqcup G_2 \mid G_1 \times G_2 \in \mathcal{G}_c \Big\}$$

dominates $\mathcal{K}(X)$.

We now consider the net $\mu_\lambda := (\iota_1)_\sharp \mu_{1,\lambda} + (\iota_2)_\sharp \mu_{2,\lambda}$ in $\mathcal{M}_+(X)$; equivalently,

$$\mu_\lambda(B) := \mu_{1,\lambda}(B \cap X_1) + \mu_{2,\lambda}(B \cap X_2)$$

for every Borel set B of X. It is immediate to check that μ_λ narrowly converges to $\mu := (\iota_1)_\sharp \mu_1 + (\iota_2)_\sharp \mu_2$. Following Theorem A.3(1) we can find a finite subset $\hat{\mathcal{G}}' = \{G_{1,j} \sqcup G_{2,j}\}_{j=1}^J$ of $\hat{\mathcal{G}}_c$ such that

$$\limsup_{\lambda \in \mathbb{L}} \min_{G \in \hat{\mathcal{G}}'} \mu_\lambda(X \setminus G) \le \epsilon. \tag{16}$$

On the other hand we observe that, for every $\lambda \in \mathbb{L}$ and $j \in \{1, \cdots, J\}$, it holds

$$\begin{aligned}\boldsymbol{\gamma}_\lambda(X_1 \times X_2 \setminus G_{1,j} \times G_{2,j})) &\le \boldsymbol{\gamma}_\lambda((X_1 \setminus G_{1,j}) \times X_2) + \boldsymbol{\gamma}_\lambda(X_1 \times (X_2 \setminus G_{2,j}))\\ &= \mu_{1,\lambda}(X_1 \setminus G_{1,j}) + \mu_{2,\lambda}(X_2 \setminus G_{2,j}) = \mu_\lambda(X \setminus G_{1,j} \sqcup G_{2,j}),\end{aligned}$$

so that, setting $\mathcal{G}' := \{G_{1,j} \times G_{2,j}\}_{j=1}^J$, (16) yields

$$\limsup_{\lambda \in \mathbb{L}} \min_{G \in \mathcal{G}'} \boldsymbol{\gamma}_\lambda(X_1 \times X_2 \setminus G) \le \limsup_{\lambda \in \mathbb{L}} \min_{G \in \hat{\mathcal{G}}'} \mu_\lambda(X \setminus G) \le \epsilon. \qquad \square$$

Arguing as in the proof of Corollary 3.2 we eventually obtain the corresponding characterization of compactness in $\mathcal{M}_+(X_1 \times X_2)$.

Corollary A.6. (Compactness from compact marginals)
Let X_i, $i = 1, 2$ be Hausdorff topological spaces and let $\mathcal{K}_i \subset \mathcal{M}_+(X_i)$ be compact in the narrow topology, $i = 1, 2$. Then the set $\mathcal{K} := \Big\{ \boldsymbol{\gamma} \in \mathcal{M}_+(X_1 \times X_2) \mid \pi^i_\# \boldsymbol{\gamma} \in \mathcal{K}_i \Big\}$ is compact in the narrow topology of $\mathcal{M}(X_1 \times X_2)$.

References

[1] A. de Acosta: *Invariance principles in probability for triangular arrays of B-valued random vectors and some applications*, Ann. Probab. 10/2 (1982) 346–373.

[2] L. Ambrosio, N. Gigli, G. Savaré: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, 2nd edition, Birkhäuser, Basel (2008).

[3] V. I. Bogachev: *Measure Theory II*, Springer, Berlin (2006).

[4] G. Bouchitté, G. Buttazzo: *Characterization of optimal shapes and masses through Monge-Kantorovich equation*, J. Eur. Math. Soc. 3/2 (2001) 139–168.

[5] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard: *Unbalanced optimal transport: dynamic and Kantorovich formulations*, J. Funct. Analysis 274/11 (2018) 3090–3123.

[6] C. Dellacherie, P.-A. Meyer: *Probabilities and Potential*, North-Holland Mathematics Studies 29, North-Holland, Amsterdam (1978).

[7] R. M. Dudley: *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics 74, revised reprint of the 1989 original, Cambridge University Press, Cambridge (2002).

[8] I. Ekeland, R. Témam: *Convex Analysis and Variational Problems*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia (1987).

[9] X. Fernique: *Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais*, in: *Seminar on Probability XV*, Univ. Strasbourg, 1979/1980, Lecture Notes in Mathematics 850, Springer, Berlin (2006) 6–10.

[10] G. B. Folland: *Real Analysis*, 2nd edition, John Wiley & Sons, New York (1999).

[11] N. Gaffke, L. Rüschendorf: *On a class of extremal problems in statistics*, Math. Operationsforsch. Statist. Ser. Optim. 12/1 (1981) 123–135.

[12] L. Kantorovitch: *On the translocation of masses*, C. R. (Doklady) Acad. Sci. URSS (N.S.) 37 (1942) 199–201; English translation in: J. Math. Sci. 133 (2006) 1381–1382.

[13] H. G. Kellerer: *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete 67/4 (1984) 399–432.

[14] Ch. Léonard: *A saddle-point approach to the Monge-Kantorovich optimal transport problem*, ESAIM Control Optim. Calc. Var. 17/3 (2011) 682–704.

[15] M. Liero, A. Mielke, G. Savaré: *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, SIAM J. Math. Analysis 48/4 (2016) 2869–2911.

[16] M. Liero, A. Mielke, G. Savaré: *Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures*, Inventiones Math. 211/3 (2018) 969–1117.

[17] J. Munkres: *Topology*, 2nd edition, Prentice Hall, Upper Saddle River (2000).

[18] S. T. Rachev, L. Rüschendorf: *Mass transportation problems. Vol. I*, Springer, New York (1998).

[19] F. Santambrogio: *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and their Applications 87, Birkhäuser, Basel (2015).

[20] L. Schwartz: *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*, Oxford University Press, Oxford (1973).

[21] F. Topsoe: *Compactness in spaces of measures*, Studia Mathematica 36 (1970) 195–212.

[22] C. Villani: *Topics in Optimal Transportation*, Graduate Studies in Mathematics 58, American Mathematical Society, Providence (2003).

[23] C. Villani: *Optimal Transport*, Grundlehren der Mathematischen Wissenschaften 338, Springer, Berlin (2009).