Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures

Matthias Liero *, Alexander Mielke †, Giuseppe Savaré ‡

August 31, 2015

Abstract

We develop a full theory for the new class of *Optimal Entropy-Transport problems* between nonnegative and finite Radon measures in general topological spaces.

They arise quite naturally by relaxing the marginal constraints typical of Optimal Transport problems: given a couple of finite measures (with possibly different total mass), one looks for minimizers of the sum of a linear transport functional and two convex entropy functionals, that quantify in some way the deviation of the marginals of the transport plan from the assigned measures.

As a powerful application of this theory, we study the particular case of *Logarithmic* Entropy-Transport problems and introduce the new *Hellinger-Kantorovich distance between measures in metric spaces*.

The striking connection between these two seemingly far topics allows for a deep analysis of the geometric properties of the new geodesic distance, which lies somehow between the well-known Hellinger-Kakutani and Kantorovich-Wasserstein distances.

Contents

*Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin; email: liero@wias-berlin.de. Partially supported by the Einstein Stiftung Berlin via the ECMath/MATHEON project SE2.

†Humboldt-Universität zu Berlin; email: mielke@wias-berlin.de, Partially supported by DFG via project C5 within CRC 1114 (Scaling cascades in complex systems) and by ERC via AdG. 267802 AnaMultiScale.

‡Università di Pavia; email: giuseppe.savare@unipv.it. Partially supported by PRIN10/11 grant from MIUR for the project *Calculus of Variations*.

1 Introduction

The aim of the present paper is twofold: In Part I we develop a full theory of the new class of *Optimal Entropy-Transport problems* between nonnegative and finite Radon measures in general topological spaces. As a powerful application of this theory, in Part II we study the particular case of *Logarithmic* Entropy-Transport problems and introduce the new *Hellinger-Kantorovich* (HK) *distance between measures in metric spaces*. The striking connection between these two seemingly far topics is our main focus, and it paves the way for a beautiful and deep analysis of the geometric properties of the geodesic HK distance, which (as our proposed name suggests) can be understood as an inf-convolution of the well-known Hellinger-Kakutani and the Kantorovich-Wasserstein distances. In fact, our approach to the theory was opposite: in trying to characterize HK, we were first led to the Logarithmic Entropy-Transport problem, see Section A.

From Transport to Entropy-Transport problems. In the classical Kantorovich formulation, Optimal Transport problems [34, 43, 2, 44] deal with minimization of a linear cost functional

$$\mathscr{C}(\boldsymbol{\gamma}) = \int_{X_1 \times X_2} \mathsf{c}(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma}(x_1, x_2), \quad \mathsf{c} : X_1 \times X_2 \to \mathbb{R}, \tag{1.1}$$

among all the transport plans, i.e. probability measures in $\mathcal{P}(X_1 \times X_2)$, $\boldsymbol{\gamma}$ whose marginals $\mu_i = \pi^i_\sharp \boldsymbol{\gamma} \in \mathcal{P}(X_i)$ are prescribed. Typically, X_1, X_2 are Polish spaces, μ_i are given Borel measures (but the case of Radon measures in Hausdorff topological spaces has also been considered, see [22, 34]), the cost function c is a lower semicontinuous (or even Borel) function, possibly assuming the value $+\infty$, and $\pi^i(x_1, x_2) = x_i$ are the projections on the i-th coordinate, so that

$$\pi^i_\sharp \boldsymbol{\gamma} = \mu_i \quad \Leftrightarrow \quad \mu_1(A_1) = \boldsymbol{\gamma}_1(A_1 \times X_2),\ \mu_2(A_2) = \boldsymbol{\gamma}_1(X_1 \times A_2) \quad \text{for every } A_i \in X_i. \tag{1.2}$$

Starting from the pioneering work of Kantorovich, an impressive theory has been developed in the last two decades: from one side, typical intrinsic questions of linear programming problems concerning duality, optimality, uniqueness and structural properties of optimal transport plans have been addressed and fully analyzed. In a parallel way, this rich general theory has been applied to many challenging problems in a variety of fields (probability and statistics, functional analysis, PDEs, Riemannian geometry, nonsmooth analysis in metric spaces, just to mention a few of them: since it is impossible here to give an even partial account of the main contributions, we refer to the books [44, 36] for a more detailed overview and a complete list of references).

The class of ***Entropy-Transport problems***, we are going to study, arises quite naturally if one tries to relax the marginal constraints $\pi^i_\sharp \boldsymbol{\gamma} = \mu_i$ by introducing suitable penalizing functionals $\mathscr{F}_i$, that quantify in some way the deviation from μ_i of the marginals $\gamma_i := \pi^i_\sharp \boldsymbol{\gamma}$ of $\boldsymbol{\gamma}$. In this paper we consider the general case of integral functionals (also called *Csiszàr f-divergences* [15]) of the form

$$\mathscr{F}_i(\gamma_i|\mu_i) := \int_{X_i} F_i(\sigma_i(x_i))\,\mathrm{d}\mu_i + \gamma_i^\perp(X_i), \quad \sigma_i = \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}, \quad \gamma_i = \sigma_i \mu_i + \gamma_i^\perp, \tag{1.3}$$

where $F_i : [0, +\infty) \to [0, +\infty]$ are given convex entropy functions, like for the logarithmic or power-like entropies

$$
\begin{aligned}
&U_p(s) := \frac{1}{p(p-1)}\big(s^p - p(s-1) + 1\big), \quad p \in \mathbb{R} \setminus \{0, 1\}, \\
&U_0(s) := s - 1 - \log s, \quad U_1(s) := s \log s - s + 1.
\end{aligned}
\tag{1.4}
$$

Notice that the presence of the singular part $\gamma_i^\perp$ in the Lebesgue decomposition of γ_i in (1.3) does not force $F_i(s)$ to be superlinear as $s \uparrow +\infty$ and allows for all the exponents p in (1.4).

Once a specific choice of entropies F_i and of finite nonnegative Radon measures $\mu_i \in \mathcal{M}(X_i)$ is given, the Entropy-Transport problem can be formulated as

$$
\mathsf{ET}(\mu_1, \mu_2) := \inf\Big\{\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) : \boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)\Big\}, \tag{1.5}
$$

where $\mathscr{E}$ is the convex functional

$$
\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) := \mathscr{F}_1(\gamma_1|\mu_1) + \mathscr{F}_2(\gamma_2|\mu_2) + \int_{X_1 \times X_2} \mathsf{c}(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma}. \tag{1.6}
$$

Notice that the entropic formulation allows for measures μ_1, μ_2 and $\boldsymbol{\gamma}$ with possibly different total mass.

The flexibility in the choice of the entropy functions F_i (which may also take the value $+\infty$) covers a wide spectrum of situations (see Section 3.3 for various examples) and in particular guarantees that (1.5) is a real generalization of the classical optimal transport problem, which can be recovered as a particular case of (1.6) when $F_i(s)$ is the indicator function of $\{1\}$ (i.e. $F_i(s)$ always takes the value $+\infty$ with the only exception of $s = 1$, where it vanishes).

Since we think that the structure (1.6) of Entropy-Transport problems will lead to new and interesting models and applications, we have tried to establish their basic theory in the greatest generality, by pursuing the same line of development of Transport problems: in particular we will obtain general results concerning existence, duality and optimality conditions.

Considering e.g. the Logarithmic Entropy case, where $F_i(s) = s \log s - (s-1)$, ***the dual formulation of*** (1.5) is given by

$$
\begin{aligned}
&\mathsf{D}(\mu_1, \mu_2) := \sup\Big\{\mathscr{D}(\varphi_1, \varphi_2|\mu_1, \mu_2) \;:\; \varphi_i : X_i \to \mathbb{R},\ \varphi_1(x_1) + \varphi_2(x_2) \le \mathsf{c}(x_1, x_2)\Big\}, \\
&\text{where } \mathscr{D}(\varphi_1, \varphi_2|\mu_1, \mu_2) := \int_{X_1}\big(1 - \mathrm{e}^{-\varphi_1}\big)\,\mathrm{d}\mu_1 + \int_{X_2}\big(1 - \mathrm{e}^{-\varphi_2}\big)\,\mathrm{d}\mu_2,
\end{aligned}
\tag{1.7}
$$

where one can immediately recognize the same convex constraint of Transport problems: the couple of dual potentials φ_i should satisfy $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$ on $X_1 \times X_2$. The main difference is due to the concavity of the objective functional

$$
(\varphi_1, \varphi_2) \mapsto \int_{X_1} \big(1 - \mathrm{e}^{-\varphi_1}\big)\,\mathrm{d}\mu_1 + \int_{X_2} \big(1 - \mathrm{e}^{-\varphi_2}\big)\,\mathrm{d}\mu_2,
$$

whose form can be explicitly calculated in terms of the Lagrangian conjugates F_i^* of the entropy functions. The change of variables $\psi_i := 1 - \mathrm{e}^{-\varphi_i}$ transforms (1.7) in the equivalent problem of maximizing the linear functional

$$(\psi_1, \psi_2) \mapsto \sum_i \int_{X_1} \psi_1 \,\mathrm{d}\mu_1 + \int_{X_2} \psi_2 \,\mathrm{d}\mu_2 \tag{1.8}$$

on the more complicated convex set

$$\Big\{(\psi_1, \psi_2) : \psi_i : X_i \to (-\infty, 1), \quad (1 - \psi_1(x_1))(1 - \psi_2(x_2)) \geq \mathrm{e}^{-\mathsf{c}(x_1,x_2)}\Big\}. \tag{1.9}$$

We will calculate the dual problem for every choice of F_i and show that its value always coincide with $\mathsf{ET}(\mu_1, \mu_2)$. The dual problem also provides ***optimality conditions,*** that involve the couple of potentials (φ_1, φ_2), the support of the optimal plan $\boldsymbol{\gamma}$ and the densities σ_i of its marginals γ_i w.r.t. μ_i. For the Logarithmic Entropy Transport problem above, they read as

$$\begin{gathered}\sigma_i > 0, \ \varphi_i = -\log \sigma_i \quad \mu_i \text{ a.e. in } X_i,\\ \varphi_1 \oplus \varphi_2 \leq \mathsf{c} \quad \text{in } X_1 \times X_2, \quad \varphi_1 \oplus \varphi_2 = \mathsf{c} \quad \boldsymbol{\gamma}\text{-a.e. in } X_1 \times X_2,\end{gathered} \tag{1.10}$$

and they are necessary and sufficient for optimality.

The study of optimality conditions reveals a different behavior between pure transport problems and the other entropic ones. In particular, the c-cyclical monotonicity of the optimal plan $\boldsymbol{\gamma}$ (which is still satisfied in the entropic case) does not play a crucial role in the construction of the potentials φ_i. When $F_i(0)$ are finite (as in the logarithmic case) it is possible to obtain a general existence result of (generalized) optimal potentials even when c takes the value $+\infty$.

A crucial feature of Entropy-Transport problems (which is not shared by the pure transport ones) concerns a ***third "homogeneous" formulation***, which exhibits new and unexpected properties. It is related to the 1-***homogeneous Marginal Perspective function***

$$H(x_1, r_1; x_2, r_2) := \inf_{\theta > 0} \Big(r_1 F_1(\theta/r_1) + r_2 F_2(\theta/r_2) + \theta \mathsf{c}(x_1, x_2)\Big) \tag{1.11}$$

and to the corresponding integral functional

$$\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) := \int_{X_1 \times X_2} H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i^{\perp}(X_i), \ \varrho_i := \frac{\mathrm{d}\mu_i}{\mathrm{d}\gamma_i}, \tag{1.12}$$

where $\mu_i = \varrho_i \gamma_i + \mu_i^{\perp}$ is the "reverse" Lebesgue decomposition of μ_i w.r.t. the marginals γ_i of $\boldsymbol{\gamma}$. We will prove that

$$\mathsf{ET}(\mu_1, \mu_2) = \min \Big\{\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)\Big\} \tag{1.13}$$

with a precise relation between optimal plans. In the Logarithmic Entropy case $F_i(s) = s \log s - (s - 1)$ the marginal perspective function H takes the particular form

$$H(x_1, r_1; x_2, r_2) = r_1 + r_2 - 2\sqrt{r_1 r_2}\, \mathrm{e}^{-\mathsf{c}(x_1,x_2)/2}, \tag{1.14}$$

which will be the starting point for understanding the deep connection with the Hellinger-Kantorovich distance. Notice that in the case when $X_1 = X_2$ and c is the singular cost

$$\mathsf{c}(x_1, x_2) := \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise,} \end{cases} \tag{1.15}$$

(1.13) provides an equivalent formulation of the Hellinger-Kakutani distance [19, 21], see also Example E.5 in Section 3.3.

Other choices, still in the simple class (1.4), give raise to "transport" versions of well known functionals (see e.g. [28] for a systematic presentation): starting from the reversed entropies $F_i(s) = s - 1 - \log s$ one gets

$$H(x_1, r_1; x_2, r_2) = r_1 \log r_1 + r_2 \log r_2 - (r_1 + r_2) \log \Big(\frac{r_1 + r_2}{2 + \mathsf{c}(x_1, x_2)}\Big), \tag{1.16}$$

which in the extreme case (1.15) reduces to the *Jensen-Shannon divergence* [29], a squared distance between measures derived from the celebrated *Kullback-Leibler divergence* [25]. The quadratic entropy $F_i(s) = \frac{1}{2}(s - 1)^2$ produces

$$H(x_1, r_1; x_2, r_2) = \frac{1}{2(r_1 + r_2)} \Big((r_1 - r_2)^2 + h(\mathsf{c}(x_1, x_2)) r_1 r_2\Big), \tag{1.17}$$

where $h(c) = c(4 - c)$ if $0 \le c \le 2$ and 4 if $c \ge 2$: Equation (1.17) can be seen as the transport variant of the *triangular discrimination* (also called symmetric χ^2-measure), based on the *Pearson* χ^2-divergence, and still obtained by (1.12) when c has the form (1.15).

Also nonsmooth cases, as for $F_i(s) = |s - 1|$ associated to the *total variation distance*, or nonsymmetric choices of F_i can be covered by the general theory. However, because of our original motivation (see Section A), Part II will focus on the case of the logarithmic entropy $F_i = U_1$, where H is given by (1.14). We will exploit its relevant geometric applications, reserving the other examples for future investigations.

From the Kantorovich-Wasserstein distance to the Hellinger-Kantorovich distance. From the analytic-geometric point of view, one of the most interesting cases of transport problems occurs when $X_1 = X_2 = X$ coincide and the cost functional $\mathscr{C}$ is induced by a distance d on X: in the quadratic case, the minimum value of (1.1) for given measures μ_1, μ_2 in the space $\mathcal{P}_2(X)$ of probability measures with finite quadratic moment defines the so called L^2-Kantorovich-Wasserstein distance

$$\mathsf{W}_{\mathsf{d}}^2(\mu_1, \mu_2) := \inf \Big\{ \int \mathsf{d}^2(x_1, x_2) \,\mathrm{d}\boldsymbol{\gamma}(x_1, x_2) : \boldsymbol{\gamma} \in \mathcal{P}(X \times X),\ \pi^i_\sharp \boldsymbol{\gamma} = \mu_i \Big\}, \tag{1.18}$$

which metrizes the weak convergence (with quadratic moments) of probability measures. The metric space $(\mathcal{P}_2(X), \mathsf{W}_{\mathsf{d}})$ inherits many geometric features from the underlying (X, d) (as separability, completeness, length and geodesic properties, positive curvature in the Alexandrov sense, see [2]). Its dynamic characterization in terms of the continuity equation [7] and its dual formulation in terms of the Hopf-Lax formula and the corresponding (sub-)solutions of the Hamilton-Jacobi equation [33] lie at the core of the applications to gradient flows and partial differential equations of diffusion type [2]. Finally, the behavior of entropy functionals as (1.3) along geodesics in $(\mathcal{P}_2(X), \mathsf{W}_\mathsf{d})$ [32, 33, 14] encodes a valuable geometric information, with relevant applications to Riemannian geometry and to the recent theory of metric-measure spaces with Ricci curvature bounded from below [41, 42, 31, 3, 4, 5, 18].

It has been a challenging question to find a corresponding distance (enjoying analogous deep geometric properties) between finite positive Borel measures with arbitrary mass in $\mathcal{M}(X)$. In the present paper we will show that by choosing the particular cost function

$$\mathsf{c}(x_1,x_2) := \ell(\mathsf{d}(x_1,x_2)), \quad \text{where} \quad \ell(\mathsf{d}) := \begin{cases} -\log\big(\cos^2(\mathsf{d})\big) & \text{if } \mathsf{d} < \pi/2, \\ +\infty & \text{otherwise,} \end{cases} \tag{1.19}$$

the corresponding Logarithmic-Entropy Transport problem

$$\mathsf{LET}(\mu_1,\mu_2) := \min_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \sum_i \int_X \big(\sigma_i \log\sigma_i - \sigma_i + 1\big)\,\mathrm{d}\mu_i + \int_{X^2} \ell\big(\mathsf{d}(x_1,x_2)\big)\,\mathrm{d}\boldsymbol{\gamma}, \quad \sigma_i = \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}, \tag{1.20}$$

provides a (squared) distance $\mathsf{HK}^2(\mu_1,\mu_2)$ in $\mathcal{M}(X)$ that can play the same fundamental role like the Kantorovich-Wasserstein distance for $\mathcal{P}_2(X)$.

Here is a schematic list of our main results:

(i) The representation (1.13) based on the Marginal Perspective function (1.14) yields

$$\mathsf{LET}(\mu_1,\mu_2) = \min\Big\{ \int \Big(\varrho_1 + \varrho_2 - 2\varrho_1\varrho_2 \cos(\mathsf{d}(x_1,x_2)\wedge\pi/2)\Big)\,\mathrm{d}\boldsymbol{\gamma} : \varrho_i = \frac{\mathrm{d}\mu_i}{\mathrm{d}\gamma_i}\Big\}. \tag{1.21}$$

(ii) By performing the rescaling $r_i \mapsto r_i^2$ we realize that the function $H(x_1, r_1^2; x_2, r_2^2)$ is strictly related to the squared (semi)-distance

$$\mathsf{d}_\mathfrak{C}^2(x_1,r_1;x_2,r_2) := r_1^2 + r_2^2 - 2r_1r_2\cos(\mathsf{d}(x_1,x_2)\wedge\pi), \quad (x_i,r_i)\in X\times\mathbb{R}_+ \tag{1.22}$$

which is the so-called *cone distance* in the metric cone $\mathfrak{C}$ over X, cf. [9]. The latter is the quotient space of $X\times\mathbb{R}_+$ obtained by collapsing all the points $(x,0)$, $x\in X$, in a single point $\mathfrak{o}$, called the vertex of the cone. We introduce the notion of "2-homogeneous marginal"

$$\mu = \mathfrak{h}^2\alpha := \pi^x_\sharp(r^2\alpha), \qquad \int_X \zeta(x)\,\mathrm{d}\mu = \int_\mathfrak{C} \zeta(x) r^2\,\mathrm{d}\alpha(x,r) \quad \text{for every } \zeta\in \mathrm{C}_b(X), \tag{1.23}$$

to "project" measures $\alpha\in\mathcal{M}(\mathfrak{C})$ on measures $\mu\in\mathcal{M}(X)$. Conversely, there are many ways to "lift" a measure $\mu\in\mathcal{M}(X)$ to $\alpha\in\mathcal{M}(\mathfrak{C})$ (e.g. by taking $\alpha := \mu\otimes\delta_1$). It turns out that the best Kantorovich-Wasserstein distance between all the possible lifts of μ_1,μ_2 in $\mathcal{P}_2(\mathfrak{C})$ yields an equivalent variational representation of the HK distance, i.e.

$$\mathsf{HK}(\mu_1,\mu_2) = \min\Big\{ \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_1,\alpha_2) : \alpha_i\in\mathcal{P}_2(\mathfrak{C}),\ \mathfrak{h}^2\alpha_i = \mu_i \Big\}. \tag{1.24}$$

In particular, (1.24) shows that in the case of concentrated measures

$$\mathsf{LET}(a_1\delta_{x_1}, a_2\delta_{x_2}) = \mathsf{HK}^2(a_1\delta_{x_1}, a_2\delta_{x_2}) = \mathsf{d}_\mathfrak{C}^2(x_1,a_1;x_2,a_2). \tag{1.25}$$

Notice that (1.24) resembles the very definition (1.18) of the Kantorovich-Wasserstein distance, where now the role of the marginals $\pi^i_\sharp$ is replaced by the homogeneous marginals $\mathfrak{h}^2$. It is a nontrivial part of the equivalence statement to check that the difference between the cut-off thresholds ($\pi/2$ in (1.21) and π in (1.22) does not affect the identity $\mathsf{LET} = \mathsf{HK}^2$.

(iii) By refining the representation formula (1.24) by a suitable rescaling and gluing technique we can prove that $(\mathcal{M}(X), \mathsf{HK})$ is a geodesic metric space, a property that it is absolutely not obvious from the LET-representation and depends on a subtle interplay of the entropy functions $F_i(\sigma) = \sigma log\sigma - \sigma + 1$ and the cost function c from (1.19). We show that the metric induces the weak convergence of measures in duality with bounded and continuous functions. It also inherits the separability, completeness, length and geodesic properties from the correspondent ones of the underlying space (X, d). On top of that, we will prove a precise superposition principle (in the same spirit of the Kantorovich-Wasserstein one [2, Sect.8],[30]) for general absolutely continuous curves in $(\mathcal{M}(X), \mathsf{HK})$ in terms of dynamic plans in $\mathfrak{C}$: as a byproduct, we can give a precise characterization of absolutely continuous curves and geodesics as homogeneous marginals of corresponding curves in $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_\mathfrak{C}})$. An interesting consequence of these results concerns the lower curvature bound of $(\mathcal{M}(X), \mathsf{HK})$ in the sense of Alexandrov: it is a positively curved space if and only if (X, d) is a geodesic space with curvature ≥ 1.

(iv) The dual formulation of the LET problem provides a dual characterization of HK, viz.

$$\frac{1}{2}\mathsf{HK}^2(\mu_1, \mu_2) = \sup\Big\{ \int \mathscr{P}_1\xi \,\mathrm{d}\mu_2 - \int \xi \,\mathrm{d}\mu_1 : \xi \in \mathrm{Lip}_b(X),\ \inf_X \xi > -1/2 \Big\}, \quad (1.26)$$

where $(\mathscr{P}_t)_{0\leq t\leq 1}$ is given by the inf-convolution

$$\mathscr{P}_t\xi(x) := \inf_{x'\in X} \frac{\xi(x')}{1+2t\xi(x')} + \frac{\sin^2(\mathsf{d}_{\pi/2}(x,x'))}{2+4t\xi(x')} = \inf_{x'\in X} \frac{1}{t}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x,x'))}{1+2t\xi(x')}\Big).$$

(v) By exploiting the Hopf-Lax representation formula for the Hamilton-Jacobi equation in $\mathfrak{C}$, we will show that for arbitrary initial data $\xi \in \mathrm{Lip}_b(X)$ with $\inf \xi > -1/2$ the function $\xi_t := \mathscr{P}_t\xi$ is a subsolution (a solution, if (X, d) is a length space) of

$$\partial_t^+ \xi_t(x) + \frac{1}{2}|\mathrm{D}_X \xi_t|^2(x) + 2\xi_t^2(x) \leq 0 \quad \text{pointwise in } X \times (0,1). \quad (1.27)$$

If (X, d) is a length space we thus obtain the characterization

$$\begin{aligned}\frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) = \sup\Big\{ \int_X \xi_1 \,\mathrm{d}\mu_1 - \int_0 \xi_0 \,\mathrm{d}\mu_0 : \xi \in \mathrm{C}^k([0,1]; \mathrm{Lip}_b(X)),\\ \partial_t\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) + 2\xi_t^2(x) \leq 0 \quad \text{in } X \times (0,1)\Big\},\end{aligned} \quad (1.28)$$

which reproduces, at the level of HK, the nice link between W_d and Hamilton-Jacobi equations. One of the direct applications of (1.28) is a sharp contraction property w.r.t. HK for the Heat flow in $\mathrm{RCD}(0, \infty)$ metric measure spaces (and therefore in every Riemannian manifold with nonnegative Ricci curvature).

(vi) (1.28) clarifies that the HK distance can be interpreted as a sort of inf-convolution between the Hellinger (in duality with solutions to the ODE $\partial_t\xi + 2\xi_t^2 = 0$) and the Kantorovich-Wasserstein distance (in duality with (sub-)solutions to $\partial_t\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) \le 0$). The Hellinger distance

$$\mathsf{Hell}^2(\mu_1,\mu_2) = \int_X \left(\sqrt{\varrho_1} - \sqrt{\varrho_2}\right)^2 \mathrm{d}\gamma, \quad \mu_i = \varrho_i\gamma,$$

corresponds to the HK functional generated by the discrete distance ($\mathsf{d}(x_1,x_2) = \pi/2$ if $x_1 \neq x_2$). We will prove that

$$\mathsf{HK}(\mu_1,\mu_2) \le \mathsf{Hell}(\mu_1,\mu_2), \quad \mathsf{HK}(\mu_1,\mu_2) \le \mathsf{W}_{\mathsf{d}}(\mu_1,\mu_2),$$
$$\mathsf{HK}_{n\mathsf{d}}(\mu_1,\mu_2) \uparrow \mathsf{Hell}(\mu_1,\mu_2), \quad n\mathsf{HK}_{\mathsf{d}/n} \uparrow \mathsf{W}_{\mathsf{d}}(\mu_1,\mu_2) \quad \text{as } n\uparrow\infty,$$

where $\mathsf{HK}_{n\mathsf{d}}$ (resp. $\mathsf{HK}_{\mathsf{d}/n}$) is the HK distance induced by $n\mathsf{d}$ (resp. d/n).

(vii) Combining the superposition principle and the duality with Hamilton-Jacobi equations, we eventually prove that HK admits an equivalent dynamic characterization "à la Benamou-Brenier" [7, 16] (see also the recent [23]) in $X = \mathbb{R}^d$

$$\begin{aligned}\mathsf{HK}^2(\mu_0,\mu_1) = \min\Big\{ \int_0^1 \int \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t : \mu \in \mathrm{C}([0,1];\mathcal{M}(\mathbb{R}^d)),\\ \mu_{t=i} = \mu_i,\ \partial_t\mu_t + \nabla\cdot(\boldsymbol{v}_t\mu_t) = w_t\mu_t \text{ in } \mathscr{D}'(\mathbb{R}^d\times(0,1))\Big\}.\end{aligned} \tag{1.29}$$

Moreover, for the length space $X = \mathbb{R}^d$ a curve $[0,1] \ni t \mapsto \mu(t)$ is geodesic curve w.r.t. HK if and only if the coupled system

$$\partial_t\mu_t + \nabla\cdot(\mathrm{D}_x\xi_t\mu_t) = 4\xi_t\mu_t, \quad \partial_t\xi_t + \frac{1}{2}|\mathrm{D}_x\xi^2|^2 + 2\xi_t^2 = 0 \tag{1.30}$$

holds for a suitable solution $\xi_t = \mathscr{P}_t\xi_0$. The representation (1.29) is the starting point for further investigations and examples, which we have collected in [27].

It is not superfluous to recall that the HK variational problem is just one example in the realm of Entropy-Transport problems and we think that other interesting applications can arise by different choices of entropies and cost. One of the simplest variation is to choose the (seemingly more natural) quadratic cost function $\mathsf{c}(x_1,x_2) := \mathsf{d}^2(x_1,x_2)$ instead of the more "exotic" (1.19). The resulting functional is still associated to a distance expressed by

$$\mathsf{GHK}^2(\mu_1,\mu_2) := \min\Big\{ \int \Big(r_1^2 + r_2^2 - 2r_1r_2\exp(-\mathsf{d}^2(x_1,x_2)/2)\Big)\,\mathrm{d}\boldsymbol{\alpha}\Big\} \tag{1.31}$$

where the minimum runs among all the plans $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C}\times\mathfrak{C})$ such that $\mathfrak{h}^2\pi^i_\sharp\boldsymbol{\alpha} = \mu_i$ (we propose the name "Gaussian Hellinger-Kantorovich distance"). If (X,d) is a complete, separable and length metric space, $(\mathcal{M}(X),\mathsf{GHK})$ is a complete and separable metric space, inducing the weak topology as HK. However, it is not a length space in general, and we will show that the length distance generated by GHK is precisely HK.

The plan of the paper is as follows.

Part I develops the general theory of Optimal Entropy-Transport problems. Section 2 collects some preliminary material, in particular concerning the measure-theoretic setting in arbitrary Hausdorff topological spaces (here we follow [38]) and entropy functionals. We devote some effort to deal with general functionals (allowing a singular part in the definition (1.3)) in order to include entropies which may have only linear growth. The extension to this general framework of the duality theorem 2.7 (well known in Polish topologies) requires some care and the use of lower semicontinuous test functions instead of continuous ones.

Section 3 introduces the class of Entropy-Transport problems, discussing same examples and proving a general existence result for optimal plans. The "reverse" formulation of Theorem 3.11, though simple, justifies the importance to deal with the largest class of entropies and will play a crucial role in Section 5.

Section 4 is devoted to find the dual formulation, to prove its equivalence with the primal problem (cf. Theorem 4.11), to derive sharp optimality conditions (cf. Theorem 4.6) and to prove the existence of optimal potentials in a suitable generalized sense (cf. Theorem 4.15). The particular class of "regular" problems (where the results are richer) is also studied with some details.

Section 5 introduces the third formulation (1.12) based on the marginal perspective function (1.11) and its "homogeneous" version (Section 5.2). The proof of the equivalence with the previous formulations is presented in Theorem 5.5 and Theorem 5.8. This part provides the crucial link for the further development in the cone setting.

Part II is devoted to Logarithmic Entropy-Transport (LET) problems (Section 6) and to their applications to the Hellinger-Kantorovich distance HK on $\mathcal{M}(X)$.

The Hellinger-Kantorovich distance is introduced by the lifting technique in the cone space in Section 7, where we try to follow a presentation modeled on the standard one for the Kantorovich-Wasserstein distance, independently from the results on the LET-problems. After a brief recap on the cone geometry (Section 7.1) we discuss in some detail the crucial notion of homogeneous marginals in Section 7.2 and the useful tightness conditions (Lemma 7.3) for plans with prescribed homogeneous marginals. Section 7.3 introduces the definition of the HK distance and its basic properties. The crucial rescaling and gluing techniques are discussed in Section 7.4: they lie at the core of the main metric properties of HK, leading to the proof of the triangle inequality and to the characterizations of various metric and topological properties in Section 7.5. The equivalence with the LET formulation is the main achievement of Section 7.6 (Theorem 7.21), with applications to the duality formula (Theorem 7.22), to the comparisons with the classical Hellinger and Kantorovich distances (Section 7.7) and with the Gaussian Hellinger-Kantorovich distance (Section 7.8).

The last Section of the paper collects various important properties of HK, that share a common "dynamic" flavor. After a preliminary discussion of absolutely continuous curves and geodesics in the cone space $\mathfrak{C}$ in Section 8.1, we derive the basic superposition principle in Theorem 8.4. This is the cornerstone to obtain a precise characterization of geodesics (Theorem 8.7), a sharp lower curvature bound in the Alexandrov sense (Theorem 8.9) and to prove the dynamic characterization à la Benamou-Brenier of Section 8.5. The other powerful tool is provided by the duality with subsolutions to the Hamilton-Jacobi equation (Theorem 8.13), which we derive after a preliminary characterization of metric slopes for a suitable class of test functions in $\mathfrak{C}$. One of the most striking results of Section 8.4 is the explicit representation formula for solutions to the Hamilton-Jacobi equation in X, that we obtain by a careful reduction technique from the Hopf-Lax formula in $\mathfrak{C}$. In this respect, we think that Theorem 8.12 is interesting by itself and could find important applications in different contexts. From the point of view of Entropy-Transport problems, Theorem 8.12 is particularly relevant since it provides a dynamic interpretation of the dual characterization of the LET functional. In Section 8.6 we show that in the Euclidean case $X = \mathbb{R}^d$ all geodesic curves are characterized by the system (1.30). The last Section 8.7 provides various contraction results: in particular we extend the well known contraction property of the Heat flow in spaces with nonnegative Riemannian Ricci curvature to HK.

Note during final preparation. The earliest parts of the work presented here were first presented at the ERC Workshop on Optimal Transportation and Applications in Pisa in 2012. Since then the authors developed the theory continuously further and presented results at different workshops and seminars. We refer to Section A for some remarks concerning the chronological development of our theory. In June 2015 they became aware of the parallel work [24]. Moreover, in mid August 2015 we became aware of [11, 12]. So far, these independent works are not reflected in the present version of this manuscript.

Main notation

$\mathcal{M}(X)$	finite positive Radon measures on a Hausdorff topological space X
$\mathcal{P}(X),\ \mathcal{P}_2(X)$	Radon probability measures on X (with finite quadratic moment)
$\mathcal{B}(X)$	Borel subsets of X
$T_\sharp\mu$	push forward of $\mu \in \mathcal{M}(X)$ by a map $T : X \to Y$: (2.5)
$\gamma = \sigma\mu + \mu^\perp,\ \mu = \varrho\gamma + \gamma^\perp$	Lebesgue decompositions of γ and μ, Lemma 2.3
$\mathrm{C}_b(X)$	continuous and bounded real functions on X
$\mathrm{Lip}_b(X)$, $\mathrm{Lip}_{bs}(X)$	bounded (with bounded support) Lipschitz real functions on X
$\mathrm{LSC}_b(X), \mathrm{LSC}_s(X)$	lower semicontinuous and bounded (or simple) real functions on X
$\mathrm{USC}_b(X), \mathrm{USC}_s(X)$	upper semicontinuous and bounded (or simple) real functions on X
$\mathrm{B}(X), \mathrm{B}_b(X)$	Borel (resp. bounded Borel) real functions
$\mathrm{L}^p(X,\mu),\ \mathrm{L}^p(X,\mu;\mathbb{R}^d)$	Borel μ-integrable real (or $\mathbb{R}^d$-valued) functions
$\Gamma(\mathbb{R}_+)$	set of admissible entropy functions, see (2.14), (2.15).
$F(s), F_i(s)$	admissible entropy functions.
$F^*(\phi), F_i^*(\phi)$	Legendre transform of F, F_i, see (2.18).
$F^\circ(\varphi), F_i^\circ(\varphi_i)$	concave conjugate of an entropy function, see (2.44).
$R(r), R_i(r_i)$	reversed entropies, see (2.29).
$H_c(r_1, r_2),\ H(x_1, r_1; x_2, r_2)$	marginal perspective function, see (5.1), (5.9), (5.3)
$\mathsf{c}(x_1, x_2)$	lower semicontinuous cost function defined in $\boldsymbol{X} = X_1 \times X_2$.
$\mathscr{F}(\gamma\|\mu), \mathscr{R}(\mu\|\gamma)$	entropy functionals and their reverse form, see (2.35) and (2.56)
$\mathscr{E}(\boldsymbol{\gamma}\|\mu_1, \mu_2), \mathsf{ET}(\mu_1, \mu_2)$	general Entropy-Transport functional and its minimum, see (3.4)
$\mathscr{D}(\boldsymbol{\varphi}\|\mu_1, \mu_2), \mathsf{D}(\mu_1, \mu_2)$	dual functional and its supremum, see (4.10) and (4.8)
$\boldsymbol{\Phi}, \boldsymbol{\Psi}$	set of admissible Entropy-Kantorovich potentials
$\mathsf{LET}(\mu_1, \mu_2),\ \ell(\mathsf{d})$	Logarithmic Entropy Transport functional and its cost: Section 6.1
$\mathsf{W}_\mathsf{d}(\mu_1, \mu_2)$	Kantorovich-Wasserstein distance in $\mathcal{P}_2(X)$
$\mathsf{HK}(\mu_1, \mu_2)$	Hellinger-Kantorovich distance in $\mathcal{M}(X)$: Section 7.3
$\mathsf{GHK}(\mu_1, \mu_2)$	Gaussian Hellinger-Kantorovich distance in $\mathcal{M}(X)$: Section 7.8

$(\mathfrak{C},\ \mathsf{d}_{\mathfrak{C}}),\ \mathfrak{o}$	metric cone and its vertex, see Section 7.1
$\mathfrak{C}[r]$	ball of radius r centered at $\mathfrak{o}$ in $\mathfrak{C}$
$\mathfrak{h}^2_i,\ \mathrm{dil}_{\theta,2}(\cdot)$	homogeneous marginals and dilations, see (7.14), (7.15)
$\mathfrak{H}^2_{=}(\mu_1,\mu_2),\ \mathfrak{H}^2_{\le}(\mu_1,\mu_2)$	plans in $\mathfrak{C}\times\mathfrak{C}$ with constrained homogeneous marginals, see (7.19)
$\mathrm{AC}^p([0,1];X)$	space of curves $\mathrm{x}:[0,1]\to X$ with p-integrable metric speed
$\lvert\mathrm{x}'\rvert_{\mathsf{d}}$	metric speed of a curve $\mathrm{x}\in\mathrm{AC}([a,b];(X,\mathsf{d}))$, Sect. 8.1
$\lvert\mathrm{D}_Z f\rvert,\ \lvert\mathrm{D}_Z f\rvert_a$	metric slope and asymptotic Lipschitz constant in Z, see (8.33)

Part I. Optimal Entropy-Transport problems

2 Preliminaries

2.1 Measure theoretic notation

Positive Radon measures, narrow and weak convergence, tightness. Let (X,τ) be a Hausdorff topological space. We will denote by $\mathcal{B}(X)$ the σ-algebra of its Borel sets and by $\mathcal{M}(X)$ the set of finite nonnegative *Radon* measures on X [38], i.e. σ-additive set functions $\mu:\mathcal{B}(X)\to[0,\infty)$ such that

$$\forall\, B\in\mathcal{B}(X),\ \forall\,\varepsilon>0\quad \exists\, K_\varepsilon\subset B \text{ compact such that}\quad \mu(B\setminus K_\varepsilon)\le\varepsilon. \tag{2.1}$$

Radon measures have strong continuity property with respect to monotone convergence: in particular if $(f_\lambda)_{\lambda\in\mathbb{L}}\subset\mathrm{LSC}(X)$ is a nondecreasing directed family (defined in an even uncountable directed set $\mathbb{L}$) of nonnegative and lower semicontinuous functions converging to f, we have [38, Prop. 5, p. 42]

$$\lim_{\lambda\in\mathbb{L}}\int_X f_\lambda\,\mathrm{d}\mu=\int_X f\,\mathrm{d}\mu,\quad \mu\in\mathcal{M}(X). \tag{2.2}$$

We endow $\mathcal{M}(X)$ with the *narrow* topology, the coarsest (Hausdorff) topology for which all the maps $\mu\mapsto\int_X\varphi\,\mathrm{d}\mu$ are lower semicontinuous, as $\varphi:X\to\mathbb{R}$ varies among the set $\mathrm{LSC}_b(X)$ of all bounded lower semicontinuous functions [38, p. 370, Def. 1].

Remark 2.1 (Radon versus Borel, narrow versus weak)**.** When (X,τ) is a Radon space (in particular a Polish, or Lusin or Souslin space [38, p. 122]) then every Borel measure satisfies (2.1), so that $\mathcal{M}(X)$ coincides with the set of all nonnegative and finite *Borel* measures. Narrow topology is in general stronger than the standard weak topology induced by the duality with continuous and bounded functions of $\mathrm{C}_b(X)$. However, when (X,τ) is completely regular, i.e.

$$\begin{gathered}\text{for any closed set } F\subset X \text{ and any } x_0\in X\setminus F\\ \text{there exists } f\in C_b(X) \text{ with } f(x_0)>0 \text{ and } f\equiv 0 \text{ on } F,\end{gathered} \tag{2.3}$$

(in particular when τ is metrizable), narrow and weak topology coincide [38, p. 371]. Therefore when (X,τ) is a Polish space we recover the usual setting of Borel "probability deleted measures endowed with the weak topology. □

A set $\mathcal{K} \subset \mathcal{M}(X)$ is *bounded* if $\sup_{\mu\in\mathcal{K}} \mu(X) < \infty$; it is *equally tight* if

$$\forall\, \varepsilon > 0 \quad \exists\, K_\varepsilon \subset X \text{ compact such that } \quad \mu(X \setminus K_\varepsilon) \le \varepsilon \quad \text{for every } \mu \in \mathcal{K}. \tag{2.4}$$

Compactness with respect to narrow topology is guaranteed by an extended version of *Prokhorov Theorem* [38, Thm. 3, p. 379]. Tightness of weakly convergent *sequences* in metrizable spaces is due to Le Cam [26].

Theorem 2.2. *If a subset $\mathcal{K} \subset \mathcal{M}(X)$ is bounded and equally tight then it is relatively compact with respect to the narrow topology. The converse is also true in the following cases:*
(i) (X, τ) is a locally compact or a Polish space;
(ii) (X, τ) is metrizable and $\mathcal{K} = \{\mu_n : n \in \mathbb{N}\}$ for a given weakly convergent sequence (μ_n).

If $\mu \in \mathcal{M}(X)$ and Y is another Hausdorff topological space, a map $T : X \to Y$ is *Lusin μ-measurable* [38, Ch. I, Sec. 5] if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that $\mu(X \setminus K_\varepsilon) \le \varepsilon$ and the restriction of T to K_ε is continuous. We denote by $T_\sharp\mu \in \mathcal{M}(Y)$ the push-forward Radon measure defined by

$$T_\sharp\mu(B) := \mu(T^{-1}(B)) \text{ for every } B \in \mathcal{B}(Y), \quad T : X \to Y \text{ is Lusin } \mu\text{-measurable.} \tag{2.5}$$

$\mathrm{B}(X)$ (resp. $\mathrm{B}_b(X)$) will be the linear space of real Borel (resp. bounded Borel) functions. If $\mu \in \mathcal{M}(X)$, $p \in [1, \infty]$, we will denote by $\mathrm{L}^p(X, \mu)$ the subspace of Borel p-integrable functions w.r.t. μ, without identifying μ-almost equal functions.

Lebesgue decomposition. Given $\gamma, \mu \in \mathcal{M}(X)$, we write $\gamma \ll \mu$ if $\mu(A) = 0$ yields $\gamma(A) = 0$ for every $A \in \mathcal{B}(X)$. We say that $\gamma \perp \mu$ if there exists $B \in \mathcal{B}(X)$ such that $\mu(B) = 0 = \gamma(X \setminus B)$.

Lemma 2.3 (Lebesgue decomposition)**.** *For every $\gamma, \mu \in \mathcal{M}(X)$ (with $(\gamma + \mu)(X) > 0$), there exist Borel functions $\sigma, \varrho : X \to [0, \infty)$ and a Borel partition (A, A_γ, A_μ) of X with the following properties:*

$$A = \{x \in X : \sigma(x) > 0\} = \{x \in X : \varrho(x) > 0\}, \quad \sigma \cdot \varrho \equiv 1 \quad \textit{in } A, \tag{2.6}$$
$$\gamma = \sigma\mu + \gamma^\perp, \quad \sigma \in \mathrm{L}^1_+(X, \mu), \quad \gamma^\perp \perp \mu, \quad \gamma^\perp(X \setminus A_\gamma) = \mu(A_\gamma) = 0, \tag{2.7}$$
$$\mu = \varrho\gamma + \mu^\perp, \quad \varrho \in \mathrm{L}^1_+(X, \gamma), \quad \mu^\perp \perp \gamma, \quad \mu^\perp(X \setminus A_\mu) = \gamma(A_\mu) = 0. \tag{2.8}$$

Moreover, the sets A, A_γ, A_μ and the densities σ, ϱ are uniquely determined up to $(\mu+\gamma)$-negligible sets.

Proof. Let $\theta \in \mathrm{B}(X; [0, 1])$ be the Lebesgue density of γ w.r.t. $\nu := \mu + \gamma$ (thus uniquely determined up to ν-negligible sets). The Borel partition can be defined by setting $A := \{x \in X : 0 < \theta(x) < 1\}$, $A_\gamma := \{x \in X : \theta(x) = 1\}$ and $A_\mu := \{x \in X : \theta(x) = 0\}$; by defining $\sigma := \theta/(1 - \theta)$, $\varrho := 1/\sigma = (1 - \theta)/\theta$ for every $x \in A$ and $\sigma = \varrho \equiv 0$ in $X \setminus A$, we obtain Borel functions satisfying (2.7) and (2.8).

Conversely, it is not difficult to check that starting from a decomposition as in (2.6), (2.7) and (2.8) and defining $\theta \equiv 0$ in A_μ, $\theta \equiv 1$ in A_γ and $\theta := \sigma/(1 + \sigma)$ in A we obtain a Borel function with values in $[0, 1]$ such that $\gamma = \theta(\mu + \gamma)$. □

2.2 Min-max and duality

We recall now a powerful form of von Neumann's Theorem, concerning minimax properties of convex-concave functions in convex subsets of vector spaces. The statement has the advantage of involving a minimal set of topological assumptions (we refer to [39, Thm. 3.1] for the proof, see also [8, Chapter 1, Prop. 1.1]).

Let A, B be nonempty convex sets of some vector spaces and let us suppose that A is endowed with an Hausdorff topology. Let $L : A \times B \to \mathbb{R}$ be a function such that

$$a \mapsto L(a, b) \quad \text{is convex and lower semicontinuous in } A \text{ for every } b \in B, \tag{2.9}$$

$$b \mapsto L(a, b) \quad \text{is concave in } B \text{ for every } a \in A. \tag{2.10}$$

Notice that for arbitrary functions L one always has

$$\inf_{a\in A} \sup_{b\in B} L(a, b) \geq \sup_{b\in B} \inf_{a\in A} L(a, b); \tag{2.11}$$

so that equality holds in (2.11) if $\sup_{b\in B} \inf_{a\in A} L(a, b) = +\infty$. When $\sup_{b\in B} \inf_{a\in A} L(a, b)$ is finite, we can still have equality thanks to the following result.

Theorem 2.4 (Minimax duality)**.** *If there exists* $b_\star \in B$ *and* $C > \sup_{b\in B} \inf_{a\in A} L(a, b)$ *such that*

$$\big\{a \in A : L(a, b_\star) \leq C\big\} \quad \textit{is compact in } A, \tag{2.12}$$

then

$$\inf_{a\in A} \sup_{b\in B} L(a, b) = \sup_{b\in B} \inf_{a\in A} L(a, b). \tag{2.13}$$

2.3 Entropy functions and their conjugates

Entropy functions in $[0, \infty)$**.** We say that $F : [0, \infty) \to [0, \infty]$ belongs to the class $\Gamma(\mathbb{R}_+)$ of *admissible entropy function* if it satisfies

$$F \text{ is convex and lower semicontinuous with } \quad \mathrm{Dom}(F) \cap (0, \infty) \neq \emptyset, \tag{2.14}$$

where

$$\mathrm{Dom}(F) := \{s \geq 0 : F(s) < \infty\}, \quad \text{with } s^- := \inf \mathrm{Dom}(F), \quad s^+ := \sup \mathrm{Dom}(F) > 0. \tag{2.15}$$

The recession constant F'_∞, the right derivative F'_0 at 0, and the asymptotic affine coefficient $\mathrm{aff} F_\infty$ are defined by (here $s_o \in \mathrm{Dom}(F)$)

$$F'_\infty := \lim_{s\to\infty} \frac{F(s)}{s} = \sup_{s>0} \frac{F(s) - F(s_o)}{s - s_o}, \qquad F'_0 := \begin{cases} -\infty & \text{if } F(0) = +\infty, \\ \lim_{s\downarrow 0} \frac{F(s)-F(0)}{s} & \text{otherwise,} \end{cases} \tag{2.16}$$

$$\mathrm{aff} F_\infty := \begin{cases} +\infty & \text{if } F'_\infty = +\infty, \\ \lim_{s\to\infty} \big(F'_\infty\, s - F(s)\big) & \text{otherwise.} \end{cases} \tag{2.17}$$

To avoid trivial cases, we assumed that the proper domain $\mathrm{Dom}(F)$ contains at least a strictly positive real number. $\mathrm{Dom}(F)$ is an interval of $[0, \infty)$: our main focus will concern the case when $\mathrm{Dom}(F)$ has nonempty interior and F has superlinear growth, i.e. $F'_\infty = +\infty$, but it will be useful to deal with the general class defined by (2.14).

Legendre duality. As usual, the *Legendre conjugate* function $F^*:\mathbb{R}\to(-\infty,+\infty]$ is defined by

$$F^*(\phi):=\sup_{s\geq 0}\big(s\phi-F(s)\big), \tag{2.18}$$

with proper domain $\mathrm{Dom}(F^*):=\{\phi\in\mathbb{R}:F^*(\phi)\in\mathbb{R}\}$; F^* is in fact the conjugate of the convex function $\tilde F:\mathbb{R}\to(-\infty,+\infty]$, obtained by extending F to $+\infty$ for negative arguments. Notice that

$$\inf \mathrm{Dom}(F^*)=-\infty,\quad \sup\mathrm{Dom}(F^*)=F'_\infty, \tag{2.19}$$

so that F^* is finite and continuous in $(-\infty,F'_\infty)$, nondecreasing, and satisfies

$$\lim_{\phi\downarrow-\infty}F^*(\phi)=\inf F^*=-F(0),\quad \sup F^*=\lim_{\phi\uparrow+\infty}F^*(\phi)=+\infty. \tag{2.20}$$

Concerning the behavior of F^* at the boundary of its proper domain we can distinguish a few cases:

- If $F'_0=-\infty$ (in particular when $F(0)=+\infty$) then F^* is strictly increasing in $\mathrm{Dom}(F^*)$.
- If F'_0 is finite, then F^* is strictly increasing in $[F'_0,F'_\infty)$ and takes the constant value $F(0)$ in $(-\infty,F'_0]$. Thus $F(0)$ belongs to the range of F^* only if $F'_0>-\infty$.
- If F'_∞ is finite, then $\lim_{\phi\uparrow F'_\infty}F^*(\phi)=\mathrm{aff}F_\infty$. Thus $F'_\infty\in\mathrm{Dom}(F^*)$ only if $\mathrm{aff}F_\infty<\infty$.
- The degenerate case when $F'_\infty=F'_0$ occurs only when F is linear.

If F is not linear, we always have

$$F^* \text{ is an increasing homeomorphism between } (F'_0,F'_\infty) \text{ and } (-F(0),\mathrm{aff}F_\infty) \tag{2.21}$$

with the obvious extensions to the boundaries of the intervals when F'_0 or $\mathrm{aff}F_\infty$ are finite.

By introducing the closed convex set in $\mathbb{R}^2$

$$\mathfrak{F}:=\big\{(\phi,\psi)\in\mathbb{R}^2: s\phi+\psi\leq F(s)\quad\forall\, s>0\big\}=\big\{(\phi,\psi)\in\mathbb{R}^2:\psi\leq -F^*(\phi)\big\}, \tag{2.22}$$

F can be recovered by F^* and by $\mathfrak{F}$ through the dual Fenchel-Moreau formula

$$F(s)=\sup_{\phi\in\mathbb{R}}\big(s\phi-F^*(\phi)\big)=\sup_{(\phi,\psi)\in\mathfrak{F}}s\phi+\psi. \tag{2.23}$$

Notice that $\mathfrak{F}$ satisfies the obvious monotonicity property

$$(\phi,\psi)\in\mathfrak{F},\quad \tilde\psi\leq\psi,\ \tilde\phi\leq\phi\quad\Rightarrow\quad(\tilde\phi,\tilde\psi)\in\mathfrak{F}. \tag{2.24}$$

If F is finite in a neighborhood of $+\infty$, then F^* is superlinear as $\phi\uparrow\infty$. More precisely, its asymptotic behavior as $\phi\to\pm\infty$ is related to the proper domain of F by

$$s_\pm=\lim_{\phi\to\pm\infty}\frac{F^*(\phi)}{\phi}. \tag{2.25}$$

F and F^* are also related to the subdifferential $\partial F:\mathbb{R}\to 2^{\mathbb{R}}$ of F by

$$\phi\in\partial F(s)\quad\Leftrightarrow\quad s\in\mathrm{Dom}(F),\quad \phi\in\mathrm{Dom}(F^*),\quad F(s)+F^*(\phi)=s\phi. \tag{2.26}$$

Example 2.5 (Power-like entropies)**.** An important class of entropy functions is provided by the power like functions $U_p : [0,\infty) \to [0,\infty]$, $p \in \mathbb{R}$, characterized by the conditions

$$U_p \in \mathrm{C}^\infty(0,\infty), \quad U_p(1) = U_p'(1) = 0, \quad U_p''(s) = s^{p-2}, \quad U_p(0) = \lim_{s\downarrow 0} U_p(s); \tag{2.27}$$

equivalently

$$U_p(s) = \begin{cases} \frac{1}{p(p-1)}\big(s^p - p(s-1) - 1\big) & \text{if } p \neq 0,1,\\ s\log s - (s-1) & \text{if } p = 1, \\ s - 1 - \log s & \text{if } p = 0, \end{cases} \qquad \text{if } s > 0, \tag{2.28}$$

with $U_p(0) = \frac{1}{p}$ if $p > 0$, $+\infty$ if $p \le 0$.

The corresponding Legendre conjugates are given, in terms of the dual exponent $q = p/(p-1)$, by

$$U_q^*(\phi) := \begin{cases} \dfrac{q-1}{q}\Big[\big(1 + \dfrac{\phi}{q-1}\big)_+^q - 1\Big], & \mathrm{Dom}(U_q^*) = \mathbb{R}, & \text{if } p > 1,\ q > 1 \\ \mathrm{e}^\phi - 1, & \mathrm{Dom}(U_q^*) = \mathbb{R}, & \text{if } p = 1,\ q = \infty \\ \dfrac{q-1}{q}\Big[\big(1 + \dfrac{\phi}{q-1}\big)^q - 1\Big], & \mathrm{Dom}(U_q^*) = (-\infty, 1-q), & \text{if } 0 < p < 1,\ q < 0, \\ -\log(1-\phi), & \mathrm{Dom}(U_q^*) = (-\infty, 1), & \text{if } p = 0,\ q = 0 \\ \dfrac{q-1}{q}\Big[\big(1 + \dfrac{\phi}{q-1}\big)^q - 1\Big], & \mathrm{Dom}(U_q^*) = (-\infty, 1-q], & \text{if } p < 0,\ 0 < q < 1. \end{cases}$$

Reverse entropies. Let us now introduce the reverse density function $R : [0,\infty) \to [0,\infty]$ as

$$R(r) := \begin{cases} rF(1/r) & \text{if } r > 0, \\ F'_\infty & \text{if } r = 0. \end{cases} \tag{2.29}$$

It is not difficult to check that R is a proper, convex and lower semicontinuous function, with

$$R(0) = F'_\infty, \quad R'_\infty = F(0), \quad \mathrm{aff} F_\infty = -R'_0, \quad \mathrm{aff} R_\infty = -F'_0, \tag{2.30}$$

so that $R \in \Gamma(\mathbb{R}_+)$ and the map $F \mapsto R$ is an involution of $\Gamma(\mathbb{R}_+)$. A further remarkable involution property is enjoyed by the dual convex set $\mathfrak{R} := \{(\psi,\phi) \in \mathbb{R}^2 : R^*(\psi) + \phi \le 0\}$ defined as (2.22): it is easy to check that

$$(\phi,\psi) \in \mathfrak{F} \quad \Leftrightarrow \quad (\psi,\phi) \in \mathfrak{R}. \tag{2.31}$$

It follows that the Legendre transform of R and F are related by

$$\psi \le -F^*(\phi) \quad \Leftrightarrow \quad \phi \le -R^*(\psi) \quad \Leftrightarrow \quad (\phi,\psi) \in \mathfrak{F} \qquad \text{for every } \phi,\psi \in \mathbb{R}. \tag{2.32}$$

As in (2.21) we have

$$R^* \text{ is an increasing homeomorphism between } (-\mathrm{aff} F_\infty, F(0)) \text{ and } (-F'_\infty, -F'_0). \tag{2.33}$$

A last useful identity involves the subdifferentials of F and R: for every $s, r > 0$, $sr = 1$, $\phi, \psi \in \mathbb{R}$

$$\phi \in \partial F(r),\ \psi = -F^*(\phi) \quad \Leftrightarrow \quad \psi \in \partial R(s),\ \phi = -R^*(\psi) \tag{2.34}$$

It is not difficult to check that the reverse entropy associated to U_p is U_{1-p}.

2.4 Relative entropy integral functionals

If $F \in \Gamma(\mathbb{R}_+)$ we consider the functional $\mathscr{F} : \mathcal{M}(X) \times \mathcal{M}(X) \to [0,\infty]$ defined by

$$\mathscr{F}(\gamma|\mu) := \int_X F(\sigma)\,\mathrm{d}\mu + F'_\infty\,\gamma^\perp(X), \quad \gamma = \sigma\mu + \gamma^\perp, \quad \gamma^\perp \perp \mu, \quad \sigma := \frac{\mathrm{d}\gamma}{\mathrm{d}\mu}, \tag{2.35}$$

where $\gamma = \sigma\mu + \gamma^\perp$ is the Lebesgue decomposition of γ w.r.t. μ (2.7). Notice that

$$\text{if } F \text{ is superlinear then} \quad \mathscr{F}(\gamma|\mu) = +\infty \quad \text{if } \gamma \not\ll \mu, \tag{2.36}$$

and, whenever η_0 is the null measure,

$$\mathscr{F}(\gamma|\eta_0) = F'_\infty\,\gamma(X), \tag{2.37}$$

where, as usual in measure theory, we adopted the convention $0 \cdot \infty = 0$.

In the next Lemma we will consider Borel functions $\varphi \in \mathrm{B}(X;\bar{\mathbb{R}})$ taking values in the extended real line $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ and we denote by $\bar{\mathfrak{F}}$ the closure of $\mathfrak{F}$ in $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$. Notice that

$$(\phi,\psi) \in \bar{\mathfrak{F}} \quad \Leftrightarrow \quad \begin{cases} \psi \le -F^*(\phi) & \text{if } -\infty < \phi \le F'_\infty,\ \phi < +\infty \\ \psi = -\infty & \text{if } \phi = F'_\infty = +\infty, \\ \psi \in [-\infty, F(0)] & \text{if } \phi = -\infty, \end{cases} \tag{2.38}$$

and, symmetrically by (2.31) and (2.30),

$$(\phi,\psi) \in \bar{\mathfrak{F}} \quad \Leftrightarrow \quad \begin{cases} \phi \le -R^*(\psi) & \text{if } -\infty < \psi \le F(0),\ \psi < +\infty \\ \phi = -\infty & \text{if } \psi = F(0) = +\infty, \\ \phi \in [-\infty, F'_\infty] & \text{if } \psi = -\infty; \end{cases} \tag{2.39}$$

in particular

$$(\phi,\psi) \in \bar{\mathfrak{F}} \quad \Rightarrow \quad \phi \le F'_\infty, \quad \psi \le F(0). \tag{2.40}$$

Lemma 2.6. *If $\mathscr{F}(\gamma|\mu) < \infty$, $(\phi,\psi) \in \mathrm{B}(X;\bar{\mathfrak{F}})$, $\psi_- \in \mathrm{L}^1(X,\mu)$ (resp. $\phi_- \in \mathrm{L}^1(X,\gamma)$) then $\phi_+ \in \mathrm{L}^1(X,\gamma)$ (resp. $\psi_+ \in \mathrm{L}^1(X,\mu)$) and*

$$\mathscr{F}(\gamma|\mu) - \int_X \psi\,\mathrm{d}\mu \ge \int_X \phi\,\mathrm{d}\gamma. \tag{2.41}$$

Whenever $\psi \in \mathrm{L}^1(X,\mu)$ or $\phi \in \mathrm{L}^1(X,\gamma)$, equality holds in (2.41) if and only if for the Lebesgue decomposition given by Lemma 2.3 one has

$$\phi \in \partial F(\sigma),\ \psi = -F^*(\phi)\ \textit{(equivalently, } \psi \in \partial R(\varrho),\ \phi = -R^*(\psi)\textit{)} \quad (\mu+\gamma)\textit{-a.e. in } A, \tag{2.42}$$

$$\psi = F(0) < \infty\ \ \mu^\perp\textit{-a.e. in } A_\mu, \quad \phi = F'_\infty < \infty\ \ \gamma^\perp\textit{-a.e. in } A_\gamma. \tag{2.43}$$

Proof. Let us first show that in both cases the two integrals of (2.41) are well defined (possibly taking the value $-\infty$). If $\psi_- \in \mathrm{L}^1(X,\mu)$ (in particular $\psi > -\infty$ μ-a.e.) with $(\phi,\psi) \in \bar{\mathfrak{F}}$ we use the pointwise bound $s\phi \le F(s) - \psi$ that yields $s\phi_+ \le (F(s)-\psi)_+ \le F(s) + \psi_-$ obtaining $\phi_+ \in \mathrm{L}^1(X,\gamma)$, since $(\phi,\psi) \in \bar{\mathfrak{F}}$ yields $\phi_+ \le F'_\infty$.

If $\phi_- \in \mathrm{L}^1(X,\gamma)$ (and thus $\phi > -\infty$ γ-a.e.) the analogous inequality $\psi_+ \le F(s) + s\phi_-$ yields $\psi_+ \in \mathrm{L}^1(X,\mu)$. (2.41) then follows from (2.22) and from (2.40).

Once $\phi \in \mathrm{L}^1(X,\mu)$ (or $\psi \in \mathrm{L}^1(X,\gamma)$), (2.41) can be written as

$$\int_A \Big(F(\sigma) - \sigma\phi - \psi\Big)\,\mathrm{d}\mu + \int_{A_\mu} \Big(F(0) - \psi\Big)\,\mathrm{d}\mu^\perp + \int_{A_\gamma} (F'_\infty - \phi)\,\mathrm{d}\gamma^\perp \ge 0,$$

and by (2.22) and (2.40) the equality case immediately yields that each of the three integrals of the previous formula vanishes. Since ϕ, ψ take real values $(\mu+\gamma)$-a.e. in A and belong to $\bar{\mathfrak{F}}$, the vanishing of the first integrand yields $\psi = -F^*(\sigma)$ and $\phi \in \partial F(\sigma)$ by (2.26) for μ (and thus $(\mu+\gamma)$) almost every point of A. (2.34) provides the reversed identities $\psi \in \partial R(\varrho)$, $\phi = -R^*(\psi)$.

(2.43) follows easily by the vanishing of the last two integrals and the fact that ψ is finite μ-a.e. and ϕ is finite γ-a.e. □

The next Theorem is the main result of this section. Its proof is a careful adaptation of [2, Lemma 9.4.4] to the present more general setting, which includes the sublinear case when $F'_\infty < \infty$ and the lack of complete regularity of the space. This suggests to deal with lower semicontinuous functions instead of continuous ones. We denote by $\mathrm{LSC}_s(X)$ the class of lower semicontinuous and simple functions (i.e. taking a finite number of values) and we introduce the notation $\varphi = -\phi$ and the concave function

$$F^\circ(\varphi) := -F^*(-\varphi). \tag{2.44}$$

Theorem 2.7 (Duality and lower semicontinuity)**.** *For every $\gamma, \mu \in \mathcal{M}(X)$ we have*

$$\mathscr{F}(\gamma|\mu) = \sup\Big\{\int_X \psi\,\mathrm{d}\mu + \int_X \phi\,\mathrm{d}\gamma : \phi, \psi \in \mathrm{LSC}_s(X),\ (\phi(x), \psi(x)) \in \mathfrak{F}\ \forall x \in X\Big\} \tag{2.45}$$

$$= \sup\Big\{\int_X \psi\,\mathrm{d}\mu - \int_X R^*(\psi)\,\mathrm{d}\gamma : \psi, R^*(\psi) \in \mathrm{LSC}_s(X)\Big\} \tag{2.46}$$

$$= \sup\Big\{\int_X F^\circ(\varphi)\,\mathrm{d}\mu - \int_X \varphi\,\mathrm{d}\gamma : \varphi, F^\circ(\varphi) \in \mathrm{LSC}_s(X)\Big\} \tag{2.47}$$

and the space $\mathrm{LSC}_s(X)$ *in the supremum of* (2.45)*,* (2.46) *and* (2.47) *can also be replaced by the space* $\mathrm{LSC}_b(X)$ *(resp* $\mathrm{B}_b(X)$*) of bounded l.s.c. (resp. Borel) functions.*

Remark 2.8. If (X,τ) is completely regular (recall (2.3)), then we can equivalently replace lower semicontinuous functions by continuous ones in (2.45), (2.46) and (2.47)); e.g. in the case of (2.45) we have

$$\mathscr{F}(\gamma|\mu) = \sup\Big\{\int_X \psi\,\mathrm{d}\mu + \int_X \phi\,\mathrm{d}\gamma : (\phi, \psi) \in \mathrm{C}_b(X; \mathfrak{F})\Big\}. \tag{2.48}$$

In fact, considering first (2.45), by complete regularity it is possible to express every couple ϕ, ψ of bounded lower semicontinuous functions with values in $\mathfrak{F}$ as the supremum of a directed family of continuous and bounded functions $(\phi_\alpha, \psi_\alpha)_{\alpha\in\mathbb{A}}$ which still satisfy the constraint $\mathfrak{F}$ by (2.24). We can then apply the continuity (2.2) of the integrals with respect to the Radon measures μ and γ.

In order to replace l.s.c. functions with continuous ones in (2.46) we can approximate ψ by an increasing directed family of continuous functions $(\psi_\alpha)_{\alpha\in\mathbb{A}}$. By truncation, one can always assume that $\max\psi \geq \sup\psi_\alpha \geq \inf\psi_\alpha \geq \min\psi$. Since $R^*(\psi)$ is bounded, it is easy to check that also $R^*(\psi_\alpha)$ is bounded and it is an increasing directed family converging to $R^*(\psi)$. An analogous argument works for (2.48). $\square$

Proof. Let us prove (2.45): denoting by $\mathscr{F}'$ its right-hand side, Lemma 2.6 yields $\mathscr{F} \geq \mathscr{F}'$. In order to prove the opposite inequality let $B \in \mathcal{B}(X)$ a μ-negligible Borel set where $\gamma^\perp$ is concentrated, let $A := X \setminus B$ and let $\sigma : X \to [0,\infty)$ be a Borel density for γ w.r.t. μ. We consider a countable subset $(\phi_n, \psi_n)_{n=1}^\infty$ dense in $\mathfrak{F}$ with $\psi_1 = \phi_1 = 0$, and an increasing sequence $\bar\phi_n \in (-\infty, F'_\infty)$ converging to F'_∞, with $\bar\psi_n := -F^*(\bar\phi_n)$. By (2.23)

$$F(\sigma(x)) = \lim_{N\uparrow\infty} F_N(x), \quad F_N(x) := \sup_{1\leq n\leq N} \psi_n + \sigma(x)\phi_n \quad \text{for every } x \in X,$$

so that, by Beppo Levi's monotone convergence theorem (notice that $F_N \geq F_1 = 0$),

$$\mathscr{F}(\gamma|\mu) = \lim_{N\uparrow\infty} \int_A F_N(x)\,\mathrm{d}\mu(x) + \bar\phi_N\gamma(B).$$

It is therefore sufficient to prove that

$$\mathscr{F}'(\gamma|\mu) \geq \int_A F_N(x)\,\mathrm{d}\mu(x) + \bar\phi_N\gamma(B) \quad \text{for every } N \in \mathbb{N}. \tag{2.49}$$

We fix $N \in \mathbb{N}$, we set $\phi_0 := \bar\phi_N$, $\psi_0 := \bar\psi_N$, we recursively define the Borel sets A_j, $j = 0, \ldots, N$,

$$A_0 := B, \quad A_j := \Big\{x \in A \setminus \bigcup_{i=0}^{j-1} A_i : \psi_j + \sigma(x)\phi_j = F_N(x)\Big\},\ 1 \leq j \leq N, \tag{2.50}$$

which form a Borel partition of A. Since μ and γ are Radon measures, for every $\varepsilon > 0$ we find disjoint compact sets $K_j \subset A_j$ and disjoint open sets (by the Hausdorff separation property of X) $G_j \supset K_j$ such that

$$\sum_{j=0}^N \Big(\mu(A_j \setminus K_j) + \gamma(A_j \setminus K_j)\Big) = \mu\Big(X \setminus \bigcup_{j=0}^N K_j\Big) + \gamma\Big(X \setminus \bigcup_{j=0}^N K_j\Big) \leq \varepsilon/S_N$$

where

$$S_N := \max_{0\leq n\leq N} (\phi_n - \phi_{\min}) + (\psi_n - \psi_{\min}), \quad \phi_{\min} := \min_{0\leq j\leq N} \phi_j, \quad \psi_{\min} := \min_{0\leq j\leq N} \psi_j.$$

Since $(\phi_{\min}, \psi_{\min}) \in \mathfrak{F}$ and the sets G_n are disjoint, the lower semicontinuous functions

$$\psi_N(x) := \psi_{\min} + \sum_{n=0}^N (\psi_n - \psi_{\min})\chi_{G_n}(x), \quad \phi_N(x) := \phi_{\min} + \sum_{n=0}^N (\phi_n - \phi_{\min})\chi_{G_n}(x) \tag{2.51}$$

take values in $\mathfrak{F}$ and satisfy

$$
\begin{aligned}
\int_A F_N(x)\,\mathrm{d}\mu(x) + \bar{\phi}_N\gamma(B) &= \sum_{j=1}^{N}\int_{A_j} F_N(x)\,\mathrm{d}\mu(x) + \phi_0\gamma(A_0)\\
&= \phi_{\min}\gamma(X) + \psi_{\min}\mu(X) + \sum_{j=0}^{N}\Big(\int_{A_j}(\phi_j-\phi_{\min})\,\mathrm{d}\gamma(x) + \int_{A_j}(\psi_j-\psi_{\min})\,\mathrm{d}\mu(x)\Big)\\
&\le \phi_{\min}\gamma(X) + \psi_{\min}\mu(X) + \sum_{j=0}^{N}\Big(\int_{K_j}(\phi_j-\phi_{\min})\,\mathrm{d}\gamma(x) + \int_{K_j}(\psi_j-\psi_{\min})\,\mathrm{d}\mu(x)\Big) + \varepsilon\\
&\le \int_X \phi_N(x)\,\mathrm{d}\gamma(x) + \int_X \psi_N(x)\,\mathrm{d}\mu(x) + \varepsilon.
\end{aligned}
$$

Since ε is arbitrary we obtain (2.49).

(2.46) follows directly by (2.45) and the previous Lemma 2.6: in fact, denoting by $\mathscr{F}''$ the righthand side of (2.46), Lemma 2.6 shows that $\mathscr{F}''(\gamma|\mu) \le \mathscr{F}(\gamma|\mu) = \mathscr{F}'(\gamma|\mu)$. On the other hand, if $\phi, \psi \in \mathrm{LSC}_s(X)$ with $(\phi,\psi) \in \mathfrak{F}$ then $-R^*(\psi) \ge \phi$ and $R^*(\psi) \in \mathrm{LSC}_s(X)$ since R^* is nondecreasing, does not take the value $-\infty$ and is bounded from above by $-\phi$. We thus get $\mathscr{F}''(\gamma|\mu) \ge \mathscr{F}'(\gamma|\mu)$.

In order to show (2.47) we observe that for every $\psi \in \mathrm{LSC}_s(X)$ with $R^*(\psi) \in \mathrm{LSC}_s(X)$ we can set $\varphi := R^*(\psi) \in \mathrm{LSC}_s(X)$; since $(\psi, -R^*(\psi)) \in \mathfrak{F}$ (2.32) yields $\psi \le -F^*(-\varphi) = F^\circ(\varphi)$ so that $\int F^\circ(\varphi)\,\mathrm{d}\mu - \int \varphi\,\mathrm{d}\gamma \ge \int \psi\,\mathrm{d}\mu - \int R^*(\psi)\,\mathrm{d}\gamma$. Since F° cannot take the value $+\infty$, we also have that $(-\varphi, F^\circ(\varphi)) \in \mathfrak{F}$ so that $\int F^\circ(\varphi)\,\mathrm{d}\mu - \int \varphi\,\mathrm{d}\gamma \le \mathscr{F}(\gamma|\mu)$ by Lemma 2.6.

When one replaces $\mathrm{LSC}_s(X)$ with $\mathrm{LSC}_b(X)$ or $\mathrm{B}_b(X)$ in (2.45), the supremum is taken on a larger set, so that the righthand side of (2.45) cannot decrease; on the other hand, Lemma 2.6 shows that $\mathscr{F}(\gamma|\mu)$ still provides an upper bound even if ϕ, ψ are in $\mathrm{B}_b(X)$, thus duality also holds in this case. The same argument applies to (2.46) or (2.47). □

Corollary 2.9. *The functional $\mathscr{F}$ is jointly convex and lower semicontinuous in $\mathcal{M}(X) \times \mathcal{M}(X)$; more generally, if $F_n \in \Gamma(\mathbb{R}_+)$, $n \in \mathbb{N}$, is an increasing sequence pointwise converging to F and $(\mu,\gamma) \in \mathcal{M}(X) \times \mathcal{M}(X)$ is the narrow limit of a sequence $(\mu_n,\gamma_n) \in \mathcal{M}(X) \times \mathcal{M}(X)$, then the corresponding entropy functionals $\mathscr{F}_n, \mathscr{F}$ satisfy*

$$\liminf_{n\to\infty} \mathscr{F}_n(\gamma_n|\mu_n) \ge \mathscr{F}(\gamma|\mu). \tag{2.52}$$

Proof. The lower semicontinuity of $\mathscr{F}$ follows by (2.45), which provides a representation of $\mathscr{F}$ as the supremum of a family of lower semicontinuous functionals for the narrow topology.

In order to prove (2.52) it is sufficient to check that for every $\gamma, \mu \in \mathcal{M}(X)$

$$\lim_{n\to\infty} \mathscr{F}_n(\gamma|\mu) = \mathscr{F}(\mu|\nu); \tag{2.53}$$

this formula follows easily by the monotonicity of the convex sets $\mathfrak{F}_n$ (associated to F_n by (2.22)) $\mathfrak{F}_n \subset \mathfrak{F}_{n+1}$ and by the fact that $\mathfrak{F} = \cup_n \mathfrak{F}_n$, since F_n^* is pointwise decreasing to F^*. Thus for every couple of simple and lower semicontinuous functions (ϕ,ψ) taking values in $\mathfrak{F}$ we have $(\psi(x),\phi(x)) \in \mathfrak{F}_N$ for every $x \in X$ and a sufficiently large N so that

$$\liminf_{n\to\infty} \mathscr{F}_n(\gamma|\mu) \ge \int_X \psi\,\mathrm{d}\mu + \int_X \phi\,\mathrm{d}\gamma.$$

Since ϕ, ψ are arbitrary we conclude applying the duality formula (2.45). $\square$

Proposition 2.10 (Boundedness and tightness)**.** *If $\mathcal{K} \subset \mathcal{M}(X)$ is bounded and $F'_\infty > 0$, for every $C \geq 0$ the sublevels of $\mathscr{F}$*

$$\Xi_C := \Big\{\gamma \in \mathcal{M}(X) : \mathscr{F}(\gamma|\mu) \leq C \textit{ for some } \mu \in \mathcal{K}\Big\}, \tag{2.54}$$

are bounded; if moreover $\mathcal{K}$ is equally tight and $F'_\infty = \infty$, then the sets Ξ_C are equally tight.

Proof. Concerning the properties of Ξ_C, we will use the inequality

$$\lambda\gamma(B) \leq \mathscr{F}(\gamma|\mu) + F^*(\lambda)\mu(B) \quad \text{for every } \lambda \in (0, F'_\infty),\ B \in \mathcal{B}(X), \tag{2.55}$$

which follows easily by integrating the Young inequality

$$\lambda\sigma \leq F(\sigma) + F^*(\lambda), \quad \lambda > 0,\ \gamma = \sigma\mu + \gamma^\perp,$$

in B with respect to μ and by observing that

$$\lambda\gamma(B) = \lambda\int_B \sigma\,\mathrm{d}\mu + \lambda\gamma^\perp(B) \leq \lambda\int_B \sigma\,\mathrm{d}\mu + F'_\infty\gamma^\perp(B) \quad \text{if } 0 < \lambda < F'_\infty.$$

Choosing first $B = X$ in (2.55) and an arbitrary λ in $(0, F'_\infty)$ (notice that $F^*(\lambda) < \infty$ thanks to (2.19)) we immediately get a uniform bound of $\gamma(X)$ for every $\gamma \in \Xi_C$.

In order to prove the tightness when $F'_\infty = \infty$, whenever $\varepsilon > 0$ is given, we can choose $\lambda = 2C/\varepsilon$ and $\eta > 0$ so small that $\eta F^*(\lambda)/\lambda \leq \varepsilon/2$, and then a compact set $K \subset X$ such that $\mu(X \setminus K) \leq \eta$ for every $\mu \in \mathcal{K}$. (2.55) shows that $\gamma(X \setminus K) \leq \varepsilon$ for every $\gamma \in \Xi$. $\square$

We conclude this section with a useful representation of $\mathscr{F}$ in terms of the reverse entropy R (2.29) and the corresponding functional $\mathscr{R}$.

Lemma 2.11. *For every $\gamma, \mu \in \mathcal{M}(X)$ we have*

$$\mathscr{R}(\mu|\gamma) = \int_X R(\varrho(x))\,\mathrm{d}\gamma(x) + R_\infty\,\mu^\perp(X), \tag{2.56}$$

where $\mu = \varrho\gamma + \mu^\perp$ is the reverse Lebesgue decomposition given by (2.8)*. In particular*

$$\mathscr{F}(\gamma|\mu) = \mathscr{R}(\mu|\gamma). \tag{2.57}$$

Proof. It is an immediate consequence of (2.45) and (2.31). $\square$

3 Optimal Entropy-Transport problems

3.1 The basic setting

Let us fix the basic set of data for Entropy-Transport problems. We are given

- two Hausdorff topological spaces (X_i, τ_i), $i = 1, 2$, with Cartesian product $\boldsymbol{X} := X_1 \times X_2$ and canonical projections $\pi^i : \boldsymbol{X} \to X_i$;

- two entropy functions $F_i \in \Gamma(\mathbb{R}_+)$, thus satisfying (2.14);

- a proper lower semicontinuous cost function $\mathsf{c} : \boldsymbol{X} \to [0, +\infty]$;

- a couple of nonnegative Radon measures $\mu_i \in \mathcal{M}(X_i)$ with finite mass $m_i := \mu_i(X_i)$ satisfying the compatibility condition

$$J := \Big(m_1 \operatorname{Dom}(F_1)\Big) \cap \Big(m_2 \operatorname{Dom}(F_2)\Big) \neq \emptyset. \tag{3.1}$$

We will often assume that the above basic setting is also *coercive*: this means that *at least one* of the following coercivity conditions holds:

$$F_i \text{ are superlinear, i.e. } (F_i)'_\infty = +\infty; \tag{3.2a}$$
$$(F_1)'_\infty + (F_2)'_\infty + \inf \mathsf{c} > 0 \text{ and } \mathsf{c} \text{ has compact sublevels.} \tag{3.2b}$$

For every $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ we set $\gamma_i := \pi^i_\sharp \boldsymbol{\gamma}$ and as for (2.35)

$$\mathscr{F}_i(\boldsymbol{\gamma}|\mu_i) := \int_{X_i} F_i\Big(\frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}\Big)\,\mathrm{d}\mu_i + (F_i)'_\infty \gamma_i^\perp(X),\ \ \gamma_i = \pi^i_\sharp \boldsymbol{\gamma} = \sigma_i \mu_i + \gamma_i^\perp, \quad \sigma_i := \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}, \tag{3.3}$$

and we introduce the Entropy-Transport functional as

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) := \sum_i \mathscr{F}_i(\boldsymbol{\gamma}|\mu_i) + \int_{\boldsymbol{X}} \mathsf{c}(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma}(x_1, x_2), \tag{3.4}$$

possibly taking the value $+\infty$. Our basic setting is *feasible* if the functional $\mathscr{E}$ is not identically $+\infty$, i.e. there exists at least one plan $\boldsymbol{\gamma}$ with $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$.

3.2 The primal formulation of the Optimal Entropy-Transport problem

In the basic setting described in the previous section 3.1, we want to investigate the following problem.

Problem 3.1 (Entropy-Transport minimization)**.** *Given $\mu_i \in \mathcal{M}(X_i)$ find $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ minimizing $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$, i.e.*

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \mathsf{ET}(\mu_1, \mu_2) := \inf_{\boldsymbol{\sigma} \in \mathcal{M}(X_1 \times X_2)} \mathscr{E}(\boldsymbol{\sigma}|\mu_1, \mu_2). \tag{3.5}$$

We denote by $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2) \subset \mathcal{M}(\boldsymbol{X})$ *the collection of all the minimizers of* (3.5).

Remark 3.2 (Feasibility conditions)**.** Problem 3.1 is feasible if there exists at least one plan $\boldsymbol{\gamma}$ with $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$. Notice that this is always the case when

$$F_i(0) < \infty, \quad i = 1, 2, \tag{3.6}$$

since among the competitors one can choose the null plan $\boldsymbol{\eta}$, so that

$$\mathsf{ET}(\mu_1,\mu_2) \le \mathscr{E}(\boldsymbol{\eta}|\mu_1,\mu_2) = F_1(0)\mu_1(X) + F_2(0)\mu_2(X). \tag{3.7}$$

More generally, thanks to (3.1) a sufficient condition for feasibility in the non-degenerate case $m_1m_2 \neq 0$ is

$$\mathsf{c}(x_1,x_2) \le B_1(x_1) + B_2(x_2), \quad B_i \in \mathrm{L}^1(X_i,\mu_i); \tag{3.8}$$

in fact, the plans

$$\boldsymbol{\gamma} = \frac{\theta}{m_1m_2}\mu_1\otimes\mu_2 \quad \text{with } \theta \in J \quad \text{given by (3.1)} \tag{3.9}$$

are Radon [38, Thm. 17, p. 63], have finite cost and provide the estimate

$$\mathsf{ET}(\mu_1,\mu_2) \le m_1F_1(\theta/m_1) + m_2F_2(\theta/m_2) + \theta\sum_i {m_i}^{-1}\|B_i\|_{\mathrm{L}^1(X_i,\mu_i)}, \quad \text{for every } \theta\in J. \tag{3.10}$$

Notice that (3.1) is also necessary for feasibility: in fact, setting $m_{i,n} := m_i + \gamma_i^\perp(X_i)/n$, the convexity of F_i, the definition (2.16) of $(F_i)'_\infty$, and Jensen's inequality provide

$$\begin{aligned}\mathscr{F}_i(\boldsymbol{\gamma}|\mu_i) &= \int_{X_i} F_i(\sigma_i)\,\mathrm{d}\mu_i + \lim_{n\uparrow\infty}\int_{X_i} F_i(n)\,\mathrm{d}(n^{-1}\gamma_i^\perp) \ge \lim_{n\to\infty} m_{i,n}F_i\big(\gamma_i(X_i)/m_{i,n}\big)\\ &\ge m_iF_i(m/m_i), \qquad m := \gamma_i(X_i) = \boldsymbol{\gamma}(\boldsymbol{X}),\end{aligned} \tag{3.11}$$

so that, whenever $\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) < \infty$,

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) \ge m\inf\mathsf{c} + m_1F_1(m/m_1) + m_2F_2(m/m_2), \quad m := \boldsymbol{\gamma}(\boldsymbol{X}), \tag{3.12}$$

and therefore

$$\boldsymbol{\gamma}(\boldsymbol{X}) \in \big(m_1\,\mathrm{Dom}(F_1)\big) \cap \big(m_2\,\mathrm{Dom}(F_2)\big) = J. \tag{3.13}$$

We will often reinforce (3.1) by assuming that at least one of the domains of the entropies F_i has nonempty interior, containing a point of the other domain:

$$\Big(\mathrm{int}\,\big(m_1\mathrm{Dom}(F_1)\big) \cap m_2\mathrm{Dom}(F_2)\Big) \cup \Big(m_1\mathrm{Dom}(F_1) \cap \mathrm{int}\,\big(m_2\mathrm{Dom}(F_2)\big)\Big) \neq \emptyset. \tag{3.14}$$

(3.14) is surely satisfied if J has nonempty interior, i.e. $\max(m_1s_1^-, m_2s_2^-) < \min(m_1s_1^+, m_2s_2^+)$, where $s_i^- = \inf\mathrm{Dom}(F_i)$, $s_i^+ := \sup\mathrm{Dom}(F_i)$. □

We also observe that whenever $\mu_i(X_i) = 0$ then the null plan $\boldsymbol{\gamma} = \boldsymbol{\eta}_0$ provides the trivial solution to Problem 3.1. Another trivial case occurs when $F_i(0) < \infty$ and F_i are nondecreasing in $\mathrm{Dom}(F_i)$ (in particular when $F_i(0) = 0$). Then it is clear that the null plan is a minimizer and $\mathsf{ET}(\mu_1,\mu_2) = F_1(0)m_1 + F_2(0)m_2$.

3.3 Examples

Let us consider a few particular cases:

E.1 $\mathsf{c} \equiv 0$. Since F_i are convex, in this case the minimum is attained when the marginals γ_i have constant densities. Setting $\sigma_i \equiv \theta/m_i$ in order to have $m_1\sigma_1 = m_2\sigma_2$, we thus have

$$\mathsf{ET}(\mu_1, \mu_2) = H_0(m_1, m_2) := \min\Big\{m_1 F_1(\theta/m_1) + m_2 F_2(\theta/m_2) : \theta \geq 0\Big\}. \tag{3.15}$$

E.2 **Entropy-potential problems**: if $\mu_2 \equiv \eta_0$ then setting $V(x_1) := \inf_{x_2 \in X_2} \mathsf{c}(x_1, x_2)$ we easily get

$$\mathsf{ET}(\mu, 0) = \inf_{\gamma \in \mathcal{M}(X_1)} \mathscr{F}_1(\gamma|\mu) + \int_{X_1} V \,\mathrm{d}\gamma + (F_2)'_\infty \gamma(X_1). \tag{3.16}$$

E.3 **Pure transport problems:** $F_i(r) = \mathrm{I}_1(r) = \begin{cases} 0 & \text{if } r = 1 \\ +\infty & \text{otherwise.} \end{cases}$

In this case any feasible plan $\boldsymbol{\gamma}$ should have μ_1 and μ_2 as marginals and the functional just reduces to the pure transport part

$$\mathsf{T}(\mu_1, \mu_2) = \min\Big\{\int_{X_1 \times X_2} \mathsf{c} \,\mathrm{d}\boldsymbol{\gamma} : \quad \pi^i_\sharp \boldsymbol{\gamma} = \mu_i\Big\}. \tag{3.17}$$

As a necessary condition for feasibility we get $\mu_1(X_1) = \mu_2(X_2)$. A situation equivalent to the optimal transport case occurs when

$$\operatorname{int}\big(m_1 \mathrm{Dom}(F_1)\big) \cap m_2 \mathrm{Dom}(F_2) = m_1 \mathrm{Dom}(F_1) \cap \operatorname{int}\big(m_2 \mathrm{Dom}(F_2)\big) = \emptyset, \tag{3.18}$$

so that the set J defined by (3.1) contains only one point θ which separates $m_1\mathrm{Dom}(F_1)$ and $m_2\mathrm{Dom}(F_2)$:

$$\theta = m_1 s_1^+ = m_2 s_2^- \quad \text{os} \quad \theta = m_1 s_1^- = m_2 s_2^+. \tag{3.19}$$

It is not difficult to check that in this case

$$\mathsf{ET}(\mu_1, \mu_2) = m_1 F_1(\theta/m_1) + m_2 F_2(\theta/m_2) + \mathsf{T}(\mu_1, \mu_2). \tag{3.20}$$

E.4 **Optimal transport with density constraints:** $F_i(r) := \mathrm{I}_{[a_i, b_i]}(r)$, $a_i \leq 1 \leq b_i$. E.g. when $a_i = 1$, $b_i = \infty$ we have

$$\mathsf{ET}(\mu_1, \mu_2) = \min\Big\{\int_{X_1 \times X_2} \mathsf{c} \,\mathrm{d}\boldsymbol{\gamma} : \quad \pi^i_\sharp \boldsymbol{\gamma} \geq \mu_i\Big\} \tag{3.21}$$

and for $[a_1, b_1] = [0, 1]$ and $[a_2, b_2] = [1, +\infty]$ we get

$$\mathsf{ET}(\mu_1, \mu_2) = \min\Big\{\int_{X_1 \times X_2} \mathsf{c} \,\mathrm{d}\boldsymbol{\gamma} : \quad \pi^1_\sharp \boldsymbol{\gamma} \leq \mu_1, \ \pi^2_\sharp \boldsymbol{\gamma} \geq \mu_2\Big\}, \tag{3.22}$$

whose feasibility requires $\mu_2(X_2) \geq \mu_1(X_1)$.

E.5 **Pure entropy problems:** $X_i = X$, $(F_i)'_\infty = +\infty$, $\mathsf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2 \\ +\infty & \text{otherwise.} \end{cases}$

In this case the marginals of $\boldsymbol{\gamma}$ coincide: we denote them by γ. We can write the density of γ w.r.t. any measure μ such that $\mu_i \ll \mu$ (say, e.g., $\mu = \mu_1 + \mu_2$) as $\boldsymbol{\gamma} = \vartheta\mu$ and then $\mu_i = \vartheta_i \mu$. Since $\boldsymbol{\gamma} \ll \mu_i$ we have $\vartheta(x) = 0$ for μ-a.e. x where $\vartheta_1(x)\vartheta_2(x) = 0$. Thus $\sigma_i = \vartheta/\vartheta_i$ is well defined and we have

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \int_X \Big(\vartheta_1 F_1(\vartheta/\vartheta_1) + \vartheta_2 F_2(\vartheta/\vartheta_2)\Big)\,\mathrm{d}\mu, \tag{3.23}$$

with the convention that $\vartheta_i F_i(\vartheta/\vartheta_i) = 0$ if $\vartheta = \vartheta_i = 0$. Since we expressed everything in terms of μ, by recalling the definition of the function H_0 given in (3.15) we get

$$\mathsf{ET}(\mu_1, \mu_2) = \int_X H_0\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}, \frac{\mathrm{d}\mu_2}{\mathrm{d}\mu}\Big)\,\mathrm{d}\mu, \quad \mu_i \ll \mu. \tag{3.24}$$

In the Hellinger case $F_i(s) = U_1(s) = s\log s - s + 1$ a simple calculation yields

$$H_0(\theta_1, \theta_2) = \theta_1 + \theta_2 - 2\sqrt{\theta_1\theta_2} = \Big(\sqrt{\theta_1} - \sqrt{\theta_2}\Big)^2. \tag{3.25}$$

In the Jensen-Shannon case, where $F_i(s) = U_0(s) = s - 1 - \log s$, we obtain

$$H_0(\theta_1; \theta_2) = \theta_1 \log\Big(\frac{2\theta_1}{\theta_1 + \theta_2}\Big) + \theta_2 \log\Big(\frac{2\theta_2}{\theta_1 + \theta_2}\Big);$$

Two other interesting examples are provided by the quadratic case $F_i(s) = \frac{1}{2}(s-1)^2$ and by the nonsmooth "piecewise affine" case $F_i(s) = |s-1|$, for which we respectively obtain

$$H_0(\theta_1, \theta_2) = \frac{1}{2(\theta_1 + \theta_2)}(\theta_1 - \theta_2)^2, \quad H_0(\theta_1, \theta_2) = |\theta_1 - \theta_2|.$$

E.6 **Regular entropy-transport problems:** they correspond to the choice of a couple of differentiable entropies F_i with $\mathrm{Dom}(F_i) \supset (0, \infty)$, as in the case of the power-like entropies U_p defined in (2.27). When they vanish (and thus have a minimum) at $s = 1$, the Entropic Optimal Transportation can be considered as a smooth relaxation of the Optimal Transport case E.3.

E.7 **Squared Hellinger-Kantorovich distances:** $X_1 = X_2$, τ is induced by a distance d, $F_1(s) = F_2(s) := U_1(s) = s\log s - s + 1$ and

$$\mathsf{c}(x_1, x_2) := \mathsf{d}^2(x_1, x_2) \quad \text{or} \quad \mathsf{c}(x_1, x_2) := -\log\Big(\cos^2\big(\mathsf{d}(x_1, x_2) \wedge \pi/2\big)\Big).$$

This case will be thoroughly studied in the second part of the present paper.

E.8 **Marginal Entropy-Transport problems:** in this case one of the two marginals of $\boldsymbol{\gamma}$ is fixed, say γ_1, by choosing $F_1(r) := \mathrm{I}_1(r)$; thus the functional minimizes the sum of the transport cost and the relative entropy of the second marginal $\mathscr{F}_2(\gamma_2|\mu_2)$ with respect to a reference measure μ_2:

$$\mathsf{ET}(\mu_1, \mu_2) = \min_{\gamma \in \mathcal{M}(X_2)} \Big\{\mathscr{F}_2(\gamma|\mu_2) + \mathsf{T}(\gamma, \mu_1)\Big\}.$$

This is the typical situation one has to solve at each iteration step of the Minimizing Movement scheme [2], when T is a (power of a) transport distance induced by c, as in the Jordan-Kinderlehrer-Otto approach [20].

E.9 **The discrete case.** Let $\mu_1 = \sum_{i=1}^m \alpha_i \delta_{x_i}$, $\mu_2 = \sum_{j=1}^N \beta_j \delta_{y_j}$ with $\alpha_i, \beta_j > 0$, and let $\mathsf{c}_{i,j} := \mathsf{c}(x_i, y_j)$. The Entropy-Transport problem for this discrete model consists in finding coefficients $\gamma_{i,j} \geq 0$ which minimize

$$\mathscr{E}(\gamma_{i,j}|\alpha_i, \beta_j) := \sum_i \alpha_i F_1\Big(\frac{\sum_j \gamma_{i,j}}{\alpha_i}\Big) + \sum_j \beta_j F_2\Big(\frac{\sum_i \gamma_{i,j}}{\beta_j}\Big) + \sum_{i,j} \mathsf{c}_{i,j}\gamma_{i,j}. \tag{3.26}$$

3.4 Existence of solutions to the primal problem

The next result provides a first general existence result for Problem 3.1 in the basic coercive setting of Section 3.1.

Theorem 3.3 (Existence of minimizers)**.** *Let us assume that Problem 3.1 is feasible (see Remark 3.2) and coercive, i.e. at least one of the following conditions hold:*

(i) the entropy functions F_i are superlinear, i.e. $(F_1)'_\infty = (F_2)'_\infty = +\infty$;

(ii) c has compact sublevels in $\boldsymbol{X}$ and $(F_1)'_\infty + (F_2)'_\infty + \inf \mathsf{c} > 0$.

Then Problem 3.1 admits at least one optimal solution. In this case $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ is a compact convex set of $\mathcal{M}(\boldsymbol{X})$.

Proof. We can apply the Direct Method of Calculus of Variations: since the map $\boldsymbol{\gamma} \mapsto \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$ is lower semicontinuous in $\mathcal{M}(X_1 \times X_2)$ by Theorem 2.7, it is sufficient to show that its sublevels are relatively compact, thus bounded and equally tight by Prokhorov Theorem 2.2. In both cases boundedness follows by the estimate (3.12), observing that for a function $F \in \Gamma(\mathbb{R}_+)$ the coercivity property $\lim_{r\to\infty} F(r) = +\infty$ is equivalent to $F'_\infty > 0$.

In case *(ii)* equal tightness is a consequence of the Markov inequality and the non-negativity of F_i: in fact, considering the compact sublevels $K_\lambda := \{(x_1, x_2) \in X_1 \times X_2 : \mathsf{c}(x_1, x_2) \leq \lambda\}$, we have

$$\boldsymbol{\gamma}(\boldsymbol{X} \setminus K_\lambda) \leq \lambda^{-1} \int \mathsf{c}\, \mathrm{d}\boldsymbol{\gamma} \leq \lambda^{-1} \mathscr{E}(\gamma|\mu_1, \mu_2) \quad \text{for every } \lambda > 0.$$

In the case *(i)*, since $\mathsf{c} \geq 0$ Proposition 2.10 shows that both the marginals of plans in a sublevel of the energy are equally tight: we thus conclude by [2, Lemma 5.2.2]. $\square$

Remark 3.4. The assumptions *(i)* and *(ii)* in the previous Theorem are almost optimal and it is possible to find counterexamples when they are not satisfied. In the case when $0 < (F_1)'_\infty + (F_2)'_\infty < \infty$ but c does not have compact sublevels one can just take $F_i(s) := U_0(s) = s - \log s - 1$, $X_i := \mathbb{R}$, $\mathsf{c}(x_1, x_2) := 3\mathrm{e}^{-x_1^2 - x_2^2}$, $\mu_i = \delta_0$. Any competitor is of the form $\boldsymbol{\gamma} := \alpha\delta_0 \otimes \delta_0 + \nu_1 \otimes \delta_0 + \delta_0 \otimes \nu_2$, with $\nu_i \in \mathcal{M}(\mathbb{R})$, $\nu_i(\{0\}) = 0$, $\nu_i(\mathbb{R}) = n_i$, so that

$$\mathscr{E}(\gamma|\mu_1, \mu_2) = F(\alpha + n_1) + F(\alpha + n_2) + 3\Big(\alpha + \int \mathrm{e}^{-x^2}\, \mathrm{d}(\nu_1 + \nu_2)\Big) + n_1 + n_2.$$

Since $\min_s F(s) + s = \log 2$ is attained at $s = 1/2$, we immediately see that

$$\mathscr{E}(\gamma|\mu_1, \mu_2) \geq 2\log 2 + \alpha + 3\int \mathrm{e}^{-x^2}\,\mathrm{d}(\nu_1 + \nu_2) \geq 2\log 2.$$

and $2\log 2$ is the infimum, obtained by choosing $\alpha = 0$ and $\nu_1 = \nu_2 = \frac{1}{2}\delta_x$, and letting $x \to \infty$. On the other hand, since $n_1 + n_2 + \alpha > 0$, the infimum can never be attained.

In the case when c has compact sublevels but $(F_1)'_\infty = (F_2)'_\infty = \min \mathsf{c} = 0$, it is sufficient to take $F_i(s) := s^{-1}$, $X_i = [-1, 1]$, $\mathsf{c}(x_1, x_2) = x_1^2 + x_2^2$, and $\mu_i = \delta_0$. Taking $\gamma_n := n\delta_0 \otimes \delta_0$ one easily checks that $\inf \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = 0$ but $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) > 0$ for every $\boldsymbol{\gamma} \in \mathcal{M}(\mathbb{R}^2)$. □

Let us briefly discuss the question of uniqueness.

Lemma 3.5 (Uniqueness of the marginals in the superlinear strictly convex case)**.** *Let us suppose that F_i are* strictly convex *functions. Then the μ_i-absolutely continuous part $\sigma_i\mu_i$ of the marginals $\gamma_i = \pi^i_\sharp\boldsymbol{\gamma}$ of any optimal plan are uniquely determined. In particular, if F_i are also superlinear, then the marginals γ_i are uniquely determined, i.e. if $\boldsymbol{\gamma}', \boldsymbol{\gamma}'' \in$* $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ *then $\pi^i_\sharp\boldsymbol{\gamma}' = \pi^i_\sharp\boldsymbol{\gamma}''$, $i = 1, 2$.*

Proof. It is sufficient to take $\boldsymbol{\gamma} = \frac{1}{2}\boldsymbol{\gamma}' + \frac{1}{2}\boldsymbol{\gamma}''$ which is still optimal in $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ since $\mathscr{E}$ is a convex functional w.r.t. $\boldsymbol{\gamma}$. We have $\pi^i_\sharp\boldsymbol{\gamma} = \gamma_i = \frac{1}{2}\gamma'_i + \frac{1}{2}\gamma''_i = \frac{1}{2}(\sigma'_i + \sigma''_i)\mu + \frac{1}{2}(\gamma'_i)^\perp + \frac{1}{2}(\gamma''_i)^\perp$ and we observe that the minimality of $\boldsymbol{\gamma}$ and the convexity of each addendum F_i in the functional yield

$$\mathscr{F}_i(\gamma_i|\mu_i) = \frac{1}{2}\mathscr{F}_i(\gamma'_i|\mu_i) + \frac{1}{2}\mathscr{F}_i(\gamma''_i|\mu_i) \quad i = 1, 2.$$

Since $\gamma_i^\perp(X_i) = \frac{1}{2}(\gamma'_i)^\perp(X_i) + \frac{1}{2}(\gamma''_i)^\perp(X_i)$ we obtain

$$\int_X \Big(F_i(\sigma_i) - \frac{1}{2}F_i(\sigma'_i) - \frac{1}{2}F_i(\sigma''_i)\Big)\,\mathrm{d}\mu_i = 0 \quad i = 1, 2.$$

Since F_i is strictly convex, the above identity implies $\sigma_i = \sigma'_i = \sigma''_i$ μ_i-a.e. in X. □

The next Corollary reduces the question of uniqueness of optimal couplings in $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ to corresponding results for the Kantorovich problem associated to the cost c.

Corollary 3.6. *Let us suppose that F_i are* superlinear strictly convex *functions and that for every couple of probability measures $\nu_i \in \mathcal{P}(X_i)$ with $\nu_i \ll \mu_i$ the optimal transport problem associated to the cost* c *(see Example* E.3 *of Section 3.3) admits a unique solution. Then* $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ *contains at most one plan.*

Proof. We can assume $m_1\, m_2 \neq 0$. It is clear that any $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ is a solution of the optimal transport problem for the cost c and given (possibly normalized) marginals γ_i. Since $\gamma_i \ll \mu_i$ we conclude. □

Example 3.7 (Uniqueness in Euclidean spaces)**.** If F_i are *superlinear strictly convex* functions, $\mathsf{c}(x, y) = h(x - y)$ for a strictly convex function $h : \mathbb{R}^d \to [0, \infty)$ and $\mu_1 \ll \mathscr{L}^d$, then Problem 3.1 admits at most one solution. It is sufficient to apply the previous corollary in conjunction with [2, Theorem 6.2.4]

Example 3.8 (Non-uniqueness of optimal couplings)**.** Consider the logarithmic density functionals $F_i(s) = U_1(s) = s\log s - s + 1$, the Euclidean space $X_1 = X_2 = \mathbb{R}^2$ and any cost c of the form $\mathsf{c}(x_1, x_2) = h(|x_1 - x_2|)$. The configuration

$$\mu_1 = \delta_{(-1,0)} + \delta_{(1,0)}, \quad \mu_2 \text{ with support in } \{0\} \times \mathbb{R} \text{ and containing at least two points,}$$

admits always an infinite number of optimal plans. In fact, we shall see that the first marginal γ_1 of any optimal plan $\boldsymbol{\gamma}$ will have full support in $(-1,0), (1,0)$, i.e. it will of the form $a\delta_{(-1,0)} + b\delta_{(1,0)}$ with strictly positive a, b, and the support of the second marginal γ_2 will be concentrated in $\{0\} \times \mathbb{R}$ and will contain at least two points. Any other plan $\boldsymbol{\sigma}$ with marginals γ_1, γ_2 will then be optimal since it can be written as the disintegration

$$\boldsymbol{\sigma} = \int_{\mathbb{R}} \Big(\alpha(y)\delta_{(-1,0)} + \beta(y)\delta_{(1,0)}\Big)\,\mathrm{d}\gamma_2(y)$$

with arbitrary nonnegative densities α, β with $\alpha + \beta = 1$ and $\int \alpha\,\mathrm{d}\gamma_2(y) = a$, $\int \beta\,\mathrm{d}\gamma_2(y) = b$. In fact, the cost contribution of $\boldsymbol{\sigma}$ to the total energy is

$$\int_{\mathbb{R}} h(\sqrt{1+y^2})\,\mathrm{d}\gamma_2(y)$$

and it is independent of the choice of α and β. □

We conclude this section by proving a simple lower semicontinuity property for ET; notice that in metrizable spaces any weakly convergent sequence of Radon measures is tight.

Lemma 3.9. *Let $\mathbb{L}$ be a directed set, $(F_i^\lambda)_{\lambda\in\mathbb{L}}$ and $(\mathsf{c}^\lambda)_{\lambda\in\mathbb{L}}$ be monotone nets of superlinear entropies and costs pointwise converging to F_i and c respectively, and let $(\mu_i^\lambda)_{\lambda\in\mathbb{L}}$ be equally tight nets of measures narrowly converging to μ_i in $\mathcal{M}(X_i)$. Denoting by ET^λ (resp. ET) the corresponding Entropy-Transport functionals induced by F_i^λ and c^λ (resp. F_i and c) we have*

$$\liminf_{\lambda\in\mathbb{L}} \mathsf{ET}^\lambda(\mu_1^\lambda, \mu_2^\lambda) \geq \mathsf{ET}(\mu_1, \mu_2). \tag{3.27}$$

Proof. Let $\boldsymbol{\gamma}^\lambda \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1^\lambda, \mu_2^\lambda) \subset \mathcal{M}(\boldsymbol{X})$ be a corresponding net of optimal plans. The thesis follows if assuming that $\mathscr{E}(\boldsymbol{\gamma}^\lambda|\mu_1^\lambda, \mu_2^\lambda) = \mathsf{ET}(\mu_1^\lambda, \mu_2^\lambda) \leq C < \infty$ we can prove that $\mathsf{ET}(\mu_1, \mu_2) \leq C$. By applying Proposition 2.10 we obtain that the sequences of marginals $\pi^i_\sharp\boldsymbol{\gamma}^\lambda$ are tight in $\mathcal{M}(X_i)$, so that the net $\boldsymbol{\gamma}^\lambda$ is also tight. By extracting a suitable subnet (not relabeled) narrowly converging to $\boldsymbol{\gamma}$ in $\mathcal{M}(\boldsymbol{X})$, we can still apply Proposition 2.10 and the lower semicontinuity of the transport part $\mathscr{F}^\lambda$ of the functional $\mathscr{E}$ to obtain $\liminf_{\lambda\in\mathbb{L}} \mathscr{F}^\lambda(\boldsymbol{\gamma}^\lambda|\mu_1^\lambda, \mu_2^\lambda) \geq \mathscr{F}(\boldsymbol{\gamma}|\mu_1, \mu_2)$. A completely analogous argument shows that $\liminf_{\lambda\in\mathbb{L}} \int \mathsf{c}^\lambda\,\mathrm{d}\boldsymbol{\gamma}^\lambda \geq \int \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma}$. □

As a simple application we prove the extremality of the class of Optimal Transport problems (see Example E.3 in Section 3.3).

Corollary 3.10. *Let $F_1, F_2 \in \Gamma(\mathbb{R}_+)$ be satisfying $F_i(r) > F_i(1) = 0$ for every $r \in [0,\infty)$, $r \neq 1$ and let ET^n be the Optimal Entropy Transport value* (3.5) *associated to (nF_1, nF_2). Then for every couple of equally tight sequences $(\mu_{1,n}, \mu_{2,n}) \subset \mathcal{M}(X_1)\times\mathcal{M}(X_2)$, $n \in \mathbb{N}$, narrowly converging to (μ_1, μ_2) we have*

$$\lim_{n\uparrow\infty} \mathsf{ET}^n(\mu_{1,n}, \mu_{2,n}) = \mathsf{T}(\mu_1, \mu_2). \tag{3.28}$$

3.5 The reverse formulation of the primal problem

Let us introduce the reverse entropy functions R_i (see (2.29))

$$R_i(r) := \begin{cases} rF_i(1/r) & \text{if } r > 0, \\ (F_i)'_\infty & \text{if } r = 0, \end{cases} \tag{3.29}$$

and let $\mathscr{R}_i$ be the corresponding integral functionals as in (2.56).

Keeping the usual notation

$$\gamma_i := \pi^i_\sharp \boldsymbol{\gamma} \in \mathcal{M}(X_i), \quad \mu_i = \varrho_i \gamma_i + \mu_i^\perp, \quad \varrho_i = \frac{\mathrm{d}\mu_i}{\mathrm{d}\gamma_i}, \tag{3.30}$$

we can thus define

$$\begin{aligned} \mathscr{R}(\mu_1, \mu_2 | \boldsymbol{\gamma}) &:= \sum_i \mathscr{R}_i(\mu_i|\gamma_i) + \int_{\boldsymbol{X}} \mathsf{c} \, \mathrm{d}\boldsymbol{\gamma} = \\ &= \int_{\boldsymbol{X}} \Big(R_1(\varrho_1(x_1)) + R_2(\varrho_2(x_2)) + \mathsf{c}(x_1, x_2) \Big) \, \mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i^\perp(X_i). \end{aligned} \tag{3.31}$$

By Lemma 2.11 we easily get the reverse formulation of the optimal Entropy-Transport problem 3.1.

Theorem 3.11. *For every $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ and $\mu_i \in \mathcal{M}(X_i)$*

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \mathscr{R}(\mu_1, \mu_2|\boldsymbol{\gamma}). \tag{3.32}$$

In particular

$$\mathsf{ET}(\mu_1, \mu_2) = \inf_{\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})} \mathscr{R}(\mu_1, \mu_2|\boldsymbol{\gamma}), \tag{3.33}$$

and $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ if and only if it minimizes $\mathscr{R}(\mu_1, \mu_2|\cdot)$ in $\mathcal{M}(\boldsymbol{X})$.

4 The dual problem

In this section we want to compute and study the dual problem and the corresponding optimality conditions for the Entropy-Transport Problem 3.1 in the basic *coercive* setting of Section 3.1.

4.1 The "inf-sup" derivation of the dual problem in the basic coercive setting

In order to write the first formulation of the dual problem we introduce the reverse entropy functions R_i defined as in (2.29) and their conjugate $R_i^* : \mathbb{R} \to (-\infty, +\infty]$ which can be expressed by

$$R_i^*(\psi) := \sup_{s>0} \big(s\psi - sF_i(1/s) \big) = \sup_{r>0} r^{-1} \big(\psi - F_i(r) \big). \tag{4.1}$$

(2.32) yields

$$(\phi, \psi) \in \mathfrak{F}_i \quad \Leftrightarrow \quad \phi \le -R_i^*(\psi) \quad (\phi, \psi) \in \mathbb{R}^2. \tag{4.2}$$

The first step is to use the dual formulation of the entropy functionals given by Theorem 2.7: we have

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2)=\int \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma}+\sup\Big\{\sum_i\int_{X_i}\psi_i\,\mathrm{d}\mu_i-\sum_i\int_{X_i}R_i^*(\psi_i)\,\mathrm{d}\gamma_i:\psi_i,R_i^*(\psi_i)\in\mathrm{LSC}_s(X_i)\Big\}.$$

It is natural to introduce the saddle function $\mathscr{L}(\boldsymbol{\gamma},\boldsymbol{\psi})$ depending on $\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})$ and $\boldsymbol{\psi}=(\psi_1,\psi_2)$ (we omit here the dependence on μ_1,μ_2)

$$\mathscr{L}(\boldsymbol{\gamma},\boldsymbol{\psi}):=\int_{\boldsymbol{X}}\Big(\mathsf{c}(x_1,x_2)-R_1^*(\psi_1(x_1))-R_2^*(\psi_2(x_2))\Big)\,\mathrm{d}\boldsymbol{\gamma}+\sum_i\int_{X_i}\psi_i\,\mathrm{d}\mu_i. \tag{4.3}$$

In order to guarantee that $\mathscr{L}$ takes real values, we consider the convex set

$$\mathrm{M}:=\Big\{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X}):\int\mathsf{c}\,\mathrm{d}\boldsymbol{\gamma}<\infty\Big\}; \tag{4.4}$$

we thus have

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2)=\sup_{\psi_i,R_i^*(\psi_i)\in\mathrm{LSC}_s(X_i)}\mathscr{L}(\boldsymbol{\gamma},\boldsymbol{\psi})$$

and the Entropy-Transport problem can be written as

$$\mathsf{ET}(\mu_1,\mu_2)=\min_{\boldsymbol{\gamma}\in\mathrm{M}}\ \sup_{\psi_i,R_i^*(\psi_i)\in\mathrm{LSC}_s(X_i)}\mathscr{L}(\boldsymbol{\gamma},\boldsymbol{\psi}). \tag{4.5}$$

We can then obtain the dual problem by interchanging the order of min and sup. Let us denote by $\varphi_1\oplus\varphi_2$ the function $(x_1,x_2)\mapsto\varphi_1(x_1)+\varphi_2(x_2)$. Since for every $\boldsymbol{\psi}=(\psi_1,\psi_2)$ with $\psi_i,R_i^*(\psi_i)\in\mathrm{LSC}_s(X_i)$

$$\inf_{\boldsymbol{\gamma}\in\mathrm{M}}\int\Big(\mathsf{c}(x_1,x_2)-R_1^*(\psi_1(x_1))-R_2^*(\psi_2(x_2))\Big)\,\mathrm{d}\boldsymbol{\gamma}=\begin{cases}0&\text{if }R_1^*(\psi_1)\oplus R_2^*(\psi_2)\le\mathsf{c},\\-\infty&\text{otherwise,}\end{cases}$$

we get

$$\inf_{\boldsymbol{\gamma}\in\mathrm{M}}\mathscr{L}(\boldsymbol{\gamma},\boldsymbol{\psi})=\begin{cases}\displaystyle\sum_i\int_{X_i}\psi_i\,\mathrm{d}\mu_i&\text{if }R_1^*(\psi_1)\oplus R_2^*(\psi_2)\le\mathsf{c},\\-\infty&\text{otherwise.}\end{cases} \tag{4.6}$$

(4.6) thus provides the dual formulation, that we will study in the next section.

4.2 Dual problem and optimality conditions

Problem 4.1 (ψ-formulation of the dual problem)**.** *Let R_i^* be the convex functions defined by* (4.1) *and let* $\boldsymbol{\Psi}$ *be the the convex set*

$$\boldsymbol{\Psi}:=\Big\{\boldsymbol{\psi}\in\mathrm{LSC}_s(X_1)\times\mathrm{LSC}_s(X_2):\ R_i^*(\psi_i)\ \textit{bounded},\ R_1^*(\psi_1)\oplus R_2^*(\psi_2)\le\mathsf{c}\Big\}. \tag{4.7}$$

The dual Entropy-Transport problem consists in finding

$$\mathsf{D}(\mu_1,\mu_2)=\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\int_{X_1}\psi_1\,\mathrm{d}\mu_1+\int_{X_2}\psi_2\,\mathrm{d}\mu_2. \tag{4.8}$$

As usual, by operating the change of variable

$$\varphi_i := -R^*(\psi_i), \quad \psi_i = F_i^\circ(\varphi_i) := -F_i^*(-\varphi_i), \tag{4.9}$$

we can obtain an equivalent formulation of the dual functional D as the supremum of the concave functionals

$$\mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2) := \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i, \tag{4.10}$$

on the simpler convex set

$$\boldsymbol{\Phi} := \Big\{\boldsymbol{\varphi} \in \mathrm{LSC}_s(X_1)\times\mathrm{LSC}_s(X_2),\ F_i^\circ(\varphi_i)\text{ bounded},\ \varphi_1\oplus\varphi_2 \le \mathsf{c}\Big\}. \tag{4.11}$$

Problem 4.2 (φ-formulation of the dual problem)**.** *Let F_i° be the concave functions defined by* (4.9) *and let $\boldsymbol{\Phi}$ be the the convex set* (4.11)*. The φ-formulation of the dual Entropy-Transport problem consists in finding*

$$\mathsf{D}'(\mu_1,\mu_2) = \sup_{\boldsymbol{\varphi}\in\boldsymbol{\Phi}} \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2) = \sup_{\boldsymbol{\varphi}\in\boldsymbol{\Phi}} \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i. \tag{4.12}$$

Proposition 4.3 (Equivalence of the dual formulations)**.** *The ψ- and the ϕ- formulations of the dual problem are equivalent,* $\mathsf{D}(\mu_1,\mu_2) = \mathsf{D}'(\mu_1,\mu_2)$.

Proof. Since R_i^* is nondecreasing, for every $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ the functions $\varphi_i := R_i^*(\psi_i)$ belong to $\mathrm{LSC}_s(X_i)$ and satisfy $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$, with $(-\varphi_i,\psi_i) \in \mathfrak{F}_i$. It then follows that $\tilde\psi_i := -F_i^*(-\varphi_i) = F_i^\circ(\varphi_i) \ge \psi_i$ are bounded, so that $(\varphi_1,\varphi_2)\in\boldsymbol{\Phi}$ and $\mathsf{D}' \ge \mathsf{D}$. An analogous argument shows the converse inequality. □

Since "$\inf\sup \ge \sup\inf$", (4.5) yields

$$\mathsf{ET}(\mu_1,\mu_2) \ge \mathsf{D}(\mu_1,\mu_2). \tag{4.13}$$

We will show in the next section that (4.13) is in fact an equality. Here we primarily discuss for which class of functions ψ_i, φ_i the dual formulations are still meaningful and which form the optimality conditions associated to the equality case in (4.13) take.

It is intended that in some cases we will also consider larger classes of potentials $\boldsymbol{\psi}$ or $\boldsymbol{\varphi}$ by allowing Borel functions with extended real values under suitable summability conditions.

I. First of all, recalling (2.20) and (2.30), we extend R^* and F° to $\bar{\mathbb{R}}$ by setting

$$R^*(-\infty) := -F'_\infty, \quad R^*(+\infty) := +\infty; \quad F^\circ(-\infty) := -\infty, \quad F^\circ(+\infty) := F(0), \tag{4.14}$$

and we observe that with the definition above and according to (2.38)–(2.39) the couples

$$(-\varphi, F^\circ(\varphi)) \text{ and } (-R^*(\psi),\psi) \text{ belong to } \bar{\mathfrak{F}} \text{ whenever } \psi \le F(0) \text{ and } \varphi \ge -F'_\infty. \tag{4.15}$$

II. We also set

$$\zeta_1 +_o \zeta_2 := \lim_{n\to\infty} (-n \vee \zeta_1 \wedge n) + (-n \vee \zeta_2 \wedge n) \quad \text{for every } \zeta_1, \zeta_2 \in \bar{\mathbb{R}}. \tag{4.16}$$

Notice that $(\pm\infty)+_o(\pm\infty) = \pm\infty$ and in the ambiguous case $+\infty-\infty$ this definition yields $(+\infty)+_o(-\infty) = 0$; we correspondingly extend the definition of $\oplus$ by setting

$$(\zeta_1 \oplus_o \zeta_2)(x_1, x_2) := \zeta_1(x_1) +_o \zeta_2(x_2) \quad \text{for every} \quad \zeta_i \in \mathrm{B}(X_i; \bar{\mathbb{R}}). \tag{4.17}$$

The following result is the natural extension of Lemma 2.6.

Proposition 4.4 (Dual lower bound for extended real valued potentials)**.** *Let $\boldsymbol{\gamma}$ be a feasible plan and let $\boldsymbol{\varphi} \in \mathrm{B}(X_1; \bar{\mathbb{R}}) \times \mathrm{B}(X_2; \bar{\mathbb{R}})$ with $\varphi_i \geq -(F_i)'_\infty$, $\varphi_1 \oplus_o \varphi_2 \leq \mathsf{c}$ with $(F_i^\circ \circ \varphi_i)_- \in \mathrm{L}^1(X_i, \mu_i)$ (resp. $(\varphi_i)_+ \in \mathrm{L}^1(X_i, \gamma_i)$).*

Then we have $(\varphi_i)_- \in \mathrm{L}^1(X_i; \gamma_i)$ (resp. $(F_i^\circ \circ \varphi_i)_+ \in \mathrm{L}^1(X_i, \mu_i)$) and

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geq \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i. \tag{4.18}$$

Remark 4.5. In a similar way, if $\boldsymbol{\psi} \in \mathrm{B}(X_1, \bar{\mathbb{R}}) \times \mathrm{B}(X_2, \bar{\mathbb{R}})$ with $\psi_i \leq F_i(0)$, $R_1^*(\psi_1) \oplus_o R_2^*(\psi_2) \leq \mathsf{c}$, and $(\psi_i)_- \in \mathrm{L}^1(X_i, \mu_i)$ (resp. $(R_i^* \circ \psi_i)_+ \in \mathrm{L}^1(X_i, \gamma_i)$), then $(R_i^* \circ \psi_i)_- \in \mathrm{L}^1(X_i, \gamma_i)$ (resp. $(\psi_i)_+ \in \mathrm{L}^1(X_i, \mu_i)$) with

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geq \sum_i \int_{X_i} \psi_i\,\mathrm{d}\mu_i. \qquad \square \tag{4.19}$$

Proof. Let us consider, e.g., the second inequality of (4.18) by assuming that $(F_i^\circ \circ \varphi_i)_- \in \mathrm{L}^1(X_i, \mu_i)$ (the calculations in the other cases are completely analogous). Applying Lemma 2.6 (with $\psi_i := F_i^\circ \circ \varphi_i$ and $\phi_i := -\varphi_i$) and (2.40) we obtain $(\varphi_i)_- \in \mathrm{L}^1(X_i, \gamma_i)$ and then

$$\begin{aligned}\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) &= \sum_i \mathscr{F}_i(\gamma_i|\mu_i) + \int_{\boldsymbol{X}} \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma} \geq \sum_i \mathscr{F}_i(\gamma_i|\mu_i) + \int_{\boldsymbol{X}} \Big(\varphi_1(x_1) +_o \varphi_2(x_2)\Big)\,\mathrm{d}\boldsymbol{\gamma} \\ &\geq \sum_i \mathscr{F}_i(\gamma_i|\mu_i) + \int_{X_i} \varphi_i\,\mathrm{d}\gamma_i \overset{(2.41)}{\geq} \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i,\end{aligned} \tag{4.20}$$

Notice that the semi-integrability of φ_i w.r.t. γ_i yields $\varphi_i(\pi^i(x_1, x_2)) > -\infty$ for $\boldsymbol{\gamma}$-a.e. $(x_1, x_2) \in \boldsymbol{X}$ so that $\varphi_1(x_1) +_o \varphi_2(x_2) = \varphi_1(x_1) + \varphi_2(x_2)$ and we can split the integral

$$+\infty > \int \Big(\sum_i \varphi_i(x_i)\Big)\,\mathrm{d}\boldsymbol{\gamma} = \sum_i \int \varphi_i(x_i)\,\mathrm{d}\boldsymbol{\gamma} = \sum_i \int \varphi_i(x_i)\,\mathrm{d}\gamma_i. \qquad \square$$

Optimality conditions. If there exists a couple $\boldsymbol{\varphi}$ as in Proposition 4.4 such that $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$ then all the above inequalities (4.20) should be identities so that we have

$$\mathscr{F}_i(\gamma_i|\mu_i) = \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i,$$

$$\int_{\boldsymbol{X}} \Big(\mathsf{c}(x_1, x_2) - (\varphi_1(x_1) +_o \varphi_2(x_2))\Big)\,\mathrm{d}\boldsymbol{\gamma} = 0,$$

and the second part of Lemma 2.6 yields

$$\varphi_1(x_1) +_o \varphi_2(x_2) = \mathsf{c}(x_1,x_2) \quad \boldsymbol{\gamma}\text{-a.e. in } \boldsymbol{X}, \tag{4.21a}$$
$$-\varphi_i \in \partial F_i(\sigma_i) \quad (\mu_i+\gamma_i)\text{-a.e. in } A_i \tag{4.21b}$$
$$\varphi_i = -(F_i)'_\infty \quad \gamma_i^\perp\text{-a.e. in } A_{\gamma_i}, \tag{4.21c}$$
$$F_i^\circ(\varphi_i) = F_i(0) \quad \mu_i^\perp\text{-a.e. in } A_{\mu_i}, \tag{4.21d}$$

where $(A_i, A_{\mu_i}, A_{\gamma_i})$ is a Borel partition related to the Lebesgue decomposition of the couple (γ_i,μ_i) as in Lemma 2.3. We will show now that the existence of a couple $\boldsymbol{\varphi}$ satisfying

$$\boldsymbol{\varphi} \in \mathrm{B}(X_1;\bar{\mathbb{R}}) \times \mathrm{B}(X_2;\bar{\mathbb{R}}), \quad \varphi_i \geq -(F_i)'_\infty, \quad \varphi_1 \oplus_o \varphi_2 \leq \mathsf{c}, \tag{4.22}$$

and the joint optimality conditions (4.21a,b,c,d) (without any integrability assumption) is also sufficient to prove that a feasible $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ is optimal.

Theorem 4.6. *Let $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ with $\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) < \infty$; if there exists a couple $\boldsymbol{\varphi}$ as in* (4.22) *which satisfies the joint optimality conditions* (4.21a,b,c,d) *then $\boldsymbol{\gamma}$ is optimal.*

Proof. We want to repeat the same calculations (4.20) of Proposition 4.4, taking care of the integrability issues. We use a clever truncation argument of [37], based on the maps

$$T_n : \mathbb{R} \to \mathbb{R}, \quad T_n(\varphi) := -n \vee \varphi \wedge n, \tag{4.23}$$

combined with a corresponding approximations of the entropies F_i given by

$$F_{i,n}(r) := \max_{|\phi|\leq n} \phi r - F_i^*(\phi). \tag{4.24}$$

Recalling (4.16), it is not difficult to check that if $\varphi_1 +_o \varphi_2 \geq 0$ we have $0 \leq T_n(\varphi_1) + T_n(\varphi_2) \uparrow \varphi_1 + \varphi_2$ as $n \uparrow \infty$, whereas $\varphi_1 +_o \varphi_2 \leq 0$ yields $0 \geq T_n(\varphi_1) + T_n(\varphi_2) \downarrow \varphi_1 + \varphi_2$. In particular if $\boldsymbol{\varphi}$ satisfies (4.22) then $T_n(\varphi_i) \in \mathrm{B}_b(X_i)$, $T_n(\varphi_i) \geq -(F_i)'_\infty$ since $(F_i)'_\infty \geq 0$ and $\varphi_i \geq -(F_i)'_\infty$, and $T_n(\varphi_1) \oplus T_n(\varphi_2) \leq \mathsf{c}$. The boundedness of $T_n(\varphi_i)$ and Proposition 4.4 yield for every $\tilde{\boldsymbol{\gamma}} \in \mathcal{M}(\boldsymbol{X})$

$$\mathscr{E}(\tilde{\gamma}|\mu_1,\mu_2) \geq \sum_i \int_{X_i} F_i^\circ(T_n(\varphi_i))\,\mathrm{d}\mu_i. \tag{4.25}$$

When $(F_i)'_\infty < \infty$, choosing $n \geq (F_i)'_\infty$ so that $T_n(\varphi_i) = \varphi_i = -(F_i)'_\infty$ $\gamma_i^\perp$-a.e., and applying (ii) of the next Lemma 4.7, we obtain

$$\begin{aligned}\int_{X_i} F_i^\circ(T_n(\varphi_i))\,\mathrm{d}\mu_i &\overset{(4.21b,d)}{=} \int_{X_i} \Big(F_{i,n}(\sigma_i) + \sigma_i T_n(\varphi_i)\Big)\,\mathrm{d}\mu_i \\ &\overset{(4.21c)}{=} \int_{X_i} F_{i,n}(\sigma_i)\,\mathrm{d}\mu_i + (F_i)'_\infty \gamma_i^\perp(X_i) + \int_{X_i} T_n(\varphi_i)\,\mathrm{d}\gamma_i,\end{aligned}$$

and the same inequality also holds when $(F_i)'_\infty = +\infty$ since in this case $\gamma_i^\perp = 0$. Summing up the two contributions we get

$$\mathscr{E}(\tilde{\gamma}|\mu_1,\mu_2) \geq \sum_i \Big(\int_{X_i} F_{i,n}(\sigma_i)\,\mathrm{d}\mu_i + (F_i)'_\infty \gamma_i^\perp(X_i)\Big) + \int_{\boldsymbol{X}} \Big(T_n(\varphi_1) \oplus T_n(\varphi_2)\Big)\,\mathrm{d}\boldsymbol{\gamma}.$$

Applying Lemma 4.7 (i) and the fact that $\varphi_1 \oplus_o \varphi_2 = \mathsf{c} \geq 0$ $\boldsymbol{\gamma}$-a.e. by (4.21a), we can pass to the limit as $n \uparrow \infty$ by monotone convergence in the right-hand side of the previous inequality, obtaining $\mathscr{E}(\tilde{\boldsymbol{\gamma}}|\mu_1,\mu_2) \geq \mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2)$. □

Lemma 4.7. *Let $F_{i,n} : [0, \infty) \to [0, \infty)$ be defined by (4.24). Then*

(i) *$F_{i,n}$ are Lipschitz, $F_{i,n}(s) \le F_i(s)$, $F_{i,n}(s) \uparrow F_i(s)$ as $n \uparrow +\infty$.*

(ii) *For every $s \in \mathrm{Dom}(F_i), \varphi_i \in \mathbb{R} \cup \{+\infty\}$*

$$\begin{aligned} -\varphi_i \in \partial F_i(s) \quad &\Rightarrow \quad -T_n(\varphi_i) \in \partial F_{i,n}(s), \\ \varphi_i = +\infty,\ s = 0 \quad &\Rightarrow \quad F_{i,n}(0) = F_i^\circ(T_n(\varphi_i)) = F_i^\circ(n). \end{aligned} \tag{4.26}$$

In particular, in both cases considered by (4.26) we have $F_{i,n}(s) = F_i^*(T_n(\varphi_i)) - sT_n(\varphi_i)$.

Proof. Property (i): by (2.23) we get $F_{i,n} \le F_i$; since $-F_i^*(0) = \inf F_i \ge 0$ we see that $F_{i,n}$ are nonnegative. Recalling that F_i^* are nondecreasing with $\mathrm{Dom}(F_i^*) \supset (-\infty, 0]$ (see Section 2.3) we also get the upper bound $F_{i,n}(s) \le ns - F_i^*(-n)$. Eventually, (4.24) defines $F_{i,n}$ as the maximum of a family of n-Lipschitz functions, so $F_{i,n}$ is n-Lipschitz.

Property (ii): notice that $F_{i,n} = \big(F_i^* + \mathrm{I}_{[-n,n]}\big)^*$ so that $(F_{i,n})^* = F_i^* + \mathrm{I}_{[-n,n]} \ge F_i^*$; it is not difficult to check that $F_i(s) = F_{i,n}(s)$ if and only if $\partial F_i(s) \cap [-n, n] \neq \emptyset$. Therefore the set $I_n := \{s \ge 0 : F_i(s) = F_{i,n}(s)\}$ is a nonempty closed interval (possibly reduced to a single point) and it is easy to see that denoting $s_n^+ := \max I_n$, $s_n^- := \min I_n$, $T_n'(s) := s_n^- \vee s \wedge s_n^+$, we have $F_{i,n}(s) = F_i(T_n'(s)) + n(s - T_n'(s))$. In particular, whenever $s \ge s_n^+$ we have $n \in \partial F_{i,n}(s)$ and similarly $-n \in \partial F_{i,n}(s)$ if $s \le s_n^-$. If s belongs to the interior of I_n, then $\partial F_i(s) = \partial F_{i,n}(s) \subset [-n, n]$.

Therefore, if $\phi_i = -\varphi_i \in \partial F_i(s)$ with $\phi_i \in [-n, n]$, we have $F_i(s) = \phi_i s - F_i^*(\phi_i) = F_{i,n}(s)$ so that $\phi_i \in \partial F_{i,n}(s)$. On the other hand, if $\partial F_i(s) \ni \phi_i > n$, s cannot belong to the interior of I_n, so that by monotonicity $s \ge s_n^+$ and $\partial F_{i,n}(s) \ni n = T_n(\phi_i) = -T_n(\varphi_i)$. The case when $\partial F_i(s) \ni \phi_i < -n$ is completely analogous.

Eventually, when $\phi_i = -\infty$ and $s = 0$ (in particular $F_i(0) = F_i^*(-\infty) < \infty$), (4.24) and the fact that F_i^* is nondecreasing yields $F_{i,n}(0) = -F_i^*(-n) = F_i^\circ(n) = F_i^\circ(T_n(\varphi_i))$. $\square$

4.3 A general duality result

The aim of this section is to show in complete generality the duality result $\mathsf{ET} = \mathsf{D}$, by using the ϕ-formulation of the dual problem (4.12), which is equivalent to (4.7) by Proposition 4.3.

We start with a simple Lemma depending on a specific feature of the Entropy functions (which fails exactly in the case of pure transport problems, see Example E.3 of Section 3.3), when

$$\Big(\mathrm{int}\,\big(m_1 \mathrm{Dom}(F_1)\big) \cap m_2 \mathrm{Dom}(F_2)\Big) \cup \Big(m_1 \mathrm{Dom}(F_1) \cap \mathrm{int}\,\big(m_2 \mathrm{Dom}(F_2)\big)\Big) \neq \emptyset. \tag{4.27}$$

Notice moreover that the couple $\varphi_i \equiv 0$ provides an obvious lower bound for $\mathsf{D}(\mu_1, \mu_2)$:

$$\mathsf{D}(\mu_1, \mu_2) \ge \mathscr{D}(0, 0|\mu_1, \mu_2) = \sum_i m_i F_i^\circ(0) = \sum_i m_i \inf F_i. \tag{4.28}$$

Lemma 4.8. *Let us suppose that* $\operatorname{int}\big(m_1\mathrm{Dom}(F_1)\big)\cap m_2\mathrm{Dom}(F_2)\neq\emptyset$*, so that*

$$\exists\, s_1^-, s_1^+ \in \mathrm{Dom}(F_1),\ s_2\in\mathrm{Dom}(F_2):\quad m_1 s_1^- < m_2 s_2 < m_1 s_1^+,\quad m_i=\mu_i(X_i), \tag{4.29}$$

and $S:=\sup\mathsf{c}<\infty$*. Then every couple* $\boldsymbol{\varphi}\in\boldsymbol{\Phi}$ *with* $\mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)\geq\sum_i m_i\inf F_i$ *satisfies*

$$\Phi_1^-\leq\sup\varphi_1\leq\Phi_1^+,\quad \Phi_1^\pm:=\frac{m_1(F_1(s_1^\pm)-\inf F_1)+m_2(F_2(s_2)-\inf F_2)+m_2s_2S}{m_2s_2-m_1s_1^\pm}. \tag{4.30}$$

Proof. (4.29), the dual bound $F_i^\circ(\varphi_i)\leq\varphi_i s_i+F_i(s_i)$ for every $s_i\in\mathrm{Dom}(F_i)$ following by (4.9), the fact that $\sup\varphi_1+\sup\varphi_2\leq S$, and the monotonicity of F° yield

$$\begin{aligned}\sum_i m_i\inf F_i\leq\mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)&\leq m_1F_1^\circ(\sup\varphi_1)+m_2F_2^\circ(S-\sup\varphi_1)\\&\leq(m_1s_1-m_2s_2)\sup\varphi_1+m_1F_1(s_1)+m_2F_2(s_2)+m_2s_2S.\end{aligned}$$

The choice $s_1:=s_1^-$ shows the upper bound in (4.30); the lower bound follows by choosing $s_1:=s_1^+$. □

The second Lemma is well known in the case of Optimal Transport problems and will provide a useful a priori estimate in the case of bounded cost functions.

Lemma 4.9. *If* $\sup\mathsf{c}=S<\infty$ *then for every couple* $\boldsymbol{\varphi}\in\boldsymbol{\Phi}$ *there exists* $\tilde{\boldsymbol{\varphi}}\in\boldsymbol{\Phi}$ *such that* $\mathscr{D}(\tilde{\boldsymbol{\varphi}}|\mu_1,\mu_2)\geq\mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)$ *and*

$$\sup\tilde\varphi_i-\inf\tilde\varphi_i\leq S,\quad 0\leq\sup\tilde\varphi_1+\sup\tilde\varphi_2\leq S. \tag{4.31}$$

If moreover (4.27) *holds, than there exist a constant* $\varphi_{\max}\geq0$ *only depending on* F_i,m_i,S *such that*

$$-\varphi_{\max}\leq\inf\tilde\varphi_i\leq\sup\tilde\varphi_i\leq\varphi_{\max}. \tag{4.32}$$

Proof. Since $\mathsf{c}\geq0$, possibly replacing φ_1 with $\tilde\varphi_1:=\varphi_1\vee(-\sup\varphi_2)$ we obtain a new couple $(\tilde\varphi_1,\varphi_2)$ with

$$\tilde\varphi_1\geq\varphi_1,\quad \tilde\varphi_1(x_1)+\varphi_2(x_2)\leq\big(\varphi_1(x_1)+\varphi_2(x_2)\big)\wedge0\leq\mathsf{c}(x_1,x_2)$$

so that $(\tilde\varphi_1,\varphi_2)\in\boldsymbol{\Phi}$ and $\mathscr{D}(\tilde\varphi_1,\varphi_2|\mu_1,\mu_2)\geq\mathscr{D}(\varphi_1,\varphi_2|\mu_1,\mu_2)$ since F_1° is nondecreasing. It is then not restrictive to assume that $\inf\varphi_1\geq-\sup\varphi_2$; a similar argument shows that we can assume $\inf\varphi_2\geq-\sup\varphi_1$. Since

$$\sup\varphi_1+\sup\varphi_2\leq S \tag{4.33}$$

we thus obtain a new couple $(\tilde\varphi_1,\tilde\varphi_2)\in\Sigma$ with

$$\mathscr{D}(\tilde\varphi_1,\tilde\varphi_2|\mu_1,\mu_2)\geq\mathscr{D}(\varphi_1,\varphi_2|\mu_1,\mu_2),\quad \sup\tilde\varphi_i-\inf\tilde\varphi_i\leq S. \tag{4.34}$$

If moreover $\sup\varphi_1+\sup\varphi_2=-\delta<0$, we could always add the constant δ to, e.g., φ_1, thus increasing the value of $\mathscr{D}$ still preserving the constraint $\boldsymbol{\Phi}$.

When (4.27) holds (e.g. in the case considered by (4.29)) the previous Lemma 4.8 provides constants $\varphi_1^\pm$ such that $\varphi_1^-\leq\sup\tilde\varphi_1\leq\varphi_1^+$. (4.31) shows that $\varphi_2^-\leq\sup\tilde\varphi_1\leq\varphi_2^+$ with $\varphi_2^-:=-\varphi_1^+$ and $\varphi_2^+:=S-\varphi_1^-$. Still applying (4.31) we obtain (4.32) with $\varphi_{\max}:=S+\varphi_1^+-\varphi_1^-$. □

Before stating the last lemma we recall the useful notion of c-transform of functions $\varphi_i : X_i \to \bar{\mathbb{R}}$ for a real valued cost $\mathsf{c} : \boldsymbol{X} \to [0,\infty)$:

$$\varphi_1^{\mathsf{c}}(x_2) := \inf_{x\in X_1} \mathsf{c}(x,x_2) - \varphi_1(x), \quad \varphi_2^{\mathsf{c}}(x_1) := \inf_{x\in X_2} \mathsf{c}(x_1,x) - \varphi_2(x). \tag{4.35}$$

It is well known that if $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$ with $\sup \varphi_i < \infty$ then

$$\varphi_i^{\mathsf{c}} \text{ are bounded}, \quad \varphi_1^{\mathsf{cc}} \oplus \varphi_1^{\mathsf{c}} \le \mathsf{c}, \quad \varphi_1^{\mathsf{cc}} \ge \varphi_1, \ \varphi_1^{\mathsf{c}} \ge \varphi_2. \tag{4.36}$$

Moreover, $\varphi_1 = \varphi_1^{\mathsf{cc}}$ if and only if $\varphi_1 = \varphi_2^{\mathsf{c}}$ for some function φ_2; in this case φ_1 is called c-concave and $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}})$ is couple of c-concave potentials.

Since F_i° are non decreasing, it is also clear that whenever $\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}$ are μ_i-measurable

$$\mathscr{D}((\varphi_1,\varphi_2)|\mu_1,\mu_2) \le \mathscr{D}((\varphi_1^{\mathsf{cc}},\varphi_2^{\mathsf{c}})|\mu_1,\mu_2) \quad \forall \boldsymbol{\varphi} \in \mathrm{B}(X_1)\times\mathrm{B}(X_2),\ \varphi_1\oplus\varphi_2 \le \mathsf{c}. \tag{4.37}$$

The next Lemma concerns the lower semicontinuity of φ_i^{c} in the case when c is simple [22], i.e. has the form

$$\mathsf{c} = \sum_{n=1}^N c_n \chi_{A_n^1\times A_n^2}, \quad A_n^i \text{ open in } X_i. \tag{4.38}$$

Lemma 4.10. *Let us assume that* c *has the form* (4.38) *and that* $\boldsymbol{\varphi} \in \mathrm{B}_s(X_1)\times\mathrm{B}_s(X_2)$ *is a couple of simple functions taking values in* $\mathrm{Dom}(F_1^\circ)\times\mathrm{Dom}(F_2^\circ)$ *and satisfying* $\varphi_1\oplus\varphi_2 \le \mathsf{c}$*. Then* $(\varphi_1^{\mathsf{cc}},\varphi_1^{\mathsf{c}}) \in \boldsymbol{\Phi}$ *with* $\mathscr{D}(\tilde{\boldsymbol{\varphi}}|\mu_1,\mu_2) \ge \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)$.

Proof. It is easy to check that $\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}$ are simple (since the infima in (4.35) are taken on a finite number of possible values); by (4.36) it is thus sufficient to check that they are lower semicontinuous functions.

We check the lower semicontinuity of φ_1^{c}, the argument for $\varphi_1^{\mathsf{cc}} = (\varphi_1^{\mathsf{c}})^{\mathsf{c}}$ is completely analogous. In fact, let us consider the sets

$$Z := \Big\{ \boldsymbol{z} = (z_n)_{n=1}^N \in \{0,1\}^N : \exists y \in X_1 \text{ such that } z_n = \chi_{A_n^1}(y), \text{ for every } n = 1,\dots,N \Big\},$$
$$Y_{\boldsymbol{z}} := \{ y \in X_1 : \chi_{A_n^1}(y) = z_n \text{ for every } n = 1,\dots N.\}$$

$(Y_{\boldsymbol{z}})_{\boldsymbol{z}\in Z}$ is a Borel partition of X_1; we define $\varphi_{\boldsymbol{z}} := \sup\{\varphi_1(y) : y \in Y_{\boldsymbol{z}}\}$.

By construction, for every $\boldsymbol{z} \in Z$ and $y \in Y_{\boldsymbol{z}}$ the map $f_{\boldsymbol{z}}(x) := \mathsf{c}(y,x) - \varphi_{\boldsymbol{z}}$ is independent of y in $Y_{\boldsymbol{z}}$ and it is lower semicontinuous w.r.t. $x \in X_2$ since c is lower semicontinuous. Since $\varphi_1^{\mathsf{c}}(x_2)$ can be represented as the infimum of a finite collection of lower semicontinuous functions

$$\varphi_1^{\mathsf{c}}(x_2) = \inf \big\{ f_{\boldsymbol{z}}(x_2) : \boldsymbol{z} \in Z \big\} \tag{4.39}$$

we get $\varphi_1^{\mathsf{c}} \in \mathrm{LSC}(X_1)$. □

Theorem 4.11. *In the basic coercive setting of Section 3.1, the Entropy-Transport functional* (3.4) *and the dual functional* (4.10) *satisfy*

$$\inf_{\boldsymbol{\gamma}\in\mathcal{M}(X_1\times X_2)} \mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) = \sup_{\boldsymbol{\varphi}\in\boldsymbol{\Phi}} \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2) \quad \textit{for every } \mu_i \in \mathcal{M}(X_i), \tag{4.40}$$

i.e. $\mathsf{ET}(\mu_1,\mu_2) = \mathsf{D}(\mu_1,\mu_2)$ *for every* $\mu_i \in \mathcal{M}(X_i)$.

Proof. It is not restrictive to assume that $\mathsf{D}(\mu_1, \mu_2)$ is finite. We proceed in various steps, considering first the case when c has compact sublevels.

Step 1: *the case of a cost* c *with compact sublevels.* We can directly apply Theorem 2.4 to the saddle functional $\mathscr{L}$ of (4.3) by choosing $A = \mathrm{M}$ given by (4.4) endowed with the narrow topology and $B = \boldsymbol{\Phi}$. (2.9) and (2.10) are clearly satisfied and the assumption $(F_1)'_\infty + (F_2)'_\infty + \min \mathsf{c} > 0$ shows that we can choose constant functions $\bar\psi_i$ with $-R^*(\bar\psi_i) = -\bar\varphi_i = \bar\phi_i \in [0, (F_i)'_\infty]$ such that

$$D = \min\Big(\mathsf{c} - (\varphi_1 \oplus \varphi_2)\Big) = \phi_1 + \phi_2 + \min \mathsf{c} > 0, \quad \bar\psi_i > -\infty.$$

Arguing as in the proof of Theorem 3.3, *(ii)*, we can immediately see that (2.12) is satisfied, since

$$\mathscr{L}(\boldsymbol{\gamma}, \bar{\boldsymbol{\psi}}) = \int_{\boldsymbol{X}} \Big(\mathsf{c} - \min \mathsf{c}\Big)\,\mathrm{d}\boldsymbol{\gamma} + D\boldsymbol{\gamma}(\boldsymbol{X}) + \sum_i \bar\psi_i \mu_i(X_i).$$

In fact, for C sufficiently big, the sublevels $\big\{\boldsymbol{\gamma} \in \mathrm{M} : \mathscr{L}(\boldsymbol{\gamma}, \bar{\boldsymbol{\psi}}) \le C\big\}$ are closed, bounded (since $D > 0$) and equally tight (by the assumption on the sublevels of c), thus narrowly compact.

In the next steps we will assume $(F_i)'_\infty = +\infty$ (so that F_i° are continuous and increasing on $\mathbb{R}$, and $F_i^\circ \circ \varphi_i \in \mathrm{LSC}_b(X_i)$ whenever $\varphi_i \in \mathrm{LSC}_b(X_i)$), and we will remove the compactness assumption on the sublevels of c.

Step 2: *the case when* μ_i *have compact support,* (4.27) *holds and the cost* c *has of the form*

$$\mathsf{c} = \sum_{n=1}^N c_n \chi_{A_n^1 \times A_n^2}, \quad c_n \ge 0,\ A_n^i \text{ open in } X_i. \tag{4.41}$$

Let us set $\tilde X_i := \mathrm{supp}(\mu_i)$. Since $(F_i)'_\infty = +\infty$ the support of all $\boldsymbol{\gamma}$ with $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$ is contained $\tilde X_1 \times \tilde X_2$ so that the minimum of the functional $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$ does not change by restricting the spaces to $\tilde X_i$. By applying the previous step to the problem stated in $\tilde X_1 \times \tilde X_2$, for every $E < \mathsf{ET}(\mu_1, \mu_2)$ we find $\boldsymbol{\varphi} \in \mathrm{LSC}_s(\tilde X_1) \times \mathrm{LSC}_s(\tilde X_2)$ such that $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$ in $\tilde X_1 \times \tilde X_2$, $F_i^\circ(\varphi_i)$ is finite, and $\sum_i \int_{\tilde X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i \ge E$.

Extending φ_i to $-\sup \mathsf{c}$ in $X_i \setminus \tilde X_i$ the value of $\mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$ does not change and we obtain a couple of simple Borel functions with $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$ in $\boldsymbol{X}$. We can eventually apply Lemma 4.10 to find $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}) \in \boldsymbol{\Phi}$ with $\mathscr{D}(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}|\mu_1, \mu_2) \ge E$.

Step 3: *we remove the assumption on the compactness of* $\mathrm{supp}(\mu_i)$.

Since μ_i are Radon, we can find two sequences of compact sets $K_{i,n} \subset X_i$ such that $\varepsilon_{i,n} := \mu_i(X_i \setminus K_{i,n}) \to 0$ as $n \to \infty$, so that $\mu_{i,n} := \chi_{K_{i,n}} \cdot \mu_i$ are narrowly converging to μ_i.

Let $E_n := \mathsf{ET}(\mu_{1,n}, \mu_{2,n})$ and let $E'_n < E_n$ with $\lim_{n\to\infty} E'_n = \liminf_{n\to\infty} E_n$. Since $\mu_{i,n}$ have compact support, by the previous step and Lemma 4.9 we can find a sequence $\boldsymbol{\varphi}_n \in \boldsymbol{\Phi}$ and a constant $\varphi_{\max}$ independent of n such that

$$\mathscr{D}(\boldsymbol{\varphi}_n|\mu_{1,n}, \mu_{2,n}) \ge E'_n, \quad \sup |\varphi_n^i| \le \varphi_{\max}.$$

This yields

$$\mathscr{D}(\boldsymbol{\varphi}_n|\mu_1, \mu_2) \ge \sum_i \int_{K_{i,n}} F_i^\circ(\varphi_{i,n})\,\mathrm{d}\mu_i + \sum_i F_i^\circ(-\varphi_{\max})\varepsilon_{i,n} \ge E'_n + \sum_i F_i^\circ(-\varphi_{\max})\varepsilon_{i,n}.$$

By Lemma 3.9 we get

$$\mathsf{ET}(\mu_1,\mu_2) \le \liminf_{n\to\infty} E_n = \lim_{n\to\infty} E'_n,$$

so that

$$\liminf_{n\to\infty} \mathscr{D}(\boldsymbol{\varphi}_n|\mu_1,\mu_2) \ge \liminf_{n\to\infty} E_n \ge \mathsf{ET}(\mu_1,\mu_2).$$

Step 4: *we remove the assumption* (4.27) *on* F_i. It is sufficient to approximate F_i by an increasing and pointwise converging sequence $F_i^n \in \Gamma(\mathbb{R}_+)$. The corresponding sequence $(F_i^n)^\star$ of conjugate concave functions is also nondecreasing and pointwise converging to F_i°. By the previous step, if $E_n < \mathscr{E}^n(\mu_1,\mu_2)$ with $\lim_{n\to\infty} E^n = \lim_{n\to\infty}\mathscr{E}^n(\mu_1,\mu_2) = \mathscr{E}(\mu_1,\mu_2)$ we can find $\boldsymbol{\varphi}_n \in \boldsymbol{\Phi}$ such that

$$E_n \le \sum_i \int_{X_i} (F_i^n)^\star(\varphi_i^n)\,\mathrm{d}\mu_i \le \sum_i \int_{X_i} F_i^\circ(\varphi_i^n)\,\mathrm{d}\mu_i.$$

Step 5: *the case of a general cost* c.

Let $\mathsf{c} : \boldsymbol{X} \to [0,\infty]$ be an arbitrary l.s.c. cost and let us denote by $(\mathsf{c}^\alpha)_{\alpha\in\mathbb{A}}$ the class of costs characterized by (4.38) and majorized by c; $\mathbb{A}$ is a directed set with the pointwise order $\le$, since join of a finite number of cost functions in $\mathbb{A}$ can still be expressed as in (4.38). It is not difficult to check that $\mathsf{c} = \sup_{\alpha\in\mathbb{A}} \mathsf{c}^\alpha = \lim_{\alpha\in\mathbb{A}} \mathsf{c}^\alpha$ so that by Lemma 3.9 $\mathsf{ET}(\mu_1,\mu_2) = \lim_{\alpha\in\mathbb{A}} \mathsf{ET}^\alpha(\mu_1,\mu_2) = \sup_{\alpha\in\mathbb{A}} \mathsf{ET}^\alpha(\mu_1,\mu_2)$, where ET^α denotes the Entropy-Transport functional associated to c^α.

Thus for every $E < \mathsf{ET}(\mu_1,\mu_2)$ we can find $\alpha \in \mathbb{A}$ such that $\mathsf{ET}^\alpha(\mu_1,\mu_2) > E$ and therefore, by the previous step, a couple $\boldsymbol{\varphi}^\alpha \in \mathrm{LSC}_s(X_1)\times \mathrm{LSC}_s(X_2)$ with $F_i^\circ(\varphi_i^\alpha)$ finite such that $\varphi_1^\alpha \oplus \varphi_2^\alpha \le \mathsf{c}^\alpha$ in $\boldsymbol{X}$ and $\mathscr{D}(\boldsymbol{\varphi}^\alpha|\mu_1,\mu_2) \ge E$. Since $\mathsf{c}^\alpha \le \mathsf{c}$ we deduce that $\boldsymbol{\varphi}^\alpha \in \boldsymbol{\Phi}$. □

Arguing as in Remark 2.8 we obtain

Corollary 4.12. *The duality formula* (4.40) *still holds if we replace the spaces of simple lower semicontinuous functions* $\mathrm{LSC}_s(X_i)$ *in the definition of* $\boldsymbol{\Phi}$ *with the spaces of bounded lower semicontinuous functions* $\mathrm{LSC}_b(X_i)$ *or with the spaces of bounded Borel functions* $\mathrm{B}_b(X_i)$.

If (X_i,τ_i) *are completely regular spaces, then we can equivalently replace lower semicontinuous functions by continuous ones, obtaining*

$$\begin{aligned}\mathsf{ET}(\mu_1,\mu_2) &= \sup\Big\{\sum_i \int_{X_i} F^\circ(\varphi_i)\,\mathrm{d}\mu_i : \varphi_i,\ F_i^\circ(\varphi_i)\in \mathrm{C}_b(X_i),\ \varphi_1\oplus\varphi_2 \le \mathsf{c}\Big\}\\ &= \sup\Big\{\sum_i \int_{X_i} \psi_i\,\mathrm{d}\mu_i : \psi_i, R_i^*(\psi_i)\in \mathrm{C}_b(X_i),\ R_1^*(\psi_1)\oplus R_2^*(\psi_2)\le \mathsf{c}\Big\}.\end{aligned} \tag{4.42}$$

Corollary 4.13 (Subadditivity of ET)**.** *The functional* ET *is convex and positively* 1*-homogeneous (in particular it is subadditive), i.e. for every* $\mu_i, \mu'_i \in \mathcal{M}(X)$ *and* $\lambda \ge 0$ *we have*

$$\mathsf{ET}(\lambda\mu_1,\lambda\mu_2) = \lambda\,\mathsf{ET}(\mu_1,\mu_2), \quad \mathsf{ET}(\mu_1+\mu'_1,\mu_2+\mu'_2) \le \mathsf{ET}(\mu_1,\mu_2) + \mathsf{ET}(\mu'_1,\mu'_2). \tag{4.43}$$

Proof. By Theorem 4.11 it is sufficient to prove the corresponding property of D, which follows immediately from its representation formula (4.8) as a supremum of linear functionals. □

4.4 Existence of optimal Entropy-Kantorovich potentials

In this section we will consider two cases, when the dual problem admits a couple of optimal Entropy-Kantorovich potentials.

The first case is completely analogous to the transport setting.

Theorem 4.14. *Let us suppose that (X_i, d_i) are complete metric space,* (4.27) *holds, and* c *is bounded and uniformly continuous with respect to the product distance* $\mathsf{d}((x_1, x_2), (x_1\, x_2)) := \sum_i \mathsf{d}_i(x_i, x_i')$ *in $\boldsymbol{X}$. Then there exists a couple of optimal Entropy-Kantorovich potentials $\boldsymbol{\varphi} \in \mathrm{C}_b(X_1) \times \mathrm{C}_b(X_2)$ satisfying*

$$\varphi_1 \oplus \varphi_2 \le \mathsf{c}, \quad \varphi_i \ge -(F_i)'_\infty, \quad \mathsf{ET}(\mu_1, \mu_2) = \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2). \tag{4.44}$$

Proof. By the boundedness and uniform continuity of c we can find a continuous and concave modulus of continuity $\omega : [0, +\infty) \to [0, +\infty)$ with $\omega(0) = 0$ such that

$$\big|\mathsf{c}(x_1', x_2) - \mathsf{c}(x_1, x_2)\big| \le \omega(\mathsf{d}_1(x_1', x_1)), \quad \big|\mathsf{c}(x_1, x_2') - \mathsf{c}(x_1, x_2)\big| \le \omega(\mathsf{d}_2(x_2', x_2)).$$

Possibly replacing the distances d_i with $\mathsf{d}_i + \omega(\mathsf{d}_i)$, we may assume that $x_1 \mapsto \mathsf{c}(x_1, x_2)$ is 1-Lipschitz w.r.t. d_1 for every $x_2 \in X_2$ and $x_2 \mapsto \mathsf{c}(x_1, x_2)$ is 1-Lipschitz with respect to d_2 for every $x_1 \in X_1$. In particular, every c-transform (4.35) of a bounded function is 1-Lipschitz (and in particular Borel).

Let $\boldsymbol{\varphi}_n$ be a maximizing sequence in Σ. By Lemma 4.9 we can assume that $\boldsymbol{\varphi}_n$ is uniformly bounded; by (4.36) and (4.37) we can also assume that $\boldsymbol{\varphi}_n$ are c-concave and thus 1-Lipschitz. If $K_{i,n}$ is a family of compact sets whose union A_i has a full $\mu_i + \gamma_i$ measure, we can thus extract a subsequence (still denote by $\boldsymbol{\varphi}_n$) pointwise convergent to $\boldsymbol{\varphi}$ in $A_1 \times A_2$. We define $\varphi_1 := \limsup_{n\to\infty} \varphi_{1,n}$, $\varphi_2 := \liminf_{n\to\infty} \varphi_{1,n}$, we obtain a family in $\mathrm{B}_b(X_1) \times \mathrm{B}_b(X_2)$ satisfying $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$, $\varphi_i \ge (F_i)'_\infty$ and

$$\mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2) = \sum_i \int_{A_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i \ge \lim_{n\to\infty} \sum_i \int_{A_i} F_i^\circ(\varphi_{i,n})\,\mathrm{d}\mu_i = \mathsf{ET}(\mu_1, \mu_2),$$

thanks to Fatou's Lemma and the fact that $F_i^*(\varphi_{i,n})$ are uniformly bounded from above. Eventually replacing (φ_1, φ_2) with $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}})$ we obtain a couple in $\mathrm{C}_b(X_1) \times \mathrm{C}_b(X_2)$ satisfying (4.44). □

The next result is of different type, since it does not require any boundedness nor regularity of c (which can also assume the value $+\infty$ in the case $F_i(0) < \infty$).

Theorem 4.15. *Let us suppose that at least one of the following two conditions hold:*

a) $F_i(0) < +\infty$

or

b) c *is everywhere finite and* (4.27) *holds.*

Then a plan $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ with finite energy $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$ is optimal if and only if there exists a couple $\boldsymbol{\varphi}$ as in (4.22) *satisfying the optimality conditions* (4.21a,b,c,d).

Proof. We already proved (Theorem 4.6) that the existence of a couple $\boldsymbol{\varphi}$ as in (4.22) satisfying (4.21a,b,c,d) yields the optimality of $\boldsymbol{\gamma}$.

Let us now assume that $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ has finite energy and is optimal. If $\mu_i \equiv \eta_0$ then also $\boldsymbol{\gamma} = 0$ and (4.21a,b,c,d) are always satisfied: we can choose $\varphi_i \equiv 0$.

We can therefore assume that at least one of the measures μ_i, say μ_2, has positive mass. Let $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1,\mu_2)$, and let us apply Theorem 4.11 to find a maximizing sequence $\boldsymbol{\varphi}_n \in \boldsymbol{\Phi}$ such that $\lim_{n\uparrow\infty} \mathscr{D}(\boldsymbol{\varphi}_n|\mu_1,\mu_2) = \mathsf{ET}(\mu_1,\mu_2)$.

Using the Borel partitions $(A_i, A_{\mu_i}, A_{\gamma_i})$ for the couples of measures γ_i, μ_i provided by Lemma 2.3 and arguing as in Proposition 4.4 we get

$$\lim_{n\to\infty} \int_{X_1\times X_2} \Big(\mathsf{c}(x_1,x_2) - \varphi_{1,n}(x_1) - \varphi_{2,n}(x_2)\Big)\,\mathrm{d}\boldsymbol{\gamma} = 0,$$
$$\lim_{n\to\infty} \int_{A_i\cup A_{\mu_i}} \Big(F_i(\sigma_i) + \sigma_i\varphi_{i,n} - F_i^\circ(\varphi_{i,n})\Big)\,\mathrm{d}\mu_i = 0,$$
$$\lim_{n\to\infty} \int_{A_{\gamma_i}} \big(\varphi_{i,n} + (F_i)'_\infty\big)\,\mathrm{d}\gamma_i^\perp = 0.$$

Since all the integrands are nonnegative, up to selecting a suitable subsequence (not relabeled) we can assume that the integrands are converging pointwise a.e. to 0. We can thus find Borel sets $A'_i \subset A_i, A'_{\mu_i} \subset A_{\mu_i}, A'_{\gamma_i} \subset A_{\gamma_i}$ and $A' \subset \boldsymbol{X}$ with $\pi^i(A') = A'_i \cup A'_{\gamma_i}$, $(\mu_i+\gamma_i)\Big((A_i\setminus A'_i)\cup(A_{\mu_i}\setminus A'_{\mu_i})\cup(A_{\gamma_i}\setminus A'_{\gamma_i})\Big) = 0$, and $\boldsymbol{\gamma}(\boldsymbol{X}\setminus A') = 0$ such that

$$\mathsf{c}(x_1,x_2) < \infty \quad \lim_{n\to\infty} \mathsf{c}(x_1,x_2) - \varphi_{1,n}(x_1) - \varphi_{2,n}(x_2) = 0 \quad \text{in } A', \tag{4.45}$$
$$F_i(\sigma_i) < \infty, \quad \lim_{n\to\infty} F_i(\sigma_i) + \sigma_i\varphi_{i,n} - F_i^\circ(\varphi_{i,n}) = 0 \quad \text{in } A'_i \cup A'_{\mu_i}, \tag{4.46}$$
$$\lim_{n\to\infty} \big(\varphi_{i,n} + (F_i)'_\infty\big) = 0 \quad \text{in } A'_{\gamma_i}. \tag{4.47}$$

For every $x_i \in X_i$ we define the Borel functions $\varphi_1(x_1) := \limsup_{n\to\infty} \varphi_{1,n}(x_1)$, $\varphi_2(x_2) := \liminf_{n\to\infty} \varphi_{2,n}(x_2)$, taking values in $\mathbb{R}\cup\{\pm\infty\}$. It is clear that the couple $\boldsymbol{\varphi} = (\varphi_1,\varphi_2)$ complies with (4.22), (4.21d) and (4.21c).

If $\boldsymbol{\gamma}(\boldsymbol{X}) = 0$ then (4.21a) and (4.21b) are trivially satisfied, so that it is not restrictive to assume $\boldsymbol{\gamma}(\boldsymbol{X}) > 0$.

If $\mu_1(X_1) = 0$ then $(F_1)'_\infty$ is finite (since $\gamma_1^\perp(X_1) = \gamma_1(X_1) = \boldsymbol{\gamma}(\boldsymbol{X}) > 0$) and $\varphi_1 \equiv (F_1)'_\infty$ on A'_{γ_1} and on A'. It follows that $\varphi_2(x_2) = \mathsf{c}(x_1,x_2) - (F_1)'_\infty \in \mathbb{R}$ on A' so that (4.21a) is satisfied. Since $\varphi_2(x_2)$ is an accumulation point of $\varphi_{2,n}(x_2)$ the next Lemma 4.19 yields $-\varphi_2(x_2) \in \partial F_2(\sigma_2(x_2))$ in A'_2 so that (4.21b) is also satisfied (in the case $i = 1$ one can choose $A'_1 = \emptyset$).

We can thus assume that $\mu_i(X_i) > 0$ and $\boldsymbol{\gamma}(\boldsymbol{X}) > 0$. In order to check (4.21a) and (4.21b) we distinguish two cases.

Case 1: c *is everywhere finite and* (4.27) *holds.* Let us first prove that $\varphi_1 < +\infty$ everywhere.

By contradiction, if there is a point $\bar{x}_1 \in X_1$ such that $\varphi_1(\bar{x}_1) = +\infty$ we deduce that $\varphi_2(x_2) = -\infty$ for every $x_2 \in X_2$.

Since the set $A'_2 \cup A'_{\mu_2}$ has positive μ_2-measure, it contains some point $\bar{x}_2$: (4.46) and the next Lemma 4.19 (with $F = F_2$, $s = \sigma_2(\bar{x}_2)$, $\phi_n := -\varphi_{2,n}(\bar{x}_2)$) yield $s_2^+ = \max \mathrm{Dom}(F_2) = \sigma_2(\bar{x}_2) < \infty$ and $\sigma_2 \equiv s_2^+$ in $A'_2 \cup A'_{\mu_2}$. We thus have $\mathrm{Dom}(F_2) \subset [0, s_2^+]$, $(F_2)'_\infty = +\infty$ and therefore $m_2 s_2^+ = \boldsymbol{\gamma}(\boldsymbol{X})$.

On the other hand, if $\varphi_2 = -\infty$ in X_2 we deduce that $\varphi_1(x_1) = +\infty$ for every $x_1 \in \pi^1(A')$. Since $(F_1)'_\infty \geq 0$, it follows that $\gamma_i(A'_{\gamma_i}) = 0$ (i.e. $\gamma_i^\perp = 0$) so that there is a point a_1 in A'_1 such that $\varphi_1(a_1) = +\infty$. Arguing as before, a further application of Lemma 4.19 yields that $\sigma_1 \equiv s_1^- = \min \mathrm{Dom}(F_1)$ μ_1-a.e. It follows that $m_1 s_1^- = \gamma_1(X_1) = \boldsymbol{\gamma}(\boldsymbol{X}) = m_2 s_2^+$, a situation that contradicts (4.27).

Since $\mu_1(X_1) > 0$ the same argument shows that $\varphi_2 < \infty$ everywhere in X_2. It follows that (4.21a) holds and $\varphi_i > -\infty$ on A'_i. Since $\varphi_i(x_i)$ is an accumulation point of $\varphi_{i,n}(x_i)$, the next Lemma 4.19 yields $-\varphi_i(x_i) \in \partial F_i(\sigma_i(x_i))$ in A'_i so that (4.21b) is also satisfied.

Case 2: $F_i(0) < \infty$. In this case F_i° are bounded from above and $\varphi_i \geq -(F_i)'_\infty$ everywhere in X_i. By Theorem 4.11 $\lim_{n\to\infty} \sum_i \int F_i^\circ(\varphi_{i,n})\,\mathrm{d}\mu_i > -\infty$, so that Fatou's Lemma yields $F_1^\circ(\varphi_1) \in \mathrm{L}^1(X_1, \mu_1)$ and $\varphi_1(x_1) > -\infty$ for μ_1-a.e. $x_1 \in X_1$, in particular for $(\mu_1 + \gamma_1)$-a.e. $x_1 \in A'_1$. Applying the next Lemma 4.19, since $\sigma_1(x_1) > 0 = \min \mathrm{Dom}(F_1)$ in A'_1, we deduce that $-\varphi_1(x_1) \in \partial F_1(\sigma_1(x_1))$ for $(\mu_1 + \gamma_1)$-a.e. $x_1 \in A'_1$, i.e. (4.21b) for $i = 1$. Since we already checked that (4.21c) and (4.21d) hold, applying Lemma 2.6 (with $\phi := -\varphi_1$ and $\psi := F_1^\circ(\varphi_1)$)) we get $\varphi_1 \in \mathrm{L}^1(X_1, \gamma_1)$, in particular $\varphi_1 \circ \pi^1 \in \mathbb{R}$ $\boldsymbol{\gamma}$-a.e. in $\boldsymbol{X}$. It follows that (4.21a) holds and $\varphi_2 \circ \pi^2 \in \mathrm{L}^1(\boldsymbol{X}, \boldsymbol{\gamma})$ so that $\varphi_2 \in \mathbb{R}$ $(\mu_2 + \gamma_2)$-a.e. in A'_2. A further application of Lemma 4.19 yields (4.21b) for $i = 2$. □

Corollary 4.16. *Let us suppose that* $\mathrm{Dom}(F_i) \supset (0, \infty)$ *and* F_i *are differentiable in* $(0, \infty)$*. A plan* $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ *with* $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$ *belongs to* $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ *if and only if there exist Borel partitions* $(A_i, A_{\mu_i}, A_{\gamma_i})$ *and corresponding Borel densities* σ_i *associated to* γ_i *and* μ_i *as in Lemma 2.3 such that setting*

$$\varphi_i(x_i) := \begin{cases} -F'_i(\sigma_i) & \text{if } x_i \in A_i, \\ -(F_i)'_0 & \text{if } x_i \in A_{\mu_i}, \\ -(F_i)'_\infty & \text{if } x_i \in X_i \setminus (A_i \cup A_{\mu_i}), \end{cases} \tag{4.48}$$

we have

$$\varphi_1 \oplus_o \varphi_2 \leq \mathsf{c} \text{ in } X_1 \times X_2, \quad \varphi_1 \oplus \varphi_2 = \mathsf{c}\ \boldsymbol{\gamma}\text{-a.e. in } (A_1 \cup A_{\gamma_1}) \times (A_2 \cup A_{\gamma_2}). \tag{4.49}$$

Proof. Since $\partial F_i(s) = \{F'_i(s)\}$ for every $s \in (0, \infty)$ and $F_i^\circ(\varphi_i) = F_i(0)$ if and only if $\varphi_i \in [-(F_i)'_0, +\infty]$, (4.49) is clearly a necessary condition for optimality, thanks to Theorem 4.15. Since $(F_i)'_0 \leq F'_i(s) \leq (F_i)'_\infty$ Theorem 4.6 shows that conditions (4.48)–(4.49) are also sufficient. □

The next result shows that (4.48)–(4.49) take an even simpler form when $-(F_i)'_0 = (F_i)'_\infty = +\infty$; in particular, by assuming that c is continuous, the support of an optimal plan $\boldsymbol{\gamma}$ cannot be too small.

Corollary 4.17 (Spread of the support)**.** *Let us suppose that*

- $\mathsf{c} : \boldsymbol{X} \to [0, \infty]$ *is continuous.*
- $\mathrm{Dom}(F_i) \supset (0, \infty)$*,* F_i *are differentiable in* $(0, \infty)$*, and* $-(F_i)'_0 = (F_i)'_\infty = \infty$*.*

$\boldsymbol{\gamma}$ *is an optimal plan if and only if* $\gamma_i \ll \mu_i$*, for every* $x_i \in \mathrm{supp}(\mu_i)$ *we have* $\mathsf{c}(x_1, x_2) = +\infty$ *if* $x_1 \in \mathrm{supp}\,\mu_1 \setminus \mathrm{supp}\,\gamma_1$ *or* $x_2 \in \mathrm{supp}\,\mu_2 \setminus \mathrm{supp}\,\gamma_2$*, and there exist Borel sets* $A_i \subset \mathrm{supp}\,\gamma_i$ *with* $\gamma_i(X_i \setminus A_i) = 0$ *and Borel densities* $\sigma_i : A_i \to (0, \infty)$ *of* γ_i *w.r.t.* μ_i *such that*

$$F'_1(\sigma_1) \oplus F'_2(\sigma_2) \geq -\mathsf{c} \text{ in } A_1 \times A_2, \quad F'_1(\sigma_1) \oplus F'_2(\sigma_2) = -\mathsf{c}\quad \boldsymbol{\gamma}\text{-a.e. in } A_1 \times A_2. \tag{4.50}$$

Remark 4.18. Apart from the case of pure transport problems (Example E.3 of Section 3.3), where the existence of Kantorovich potentials is well known (see [44, Thm. 5.10]), Theorem 4.15 covers essentially all the interesting cases, at least when the cost c takes finite values if $0 \notin \mathrm{Dom}(F_i)$. In fact, when (4.27) does not hold, it is not difficult to construct an example of optimal plan $\boldsymbol{\gamma}$ for which conditions (4.22), (4.21a), (4.21b) cannot be satisfied. Consider e.g. $X_i = \mathbb{R}$, $\mathsf{c}(x_1, x_2) := \frac{1}{2}|x_1 - x_2|^2$, $\mu_1 := \mathrm{e}^{-\sqrt{\pi}x_1^2}\mathscr{L}^1$, $\mu_2 := \mathrm{e}^{-\sqrt{\pi}(x_2+1)^2}\mathscr{L}^1$, $\mathrm{Dom}(F)_1 = [a, 1]$, $\mathrm{Dom}(F)_2 = [1, b]$ with arbitrary choice of $a \in [0, 1)$ and $b \in (1, \infty]$. Since $m_1 = m_2 = 1$ (3.18) holds and yields $\gamma_i = \mu_i$, $\sigma_i \equiv 1$, so that the optimal plan $\boldsymbol{\gamma}$ can be obtained by solving the quadratic optimal transportation problem, thus $\boldsymbol{\gamma} := \boldsymbol{t}_\sharp\mu_1$ where $\boldsymbol{t}(x) := (x, x-1)$. In this case the potentials φ_i are uniquely determined up to an additive constant $a \in \mathbb{R}$ so that we have $\varphi_1(x_1) = x_1 + a$, $\varphi_2(x_2) = -x_2 - a - \frac{1}{2}$, and it is clear that condition $-\varphi_i \in \partial F_i(1)$ corresponding to (4.21b) cannot be satisfied, since $\partial F_i(1)$ are always proper subsets of $\mathbb{R}$. We can also construct entropies such that $\partial F_i(1) = \emptyset$ (e.g. $F_1(r) = (1-r)\log(1-r) + r$, $F_2(r) = (r-1)\log(r-1) - r + 2$) so that (4.21b) can never hold, independently of the cost c. □

We conclude this section by proving the simple property on subdifferentials we used in the proof of Theorem 4.15.

Lemma 4.19. *Let $F \in \Gamma(\mathbb{R}_+)$, $s \in \mathrm{Dom}(F)$, let $\phi \in \mathbb{R} \cup \{\pm\infty\}$ be an accumulation point of a sequence $(\phi_n) \subset \mathbb{R}$ satisfying*

$$\lim_{n\to\infty} F(s) - s\phi_n + F^*(\phi_n) = 0. \tag{4.51}$$

If $\phi \in \mathbb{R}$ then $\phi \in \partial F(s)$, if $\phi = +\infty$ then $s = \max \mathrm{Dom}(F)$ and if $\phi = -\infty$ then $s = \min \mathrm{Dom}(F)$. In particular, whenever $s \in \mathrm{int}(\mathrm{Dom}(F))$ ϕ is finite.

Proof. Up to extracting a suitable subsequence, it is not restrictive to assume that ϕ is the limit of ϕ_n as $n \to \infty$. For every $w \in \mathrm{Dom}(F)$ the Young inequality $w\phi_n \le F(w) + F^*(\phi_n)$ yields

$$\limsup_{n\to\infty}(w-s)\phi_n \le \limsup_{n\to\infty} F(w) - F(s) + \Big(F(s) - s\phi_n + F^*(\phi_n)\Big) = F(w) - F(s) \tag{4.52}$$

If $\mathrm{Dom}(F) = \{s\}$ then $\partial F(s) = \mathbb{R}$ and there is nothing to prove; let thus assume that $\mathrm{Dom}(F)$ has nonempty interior.

If $\phi \in \mathbb{R}$ then $(w-s)\phi \le F(w) - F(s)$ for every $w \in \mathrm{Dom}(F)$, so that $\phi \in \partial F(s)$. Since the righthand side of (4.52) is finite for every $w \in \mathrm{Dom}(F)$, if $\phi = +\infty$ then $w \le s$ for every $w \in \mathrm{Dom}(F)$, so that $s = \max \mathrm{Dom}(F)$. An analogous argument holds when $\phi = -\infty$. □

5 "Homogeneous" formulations of optimal Entropy-Transport problems

Starting from the reverse formulation of the Entropy-Transport problem of Section 3.5 via the $\mathscr{R}$ functional (3.31), in this section we will derive further equivalent representations of the ET functional, which will also reveal new interesting properties, in particular when we will apply these results to the logarithmic Hellinger-Kantorovich functional. The advantage of the reverse formulation is that it always admits a "1-homogeneous" representation, associated to a modified cost functional that can be explicitly computed in terms of to R_i and c.

We will always tacitly assume the basic *coercive* setting of Section 3.1.

5.1 The homogeneous marginal perspective functional.

First of all we introduce the marginal perspective function H_c depending on the parameter $c \geq \inf \mathsf{c}$:

Definition 5.1 (Marginal perspective function and cost)**.** *Let $c \in [0, \infty]$; the marginal perspective function $H_c : [0, \infty) \times [0, \infty) \to [0, +\infty]$ is defined as the lower semicontinuous envelope of*

$$\tilde{H}_c(r_1, r_2) := \inf_{\theta > 0} \theta \big(R_1(r_1/\theta) + R_2(r_2/\theta) + c \big) = \inf_{\theta > 0} r_1 F_1(\theta/r_1) + r_2 F_2(\theta/r_2) + \theta c, \quad (5.1)$$

when $c < \infty$; when $c = \infty$ we set

$$H_\infty(r_1, r_2) := F_1(0) r_1 + F_2(0) r_2. \quad (5.2)$$

The induced marginal perspective cost is $H : (X_1 \times \mathbb{R}_+) \times (X_2 \times \mathbb{R}_+) \to [0, +\infty]$

$$H(x_1, r_1; x_2, r_2) := H_{\mathsf{c}(x_1, x_2)}(r_1, r_2), \quad x_i \in X_i, \ r_i \geq 0. \quad (5.3)$$

The last formula (5.2) is justified by the property $F_i(0) = (R_i)'_\infty$ and the fact that $H_c(r_1, r_2) \uparrow H_\infty(r_1, r_2)$ as $c \uparrow \infty$ for every $r_1, r_2 \in [0, \infty)$, see also the next Lemma 5.3.

Example 5.2. Let us consider the symmetric cases associated to the entropies U_p:

E.1 In the "logarithmic entropy case", which we will extensively study in Part II, we have

$$F_i(s) := U_1(s) = s \log s - (s - 1) \quad \text{and} \quad R_i(r) = U_0(r) = r - 1 - \log r.$$

A direct computation shows

$$\begin{aligned} \tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) &= r_1 + r_2 - 2\sqrt{r_1 r_2}\, \mathrm{e}^{-c/2} \\ &= \big(\sqrt{r_1} - \sqrt{r_2}\big)^2 + 2\sqrt{r_1 r_2}\, \big(1 - \mathrm{e}^{-c/2}\big). \end{aligned} \quad (5.4)$$

E.2 For $p = 0$, $F_i(s) = U_0(s) = s - \log s - 1$, and $R_i(r) = U_1(r)$ we obtain

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = r_1 \log r_1 + r_2 \log r_2 - (r_1 + r_2) \log \Big(\frac{r_1 + r_2}{2 + c} \Big). \quad (5.5)$$

E.3 In the power-like case with $p \in \mathbb{R} \setminus \{0, 1\}$ we start from

$$F_i(s) := U_p(s) = \frac{1}{p(p-1)} \big(s^p - p(s-1) - 1 \big), \quad R_i(r) = U_{1-p}(r)$$

and obtain, for $r_1, r_2 > 0$,

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = \frac{1}{p}\Big[\big(r_1 + r_2\big) - \frac{r_1\, r_2}{(r_1^{p-1} + r_2^{p-1})^{1/(p-1)}}\Big(2 - (p-1)c\Big)_+^q\Big], \quad (5.6)$$

where $q = p/(p-1)$. In fact

$$\theta\big(U_{1-p}(\tfrac{r_1}{\theta}) + U_{1-p}(\tfrac{r_2}{\theta}) + c\big) = \frac{r_1^{1-p} + r_2^{1-p}}{p(p-1)}\theta^p + \frac{1}{p}(r_1 + r_2) + \frac{1}{p-1}((p-1)c - 2)\theta)$$
$$= \frac{1}{p}(r_1 + r_2) + \frac{1}{p-1}\Big[\frac{1}{p}\Big((r_1^{1-p} + r_2^{1-p})^{1/p}\,\theta\Big)^p - \big(2 - (p-1)c\big)\theta\Big]$$

and (5.6) follows by minimizing w.r.t. θ. E.g. when $p = q = 2$

$$H_c(r_1, r_2) = \frac{1}{2}\big(r_1 + r_2\big) - \frac{1}{2}\frac{r_1 r_2}{r_1 + r_2}(2 - c)_+^2 = \frac{1}{2(r_1 + r_2)}\Big((r_1 - r_2)^2 + h(c) r_1 r_2\Big), \quad (5.7)$$

where $h(c) = c(4 - c)$ if $0 \le c \le 2$ and 4 if $c \ge 2$. For $p = -1$ and $q = 1/2$ equation (5.6) yields

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = \sqrt{(r_1^2 + r_2^2)(2 + 2c)} - \big(r_1 + r_2\big). \quad (5.8)$$

The following dual characterization of H_c nicely explains the crucial role of H_c.

Lemma 5.3 (Dual characterization of H_c)**.** *For every $c \ge 0$ the function H_c admits the dual representation*

$$H_c(r_1, r_2) = \sup\Big\{r_1\psi_1 + r_2\psi_2 : \psi_i \in \mathrm{Dom}(R_i^*),\ R_1^*(\psi_1) + R_2^*(\psi_2) \le c\Big\} \quad (5.9)$$
$$= \sup\Big\{r_1 F_1^\circ(\phi_1) + r_2 F_2^\circ(\phi_2) : \phi_i \in \mathrm{Dom}(F_i^\circ),\ \phi_1 + \phi_2 \le c\Big\}. \quad (5.10)$$

In particular it is lower semicontinuous, convex and positively 1-homogeneous (thus sublinear) with respect to (r_1, r_2), nondecreasing and concave w.r.t. c, and satisfies

$$H_c(r_1, r_2) \le H_\infty(r_1, r_2) = \sum_i F_i(0) r_i \quad \textit{for every } c \ge 0,\ r_i \ge 0. \quad (5.11)$$

Moreover,

a) the function H_c coincides with $\tilde{H}_c$ in the interior of its domain; in particular, if $F_i(0) < \infty$ then $H_c(r_1, r_2) = \tilde{H}_c(s_1, r_2)$ whenever $r_1 r_2 > 0$.

b) if $(F_1)'_\infty + (F_2)'_0 + c \ge 0$, $(F_2)'_\infty + (F_1)'_0 + c \ge 0$ then

$$H_c(r_1, r_2) = \sum_i F_i(0) r_i \quad \textit{if } r_1 r_2 = 0. \quad (5.12)$$

Proof. Since $\sup \mathrm{Dom}(R_i^*) = F_i(0)$ by (2.33), one immediately gets (5.9) in the case $c = +\infty$; we can thus assume $c < +\infty$.

It is not difficult to check that the function $(r_1, r_2, \theta) \mapsto \theta\big(R_1(r_1/\theta) + R_2(r_2/\theta) + c\big)$ is jointly convex in $[0,\infty)\times[0,\infty)\times(0,\infty)$ so that $\tilde{H}_c$ is a convex and positive 1-homogeneous function. It is also proper (i.e. it is not identically $+\infty$) thanks to (3.1). By Legendre duality [35, Thm.12.2], its lower semicontinuous envelope is given by

$$H_c(r_1, r_2) = \sup\Big\{\sum_i \psi_i r_i, \quad H_c^*(\psi_1, \psi_2) \le 0\Big\}, \tag{5.13}$$

where

$$\begin{aligned} H_c^*(\psi_1, \psi_2) &= \sup\Big\{\sum_i \psi_i r_i - \tilde{H}_c(r_1, r_2) : r_i \ge 0\Big\} = \sup_{r_i \ge 0, \theta > 0} \sum_i \Big(\psi_i r_i - \theta R_i(r_i/\theta)\Big) - c\theta \\ &= \sup_{\theta > 0} \theta\Big(\sum_i R_i^*(\psi_i) - c\Big) = \begin{cases} 0 & \text{if } R_i^*(\psi_i) < \infty, \quad \sum_i R_i^*(\psi_i) \le c \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

In order to prove point a) it is sufficient to recall that convex functions are always continuous in the interior of their domain [35, Thm. 10.1]. In particular, since for every $r_1, r_2 > 0$ $\lim_{\theta \downarrow 0} \theta\big(R_1(r_1/\theta) + R_2(r_2/\theta) + c\big) = \sum_i (R_i)'_\infty r_i = \sum_i F_i(0) r_i$ we have $\tilde{H}_c(r_1, r_2) \le \sum_i F_i(0) r_i$ so that $\tilde{H}_c$ is always finite if $F_i(0) < \infty$.

Concerning b), it is obvious when $r_1 = r_2 = 0$. When $r_1 > r_2 = 0$, the facts that $\sup \mathrm{Dom}(R_i^*) = F_i(0)$, $\lim_{r \uparrow F_i(0)} R_i^*(r) = -(F_i)'_0$ and $\inf R_i^* = -(F_i)'_\infty$ (see (2.33)) yield

$$H_c(r_1, 0) = \sup\Big\{\psi_1 r_1 : R_1^*(\psi_1) \le c - \inf R_2^*\Big\} = F_1(0) r_1;$$

an analogous formula holds when $0 = r_1 < r_2$. □

A simple consequence of Lemma 5.3 and (2.32) is the following lower bound

$$\tilde{H}_c(r_1, r_2) \ge H_c(r_1, r_2) \ge \sum_i \psi_i r_i \quad (-\varphi_i, \psi_i) \in \mathfrak{F}_i, \quad \varphi_1 + \varphi_2 \le c. \tag{5.14}$$

We can now introduce the integral functional associated to marginal perspective cost (5.3):

$$\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) := \int_{\boldsymbol{X}} H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2))\,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i^\perp(X_i), \quad \mu_i = \varrho_i \gamma_i + \mu_i^\perp, \tag{5.15}$$

where we adopted the same notation of (3.30). Let us first show that $\mathscr{H}$ is always greater than $\mathscr{D}$.

Lemma 5.4. *For every $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$, $\mu_i, \mu_i' \in \mathcal{M}(X_i)$, $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$, $\varrho_i \in \mathrm{L}^1_+(X_i, \gamma_i)$ with $\mu_i = \varrho_i \gamma_i + \mu_i'$, we have*

$$\int_{\boldsymbol{X}} H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2))\,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i'(X_i) \ge \mathscr{D}(\boldsymbol{\varphi} | \mu_1, \mu_2). \tag{5.16}$$

Proof. We get (recalling that $F_i^\circ(\varphi_i) = -F^*(-\varphi_i) \ge F_i(0)$)

$$
\begin{aligned}
&\int_{\boldsymbol{X}} H(x_1,\varrho_1(x_1);x_2,\varrho_2(x_2))\,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0)\mu_i'(X_i)\\
&\overset{(5.14)}{\ge} \int_{\boldsymbol{X}} \Big(F_1^\circ(\varphi_1(x_1))\varrho_1(x_1) + F_2^\circ(\varphi_2(x_2))\varrho_2(x_2)\Big)\,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0)\mu_i'(X_i)\\
&= \sum_i \int_{X_i} F_i^\circ(\varphi_i)\varrho_i(x_i)\,\mathrm{d}\gamma_i + \sum_i F_i(0)\mu_i'(X_i)\\
&\overset{(2.20)}{\ge} \sum_i \int_{X_i} F_i^\circ(\varphi_i)\varrho_i(x_i)\,\mathrm{d}\gamma_i + \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i' = \sum_i \int_{X_i} F_i^\circ(\varphi_i)\,\mathrm{d}\mu_i = \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2). \qquad \square
\end{aligned}
$$

An immediate consequence of the previous Lemma is the following important result concerning the marginal perspective cost functional $\mathscr{H}$ defined by (5.15).

Theorem 5.5. *For every $\mu_i \in \mathcal{M}(X_i)$, $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ and $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$ we have*

$$
\mathscr{R}(\mu_1,\mu_2|\boldsymbol{\gamma}) \ge \mathscr{H}(\mu_1,\mu_2|\boldsymbol{\gamma}) \ge \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2). \tag{5.17}
$$

In particular

$$
\mathsf{ET}(\mu_1,\mu_2) = \min_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \mathscr{H}(\mu_1,\mu_2|\boldsymbol{\gamma}), \tag{5.18}
$$

and $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1,\mu_2)$ if and only if it minimizes $\mathscr{H}(\mu_1,\mu_2|\cdot)$ in $\mathcal{M}(\boldsymbol{X})$ and satisfies

$$
H(x_1,\varrho_1(x_1);x_2,\varrho_2(x_2)) = \sum_i R_i(\varrho_i(x_i)) + \mathsf{c}(x_1,x_2) \quad \boldsymbol{\gamma}\text{-a.e. in } \boldsymbol{X}. \tag{5.19}
$$

If moreover the following condition

$$
F_i(0) = +\infty \text{ or there exists } \bar{x}_i \in X_i \text{ with } \mu_i(\{\bar{x}_i\}) = 0 \quad \text{for each } i = 1,2 \tag{5.20}
$$

is satisfied, then

$$
\mathsf{ET}(\mu_1,\mu_2) = \min\Big\{ \int_{\boldsymbol{X}} H(x_1,\varrho_1(x_1);x_2,\varrho_2(x_2))\,\mathrm{d}\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X}),\ \mu_i = \varrho_i\gamma_i \Big\}. \tag{5.21}
$$

Proof. The inequality $\mathscr{R}(\mu_1,\mu_2|\boldsymbol{\gamma}) \ge \mathscr{H}(\mu_1,\mu_2|\boldsymbol{\gamma})$ is an immediate consequence of the fact that $\sum_i R_i(r_1,r_2) + c \ge \tilde{H}_c(r_1,r_2) \ge H_c(r_1,r_2)$ for every $r_i, c \in [0,\infty]$, obtained by choosing $\theta = 1$ in (5.1). The inequality $\mathscr{H}(\mu_1,\mu_2|\boldsymbol{\gamma}) \ge \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)$ is provided by Lemma 5.4.

By using the "reverse" formulation of $\mathsf{ET}(\mu_1,\mu_2)$ in terms of the functional $\mathscr{R}(\mu_1,\mu_2|\boldsymbol{\gamma})$ given by Theorem 3.11 and applying Theorem 4.11 we obtain (5.18) and the characterization (5.19).

Finally, the identity (5.21) is obvious when $F_i(0) = +\infty$. In the general case, one immediately see that the righthand side E' of (5.21) (with "inf" instead of "min") is larger than $\mathsf{ET}(\mu_1,\mu_2)$, since the infimum of $\mathscr{H}(\mu_1,\mu_2|\cdot)$ is constrained to the smaller set of plans $\boldsymbol{\gamma}$ satisfying $\mu_i \ll \gamma_i$. On the other hand, if $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1,\mu_2)$ with $\mu_i = \varrho_i\bar{\gamma}_i + \mu_i^\perp$

and $\tilde m_i := \mu_i^\perp(X_i) > 0$, we can consider $\boldsymbol\gamma := \bar{\boldsymbol\gamma} + \frac{1}{\tilde m_1 \tilde m_2}\mu_1^\perp \otimes \mu_2^\perp$ which satisfies $\mu_i \ll \gamma_i$; by exploiting the fact that $H(x_1, r_1; x_2, r_2) \le \sum_i F_i(0) r_i$ by (5.11), we obtain

$$\begin{aligned}\mathscr{H}(\mu_1, \mu_2|\boldsymbol\gamma) &= \int_{\boldsymbol X} H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2))\,\mathrm{d}\bar{\boldsymbol\gamma} + \frac{1}{\tilde m_1 \tilde m_2}\int_{\boldsymbol X} H(x_1, \tilde m_1; x_2, \tilde m_2)\,\mathrm{d}\mu_1^\perp \otimes \mu_2^\perp \\ &\le \int_{\boldsymbol X} H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2))\,\mathrm{d}\bar{\boldsymbol\gamma} + \sum_i F_i(0)\tilde m_i = \mathscr{H}(\mu_1, \mu_2|\bar{\boldsymbol\gamma}),\end{aligned}$$

so that we have $E' \le \mathsf{ET}(\mu_1, \mu_2)$. The case when only one (say $\mu_2^\perp$) of the measures $\mu_i^\perp$ vanishes can be treated in the same way, by applying (5.20) and choosing $\boldsymbol\gamma := \bar{\boldsymbol\gamma} + \frac{1}{\tilde m_1}\mu_1^\perp \otimes \delta_{\bar x_2}$. □

Remark 5.6. Notice that (5.20) is always satisfied if the spaces X_i are uncountable. If $F_i(0) < \infty$ and X_i is countable, one can always add an isolated point $\bar x_i$ (sometimes called "cemetery") to X_i and to consider the augmented space $\bar X_i = X_i \sqcup \{\bar x_i\}$ obtained as the disjoint union of X and $\bar x_i$, with augmented cost $\bar{\mathsf c}$ which extends $\mathsf c$ to $+\infty$ on $\bar X_1 \times \bar X_2 \setminus (X_1 \times X_2)$. We can recover (5.21) by allowing $\boldsymbol\gamma$ in $\mathcal M(\bar X_1 \times \bar X_2)$. □

5.2 Entropy-transport problems with "homogeneous" marginal constraints

In this section we will exploit the 1-homogeneity of the marginal perspective function $\mathscr H$ in order to derive a last representation of the functional ET, related to the new notion of *homogeneous marginals*. We will confine our presentation to the basic, still relevant, facts, and we will devote the second part of the paper to develop a full theory for the specific case of the Logarithmic Entropy-transport case.

Homogeneous marginals. In the usual setting of Section 3.1, we consider the product spaces $Y_i := X_i \times [0, \infty)$ endowed with the product topology; we will denote by $y_i = (x_i, r_i)$, $x_i \in X_i$, $r_i \in [0, \infty)$, the generic points in Y_i, $i = 1, 2$. Projections from $\boldsymbol Y := Y_1 \times Y_2$ onto the various coordinates will be denoted by π^{y_i}, π^{x_i}, π^{r_i} with obvious meaning.

For $p > 0$ and $\boldsymbol y \in \boldsymbol Y$ we will set $|\boldsymbol y|_p^p := \sum_i |r_i|^p$ and we call $\mathcal M_p(\boldsymbol Y)$ (resp. $\mathcal P_p(\boldsymbol Y)$) the space of measures $\boldsymbol\alpha \in \mathcal M(\boldsymbol Y)$ (resp. $\mathcal P(\boldsymbol Y)$) such that

$$\int_{\boldsymbol Y} |\boldsymbol y|_p^p\,\mathrm{d}\boldsymbol\alpha < \infty. \tag{5.22}$$

If $\boldsymbol\alpha \in \mathcal M_p(\boldsymbol Y)$ the measures $r_i^p\boldsymbol\alpha$ belong to $\mathcal M(\boldsymbol Y)$: the "$p$-homogeneous" marginal $\mathrm h_i^p(\boldsymbol\alpha)$ of $\boldsymbol\alpha \in \mathcal M_p(\boldsymbol Y)$ is defined as the x_i-marginal of $r_i^p\boldsymbol\alpha$:

$$\mathrm h_i^p(\boldsymbol\alpha) := \pi_\sharp^{x_i}(r_i^p\boldsymbol\alpha) \in \mathcal M(X_i). \tag{5.23}$$

The maps $\mathrm h_i^p : \mathcal M_p(\boldsymbol Y) \to \mathcal M(X_i)$ are linear and invariant with respect to dilations: if $\vartheta : \boldsymbol Y \to (0, \infty)$ is a Borel map in $\mathrm L^p(\boldsymbol Y, \boldsymbol\alpha)$ and $\mathrm{prd}_\vartheta(\boldsymbol y) := (x_1, r_1/\vartheta(\boldsymbol y); x_2, r_2/\vartheta(\boldsymbol y))$, we set

$$\begin{gathered}\mathrm{dil}_{\vartheta,p}(\boldsymbol\alpha) := \big(\mathrm{prd}_\vartheta\big)_\sharp\big(\vartheta^p\boldsymbol\alpha\big) \quad \text{i.e.} \\ \int \varphi(\boldsymbol y)\,\mathrm{d}(\mathrm{dil}_{\vartheta,p}(\boldsymbol\alpha)) = \int \varphi(x_1, r_1/\vartheta; x_2, r_2/\vartheta)\vartheta^p(\boldsymbol y)\,\mathrm{d}\boldsymbol\alpha(\boldsymbol y) \quad \text{for every } \varphi \in \mathrm B_b(\boldsymbol Y),\end{gathered} \tag{5.24}$$

so that

$$\mathrm{h}_i^p(\mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha})) = \mathrm{h}_i^p(\boldsymbol{\alpha}). \tag{5.25}$$

In particular, if $\boldsymbol{\alpha} \in \mathcal{M}_p(\boldsymbol{Y})$ with $\boldsymbol{\alpha}(\boldsymbol{Y}) > 0$, by choosing

$$\vartheta(\boldsymbol{y}) := r^{-1} \begin{cases} |\boldsymbol{y}|_p & \text{if } |\boldsymbol{y}|_p \neq 0, \\ 1 & \text{if } |\boldsymbol{y}|_p = 0, \end{cases} \qquad r^p := \int_{\boldsymbol{Y}} |\boldsymbol{y}|_p^p \,\mathrm{d}\boldsymbol{\alpha} + \boldsymbol{\alpha}(\{|\boldsymbol{y}| = 0\}) \tag{5.26a}$$

we obtain a rescaled probability measure $\tilde{\boldsymbol{\alpha}}$ with the same homogeneous marginals as $\boldsymbol{\alpha}$ and concentrated on $\boldsymbol{Y}_{r,p} := \big\{\boldsymbol{y} \in \boldsymbol{Y} : |\boldsymbol{y}|_p \le r\big\} \subset (X \times [0, r]) \times (X \times [0, r])$:

$$\tilde{\boldsymbol{\alpha}} = \mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha}) \in \mathcal{P}_p(\boldsymbol{Y}), \quad \mathrm{h}_i^p(\tilde{\boldsymbol{\alpha}}) = \mathrm{h}_i^p(\boldsymbol{\alpha}), \quad \tilde{\boldsymbol{\alpha}}\big(\boldsymbol{Y} \setminus \boldsymbol{Y}_{r,p}\big) = 0. \tag{5.26b}$$

Entropy-transport problems with prescribed homogeneous marginals. Given $\mu_1, \mu_2 \in \mathcal{M}(X)$ we now introduce the convex sets

$$\begin{aligned} \mathcal{H}^p_{\le}(\mu_1, \mu_2) &:= \Big\{\boldsymbol{\alpha} \in \mathcal{M}_p(\boldsymbol{Y}) : \mathrm{h}_i^p(\boldsymbol{\alpha}) \le \mu_i\Big\}, \\ \mathcal{H}^p_{=}(\mu_1, \mu_2) &:= \Big\{\boldsymbol{\alpha} \in \mathcal{M}_p(\boldsymbol{Y}) : \mathrm{h}_i^p(\boldsymbol{\alpha}) = \mu_i\Big\}. \end{aligned} \tag{5.27}$$

Clearly $\mathcal{H}^p_{=}(\mu_1, \mu_2) \subset \mathcal{H}^p_{\le}(\mu_1, \mu_2)$. They are nonempty: any plan of the form

$$\boldsymbol{\alpha} = \frac{1}{a_1^p\, a_2^p} \Big(\mu_1 \otimes \delta_{a_1}\Big) \otimes \Big(\mu_2 \otimes \delta_{a_2}\Big), \quad a_1, a_2 > 0 \tag{5.28}$$

belongs to $\mathcal{H}^p_{=}(\mu_1, \mu_2)$. It is not difficult to check that $\mathcal{H}^p_{\le}(\mu_1, \mu_2)$ is also narrowly closed; on the contrary, this property does not hold for $\mathcal{H}^p_{=}(\mu_1, \mu_2)$ if $\mu_1(X_1)\mu_2(X_2) \neq 0$: it is sufficient to consider the vanishing sequence $\mathrm{dil}_{n^{-1},p}(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha} \in \mathcal{H}^p_{=}(\mu_1, \mu_2)$.

There is a natural correspondence between $\mathcal{H}^p_{\le}(\mu_1, \mu_2)$ (resp. $\mathcal{H}^p_{=}(\mu_1, \mu_2)$) and $\mathcal{H}^1_{\le}(\mu_1, \mu_2)$ (resp. $\mathcal{H}^1_{=}(\mu_1, \mu_2)$) induced by the map $\boldsymbol{Y} \ni (x_1, r_1; x_2, r_2) \mapsto (x_1, r_1^p; x_2, r_2^p)$. For plans $\boldsymbol{\alpha} \in \mathcal{H}^1_{\le}(\mu_1, \mu_2)$ we can prove a result similar to Lemma 5.4.

Lemma 5.7. *For every $\mu_i \in \mathcal{M}(X_i)$, $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$, $\boldsymbol{\alpha} \in \mathcal{H}^p_{\le}(\mu_1, \mu_2)$ with $\mu_i' := \mu_i - \mathrm{h}_i^p\boldsymbol{\alpha}$, $p \in (0, \infty)$, we have*

$$\int_{\boldsymbol{X}} H(x_1, r_1^p; x_2, r_2^p) \,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)\mu_i'(X_i) \ge \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2). \tag{5.29}$$

Proof. The calculations are quite similar to the proof of Lemma 5.4:

$$\begin{aligned} &\int_{\boldsymbol{Y}} H(x_1, r_1^p; x_2, r_2^p) \,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)\mu_i'(X_i) \\ &\overset{(5.14)}{\ge} \int_{\boldsymbol{Y}} \Big(F_1^\circ(\varphi_1(x_1))r_1^p + F_2^\circ(\varphi_2(x_2))r_2^p\Big) \,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)\mu_i'(X_i) \\ &= \sum_i \int_{X_i} F_i^\circ(\varphi_i) \,\mathrm{d}(\mathrm{h}_i^p\boldsymbol{\alpha}) + \sum_i F_i(0)\mu_i'(X_i) \\ &\overset{(2.20)}{\ge} \sum_i \int_{X_i} F_i^\circ(\varphi_i) \,\mathrm{d}(\mathrm{h}_i^p\boldsymbol{\alpha}) + \sum_i \int_{X_i} F_i^\circ(\varphi_i) \,\mathrm{d}\mu_i' = \sum_i \int_{X_i} F_i^\circ(\varphi_i) \,\mathrm{d}\mu_i = \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2). \qquad \square \end{aligned}$$

Theorem 5.8. *For every $\mu_i \in \mathcal{M}(X_i)$, $p \in (0,\infty)$ we have*

$$\mathsf{ET}(\mu_1,\mu_2) = \min_{\boldsymbol{\alpha} \in \mathcal{H}^p_{\le}(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} \Big(\sum_i R_i(r_i^p) + \mathsf{c}(x_1,x_2)\Big)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)(\mu_i - \mathrm{h}_i^p(\boldsymbol{\alpha}))(X_i) \tag{5.30}$$

$$= \min_{\boldsymbol{\alpha} \in \mathcal{H}^p_{\le}(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} H(x_1,r_1^p;x_2,r_2^p)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)(\mu_i - \mathrm{h}_i^p(\boldsymbol{\alpha}))(X_i) \tag{5.31}$$

$$= \min_{\boldsymbol{\alpha} \in \mathcal{H}^p_{=}(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} H(x_1,r_1^p;x_2,r_2^p)\,\mathrm{d}\boldsymbol{\alpha}. \tag{5.32}$$

Moreover, every plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1,\mu_2)$ (resp. optimal for (5.18) or for (5.21)) with $\mu_i = \varrho_i\gamma_i + \mu_i^\perp$, the plan $\boldsymbol{\alpha} := (x_1, \varrho_1^{1/p}(x_1); x_2, \varrho_2^{1/p}(x_2))_\sharp\boldsymbol{\gamma}$ realizes the minimum of (5.30) (resp. (5.31) or (5.32)).

Remark 5.9. When $F_i(0) = +\infty$ (5.30) and (5.31) simply read as

$$\mathsf{ET}(\mu_1,\mu_2) = \min_{\boldsymbol{\alpha} \in \mathcal{H}^p_{=}(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} \Big(\sum_i R_i(r_i^p) + \mathsf{c}(x_1,x_2)\Big)\,\mathrm{d}\boldsymbol{\alpha} \tag{5.33}$$

$$= \min_{\boldsymbol{\alpha} \in \mathcal{H}^1_{=}(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} H(x_1,r_1^p;x_2,r_2^p)\,\mathrm{d}\boldsymbol{\alpha}. \qquad \square \tag{5.34}$$

Proof of Theorem 5.8. Let us denote by E' (resp. E'', E''') the right-hand side of (5.30) (resp. of (5.31), (5.32)), where "min" has been replaced by "inf". If $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ and $\mu_i = \varrho_i\gamma_i + \mu_i^\perp$ (in the case of (5.32) $\mu_i^\perp = 0$) is the usual Lebesgue decomposition as in (3.30), we can consider the plan $\boldsymbol{\alpha} := (x_1, \varrho_1^{1/p}(x_1); x_2, \varrho_2^{1/p}(x_2))_\sharp\boldsymbol{\gamma}$.

Since the map $(\varrho_1^{1/p}, \varrho_2^{1/p}) : \boldsymbol{X} \to \mathbb{R}^2$ is Borel and takes values in a metrizable and separable space, it is Lusin $\boldsymbol{\gamma}$-measurable [38, Thm 5, p. 26], so that $\boldsymbol{\alpha}$ is a Radon measure in $\mathcal{M}(\boldsymbol{Y})$. We easily get for every nonnegative $\phi_i \in \mathrm{B}_b(X_i)$

$$\int \phi_i(x_i) r_i^p\,\mathrm{d}\boldsymbol{\alpha} = \int \varrho_i(x_i)\phi_i(x_i)\,\mathrm{d}\boldsymbol{\gamma} = \int \varrho_i\phi_i\,\mathrm{d}\gamma_i \le \int \phi_i\,\mathrm{d}\mu_i,$$

so that $\boldsymbol{\alpha} \in \mathcal{H}^p_{\le}(\mu_1,\mu_2)$, $\mathrm{h}_i^p\boldsymbol{\alpha} = \gamma_i$ and

$$\begin{aligned}\mathscr{R}(\mu_1,\mu_2|\boldsymbol{\gamma}) &= \int_{\boldsymbol{X}} \Big(\sum_i R_i(\varrho_i(x_i)) + \mathsf{c}(x_1,x_2)\Big)\,\mathrm{d}\boldsymbol{\gamma} + \sum_i F_i(0)\mu_i^\perp(X_i) \\ &= \int_{\boldsymbol{Y}} \sum_i R_i(r_i^p) + \mathsf{c}(x_1,x_2)\Big)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i F_i(0)(\mu_i - \mathrm{h}_i^p\boldsymbol{\alpha})(X_i) \ge E';\end{aligned}$$

taking the infimum w.r.t. $\boldsymbol{\gamma}$ and recalling (3.33) we get $\mathsf{ET}(\mu_1,\mu_2) \ge E'$. Since $\sum_i R_i(r_i^p) + \mathsf{c}(x_1,x_2) \ge H(x_1,r_1^p;x_2,r_2^p)$ it is also clear that $E' \ge E''$.

On the other hand, Lemma 5.7 shows that $E'' \ge \mathscr{D}(\boldsymbol{\varphi}|\mu_1,\mu_2)$ for every $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$: applying Theorem 4.11 we get $\mathsf{ET}(\mu_1,\mu_2) = E' = E''$. Concerning E''' it is clear that $E''' \ge E'' = \mathsf{ET}(\mu_1,\mu_2)$; when (5.20) hold, by choosing $\boldsymbol{\alpha}$ induced by a minimizer of (5.21) we get the opposite inequality $E''' \le \mathsf{ET}(\mu_1,\mu_2)$.

If (5.20) does not hold, we can still apply a slight modification of the argument at the end of the proof of Theorem 5.5. The only case to consider is when only one of the two measures $\mu_i^\perp$ vanishes: just to fix the ideas, let us suppose that $\tilde m_1 = \mu_1^\perp(X_1) > 0 = \mu_2^\perp(X_2)$. If $\bar{\boldsymbol\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1,\mu_2)$ and $\bar{\boldsymbol\alpha}$ is obtained as above, we can just set $\boldsymbol\alpha := \bar{\boldsymbol\alpha} + (\mu_1^\perp \times \delta_1) \times (\nu \times \delta_0)$ for an arbitrary $\nu \in \mathcal{P}(X_2)$. It is clear that $\mathrm{h}_i^p\boldsymbol\alpha = \mu_i$ and

$$\begin{aligned}\int_{\boldsymbol Y} H(x_1,r_1^p;x_2,r_2^p)\,\mathrm{d}\boldsymbol\alpha &= \int_{\boldsymbol X} H(x_1,\varrho_1(x_1);x_2,\varrho_2(x_2))\,\mathrm{d}\bar{\boldsymbol\gamma} + \int_{\boldsymbol X} H(x_1,1;x_2,0)\,\mathrm{d}\mu_1^\perp\otimes\nu\\ &\overset{(5.11)}{\le} \int_{\boldsymbol X} H(x_1,\varrho_1(x_1);x_2,\varrho_2(x_2))\,\mathrm{d}\bar{\boldsymbol\gamma} + F_1(0)\tilde m_1 = \mathscr{H}(\mu_1,\mu_2|\bar{\boldsymbol\gamma}) = \mathsf{ET}(\mu_1,\mu_2),\end{aligned}$$

which yields $E''' \le \mathsf{ET}(\mu_1,\mu_2)$. □

Remark 5.10 (Rescaling invariance)**.** By recalling (5.26a,b) and exploiting the 1-homogeneity of H it is not restrictive to solve the minimum problem (5.31) in the smaller class of probability plans concentrated in

$$\boldsymbol Y_{r,p} := \big\{(x_1,r_1;x_2,r_2) \in \boldsymbol Y : r_1^p + r_2^p \le r^p\big\},\quad r^p = \sum_i \mu_i(X_i).$$

Notice that it is not restrictive to assume that $\boldsymbol\alpha(\{\boldsymbol y \in \boldsymbol Y : |\boldsymbol y| = 0\}) = 0$ since $H(x_1,0;x_2,0) = 0$ for every $x_i \in X_i$. □

Part II. The Logarithmic Entropy-Transport problem and the Hellinger-Kantorovich distance

6 The Logarithmic Entropy-Transport (LET) problem

Starting from this section we will study a particular Entropy-Transport problem, whose structure reveals surprising properties.

6.1 The metric setting for Logarithmic Entropy-Transport problems.

Let (X,τ) be a Hausdorff topological space endowed with an extended distance function $\mathsf{d} : X \times X \to [0,\infty]$ which is lower semicontinuous w.r.t. τ; we refer to (X,τ,d) as an extended metric-topological space. In the most common situations, d will take finite values, (X,d) will be separable and complete and τ will be the topology induced by d; nevertheless, there are interesting applications where non separable extended distances play an important role, so that it will be useful to deal with an auxiliary topology, see e.g. [3, 1].

From now on we suppose that $X_1 = X_2 = X$, we choose the logarithmic entropies

$$F_i(s) = U_1(s) := s\log s - (s-1), \tag{6.1}$$

and a cost c depending on the distance d through the function $\ell : [0,\infty] \to [0,\infty]$

$$\mathsf{c}(x_1,x_2) := \ell\big(\mathsf{d}(x_1,x_2)\big), \qquad \ell(d) := \begin{cases} \log(1+\tan^2(d)) & \text{if } d \in [0,\pi/2), \\ +\infty & \text{if } d \geq \pi/2, \end{cases} \tag{6.2}$$

so that

$$\mathsf{c}(x_1,x_2) = \begin{cases} -\log\big(\cos^2(\mathsf{d}(x_1,x_2))\big) & \text{if } \mathsf{d}(x_1,x_2) < \pi/2 \\ +\infty & \text{otherwise.} \end{cases} \tag{6.3}$$

Let us collect a few key properties that will be relevant in the sequel.

LE.1 F_i are superlinear, regular, strictly convex, with $\mathrm{Dom}(F_i) = [0,\infty)$, $F_i(0) = 1$, and $(F_i)'_0 = -\infty$. When $s > 0$ $\partial F_i(s) = \{\log s\}$.

LE.2 $R_i(r) = rF_i(1/r) = r - 1 - \log r$; $R_i(0) = +\infty$, $(R_i)'_\infty = 1$.

LE.3 $F_i^*(\phi) = \exp(\phi) - 1$, $F_i^\circ(\varphi) = 1 - \exp(-\varphi)$; $\mathrm{Dom}(F_i^*) = \mathrm{Dom}(F_i^\circ) = \mathbb{R}$.

LE.4 $R_i^*(\psi) = -\log(1-\psi)$ if $\psi < 1$; $R_i^*(\psi) = +\infty$ if $\psi \geq 1$.

LE.5 The function ℓ can be characterized as the unique solution of the differential equation

$$\ell''(d) = 2\exp(\ell(d)), \quad \ell(0) = \ell'(0) = 0, \tag{6.4}$$

since it satisfies

$$\ell(d) = -\log\big(\cos^2(d)\big) = 2\int_0^d \tan(s)\,\mathrm{d}s, \quad d \in [0,\pi/2), \tag{6.5}$$

so that

$$\ell(d) \geq d^2, \quad \ell'(d) = 2\tan d \geq 2d, \quad \ell''(d) = 2(1+\tan^2(d)) = 2\exp(\ell(d)) \geq 2. \tag{6.6}$$

In particular ℓ is strictly increasing and uniformly 2-convex. It is not difficult to check that $\sqrt{\ell}$ is also convex: this property is equivalent to $2\ell\ell'' \geq (\ell')^2$ and a direct calculation shows

$$2\ell\ell'' - (\ell')^2 = 4\log(1+\tan^2(d))(1+\tan^2(d)) - 4\tan^2(d) \geq 0$$

since $(1+r)\log(1+r) \geq r$.

LE.6 $H_c(r_1,r_2) = r_1 + r_2 - 2\sqrt{r_1 r_2}\exp(-c/2)$ if $c < \infty$, so that

$$H(x_1,r_1;x_2,r_2) = r_1 + r_2 - 2\sqrt{r_1 r_2}\cos\big(\mathsf{d}_{\pi/2}(x_1,x_2)\big), \tag{6.7}$$

where we set

$$\mathsf{d}_a(x_1,x_2) := \mathsf{d}(x_1,x_2) \wedge a, \quad x_i \in X,\ a \geq 0. \tag{6.8}$$

Since the function

$$H(x_1,r_1^2;x_2,r_2^2) = r_1^2 + r_2^2 - 2r_1r_2\cos(\mathsf{d}_{\pi/2}(x_1,x_2)) \tag{6.9}$$

will have an important geometric interpretation (see Section 7.1), in the following we will choose the exponent $p = 2$.

We keep the usual notation $\boldsymbol{X} = X \times X$, identifying X_1 and X_2 with X and letting the index i run between 1 and 2, e.g. for $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ the marginals are denoted by $\gamma_i = (\pi^i)_\sharp\boldsymbol{\gamma}$.

Problem 6.1 (The Logarithmic Entropy-Transport problem)**.** *Let (X, τ, d) be an extended metric-topological space, ℓ and c be as in* (6.2)*. Given $\mu_i \in \mathcal{M}(X)$ find $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ minimizing*

$$\mathsf{LET}(\mu_1, \mu_2) = \min_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \sum_i \int_X \big(\sigma_i \log \sigma_i - \sigma_i + 1\big)\,\mathrm{d}\mu_i + \int_{\boldsymbol{X}} \ell\big(\mathsf{d}(x_1, x_2)\big)\,\mathrm{d}\boldsymbol{\gamma}, \quad \sigma_i = \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}. \tag{6.10}$$

We denote by $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ *the set of all the minimizers of* (6.10)*.*

6.2 The Logarithmic Entropy-Transport problem: main results

We collect in the next Theorem the main properties of the Logarithmic Entropy-Transport (LET) problem.

Theorem 6.2 (Direct formulation of the LET problem)**.** *Let $\mu_i \in \mathcal{M}(X)$ be given and let $\ell, \mathsf{d}_{\pi/2}$ be defined as in* (6.2) *and* (6.8)*.*
a) Existence of optimal plans. *There exists an optimal plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ solving Problem 6.1. The set $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ is convex and compact in $\mathcal{M}(\boldsymbol{X})$, LET is a convex and positively 1-homogeneous functional (see* (4.43)*) satisfying $0 \le \mathsf{LET}(\mu_1, \mu_2) \le \sum_i \mu_i(X)$.*
b) Reverse formulation. *The functional LET has the equivalent reverse formulation as*

$$\begin{aligned}\mathsf{LET}(\mu_1, \mu_2) = \min\Big\{ \sum_i \Big(\mu_i^\perp(X) + \int_X \big(\varrho_i - 1 - \log \varrho_i\big)\,\mathrm{d}\gamma_i\Big) + \int_{\boldsymbol{X}} \ell\big(\mathsf{d}(x_1, x_2)\big)\,\mathrm{d}\boldsymbol{\gamma} : \\ \boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X}), \quad \mu_i = \varrho_i\gamma_i + \mu_i^\perp \Big\},\end{aligned} \tag{6.11}$$

and $\bar{\boldsymbol{\gamma}}$ is an optimal plan in $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ if and only if it minimizes (6.11)*.*
c) The homogeneous perspective formulation. *$\mathsf{LET}(\mu_1, \mu_2)$ can be equivalently characterized as*

$$\begin{aligned}&\min_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \sum_i \mu_i^\perp(X) + \int_{\boldsymbol{X}} \Big(\varrho_1(x_1) + \varrho_2(x_2) - 2\sqrt{\varrho_1(x_1)\varrho_2(x_2)}\cos(\mathsf{d}_{\pi/2}(x_1, x_2))\Big)\,\mathrm{d}\boldsymbol{\gamma} \\ &= \sum_i \mu_i(X) - 2\max_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \int_{\boldsymbol{X}} \sqrt{\varrho_1(x_1)\varrho_2(x_2)}\cos(\mathsf{d}_{\pi/2}(x_1, x_2))\,\mathrm{d}\boldsymbol{\gamma},\end{aligned} \tag{6.12}$$

where $\gamma_i = \varrho_i\mu_i + \mu_i^\perp$, and every plan $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ provides a solution to (6.12)*.*

Proof. The variational problem (6.10) fits in the class considered by Problem 3.1, in the basic coercive setting of Section 3.1 since the logarithmic entropy (6.1) is superlinear with domain $[0, \infty)$. The problem is always feasible since $U_1(0) = 1$ so that (3.6) holds.

a) follows by Theorem 3.3(i); the upper bound of LET is a particular case of (3.7), and its convexity and 1-homogeneity follows by Corollary 4.13.

b) is a consequence of Theorem 3.11.

c) is an application of Theorem 5.5 and (6.7). □

We consider now the dual representation of LET; recall that $\mathrm{LSC}_s(X)$ denotes the space of simple (i.e. taking a finite number of values) lower semicontinuous functions and for a couple $\phi_i : X \to \mathbb{R}$ the symbol $\phi_1 \oplus \phi_2$ denotes the function $(x_1, x_2) \mapsto \phi_1(x_1) + \phi_2(x_2)$ defined in $\boldsymbol{X}$.

Theorem 6.3 (Dual formulation and optimality conditions)**.**
a) The dual problem. *For every $\mu_i \in \mathcal{M}(X)$ we have*

$$\mathsf{LET}(\mu_1, \mu_2) = \sup\Big\{ \sum_i \int_X \big(1 - \mathrm{e}^{-\varphi_i}\big)\,\mathrm{d}\mu_i : \varphi_i \in \mathrm{LSC}_s(X),\ \varphi_1 \oplus \varphi_2 \le \ell(\mathsf{d})\Big\}, \tag{6.13}$$

$$= \sup\Big\{ \sum_i \int_X \psi_i\,\mathrm{d}\mu_i : \psi_i \in \mathrm{LSC}_s(X),\quad \sup_X \psi_i < 1,$$

$$(1 - \psi_1(x_1))(1 - \psi_2(x_2)) \ge \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \text{ in } \boldsymbol{X}\Big\}. \tag{6.14}$$

The same identities hold if the space $\mathrm{LSC}_s(X)$ *is replaced by* $\mathrm{LSC}_b(X)$ *or* $\mathrm{B}_b(X)$ *in* (6.13) *and* (6.14)*. When the topology τ is completely regular (in particular when d is a distance and τ is induced by d) the space* $\mathrm{LSC}_s(X)$ *can be replaced by* $\mathrm{C}_b(X)$ *as well.*
b) Optimality conditions. *Let us assume that d is continuous. A plan $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ is optimal if and only if its marginals γ_i are absolutely continuous w.r.t. μ_i, $\int_{\boldsymbol{X}} \ell(\mathsf{d})\,\mathrm{d}\boldsymbol{\gamma} < \infty$,*

$$\mathsf{d} \ge \pi/2 \quad \text{in} \quad \Big(\big(\operatorname{supp}\mu_1 \setminus \operatorname{supp}\gamma_1\big) \times \operatorname{supp}\mu_2\Big) \bigcup \Big(\operatorname{supp}\mu_1 \times \big(\operatorname{supp}\mu_2 \setminus \operatorname{supp}\gamma_2\big)\Big), \tag{6.15}$$

and there exist Borel sets $A_i \subset \operatorname{supp}\gamma_i$ with $\gamma_i(X \setminus A_i) = 0$ and Borel densities $\sigma_i : A_i \to (0, \infty)$ of γ_i w.r.t. μ_i such that

$$\sigma_1(x_1)\sigma_2(x_2) \ge \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \quad \text{in } A_1 \times A_2, \tag{6.16}$$

$$\sigma_1(x_1)\sigma_2(x_2) = \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \quad \boldsymbol{\gamma}\text{-a.e. in } A_1 \times A_2. \tag{6.17}$$

c) $\ell(\mathsf{d})$-cyclical monotonicity. *Every optimal plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ is a solution of the optimal transport problem with cost $\ell(\mathsf{d})$ between its marginals γ_i. In particular it is $\ell(\mathsf{d})$-cyclically monotone, i.e. it is concentrated on a Borel set $G \subset \boldsymbol{X}$ ($G = \operatorname{supp}(\boldsymbol{\gamma})$ when d is continuous) such that for every choice of $(x_1^n, x_2^n)_{n=1}^N \subset G$ and every permutation $\kappa : \{1, \ldots, N\} \to \{1, \ldots, N\}$*

$$\Pi_{n=1}^N \cos^2(\mathsf{d}_{\pi/2}(x_1^n, x_2^n)) \ge \Pi_{n=1}^N \cos^2(\mathsf{d}_{\pi/2}(x_1^n, x_2^{\kappa(n)})). \tag{6.18}$$

d) Generalized potentials. *If $\boldsymbol{\gamma}$ is optimal and A_i, σ_i are defined as in b) above, the Borel potentials $\varphi_i, \psi_i : X \to \overline{\mathbb{R}}$*

$$\varphi_i := \begin{cases} -\log\sigma_i & \text{in } A_i, \\ -\infty & \text{in } X \setminus \operatorname{supp}\mu_i, , \\ +\infty & \text{otherwise}, \end{cases} \qquad \psi_i := \begin{cases} 1 - \sigma_i & \text{in } A_i, \\ -\infty & \text{in } X \setminus \operatorname{supp}\mu_i, \\ 1 & \text{otherwise}, \end{cases} \tag{6.19}$$

satisfy $\varphi_1 \oplus_o \varphi_2 \le \ell(\mathsf{d})$, and the optimality conditions (4.21ab,c,d) *(with the analogous properties for ψ_i). Moreover $\mathrm{e}^{-\varphi_i}, \psi_i \in \mathrm{L}^1(X, \mu_i)$ and*

$$\mathsf{LET}(\mu_1, \mu_2) = \sum_i \int_X \big(1 - \mathrm{e}^{-\varphi_i}\big)\,\mathrm{d}\mu_i = \sum_i \int_X \psi_i\,\mathrm{d}\mu_i = \sum_i \mu_i(X) - 2\boldsymbol{\gamma}(\boldsymbol{X}). \tag{6.20}$$

Proof. (6.13) follows by Theorem 4.11, recalling the definition (4.11) of $\boldsymbol{\Phi}$ and the fact that $F_i^\circ(\varphi) = 1 - \exp(-\varphi)$.

(6.14) follows from Proposition 4.3 and the fact that $R_i^*(\psi) = -\log(1-\psi)$. Notice that the definition (4.7) of $\boldsymbol{\Psi}$ ensures that we can restrict the supremum in (6.14) to functions ψ_i with $\sup_X \psi_i < 1$. We have discussed the possibility to replace $\mathrm{LSC}_s(X)$ with $\mathrm{LSC}_b(X)$, $\mathrm{B}_b(X)$ or $\mathrm{C}_b(X)$ in Corollary 4.12.

The statement of point b) follows by Corollary 4.17; notice that a plan with finite energy $\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) < \infty$ always satisfies $\int_{\boldsymbol{X}} \ell(\mathsf{d}) < \infty$; conversely, if the latter integrability property holds, (6.17) and the fact that $\int_{A_i} (\log\sigma_i)_- \,\mathrm{d}\gamma_i = \int_{A_i} \sigma_i(\log\sigma_i)_- \,\mathrm{d}\mu_i < \infty$ yields $\mathscr{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) < \infty$.

Point c) is an obvious consequence of the optimality of $\boldsymbol{\gamma}$.

Point d) can be easily deduced by b) or by applying Theorem 4.15. □

Corollary 6.4 (Monotonicity of optimal plans in $\mathbb{R}$)**.** When $X = \mathbb{R}$ with the usual distance, the support of every optimal plan $\boldsymbol{\gamma}$ is a monotone set, i.e.

$$(x_1,x_2),\ (x_1',x_2') \in \mathrm{supp}(\boldsymbol{\gamma}),\ x_1 < x_1' \quad\Rightarrow\quad x_2 \le x_2'. \tag{6.21}$$

Proof. Since the function ℓ is uniformly convex, (6.18) is equivalent to monotonicity. □

Corollary 6.5. *For every $\mu_i \in \mathcal{M}(X)$ we have*

$$\mathsf{LET}(\mu_1,\mu_2) = \sum_i \mu_i(X) - 2\max\Big\{\boldsymbol{\gamma}(\boldsymbol{X}):\ \boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X}),\ \gamma_i = \sigma_i\mu_i,$$
$$\sigma_1(x_1)\sigma_2(x_2) \le \cos^2(\mathsf{d}_{\pi/2}(x_1,x_2))\ \boldsymbol{\gamma}\text{-a.e. in } \boldsymbol{X}\Big\}. \tag{6.22}$$

Proof. Let us denote by M' the right-hand side and let $\boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X})$ be a plan satisfying the conditions of (6.22). If A_i are Borel sets with $\gamma_i(X \setminus A_i) = 0$ and $\sigma_i : X \to (0,\infty)$ are Borel densities of γ_i w.r.t. μ_i, we have $\varrho_i(x_i) = 1/\sigma_i(x_i)$ in A_i so that $\sigma_1(x_1)\sigma_2(x_2) \le \cos^2(\mathsf{d}_{\pi/2}(x_1,x_2))$ yields $\varrho_1(x_1)\varrho_2(x_2)\cos^2(\mathsf{d}_{\pi/2}(x_1,x_2)) \ge 1$. Since $(\log\varrho_i)_+ \in \mathrm{L}^1(X,\gamma_i)$ we have

$$\begin{aligned}&\sum_i \Big(\mu_i^\perp(X) + \int_X \big(\varrho_i - 1 - \log\varrho_i\big)\,\mathrm{d}\gamma_i\Big) + \int_{\boldsymbol{X}} \ell\big(\mathsf{d}(x_1,x_2)\big)\,\mathrm{d}\boldsymbol{\gamma}\\ =&\sum_i \big(\mu_i(X) - \gamma_i(X)\big) - \int_{\boldsymbol{X}} \log\big(\varrho_1(x_1)\varrho_2(x_2)\cos^2(\mathsf{d}_{\pi/2}(x_1,x_2))\big)\,\mathrm{d}\boldsymbol{\gamma} \le \sum_i \mu_i(X) - 2\boldsymbol{\gamma}(\boldsymbol{X}).\end{aligned}$$

By (6.11) we get $M' \ge \mathsf{LET}(\mu_1,\mu_2)$. On the other hand, choosing any $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$ (6.17) shows that $\bar{\boldsymbol{\gamma}}$ is an admissible competitor for (6.22) and (6.20) shows that $M' = \mathsf{LET}(\mu_1,\mu_2)$. □

Remark 6.6. Notice that the concave function $\sum_i \mu_i(X) - \mathsf{LET}(\mu_1,\mu_2)$ can be represented as

$$\sum_i \mu_i(X) - \mathsf{LET}(\mu_1,\mu_2) = 2 \max_{\boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X})} \int_{\boldsymbol{X}} \sqrt{\varrho_1(x_1)\varrho_2(x_2)}\cos(\mathsf{d}_{\pi/2}(x_1,x_2))\,\mathrm{d}\boldsymbol{\gamma} \tag{6.23}$$

$$= \inf\Big\{\sum_i \int_X \mathrm{e}^{-\varphi_i}\,\mathrm{d}\mu_i : \varphi_i \in \mathrm{LSC}_s(X),\ \varphi_1 \oplus \varphi_2 \le \ell(\mathsf{d})\Big\} \tag{6.24}$$

$$= \inf\Big\{\sum_i \int_X \tilde\psi_i\,\mathrm{d}\mu_i : \tilde\psi_i \in \mathrm{USC}_s(X),\ \inf_X \tilde\psi_i > 0,$$

$$\psi_1(x_1)\psi_2(x_2) \ge \cos^2(\mathsf{d}_{\pi/2}(x_1,x_2)) \text{ in } \boldsymbol{X}\Big\} \tag{6.25}$$

$$= 2\max\Big\{\boldsymbol{\gamma}(\boldsymbol{X}) :\ \boldsymbol{\gamma}\in\mathcal{M}(\boldsymbol{X}),\ \gamma_i = \sigma_i\mu_i,$$

$$\sigma_1(x_1)\sigma_2(x_2) \le \cos^2(\mathsf{d}_{\pi/2}(x_1,x_2))\ \boldsymbol{\gamma}\text{-a.e. in } \boldsymbol{X}\Big\}. \tag{6.26}$$

□

The next result concerns uniqueness of the optimal plan $\boldsymbol{\gamma}$ in the Euclidean case $X = \mathbb{R}^d$. We will use the notion of approximate differential (denoted by $\tilde{\mathrm{D}}$), see e.g. [2, Def. 5.5.1].

Theorem 6.7 (Uniqueness)**.** *Let $\mu_i \in \mathcal{M}(X)$, $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$.*

(i) The marginals $\gamma_i = \pi^i_\sharp\boldsymbol{\gamma}$ are uniquely determined.

(ii) If $X = \mathbb{R}$ with the usual distance then $\boldsymbol{\gamma}$ is the unique element of $\mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$.

(iii) If $X = \mathbb{R}^d$ with the usual distance, $\mu_1 \ll \mathscr{L}^d$ is absolutely continuous, and $A_i \subset \mathbb{R}^d$ and $\sigma_i : A_i \to (0,\infty)$ are as in Theorem 6.3 b), then σ_1 is approximately differentiable at $\boldsymbol{\gamma}_1$-a.e. point of A_1 and $\boldsymbol{\gamma}$ is the unique element of $\mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$; it is concentrated on the graph of a function $\boldsymbol{t} : \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\boldsymbol{t}(x_1) = x_1 + \frac{\arctan(|\boldsymbol{\xi}(x_1)|)}{|\boldsymbol{\xi}(x_1)|}\boldsymbol{\xi}(x_1), \quad \boldsymbol{\xi}(x_1) = -\frac{1}{2}\tilde{\mathrm{D}}\log\sigma_1(x_1)\ \gamma_1\text{-a.e. in } A_1. \tag{6.27}$$

Proof. (i) follows directly from Lemma 3.5.

(ii) follows by Theorem 6.3(c), since whenever the marginals γ_i are fixed there is only one plan with monotone support in $\mathbb{R}$.

In order to prove (iii) we adapt the argument of [2, Thm. 6.2.4] to our singular setting, where the cost c can take the value $+\infty$.

Let $A_i \subset \mathbb{R}^d$ and $\sigma_i : A_i \to (0,\infty)$ as in Theorem 6.3 b). Since $\mu_1 = u\mathscr{L}^d \ll \mathscr{L}^d$, up to removing a μ_1-negligible set (and thus γ_1-negligible) from A_1, it is not restrictive to assume that $u(x_1) > 0$ everywhere in A_1, so that the classes of $\mathscr{L}^d$- and γ_1-negligible subsets of A_1 coincide. For every $n \in \mathbb{N}$ we define

$$A_{2,n} := \{x_2 \in A_2 : \sigma_2(x_2) \ge 1/n\}, \quad s_n(x_1) := \sup_{x_2\in A_{2,n}} \cos^2(|x_1 - x_2|)/\sigma_2(x_2). \tag{6.28}$$

The functions s_n are uniformly bounded and Lipschitz in $\mathbb{R}^d$ and therefore differentiable $\mathscr{L}^d$-a.e. by Rademacher's Theorem. Since $\gamma_1 \ll \mu_1$ and μ_1 is absolutely continuous w.r.t. $\mathscr{L}^d$ we deduce that s_n are differentiable γ_1-a.e. in A_1.

By (6.16) we have $\sigma_1(x_1) \geq s_n(x_1)$ in A_1; (6.17) yields that for γ_1-a.e. $x_1 \in A_1$ there exists $x_2 \in A_2$ such that $|x_1 - x_2| < \pi/2$ and $\sigma_1(x_1) = \cos^2(|x_1 - x_2|)/\sigma_2(x_2)$ so that $\sigma_1(x_1) = s_n(x_1)$ for n sufficiently big. The family of sets $\{x_1 \in A_1 : \sigma_1(x_1) > s_n(x_1)\}$ is thus decreasing (since s_n is increasing and dominated by σ_1) and has $\mathscr{L}^d$-negligible intersection.

It follows that γ_1-a.e. $x_1 \in A_1$ is a point of $\mathscr{L}^d$-density 1 of $\{x_1 \in A_1 : \sigma_1(x_1) = s_n(x_1)\}$ for some $n \in \mathbb{N}$ and s_n is differentiable at x_1. Let us denote by A_1' such a set: σ_1 is approximately differentiable at every $x_1 \in A_1'$ with approximate differential $\tilde{\mathrm{D}}\sigma_1(x_1)$ equal to $\mathrm{D}s_n(x_1)$ for n sufficiently big.

Suppose now that $x_1 \in A_1'$ and $\sigma_1(x_1) = \cos^2(|x_1 - x_2|)/\sigma_2(x_2)$ for some $x_2 \in A_2$; since the map $x_1' \mapsto \cos^2(|x_1' - x_2|)/\sigma_1(x_1')$ attains its maximum at $x_1' = x_1$ we deduce that

$$\tan(|x_1 - x_2|)\frac{x_1 - x_2}{|x_1 - x_2|} = -\frac{1}{2}\tilde{\mathrm{D}}\log\sigma_1(x_1),$$

so that x_2 is uniquely determined and we get (6.27). □

We conclude this section with the last representation formula for $\mathsf{LET}(\mu_1, \mu_2)$ given in terms of plans in $\boldsymbol{Y} := Y \times Y$ with $Y := X \times [0, \infty)$ with constraints on the homogeneous marginals, keeping the notation of Section 5.2. Even if it seems the most complicated one, it will provide the natural point of view in order to study the metric properties of the LET functional.

Theorem 6.8. *For every $\mu_i \in \mathcal{M}(X)$ we have*

$$\begin{aligned}\mathsf{LET}(\mu_1, \mu_2) = \min\Big\{ \int_{\boldsymbol{Y}} \Big(r_1^2 + r_2^2 - 2r_1r_2\cos(\mathsf{d}_{\pi/2}(x_1, x_2))\Big)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i (\mu_i - \mathrm{h}_i^2\boldsymbol{\alpha})(X) :& \\ \boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{Y}),\ \mathrm{h}_i^2\boldsymbol{\alpha} \leq \mu_i\Big\}& \qquad (6.29)\\ = \sum_i \mu_i(X) - 2\max_{\boldsymbol{\alpha}\in\mathcal{H}^2_{\leq}(\mu_1,\mu_2)} \int_{\boldsymbol{X}} r_1r_2\cos(\mathsf{d}_{\pi/2}(x_1, x_2))\,\mathrm{d}\boldsymbol{\alpha}& \qquad (6.30)\\ = \min\Big\{ \int_{\boldsymbol{Y}} \Big(r_1^2 + r_2^2 - 2r_1r_2\cos(\mathsf{d}_{\pi/2}(x_1, x_2))\Big)\,\mathrm{d}\boldsymbol{\alpha} :& \\ \boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{Y}),\ \mathrm{h}_i^2\boldsymbol{\alpha} = \mu_i\Big\}& \qquad (6.31)\end{aligned}$$

Moreover, for every plan $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{LET}}\mu_1\mu_2$ and every couple of Borel densities ϱ_i as in (6.11) *the plan $\bar{\boldsymbol{\alpha}} := (x_1, \sqrt{\varrho_1(x_1)}; x_2, \sqrt{\varrho_2(x_2)})_\sharp\bar{\boldsymbol{\gamma}}$ is optimal for* (6.29) *and* (6.30).

Proof. (6.29) (resp. (6.31)) follows directly by (5.31) (resp. (5.32)) of Theorem 5.8; (6.30) is just a different form for (6.29). □

7 The metric side of the LET-functional: the Hellinger-Kantorovich distance

In this section we want to show that the functional

$$\mathsf{HK}(\mu_1, \mu_2) := \sqrt{\mathsf{LET}(\mu_1, \mu_2)} \tag{7.1}$$

defines a distance in $\mathcal{M}(X)$. This property is strongly related to the fact that the function $(x_1, r_1; x_2, r_2) \mapsto \big(H(x_1, r_1^2; x_2, r_2^2)\big)^{1/2}$ is a (possibly extended) semidistance in $Y = X \times [0, \infty)$.

In the next section we will briefly study this function and the induced metric space (the so-called *cone* $\mathfrak{C}$ *on* X, [9, Section 3.6] obtained by taking the quotient w.r.t. the equivalent classes of points with distance 0.

7.1 The cone construction

In the extended metric-topological space (X, τ, d) of Section 6.1, we will denote by $\mathsf{d}_a := \mathsf{d} \wedge a$ the truncated distance and by $y = (x, r)$, $x \in X$, $r \in [0, \infty)$, the generic points of $Y := X \times [0, \infty)$.

It is not difficult to show that the function $\mathsf{d}_{\mathfrak{C}} : Y \times Y \to [0, \infty)$

$$\mathsf{d}_{\mathfrak{C}}^2((x_1, r_1), (x_2, r_2)) := r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_\pi(x_1, x_2)) \tag{7.2}$$

is nonnegative, symmetric, and satisfies the triangle inequality (see e.g. [9, Prop. 3.6.13]). We also notice that

$$\mathsf{d}_{\mathfrak{C}}^2(y_1, y_2) = |r_1 - r_2|^2 + 4r_1 r_2 \sin^2\big(\mathsf{d}_\pi(x_1, x_2)/2\big), \tag{7.3}$$

so that

$$\max\Big(|r_1 - r_2|, \frac{2}{\pi}\sqrt{r_1 r_2}\,\mathsf{d}_\pi(x_1, x_2)\Big) \le \mathsf{d}_{\mathfrak{C}}(y_1, y_2) \le |r_1 - r_2| + \sqrt{r_1 r_2}\,\mathsf{d}_\pi(x_1, x_2). \tag{7.4}$$

It should be clear that $\mathsf{d}_{\mathfrak{C}}$ induces a true distance in the quotient space $\mathfrak{C} = Y/\sim$ where

$$y_1 \sim y_2 \quad \Leftrightarrow \quad r_1 = r_2 = 0 \quad \text{or} \quad r_1 = r_2,\ x_1 = x_2. \tag{7.5}$$

Equivalence classes are usually denoted by $\mathfrak{y} = [y] = [x, r]$ the vertex $[x, 0]$ plays a distinguished role and it is denoted by $\mathfrak{o}$, its complement is the open set $\mathfrak{C}_{\mathfrak{o}} = \mathfrak{C} \setminus \{\mathfrak{o}\}$. On $\mathfrak{C}$ we introduce a topology $\tau_{\mathfrak{C}}$, which is in general weaker than the canonical quotient topology: $\tau_{\mathfrak{C}}$ neighborhoods of points in $\mathfrak{C}_{\mathfrak{o}}$ coincide with neighborhoods in Y, whereas the sets

$$\{[x, r] : 0 \le r < \varepsilon\} = \{\mathfrak{y} \in \mathfrak{C} : \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}, \mathfrak{o}) < \varepsilon\}, \quad \varepsilon > 0, \tag{7.6}$$

provide a system of open neighborhoods of $\mathfrak{o}$. $\tau_{\mathfrak{C}}$ coincides with the quotient topology when X is compact.

It is easy to check that $(\mathfrak{C}, \tau_{\mathfrak{C}})$ is an Hausdorff topological space and $\mathsf{d}_{\mathfrak{C}}$ is $\tau_{\mathfrak{C}}$-lower semicontinuous. If τ is induced by d then $\tau_{\mathfrak{C}}$ is induced by $\mathsf{d}_{\mathfrak{C}}$. If (X, d) is complete (resp. separable), then $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ is also complete (resp. separable).

Perhaps the simplest example is provided by the unit sphere $X = \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$ in $\mathbb{R}^d$ endowed with the intrinsic Riemannian distance: the corresponding cone $\mathfrak{C}$ is precisely $\mathbb{R}^d$.

We will call $\mathfrak{p} : Y \to \mathfrak{C}$ the canonical projection, $\mathfrak{p}(x, r) = [x, r]$. Clearly $\mathfrak{p}$ is continuous and is an homeomorphism between $Y \setminus (X \times \{0\})$ and $\mathfrak{C}_{\mathfrak{o}}$. A right inverse $\mathsf{y} : \mathfrak{C} \to Y$ of the map $\mathfrak{p}$ can be obtained by fixing a point $\bar{x} \in X$ and defining

$$\mathsf{r} : \mathfrak{C} \to [0, \infty), \ \ \mathsf{r}[x, r] = r; \qquad \mathsf{x} : \mathfrak{C} \to X, \ \ \mathsf{x}[x, r] = \begin{cases} x & \text{if } r > 0, \\ \bar{x} & \text{if } r = 0. \end{cases} \qquad \mathsf{y} := (\mathsf{x}, \mathsf{r}). \tag{7.7}$$

Notice that r is continuous and x is continuous in $\mathfrak{C}_{\mathfrak{o}}$.

A continuous rescaling product from $\mathfrak{C} \times [0, \infty)$ to $\mathfrak{C}$ can be defined by

$$\mathfrak{y} \cdot \lambda := \begin{cases} \mathfrak{o} & \text{if } \mathfrak{y} = \mathfrak{o}, \\ [x, \lambda r] & \text{if } \mathfrak{y} = [x, r], \ s > 0. \end{cases} \tag{7.8}$$

We conclude this introductory section by a simple characterization of $\tau_{\mathfrak{C}}$ compact sets in $\mathfrak{C}$:

Lemma 7.1 (Compact sets in $\mathfrak{C}$)**.** *A closed set K of $\mathfrak{C}$ is compact if and only if there is $r_0 > 0$ such that its upper sections*

$$K(\rho) := \{x \in X : [x, r] \in K \textit{ for some } r \geq \rho\}$$

are empty for $\rho > r_0$ and compact in X for $0 < \rho \leq r_0$.

Proof. It is easy to check that the condition is necessary. In order to show the sufficiency, let $\mathfrak{y}_\lambda = [x_\lambda, r_\lambda]$, $\lambda \in \mathbb{L}$, be a net in K and let $\rho := \inf_\lambda r_\lambda$. If $\rho = 0$ then we can find a subnet $(r_{\lambda'})_{\lambda' \in \mathbb{L}'}$ converging to 0 and therefore $\mathfrak{y}_{\lambda'} \to \mathfrak{o}$. If $\rho > 0$, then $x_\lambda \in K(\rho)$ and $r_\lambda \in [\rho, r_0]$ so that (r_λ, r_λ) admits a converging subnet in Y whose projection converges in $\mathfrak{C}$. $\square$

Remark 7.2 (Two different truncations)**.** Notice that in the constitutive formula defining $\mathsf{d}_{\mathfrak{C}}$ we used the truncated distance d_π with upper threshold π, whereas in Theorem 6.8 an analogous formula with $\mathsf{d}_{\pi/2}$ and threshold $\pi/2$ played a crucial role. We could then consider the distance on $\mathfrak{C}$

$$\begin{aligned} \mathsf{d}^2_{\pi/2,\mathfrak{C}}([x_1, r_1], [x_2, r_2]) &:= r_1^2 + r_2^2 - 2 r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \qquad &(7.9\text{a}) \\ &= |r_1 - r_2|^2 + 4 r_1 r_2 \sin^2(\mathsf{d}_{\pi/2}(x_1, x_2)/2) \qquad &(7.9\text{b}) \end{aligned}$$

which satisfy

$$\mathsf{d}_{\pi/2,\mathfrak{C}} \leq \mathsf{d}_{\mathfrak{C}} \leq \sqrt{2}\, \mathsf{d}_{\pi/2,\mathfrak{C}}, \ \ \mathfrak{C}' := \big\{\mathsf{d}_{\pi/2,\mathfrak{C}} < \mathsf{d}_{\mathfrak{C}}\big\} = \big\{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}_{\mathfrak{o}} \times \mathfrak{C}_{\mathfrak{o}} : \mathsf{d}(\mathsf{x}_1, \mathsf{x}_2) > \pi/2\big\}. \tag{7.10}$$

The notation (7.9a) is justified by the fact that $\mathsf{d}_{\pi/2,\mathfrak{C}}$ is still a cone distance associated to the metric space $(X, \mathsf{d}_{\pi/2})$, since obviously $(\mathsf{d}_{\pi/2})_\pi = (\mathsf{d}_{\pi/2}) \wedge \pi/2 = \mathsf{d}_{\pi/2}$. From the geometric point of view, the choice of $\mathsf{d}_{\mathfrak{C}}$ is the most natural one, since it preserves important metric properties concerning geodesics (see [9, Thm. 3.6.17] and the next section

8.1) and curvature (see [9, Sect. 4.7] and the next section 8.3). On the other hand, the choice of $\mathsf{d}_{\pi/2}$ is crucial for its link with the function H of (6.9), with Entropy-Transport problems, and with a representation property for the Hopf-Lax formula that we will see in the next sections. Notice that the 1-homogeneous formula (6.7) would not be convex in (r_1, r_2) if one considers d_π instead of $\mathsf{d}_{\pi/2}$. Nevertheless, we will prove in Section 7.3 the remarkable fact that both d_π and $\mathsf{d}_{\pi/2}$ will lead to the same distance between positive measures. □

7.2 Radon measures in the cone $\mathfrak{C}$ and homogeneous marginals

It is clear that any measure $\nu \in \mathcal{M}(\mathfrak{C})$ can be lifted to a measure $\bar\nu \in \mathcal{M}(Y)$ such that $\mathfrak{p}_\sharp \bar\nu = \nu$: it is sufficient to take $\bar\nu = \mathsf{y}_\sharp \nu$ where y is a right inverse of $\mathfrak{p}$ defined as in (7.7).

We call $\mathcal{M}_2(\mathfrak{C})$ (resp. $\mathcal{P}_2(\mathfrak{C})$) the space of measures $\nu \in \mathcal{M}(\mathfrak{C})$ (resp. $\nu \in \mathcal{P}(\mathfrak{C})$) such that

$$\int_{\mathfrak{C}} \mathsf{r}^2 \,\mathrm{d}\nu = \int_{\mathfrak{C}} \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o}) \,\mathrm{d}\nu = \int_Y r^2 \,\mathrm{d}\bar\nu < \infty, \quad \bar\nu = \mathsf{y}_\sharp \nu. \tag{7.11}$$

Measures in $\mathcal{M}_2(\mathfrak{C})$ thus correspond to images $\mathfrak{p}_\sharp \bar\nu$ of measures $\bar\nu \in \mathcal{M}_2(Y)$ and have finite second moment w.r.t. the distance $\mathsf{d}_{\mathfrak{C}}$ (this justifies the index 2 in $\mathcal{M}_2(\mathfrak{C})$). Notice moreover that the measure $s^2\bar\nu$ does not charge $X \times \{0\}$ and it is independent of the choice of the point $\bar x$ in (7.7).

The above considerations can be easily extended to plans in the product spaces $\mathfrak{C}^{\otimes N}$ (where typically $N = 2$, but also the general case will turn out to be useful later on). To clarify the notation, we will denote by $\boldsymbol{\mathfrak{y}} = (\mathfrak{y}_i)_{i=1}^N = ([x_i, r_i])_{i=1}^N$ a point in $\mathfrak{C}^{\otimes N}$ and we will set $\mathsf{r}_i(\boldsymbol{\mathfrak{y}}) = \mathsf{r}(\mathfrak{y}_i) = r_i$, $\mathsf{x}_i(\boldsymbol{\mathfrak{y}}) = \mathsf{x}(\mathfrak{y}_i) \in X$. Projections on the i-coordinate from $\mathfrak{C}^{\otimes N}$ to $\mathfrak{C}$ are usually denoted by π^i or $\pi^{\mathfrak{y}_i}$, $\boldsymbol{\mathfrak{p}} = \mathfrak{p}^{\otimes N} : (Y)^{\otimes N} \to \mathfrak{C}^{\otimes N}$, $\mathbf{y} = \mathsf{y}^{\otimes N} : \mathfrak{C}^{\otimes N} \to (Y)^{\otimes N}$ are the Cartesian products of the projections and of the lifts.

Recall that the L^2-Kantorovich-Wasserstein (extended) distance $\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}$ in $\mathcal{M}_2(\mathfrak{C})$ induced by $\mathsf{d}_{\mathfrak{C}}$ is defined by

$$\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}^2(\nu_1, \nu_2) := \min\Big\{ \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) \,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{\mathfrak{C}}),\ \pi^{\mathfrak{y}_i}_\sharp \boldsymbol{\alpha} = \nu_i \Big\}, \tag{7.12}$$

with the convention that $\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\nu_1, \nu_2) = +\infty$ if $\nu_1(\mathfrak{C}) \neq \nu_2(\mathfrak{C})$ and thus the minimum in (7.12) is taken on an empty set. We want to mimic the above definition, replacing the usual marginal conditions in (7.12) with the homogeneous marginals $\mathfrak{h}_i^2$ which we are going to define.

Let us consider now a plan $\boldsymbol{\alpha}$ in $\mathcal{M}(\mathfrak{C}^{\otimes N})$ with $\bar{\boldsymbol{\alpha}} = \mathbf{y}_\sharp \boldsymbol{\alpha} \in \mathcal{M}(Y^{\otimes N})$: we say that $\boldsymbol{\alpha}$ lies in $\mathcal{M}_2(\mathfrak{C}^{\otimes N})$ if

$$\int_{\mathfrak{C}^{\otimes N}} \sum_i \mathsf{r}_i^2 \,\mathrm{d}\boldsymbol{\alpha} = \int_{Y^{\otimes N}} \sum_i r_i^2 \,\mathrm{d}\bar{\boldsymbol{\alpha}} < \infty. \tag{7.13}$$

Its “canonical” marginals in $\mathcal{M}(\mathfrak{C})$ are $\boldsymbol{\alpha}_i = \pi^{\mathfrak{y}_i}_\sharp \boldsymbol{\alpha}$, whereas the “homogeneous” marginals correspond to (5.23):

$$\mathfrak{h}_i^2(\boldsymbol{\alpha}) := (\mathsf{x}_i)_\sharp(\mathsf{r}_i^2 \boldsymbol{\alpha}) = \pi^{x_i}_\sharp(r_i^2 \bar{\boldsymbol{\alpha}}) = \mathrm{h}_i^2(\bar{\boldsymbol{\alpha}}) \in \mathcal{M}(X), \quad \bar{\boldsymbol{\alpha}} := \mathbf{y}_\sharp \boldsymbol{\alpha}. \tag{7.14}$$

We will omit the index i when $N = 1$. Notice that $\mathsf{r}_i^2 \boldsymbol{\alpha}$ does not charge $(\pi^i)^{-1}(\mathfrak{o})$ (similarly, $r_i^2 \bar{\boldsymbol{\alpha}}$ does not charge $Y^{\otimes i-1} \times \{(\bar x, 0)\} \times Y^{\otimes N-i}$) so that (7.14) is independent of the choice of the point $\bar{x}$ in (7.7). As for (5.25), the homogeneous marginals on the cone are also invariant with respect to dilations: if $\vartheta : \mathfrak{C}^{\otimes N} \to (0,\infty)$ is a Borel map in $\mathrm{L}^2(\mathfrak{C}^{\otimes N}, \boldsymbol{\alpha})$ we set

$$\big(\mathrm{prd}_\vartheta(\mathfrak{y})\big)_i := \mathfrak{y}_i \cdot \vartheta^{-1}(\mathfrak{y}), \qquad \mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha}) := \vartheta^2\,(\mathrm{prd}_\vartheta)_\sharp \boldsymbol{\alpha}, \tag{7.15}$$

so that

$$\mathfrak{h}_i^2(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})) = \mathfrak{h}_i^2(\boldsymbol{\alpha}) \quad \text{for every } \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N}). \tag{7.16}$$

As for the canonical marginals, a uniform control of the homogeneous marginals is sufficient to get equal tightness. We state this result for an arbitrary number of components; notice that we are not claiming any closedness of the involved sets.

Lemma 7.3 (Homogeneous marginals and tightness)**.** *Let $\mathcal{K}_i$, $i = 1, \cdots, N$, be a finite collection of bounded and equally tight sets in $\mathcal{M}(X)$. The set*

$$\big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^N) : \mathfrak{h}_i^2 \boldsymbol{\alpha} \in \mathcal{K}_i\big\} \tag{7.17}$$

is equally tight in $\mathcal{M}(\mathfrak{C}^N)$.

Proof. By applying [2, Lemma 5.2.2], it is sufficient to consider the case $N = 1$: given a bounded and equally tight set $\mathcal{K} \subset \mathcal{M}(X)$ we prove that $\mathcal{H} := \big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}) : \mathfrak{h}^2\boldsymbol{\alpha} \in \mathcal{K}\big\}$ is equally tight. For $A \subset X$, $R \subset (0,\infty)$ we will use the shorter notation $A \times_{\mathfrak{C}} R$ for $\mathfrak{p}(A \times R) \subset \mathfrak{C}$; if A and R are compact, than $A \times_{\mathfrak{C}} R$ is compact in $\mathfrak{C}$.

Let $M := \sup_{\mu\in\mathcal{K}} \mu(X) < \infty$; since $\mathcal{K}$ is tight, we can find an increasing sequence of compact sets $K_n \subset X$ such that $\mu(X \setminus K_n) \le 8^{-n}$ for every $\mu \in \mathcal{K}$. For an integer $m \in \mathbb{N}$ we then consider the compact sets $\mathfrak{K}_m \subset \mathfrak{C}$ defined by

$$\mathfrak{K}_m = \{\mathfrak{o}\} \cup K_m \times_{\mathfrak{C}} [2^{-m}, 2^m] \cup \Big(\bigcup_{n=1}^{\infty} K_{n+m} \times_{\mathfrak{C}} [2^{-n}, 2^{-n+1}]\Big). \tag{7.18}$$

If $K_\infty = \bigcup_{n=1}^\infty K_n$, so that $\mu(X \setminus K_\infty) = 0$, we have

$$\mathfrak{C} \setminus \mathfrak{K}_m \subset K_m \times_{\mathfrak{C}} (2^m, \infty) \cup \Big(\bigcup_{n=1}^{\infty} (K_{n+m} \setminus K_{n+m-1}) \times_{\mathfrak{C}} (2^{-n+1}, \infty)\Big) \cup (X \setminus K_\infty) \times_{\mathfrak{C}} (0,\infty).$$

Since for every $\boldsymbol{\alpha} \in \mathcal{H}$ with $\mathfrak{h}^2\boldsymbol{\alpha} = \mu$ and every $A \in \mathcal{B}(X)$

$$\boldsymbol{\alpha}(A \times_{\mathfrak{C}} (s,\infty)) \le s^{-2}\mu(A) \le s^{-2}M, \quad \boldsymbol{\alpha}\big((X \setminus K_\infty) \times_{\mathfrak{C}} (0,\infty)\big) = 0,$$

we get

$$\begin{aligned} \boldsymbol{\alpha}(\mathfrak{C} \setminus \mathfrak{K}_m) &\le M\,4^{-m} + \sum_{n=1}^{\infty} \boldsymbol{\alpha}\big((X \setminus K_{n+m-1}) \times_{\mathfrak{C}} (2^{-n+1}, \infty)\big) \le \\ &\qquad M\,4^{-m} + \sum_{n=1}^{\infty} 4^{n-1} 8^{1-n-m} \le 4^{-m}\Big(M + \sum_{n=1}^{\infty} 4^{-n}\Big) \le 4^{-m}\big(1 + M)\big), \end{aligned}$$

for every $\boldsymbol{\alpha} \in \mathcal{H}$. Since $\mathfrak{K}_m$ are compact we obtain the tightness. □

7.3 The Hellinger-Kantorovich problem

In this section we will always consider $N = 2$, keeping the shorter notation $\boldsymbol{Y} = Y^{\otimes 2}$, $\boldsymbol{\mathfrak{C}} = \mathfrak{C}^{\otimes 2}$. As for (5.27), for every $\mu_1, \mu_2 \in \mathcal{M}_2(X)$ we denote by

$$\mathfrak{H}^2_{\le}(\mu_1,\mu_2) := \Big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\boldsymbol{Y}) : \mathfrak{h}^2_i\boldsymbol{\alpha} \le \mu_i\Big\}, \quad \mathfrak{H}^2_{=}(\mu_1,\mu_2) := \Big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\boldsymbol{\mathfrak{C}}) : \mathfrak{h}^2_i\boldsymbol{\alpha} = \mu_i\Big\}. \tag{7.19}$$

They are the images of $\mathcal{H}^2_{\le}(\mu_1,\mu_2)$ and $\mathcal{H}^2_{=}(\mu_1,\mu_2)$ through the projections $\mathfrak{p}_\sharp$; in particular they always contain plans $\mathfrak{p}_\sharp\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is given by (5.28). The condition $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\le}(\mu_1,\mu_2)$ is equivalent to ask that

$$\int \mathsf{r}_i^2\varphi(\mathsf{x}_i)\,\mathrm{d}\boldsymbol{\alpha} \le \int \varphi\,\mathrm{d}\mu_i \quad \text{for every nonnegative } \varphi \in \mathrm{B}_b(X). \tag{7.20}$$

We can thus define the following minimum problem:

Problem 7.4 (The Hellinger-Kantorovich problem)**.** *Given $\mu_1, \mu_2 \in \mathcal{M}(X)$ find an optimal plan $\boldsymbol{\alpha}_{\rm opt} \in \mathcal{H}^2_{=}(\mu_1,\mu_2) \subset \mathcal{M}_2(\boldsymbol{\mathfrak{C}})$ solving the minimum problem*

$$\min\Big\{\int \mathsf{d}^2_{\mathfrak{C}}(\mathfrak{y}_1,\mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}_2(\boldsymbol{\mathfrak{C}}),\ \mathfrak{h}^2_i\boldsymbol{\alpha} = \mu_i\Big\}. \tag{7.21}$$

We denote by $\mathrm{Opt}_{\mathsf{HK}}(\mu_1,\mu_2) \subset \mathcal{M}(\boldsymbol{\mathfrak{C}})$ the collection of all the optimal plans $\boldsymbol{\alpha}$ realizing the minimum in (7.21) *and by $\mathsf{HK}^2(\mu_1,\mu_2)$ the value of the minimum in* (7.21) *(whose existence is guaranteed by the next Theorem 7.7).*

Remark 7.5 (Lifting of plans in Y)**.** Since any plan $\boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{\mathfrak{C}})$ can be lifted to a plan $\bar{\boldsymbol{\alpha}} = \mathsf{y}_\sharp\boldsymbol{\alpha} \in \mathcal{P}(Y \times Y)$ such that $\mathfrak{p}_\sharp\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ the previous problem 7.4 is also equivalent to find

$$\min\Big\{\int \mathsf{d}^2_{\mathfrak{C}}(y_1,y_2)\,\mathrm{d}\bar{\boldsymbol{\alpha}} : \bar{\boldsymbol{\alpha}} \in \mathcal{M}(Y \times Y), \quad \mathrm{h}^2_i\bar{\boldsymbol{\alpha}} = \mu_i\Big\}. \tag{7.22}$$

The advantage to work in the quotient space $\mathfrak{C}$ is to gain compactness, as the next Theorem 7.7 will show. □

Remark 7.6 (Rescaling invariance)**.** Let us set

$$\mathfrak{C}[R] := \big\{[x,r] \in \mathfrak{C} : r \le R\big\}, \quad \boldsymbol{\mathfrak{C}}[R] := \mathfrak{C}[R] \times \mathfrak{C}[R]. \tag{7.23}$$

It is not restrictive to solve the previous problem 7.4 by also assuming that $\boldsymbol{\alpha}$ is a probability plan in $\mathcal{P}(\boldsymbol{\mathfrak{C}})$ concentrated on $\boldsymbol{\mathfrak{C}}[R]$ with $R^2 = \sum_i \mu_i(X)$, i.e.

$$\mathsf{HK}^2(\mu_1,\mu_2) = \min_{\boldsymbol{\alpha}\in C}\int \mathsf{d}^2_{\mathfrak{C}}\,\mathrm{d}\boldsymbol{\alpha}, \qquad C := \Big\{\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{\mathfrak{C}}) : \mathfrak{h}^2_i\boldsymbol{\alpha} = \mu_i,\ \boldsymbol{\alpha}\big(\boldsymbol{\mathfrak{C}} \setminus \boldsymbol{\mathfrak{C}}[R]\big) = 0\Big\}. \tag{7.24}$$

In fact the functional $\mathsf{d}^2_{\mathfrak{C}}$ and the constraints have a natural scaling invariance induced by the dilation maps defined by (7.15). Since

$$\int \mathsf{d}^2_{\mathfrak{C}}\,\mathrm{d}(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})) = \int \vartheta^2\mathsf{d}^2_{\mathfrak{C}}([x_1,r_1/\vartheta];[x_2,r_2/\vartheta])\,\mathrm{d}\boldsymbol{\alpha} = \int \mathsf{d}^2_{\mathfrak{C}}\,\mathrm{d}\boldsymbol{\alpha}, \tag{7.25}$$

choosing ϑ as in (5.26a) with $p = 2$ and restricting $\boldsymbol{\alpha}$ to $\boldsymbol{\mathfrak{C}}\setminus\{(\mathfrak{o},\mathfrak{o})\}$ we obtain a probability plan $\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})$ in $\mathcal{H}^2_{=}(\mu_1,\mu_2)$ concentrated in $\boldsymbol{\mathfrak{C}}[R] \setminus \{(\mathfrak{o},\mathfrak{o})\}$ with the same cost $\int \mathsf{d}^2_{\mathfrak{C}}\,\mathrm{d}\boldsymbol{\alpha}$. □

In order to show that Problem 7.4 has a solution we can then use the formulation (7.24) and prove that the set C where the minimum is settled is narrowly compact in $\mathcal{P}(\boldsymbol{\mathfrak{C}})$. Notice that the analogous property would not be true in $\mathcal{P}(Y \times Y)$ (unless X is compact) since measures concentrated in $(X \times \{0\}) \times (X \times \{0\})$ would be out of control. Also the constraints $\mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i$ would not be preserved by narrow convergence, if one allows for arbitrary plans in $\mathcal{P}(\boldsymbol{\mathfrak{C}})$ as in (7.21).

Theorem 7.7 (Existence of optimal plans for the HK problem)**.** *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ the Hellinger-Kantorovich problem 7.4 always admits a solution $\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{\mathfrak{C}})$ concentrated on $\boldsymbol{\mathfrak{C}}[R] \setminus \{(\mathfrak{o}, \mathfrak{o})\}$ with $R^2 = \sum_i \mu_i(X)$.*

Proof. By Remark 7.6 it is not restrictive to look for minimizers $\boldsymbol{\alpha}$ of (7.24). Since $\boldsymbol{\mathfrak{C}}[R]$ is closed in $\boldsymbol{\mathfrak{C}}$ and the maps r_i^2 are continuous and bounded in $\boldsymbol{\mathfrak{C}}[R]$, C is clearly narrowly closed. By Lemma 7.3, C is also equally tight in $\mathcal{P}(\boldsymbol{\mathfrak{C}})$, thus narrowly compact by Theorem 2.2. Since the $\mathsf{d}_{\mathfrak{C}}^2$ is lower semicontinuous in $\boldsymbol{\mathfrak{C}}$, the existence of a minimizer of (7.24) then follows by the direct method of Calculus of Variations. □

We can also prove an interesting characterization of HK in terms of the L^2-Kantorovich-Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$ given by (7.12). An even deeper connection will be discussed in the next section, see Corollary 7.14.

Corollary 7.8 (HK and the Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$)**.** *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}(\mu_1, \mu_2) = \min \Big\{ \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) : \alpha_i \in \mathcal{P}_2(\mathfrak{C}), \quad \mathfrak{h}^2 \alpha_i = \mu_i \Big\}, \tag{7.26}$$

and there exist optimal measures $\bar{\alpha}_i$ for (7.26) concentrated on $\mathfrak{C}[R]$ with $R^2 = \sum_i \mu_i(X)$. In particular the map $\mathfrak{h}^2 : \mathcal{P}_2(\mathfrak{C}) \to \mathcal{M}(X)$ is a contraction, i.e.

$$\mathsf{HK}(\mathfrak{h}^2 \alpha_1, \mathfrak{h}^2 \alpha_2) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) \quad \textit{for every } \alpha_i \in \mathcal{P}_2(\mathfrak{C}). \tag{7.27}$$

Proof. If $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$ with $\mathfrak{h}^2 \alpha_i = \mu_i$ then any Kantorovich-Wasserstein optimal plan $\boldsymbol{\alpha} \in \mathcal{P}(\mathfrak{C} \times \mathfrak{C})$ for (7.12) with marginals α_i clearly belongs to $\mathfrak{H}^2_=(\mu_1, \mu_2)$ and yields the bound $\mathsf{HK}(\mu_1, \mu_2) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2)$. On the other hand, if $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}} \mu_1 \mu_2$ is an optimal solution for (7.21) and $\alpha_i := \pi^i \boldsymbol{\alpha} \in \mathcal{P}_2(\mathfrak{C})$ are its marginals, we have $\mathsf{HK}(\mu_1, \mu_2) \ge \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2)$, so that α_i realize the minimum for (7.26). □

We conclude this section with two simple properties of the HK functional. We denote by η_0 the null measure.

Lemma 7.9 (Subadditivity of HK^2)**.** *The functional HK^2 satisfies*

$$\mathsf{HK}^2(\mu, \eta_0) = \mu(X), \qquad \mathsf{HK}^2(\mu_1, \mu_2) \le \mu_1(X) + \mu_2(X) \quad \textit{for every } \mu, \mu_i \in \mathcal{M}(X), \tag{7.28}$$

and it is subadditive, i.e. for every $\mu_i, \mu_i' \in \mathcal{M}(X)$ we have

$$\mathsf{HK}^2(\mu_1 + \mu_1', \mu_2 + \mu_2') \le \mathsf{HK}^2(\mu_1, \mu_2) + \mathsf{HK}^2(\mu_1', \mu_2'). \tag{7.29}$$

Proof. (7.28) is obvious. If $\boldsymbol{\alpha} \in \mathfrak{H}^2_=(\mu_1, \mu_2)$ and $\boldsymbol{\alpha}' \in \mathfrak{H}^2_=(\mu_1', \mu_2')$ it is easy to check that $\boldsymbol{\alpha} + \boldsymbol{\alpha}' \in \mathfrak{H}^2_=(\mu_1 + \mu_1', \mu_2 + \mu_2')$. Since the cost functional is linear with respect to the plan, we get (7.29). □

Subsequently we will use "$\llcorner$" for the restriction of measures.

Lemma 7.10 (A formulation with relaxed constraints). *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}^2(\mu_1,\mu_2) = \min_{\boldsymbol{\alpha}\in\mathfrak{H}^2_{\le}(\mu_1,\mu_2)} \Big\{ \int \mathsf{d}^2_{\mathfrak{C}}(\mathfrak{y}_1,\mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i \big(\mu_i - \mathfrak{h}^2_i\boldsymbol{\alpha}\big)(X) \Big\} \tag{7.30a}$$

$$= \mu_1(X) + \mu_2(X) - \max_{\boldsymbol{\alpha}\in\mathfrak{H}^2_{\le}(\mu_1,\mu_2)} \Big\{ 2\int \mathsf{r}_1\,\mathsf{r}_2 \cos(\mathsf{d}_\pi(\mathsf{x}_1,\mathsf{x}_2))\,\mathrm{d}\boldsymbol{\alpha} \Big\}. \tag{7.30b}$$

Moreover

(i) equations (7.30a)–(7.30b) *share the same class of optimal plans.*

(ii) A plan $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\le}(\mu_1,\mu_2)$ is optimal for (7.30a)–(7.30b) *if and only if the plan $\boldsymbol{\alpha}_{\mathfrak{o}} := \boldsymbol{\alpha}\llcorner(\mathfrak{C}_{\mathfrak{o}}\times\mathfrak{C}_{\mathfrak{o}})$ is optimal as well.*

(iii) If $\boldsymbol{\alpha}$ is optimal for (7.30a)–(7.30b) *with $\mu_i' := \mu_i - \mathfrak{h}^2_i\boldsymbol{\alpha}$ then $\tilde{\boldsymbol{\alpha}} := \boldsymbol{\alpha} + \boldsymbol{\alpha}'$, where $\boldsymbol{\alpha}' \in \mathfrak{H}^2_{=}(\mu_1',\mu_2')$, is an optimal plan of* $\mathrm{Opt}_{\mathsf{HK}}(\mu_1,\mu_2)$.

(iv) A plan $\boldsymbol{\alpha} \in \mathfrak{H}^2_{=}(\mu_1,\mu_2)$ belongs to $\mathrm{Opt}_{\mathsf{HK}}(\mu_1,\mu_2)$ *if and only if $\boldsymbol{\alpha}_{\mathfrak{o}} := \boldsymbol{\alpha}\llcorner(\mathfrak{C}_{\mathfrak{o}}\times\mathfrak{C}_{\mathfrak{o}})$ is optimal for* (7.30a)–(7.30b).

Proof. (7.30a) and (7.30b) are just a different way to write the same functional (this proves (i)), since for every $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\le}(\mu_1,\mu_2)$

$$\int \mathsf{d}^2_{\mathfrak{C}}\,\mathrm{d}\boldsymbol{\alpha} + \sum_i \big(\mu_i - \mathfrak{h}^2_i\boldsymbol{\alpha}\big)(X) = \sum_i \mu_i(X) - 2\int \mathsf{r}_1\,\mathsf{r}_2\cos(\mathsf{d}_\pi(\mathsf{x}_1,\mathsf{x}_2))\,\mathrm{d}\boldsymbol{\alpha}; \tag{7.31}$$

so it is sufficient to prove (7.30a). The inequality $\ge$ is obvious, since $\mathfrak{H}^2_{\le}(\mu_1,\mu_2) \supset \mathfrak{H}^2_{=}(\mu_1,\mu_2)$ and for every $\boldsymbol{\alpha} \in \mathfrak{H}^2_{=}(\mu_1,\mu_2)$ the term $\sum_i \big(\mu_i - \mathfrak{h}^2_i\boldsymbol{\alpha}\big)(X)$ vanishes.

On the other hand, whenever $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\le}(\mu_1,\mu_2)$, setting $\mu_i'' := \mathfrak{h}^2_i\boldsymbol{\alpha} \in \mathcal{M}(X)$, $\mu_i' := \mu_i - \mu_i''$ and observing that $\boldsymbol{\alpha} \in \mathfrak{H}^2_{=}(\mu_1'',\mu_2'')$ we get

$$\begin{aligned}\int \mathsf{d}^2_{\mathfrak{C}}(\mathfrak{y}_1,\mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} + \sum_i \big(\mu_i - \mathfrak{h}^2_i\boldsymbol{\alpha}\big)(X) &\ge \mathsf{HK}^2(\mu_1'',\mu_2'') + \mu_1'(X) + \mu_2'(X)\\ &\overset{(7.28)}{\ge} \mathsf{HK}^2(\mu_1',\mu_2') + \mathsf{HK}^2(\mu_1'',\mu_2'') \overset{(7.29)}{\ge} \mathsf{HK}^2(\mu_1,\mu_2).\end{aligned}$$

The same calculations also prove the point (iii).

In order to check (ii) it is sufficient to observe that the integrand in (7.30b) vanishes on $(\mathfrak{C}) \setminus (\mathfrak{C}_{\mathfrak{o}}\times\mathfrak{C}_{\mathfrak{o}})$.

Finally, if $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_1,\mu_2)$ is optimal for (7.21), then by the consideration above it is optimal for (7.30b) and therefore (ii) shows that $\boldsymbol{\alpha}_{\mathfrak{o}}$ is optimal as well. The converse implication follows by (iii). $\square$

7.4 Gluing lemma and triangle inequality

In this section we will prove that HK satisfies the triangle inequality and therefore is a distance on $\mathcal{M}(X)$. The main technical step is provided by the following useful property for plans in $\mathcal{M}(\mathfrak{C}^{\otimes N})$ with given homogeneous marginals, which is a simple application of the rescaling technique.

Lemma 7.11. *Let $\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$, $N \geq 2$ be a plan satisfying*

$$\mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i \in \mathcal{M}(X) \quad i = 1, \ldots, N, \qquad a_i = \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i) \,\mathrm{d}\boldsymbol{\alpha} \qquad i = 2, \ldots, N, \tag{7.32}$$

and let $j \in \{1, \ldots, N\}$ be fixed. It is possible to find a new plan $\bar{\boldsymbol{\alpha}} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$ which still satisfies (7.32) *and*

$$\pi^j_\sharp(\bar{\boldsymbol{\alpha}}) = \delta_{\mathfrak{o}} + \mathfrak{p}_\sharp(\mu_j \otimes \delta_1). \tag{7.33}$$

Proof. By possibly adding $\otimes^N \delta_{\mathfrak{o}}$ to $\boldsymbol{\alpha}$ (which does not modify (7.32)) we may suppose that

$$\omega_j := \boldsymbol{\alpha}\big(\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \pi^j(\mathfrak{y}) = \mathfrak{o}\}\big) \geq 1.$$

It is sufficient to rescale $\boldsymbol{\alpha}$ by the function

$$\vartheta(\mathfrak{y}) := \begin{cases} \mathsf{r}_j(\mathfrak{y}) & \text{if } \mathfrak{y}_j \neq \mathfrak{o}, \\ \omega_j^{-1/2} & \text{otherwise.} \end{cases} \tag{7.34}$$

With the notation of (7.15) we set $\bar{\boldsymbol{\alpha}} := \mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})$ and we decompose $\boldsymbol{\alpha}$ in the sum $\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \boldsymbol{\alpha}''$ where $\boldsymbol{\alpha}' = \boldsymbol{\alpha} \llcorner \{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \pi^j(\mathfrak{y}) = \mathfrak{o}\}$. For every $\zeta \in \mathrm{B}_b(\mathfrak{C})$ we have

$$\begin{aligned}\int \zeta(\mathfrak{y}_j)\,\mathrm{d}\bar{\boldsymbol{\alpha}} &= \int \zeta(\mathfrak{y}_j \cdot \vartheta^{-1}(\mathfrak{y}))\vartheta^2(\mathfrak{y})\,\mathrm{d}\boldsymbol{\alpha} = \int \zeta(\mathfrak{o})\omega_j^{-1}\,\mathrm{d}\boldsymbol{\alpha}' + \int \zeta([x_j, r_j/\vartheta(\mathfrak{y})])\vartheta^2(\mathfrak{y})\,\mathrm{d}\boldsymbol{\alpha}'' \\ &= \zeta(\mathfrak{o}) + \int \zeta([x_j, 1])\mathsf{r}_j^2 \mathrm{d}\boldsymbol{\alpha}'' = \zeta(\mathfrak{o}) + \int \zeta([x_j, 1])\mathsf{r}_j^2 \mathrm{d}\boldsymbol{\alpha} = \zeta(\mathfrak{o}) + \int \zeta \circ \mathfrak{p}\mathrm{d}(\mu_j \otimes \delta_1)\end{aligned}$$

which yields (7.33). $\square$

We can now prove a general form of the so-called "gluing lemma" that is the natural extension of the well known result for transport problems (wee e.g. [2, Lemma 5.3.4]). Here its formulation is strongly related to the rescaling invariance of optimal plans given by Lemma 7.11.

Lemma 7.12 (Gluing lemma)**.** *Let us consider a finite collection of measures $\mu_i \in \mathcal{M}(X)$ and coefficients $\theta_i > 0$, $i = 1, \ldots, N$, $N \geq 2$ with*

$$\Theta^2 := \Big(\sum_i \theta_i^{-1}\Big) \cdot \Big(\theta_1 \mu_1(X) + \sum_{i=2}^N \theta_i \,\mathsf{HK}^2(\mu_{i-1}, \mu_i)\Big), \qquad M^2 := \sum_{i=1}^N \mu_i(X). \tag{7.35}$$

Then there exist plans $\boldsymbol{\alpha}_k \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$, $k = 1, 2$, such that

$$\mathfrak{h}_i^2 \boldsymbol{\alpha}_k = \mu_i, \quad i = 1, \ldots, N; \qquad \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i)\,\mathrm{d}\boldsymbol{\alpha}_k = \mathsf{HK}^2(\mu_{i-1}, \mu_i), \quad i = 2, \ldots, N, \tag{7.36}$$

and moreover

$$\boldsymbol{\alpha}_1 \textit{ is concentrated on } \Big\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sum_i \mathsf{r}_i^2(\mathfrak{y}) \le M^2\Big\}, \tag{7.37}$$

$$\boldsymbol{\alpha}_2 \textit{ is concentrated on } \Big\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sup_i \mathsf{r}_i(\mathfrak{y}) \le \Theta\Big\} = \big(\mathfrak{C}[\Theta]\big)^{\otimes N}. \tag{7.38}$$

Proof. We first construct a plan $\boldsymbol{\alpha}$ satisfying (7.36); further suitable rescalings will provide $\boldsymbol{\alpha}_k$ satisfying (7.37) or (7.38). In order to clarify the argument, we consider N-copies $X_1, X_2, \ldots, X_N$ of X (and for $\mathfrak{C}$ in a similar way) so that $X^{\otimes N} = \prod_{i=1}^N X_i$

We argue by induction; the starting case $N = 2$ is covered by Theorem 7.7 and Lemma 7.11. Let us then discuss the induction step, by assuming that the thesis holds for N and proving it for $N+1$. We can thus find an optimal plan $\boldsymbol{\alpha}^N$ such that (7.36) hold, and another optimal plan $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_N, \mu_{N+1})$ for the couple μ_N, μ_{N+1}. Applying Lemma 7.11 to $\boldsymbol{\alpha}^N$ (with $j = N$) and to $\boldsymbol{\alpha}$ (with $j = 1$) we can also assume that

$$\pi^N_\sharp(\boldsymbol{\alpha}^N) = \delta_{\mathfrak{o}} + \mathfrak{p}_\sharp(\mu_N \otimes \delta_1) = \pi^1_\sharp(\boldsymbol{\alpha}).$$

Therefore we can apply the standard gluing Lemma in $\big(\prod_{i=1}^{N-1} \mathfrak{C}_i\big), \mathfrak{C}_N, \mathfrak{C}_{N+1}$ (see e.g. [2, Lemma 5.3.2] and [1, Lemma 2.2] in the case of arbitrary topological spaces) obtaining a new plan $\boldsymbol{\alpha}^{N+1}$ satisfying $\pi^{1,2,\cdots,N}_\sharp \boldsymbol{\alpha}^{N+1} = \boldsymbol{\alpha}^N$ and $\pi^{N,N+1}\boldsymbol{\alpha}^{N+1} = \boldsymbol{\alpha}$. In particular, $\boldsymbol{\alpha}^{N+1}$ satisfies (7.36).

A further application of the argument of Remark 7.6, with ϑ as in (5.26a) yields a plan $\boldsymbol{\alpha}_1$ satisfying also (7.37).

In order to obtain $\boldsymbol{\alpha}_2$, since we can assume that $\boldsymbol{\alpha}(\{|\mathfrak{y}| = 0\}) = 0$, we can set $\boldsymbol{\alpha}_2 = \mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})$ use the rescaling function

$$\vartheta(\mathfrak{y}) := r^{-1}|\mathfrak{y}|_\infty = r^{-1}\sup_i \mathsf{r}_i(\mathfrak{y}), \quad r^2 := \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_\infty^2 \,\mathrm{d}\boldsymbol{\alpha},$$

and we have to estimate r. We can use the inequality

$$\begin{aligned}
\mathsf{r}_n &\le \mathsf{r}_1 + \sum_{i=2}^n |\mathsf{r}_i - \mathsf{r}_{i-1}| \le \Big(\sum_{i=1}^n \theta_i^{-1}\Big)^{1/2}\Big(\theta_1 \mathsf{r}_1^2 + \sum_{i=2}^n \theta_i |\mathsf{r}_i - \mathsf{r}_{i-1}|^2\Big)^{1/2} \\
&\le \Big(\sum_{i=1}^N \theta_i^{-1}\Big)^{1/2}\Big(\theta_1 \mathsf{r}_1^2 + \sum_{i=2}^N \theta_i \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_{i-1})\Big)^{1/2}
\end{aligned}$$

which yields

$$r^2 = \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_\infty^2 \,\mathrm{d}\boldsymbol{\alpha} \le \Big(\sum_{i=1}^N \theta_i^{-1}\Big)\int_{\mathfrak{C}^{\otimes N}} \Big(\theta_1 \mathsf{r}_1^2 + \sum_{i=2}^N \theta_i \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_{i-1})\Big)\,\mathrm{d}\boldsymbol{\alpha} = \Theta^2. \qquad \square$$

Remark 7.13. In a completely similar way (see [2, Lemma 5.3.4]), for every a finite collection of measures $\mu_i \in \mathcal{M}(X)$ and coefficients $\theta_i > 0$, $i = 1, \ldots, N$ $N \ge 2$, there exists a plan $\boldsymbol{\beta} \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ concentrated on $\big\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sup_i \mathsf{r}_i(\mathfrak{y}) \le \Xi\big\}$ with

$$\Xi := \Big(\sum_i \theta_i^{-1}\Big) \cdot \Big(\theta_1 \mu_1(X) + \sum_{i=2}^N \theta_i \mathsf{HK}^2(\mu_1, \mu_i)\Big), \tag{7.39}$$

such that

$$\mathfrak{h}_i^2\boldsymbol{\beta} = \mu_i \quad i = 1, \ldots, N; \qquad \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_i)\,\mathrm{d}\boldsymbol{\beta} = \mathsf{HK}^2(\mu_1, \mu_i) \quad i = 1, \ldots, N. \qquad \square \tag{7.40}$$

Arguing as in the proof of Corollary 7.8 one can immediately prove the following result.

Corollary 7.14. *For every finite collection of measures $\mu_i \in \mathcal{M}(X)$, $i = 1, \cdots, N$, there exist $\alpha_i, \beta_i \in \mathcal{P}_2(\mathfrak{C})$ with α_i concentrated in $\mathfrak{C}[r]$ where $r = \min(M, \Theta)$ is given as in (7.35) and β_i concentrated in $\mathfrak{C}[\Xi]$ given by (7.39) such that*

$$\mathfrak{h}^2\alpha_i = \mu_i \qquad i = 1, \cdots, N; \qquad \mathsf{HK}(\mu_i, \mu_{i+1}) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_i, \alpha_{i+1}) \qquad i = 2, \cdots, N \tag{7.41}$$

$$\mathfrak{h}^2\beta_i = \mu_i \qquad i = 1, \cdots, N; \qquad \mathsf{HK}(\mu_1, \mu_i) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\beta_1, \beta_i) \qquad i = 2, \cdots, N. \tag{7.42}$$

Corollary 7.15 (HK is a distance)**.** HK *is a distance on $\mathcal{M}(X)$; in particular, for every $\mu_1, \mu_2, \mu_3 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}(\mu_1, \mu_3) \le \mathsf{HK}(\mu_1, \mu_2) + \mathsf{HK}(\mu_2, \mu_3). \tag{7.43}$$

Proof. It is immediate to check that HK is symmetric and $\mathsf{HK}(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$. In order to check the triangle inequality (7.43) it is sufficient to apply the previous corollary 7.14 to find measures $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$, $i = 1, 2, 3$, such that $\mathfrak{h}^2\alpha_i = \mu_i$ and $\mathsf{HK}(\mu_1, \mu_2) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2)$ and $\mathsf{HK}(\mu_2, \mu_3) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_2, \alpha_3)$. Applying the triangle inequality for $\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}$ we obtain

$$\mathsf{HK}(\mu_1, \mu_3) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_3) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) + \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_2, \alpha_3) = \mathsf{HK}(\mu_1, \mu_2) + \mathsf{HK}(\mu_2, \mu_3). \qquad \square$$

As a consequence of the previous results, the map $\mathfrak{h}^2 : \mathcal{P}_2(\mathfrak{C}) \to \mathcal{M}(X)$ is a metric submersion.

7.5 Metric and topological properties

In this section we will assume that the topology τ is induced by d and (X, d) is separable, so that also $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ is separable. Notice that in this case there is no difference between weak and narrow topology in $\mathcal{M}(X)$. Moreover, since X is separable then $\mathcal{M}(X)$ is metrizable, so that converging sequences are sufficient to characterize the weak-narrow topology.

It turns out [2, Chap. 7] that $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}})$ is a separable metric space: convergence of a sequence $(\alpha_n)_{n\in\mathbb{N}}$ to a limit measure α in $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}})$ corresponds to weak-narrow convergence in $\mathcal{P}(\mathfrak{C})$ and convergence of the quadratic moments, or, equivalently, to convergence of integrals of continuous functions with quadratic growth, i.e.

$$\lim_{n\to\infty} \int_{\mathfrak{C}} \varphi\,\mathrm{d}\alpha_n = \int_{\mathfrak{C}} \varphi\,\mathrm{d}\boldsymbol{\alpha} \quad \text{for every } \varphi \in \mathrm{C}(\mathfrak{C}) \text{ with } |\varphi(\mathfrak{y})| \le A + B\mathsf{r}^2(\mathfrak{y}), \tag{7.44}$$

for some constants $A, B \ge 0$ depending on φ. Notice that $\mathsf{r}^2(\mathfrak{y}) = \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o})$.

Theorem 7.16 (HK metrizes the weak topology on $\mathcal{M}(X)$)**.** HK *induces the weak-narrow topology on $\mathcal{M}(X)$: a sequence $(\mu_n)_{n\in\mathbb{N}} \in \mathcal{M}(X)$ converges to a measure μ in $(\mathcal{M}, \mathsf{HK})$ if and only if $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ in duality with continuous and bounded functions.*

In particular, the metric space $(\mathcal{M}(X), \mathsf{HK})$ is separable.

Proof. Let us first suppose that $\lim_{n\to\infty} \mathsf{HK}(\mu_n, \mu) = 0$. We argue by contradiction and we assume that there exists a function $\zeta \in \mathrm{C}_b(X)$ and a subsequence (still denoted by μ_n) such that

$$\inf_n \Big| \int_X \zeta \,\mathrm{d}\mu_n - \int_X \zeta \,\mathrm{d}\mu \Big| > 0. \tag{7.45}$$

The first estimate of (7.28) and the triangle inequality show that

$$\limsup_{n\to\infty} \mu_n(X) \le \limsup_{n\to\infty} \big(\mathsf{HK}(\mu_n, \mu) + \mathsf{HK}(\mu, \eta_0)\big)^2 = \mu(X),$$

so that $\sup_n \mu_n(X) = M < \infty$. By Corollary 7.8 we can find measures $\alpha_n, \alpha_n' \in \mathcal{P}_2(\mathfrak{C})$ concentrated on $\mathfrak{C}[\sqrt{2M}]$ such that

$$\mathfrak{h}^2\alpha_n = \mu, \quad \mathfrak{h}^2\alpha_n' = \mu_n, \quad \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_n, \alpha_n') = \mathsf{HK}(\mu, \mu_n).$$

By Lemma 7.3 the sequence $(\alpha_n)_{n\in\mathbb{N}}$ is equally tight in $\mathcal{P}_2(\mathfrak{C})$; since it is also uniformly bounded there exists a subsequence $k \mapsto n_k$ such that α_{n_k} weakly converges to a limit $\alpha \in \mathcal{P}_2(\mathfrak{C})$. Since α_n is concentrated on $\mathfrak{C}[\sqrt{2M}]$ we also have $\lim_{k\to\infty} \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_{n_k}, \alpha) = 0$ and therefore $\mathfrak{h}^2\alpha = \mu$, $\lim_{k\to\infty} \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_{n_k}', \alpha) = 0$.

We thus have

$$\lim_{k\to\infty} \int_X \zeta(x) \,\mathrm{d}\mu_{n_k} = \lim_{k\to\infty} \int_\mathfrak{C} \zeta(\mathsf{x})\mathsf{r}^2 \,\mathrm{d}\alpha_{n_k}' = \int_\mathfrak{C} \zeta(\mathsf{x})\mathsf{r}^2 \,\mathrm{d}\alpha = \int_X \zeta(x) \,\mathrm{d}\mu$$

which contradicts (7.45).

In order to prove the converse implication, let us suppose that μ_n is converging weakly to μ in $\mathcal{M}(X)$. If μ is the null measure, then $\lim_{n\to\infty} \mu_n(X) = 0$ so that $\lim_{n\to\infty} \mathsf{HK}(\mu_n, \mu) = 0$ by (7.28).

So we can suppose that $m := \mu(X) > 0$ and, definitely, $m_n := \mu_n(X) \ge m/2 > 0$. We can then consider the measures $\alpha_n, \alpha \in \mathcal{P}(\mathfrak{C})$ given by

$$\alpha_n := \mathfrak{p}_\sharp\Big(m_n^{-1}\mu_n \otimes \delta_{\sqrt{m_n}}\Big), \quad \alpha := \mathfrak{p}_\sharp\Big(m^{-1}\mu \otimes \delta_{\sqrt{m}}\Big).$$

Since $\mathfrak{h}^2\alpha_n = \mu_n$ and $\mathfrak{h}^2\alpha = \mu$, by (7.27) we have $\mathsf{HK}(\mu_n, \mu) \le \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_n, \alpha)$. Since $m_n^{-1}\mu_n$ is weakly converging to $m^{-1}\mu$ in $\mathcal{P}(X)$ and $m_n \to m$, it is easy to check that $m_n^{-1}\mu_n \otimes \delta_{\sqrt{m_n}}$ weakly converges to $m^{-1}\mu \otimes \delta_{\sqrt{m}}$ in $\mathcal{P}(Y)$ and therefore α_n weakly converges to α in $\mathcal{P}(\mathfrak{C})$ by the continuity of the projection $\mathfrak{p}$. In order to prove that $\mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_n, \alpha) \to 0$ it is then sufficient to prove the convergence of their quadratic moments with respect to the vertex $\mathfrak{o}$: this is is immediate since

$$\lim_{n\to\infty} \int \mathsf{d}_\mathfrak{C}^2(\mathfrak{y}, \mathfrak{o}) \,\mathrm{d}\alpha_n = \lim_{n\to\infty} \int \mathsf{r}^2 \,\mathrm{d}\alpha_n = \lim_{n\to\infty} m_n = m = \int \mathsf{d}_\mathfrak{C}^2(\mathfrak{y}, \mathfrak{o}) \,\mathrm{d}\alpha. \qquad \square$$

Corollary 7.17 (Compactness)**.** *If (X, d) is a compact metric space then $(\mathcal{M}(X), \mathsf{HK})$ is a proper metric space, i.e. every bounded set is relatively compact.*

Proof. It is sufficient to notice that a set $\mathcal{C} \subset \mathcal{M}(X)$ is bounded w.r.t. HK if and only if $\sup_{\mu\in\mathcal{C}} \mu(X) < \infty$. $\square$

Theorem 7.18 (Completeness of $(\mathcal{M}(X), \mathsf{HK})$)**.** *If (X, d) is complete than the metric space $(\mathcal{M}(X), \mathsf{HK})$ is complete.*

Proof. We have to prove that every Cauchy sequence $(\mu_n)_{n\in\mathbb{N}}$ in $(\mathcal{M}(X), \mathsf{HK})$ admits a convergent subsequence. By exploiting the Cauchy property, we can find an increasing sequence of integers $k \mapsto n(k)$ such that $\mathsf{HK}(\mu_m, \mu'_m) \le 2^{-k}$ whenever $m, m' \ge n(k)$ and we consider the subsequence $\mu'_i := \mu_{n(i)}$. By choosing $\theta_i := 2^{i-1}$ so that

$$\begin{aligned}\Theta_N^2 &= \Big(\sum_{i=1}^N \theta_i^{-1}\Big)\Big(\theta_1\mu_1(X) + \sum_{i=2}^N \theta_i \mathsf{HK}^2(\mu_{n(i)}, \mu_{n(i-1)})\Big)\\ &\le 2\Big(\mu_1(X) + \sum_{i=2}^\infty 2^{-i+1}\Big) \le 2(\mu_1(X)+1),\end{aligned}$$

and by applying the gluing Lemma 7.12, for every $N > 0$ we can find measures $\alpha_i^N \in \mathcal{P}_2(\mathfrak{C})$, $i = 1, \cdots, N$, concentrated on $\mathfrak{C}[\Theta]$ with $\Theta := 2(\mu_1(X) + 1)$, such that $\mathfrak{h}^2\alpha_i^N = \mu'_i$ and

$$\mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_i^N, \alpha_{i-1}^N) = \mathsf{HK}(\mu'_i, \mu'_{i-1}).$$

For every i the sequence $N \mapsto \alpha_i^N \in \mathcal{P}_2(\mathfrak{C})$ is tight by Lemma 7.3 and it is concentrated on the bounded set $\mathfrak{C}[\Theta]$, so that by Prokhorov Theorem it is relatively compact in $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_\mathfrak{C}})$.

By a standard diagonal argument, we can find a further increasing subsequence $m \mapsto N(m)$ and limit measures $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$ such that $\lim_{m\to\infty} \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_i^{N(m)}, \alpha_i) = 0$. The convergence with respect to $\mathsf{W}_{\mathsf{d}_\mathfrak{C}}$ yields that

$$\mathfrak{h}^2\alpha_i = \mu_i, \quad \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_i, \alpha_{i-1}) = \mathsf{HK}(\mu'_i, \mu'_{i-1}) \le 2^{i-1}.$$

It follows that $i \mapsto \alpha_i$ is a Cauchy sequence in $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_\mathfrak{C}})$ which is a complete metric space and therefore there exists $\alpha \in \mathcal{P}_2(\boldsymbol{Y})$ such that $\lim_{i\to\infty} \mathsf{W}_{\mathsf{d}_\mathfrak{C}}(\alpha_i, \alpha) = 0$. Setting $\mu := \mathfrak{h}^2\alpha \in \mathcal{M}(X)$ we thus obtain $\lim_{i\to\infty} \mathsf{HK}(\mu'_i, \mu) = 0$. □

We conclude this section by proving a simple comparison estimate of HK with the *Bounded Lipschitz* metric (cf. [17, Sec. 11.3]) defined via

$$\mathsf{BL}(\mu_1, \mu_2) := \sup\Big\{\int \zeta \,\mathrm{d}(\mu_1 - \mu_2) : \zeta \in \mathrm{Lip}_b(X), \quad \sup_X |\zeta| + \mathrm{Lip}(\zeta, X) \le 1\Big\}; \tag{7.46}$$

we do not claim that the constant C below is optimal.

Proposition 7.19. *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{BL}(\mu_1, \mu_2) \le C\Big(\sum_i \mu_i(X)\Big)^{1/2} \mathsf{HK}(\mu_1, \mu_2), \quad C := \sqrt{2 + \pi^2/2}. \tag{7.47}$$

Proof. Let $\xi \in \mathrm{Lip}_b(X)$ with $\sup_X |\xi| + \mathrm{Lip}(\xi, X) \le 1$ and let $\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{\mathfrak{C}})$ optimal for (7.24) and concentrated on $\boldsymbol{\mathfrak{C}}[R]$ with $R^2 := \mu_1(X_1) + \mu_2(X_2)$. Notice that

$$|\xi(x_1) - \xi(x_2)| \le \max(\mathsf{d}(x_1, x_2), 2) \le 2\mathsf{d}_2(x_1, x_2) \le 2\mathsf{d}_\pi(x_1, x_2) \le 2\pi \sin(\mathsf{d}_\pi(x_1, x_2)/2)$$

We consider the function $\zeta : \mathfrak{C} \to \mathbb{R}$ defined by $\zeta(\mathfrak{y}) := \xi(\mathsf{x})\mathsf{r}^2$. Hence, ζ satisfies

$$\begin{aligned}\Big|\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)\Big| &\le |\xi(\mathsf{x}_1) - \xi(\mathsf{x}_2)|\mathsf{r}_1\mathsf{r}_2 + \big(|\xi(\mathsf{x}_1)|\mathsf{r}_1 + |\xi(\mathsf{x}_2)|\mathsf{r}_2\big)|\mathsf{r}_1 - \mathsf{r}_2| \\ &\le 2\pi \sin(\mathsf{d}_\pi(\mathsf{x}_1, \mathsf{x}_2)/2)\mathsf{r}_1\mathsf{r}_2 + (\mathsf{r}_1 + \mathsf{r}_2)|\mathsf{r}_1 - \mathsf{r}_2| \\ &\overset{(7.3)}{\le} \sqrt{(\mathsf{r}_1 + \mathsf{r}_2)^2 + \pi^2\mathsf{r}_1\mathsf{r}_2}\ \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) \le C\sqrt{\mathsf{r}_1^2 + \mathsf{r}_2^2}\, \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\end{aligned}$$

Since the optimal plan $\boldsymbol{\alpha}$ is concentrated on $\{\mathsf{r}_1^2 + \mathsf{r}_2^2 \le R^2\}$ we obtain

$$\begin{aligned}\Big|\int_X \xi \,\mathrm{d}(\mu_1 - \mu_2)\Big| &= \Big|\int \zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha}\Big| \le \int |\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)|\,\mathrm{d}\boldsymbol{\alpha} \\ &\le CR \int \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} \le CR\,\mathsf{HK}(\mu_1, \mu_2). \qquad \square\end{aligned}$$

7.6 Hellinger-Kantorovich distance and Entropy-Transport functionals

In this section we will establish our main result connecting HK with LET.

It is clear that the definition of HK does not change if we replace the distance d of X by its truncation $\mathsf{d}_\pi = \mathsf{d} \wedge \pi$. It is less obvious that we can even replace the threshold π with $\pi/2$ and use the distance $\mathsf{d}_{\pi/2,\mathfrak{C}}$ of Remark 7.2 in the formulation of the Hellinger-Kantorovich Problem 7.4. This property is related to the particular structure of the homogeneous marginals (which are not affected by masses concentrated in the vertex $\mathfrak{o}$ of the cone $\mathfrak{C}$) and it will provide the crucial piece of information to connect the HK and the LET functionals.

In order to prove it, we consider the partition $(\boldsymbol{\mathfrak{C}}', \boldsymbol{\mathfrak{C}}'')$ of $\boldsymbol{\mathfrak{C}} = \mathfrak{C} \times \mathfrak{C}$, where $\boldsymbol{\mathfrak{C}}'$ is defined by (7.10) and $\boldsymbol{\mathfrak{C}}'' := \boldsymbol{\mathfrak{C}} \setminus \boldsymbol{\mathfrak{C}}' = \big\{\mathsf{d}_{\pi/2,\mathfrak{C}} = \mathsf{d}_{\mathfrak{C}}\big\}$. Observe that

$$\boldsymbol{\mathfrak{C}}''_{\mathfrak{o}} := \boldsymbol{\mathfrak{C}}'' \cap (\mathfrak{C}_{\mathfrak{o}} \times \mathfrak{C}_{\mathfrak{o}}) = \big\{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}_{\mathfrak{o}} \times \mathfrak{C}_{\mathfrak{o}} : \mathsf{d}(\mathsf{x}_1, \mathsf{x}_2) \le \pi/2\big\} \tag{7.48}$$

and $\boldsymbol{\alpha}(\boldsymbol{\mathfrak{C}}') = 0$ if and only if $\boldsymbol{\alpha}_{\mathfrak{o}} = \boldsymbol{\alpha} \llcorner (\mathfrak{C}_{\mathfrak{o}} \times \mathfrak{C}_{\mathfrak{o}})$ is concentrated on $\boldsymbol{\mathfrak{C}}''_{\mathfrak{o}}$. We introduce the continuous maps

$$\mathfrak{g}_i : \boldsymbol{\mathfrak{C}} \to \boldsymbol{\mathfrak{C}}, \quad \mathfrak{g}_1(\mathfrak{y}_1, \mathfrak{y}_2) := (\mathfrak{y}_1, \mathfrak{o}), \quad \mathfrak{g}_2(\mathfrak{y}_1, \mathfrak{y}_2) := (\mathfrak{o}, \mathfrak{y}_2). \tag{7.49}$$

Lemma 7.20 (Plan restriction)**.** *For every $\boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{\mathfrak{C}})$ the plan*

$$\bar{\boldsymbol{\alpha}} := \boldsymbol{\alpha}'' + (\mathfrak{g}_1)_\sharp \boldsymbol{\alpha}' + (\mathfrak{g}_2)_\sharp \boldsymbol{\alpha}' \quad \textit{with} \quad \boldsymbol{\alpha}' := \boldsymbol{\alpha} \llcorner \boldsymbol{\mathfrak{C}}', \quad \boldsymbol{\alpha}'' := \boldsymbol{\alpha} \llcorner \boldsymbol{\mathfrak{C}}'', \tag{7.50}$$

is concentrated on $\boldsymbol{\mathfrak{C}}''$, has the same homogeneous marginals of $\boldsymbol{\alpha}$, i.e. $\mathfrak{h}_i^2 \bar{\boldsymbol{\alpha}} = \mathfrak{h}_i^2 \boldsymbol{\alpha}$, and

$$\int_{\boldsymbol{\mathfrak{C}}} \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\bar{\boldsymbol{\alpha}} = \int_{\boldsymbol{\mathfrak{C}}} \mathsf{d}_{\pi/2,\mathfrak{C}}^2 \,\mathrm{d}\bar{\boldsymbol{\alpha}} \le \int_{\boldsymbol{\mathfrak{C}}} \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\alpha}, \tag{7.51}$$

where the inequality is strict if $\boldsymbol{\alpha}(\boldsymbol{\mathfrak{C}}') > 0$. In particular for every $\mu_1, \mu_2 \in \mathcal{M}(X)$

$$\mathsf{HK}^2(\mu_1, \mu_2) = \min\Big\{\int \mathsf{d}_{\pi/2,\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}_2(\boldsymbol{\mathfrak{C}}),\ \mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i\Big\}. \tag{7.52}$$

Proof. For every $\zeta \in \mathrm{B}_b(X)$, since $\mathsf{r}_1 \circ \mathfrak{g}_2 = 0$ and $\mathsf{r}_1 \circ \mathfrak{g}_1 = \mathsf{r}_1$, we have

$$\int \zeta \,\mathrm{d}(\mathfrak{h}_1^2 \bar{\boldsymbol{\alpha}}) = \int \zeta(\mathsf{x}_1)\mathsf{r}_1^2 \,\mathrm{d}\bar{\boldsymbol{\alpha}} = \int \zeta(\mathsf{x}_1)\mathsf{r}_1^2 \,\mathrm{d}\boldsymbol{\alpha}'' + \sum_k \int \zeta(\mathsf{x}_1(\mathfrak{g}_k))\mathsf{r}_1(\mathfrak{g}_k)^2 \,\mathrm{d}\boldsymbol{\alpha}'$$
$$= \int \zeta(\mathsf{x}_1)\mathsf{r}_1^2 \,\mathrm{d}\boldsymbol{\alpha}'' + \int \zeta(\mathsf{x}_1)\mathsf{r}_1^2 \,\mathrm{d}\boldsymbol{\alpha}' = \int \zeta(\mathsf{x}_1)\mathsf{r}_1^2 \,\mathrm{d}\boldsymbol{\alpha} = \int \zeta \,\mathrm{d}(\mathfrak{h}_1^2 \boldsymbol{\alpha}),$$

so that $\mathfrak{h}_1^2 \bar{\boldsymbol{\alpha}} = \mathfrak{h}_1^2 \boldsymbol{\alpha}$; a similar calculation holds for $\mathfrak{h}_2^2$ so that $\bar{\boldsymbol{\alpha}} \in \mathfrak{H}^2_{=}(\mu_1, \mu_2)$. Moreover, if $(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}'$ we easily get

$$\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) > \mathsf{r}_1^2 + \mathsf{r}_2^2 = \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{g}_1(\mathfrak{y}_1, \mathfrak{y}_2)) + \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{g}_2(\mathfrak{y}_1, \mathfrak{y}_2))$$

so that whenever $\boldsymbol{\alpha}(\mathfrak{C}') > 0$ we get

$$\int \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\bar{\boldsymbol{\alpha}} = \int \big(\mathsf{d}_{\mathfrak{C}}^2 \circ \mathfrak{g}_1 + \mathsf{d}_{\mathfrak{C}}^2 \circ \mathfrak{g}_2\big) \,\mathrm{d}\boldsymbol{\alpha}' + \int \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\alpha}'' < \int \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\alpha}' + \int \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\alpha}'' = \int \mathsf{d}_{\mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\alpha},$$

which proves (7.51) and characterizes the equality case. (7.52) then follows by (7.51) and the fact that the homogeneous marginals of $\bar{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$ coincide. □

We introduce the open set

$$\mathfrak{G} := \Big\{([x_1, r_1], [x_2, r_2]) \in \mathfrak{C} : r_1 r_2 \neq 0,\ \mathsf{d}(x_1, x_2) < \pi/2\Big\}.$$

Notice that $\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) > 0$ in $\mathfrak{G}$.

Theorem 7.21 ($\mathsf{HK}^2 = \mathsf{LET}$)**.** *For every $\mu_i \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}^2(\mu_1, \mu_2) = \mathsf{LET}(\mu_1, \mu_2), \tag{7.53}$$

and any optimal solution $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ of Problem 7.4 or of (7.30a,b) *satisfies $\boldsymbol{\alpha}(\mathfrak{C}') = 0$. Moreover*

(i) $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ is an optimal solution of (7.30a,b) *if and only if $\boldsymbol{\alpha}(\mathfrak{C}') = 0$ and $\mathsf{y}_\sharp(\boldsymbol{\alpha} \llcorner \mathfrak{C}_o \times \mathfrak{C}_o)$ is an optimal plan for* (6.29)–(6.30)*.*

(ii) $\bar{\boldsymbol{\alpha}} \in \mathcal{M}(\boldsymbol{Y})$ is any optimal plan for (6.31) *if and only if $\boldsymbol{\alpha} := \mathfrak{p}_\sharp \bar{\boldsymbol{\alpha}}$ is an optimal solution for the Hellinger-Kantorovich Problem 7.4.*

(iii) If $\boldsymbol{\gamma} \in \mathcal{M}(X \times X)$ belongs to $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ and $\varrho_i : X \to [0, \infty)$ are Borel maps so that $\mu_i = \varrho_i \gamma_i + \mu_i^\perp$, then $\boldsymbol{\beta} := \big(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2))\big)_\sharp \boldsymbol{\gamma}$ is an optimal plan for (7.30a)–(7.30b)*, and it satisfies $\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) = 1$ $\boldsymbol{\beta}$-a.e.; in particular $\boldsymbol{\beta}$ is concentrated on $\mathfrak{G}$.*

(iv) If $\boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{Y})$ is an optimal plan for Problem 7.4 then $\tilde{\boldsymbol{\alpha}} := \boldsymbol{\alpha} \llcorner \mathfrak{G}$ is an optimal plan for (7.30a,b)*. Moreover,*
- the plan $\boldsymbol{\beta} := \mathrm{dil}_{\vartheta,2}(\tilde{\boldsymbol{\alpha}})$, with $\vartheta := \big(\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2))\big)^{1/2}$, is an optimal plan satisfying $\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) = 1$ $\boldsymbol{\beta}$-a.e.
- $\boldsymbol{\gamma} := (\mathsf{x}_1, \mathsf{x}_2)_\sharp \boldsymbol{\beta}$ belongs to $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$,
- $\boldsymbol{\beta} = \big(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2))\big)_\sharp \boldsymbol{\gamma}$.

Proof. (7.53) and the first statement immediately follow by combining the previous Lemma 7.20 with Remark 7.5 and (6.31).

If $\boldsymbol{\alpha}$ is an optimal plan for the formulation (7.30a,b) we can apply Lemma 7.10(iii) to find $\tilde{\boldsymbol{\alpha}} \geq \boldsymbol{\alpha}$ optimal for (7.21), so that $\boldsymbol{\alpha}(\mathfrak{C}') \leq \tilde{\boldsymbol{\alpha}}(\mathfrak{C}') = 0$.

Since all the optimal plans for HK do not charge $\mathfrak{C}'$, combining Lemma 7.10, Remark 7.5 and Theorems 6.3, 6.8 we can easily prove the other statements (i), (ii), (iii).

Concerning (iv), the optimality of $\tilde{\boldsymbol{\alpha}}$ is obvious from the formulation (7.30b) and the optimality of $\boldsymbol{\beta}$ follows from the invariance of (7.30b) with respect to dilations. We notice that $\boldsymbol{\beta}$-almost everywhere in $\mathfrak{G}$ we have

$$\begin{aligned}\sum_i U_0(\mathsf{r}_i^2) + \mathsf{c}(\mathsf{x}_1, \mathsf{x}_2) &= \sum_i \mathsf{r}_i^2 - 1 - \log \mathsf{r}_i^2 - \log(\cos^2(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)))\\ &= \sum_i \mathsf{r}_i^2 - 2 - 2\log(\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)))\\ &= \mathsf{r}_1^2 + \mathsf{r}_2^2 - 2\mathsf{r}_1\mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2))\end{aligned}$$

so that by (7.30a)

$$\int \Big(\sum_i U_0(\mathsf{r}_i^2) + \mathsf{c}(\mathsf{x}_1, \mathsf{x}_2)\Big)\,\mathrm{d}\boldsymbol{\beta} + \sum_i \big(\mu_i(X) - \mathfrak{h}_i^2\boldsymbol{\beta}(X)\big) = \mathsf{HK}^2(\mu_1, \mu_2). \tag{7.54}$$

Let us now set $\boldsymbol{\gamma} := (\mathsf{x}_1, \mathsf{x}_2)_\sharp\boldsymbol{\beta} \in \mathcal{M}(X \times X)$, $\beta_i := \pi^i_\sharp\boldsymbol{\beta} \in \mathcal{M}(\mathfrak{C})$, $\gamma_i := \pi^i_\sharp\boldsymbol{\gamma} = (\mathsf{x}_i)_\sharp\boldsymbol{\beta} = \mathsf{x}_\sharp\beta_i \in \mathcal{M}(X)$, $\tilde{\mu}_i := \mathfrak{h}_i^2\boldsymbol{\beta} = (\mathsf{x}_i)_\sharp(r_i^2\boldsymbol{\gamma}) = \mathsf{x}_\sharp(r^2\beta_i)$ and let us denote by $(\beta_{i,x_i})_{x_i\in X}$ the disintegration of β_i with respect to γ_i. We obviously have

$$\int \mathsf{c}(\mathsf{x}_1, \mathsf{x}_2)\,\mathrm{d}\boldsymbol{\beta} = \int \mathsf{c}(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma},$$

and for every $\zeta \in \mathrm{B}_b(X)$

$$\int_X \zeta\,\mathrm{d}\tilde{\mu}_i = \int_{\mathfrak{C}} \zeta(\mathsf{x}) r^2\,\mathrm{d}\beta_i = \int_X \Big(\int_{\mathfrak{C}} \zeta(\mathsf{x}) r^2\,\mathrm{d}\beta_{i,x}\Big)\,\mathrm{d}\gamma_i = \int_X \zeta(x)\Big(\int_{\mathfrak{C}} r^2\,\mathrm{d}\beta_{i,x}\Big)\,\mathrm{d}\gamma_i$$

so that

$$\mu_i' = \varrho_i\gamma_i \leq \mu_i, \quad \varrho_i(x) := \int_{\mathfrak{C}} r^2\,\mathrm{d}\beta_{i,x}.$$

Applying Jensen inequality we obtain

$$\begin{aligned}\int U_0(\mathsf{r}_i^2)\,\mathrm{d}\boldsymbol{\beta} = \int U_0(\mathsf{r}_i^2)\,\mathrm{d}\beta_i = \int \Big(\int U_0(r_i^2)\,\mathrm{d}\beta_{i,x_i}(r_i)\Big)\,\mathrm{d}\gamma_i \geq \int U_0\Big(\int r_i^2\,\mathrm{d}\beta_{i,x_i}(r_i)\Big)\,\mathrm{d}\gamma_i\\ = \int U_0\big(\varrho_i(x)\big)\,\mathrm{d}\gamma_i\end{aligned}$$

so that (7.54) yields

$$\mathsf{HK}^2(\mu_1, \mu_2) \geq \sum_i \int_X U_0(\varrho_i)\,\mathrm{d}\gamma_i + \int_{X\times X} \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma} + \sum_i \nu_i(X)$$

with $\nu_i := \mu_i - \mu'_i \in \mathcal{M}(X)$. Since $\mu_i = \tilde\varrho_i\gamma_i + \mu_i^\perp = \varrho_i\gamma_i + \nu_i$, we get $\nu_i = \mu_i^\perp + (\tilde\varrho_i - \varrho_i)\gamma_i \geq \mu_i^\perp$ and the monotonicity of the logarithm yield

$$
\begin{aligned}
\mathsf{HK}^2(\mu_1,\mu_2) &\geq \sum_i \Big(\int_X U_0(\varrho_i)\,\mathrm{d}\gamma_i + \nu_i(X)\Big) + \int \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma}\\
&= \sum_i \Big(\int_X \Big(U_0(\varrho_i) + \tilde\varrho_i - \varrho_i\Big)\,\mathrm{d}\gamma_i + \mu_i^\perp(X)\Big) + \int \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma}\\
&\geq \sum_i \Big(\int_X U_0(\tilde\varrho_i)\,\mathrm{d}\gamma_i + \mu_i^\perp(X)\Big) + \int \mathsf{c}\,\mathrm{d}\boldsymbol{\gamma} \geq \mathsf{LET}(\mu_1,\mu_2),
\end{aligned}
$$

where the first inequality is strict if $\nu_i \neq \mu_i^\perp$ so that $\tilde\varrho_i > \varrho_i$ on some set with positive γ_i-measure.

By the first statement of the Theorem it follows that $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$, $\tilde\varrho_i \equiv \varrho_i$ and all the inequalities are in fact identities. Since U_0 is strictly convex we also have that β_{i,x_i} is concentrated on $\sqrt{\varrho_i(x_i)}$ so that $\boldsymbol{\beta} = \big(\mathfrak{p}\circ(x_1,\varrho_1^{1/2}(x_1);x_2,\varrho_2^{1/2}(x_2))\big)_\sharp\boldsymbol{\gamma}$. □

We observe that the system $(\boldsymbol{\gamma},\varrho_1,\varrho_2)$ provided by the previous Theorem enjoys a few remarkable properties, that are not obvious from the original Hellinger-Kantorovich formulation.

a) First of all, the annihilated part $\mu_i^\perp$ of the measures μ_i is concentrated on the set

$$M_{i,j} := \{x_i \in X : \mathsf{d}(x_i, \mathrm{supp}(\mu_j)) \geq \pi/2\}$$

When $\mu_i(M_{i,j}) = 0$ then $\mu_i \ll \gamma_i$.

b) As a second property, an optimal plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1,\mu_2)$ provides an optimal plan $\boldsymbol{\alpha} = \big(\mathfrak{p}\circ(x_1,\varrho_1^{1/2}(x_1);x_2,\varrho_2^{1/2}(x_2))\big)_\sharp\boldsymbol{\gamma}$ which is concentrated on the graph of the map $(\varrho_1^{1/2}(x_1);\varrho_2^{1/2}(x_2))$ from $X\times X$ to $\mathbb{R}_+\times\mathbb{R}_+$; the maps ϱ_i are independent, in the sense that ϱ_i only depends on x_i.

c) A third important application of Theorem 7.21 is the duality formula for the HK functional which directly follows from (6.14) of Theorem 6.3. We will state it in a slightly different form in the next theorem, whose interpretation will be clearer in the light of Section 8.4. It is based on the inf-convolution formula

$$\mathscr{P}_1\xi(x) = \inf_{x'\in X} \frac{\xi(x')}{1+2\xi(x')} + \frac{\sin^2(\mathsf{d}_{\pi/2}(x,x'))}{2(1+2\xi(x'))} = \inf_{x'\in X}\frac{1}{2}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x,x'))}{1+2\xi(x')}\Big). \tag{7.55}$$

where $\xi \in \mathrm{B}(X)$ with $\xi > -1/2$.

Theorem 7.22 (Duality formula for HK)**.**

(i) If $\xi \in \mathrm{B}_b(X)$ with $\inf_X \xi > -1/2$ then the function $\mathscr{P}_1\xi$ defined by (7.55) belongs to $\mathrm{Lip}_b(X)$, satisfies $\sup_X \mathscr{P}_1\xi < 1/2$, and admits the equivalent representation

$$\mathscr{P}_1\xi(x) = \inf_{x'\in B_{\pi/2}(x)} \frac{1}{2}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x,x'))}{1+2\xi(x')}\Big). \tag{7.56}$$

In particular if ξ has bounded support than $\mathscr{P}_1\xi \in \mathrm{Lip}_{bs}(X)$, the space of Lipschitz functions with bounded support.

(ii) Let us suppose that (X, d) is a separable metric space and τ is induced by d. For every $\mu_0, \mu_1 \in \mathcal{M}(X)$ we have

$$\frac{1}{2}\mathsf{HK}^2(\mu_0,\mu_1) = \sup\Big\{ \int \mathscr{P}_1\xi \,\mathrm{d}\mu_1 - \int \xi \,\mathrm{d}\mu_0 : \xi \in \mathrm{Lip}_{bs}(X),\ \inf_X \xi > -1/2 \Big\}. \quad (7.57)$$

Proof. Let us first observe that if

$$-\frac{1}{2} < a \le \xi \le b \text{ in } X \quad \Rightarrow \quad \frac{a}{1+2a} \le \mathscr{P}_1\xi \le \frac{b}{1+2b} \text{ in } X. \quad (7.58)$$

Since $1/(1+2\xi(x')) \le 1/(1+2a)$ for every $x' \in X$, the function $\mathscr{P}_1\xi$ is also Lipschitz, because it is the infimum of a family of uniformly Lipschitz functions.

Since

$$\frac{1}{2}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x,x'))}{1+\xi(x')}\Big) = \frac{1}{2} > \frac{b}{1+2b} \quad \text{if } \mathsf{d}(x,x') \ge \pi/2, \quad (7.59)$$

we immediately obtain (7.56). In particular, we have

$$\xi \equiv 0 \quad \text{in } X \setminus B \quad \Rightarrow \quad \mathscr{P}_1\xi \equiv 0 \quad \text{in } \{x \in X : \mathsf{d}(x,B) \ge \pi/2\}. \quad (7.60)$$

Let us now prove statement (ii). We denote by E the the right-hand side of (7.57) and by E' the analogous expression where ξ runs in $\mathrm{C}_b(X)$:

$$E' := 2\sup\Big\{ \int \mathscr{P}_1\xi \,\mathrm{d}\mu_1 - \int \xi \,\mathrm{d}\mu_0 : \xi \in \mathrm{C}_b(X),\ \inf_X \xi > -1/2 \Big\}. \quad (7.61)$$

It is clear that $E' \ge E$. If $\xi \in \mathrm{C}_b(X)$ with $\inf \xi > -1/2$, setting $\psi_1(x_1) := -2\xi(x_1)$, $\psi_2(x_2) := 2(\mathscr{P}_1\xi)(x_2)$, we know that $\sup_X \psi_2 < 1$ and $\psi_2 \in \mathrm{Lip}_b(X)$. Thus, ψ_1 and ψ_2 are continuous and satisfy

$$\big(1-\psi_2(x_2)\big)\big(1-\psi_1(x_1)\big) \ge \cos^2(\mathsf{d}_{\pi/2}(x_1,x_2)).$$

The couple (ψ_1,ψ_2) is therefore admissible for (6.14) (with $\mathrm{C}_b(X)$ instead of $\mathrm{LSC}_s(X)$: notice that τ is metrizable and therefore completely regular), so that $\mathsf{HK}^2(\mu_0,\mu_1) \ge E'$.

On the other hand, if $(\psi_1,\psi_2) \in \mathrm{C}_b(X) \times \mathrm{C}_b(X)$ with $\sup_X \psi_i < 1$, setting $\xi_1 = -\frac{1}{2}\psi_1$ and $\tilde\xi_2 := \mathscr{P}_1(-\xi_1)$ we see that $2\tilde\xi_2 \ge \psi_2$ giving $E' \ge \mathsf{HK}^2(\mu_0,\mu_1)$, and $E = E'$ follows.

To show that $E = E'$ in the general case, we approximate $\psi \in \mathrm{C}_b(X)$ with $\inf_X \psi > -1$ by a decreasing sequence of Lipschitz and bounded functions (e.g. by taking $\psi_n(x) := \sup_y \psi(y) - n\mathsf{d}_\pi(x,y))$ and use that the supremum in (7.61) does not change if we restrict it to $\mathrm{Lip}_b(X)$.

Let now ξ be Lipschitz and valued in $[a,b]$ with $-1/2 < a \le 0 \le b$. Taking the increasing sequence of nonnegative cut-off functions $\zeta_n(x) := 0 \vee \big(n - \mathsf{d}(x,\bar x)\big) \wedge 1$ which are uniformly 1-Lipschitz, have bounded support and satisfy $\zeta_n \uparrow 1$ as $n \to \infty$, it is easy to check that $\xi_n := \zeta_n\xi$ belong to $\mathrm{Lip}_{bs}(X)$ and take values in the interval $[a,b]$ so that $\frac{a}{1+2a} \le \mathscr{P}_1\xi_n \le \frac{b}{1+2b}$ for every $n \in \mathbb{N}$.

Since $\xi_n(x) = 0$ if $\mathsf{d}(x,\bar x) \ge n$ and $\xi_n(x) = \xi(x)$ if $\mathsf{d}(x,\bar x) \le n-1$, by (7.56) we get

$$\mathscr{P}_1\xi_n(x) = 0 \quad \text{if } x \ge n + \pi/2, \quad \mathscr{P}_1\xi_n(x) = \mathscr{P}_1\xi(x) \quad \text{if } x < n - 1 - \pi/2. \quad (7.62)$$

Thus $\mathscr{P}_1\xi_n \in \mathrm{Lip}_{bs}(X)$ and the Lebesgue Dominated Convergence theorem shows that

$$\lim_{n\to\infty} \int_X \mathscr{P}_1\xi_n \,\mathrm{d}\mu_1 - \int_X \xi_n \,\mathrm{d}\mu_0 = \int_X \mathscr{P}_1\xi \,\mathrm{d}\mu_1 - \int_X \xi \,\mathrm{d}\mu_0. \qquad \square$$

7.7 Limiting cases: recovering the Hellinger and the Kantorovich–Wasserstein distance

In this section we will show that we can recover the Hellinger-Kakutani and the Kantorovich-Wasserstein distance by suitably rescaling the HK functional.

The Hellinger-Kakutani distance. As we have seen in Example E.5 of Section 3.3, the Hellinger-Kakutani distance between two measures $\mu_1, \mu_2 \in \mathcal{M}(X)$ can be obtained as a limiting case when the space X is endowed with the discrete distance

$$\mathsf{d}_{\mathsf{Hell}}(x_1, x_2) := \begin{cases} a & \text{if } x_1 \neq x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases} \qquad \text{with } a \in [\pi, +\infty]. \tag{7.63}$$

The induced cone distance in this case is

$$\mathsf{d}^2_{\mathfrak{C}}([x_1, r_1], [x_2, r_2]) = \begin{cases} (r_1 - r_2)^2 & \text{if } x_1 = x_2, \\ r_1^2 + r_2^2 & \text{if } x_1 \neq x_2. \end{cases} \tag{7.64}$$

and the induced cost function for the Entropy-Transport formalism is given by

$$\mathsf{c}_{\mathsf{Hell}}(x_1, x_2) := \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases} \tag{7.65}$$

Recalling (3.24)–(3.25) we get

$$\mathsf{Hell}^2(\mu_1, \mu_2) = \mathsf{LET}_{\mathsf{Hell}}(\mu_1, \mu_2) = \int_X \left(\sqrt{\varrho_1} - \sqrt{\varrho_2}\right)^2 \mathrm{d}\gamma \quad \gamma \in \mathcal{M}(X),\ \mu_i = \varrho_i \gamma \ll \gamma. \tag{7.66}$$

Since $\mathsf{c}_{\mathsf{Hell}} \geq \mathsf{c} = \ell(\mathsf{d})$ for every distance function on X, we always have the upper bound

$$\mathsf{HK}(\mu_1, \mu_2) \leq \mathsf{Hell}(\mu_1, \mu_2) \quad \text{for every } \mu_1, \mu_2 \in \mathcal{M}(X). \tag{7.67}$$

Applying Lemma 3.9 we easily get

Theorem 7.23 (Convergence of HK to Hell)**.** *Let (X, τ, d) be an extended metric topological space and let $\mathsf{HK}_{\lambda\mathsf{d}}$ be the Hellinger-Kantorovich distances in $\mathcal{M}(X)$ induced by the distances $\mathsf{d}_\lambda := \lambda\mathsf{d}$, $\lambda > 0$. For every couple $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}_{\lambda\mathsf{d}}(\mu_1, \mu_2) \uparrow \mathsf{Hell}(\mu_1, \mu_2) \quad \text{as } \lambda \uparrow \infty. \tag{7.68}$$

The Kantorovich–Wasserstein distance. Let us first observe that whenever $\mu_1, \mu_2 \in \mathcal{M}(X)$ have the same mass their HK-distance is always bounded form above by the Kantorovich-Wasserstein distance W_d (the upper bound is trivial when $\mu_1(X) \neq \mu_2(X)$, since in this case $\mathsf{W}_\mathsf{d}(\mu_1, \mu_2) = +\infty$).

Proposition 7.24. *For every couple $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}(\mu_1, \mu_2) \leq \mathsf{W}_{\mathsf{d}_{\pi/2}}(\mu_1, \mu_2) \leq \mathsf{W}_\mathsf{d}(\mu_1, \mu_2). \tag{7.69}$$

Proof. It is not restrictive to assume that $\mathsf{W}^2_{\mathsf{d}_{\pi/2}}(\mu_1,\mu_2)=\int \mathsf{d}^2_{\pi/2}\boldsymbol{\gamma}<\infty$ for an optimal plan $\boldsymbol{\gamma}$ with marginals μ_i. We then define the plan $\boldsymbol{\alpha}:=\mathfrak{s}_\sharp\boldsymbol{\gamma}\in\mathcal{M}(\mathfrak{C}\times\mathfrak{C})$ where $\mathfrak{s}(x_1,x_2):=([x_1,1],[x_2,1])$, so that $\mathfrak{h}^2_i\boldsymbol{\alpha}=\mu_i$; by using (7.52) and (7.3) we obtain

$$\mathsf{HK}^2(\mu_1,\mu_2)\le 4\int_{\mathfrak{C}}\sin^2(\mathsf{d}_{\pi/2}(\mathsf{x}_1,\mathsf{x}_2)/2)\,\mathrm{d}\boldsymbol{\alpha}\le\int_{\boldsymbol{X}}\mathsf{d}^2_{\pi/2}(x_1,x_2)\,\mathrm{d}\boldsymbol{\gamma}\le\mathsf{W}^2_{\mathsf{d}_{\pi/2}}(\mu_1,\mu_2).\qquad\square$$

In order to recover the Kantorovich-Wasserstein distance we perform a simultaneous scaling, by taking the limit of $n\mathsf{HK}_{\mathsf{d}/n}$ where $\mathsf{HK}_{\mathsf{d}/n}$ is induced by the distance d/n.

Theorem 7.25 (Convergence of HK to W)**.** *Let (X,τ,d) be an extended metric topological space and let $\mathsf{HK}_{\mathsf{d}/\lambda}$ be the Hellinger-Kantorovich distances in $\mathcal{M}(X)$ induced by the distances $\lambda^{-1}\mathsf{d}$, $\lambda>0$. For every couple $\mu_1,\mu_2\in\mathcal{M}(X)$ we have*

$$\lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1,\mu_2)\uparrow\mathsf{W}_{\mathsf{d}}(\mu_1,\mu_2)\quad\textit{as }\lambda\uparrow\infty.\tag{7.70}$$

Proof. Let us denote by $\mathsf{LET}_\lambda=\mathsf{HK}^2_{\mathsf{d}/\lambda}$ the optimal value of the LET-problem associated to the distance d/λ. Since the Kantorovich-Wasserstein distance is invariant by the rescaling $\lambda\mathsf{W}_{\mathsf{d}/\lambda}=\mathsf{W}_{\mathsf{d}}$, (7.69) shows that $\lambda\mathsf{HK}_{\mathsf{d}/\lambda}\le\mathsf{W}_{\mathsf{d}}$.

Since $x\mapsto\sin(x\wedge\pi/2)$ is concave in $[0,\infty)$, the function $x\mapsto\sin(x\wedge\pi/2)/x$ is decreasing in $[0,\infty)$, so that $\alpha\sin((d/\alpha)\wedge\pi/2)\le\lambda\sin((d/\lambda)\wedge\pi/2)$ for every $d\ge 0$ and $\alpha<\lambda$. Combining (7.52) with (7.9b) we obtain that the map $\lambda\mapsto\lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1,\mu_2)$ is nondecreasing.

It remains to prove that $L:=\lim_{\lambda\to\infty}\lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1,\mu_2)=\sup_{\lambda\ge1}\lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1,\mu_2)\ge\mathsf{W}_{\mathsf{d}}(\mu_1,\mu_2)$; it is not restrictive to assume that L is finite.

Let $\boldsymbol{\gamma}_\lambda$ be an optimal plan for $\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1,\mu_2)$ with marginals $\gamma_{\lambda,i}=\pi^i_\sharp\boldsymbol{\gamma}_\lambda$. We denote by $\mathscr{F}$ the entropy functionals associated to logarithmic entropy $F(s)=U_1(s)$ and by $\mathscr{G}$ the entropy functionals associated to $F(s):=\mathrm{I}_1(s)$ as in Example E.3 of Section 3.3. Since the transport part of the LET-functional is associated to the costs

$$\mathsf{c}_\lambda(x_1,x_2)=\lambda^2\ell(\mathsf{d}(x_1,x_2)/\lambda)\overset{(6.6)}{\ge}\mathsf{d}^2(x_1,x_2),$$

we obtain the estimate

$$L^2\ge\lambda^2\mathsf{LET}_\lambda(\mu_1,\mu_2)\ge\sum_i\lambda^2\mathscr{F}(\gamma_{\lambda,i}|\mu_i)+\int_{\boldsymbol{X}}\mathsf{d}^2(x_1,x_2)\,\mathrm{d}\boldsymbol{\gamma}_\lambda.\tag{7.71}$$

Proposition 2.10 shows that the family of plans $(\boldsymbol{\gamma}_\lambda)_{\lambda\ge1}$ is relatively compact with respect to narrow convergence in $\mathcal{M}(X\times X)$. Since $\lambda^2F(s)\uparrow\mathrm{I}_1(s)$, passing to the limit along a suitable subnet $(\lambda(\alpha))_{\alpha\in\mathbb{A}}$ parametrized by a directed set $\mathbb{A}$, and applying Corollary 2.9 we get a limit plan $\boldsymbol{\gamma}\in\mathcal{M}(X\times X)$ with marginals γ_i such that

$$\sum_i\mathscr{G}(\gamma_i|\mu_i)=0,\quad\text{i.e.}\quad\gamma_i=\mu_i.\tag{7.72}$$

(7.72) yields in particular that $\mu_1(X)=\mu_2(X)$. Since d is lower semicontinuous, narrow convergence of $\boldsymbol{\gamma}_{\lambda(\alpha)}$ and (7.71) also yield

$$L^2\ge\liminf_{\alpha\in\mathbb{A}}\int_{\boldsymbol{X}}\mathsf{d}^2(x_1,x_2)\,\mathrm{d}\boldsymbol{\gamma}_{\lambda(\alpha)}\ge\int_{\boldsymbol{X}}\mathsf{d}^2(x_1,x_2)\,\mathrm{d}\boldsymbol{\gamma}\ge\mathsf{W}^2_{\mathsf{d}}(\mu_1,\mu_2).\qquad\square$$

7.8 The Gaussian Hellinger-Kantorovich distance

We conclude this general introduction to the Hellinger-Kantorovich distance by discussing another interesting example.

We consider the inverse function $g : \mathbb{R}_+ \to [0, \pi/2)$ of $\sqrt{\ell}$:

$$g(z) := \arccos(\mathrm{e}^{-z^2/2}), \quad \text{satisfying} \quad g(0) = 0,\ g'(0) = 1,\ \ell(g(d)) = d^2; \tag{7.73}$$

since $\sqrt{\ell}$ is a convex function, g is a concave increasing function in $[0, \infty)$ with $g(z) \le z$ and $\lim_{z\to\infty} g(z) = \pi/2$.

It follows that $\mathsf{g} := g \circ \mathsf{d}$ is a distance in X, inducing the same topology of d. We can thus introduce the HK_g distance associated to g. The corresponding distance on $\mathfrak{C}$ is given by

$$\mathsf{g}_\mathfrak{C}(\mathfrak{y}_1, \mathfrak{y}_2) := \mathsf{r}_1^2 + \mathsf{r}_2^2 - 2\mathsf{r}_1\mathsf{r}_2 \exp(-\mathsf{d}^2(\mathsf{x}_1, \mathsf{x}_2)/2). \tag{7.74}$$

Notice that $\mathsf{g}_\mathfrak{C} \le \mathsf{d}_\mathfrak{C}$.

Theorem 7.26 (The Gaussian Hellinger-Kantorovich distance)**.** *The functional*

$$\mathsf{GHK}^2(\mu_1, \mu_2) := \mathsf{HK}_\mathsf{g}^2(\mu_1, \mu_2) = \min\Big\{ \int \mathsf{g}_\mathfrak{C}^2(\mathfrak{y}_1, \mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}(\boldsymbol{\mathfrak{C}}) :\ \mathfrak{h}_i^2\boldsymbol{\alpha} = \mu_i \Big\} \tag{7.75}$$

defines a distance on $\mathcal{M}(X)$ dominated by HK. If (X, d) is separable (resp. complete) then $(\mathcal{M}(X), \mathsf{GHK})$ is a separable (resp. complete) metric space, whose topology coincides with the weak convergence. We also have

$$\begin{aligned} \mathsf{GHK}^2(\mu_1, \mu_2) &= \min\Big\{ \sum_i \mathscr{F}(\gamma_i|\mu_i) + \int_{\boldsymbol{X}} \mathsf{d}^2(x_1, x_2)\,\mathrm{d}\boldsymbol{\gamma} :\ \boldsymbol{\gamma} \in \mathcal{M}(\boldsymbol{X}) \Big\} \\ &= \sup\Big\{ \sum_i \int \Big(1 - \mathrm{e}^{-\varphi_i}\Big)\,\mathrm{d}\mu_i :\ \varphi_1 \oplus \varphi_2 \le \mathsf{d}^2 \Big\}. \end{aligned} \tag{7.76}$$

We shall see in the next Section 8.2 that HK is the length distance induced by GHK.

8 Dynamic interpretation of the Hellinger-Kantorovich distance

8.1 Absolutely continuous curves and geodesics in the cone $\mathfrak{C}$

Absolutely continuous curves and metric derivative. If (Z, d_Z) is a (possibly extended) metric space and I is an interval of $\mathbb{R}$, a curve $\mathrm{z} : I \to Z$ is absolutely continuous if there exists $m \in \mathrm{L}^1(I)$ such that

$$\mathsf{d}_Z(\mathrm{z}(t_0), \mathrm{z}(t_1)) \le \int_{t_0}^{t_1} m(t)\,\mathrm{d}t \quad \text{whenever } t_0, t_1 \in I,\ t_0 < t_1. \tag{8.1}$$

Its metric derivative $|\mathrm{z}'|_{\mathsf{d}_Z}$ (we will omit the index d_Z when the choice of the metric is clear from the context) is the Borel function defined by

$$|\mathrm{z}'|_{\mathsf{d}_Z}(t) := \limsup_{h\to 0} \frac{\mathsf{d}_Z(\mathrm{z}(t+h), \mathrm{z}(t))}{|h|} \tag{8.2}$$

and it is possible to show (see [2]) that the lim sup above is in fact a limit for $\mathscr{L}^1$-a.e. points in I and it provides the minimal (up to possible modifications in $\mathscr{L}^1$-negligible sets) function m for which (8.1) holds. We will denote by $\mathrm{AC}^p(I;Z)$ the class of all absolutely continuous curves $\mathrm{z} : I \to Z$ with $|\mathrm{z}'| \in \mathrm{L}^p(I)$; when I is an open set of $\mathbb{R}$, we will also consider the local space $\mathrm{AC}^p_{loc}(I;Z)$. If Z is complete and separable then $\mathrm{AC}^p([0,1];Z)$ is a Borel set in the space $\mathrm{C}([0,1];Z)$ endowed with the topology of uniform convergence (this property can be extended to the framework of extended metric-topological spaces, see [3]).

$\mathrm{z} : [0,1] \to Z$ is a (minimal, constant speed) geodesic if

$$\mathsf{d}_Z(\mathrm{z}(t_0), \mathrm{z}(t_1)) = |t_1 - t_0| \mathsf{d}_Z(\mathrm{z}(0), \mathrm{z}(1)) \quad \text{for every } t_0, t_1 \in [0,1]. \tag{8.3}$$

In particular z is Lipschitz and $|\mathrm{z}'| \equiv 1$ in $[0,1]$. We denote by $\mathrm{Geo}(Z) \subset \mathrm{C}([0,1];Z)$ the closed subset of all the geodesics.

A metric space (Z, d_Z) is called a length (or intrinsic) space if the distance between arbitrary couples of points can be obtained as the infimum of the length of the absolutely continuous curves connecting them. It is called a geodesic (or strictly intrinsic) space if every couple of points z_0, z_1 at finite distance can be joined by a geodesic.

Geodesics in $\mathfrak{C}$. If (X, d) is a geodesic (resp. length) space, then also $\mathfrak{C}$ is a geodesic (resp. length) space [9, Sec. 3.6]. A geodesic connecting a point $\mathfrak{y} = [x, r]$ with $\mathfrak{o}$ is just

$$\mathfrak{y}(t) = [x, tr] = \mathfrak{y} \cdot t, \quad t \in [0,1]. \tag{8.4}$$

If $x_1, x_2 \in X$ with $\mathsf{d}(x_1, x_2) \geq \pi$, then a geodesic between $\mathfrak{y}_i = [x_i, r_i]$ can be easily obtained by joining two geodesics connecting $\mathfrak{y}_i$ to $\mathfrak{o}$ as before; observe that in this case $\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) = r_1 + r_2$.

In the case when $\mathsf{d}(x_1, x_2) < \pi$ and $r_1, r_2 > 0$, every geodesic $\mathfrak{y} : I \to \mathfrak{C}$ connecting $\mathfrak{y}_1$ to $\mathfrak{y}_2$ is associated to a geodesic x in X joining x_1 to x_2 and parametrized with unit speed in the interval $[0, \mathsf{d}(x_1, x_2)]$: we write the curve in $\mathbb{C}$ connecting $z_1 = r_1$ to $z_2 = r_2 \exp(i\mathsf{d}(x_1, x_2))$ given by

$$\mathrm{z}(t) = r(t)\exp(i\theta(t)), \qquad \begin{cases} r^2(t) = (1-t)^2 r_1^2 + t^2 r_2^2 + 2t(1-t) r_1 r_2 \cos(\mathsf{d}(x_1, x_2)) \\ \cos(\theta(t)) = \dfrac{(1-t) r_1 + t r_2 \cos(\mathsf{d}(x_1, x_2))}{r(t)}, \quad \theta(t) \in [0, \pi], \end{cases} \tag{8.5}$$

and then we have

$$\mathfrak{y}(t) = [\mathrm{x}(\theta(t)), r(t)]. \tag{8.6}$$

Absolutely continuous curves in $\mathfrak{C}$. We want to obtain now a simple characterizations of absolutely continuous curves in $\mathfrak{C}$. If $t \mapsto \mathfrak{y}(t)$ is a continuous curve in $\mathfrak{C}$, with $t \in [0,1]$, is clear that $\mathrm{r}(t) := \mathsf{r}(\mathfrak{y}(t))$ is a continuous curve with values in $[0, \infty)$; we can then consider the open set $O_{\mathrm{r}} = \mathrm{r}^{-1}\big((0,\infty)\big)$ and the map $\mathrm{x} : [0,1] \to X$ defined by $\mathrm{x}(t) := \mathsf{x}(\mathfrak{y}(t))$, whose restriction to O_{r} is also continuous. Thus any continuous curve $\mathfrak{y} : I \to \mathfrak{C}$ can be lifted to a couple of maps $\mathrm{y} = \mathsf{y} \circ \mathfrak{y} = (\mathrm{x}, \mathrm{r}) : [0,1] \to Y$ with r continuous and x continuous on O_{r} and constant on its complement. Conversely, is clear that starting from a couple $\mathrm{y} = (\mathrm{x}, \mathrm{r})$ as above, then $\mathfrak{y} = \mathfrak{p} \circ \mathrm{y}$ is continuous in $\mathfrak{C}$. We thus introduce the set

$$\widetilde{\mathrm{C}}([0,1]; Y) := \big\{\mathrm{y} = (\mathrm{x}, \mathrm{r}) : [0,1] \to Y : \mathrm{r} \in \mathrm{C}([0,1]; \mathbb{R}_+), \quad \mathrm{x}|_{O_\mathrm{r}} \text{ is continuous}\big\} \tag{8.7}$$

and the analogous

$$\begin{aligned}\widetilde{\mathrm{AC}}^p([0,1]; Y) := \Big\{&\mathrm{y} = (\mathrm{x}, \mathrm{s}) : \mathrm{r} \in \mathrm{AC}^p([0,1]; \mathbb{R}_+), \\ &\qquad \mathrm{x}|_{O_\mathrm{r}} \in \mathrm{AC}^p_{loc}(O_\mathrm{r}; X), \quad \mathrm{r}|\mathrm{x}'| \in \mathrm{L}^p(O_\mathrm{r})\Big\}.\end{aligned} \tag{8.8}$$

If $\mathrm{y} = (\mathrm{x}, \mathrm{r}) \in \widetilde{\mathrm{AC}}^p([0,1]; Y)$ we define the Borel map $|\mathrm{y}'| : [0,1] \to \mathbb{R}_+$ by

$$|\mathrm{y}'|^2(t) := |\mathrm{r}'(t)|^2 + \mathrm{r}^2(t)|\mathrm{x}'|^2_{\mathsf{d}}(t) \quad \text{if } t \in O_\mathrm{r}, \quad |\mathrm{y}'|(t) = 0 \text{ otherwise.} \tag{8.9}$$

For absolutely continuous curves the following characterization holds:

Lemma 8.1. *Let* $\mathfrak{y} \in \mathrm{C}([0,1]; \mathfrak{C})$ *lifted to* $\mathrm{y} = \mathsf{y} \circ \mathfrak{y} \in \widetilde{\mathrm{C}}([0,1]; Y)$. $\mathfrak{y} \in \mathrm{AC}^p(I; \mathfrak{C})$ *if and only if* $\mathrm{y} = (\mathrm{x}, \mathrm{r}) \in \widetilde{\mathrm{AC}}^p([0,1]; Y)$ *and*

$$|\mathfrak{y}'|_{\mathsf{d}_\mathfrak{C}}(t) = |\mathrm{y}'|(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in [0,1]. \tag{8.10}$$

Proof. By (7.4) one immediately gets that if $\mathfrak{y} = \mathfrak{p} \circ \mathrm{y} \in \mathrm{AC}^p([0,1]; \mathfrak{C})$ then s belongs to $\mathrm{AC}^p([0,1]; \mathbb{R})$ and $\mathrm{x} \in \mathrm{AC}^p_{\mathrm{loc}}(O_\mathrm{r}; X)$. Since $\mathfrak{y}$ is absolutely continuous, we can evaluate the metric derivative at a.e. $t \in O_\mathrm{r}$ where also r' and $|\mathrm{x}'|$ exist: starting from (7.3) we get the limit

$$\begin{aligned}\lim_{h \downarrow 0} \frac{\mathsf{d}^2_\mathfrak{C}(\mathfrak{y}(t+h), \mathfrak{y}(t))}{h^2} &= \lim_{h \downarrow 0} \frac{|\mathrm{r}(t+h) - \mathrm{r}(t)|^2 + 4\mathrm{r}(t+h)\mathrm{r}(t)\sin^2(\frac{1}{2}\mathsf{d}_\pi(\mathrm{x}(t+h), \mathrm{x}(t)))}{h^2} \\ &= |\mathrm{r}'(t)|^2 + s(t)|\mathrm{x}'|^2_\mathsf{d}(t)\end{aligned}$$

which provides (8.10).

On the other hand, the same calculations show that if the lifting y belongs to $\widetilde{\mathrm{AC}}^p([0,1]; Y)$ then the restriction of $\mathfrak{y}$ to each connected component of O_r is absolutely continuous with metric velocity given by (8.10) in $\mathrm{L}^p(0,1)$. Since $\mathfrak{y}$ is globally continuous and constant in $[0,1] \setminus O_\mathrm{r}$, we conclude that $\mathfrak{y} \in \mathrm{AC}^p([0,1]; \mathfrak{C})$. □

As a consequence, in a length space, we get the variational representation formula

$$\begin{aligned}\mathsf{d}^2_\mathfrak{C}(\mathfrak{y}_0, \mathfrak{y}_1) = \inf \Big\{ \int_{[0,1]\cap\{\mathrm{r}>0\}} &\Big(\mathrm{r}^2(t)|\mathrm{x}'|^2_\mathsf{d}(t) + |\mathrm{r}'(t)|^2\Big)\,\mathrm{d}t : \\ &(\mathrm{x}, \mathrm{r}) \in \widetilde{\mathrm{AC}}^2([0,1]; Y), \ [\mathrm{x}(i), \mathrm{r}(i)] = \mathfrak{y}_i, \ i = 0, 1\Big\}.\end{aligned} \tag{8.11}$$

Remark 8.2 (The Euclidean case)**.** If $X = \mathbb{R}^d$ with the usual Euclidean distance $\mathsf{d}(x_1, x_2) := |x_1 - x_2|$ and $\mathfrak{y} = [\mathrm{x}, \mathrm{r}] \in \mathrm{AC}^2([0,1]; \mathfrak{C})$, we can define a Borel vector field $\mathfrak{y}'_\mathfrak{C} : [0,1] \to \mathbb{R}^{d+1}$ by

$$\mathfrak{y}'_\mathfrak{C}(t) := \begin{cases} (\mathrm{r}(t)\mathrm{x}'(t), \mathrm{r}'(t)) & \text{whenever } \mathrm{r}(t) \neq 0 \text{ and the derivatives exist,} \\ (0,0) & \text{otherwise.} \end{cases} \tag{8.12}$$

Then, (8.10) yields $|\mathfrak{y}'|_{\mathsf{d}_{\mathfrak{C}}}(t) = |\mathfrak{y}'_{\mathfrak{C}}(t)|_{\mathbb{R}^{d+1}}$ for $\mathscr{L}^1$-a.e. $t \in (0,1)$. If moreover $\psi \in \mathrm{C}^1(\mathbb{R}^d \times [0,1])$ and $\zeta([x,r],t) := \frac{1}{2}\psi(x,t)r^2$, $\partial_t\zeta([x,r],t) := \frac{1}{2}\partial_t\psi(x,t)r^2$, by defining the Borel map $\mathrm{D}_{\mathfrak{C}}\zeta : \mathfrak{C} \to (\mathbb{R}^{d+1})^*$

$$\mathrm{D}_{\mathfrak{C}}\zeta(\mathfrak{y},t) := \begin{cases} (\frac{1}{2}r\mathrm{D}_x\psi(x,t), r\psi(x,t)) & \text{if } \mathfrak{y} \neq \mathfrak{o}, \\ (0,0) & \text{otherwise,} \end{cases} \tag{8.13}$$

then the map $t \mapsto \zeta(\mathfrak{y}(t),t)$ is absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(\mathfrak{y}(t),t) = \frac{1}{2}\partial_t\zeta(\mathfrak{y}(t),t) + \langle \mathrm{D}_{\mathfrak{C}}\zeta(\mathfrak{y}(t),t), \mathfrak{y}'_{\mathfrak{C}}(t)\rangle_{\mathbb{R}^{d+1}} \quad \mathscr{L}^1\text{-a.e. in } (0,1). \qquad \square \tag{8.14}$$

8.2 Lifting of absolutely continuous curves and geodesics

Dynamic plans and time-dependent marginals Let (Z,d_Z) be a complete and separable metric space. A dynamic plan $\boldsymbol{\pi}$ in Z is a probability measure in $\mathcal{P}(\mathrm{C}(I;Z))$; we say that $\boldsymbol{\pi}$ has finite 2-energy if it is concentrated on $\mathrm{AC}^2(I;Z)$ and

$$\int \Big(\int_0^1 |\mathrm{z}'|^2_{\mathsf{d}_Z}(t)\,\mathrm{d}t \Big)\,\mathrm{d}\boldsymbol{\pi}(\mathrm{z}) < \infty. \tag{8.15}$$

We denote by e_t the evaluation map in $\mathrm{C}(I;Z)$ given by $\mathsf{e}_t(\mathrm{z}) := \mathrm{z}(t)$. If $\boldsymbol{\pi}$ is a dynamic plan, $\alpha_t = (\mathsf{e}_t)_\sharp\boldsymbol{\pi} \in \mathcal{M}(Z)$ is its marginal at time $t \in I$ and the curve $t \mapsto \alpha_t$ belongs to $\mathrm{C}(I;(\mathcal{M}(Z),\mathsf{W}_{\mathsf{d}_Z}))$. If moreover $\boldsymbol{\pi}$ is a dynamic plan with finite 2-energy, then $\alpha \in \mathrm{AC}^2(I;(\mathcal{M}(Z),\mathsf{W}_{\mathsf{d}_Z}))$.

We say that $\boldsymbol{\pi}$ is an *optimal geodesic plan* between $\alpha_0,\alpha_1 \in \mathcal{P}(Z)$ if it is a dynamic plan concentrated on $\mathrm{Geo}(Z)$ and

$$(\mathsf{e}_i)_\sharp\boldsymbol{\pi} = \alpha_i,\ i=0,1, \quad \int \mathsf{d}^2_Z(\mathrm{z}(0),\mathrm{z}(1))\,\mathrm{d}\boldsymbol{\pi}(\mathrm{z}) = \iint_0^1 |\mathrm{z}'|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\mathrm{z}) = \mathsf{W}^2_{\mathsf{d}_Z}(\alpha_0,\alpha_1). \tag{8.16}$$

When $Z = \mathfrak{C}$ we will denote by $\mathfrak{h}^2_t = \mathfrak{h}^2 \circ (\mathsf{e}_t)_\sharp$ the homogeneous marginal at time $t \in I$. Since $\mathfrak{h}^2$ is 1-Lipschitz, it follows that the curve $\mu_t := \mathfrak{h}^2\alpha_t = \mathfrak{h}^2_t\boldsymbol{\pi}$ belongs to $\mathrm{AC}^2(I;(\mathcal{M}(X),\mathsf{HK}))$ and moreover

$$|\mu'_t|^2_{\mathsf{HK}} \le \int |\mathfrak{y}'|^2_{\mathsf{d}_{\mathfrak{C}}}(t)\,\mathrm{d}\boldsymbol{\pi}(\mathfrak{y}) \quad \text{for a.e. } t \in (0,1). \tag{8.17}$$

A simple consequence of this property is that $(\mathcal{M}(X),\mathsf{HK})$ inherits the length (or geodesic) property of (X,d).

Proposition 8.3. $(\mathcal{M}(X),\mathsf{HK})$ *is a length (resp. geodesic) space if and only if* (X,d) *is a length (resp. geodesic) space.*

Proof. Let us first suppose that (X,d) is a length space (the argument in the geodesic case is completely equivalent) and let $\mu_i \in \mathcal{M}(X)$. By Corollary 7.8 we find $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$ such that $\mathfrak{h}^2\alpha_i = \mu_i$ and $\mathsf{HK}(\mu_1,\mu_2) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1,\alpha_2)$. Since $\mathfrak{C}$ is a length space, it is well known [41] that $\mathcal{P}_2(\mathfrak{C})$ is length so that for every $\kappa > 1$ there exists $\alpha \in \mathrm{Lip}([0,1];(\mathcal{P}_2(\mathfrak{C}),\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}))$

connecting α_1 to α_2 such that $|\alpha'|_{\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}} \leq \kappa\,\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2)$. Setting $\mu_t := \mathfrak{h}^2\alpha_t$ we obtain a Lipschitz curve connecting μ_1 to μ_2 with length less than $\kappa\,\mathsf{HK}(\mu_1, \mu_2)$.

The converse property is a consequence of the next representation Theorem 8.4 and the fact that if $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}})$ is length (resp. geodesic) then $\mathfrak{C}$ and thus X are length (resp. geodesic). □

We want to prove a converse representation result; the argument only depends on the metric properties of the Lipschitz submersion $\mathfrak{h}$.

Theorem 8.4. *Let $(\mu_t)_{t\in[0,1]}$ be an absolutely continuous curve in* $\mathrm{AC}([0,1]; (\mathcal{M}(X), \mathsf{HK}))$ *with*

$$\Theta^2 := 2\Big(\mu_0(X) + \int_0^1 |\mu'|^2_{\mathsf{HK}}\,\mathrm{d}t\Big). \tag{8.18}$$

Then there exists a dynamic plan $\boldsymbol{\pi} \in \mathcal{P}(\mathrm{AC}^2([0,1];\mathfrak{C}))$ and a curve $\alpha \in \mathrm{AC}^2([0,1]; (\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}))$ with $\alpha_t = (\mathsf{e}_t)_\sharp\boldsymbol{\pi}$ concentrated on $\mathfrak{C}[\Theta]$, such that

$$\mu_t = \mathfrak{h}^2_t\boldsymbol{\pi} = \mathfrak{h}^2\alpha_t \ \textit{in } [0,1], \quad |\mu_t'|^2_{\mathsf{HK}} = |\alpha_t'|^2_{\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}} = \int |\mathfrak{y}'|^2_{\mathsf{d}_{\mathfrak{C}}}(t)\,\mathrm{d}\boldsymbol{\pi}(\mathfrak{y}) \quad \textit{for a.e. } t \in (0,1). \tag{8.19}$$

Proof. By Lisini's lifting Theorem [30, Theorem 5] it is sufficient to prove that there exists a curve $\alpha \in \mathrm{AC}^2([0,1]; (\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}))$ such that $\mu_t = \mathfrak{h}(\alpha_t)$ and $|\mu_t'| = |\alpha_t'|$ a.e. in $(0,1)$. By a standard reparametrization technique, we may assume that μ is Lipschitz continuous and $|\mu_t'| = L$.

We divide the interval $I = [0,1]$ in 2^N-intervals $I_i^N := [t_{i-1}^N, t_i^N]$, $t_i^N := i\,2^{-N}$, $i = 1, \cdots, 2^N$ of size 2^{-N} and we set $\mu_i^N := \mu_{t_i^N}$.

We can apply the Gluing Lemma 7.12 (starting from $i = 0$ to 2^N with $\theta_0 := 1$, $\theta_i = 2^N$ for $i \geq 1$) to obtain measures $\alpha_i^N \in \mathcal{P}_2(\mathfrak{C})$ such that

$$\mathfrak{h}(\alpha_i^N) = \mu_i^N, \quad W_{\mathsf{d}_{\mathfrak{C}}}(\alpha_i^N, \alpha_{i+1}^N) = \mathsf{HK}(\mu_i^N, \mu_{i+1}^N) \leq L2^{-N}, \tag{8.20}$$

and concentrated on $\mathfrak{C}[\Theta_N]$ where

$$\Theta_N^2 = 2\Big(\mu_0(X) + 2^N \sum_{i=1}^{2^N} \mathsf{HK}^2(\mu_{i-1}^N, \mu_i^N)\Big) \leq 2\Big(\mu_0(X) + 2^N \sum_{i=1}^{2^N} 2^{-N} \int_{I_i^N} |\mu'|^2_{\mathsf{HK}}\,\mathrm{d}t\Big) = \Theta^2.$$

Thus if t is a dyadic point, we obtain a sequence of probability measures $\alpha^N(t) \in \mathcal{P}_2(\mathfrak{C})$ concentrated on $\mathfrak{C}[\Theta]$ with $\mathfrak{h}(\alpha^N(t)) = \mu_t$ and such that $W_{\mathsf{d}_{\mathfrak{C}}}(\alpha^N(t), \alpha^N(s)) \leq L|t-s|$ if $s = m2^{-N}, t = n2^{-N}$ are dyadic points in the same grid. By the compactness lemma 7.3 and a standard diagonal argument, we can extract a subsequence $N(k)$ such that $\alpha_{N(k)}(t)$ converges to $\alpha(t)$ in $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}})$ for every dyadic point t. Since $W_{\mathsf{d}_{\mathfrak{C}}}(\alpha(s), \alpha(t)) \leq |t-s|$ for every dyadic s, t, we can extend α to a L-Lipschitz curve, still denoted by α, which satisfies $\mathfrak{h}(\alpha(t)) = \mu(t)$. Since $\mathfrak{h}$ is 1-Lipschitz, we conclude that $|\alpha_t'| = |\mu_t'|$ a.e. in $(0,1)$. □

Corollary 8.5. *Let $(\mu_t)_{t\in[0,1]}$ be an absolutely continuous curve in* $\mathrm{AC}^2([0,1]; (\mathcal{M}(X), \mathsf{HK}))$ *and let Θ as in* (8.18)*. Then there exists a dynamic plan $\tilde{\boldsymbol{\pi}}$ in $\mathcal{P}(\widetilde{\mathrm{C}}([0,1];Y))$ concentrated* *on* $\widetilde{\mathrm{AC}}^2([0,1];Y)$ *such that* $\alpha_t = (\mathsf{e}_t)_\sharp\boldsymbol{\pi}$ *is concentrated in* $X \times [0,\Theta]$, $\mu_t = \mathrm{h}^2((\mathsf{e}_t)_\sharp\boldsymbol{\pi})$ *and, recalling* (8.9),

$$|\mu_t'|^2_{\mathsf{HK}} = \int |\mathrm{y}'|^2(t)\,\mathrm{d}\boldsymbol{\pi}(\mathrm{y}) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in [0,1]. \tag{8.21}$$

Remark 8.6. Since every absolutely continuous curve $\mu \in \mathrm{AC}([0,1];(\mathcal{M}(X),\mathsf{HK}))$ admits a Lipschitz reparametrization (see, e.g., [2, Lemma 1.1.4]), by applying Theorem 8.4 it is not difficult to check that there exists a curve $\alpha \in \mathrm{AC}([0,1];(\mathcal{P}_2(\mathfrak{C}),\mathsf{W}_{\mathsf{d}_\mathfrak{C}}))$ such that $\mu_t = \mathfrak{h}^2\alpha_t$ and $|\mu_t'|_{\mathsf{HK}} = |\alpha_t'|_{\mathsf{W}_{\mathsf{d}_\mathfrak{C}}}$ for every $t \in [0,1]$. □

Another important consequence of the previous representation result is a precise characterization of the geodesics in $(\mathcal{M}(X),\mathsf{HK})$.

Theorem 8.7 (Geodesics in $(\mathcal{M}(X),\mathsf{HK})$)**.**

(i) If $(\mu_t)_{t\in[0,1]}$ *is a geodesic in* $(\mathcal{M}(X),\mathsf{HK})$ *then there exists an optimal geodesic plan* $\boldsymbol{\pi}$ *in* $\mathcal{P}(\mathrm{Geo}(\mathfrak{C}))$ *(recall* (8.16)*) such that*
(a) $\boldsymbol{\pi}$*-a.e. curve* $\mathfrak{y}$ *is a geodesic in* $\mathfrak{C}$*,*
(b) $[0,1] \ni t \mapsto \alpha_t := (\mathsf{e}_t)_\sharp\boldsymbol{\pi}$ *is a geodesic in* $(\mathcal{P}_2(\mathfrak{C}),\mathsf{W}_{\mathsf{d}_\mathfrak{C}})$*, where all* α_t *are concentrated on* $\mathfrak{C}[\Theta]$*, where* $\Theta^2 = 2(\mu_0(X) + \mathsf{HK}^2(\mu_0,\mu_1))$*,*
(c) $\mu_t = \mathfrak{h}^2_t\boldsymbol{\pi} = \mathfrak{h}^2\alpha_t$ *for every* $t \in [0,1]$ *and* $(\mathsf{e}_s,\mathsf{e}_t)_\sharp\boldsymbol{\pi} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_s,\mu_t)$ *if* $0 \le s < t \le 1$*.*

(ii) If (X,d) *is a geodesic space, for every* $\mu_0,\mu_1 \in \mathcal{M}(X)$ *and every* $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_0,\mu_1)$ *there exists an optimal geodesic plan* $\boldsymbol{\pi} \in \mathcal{P}(\mathrm{Geo}(\mathfrak{C}))$ *such that* $(\mathsf{e}_s,\mathsf{e}_t)_\sharp\boldsymbol{\pi} = \boldsymbol{\alpha}$*.*

Proof. The statement (i) is an immediate consequence of Theorem 8.4 to the geodesic case.

Statement (ii) is a well known property of the Kantorovich-Wasserstein space $(\mathfrak{C},\mathsf{W}_{\mathsf{d}_\mathfrak{C}})$ in the case when $\mathfrak{C}$ is geodesic. □

Theorem 8.4 also clarifies the relation between GHK and HK.

Corollary 8.8. *If* (X,d) *is separable and complete then* $\mathrm{AC}^2([0,1];(\mathcal{M}(X),\mathsf{GHK}))$ *coincides with* $\mathrm{AC}^2([0,1];(\mathcal{M}(X),\mathsf{HK}))$ *and for every curve* $\mu \in \mathrm{AC}^2([0,1];(\mathcal{M}(X),\mathsf{GHK}))$ *we have*

$$|\mu'|_{\mathsf{GHK}}(t) = |\mu'|_{\mathsf{HK}}(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in [0,1]. \tag{8.22}$$

In particular if (X,d) *is a length metric space then* HK *is the length distance generated by* GHK*.*

Proof. Since $\mathsf{GHK} \le \mathsf{HK}$ it is clear that $\mathrm{AC}^2([0,1];(\mathcal{M}(X),\mathsf{HK})) \subset \mathrm{AC}^2([0,1];(\mathcal{M}(X),\mathsf{GHK}))$.

In order to prove the opposite inclusion and (8.22) it is sufficient to notice that the classes of absolutely continuous curves in $\mathfrak{C}$ w.r.t. $\mathsf{d}_\mathfrak{C}$ and $\mathsf{g}_\mathfrak{C}$ coincide with equal metric derivatives $|\mathfrak{y}'|_{\mathsf{d}_\mathfrak{C}} = |\mathfrak{y}'|_{\mathsf{g}_\mathfrak{C}}$. Since $\mathsf{GHK} = \mathsf{HK}_\mathsf{g}$ is the Hellinger-Kantorovich distance induced by g, the thesis follows by (8.19) of Theorem 8.4. □

8.3 Lower curvature bound in the sense of Alexandrov

Let us first recall two possible definitions of Positively Curved (PC) space in the sense of Alexandrov, referring to [9] and to [10] for other equivalent definitions and for the more general case of spaces with curvature $\geq k$.

According to Sturm [40], a metric space (Z, d_Z) is a Positively Curved (PC) metric space in the large if for every choice of points $z_0, z_1, \cdots, z_N \in Z$ and coefficients $\lambda_1, \cdots, \lambda_N \in (0, +\infty)$ we have

$$\sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{d}_Z^2(z_i, z_j) \leq 2 \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{d}_Z^2(z_0, z_j) \tag{8.23}$$

If every point of Z has a neighborhood which is PC then we say that Z is locally positively curved.

When the space Z is geodesic, the above (local and global) definitions coincide with the corresponding one given by Alexandrov (based on triangle comparison): for every choice of $z_0, z_1, z_2 \in Z$ and every point z_t such that $\mathsf{d}_Z(z_t, z_i) = |i - t| \mathsf{d}_Z(z_0, z_1)$ we have

$$\mathsf{d}_Z^2(z_2, z_t) \geq (1-t)\mathsf{d}_Z^2(z_2, z_0) + t\mathsf{d}_Z^2(z_2, z_1) - 2t(1-t)\mathsf{d}_Z^2(z_0, z_1). \tag{8.24}$$

When Z is also complete, the local and the global definition are equivalent.

Theorem 8.9. *Let (X, d) be a metric space.*

(i) If $X \subset \mathbb{R}$ is convex (i.e. an interval) endowed with the standard distance, then $(\mathcal{M}(X), \mathsf{HK})$ is a PC space.

(ii) If $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ is a PC space in the large according to (8.23) then $(\mathcal{M}(X), \mathsf{HK}(X))$ is a PC space.

(iii) If (X, d) is separable, complete and geodesic, then $(\mathcal{M}(X), \mathsf{HK}(X))$ is a PC space if and only if (X, d) has locally curvature ≥ 1.

Before we go into the proof of this result, we highlight that for a compact convex subset $\Omega \subset \mathbb{R}^d$ with $d \geq 2$ equipped with the Euclidean distance, the space $(\mathcal{M}(X), \mathsf{HK}(X))$ is not PC, see [27, Sect. 5.6] for an explicit construction showing the semiconcavity of the squared distance fails.

Proof. Let us first prove statement (ii). If $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ is a PC space then also $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}})$ is a PC space [41]. Applying Corollary 7.14, for every choice of $\mu_i \in \mathcal{M}(X)$, $i = 0, \ldots, N$, we can then find measures $\beta_i \in \mathcal{P}_2(\mathfrak{C})$ such that

$$\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\beta_0, \beta_i) = \mathsf{HK}(\beta_0, \beta_i) \quad i = 1, \cdots, N; \tag{8.25}$$

here it is crucial that β_0 is the same for every i. It then follows that

$$\sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{HK}^2(\beta_i, \beta_j) \leq \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}^2(\beta_i, \beta_j) \leq 2 \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}^2(\beta_0, \beta_i) = 2 \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{HK}^2(\beta_0, \beta_i).$$

Let us now consider (iii) "⇒": If $(\mathcal{M}(X), \mathsf{HK})$ is PC, we have to prove that (X, d) has locally curvature ≥ 1. By Theorem [9, Thm. 4.7.1] it is sufficient to prove that $\mathfrak{C} \setminus \{\mathfrak{o}\}$ is locally PC to conclude that $(\mathfrak{C}, \mathsf{d})$ has locally curvature ≥ 1. We thus select points $\mathfrak{y}_i = [x_i, r_i]$, $i = 0, 1, 2$, in a sufficiently small neighborhood of $\mathfrak{y} = [x, r]$ with $r > 0$, so that $\mathsf{d}(x_i, x_j) < \pi/2$ for every i, j and $r_i, r_j > 0$. We also consider a geodesic $\mathfrak{y}_t = [x_t, s_t]$, $t \in [0, 1]$, connecting $\mathfrak{y}_0$ to $\mathfrak{y}_1$, thus satisfying $\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_t, \mathfrak{y}_i) = |i - t| \mathsf{d}(\mathfrak{y}_0, \mathfrak{y}_1)$, $i = 0, 1$.

Setting $\mu_i := r_i \delta_{x_i}$, $\mu_t := s_t \delta_{x_t}$, it is easy to check (cf. [27, Sect. 3.3.1]) that

$$\mathsf{HK}(\mu_i, \mu_j) = \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_j),\ i, j \in \{0, 1, 2\}, \quad \mathsf{HK}(\mu_t, \mu_i) = |i - t| \mathsf{HK}(\mu_0, \mu_1). \tag{8.26}$$

We can thus apply (8.24) to $\mu_0, \mu_1, \mu_2, \mu_t$ and obtain the corresponding inequality for $\mathfrak{y}_0, \mathfrak{y}_1, \mathfrak{y}_2, \mathfrak{y}_t$.

(iii) "⇐": In order to prove the converse property we apply Remark 7.13: $\mu_0, \mu_1, \mu_2, \mu_3 = \mu_t \in \mathcal{M}(X)$, $t \in [0, 1]$, with $\mathsf{HK}(\mu_3, \mu_i) = |i - t| \mathsf{HK}(\mu_0, \mu_1)$, we find a plan $\boldsymbol{\alpha} \in \mathcal{P}(X_0 \times X_1 \times X_2 \times X_3)$ (with the usual convention to use copies of X) such that

$$\mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i, \qquad \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_j)\, \mathrm{d}\boldsymbol{\alpha} = \mathsf{HK}^2(\mu_i, \mu_j) \quad \text{for } (i, j) \in A = \{(0, 3), (1, 3), (2, 3)\}. \tag{8.27}$$

The triangle inequality, the elementary inequality $t(1-t)(a+b)^2 \leq (1-t)a^2 + tb^2$, and the very definition of HK yield for $t \in (0, 1)$

$$\begin{aligned} t(1-t)\mathsf{HK}^2(\mu_0, \mu_1) &\leq t(1-t) \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1)\, \mathrm{d}\boldsymbol{\alpha} \leq \int t(1-t)\big((\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) + \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1)\big)^2\, \mathrm{d}\boldsymbol{\alpha} \\ &\leq \int (1-t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_3) + t\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_3, \mathfrak{y}_1)\, \mathrm{d}\boldsymbol{\alpha} = (1-t)\mathsf{HK}^2(\mu_0, \mu_3) + t\mathsf{HK}^2(\mu_3, \mu_1) \\ &= t(1-t)\mathsf{HK}^2(\mu_0, \mu_1). \end{aligned}$$

This series of inequalities shows in particular that

$$(1-t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_3) + t\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_3, \mathfrak{y}_1) = t(1-t)\big(\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) + \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1)\big)^2 = t(1-t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \quad \boldsymbol{\alpha}\text{-a.e.}$$

so that

$$\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) = t\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1), \quad \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1) = (1-t)\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) \quad \boldsymbol{\alpha}\text{-a.e.},$$

and that $\pi_\sharp^{\mathfrak{y}_0, \mathfrak{y}_1} \boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_0, \mu_1)$, so that (8.27) holds for $(i, j) \in A' = A \cup \{(0, 1)\}$.

By Theorem 7.21 we deduce that

$$\mathsf{d}(\mathsf{x}_i, \mathsf{x}_j) \leq \pi/2 \quad \boldsymbol{\alpha}\text{-a.e. for } (i, j) \in A'.$$

If one of the point $\mathfrak{y}_i$, $i = 0, 1, 2$, is the vertex $\mathfrak{o}$, then it is not difficult to check by a direct computation that

$$\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_3) \geq (1-t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_0) + t\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_1) - 2t(1-t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1). \tag{8.28}$$

When $\mathfrak{y}_i \in \mathfrak{C} \setminus \{\mathfrak{o}\}$ for every $i = 0, 1, 2$, since $\mathsf{d}(\mathsf{x}_0, \mathsf{x}_1) + \mathsf{d}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{d}(\mathsf{x}_2, \mathsf{x}_0) \leq \frac{3}{2}\pi < 2\pi$, Theorem [9, Thm. 4.7.1] yields (8.28) thanks to the curvature assumption on X. Integrating (8.28) w.r.t. $\boldsymbol{\alpha}$, by taking into account (8.27), the fact that $(\pi^0, \pi^1)_\sharp \boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_0, \mu_1)$, and

$$\int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_i)\, \mathrm{d}\boldsymbol{\alpha} \geq \mathsf{HK}^2(\mu_2, \mu_i), \quad i = 0, 1,$$

we get

$$\mathsf{HK}^2(\mu_2,\mu_3) \geq (1-t)\mathsf{HK}^2(\mu_2,\mu_0) + t\mathsf{HK}^2(\mu_2,\mu_1) - 2t(1-t)\mathsf{HK}^2(\mu_0,\mu_1).$$

Finally, statement (i) is just a particular case of (iii). □

As simple applications of the Theorem above we obtain that $\mathcal{M}(\mathbb{R})$ and $\mathcal{M}(\mathbb{S}^{d-1})$ endowed with HK are Positively Curved spaces.

8.4 Duality and Hamilton-Jacobi equation

In this section we will show the connections of the duality formula of Theorem 7.22 with Lipschitz subsolutions of the Hamilton-Jacobi equation in $X \times (0,1)$

$$\partial_t \xi_t + \frac{1}{2}|\mathrm{D}_X \xi_t|^2 + 2\xi_t^2 = 0 \tag{8.29}$$

and its counterpart in the cone space

$$\partial_t \zeta_t + \frac{1}{2}|\mathrm{D}_{\mathfrak{C}} \zeta_t|^2 = 0. \tag{8.30}$$

At a formal level, it is not difficult to check that solutions to (8.29) corresponds to the special class of solutions to (8.30) of the form

$$\zeta_t([x,r]) := \xi_t(x) r^2; \tag{8.31}$$

this is due to the formula (still formal at this level)

$$|\mathrm{D}_{\mathfrak{C}} \zeta|^2 = \frac{1}{r^2}|\mathrm{D}_X \zeta|^2 + |\partial_s \zeta|^2 = |\mathrm{D}_X \xi|^2 r^2 + 4\xi^2 r^2 \quad \text{if } \zeta = \xi\, r^2. \tag{8.32}$$

Since the Kantorovich-Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$ can be defined in duality with subsolutions to (8.30) via the Hopf-Lax formula and 2-homogeneous marginals are modeled on test functions as (8.31), we can expect to obtain a dual representation for the Hellinger-Kantorovich distance on $\mathcal{M}(X)$ by studying the Hopf-Lax formula for initial data of the form $\zeta_0(x,r) = \xi_0(x) r^2$.

Slope and asymptotic Lipschitz constant. In order to give a metric interpretation to (8.29) and (8.30), let us first recall that for a locally Lipschitz function $f : Z \to \mathbb{R}$ defined in a metric space (Z, d_Z) the *metric slope* $|\mathrm{D}_Z f|$ and the *asymptotic Lipschitz constant* $|\mathrm{D}_Z f|_a$ are defined by

$$|\mathrm{D}_Z f|(z) := \limsup_{x\to z} \frac{|f(x)-f(z)|}{\mathsf{d}_Z(x,z)}, \qquad |\mathrm{D}_Z f|_a(z) := \lim_{r\downarrow 0} \sup_{\substack{x,y\in B_r(z)\\ y\neq x}} \frac{|f(y)-f(x)|}{\mathsf{d}_Z(x,y)} \tag{8.33}$$

with the convention that $|\mathrm{D}_Z f|(z) = |\mathrm{D}_Z f|_a(z) = 0$ whenever z is an isolated point. $|\mathrm{D}_Z f|_a$ can also be defined as the minimal constant $L \geq 0$ such that there exists a function $G_L : Z \to [0,\infty)$ satisfying

$$|f(x) - f(y)| \leq G_L(x,y)\mathsf{d}_Z(x,y), \quad \limsup_{x,y\to z} G_L(x,y) \leq L. \tag{8.34}$$

Notice that $|\mathrm{D}_Z f|_a$ is an u.s.c function; when Z is a length space, it is the upper semicontinuous envelope of the metric slope $|\mathrm{D}_Z f|$. We will often write $|\mathrm{D} f|$, $|\mathrm{D} f|_a$ whenever the space Z will be clear from the context.

Remark 8.10. Notice that the notion of locally Lipschitz function and $|\mathrm{D}_Z f|_a$ does not change if we replace the distance d_Z with a distance $\tilde{d}_Z$ of the form

$$\tilde{\mathsf{d}}_Z(z_1,z_2) := h(\mathsf{d}_Z(z_1,z_2))\ z_1,z_2\in Z,\quad h:[0,\infty)\to[0,\infty)\text{ concave, }\lim_{r\downarrow 0}\frac{h(r)}{r}=1. \tag{8.35}$$

In particular the truncated distances $\mathsf{d}_Z\wedge\kappa$, $\kappa>0$, the distances $a\sin((\mathsf{d}_Z\wedge\kappa)/a)$, $\kappa\in(0,a\pi/2]$, and the distance $\mathsf{g}=g(\mathsf{d})$ given by (7.73) yield the same asymptotic Lipschitz constant.

In the case of the cone space $\mathfrak{C}$ it is not difficult to see that the distance $\mathsf{d}_{\mathfrak{C}}$ and $\mathsf{d}_{\pi/2,\mathfrak{C}}$ coincide in suitably small neighborhoods of every point $\mathfrak{y}\in\mathfrak{C}\setminus\{\mathfrak{o}\}$, so that they induce the same asymptotic Lipschitz constants in $\mathfrak{C}\setminus\{\mathfrak{o}\}$. The same property holds for $\mathsf{g}_{\mathfrak{C}}$. In the case of the vertex $\mathfrak{o}$ (7.10) yields

$$|\mathrm{D}_{\mathfrak{C}}f|_a(\mathfrak{o})\le|\mathrm{D}_{(\mathfrak{C},\mathsf{d}_{\pi/2,\mathfrak{C}})}f|_a(\mathfrak{o})\le\sqrt{2}\,|\mathrm{D}_{\mathfrak{C}}f|_a(\mathfrak{o}).\qquad\square \tag{8.36}$$

We can now show that the asymptotic Lipschitz constant satisfies (8.32).

Lemma 8.11. *Let $\xi:X\to\mathbb{R}$ and let $\zeta:\mathfrak{C}\to\mathbb{R}$ be defined by $\zeta([x,r]):=\xi(x)r^2$.*

(i) If ζ is $\mathsf{d}_{\mathfrak{C}}$-Lipschitz in $\mathfrak{C}[R]$ then $\xi\in\mathrm{Lip}_b(X)$ and

$$\sup_X|\xi|\le\frac{1}{R^2}\sup_{\mathfrak{C}[R]}|\zeta|\le\frac{1}{R}\mathrm{Lip}(\zeta,\mathfrak{C}[R]),\quad\mathrm{Lip}(\xi,X)\le\frac{1}{R}\mathrm{Lip}(\zeta,\mathfrak{C}[R]). \tag{8.37}$$

(ii) If $\xi\in\mathrm{Lip}_b(X)$ then ζ is $\mathsf{d}_{\mathfrak{C}}$-Lipschitz in $\mathfrak{C}[R]$ for every $R>0$ and

$$\sup_{\mathfrak{C}[R]}|\zeta|\le R^2\sup_X|\xi|,\quad\mathrm{Lip}^2(\zeta,\mathfrak{C}[R])\le R^2\Big(\mathrm{Lip}^2(\xi,(X,\tilde{\mathsf{d}}))+4\sup_X|\xi|^2\Big) \tag{8.38}$$

where $\tilde{\mathsf{d}}:=2\sin(\mathsf{d}_\pi/2)$.

(iii) In case (i) or (ii) we have, for every $x\in X$ and $r\ge 0$,

$$|\mathrm{D}_{\mathfrak{C}}\zeta|_a^2([x,r])=\begin{cases}\Big(|\mathrm{D}_X\xi|_a^2(x)+4\xi^2(x)\Big)r^2 & \text{if } r>0,\\ 0 & \text{if } r=0,\end{cases} \tag{8.39}$$

and a similar formula holds for the metric slope $|\mathrm{D}_{\mathfrak{C}}\zeta|([x,r])$. Equation (8.39) *also holds for the distance $\mathsf{d}_{\pi/2,\mathfrak{C}}$.*

Proof. As usual we set $\mathfrak{y}_i=[x_i,r_i]$, $\mathfrak{y}=[x,r]$.

Let us check the first statement (i). If ζ is locally Lipschitz then $|\xi(x)|=\frac{1}{R^2}|\zeta([x,R])-\zeta([x,0])|\le\frac{1}{R}\mathrm{Lip}(\zeta;\mathfrak{C}[R])$ for every R sufficiently small, so that ξ is uniformly bounded. Moreover, for every $R>0$ we have

$$R^2|\xi(x_1)-\xi(x_2)|\le|\zeta(x_1,R)-\zeta(x_2,R)|\le\mathrm{Lip}(\zeta;\mathfrak{C}[R])R\tilde{\mathsf{d}}(x_1,x_2)\le\mathrm{Lip}(\zeta;\mathfrak{C}[R])R\mathsf{d}(x_1,x_2),$$

so that ξ is uniformly Lipschitz and (8.37) holds.

Concerning (ii), if $\xi \in \mathrm{Lip}_b(X)$ we set $\sup|\xi| = S$, $\mathrm{Lip}(\xi, (X, \tilde{\mathsf{d}})) = L$, and we use the identity

$$\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2) = (\xi(x_1) - \xi(x_2))r_1r_2 + \xi(x)r(r_1 - r_2) + \omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y})(r_1 - r_2), \tag{8.40}$$
$$\text{where } \omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y}) := 2r_1\xi(x_1) + 2r_2\xi(x_2) - 4r\xi(x) \text{ with } \lim_{\mathfrak{y}_1,\mathfrak{y}_2\to\mathfrak{y}} \omega(\mathfrak{y}_1, \mathfrak{y}_2) = 0.$$

Since $|\omega(\mathfrak{y}_1, \mathfrak{y}_2; 0)| \le 4RS$ if $\mathfrak{y}_i \in \mathfrak{C}[R]$, equation (8.40) with $r = 0$ yields

$$|\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)| \le L\tilde{\mathsf{d}}(x_1, x_2)r_1r_2 + 4RS|r_1 - r_2| \le 2\Big(L^2 + 4r^2\Big)^{1/2} R\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2).$$

Letting $R \downarrow 0$ the inequality above also proves (8.39) in the case $r = 0$.

In order to prove (8.39) when $r \neq 0$ let us set $L_{\mathfrak{C}} := |\mathrm{D}_{\mathfrak{C}}\zeta|_a^2([x, r])$, $L_X := |\mathrm{D}_X\xi|_a(x)$, and let G_L be a function satisfying (8.34) with respect to the distance $\tilde{\mathsf{d}}$ (see Remark 8.10). Equation (8.40) yields

$$\begin{aligned}|\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)| &\le G_L(x_1, x_2)\tilde{\mathsf{d}}(x_1, x_2)r_1r_2 + \big(2|\xi(x)|r + |\omega(\mathfrak{y}_1, \mathfrak{y}_2)|\big)|r_1 - r_2| \\ &\le \Big(G_L^2(x_1, x_2)r_1r_2 + \big(2|\xi(x)|r + |\omega(\mathfrak{y}_1, \mathfrak{y}_2)|\big)^2\Big)^{1/2} \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2).\end{aligned}$$

Passing to the limit as $\mathfrak{y}_1, \mathfrak{y}_2 \to \mathfrak{y}$ (and using the fact that $x_1, x_2 \to x$ since $r \neq 0$) we obtain

$$L_{\mathfrak{C}} \le r\Big(L_X^2 + 4|\xi(x)|^2\Big)^{1/2}$$

In order to prove the converse inequality we observe that for every $L' < L_X$ there exist two sequences of points $(x_{i,n})_{n\in\mathbb{N}}$ converging to x w.r.t. d such that $\frac{\xi(x_{1,n})-\xi(x_{2,n})}{\tilde{\mathsf{d}}(x_{1,n},x_{2,n})} \ge L'$. Let us set $\delta_n := \tilde{\mathsf{d}}(x_{1,n}, x_{2,n}) \to 0$ and choose $r_{1,n} := r$, $r_{2,n} = r(1 + \lambda\delta_n)$ for an arbitrary constant $\lambda \in \mathbb{R}$ with the same sign of $\xi(x)$. Again applying (8.40) we get

$$L_{\mathfrak{C}} \ge \liminf_{n\to\infty} \frac{|\zeta(\mathfrak{y}_{1,n}) - \zeta(\mathfrak{y}_{2,n})|}{\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_{1,n}, \mathfrak{y}_{2,n})} \ge \liminf_{n\to\infty} \frac{L'\delta_n r^2 + 2|\xi(x)|r^2|\lambda|\delta_n + o(\delta_n)}{\sqrt{\lambda^2r^2\delta_n^2 + r^2\delta_n^2 + o(\delta_n)}} = r\frac{L' + 2|\xi(x)|\,|\lambda|}{\sqrt{\lambda^2 + 1}}.$$

Optimizing with respect to λ we obtain

$$L_{\mathfrak{C}}^2 \ge r^2\big((L')^2 + 4|\xi(x)|^2\big) \quad \text{for every } L' \le L_X.$$

The arguments for proving (8.39) for metric slopes are completely analogous. □

Hopf-Lax formula and subsolutions to metric HJ equation in the cone $\mathfrak{C}$. Whenever $f \in \mathrm{Lip}_b(\mathfrak{C})$ the Hopf-Lax formula

$$\mathscr{Q}_t f(\mathfrak{y}) := \inf_{\mathfrak{y}'\in\mathfrak{C}} f(\mathfrak{y}') + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{y}') \quad \mathfrak{y} \in \mathfrak{C},\ t > 0, \tag{8.41}$$

provides a function $t \mapsto \mathscr{Q}_t f$ which is Lipschitz from $[0, \infty)$ to $\mathrm{C}_b(\mathfrak{C})$, satisfies the a-priori bounds

$$\inf_{\mathfrak{C}} f \le \mathscr{Q}_\cdot f \le \sup_{\mathfrak{C}} f, \quad \mathrm{Lip}(\mathscr{Q}_t f; \mathfrak{C}) \le 2\,\mathrm{Lip}(f, \mathfrak{C}), \tag{8.42}$$

and solves

$$\partial_t^+ \mathscr{Q}_t f(\mathfrak{z}) + \frac{1}{2}|\mathrm{D}_{\mathfrak{C}}\mathscr{Q}_t f|_a^2(\mathfrak{z}) \le 0 \quad \text{for every } \mathfrak{z} \in \mathfrak{C},\ t>0, \tag{8.43}$$

where ∂_t^+ denotes the partial right derivative w.r.t. t. It is also possible to prove that for every $\mathfrak{y} \in \mathfrak{C}$ the time derivative of $\mathscr{Q}_t f(\mathfrak{y})$ exists with possible countable exceptions and that (8.43) is in fact an equality if $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ is a length space, a property that always holds if (X, d) is a length metric space.

Let us state now our main result:

Theorem 8.12 (Metric subsolution of the Hamilton-Jacobi equation in X)**.** *Let $\xi \in \mathrm{Lip}_b(X)$ satisfy the uniform lower bound $P := \inf_X \xi + 1/2 > 0$ and let us set $\zeta([x,r]) := \xi(x)r^2$. Then, for every $t \in [0,1]$ we have*

$$\begin{aligned}
&\mathscr{Q}_t\zeta([x,r]) = \xi_t(x)r^2 \quad \textit{where} \quad \xi_t(x) := \mathscr{P}_t\xi(x),\\
&\mathscr{P}_t\xi(x) := \inf_{x'\in X} \frac{\xi(x')}{1+2t\xi(x')} + \frac{\sin^2(\mathsf{d}_{\pi/2}(x,x'))}{2t(1+2t\xi(x'))} = \inf_{x'\in X} \frac{1}{2t}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x,x'))}{1+2t\xi(x')}\Big).
\end{aligned} \tag{8.44}$$

Moreover, for every $R>0$ we have

$$\xi_t(x)r^2 = \inf_{\mathfrak{y}'=[x',r']\in\mathfrak{C}[R]} \xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r']) \quad \textit{for every } x\in X,\ r\le PR. \tag{8.45}$$

The map $t \mapsto \xi_t$ is Lipschitz from $[0,1]$ to $\mathrm{C}_b(X)$ with $\xi_t \in \mathrm{Lip}_b(X)$ for every $t \in [0,1]$ and is a subsolution to the Hamilton-Jacobi equation

$$\partial_t^+\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|_a^2(x) + 2\xi_t^2(x) \le 0 \quad \textit{for every } x\in X,\ t\in[0,1]. \tag{8.46}$$

For every $x \in X$ the map $t \mapsto \xi_t(x)$ is time differentiable with at most countable exceptions. If eventually (X,d) is a length metric space then (8.46) is an identity.

Notice that when $\xi(x) \equiv \xi$ is constant, (8.44) reduces to $\mathscr{P}_t\xi = \xi/(1+2t\xi)$ which is the solution to the elementary differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\xi + 2\xi^2 = 0$.

Proof. Let us observe that $\inf_{t\in[0,1],z\in X}(1+2t\xi(z)) = P > 0$. A simple calculation shows

$$\begin{aligned}
\xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r']) &= \frac{1}{2t}\Big((1+2t\xi(x'))(r')^2 + r^2 - 2r\,r'\cos(\mathsf{d}_\pi(x,x'))\Big)\\
&= \frac{1}{2t(1+2t\xi(x'))}\Big[\Big((1+2t\xi(x'))r' - \cos(\mathsf{d}_\pi(x,x'))r\Big)^2 + r^2\Big(2t\xi(x') + \sin^2(\mathsf{d}_\pi(x,x'))\Big)\Big].
\end{aligned}$$

Hence, if we choose

$$r' = r'(x,x',r) := \begin{cases} r\cos(\mathsf{d}_\pi(x,x'))/(1+2t\xi(x')) & \text{if } \mathsf{d}(x,x') \le \pi/2\\ 0 & \text{otherwise,}\end{cases}$$

we find (notice the truncation at $\pi/2$ instead of π)

$$\inf_{r'\ge 0} \xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r']) = \frac{r^2}{2t(1+2t\xi(x'))}\Big(2t\xi(x') + \sin^2(\mathsf{d}_{\pi/2}(x,x'))\Big) \tag{8.47}$$

which yields (8.44) and (8.45).

Equation (8.45) also shows that the function $\zeta_t = \xi_t(x)r^2$ coincides on $\mathfrak{C}[PR]$ with the solution ζ_t^R given by the Hopf-Lax formula in the metric space $\mathfrak{C}[R]$. Since the initial datum ζ is bounded and Lipschitz on $\mathfrak{C}[R]$ we deduce that ζ_t^R is bounded and Lipschitz, so that $t \mapsto \xi_t$ is bounded and Lipschitz in X by Lemma 8.11.

Equation (8.46) and the other regularity properties then follow by (8.39) and the general properties of the Hopf-Lax formula in $\mathfrak{C}[R]$. □

Duality between the Hellinger-Kantorovich distance and subsolutions to the Hamilton-Jacobi equation. We conclude this section with the main application of the above results to the Hellinger-Kantorovich distance.

Theorem 8.13. *Let us suppose that (X, d) is a complete and separable metric space.*

(i) If $\mu \in \mathrm{AC}^2([0,1]; (\mathcal{M}(X), \mathsf{HK}))$ and $\xi : [0,1] \to \mathrm{Lip}_b(X)$ is uniformly bounded, Lipschitz w.r.t. the uniform norm, and satisfies (8.46), then the curve $t \mapsto \int \xi_t \,\mathrm{d}\mu_t$ is absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_X \xi_t \,\mathrm{d}\mu_t \le \frac{1}{2}|\mu_t'|^2_{\mathsf{HK}} \tag{8.48}$$

(ii) If (X, d) is a length space then for every μ_0, μ_1 and $k \in \mathbb{N} \cup \{\infty\}$

$$\begin{aligned}\frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) = \sup \Big\{ \int_X \xi_1 \,\mathrm{d}\mu_1 - \int_0 \xi_0 \,\mathrm{d}\mu_0 : \quad \xi \in \mathrm{C}^k([0,1]; \mathrm{Lip}_b(X)),& \\ \partial_t \xi_t(x) + \frac{1}{2}|\mathrm{D}_X \xi_t|^2(x) + 2\xi_t^2(x) \le 0 \textit{ in } X \times (0,1) \Big\}.&\end{aligned} \tag{8.49}$$

Moreover, in the above formula we can also take the supremum over functions $\xi \in \mathrm{C}^k([0,1]; \mathrm{Lip}_b(X))$ with bounded support.

Proof. If ξ satisfies (8.46) then setting $\zeta_t([x,r]) := \xi_t(x)r^2$ we obtain a family of functions $t \mapsto \zeta_t$, $t \in [0,1]$, whose restriction to every $\mathfrak{C}[R]$ is uniformly bounded and Lipschitz, and it is Lipschitz continuous with respect to the uniform norm of $\mathrm{C}_b(\mathfrak{C}[R])$. By Lemma 8.11 teh function ζ solves

$$\partial_t^+ \zeta_t + \frac{1}{2}|\mathrm{D}_{\mathfrak{C}} \zeta_t|_a^2 \le 0 \quad \text{in } \mathfrak{C} \times (0,1).$$

According to Theorem 8.4 we find $\theta > 0$ a curve $\alpha \in \mathrm{AC}^2([0,1]; (\mathcal{P}_2(\mathfrak{C}[\theta]), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}))$ satisfying (8.19). Applying the results of [6, Sect. 6], the map $t \mapsto \int_{\mathfrak{C}} \zeta_t \,\mathrm{d}\alpha_t$ is absolutely continuous with

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathfrak{C}} \zeta_t \,\mathrm{d}\alpha_t \le \frac{1}{2}|\alpha_t'|^2_{\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}} \quad \mathscr{L}^1\text{-a.e. in (0,1).}$$

Since $\int_{\mathfrak{C}} \zeta_t \,\mathrm{d}\alpha_t = \int_X \xi_t \,\mathrm{d}\mu_t$ we obtain (8.48).

Let us now prove (ii). As a first step, denoting by S the right-hand side of (8.49), we prove that $\mathsf{HK}^2(\mu_0, \mu_1) \ge S$.

Notice that if $\xi \in \mathrm{C}^1([0,1]; \mathrm{Lip}_b(X))$ satisfies the pointwise inequality

$$\partial_t \xi_t(x) + \frac{1}{2}|\mathrm{D}_X \xi_t|^2(x) + 2\xi_t^2(x) \le 0, \tag{8.50}$$

then it also satisfies (8.46) since (8.50) yields, for every $t \in (0,1)$, the relation

$$\frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) \le -\Big(\partial_t\xi_t(x) + 2\xi_t^2(x)\Big) \quad \text{for every } x \in X, \tag{8.51}$$

and the right hand side is bounded and continuous in X. Equation (8.51) thus yields the same inequality for the upper semicontinuous envelope of $|\mathrm{D}_X\xi_t|$ and this function coincides with $|\mathrm{D}_X\xi_t|_a$ since X is a length space.

We can therefore apply the previous point (i) by choosing $\lambda > 1$ and a Lipschitz curve $\mu : [0,1] \to \mathcal{M}(X)$ joining μ_0 to μ_1 with metric velocity $|\mu_t'|_{\mathsf{HK}} \le \lambda\mathsf{HK}(\mu_0,\mu_1)$, whose existence is guaranteed by the length property of X and a standard rescaling technique. Relation (8.48) yields

$$2\int_X \xi_1\,\mathrm{d}\mu_1 - 2\int_X \xi_0\,\mathrm{d}\mu_0 \le \int_0^1 |\mu_t'|^2_{\mathsf{HK}}\,\mathrm{d}t \le \lambda^2\mathsf{HK}^2(\mu_0,\mu_1).$$

Since $\lambda > 1$ is arbitrary, we get $\mathsf{HK}^2(\mu_0,\mu_1) \ge S$.

In order to prove the converse inequality we fix $\eta > 0$ and apply the duality Theorem 7.22 to get $\xi_0 \in \mathrm{Lip}_{bs}(X)$ (the space of Lipschitz functions with bounded support) with $\inf \xi_0 > -1/2$ such that

$$2\int_X \mathscr{P}_1\xi_0\,\mathrm{d}\mu_1 - 2\int_X \xi_0\,\mathrm{d}\mu_0 \ge \mathsf{HK}^2(\mu_0,\mu_1) - \eta. \tag{8.52}$$

Setting $\xi_t := \mathscr{P}_t\xi_0$ we find a solution to (8.46) which has bounded support, is uniformly bounded in $\mathrm{Lip}_b(X)$ and Lipschitz with respect to the uniform norm. We have to show that $(\xi_t)_{t\in[0,1]}$ can be suitably approximated by smoother solutions $\xi^\varepsilon \in \mathrm{C}^\infty([0,1];\mathrm{Lip}_b(X))$, $\varepsilon > 0$, in such a way that $\int \xi_i^\varepsilon\,\mathrm{d}\mu_i \overset{\varepsilon\downarrow 0}{\to} \int \xi_i\,\mathrm{d}\mu_i$, $i = 0,1$.

We use an argument of [1]: by approximating ξ_t with $\lambda\xi(\lambda t+(1-\lambda)/2, x)$, $\lambda < 1$ and passing to the limit as $\lambda \uparrow 1$ it is not restrictive to assume that ξ is defined in a larger interval $[a,b]$, with $a < 0, b > 1$. Now, a time convolution is well defined on $[0,1]$, for which we use a symmetric, nonnegative kernel $\kappa \in \mathrm{C}_\mathrm{c}^\infty(\mathbb{R})$ with integral 1 defined via

$$\xi_t^\varepsilon(x) := (\xi_{(\cdot)}(x) * \kappa_\varepsilon)_t = \int_\mathbb{R} \xi_w(x)\kappa_\varepsilon(t-w)\,\mathrm{d}w, \quad \text{where } \kappa_\varepsilon(t) := \varepsilon^{-1}\kappa(t/\varepsilon), \tag{8.53}$$

yields a curve $\xi^\varepsilon \in \mathrm{C}^\infty([0,1];\mathrm{Lip}_b(X))$ satisfying

$$\partial_t\xi_t^\varepsilon + \frac{1}{2}\big(|\mathrm{D}_X\xi_{(\cdot)}|^2\big) * \kappa_\varepsilon + 2\big(\xi_{(\cdot)}^2\big) * \kappa_\varepsilon \le 0 \quad \text{in } X \times [0,1].$$

By Jensen inequality $\xi_{(\cdot)}^2 * \kappa_\varepsilon \ge (\xi_{(\cdot)} * \kappa_\varepsilon)^2$, $|\mathrm{D}_X\xi_{(\cdot)}|^2 * \kappa_\varepsilon \ge (|\mathrm{D}_X\xi_{(\cdot)}| * \kappa_\varepsilon)^2$. Moreover, applying the next Lemma 8.14 we also get $|\mathrm{D}_X\xi_{(\cdot)}| * \kappa_\varepsilon \ge |\mathrm{D}_X\xi_t^\varepsilon\xi_{(\cdot)}|$ so that ξ_t^ε satisfies (8.50). Since $\xi_t^\varepsilon \to \xi_t$ uniformly in X for every $t \in [0,1]$, we easily get

$$S \ge \lim_{\varepsilon\downarrow 0} 2\Big(\int_X \xi_1^\varepsilon\,\mathrm{d}\mu_1 - \int_X \xi_0^\varepsilon\,\mathrm{d}\mu_0\Big) \ge \mathsf{HK}^2(\mu_0,\mu_1) - \eta.$$

Since $\eta > 0$ is arbitrary the proof of (ii) is complete. □

Lemma 8.14. *Let (X, d) be a separable metric space, let $(\Omega, \mathcal{B}, \pi)$ be a measure space with $\pi(\Omega) < \infty$ and let $\xi_\omega \in \mathrm{Lip}_b(X)$, $\omega \in \Omega$, be a family of uniformly bounded functions such that $\sup_{\omega\in\Omega} \mathrm{Lip}(\xi_\omega; X) < \infty$ and $\omega \mapsto \xi_\omega(x)$ is $\mathcal{B}$-measurable for every $x \in X$. Then the function $\xi(x) := \int_\Omega \xi_\omega(x)\,\mathrm{d}\pi(\omega)$ belongs to $\mathrm{Lip}_b(X)$, for every $x \in X$ the maps $\omega \mapsto |\mathrm{D}_X\xi_\omega|(x)$ and $\omega \mapsto |\mathrm{D}_X\xi_\omega|_a(x)$ are $\mathcal{B}$-measurable and for every $x \in X$ they satisfy*

$$|\mathrm{D}_X\xi|_a(x) \le \int_X |\mathrm{D}_X\xi_\omega|_a(x)\,\mathrm{d}\pi(\omega), \quad |\mathrm{D}_X\xi|(x) \le \int_X |\mathrm{D}_X\xi_\omega|(x)\,\mathrm{d}\pi(\omega). \tag{8.54}$$

Proof. The fact that $\xi_\omega \in \mathrm{Lip}_b(X)$ is obvious. Whenever $x \in X$ is fixed, thanks to the expression (8.33) for $|\mathrm{D}_X\xi|_a(x)$, it is sufficient to prove that for every $r > 0$ the map $\omega \mapsto s_{r,\omega}(x) := \sup_{y\neq z\in B_r(x)} |\xi_\omega(y) - \xi_\omega(z)|/\mathsf{d}(y,z)$ is $\mathcal{B}$-measurable: this property follows by the continuity of ξ_ω and the separability of X, so that it is possible to restrict the supremum to a countable dense collection of points $\tilde{B}_r(x)$ in $B_r(x)$. An analogous argument holds for $|\mathrm{D}_X\xi_\omega|$.

Taking the supremum of the inequality

$$\frac{|\xi(y) - \xi(z)|}{\mathsf{d}(y,z)} \le \int_\Omega \frac{|\xi_\omega(y) - \xi_\omega(z)|}{\mathsf{d}(y,z)}\,\mathrm{d}\pi(\omega), \quad y \neq z,$$

with respect to $y, z \in \tilde{B}_r(x)$ and $y \neq z$, we obtain

$$\sup_{y\neq z\in B_r(x)} \frac{|\xi(y) - \xi(z)|}{\mathsf{d}(y,z)} \le \int_\Omega s_{r,\omega}(x)\,\mathrm{d}\pi(\omega).$$

A further limit as $r \downarrow 0$ and the application of the Lebesgue Dominated convergence Theorem yields the first inequality of (8.54). The argument to prove the second inequality is completely analogous. □

When $X = \mathbb{R}^d$ the characterization (8.49) holds for an even smoother class of subsolutions.

Corollary 8.15. *Let $X = \mathbb{R}^d$ endowed with the Euclidean distance. Then*

$$\begin{aligned}\mathsf{HK}^2(\mu_0,\mu_1) = 2\sup\Big\{ \int_X \xi_1\,\mathrm{d}\mu_1 - \int_X \xi_0\,\mathrm{d}\mu_0 : \xi \in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d\times[0,1]),& \\ \partial_t\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) + 2\xi_t^2(x) \le 0 \quad \textit{in } X\times(0,1)\Big\}.&\end{aligned} \tag{8.55}$$

Proof. We just have to check that the supremum of (8.49) does not change if we substitute $\mathrm{C}^\infty([0,1];\mathrm{Lip}_{bs}(\mathbb{R}^d))$ with $\mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d\times[0,1])$. This can be achieved by approximating any subsolution $\xi \in \mathrm{C}^\infty([0,1];\mathrm{Lip}_{bs}(\mathbb{R}^d))$ via convolution in space with a smooth kernel with compact support, which still provides a subsolution thanks to Lemma 8.14. □

8.5 The dynamic interpretation of the Hellinger-Kantorovich distance "à la Benamou-Brenier"

In this section we will apply the superposition principle of Theorem 8.4 and the duality result 8.13 with subsolutions of the Hamilton-Jacobi equation to quickly derive a dynamic formulation “à la Benamou-Brenier” [7, 33], [2, Sect. 8] of the Hellinger-Kantorovich distance, which has also been considered in the recent [23]. In order to keep the exposition simpler, we will consider the case $X = \mathbb{R}^d$ with the canonical Euclidean distance $\mathsf{d}(x_1, x_2) := |x_1 - x_2|$, but the result can be extended to more general Riemannian and metric settings, e.g. arguing as in [6, Sect. 6]. A different approach, based on suitable representation formulae for the continuity equation, is discussed in [27].

Our starting point is provided by a suitable class of linear continuity equations with reaction. In the following we will denote by $\mu_I \in \mathcal{M}(\mathbb{R}^d \times [0,1])$ the measure

$$\int \xi \,\mathrm{d}\mu_I := \int_0^1 \int_X \xi_t(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \tag{8.56}$$

induced by a curve $\mu \in \mathrm{C}^0([0,1]; \mathcal{M}(\mathbb{R}^d))$.

Definition 8.16. *Let $\mu \in \mathrm{C}^0([0,1]; \mathcal{M}(\mathbb{R}^d))$, $(\boldsymbol{v}, w) : \mathbb{R}^d \times (0,1) \to \mathbb{R}^{d+1}$ be a Borel vector field in $\mathrm{L}^2(\mathbb{R}^d \times (0,1), \mu_I; \mathbb{R}^{d+1})$ thus satisfying*

$$\int_0^1 \int_X \Big(|\boldsymbol{v}_t(x)|^2 + w_t^2(x)\Big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t = \int |(\boldsymbol{v}, w)|^2\,\mathrm{d}\mu_I < \infty. \tag{8.57}$$

We say that μ satisfies the continuity equation with reaction governed by the field $(\boldsymbol{v}, w)$ if

$$\partial_t \mu_t + \nabla\cdot(\boldsymbol{v}_t \mu_t) = w_t \mu_t \quad \textit{holds in the sense of distributions of } \mathbb{R}^d \times (0,1), \tag{8.58}$$

i.e. for every test function $\xi \in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d \times (0,1))$

$$\int_0^1 \int_{\mathbb{R}^d} \Big(\partial_t \xi_t(x) + \mathrm{D}_x \xi_t(x) \boldsymbol{v}_t(x) + \xi_t(x) w_t(x)\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t = 0. \tag{8.59}$$

An equivalent formulation [2, Sect. 8.1] of (8.58) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X \xi(x)\,\mathrm{d}\mu_t(x) = \int_X \Big(\mathrm{D}_x\xi(x)\boldsymbol{v}_t(x) + \xi(x) w_t(x)\Big)\,\mathrm{d}\mu_t \quad \text{in } \mathscr{D}'(0,1), \tag{8.60}$$

for every $\xi \in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d)$. We have a first representation result for absolutely continuous curves.

Theorem 8.17. *Let $(\mu_t)_{t\in[0,1]}$ be an absolutely continuous curve in $\mathrm{AC}^2([0,1]; (\mathcal{M}(\mathbb{R}^d), \mathsf{HK}))$. Then μ satisfies the continuity equation with reaction (8.58) governed by a Borel vector field $(\boldsymbol{v}, w) \in \mathrm{L}^2(\mathbb{R}^d \times (0,1), \mu_I; \mathbb{R}^{d+1})$ such that*

$$(\boldsymbol{v}_t, w_t) \in \mathrm{L}^2(\mathbb{R}^d; \mu_t), \qquad \int \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t \le |\mu_t'|^2 \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (0,1). \tag{8.61}$$

Proof. We will denote by I the interval $[0,1]$ endowed with the Lebesgue measure $\lambda = \mathscr{L}^1 \llcorner [0,1]$ and we consider the maps $\mathsf{x}_I : \mathrm{C}(I; \mathfrak{C}) \times I \to X \times I$, $\mathsf{R} : \mathrm{C}(I; \mathfrak{C}) \times I \to \mathbb{R}_+$ defined by $\mathsf{x}_I(\mathrm{z}, t) := (\mathsf{x}(\mathrm{z}(t)), t)$ and $\mathsf{R}(\mathrm{z}, t) := \mathsf{r}(\mathrm{z}(t))$.

Let $\boldsymbol{\pi}$ be a dynamic plan in $\mathfrak{C}$ representing μ_t as in Theorem 8.4. We consider the deformed dynamic plan $\boldsymbol{\pi}_I := (\mathsf{R}^2\boldsymbol{\pi}) \otimes \lambda$, the measure $\mu_I := (\mathsf{x}_I)_\sharp \boldsymbol{\pi}_I$ and the disintegration

$(\tilde{\boldsymbol{\pi}}_{x,t})_{(x,t)\in X\times I}$ of $\boldsymbol{\pi}_I$ with respect to μ_I. Notice that $\tilde{\boldsymbol{\pi}} \le \Theta\boldsymbol{\pi}$, where Θ is given by (8.18), andAlex: Now better, but RHS cannot have time dependence, while μ_I still depends on t.

$$\mu_I = \int_0^1 (\mu_t \otimes \delta_t)\,\mathrm{d}\lambda(t), \tag{8.62}$$

coincides with (8.56), since for every $\xi \in \mathrm{B}_b(X\times I)$ we have

$$\int \xi\,\mathrm{d}\mu_I = \int \xi(\mathsf{x}(\mathrm{z}(t)),t)\mathsf{r}^2(\mathrm{z}(t))\,\mathrm{d}\boldsymbol{\pi}_I(\mathrm{z},t) = \int_0^1\int_X \xi_t(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.$$

Let $\boldsymbol{u} \in \mathrm{L}^2(\mathrm{AC}^2(I;\mathfrak{C})\times I;\boldsymbol{\pi}\otimes\lambda;\mathbb{R}^{d+1})$ be the Borel vector field $\boldsymbol{u}(\mathfrak{y},t) := \mathfrak{y}'_{\mathfrak{C}}(t)$ for every curve $\mathfrak{y} \in \mathrm{AC}^2(I;\mathfrak{C})$ and $t\in I$, where $\mathfrak{y}'_{\mathfrak{C}}$ is defined as in (8.12). By taking the density of the vector measure $(\mathsf{x}_I)_\sharp(\boldsymbol{u}\boldsymbol{\pi}_I)$ with respect to μ_I we obtain a Borel vector field $\boldsymbol{u}_I = (\boldsymbol{v},w) \in \mathrm{L}^2(X\times I;\mu_I;\mathbb{R}^{d+1})$ which satisfies

$$\boldsymbol{u}_I(x,t) = \int \boldsymbol{u}\,\mathrm{d}\boldsymbol{\pi}_{x,t} \quad \text{for } \mu_I\text{-a.e. } (x,t)\in X\times I. \tag{8.63}$$

By choosing a test function $\zeta([x,r],t) := \xi(x)\eta(t)r^2$ with $\xi\in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d), \eta\in\mathrm{C}^\infty_\mathrm{c}(I)$ we getCheck following for "2ξ"

$$\begin{aligned}
-\int_0^1 \eta' \int_{\mathbb{R}^d}\xi\,\mathrm{d}\mu_t\,\mathrm{d}t &= -\int_{\mathbb{R}^d\times I}\eta'(t)\xi(x)\,\mathrm{d}\mu_I = -\int \xi(\mathsf{x}(\mathfrak{y}(t))\,\mathsf{r}^2(\mathfrak{y}(t))\eta'(t)\,\mathrm{d}(\boldsymbol{\pi}\otimes\lambda)\\
&= -\int \partial_t\zeta(\mathfrak{y}(t),t)\,\mathrm{d}(\boldsymbol{\pi}\otimes\lambda) = \int\Big(-\frac{\mathrm{d}}{\mathrm{d}t}\zeta(\mathrm{y}(t),t) + \langle \mathrm{D}_{\mathfrak{C}}\zeta(\mathrm{y}(t),t),\mathrm{y}'_{\mathfrak{C}}(t)\rangle\Big)\,\mathrm{d}(\boldsymbol{\pi}\otimes\lambda)\\
&= \int\Big(\int_0^1 -\frac{\mathrm{d}}{\mathrm{d}t}\zeta(\mathrm{y}(t),t)\,\mathrm{d}t\Big)\,\mathrm{d}\boldsymbol{\pi} + \int \langle(\mathrm{D}_X\xi(\mathsf{x}_I),2\xi(\mathsf{x}_I)),\boldsymbol{u}\rangle\mathsf{R}^2\,\mathrm{d}(\boldsymbol{\pi}\otimes\lambda)\\
&= \int \eta(t)\langle(\mathrm{D}_X\xi(x),2\xi(x)),\boldsymbol{u}_I\rangle\,\mathrm{d}\mu_I\\
&= \int_0^1\eta(t)\int_{\mathbb{R}^d}\Big(\langle \mathrm{D}_X\xi(x),\boldsymbol{v}_t(x)\rangle + \xi(x)w_t(x)\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t. \qquad \square
\end{aligned}$$

Theorem 8.18. *Let $(\mu_t)_{t\in[0,1]}$ be a continuous curve in $\mathcal{M}(\mathbb{R}^d)$ that solves the continuity equation with reaction (8.58) governed by the Borel vector field $(\boldsymbol{v},w)\in L^2(\mathbb{R}^d\times[0,1],\mu_I;\mathbb{R}^{d+1})$, μ_I given by (8.56). Then $\mu\in\mathrm{AC}^2([0,1];(\mathcal{M}(\mathbb{R}^d),\mathsf{HK}))$ and*

$$|\mu_t'|^2 \le \int\Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t\in(0,1). \tag{8.64}$$

Proof. The simple scaling $\xi(t,x)\to(b-a)\xi(a+(b-a)t,x)$ transforms any subsolution of the Hamilton-Jacobi equation in $[0,1]$ to a subsolution of the same equation in $[a,b]$. Thus,

$$\begin{aligned}\mathsf{HK}^2(\mu_0,\mu_1) = 2(b-a)\sup\Big\{\int_X\xi_b\,\mathrm{d}\mu_1 - \int_X\xi_a\,\mathrm{d}\mu_0 :\quad &\xi\in\mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d\times[a,b]),\\ &\partial_t\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) + 2\xi_t^2(x)\le 0 \text{ in } \mathbb{R}^d\times(a,b)\Big\}.\end{aligned} \tag{8.65}$$

Let $\xi \in \mathrm{C}_c^\infty(\mathbb{R}^d \times [0,1])$ be a subsolution to the Hamilton-Jacobi equation $\partial_t \xi + \frac{1}{2}|\mathrm{D}\xi|^2 + 2\xi^2 \le 0$ in $\mathbb{R}^d \times [0,1]$. By a standard argument (see [2, Lem. 8.1.2]), the integrability (8.58) and the weak continuity of $t \mapsto \mu_t$ yield

$$\begin{aligned}
2\int_{\mathbb{R}^d} \xi_{t_1}\,\mathrm{d}\mu_{t_1} - 2\int_{\mathbb{R}^d} \xi_{t_0}\,\mathrm{d}\mu_{t_0} &= 2\int_{t_0}^{t_1}\int_{\mathbb{R}^d} \Big(\partial_t \xi_t + \langle \mathrm{D}_x \xi_t, \boldsymbol{v}_t\rangle + \xi_t w_t\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t \\
&\le 2\int_{t_0}^{t_1}\int_{\mathbb{R}^d} \Big(-\frac{1}{2}|\mathrm{D}\xi_t|^2 - 2\xi_t^2 + \langle \mathrm{D}_x \xi_t, \boldsymbol{v}_t\rangle + \xi_t w_t\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t \\
&\le \int_{t_0}^{t_1}\int_{\mathbb{R}^d} \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t.
\end{aligned}$$

Applying Corollary 8.15 and (8.65) we find

$$\mathsf{HK}^2(\mu_{t_0}, \mu_{t_1}) \le (t_1 - t_0)\int_{t_0}^{t_1}\int_{\mathbb{R}^d} \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t \quad \text{for every } 0 \le t_0 < t_1 \le 1,$$

which yields (8.64). □

Combining Theorems 8.17 and 8.18 with Theorem 8.4 and the geodesic property of $(\mathcal{M}(\mathbb{R}^d), \mathsf{HK})$ we immediately have:

Corollary 8.19 (Representation of HK à la Benamou-Brenier)**.** *For every $\mu_0, \mu_1 \in \mathcal{M}(\mathbb{R}^d)$ we have*

$$\begin{aligned}
\mathsf{HK}^2(\mu_0, \mu_1) = \min\Big\{ \int_0^1 \int \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t : \mu \in \mathrm{C}([0,1]; \mathcal{M}(\mathbb{R}^d)), \quad \mu_{t=i} = \mu_i, \\
\partial_t \mu_t + \nabla \cdot (\boldsymbol{v}_t \mu_t) = w_t \mu_t \quad \textit{in } \mathscr{D}'(\mathbb{R}^d \times (0,1))\Big\}.
\end{aligned} \tag{8.66}$$

The Borel vector field $(\boldsymbol{v}, w)$ realizing the minimum in (8.66) *is uniquely determined μ_I-a.e. in $\mathbb{R}^d \times (0,1)$.*

8.6 Geodesics in $\mathcal{M}(\mathbb{R}^d)$

As in the case of the Kantorovich-Wasserstein distance, one may expect that geodesics $(\mu_t)_{t\in[0,1]}$ in $(\mathcal{M}(\mathbb{R}^d), \mathsf{HK})$ can be characterized by the system [27, Sect. 5]

$$\partial_t \mu_t + \nabla \cdot (\mu_t \mathrm{D}\xi_t) = 4\xi_t \mu_t, \quad \partial_t \xi_t + \frac{1}{2}|\mathrm{D}\xi_t|^2 + 2\xi_t^2 = 0. \tag{8.67}$$

In order to give a precise meaning to (8.67) we first have to select an appropriate regularity for ξ_t. Notice that we cannot expect C^1 smoothness for solutions of the Hamilton-Jacobi equation (8.67) (in contrast with subsolutions, that can be regularized as in Corollary 8.15) and on the other hand the $\mathscr{L}^d$ a.e. differentiability of Lipschitz functions guaranteed by Rademacher's Theorem is not sufficient if we want to consider arbitrary measures μ_t that could be singular with respect $\mathscr{L}^d$.

A convenient choice for our aims is provided by locally Lipschitz functions which are strictly differentiable at μ_I-a.e. points, where μ_I has been defined by (8.56): recall that a function $f : \mathbb{R}^d \to \mathbb{R}$ is strictly differentiable at $x \in \mathbb{R}^d$ if there exists $\mathrm{D}f(x) \in (\mathbb{R}^d)^*$ such that

$$\lim_{\substack{x',x''\to x\\ x'\neq x''}} \frac{f(x')-f(x'')-\mathrm{D}f(x)(x'-x'')}{|x'-x''|} = 0. \tag{8.68}$$

It is possible to prove [13, Prop. 2.2.4] that a locally Lipschitz function f is strictly differentiable at x if and only if the Clarke subgradient [13, Sect. 2.1] of f at x reduces to the singleton $\{\mathrm{D}f(x)\}$. In particular, denoting by $D \subset \mathbb{R}^d$ the set where f is differentiable and denoting by κ_ε a smooth convolution kernel as in (8.53), [13, Thm. 2.5.1] and Rademacher Theorem yield

$$\lim_{\substack{x'\to x\\ x'\in D}} \mathrm{D}f(x') = \mathrm{D}f(x), \quad \lim_{\varepsilon\downarrow 0} \mathrm{D}(f*\kappa_\varepsilon)(x) = \mathrm{D}f(x), \tag{8.69}$$

whenever x is a point of strict differentiability for f.

Theorem 8.20. *A weakly continuous curve $\mu \in \mathrm{C}^0([0,1];\mathcal{M}(\mathbb{R}^d))$ is a geodesic w.r.t. the HK distance if and only if there exists a map $\xi \in \mathrm{Lip}_{\mathrm{loc}}((0,1);\mathrm{C}_b(\mathbb{R}^d))$ such that*

(i) $\xi_t \in \mathrm{Lip}_b(\mathbb{R}^d)$ *for every* $t \in (0,1)$ *with* $t \mapsto \mathrm{Lip}(\xi_t,\mathbb{R}^d)$ *locally bounded in* $(0,1)$.

(ii) ξ *is right-differentiable w.r.t.* t *in* $\mathbb{R}^d \times (0,1)$ *and it is strictly differentiable w.r.t.* x *at* μ_I*-a.e.* $(x,t) \in \mathbb{R}^d \times (0,1)$.

(iii) The map ξ satisfies

$$\partial_t^+ \xi_t + \frac{1}{2}|\mathrm{D}_x\xi_t|_a^2(x) + 2\xi_t^2(x) = 0 \quad \textit{in } \mathbb{R}^d \times (0,1), \tag{8.70}$$

(iv) The curve $(\mu_t)_{t\in[0,1]}$ solves the continuity equation governed by $(\mathrm{D}_x\xi,\xi)$ in every compact interval of $(0,1)$, i.e.

$$\partial_t\mu_t + \nabla\cdot(\mu_t \mathrm{D}\xi_t) = 4\xi_t\mu_t \quad \textit{in } \mathscr{D}'(\mathbb{R}^d\times(0,1)). \tag{8.71}$$

Proof. The proof splits into a sufficiency and a necessity part, the latter having several steps.

Sufficiency. Let us suppose that μ,ξ satisfy conditions $(i),\dots,(iv)$.

Applying Lemma 8.21 below with $\boldsymbol{v} = \mathrm{D}_x\xi$ and $w = 4\xi$ and observing that $|\mathrm{D}\xi_t|_a(x) = |\mathrm{D}_x\xi_t(x)|$ at every point x of strict differentiability of ξ_t, we get, for every $0<a<b<1$,

$$\begin{aligned}
2\int_{\mathbb{R}^d}\xi_b\,\mathrm{d}\mu_b - 2\int_{\mathbb{R}^d}\xi_a\,\mathrm{d}\mu_a &\geq 2\int_{\mathbb{R}^d\times(a,b)}\Big(\partial_t^+\xi + |\mathrm{D}_x\xi_t(x)|^2 + 4\xi_t^2(x)\Big)\,\mathrm{d}\mu_I\\
&\overset{(8.70)}{=} 2\int_{\mathbb{R}^d\times(a,b)}\Big(\frac{1}{2}|\mathrm{D}_x\xi_t(x)|^2 + 2\xi_t^2(x)\Big)\,\mathrm{d}\mu_I \overset{(8.64)}{\geq} \int_a^b |\mu_t'|^2\,\mathrm{d}t\\
&\geq \frac{1}{b-a}\mathsf{HK}^2(\mu_a,\mu_b).
\end{aligned}$$

On the other hand, since $\mathbb{R}^d$ is a length space, Theorem 8.13 yields

$$\frac{1}{b-a}\mathsf{HK}^2(\mu_a,\mu_b) \geq 2\int_{\mathbb{R}^d}\xi_b\,\mathrm{d}\mu_b - 2\int_{\mathbb{R}^d}\xi_a\,\mathrm{d}\mu_a$$

so that all the above inequalities are in fact identities and

$$\mathsf{HK}(\mu_a, \mu_b) = (b-a)\,|\mu_t'| \quad \mathscr{L}^1\text{-a.e. in } [a,b],$$

which shows that μ is a geodesic. Passing to the limit as $a \downarrow 0$ and $b \uparrow 1$ we conclude the proof of the first part of the Theorem.

Necessity. Let $(\mu_t)_{t\in[0,1]}$ be a HK-geodesic in $\mathcal{M}(\mathbb{R}^d)$ connecting μ_0 to μ_1; applying Theorem 8.17 we can find a Borel vector field $(\boldsymbol{v}, w) \in \mathrm{L}^2(\mathbb{R}^d \times (0,1), \mu_I; \mathbb{R}^{d+1})$ such that (8.58) and (8.61) hold. We also consider an optimal plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$.

Let $\psi_1, \psi_2 : \mathbb{R}^d \to [-\infty, 1]$ be a pair of optimal potentials given by Theorem 6.3 d) and let us set $\xi := -\frac{1}{2}\psi_1$, $\xi_t := \mathscr{P}_t\xi$, $t \in (0,1)$. Even if we are considering more general initial data $\xi \in \mathrm{B}(X; [-1/2, +\infty])$ in (8.44), it is not difficult to check that the same statement of Theorem 8.12 holds in every subinterval $[a,b]$ with $0 < a < b < 1$ and

$$\lim_{t\downarrow 0} \mathscr{P}_t\xi(x) = \sup_{t>0} \mathscr{P}_t\xi(x) = \xi_*(x), \quad \text{where} \quad \xi_*(x) := \lim_{r\downarrow 0} \inf_{x'\in B_r(x)} \xi(x') \tag{8.72}$$

is the l.s.c. envelope of ξ. Moreover, setting

$$\xi_1(x) = \mathscr{P}_1\xi(x) := \lim_{t\uparrow 1} \xi_t(x) = \inf_{0<t<1} \xi_t(x), \tag{8.73}$$

the function ξ_1 is u.s.c., and optimality yields

$$\frac{1}{2}\psi_2(x) = \xi_1(x) \quad \text{for } \gamma_2\text{-a.a. } x \in \mathbb{R}^d. \tag{8.74}$$

By introducing the semigroup $\bar{\mathscr{P}}_t\xi := -\mathscr{P}_t(-\bar{\xi})$ and reversing time, we can define

$$\bar{\xi}_t := \bar{\mathscr{P}}_{1-t}. \tag{8.75}$$

By using the link with the Hopf-Lax semigroup in $\mathfrak{C}$ given by Theorem 8.12, the optimality of (ψ_1, ψ_2) and arguing as in [44, Thm. 7.36], it is not difficult to check that

$$\bar{\xi}_t \le \xi_t \quad \text{in } \mathbb{R}^d, \quad \bar{\xi}_0 = \xi_0 = \frac{1}{2}\psi_1 \quad \mu_0\text{-a.e. in } \mathbb{R}^d. \tag{8.76}$$

Notice that the function $x \mapsto -\cos^2(|x - x'| \wedge \pi/2)$ has bounded first and second derivatives, so it is semiconcave. It follows that the map $x \mapsto \xi_t(x)$ is semiconcave for every $t \in (0,1)$ and $x \mapsto \bar{\xi}_t(x)$ is semiconvex.

Since $t \mapsto \int \xi_t\,\mathrm{d}\mu_t$ and $t \mapsto \int \bar{\xi}_t\,\mathrm{d}\mu_t$ are absolutely continuous in $(0,1)$, Theorem 8.13(i) yields that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \xi_t\,\mathrm{d}\mu_t \le \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1), \tag{8.77}$$

so that

$$\int \xi_b\,\mathrm{d}\mu_b - \int \xi_a\,\mathrm{d}\mu_a \le \frac{b-a}{2}\mathsf{HK}^2(\mu_0, \mu_1).$$

Passing to the limit first as $a \downarrow 0$ and then as $b \uparrow 1$ by monotone convergence (notice that $\xi_t \le 1/2$)12 $\rightsquigarrow 1/2$ and using optimality once again we obtain

$$\begin{aligned}\mathsf{HK}^2(\mu_0, \mu_1) &= \int \xi_1\,\mathrm{d}\mu_0 + \int \xi_2\,\mathrm{d}\mu_1 = 2\int \xi_1\,\mathrm{d}\mu_1 - 2\int \xi_0\,\mathrm{d}\mu_0 \\ &= \lim_{a\downarrow 0, b\uparrow 1} 2\Big(\int \xi_b\,\mathrm{d}\mu_b - \int \xi_a\,\mathrm{d}\mu_a\Big).\end{aligned} \tag{8.78}$$

It follows by (8.77) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \xi_t\,\mathrm{d}\mu_t = \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0,\mu_1) \quad \text{in } (0,1). \tag{8.79}$$

Reversing time, a similar argument yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \bar\xi_t\,\mathrm{d}\mu_t = \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0,\mu_1) \quad \text{in } (0,1). \tag{8.80}$$

We have proved that the maps $t\mapsto \int \xi_t\,\mathrm{d}\mu_t$ and $t\mapsto \int\bar\xi_t\,\mathrm{d}\mu_t$ are affine in $[0,1]$ and coincide at $t=0,1$ so that

$$\int \xi_t\,\mathrm{d}\mu_t = \int \bar\xi_t\,\mathrm{d}\mu_t \quad \text{for every } t\in[0,1]. \tag{8.81}$$

Recalling (8.76), we deduce that the complement of the set $Z_t := \{x\in\mathbb{R}^d : \xi_t(x)=\bar\xi_t(x)\}$ is μ_t-negligible; since for $t\in(0,1)$ ξ_t is Lipschitz and semiconcave (thus everywhere superdifferentiable) and $\bar\xi_t$ is Lipschitz and semiconvex (thus everywhere subdifferentiable) we conclude that ξ_t is strictly differentiable in Z_t, so that it satisfies conditions (i), (ii).

Since (iii) is guaranteed by Theorem 8.12 ($\mathbb{R}^d$ is a length space), it remains to check (8.71). We apply the following Lemma 8.21 by observing that [3, Prop. 3.2,3.3] and Theorem 8.12 yield

$$\limsup_{x'\to x}\partial_t^+\xi_t(x') \le \limsup_{x'\to x}\partial_t^-\xi_t(x') \le \partial_t^-\xi_t(x), \quad \partial_t^-\xi_t(x)=\partial_t^+\xi_t(x) \quad \mu_I\text{-a.e.}$$

so that (8.82) holds as an identity.

Recalling that $|\mathrm{D}\xi_t|_a^2(x) = |\mathrm{D}_x\xi_t(x)|^2$ at every point of Z_t, for every $0<a<b<1$ we have

$$\begin{aligned}
\frac{b-a}{2}\mathsf{HK}^2(\mu_0,\mu_1) &= \int_{\mathbb{R}^d}\xi_b\,\mathrm{d}\mu_b - \int_{\mathbb{R}^d}\xi_a\,\mathrm{d}\mu_a \overset{(8.82)}{=} \int_{\mathbb{R}^d\times(a,b)}\Big(\partial_t^+\xi + \mathrm{D}_x\xi\,\boldsymbol{v} + \xi w\Big)\,\mathrm{d}\mu_I\\
&= \int_{\mathbb{R}^d\times(a,b)}\Big(-\frac{1}{2}|\mathrm{D}_x\xi_t|^2 - 2\xi_t^2 + \mathrm{D}_x\xi\,\boldsymbol{v} + \xi w\Big)\,\mathrm{d}\mu_I\\
&= \int_{\mathbb{R}^d\times(a,b)}\Big(-\frac{1}{2}|\mathrm{D}_x\xi_t-\boldsymbol{v}|^2 - 2(\xi_t-\frac{1}{4}w)^2 + \frac{1}{2}|\boldsymbol{v}|^2 + \frac{1}{8}w^2\Big)\,\mathrm{d}\mu_I\\
&\overset{(8.61)}{\le} -\int_{\mathbb{R}^d\times(a,b)}\Big(\frac{1}{2}|\mathrm{D}_x\xi_t-\boldsymbol{v}|^2 + 2(\xi_t-\frac{1}{4}w)^2\Big)\,\mathrm{d}\mu_I + \frac{1}{2}\int_a^b|\mu_t'|^2\,\mathrm{d}t
\end{aligned}$$

We deduce that $\boldsymbol{v}=\mathrm{D}_x\xi$ and $w=4\xi$ holds μ_I-a.e. □

Lemma 8.21. *Let $\mu\in\mathrm{AC}^2_{\mathrm{loc}}(0,1);(\mathcal{M}(\mathbb{R}^d),\mathsf{HK}))$ be satisfying the continuity equation with reaction (8.58) governed by the field $(\boldsymbol{v},w)\in L^2(\mathbb{R}^d\times(a,b),\mu_I)$ for every $[a,b]\subset(0,1)$. If $\xi\in\mathrm{Lip}_{\mathrm{loc}}((0,1);\mathrm{C}_b(\mathbb{R}^d))$ satisfies conditions (i, ii, iii) of Theorem 8.20 then for every $0<a\le b<1$*

$$\int_{\mathbb{R}^d}\xi_b\,\mathrm{d}\mu_b - \int_{\mathbb{R}^d}\xi_a\,\mathrm{d}\mu_a \ge \int_{\mathbb{R}^d\times(a,b)}\Big(\partial_t^+\xi + \mathrm{D}_x\xi\,\boldsymbol{v} + \xi w\Big)\,\mathrm{d}\mu_I. \tag{8.82}$$

If moreover $\limsup_{x'\to x}\partial_t^+\xi_t(x') \le \partial_t^+\xi_t(x)$ for μ_I-a.e. $(x,t)\in\mathbb{R}^d\times(0,1)$, then equality holds in (8.82).

Proof. We fix a compact interval $[a,b]\subset(0,1)$ and we set $M=\max_{t\in[a,b]}\mu_t(\mathbb{R}^d)$ and $L:=\mathrm{Lip}(\xi;\mathbb{R}^d\times[a,b])+\sup_{\mathbb{R}^d\times[a,b]}|\xi|$.

We regularize ξ by space convolution as in (8.53) by setting $\xi^\varepsilon:=\xi*\kappa_\varepsilon$ and we will perform a further regularization in time:

$$\xi^{\varepsilon,\tau}_t(x):=\frac{1}{\tau}\int_0^\tau \xi^\varepsilon_{t+r}(x)\,\mathrm{d}r,\quad \tau>0. \tag{8.83}$$

Since $\xi^{\varepsilon,\tau}\in \mathrm{C}^1(\mathbb{R}^d\times[a,b])$, arguing as in the proof of Theorem 8.18 we get for every $\varepsilon>0$ and $\tau\in(0,1-b)$

$$\int_{\mathbb{R}^d}\xi^{\varepsilon,\tau}_b\,\mathrm{d}\mu_b-\int_{\mathbb{R}^d}\xi^{\varepsilon,\tau}_a\,\mathrm{d}\mu_a=\int_{\mathbb{R}^d\times(a,b)}\Big(\partial_t\xi^{\varepsilon,\tau}+\mathrm{D}_x\xi^{\varepsilon,\tau}\,\boldsymbol{v}+\xi^{\varepsilon,\tau}w\Big)\,\mathrm{d}\mu_I. \tag{8.84}$$

We first pass to the limit as $\tau\downarrow 0$, observing that $\xi^{\varepsilon,\tau}\to\xi^\varepsilon$ uniformly since ξ^ε is bounded and Lipschitz; similarly, since $\mathrm{D}\xi^{\varepsilon,\tau}=(\mathrm{D}\xi^\varepsilon)^\tau$ and $\mathrm{D}\xi^\varepsilon$ is bounded and Lipschitz, we have $\mathrm{D}\xi^{\varepsilon,\tau}\to\mathrm{D}\xi^\varepsilon$ uniformly. Finally, since

$$\partial_t\xi^{\varepsilon,\tau}_t(x)=\frac{1}{\tau}(\xi^\varepsilon_{t+\tau}(x)-\xi^\varepsilon_t(x))=\int_{\mathbb{R}^d}\frac{1}{\tau}(\xi^\varepsilon_{t+\tau}(x')-\xi^\varepsilon_t(x'))\kappa_\varepsilon(x-x')\,\mathrm{d}x',$$

we have

$$\lim_{\tau\downarrow 0}\partial_t\xi^{\varepsilon,\tau}_t(x)=\partial_t^+\xi^\varepsilon_t(x)=((\partial_t^+\xi)*\kappa_\varepsilon)(x)\quad\text{for every }x\in\mathbb{R}^d. \tag{8.85}$$

Applying Lebesgue's Dominated Convergence Theorem we get

$$\int_{\mathbb{R}^d}\xi^\varepsilon_b\,\mathrm{d}\mu_b-\int_{\mathbb{R}^d}\xi^\varepsilon_a\,\mathrm{d}\mu_a=\int_{\mathbb{R}^d\times(a,b)}\Big(\partial_t^+\xi^\varepsilon+\mathrm{D}_x\xi^\varepsilon\,\boldsymbol{v}+\xi^\varepsilon w\Big)\,\mathrm{d}\mu_I. \tag{8.86}$$

(8.82) will follow by passing to the limit as $\varepsilon\downarrow 0$ in (8.86). First of all, we observe that ξ^ε converges uniformly to ξ since ξ is bounded and Lipschitz; moreover since $\lim_{\varepsilon\downarrow 0}\mathrm{D}_x\xi^\varepsilon_t(x)=\mathrm{D}_x\xi_t(x)$ at every point $x\in\mathbb{R}^d$ where ξ_t is strictly differentiable, we get

$$|\mathrm{D}_x\xi^\varepsilon\,\boldsymbol{v}|\le L|\boldsymbol{v}|\in \mathrm{L}^1(\mathbb{R}^d\times(a,b);\mu_I),\quad \lim_{\varepsilon\downarrow 0}\mathrm{D}_x\xi^\varepsilon=\mathrm{D}_x\xi\quad \mu_I\text{-a.e. in }\mathbb{R}^d\times[a,b]$$

so that

$$\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^d}\xi^\varepsilon_{a,b}\,\mathrm{d}\mu_{a,b}=\int_{\mathbb{R}^d}\xi_{a,b}\,\mathrm{d}\mu_{a,b},\quad \int\limits_{\mathbb{R}^d\times(a,b)}\Big(\mathrm{D}_x\xi^\varepsilon\,\boldsymbol{v}+\xi^\varepsilon w\Big)\,\mathrm{d}\mu_I=\int\limits_{\mathbb{R}^d\times(a,b)}\Big(\mathrm{D}_x\xi\,\boldsymbol{v}+\xi w\Big)\,\mathrm{d}\mu_I.$$

Finally, (8.70) shows that for every $t\in[0,1]$ the map $x\mapsto\partial_t^+\xi_t(x)$ is lower semicontinuous in $\mathbb{R}^d$, so that

$$\liminf_{\varepsilon\downarrow 0}\partial_t^+\xi^\varepsilon_t(x)\ge\partial_t^+\xi_t(x)\quad\text{for every }x\in\mathbb{R}^d;$$

since $\partial_t^+\xi^\varepsilon_t$ is also uniformly bounded Fatou's Lemma yields

$$\liminf_{\varepsilon\downarrow 0}\int_{\mathbb{R}^d\times(a,b)}\partial_t^+\xi^\varepsilon_t\,\mathrm{d}\mu_I\ge\int_{\mathbb{R}^d\times(a,b)}\partial_t^+\xi_t\,\mathrm{d}\mu_I.$$

The equality case follow by applying the Lebesgue's Dominated Convergence Theorem. □

8.7 Contraction properties: convolution and Heat equation in $\mathrm{RCD}(0,\infty)$ metric-measure spaces.

We conclude this paper with a few applications concerning contraction properties of the HK distance. The first one concerns the behavior with respect 1-Lipschitz maps.

Lemma 8.22. *Let* (X, d_X), (Y, d_Y) *be separable metric spaces and let* $f : X \to Y$ *be a* 1*-Lipschitz map. Then* $f_\sharp : \mathcal{M}(X) \to \mathcal{M}(Y)$ *is* 1*-Lipschitz w.r.t.* HK*:*

$$\mathsf{HK}(f_\sharp\mu_1, f_\sharp\mu_2) \le \mathsf{HK}(\mu_1, \mu_2). \tag{8.87}$$

Proof. It is sufficient to observe that the map $\mathfrak{f} : \mathfrak{C}_X \mapsto \mathfrak{C}_Y$ defined by $\mathfrak{f}([x,r]) := [f(x), s]$ satisfies $\mathsf{d}_{\mathfrak{C}_Y}(\mathfrak{f}([x_1,r_1]), \mathfrak{f}([x_2,r_2])) \le \mathsf{d}_{\mathfrak{C}_X}([x_1,r_1],[x_2,r_2])$ for every $[x_i,r_i] \in \mathfrak{C}_X$. $\mathfrak{f}_\sharp$ is thus a contraction from $(\mathcal{P}_2(\mathfrak{C}_X), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}_X}})$ to $(\mathcal{P}_2(\mathfrak{C}_Y), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}_Y}})$ and therefore $f_\sharp$ satisfies (8.87). □

A second application concerns convolutions in $\mathbb{R}^d$.

Theorem 8.23. *Let* $X = \mathbb{R}^d$ *with the Euclidean distance and let* $\nu \in \mathcal{M}(\mathbb{R}^d)$*. The map* $\mu \mapsto \mu * \nu$ *is contractive w.r.t.* HK *if* $\nu(\mathbb{R}^d) = 1$ *and, more generally,*

$$\mathsf{HK}^2(\mu_1 * \nu, \mu_2 * \nu) \le \nu(\mathbb{R}^d)\mathsf{HK}^2(\mu_1, \mu_2) \quad \text{for every } \mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d). \tag{8.88}$$

Proof. The previous Lemma shows that HK is invariant by isometries, in particular translations in $\mathbb{R}^d$, so that

$$\mathsf{HK}(\mu_1 * \delta_x, \mu_2 * \delta_x) = \mathsf{HK}(\mu_1, \mu_2) \quad \text{for every } \mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d),\ x \in \mathbb{R}^d.$$

By the subadditivity property it follows that if $\nu = \sum_k a_k \delta_{x_k}$ for some $a_k \ge 0$

$$\begin{aligned}\mathsf{HK}^2(\mu_1 * \nu, \mu_2 * \nu) &= \mathsf{HK}^2\Big(\sum_k a_k \mu_1 * \delta_{x_k}, \sum_k a_k \mu_2 * \delta_{x_k}\Big)\\ &\le \sum_k a_k \mathsf{HK}^2(\mu_1 * \delta_{x_k}, \mu_2 * \delta_{x_k}) = \sum_k a_k \mathsf{HK}^2(\mu_1, \mu_2) = \nu(\mathbb{R}^d)\mathsf{HK}^2(\mu_1, \mu_2).\end{aligned}$$

The general case then follows by approximating ν by a sequence of discrete measure ν_n converging to ν in $\mathcal{M}(\mathbb{R}^d)$ and observing that $\mu_i * \nu_n \to \mu_i * \nu$ weakly in $\mathcal{M}(\mathbb{R}^d)$. Since HK is weakly continuous we obtain (8.88). □

An easy application of the previous result is the contraction property of the (adjoint) Heat semigroup $(P_t^*)_{t\ge 0}$ in $\mathbb{R}^d$ with respect to HK. In fact, we can prove a much more general result for the Heat flow in $\mathrm{RCD}(0,\infty)$ metric measure spaces (X, d, m) [4, 5]. It covers the case of the semigroups $(P_t)_{t\ge0}$ generated by the Heat equation on a open convex domain $\Omega \subset \mathbb{R}^d$ with homogenous Neumann conditions

$$\partial_t u = \Delta u \quad \text{in } \Omega \times (0,\infty), \qquad \partial_n u = 0 \quad \text{on } \partial\Omega \times (0,\infty),$$

the Heat equation on a complete Riemannian manifold $(\mathbb{M}^d, g)$ with nonnegative Ricci curvature defined by

$$\partial_t u = \Delta_g u \quad \text{in } \mathbb{M}^d \times (0,\infty),$$

where Δ_g is the usual Laplace-Beltrami operator, and the Fokker-Planck equation in $\mathbb{R}^d$ generated by the gradient of a convex potentials $V : \mathbb{R}^d \to \mathbb{R}$

$$\partial_t u = \Delta_g u - \nabla \cdot (u\, \mathrm{D}V) \quad \text{in } \mathbb{R}^d \times (0,\infty).$$

Theorem 8.24. *Let (X, d, m) be a complete and separable metric-measure space with nonnegative Riemannian Ricci Curvature, i.e. satisfying the* $\mathrm{RCD}(0,\infty)$ *condition, and let $(P_t^*)_{t\ge 0} : \mathcal{M}(X) \to \mathcal{M}(X)$ be the Heat semigroup in the measure setting. Then*

$$\mathsf{HK}(P_t^*\mu_1, P_t^*\mu_2) \le \mathsf{HK}(\mu_1,\mu_2) \quad \textit{for every } \mu_1,\mu_2 \in \mathcal{M}(X). \tag{8.89}$$

Proof. Recall that in $\mathrm{RCD}(0,\infty)$ metric measure spaces the L^2-gradient flow of the Cheeger energy induces a symmetric Markov semigroup $(P_t)_{t\ge 0}$ in $L^2(X, m)$ which has a pointwise version satisfying the Feller regularization property $P_t(\mathrm{B}_b(X)) \subset \mathrm{Lip}_b(X)$ for $t > 0$, and the estimate

$$|\mathrm{D}_X P_t f|^2(x) \le P_t\big(|\mathrm{D}_X f|^2\big)(x) \quad \text{for every } f \in \mathrm{Lip}_b(X),\ x \in X,\ t \ge 0. \tag{8.90}$$

Its adjoint $(P_t^*)_{t\ge 0}$ coincides with the Kantorovich-Wasserstein gradient flow in $\mathcal{P}_2(X)$ of the Entropy Functional $\mathscr{F}(\cdot|m)$ where $\mathscr{F}$ is induced by $F(s) = U_1(s)$ and defines a semigroup in $\mathcal{M}(X)$ by the formula

$$\int_X f \,\mathrm{d}(P_t^*\mu) = \int_X P_t f \,\mathrm{d}\mu \quad \text{for every } f \in \mathrm{B}_b(X),\ \mu \in \mathcal{M}(X). \tag{8.91}$$

In order to prove (8.89) we use (8.49) (RCD-spaces satisfy the length property) and apply P_t to a subsolution $(\psi_\theta)_{\theta\in[0,1]}$ in $\mathrm{C}^1([0,1]; \mathrm{Lip}_b(X))$ of the Hamilton-Jacobi equation

$$\partial_\theta \psi_\theta + \frac{1}{4}|\mathrm{D}_X\psi_\theta|^2 + \psi_\theta^2 \le 0 \quad \text{in } X \times (0,1). \tag{8.92}$$

Since P_t is a linear and continuous map from $\mathrm{Lip}_b(X)$ to $\mathrm{Lip}_b(X)$ the curve $\theta \mapsto \psi_{\theta,t} := P_t(\psi_\theta)$ belongs to $\mathrm{C}^1([0,1]; \mathrm{Lip}_b(X))$. (8.90) and the Markov property yield

$$|\mathrm{D}_X P_t\psi_\theta|^2(x) \le P_t\big(|\mathrm{D}_X\psi_\theta|^2\big)(x), \quad (P_t\psi_\theta)^2(x) \le P_t(\psi_\theta^2)(x) \quad \text{for every } x \in X,\ \theta \in [0,1],\ t \ge 0.$$

We thus get for every $t \ge 0$

$$\partial_\theta \psi_{\theta,t} + \frac{1}{4}|\mathrm{D}_X\psi_{\theta,t}|^2 + \psi_{\theta,t}^2 \le 0 \quad \text{in } X \times (0,1)$$

and therefore

$$\int_X \psi_1 \,\mathrm{d}(P_t^*)\mu_1 - \int_X \psi_0 \,\mathrm{d}(P_t^*)\mu_0 = \int_X P_t\psi_1 \,\mathrm{d}\mu_1 - \int_X P_t\psi_0 \,\mathrm{d}\mu_0 \le \mathsf{HK}^2(\mu_1,\mu_0).$$

We conclude by taking the supremum with respect to all the subsolutions of (8.92) in $\mathrm{C}^1([0,1]; \mathrm{Lip}_b(X))$ and applying (8.49). □

A On the chronological development of our theory

In this section we give a brief account of the order in which we developed the different parts of the theory. The beginning was the mostly formal work in [27] on reaction-diffusion systems, where a distance on vectors $\boldsymbol{u}$ of densities over a domain $\Omega \subset \mathbb{R}^d$ was formally defined in the Benamou-Brenier sense via

$$\mathsf{d}(\boldsymbol{u}_0, \boldsymbol{u}_1)^2 = \inf \int_0^1 \int_\Omega \Xi_t : \mathbb{M}_{\mathrm{diff}}(\boldsymbol{u}_t)\Xi_t + \boldsymbol{\xi}_t \cdot \mathbb{K}_{\mathrm{react}}(\boldsymbol{u}_t)\boldsymbol{\xi}_t \mathrm{d}x\mathrm{d}t$$

under the constraint of the continuity equation $\partial_t \boldsymbol{u}_t + \mathrm{div}\big(\mathbb{M}_{\mathrm{diff}}(\boldsymbol{u}_t)\Xi_t\big) = \mathbb{K}_{\mathrm{react}}(\boldsymbol{u}_t)\boldsymbol{\xi}_t$. The central question was and still is the understanding of diffusion equations with reactions in the gradient-flow form $\partial_t \boldsymbol{u} = \mathrm{div}\Big(\mathbb{M}_{\mathrm{diff}}(\boldsymbol{u})\nabla\delta\mathcal{F}(\boldsymbol{u})\Big) - \mathbb{K}_{\mathrm{react}}(\boldsymbol{u})\delta\mathcal{F}(\boldsymbol{u})$, see [27, Sect. 5.1].

It was natural to treat the scalar case first and to restrict to the case where both mobility operator $\mathbb{M}_{\mathrm{diff}}(u)$ and $\mathbb{K}_{\mathrm{react}}(u)$ are linear in u. Only in that case the formally derived system (1.30) for the geodesics (u_t, ξ_t) decouples in the sense that ξ_t solves an Hamilton-Jacobi equation that does not depend on u. Choosing $\mathbb{M}_{\mathrm{diff}}(u) = \alpha u$ and $\mathbb{K}_{\mathrm{react}}(u) = \beta u$ with $\alpha, \beta \geq 0$, the relevant Hamilton-Jacobi equation reads

$$\partial_t \xi_t + \frac{\alpha}{2}|\mathrm{D}_x \xi_t|^2 + \frac{\beta}{2}\xi_t^2 = 0.$$

As in the other parts of this paper, we restrict to the case $\alpha = 1$ and $\beta = 4$ subsequently, but refer to [27] for the general case. Thus, the conjectured characterization (8.49) was first presented in Pisa at the Workshop "Optimal Transportation and Applications" in November 2012.

During a visit of the second author in Pavia, the generalized Hopf-Lax formula via the nonlinear convolution $\mathscr{P}_t$ (cf. (8.44)) was derived via the classical method of characteristics. This led to the unsymmetric representation (1.26) for HK. To symmetrize this relation we used that $\mathscr{P}_1\xi(x) = \inf \Phi(\xi(y), |y{-}x|)$ with $\Phi(z, R) = \frac{1}{2}\big(1{-}\frac{A(R)}{1+2z}\big)$, where $A(R) = \cos^2\big(R\wedge(\pi/2)\big)$. Setting $\psi_0 = -2\xi_0$ and $\psi_1 = 2\xi_1 = 2\mathscr{P}_1$, we have the equivalence

$$\xi_1 = \mathscr{P}_1\xi_0 \quad \Longleftrightarrow \quad (1{-}\psi_0(x_0))(1{-}\psi_1(x_1)) \geq A(|x_0{-}x_1|) \text{ for all } x_i.$$

Setting $\varphi_i = -\log(1{-}\psi_i)$ we arrived at the cost function

$$\mathsf{c}(x_0, x_1) = -\log A(|x_0{-}x_1|) = \begin{cases} -2\log\big(\cos|x_0{-}x_1|\big) & \text{for } |x_0{-}x_1| < \pi/2, \\ \infty & \text{otherwise,} \end{cases}$$

for the first time and obtained the characterization (1.7), namely

$$\mathsf{HK}(\mu_0, \mu_1)^2 = \mathsf{D}(\mu_0, \mu_1) = \sup\big\{\mathscr{D}(\varphi_0, \varphi_1|\mu_0, \mu_1) \ : \ \varphi_0 \oplus \varphi_1 \leq \mathsf{c}\big\}.$$

It was then easy to dualize $\mathscr{D}$, and the Logarithmic Entropy functional LET in (1.20) was derived in July 2013.

While the existence of minimizers for $\mathsf{LET}(\mu_0, \mu_1) = \min \mathscr{E}(\gamma|\mu_0, \mu_1)$ was easily obtained, it was not clear at all, why and how HK defined via $\mathsf{HK}^2(\mu_0, \mu_1) = \min \mathscr{E}(\cdot|\mu_0, \mu_1)$ generates a geodesic distance. The only thing which could easily be checked was that the minimum was consistent with the distance between two Dirac masses, which could easily be calculated via the dynamic formulation.

So, in parallel we tried to develop the dynamic approach, which was not too successful at the early stages. Only after realizing and exploiting the connection to the cone distance in Summer and Autumn of 2013 we were able to connect LET systematically with the dynamic approach. The crucial and surprising observation was that optimal plans for $\mathscr{E}$ and lifts of measures $\mu \in \mathcal{M}(X)$ to measures λ on the cone $\mathfrak{C}$ could be identified by exploiting the optimality conditions systematically. Corresponding results were presented in workshops on Optimal Transport in Banff (June 2014) and Pisa (November 2014).

Already at the Banff workshop, the general structure of the primal and dual Entropy-Transport problem as well as the homogeneous perspective formulation were presented. Several examples and refinements where developed afterwards. The most recent part from Summer 2015 concerns our Hamilton-Jacobi equation in general metric spaces (X, d) and the induced cone $\mathfrak{C}$ (cf. Section 8.4) and the derivation of the geodesic equations in $\mathbb{R}^d$ (cf. Section 8.6). This last achievement now closes the circle, by showing that all the initial steps, which were done on a formal level in 2012 and the first half of 2013, have indeed a rigorous interpretation.

References

[1] L. Ambrosio, M. Erbar, and G. Savaré, *Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces*, Arxiv preprint, arXiv 1506.05932, (2015).

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[3] ———, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math., 195 (2014), pp. 289–391.

[4] ———, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Duke Math. J., 163 (2014), pp. 1405–1490.

[5] ———, *Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds*, Ann. Probab., 43 (2015), pp. 339–404.

[6] L. Ambrosio, A. Mondino, and G. Savaré, *Nonlinear diffusion equations and curvature conditions in metric measure spaces*, In preparation, (2015).

[7] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[8] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[9] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

[10] Y. Burago, M. Gromov, and G. Perel′man, *A. D. Aleksandrov spaces with curvatures bounded below*, Uspekhi Mat. Nauk, 47 (1992), pp. 3–51, 222.

[11] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *An interpolating distance between optimal transport and Fisher–Rao*, arXiv:1506.06430v2, (2015).

[12] ———, *Unbalanced optimal transport: geometry and Kantorovich formulation*, arXiv:1508.05216v1, (2015).

[13] F. H. Clarke, *Optimization and nonsmooth analysis*, John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.

[14] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math., 146 (2001), pp. 219–257.

[15] I. Csiszár, *Information-type measures of difference of probability distributions and indirect observations*, Studia Sci. Math. Hungar., 2 (1967), pp. 299–318.

[16] J. Dolbeault, B. Nazaret, and G. Savaré, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations, 34 (2009), pp. 193–231.

[17] R. M. Dudley, *Real analysis and probability*, vol. 74 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[18] M. Erbar, K. Kuwada, and K.-T. Sturm, *On the Equivalence of the Entropic Curvature-Dimension Condition and Bochner's Inequality on Metric Measure Spaces*, ArXiv e-prints, (2013).

[19] E. Hellinger, *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen*, J. Reine Angew. Math., 136 (1909), pp. 210–271.

[20] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17 (electronic).

[21] S. Kakutani, *On equivalence of infinite product measures*, Ann. of Math. (2), 49 (1948), pp. 214–224.

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

[23] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov, *A new optimal transport distance on the space of finite radon measures*, ArXiv preprint, arxiv 1505.07746v2, (2015).

[24] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov, *A new optimal transport distance on the space of finite radon measures*, arXiv 1505.07746v2, (2015).

[25] S. Kullback and R. A. Leibler, *On information and sufficiency*, Ann. Math. Statistics, 22 (1951), pp. 79–86.

[26] L. LeCam, *Convergence in distribution of stochastic processes*, Univ. Calif. Publ. Statist., 2 (1957), pp. 207–236.

[27] M. Liero, A. Mielke, and G. Savaré, *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, in preparation, (2015).

[28] F. Liese and I. Vajda, *On divergences and informations in statistics and information theory*, IEEE Trans. Inform. Theory, 52 (2006), pp. 4394–4412.

[29] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. Inform. Theory, 37 (1991), pp. 145–151.

[30] S. Lisini, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 85–120.

[31] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169 (2009), pp. 903–991.

[32] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[33] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal., 173 (2000), pp. 361–400.

[34] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I–II*, Probability and its Applications, Springer-Verlag, New York, 1998.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[36] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Springer, 2015.

[37] W. Schachermayer and J. Teichmann, *Characterization of optimal transport plans for the Monge-Kantorovich problem*, Proc. Amer. Math. Soc., 137 (2009), pp. 519–529.

[38] L. Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973. Tata Institute of Fundamental Research Studies in Mathematics, No. 6.

[39] S. Simons, *Minimax and monotonicity*, vol. 1693 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1998.

[40] K.-T. Sturm, *Metric spaces of lower bounded curvature*, Exposition. Math., 17 (1999), pp. 35–47.

[41] ——, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

[42] ——, *On the geometry of metric measure spaces. II*, Acta Math., 196 (2006), pp. 133–177.

[43] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[44] ——, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.