CrossMark

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

Matthias Liero[1] · **Alexander Mielke[1,2]** ·
Giuseppe Savaré[3]

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. These problems arise quite naturally by relaxing the marginal constraints typical of Optimal Transport problems: given a pair 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, which quantify in some way the deviation of the marginals of the transport plan from the assigned measures. As a powerful application of this theory,

✉ Giuseppe Savaré
    giuseppe.savare@unipv.it

    Matthias Liero
    liero@wias-berlin.de

    Alexander Mielke
    mielke@wias-berlin.de

[1] Weierstraß-Institut Für Angewandte Analysis Und Stochastik, Berlin, Germany

[2] Humboldt-Universität Zu, Berlin, Germany

[3] Dipartimento Di Matematica "F. Casorati", Università Di Pavia, Pavia, Italy

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

Main notation

$\mathcal{M}(X)$	Finite positive Radon measures on a Hausdorff topological space X
η_0	The null measure in some $\mathcal{M}(X)$
$\mathcal{P}(X),\ \mathcal{P}_2(X)$	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 $T : X \to Y$, see (2.5)
$\gamma = \sigma\mu + \mu^\perp,\ \mu = \varrho\gamma + \gamma^\perp$	Lebesgue decompositions of γ and μ, Lemma 2.3
$\mu \llcorner A$	Restriction of a measure to A: $\mu \llcorner A(B) := \mu(A \cap B)$, Sect. 2.1
$C_b(X)$	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 (2.13), (2.14)
$F(s),\ F_i(s)$	Admissible entropy functions
$F^*(\phi),\ F_i^*(\phi)$	Legendre transform of F, F_i, see (2.16)
$F^\circ(\varphi),\ F_i^\circ(\varphi_i)$	Concave conjugate of an entropy function, see (2.45)
$R(r),\ R_i(r_i)$	Reversed entropies, see (2.28)
$\mathrm{D}(F)$	Proper domain $\{F < \infty\}$ of an extended real function F, see (2.14)

$\mathring{\mathrm{D}}(F)$	The interior of D(F)
$F'_\infty, F'_0, \mathrm{aff}\, F_\infty$	Recession, derivative at 0, and asymptotic coefficient of F, see (2.15)
I_K	The indicator function of a set K (0 on K and $+\infty$ outside)
χ_A	The characteristic function of a set A (1 on A and 0 outside)
$H_c(r_1, r_2),\ H(x_1, r_1; x_2, r_2)$	Marginal perspective function, see (5.1), (5.10), (5.3)
$\mathsf{c}(x_1, x_2)$	Lower semicontinuous cost function defined in $X = X_1 \times X_2$
$\mathsf{d}(x_1, x_2),\ \mathsf{d}_a(x_1, x_2)$	A metric on a space X and its truncated version $\mathsf{d} \wedge a$, see (6.8)
$\mathscr{F}(\gamma|\mu),\ \mathscr{R}(\mu|\gamma)$	Entropy functionals and their reverse form, see (2.35) and (2.57)
$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2),\ \mathsf{ET}(\mu_1, \mu_2)$	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, see Sect. 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 $\mathcal{M}(X)$, see Section 7.3
$\mathsf{GHK}(\mu_1, \mu_2)$	Gaussian Hellinger–Kantorovich distance in $\mathcal{M}(X)$, see 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 $\mathfrak{C}$
$\mathfrak{h}_i^2,\ \mathrm{dil}_{\theta,2}(\cdot)$	Homogeneous marginals and dilations, see (7.16), (7.17)
$\mathfrak{H}_=^2(\mu_1, \mu_2),\ \mathfrak{H}_\le^2(\mu_1, \mu_2)$	Plans in $\mathfrak{C} \times \mathfrak{C}$ with constrained homogeneous marginals, see (7.21)
$\mathrm{AC}^p([0, 1]; X)$	Space of curves $\mathrm{x} : [0, 1] \to X$ with p-integrable metric speed
$|\mathrm{x}'|_\mathsf{d}$	Metric speed of a curve $\mathrm{x} \in \mathrm{AC}([a, b]; (X, \mathsf{d}))$, see Sect. 8.1
$|\mathsf{D}_Z f|,\ |\mathsf{D}_Z f|_a$	Metric slope and asymptotic Lipschitz constant in Z, see (8.37)

When no ambiguity is possible, we will often adopt the convention to write the integral of a composition of functions as

$$\int F(\phi)\,\mathrm{d}\mu = \int (F \circ \phi)\,\mathrm{d}\mu = \int F(\phi(x))\,\mathrm{d}\mu(x).$$

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, see Remark 8.19 for a discussion of inf-convolutions of 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 [2,40,49,50] deal with minimization of a linear cost functional

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

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 [26,40]), 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

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

Starting from the pioneering work of Kantorovich, an impressive theory has been developed in the last two decades: from one side, typical intrin-

sic 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 [42,50] 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 $\pi_\sharp^i \gamma = \mu_i$ by introducing suitable penalizing functionals $\mathscr{F}_i$, that quantify in some way the deviation from μ_i of the marginals $\gamma_i := \pi_\sharp^i \gamma$ of γ. In this paper we consider the general case of integral functionals (also called *Csiszàr f-divergences* [17]) of the form

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

where $F_i : [0, \infty) \to [0, \infty]$ are given convex entropy functions and $(F_i)'_\infty$ are their recession constants, see (2.15). Typical examples are the logarithmic or power-like entropies

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

or for the total variation functional corresponding to the nonsmooth entropy $V(s) := |s - 1|$, considered in [38]. We shall see 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

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

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

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

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 Sect. 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

$$\mathsf{D}(\mu_1, \mu_2) := \sup \Big\{ \mathscr{D}(\varphi_1, \varphi_2 | \mu_1, \mu_2) : \ \varphi_i \colon X_i \to \mathbb{R},$$
$$\varphi_1(x_1) + \varphi_2(x_2) \leq \mathsf{c}(x_1, x_2) \Big\},$$
$$\text{where } \mathscr{D}(\varphi_1, \varphi_2 | \mu_1, \mu_2) := \int_{X_1} \left(1 - \mathrm{e}^{-\varphi_1}\right) \mathrm{d}\mu_1 + \int_{X_2} \left(1 - \mathrm{e}^{-\varphi_2}\right) \mathrm{d}\mu_2, \tag{1.7}$$

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

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

whose form can be explicitly calculated in terms of the Lagrangian conjugates F_i^* of the entropy functions. Thus (1.7) consists in the supremum of a concave functional on a convex set described by a system of affine inequalities.

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

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

It will be useful to have both the representations at our disposal: (1.7) naturally appears from the application of the von Neumann min–max principle from a saddle point formulation of the primal problem (1.5). Moreover, (1.8)–(1.10) will play an important role in the dynamic version of a particular case of ET, the Hellinger–Kantorovich distance that we will introduce later on.

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 pair 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

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

and they are necessary and sufficient for optimality.

The study of optimality conditions reveals a different behavior between pure transport problems and 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, in particular concerning the metric and dynamical aspects of such problems. It is related to the 1-***homogeneous Marginal Perspective function***

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

and to the corresponding integral functional

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

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

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

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

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

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

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

(1.13) provides an equivalent formulation of the Hellinger–Kakutani distance [22,25], see also Example E.5 in Sect. 3.3.

Other choices, still in the simple class (1.4), give raise to "transport" versions of well known functionals (see e.g. [31] for a systematic presentation): starting from the reversed entropies $F_i(s) = s - 1 - \log s$ one gets

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

which in the extreme case (1.15) reduces to the *Jensen–Shannon divergence* [32], a squared distance between measures derived from the celebrated *Kullback-Leibler divergence* [28]. The quadratic entropy $F_i(s) = \frac{1}{2}(s - 1)^2$ produces

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

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* [31], and still obtained by (1.12) when c has the form (1.15).

Also nonsmooth cases, as for $V(s) = |s-1|$ associated to the *total variation distance* (or nonsymmetric choices of F_i) can be covered by the general theory. In the case of $F_i(s) = V(s)$ the marginal perspective function is

$$H(x_1, r_1; x_2, r_2) = r_1 + r_2 - \left(2 - \mathsf{c}(x_1, x_2)\right)_+ (r_1 \wedge r_2)$$
$$= |r_2 - r_1| + (\mathsf{c}(x_1, x_2) \wedge 2)(r_1 \wedge r_2);$$

when $X_1 = X_2 = \mathbb{R}^d$ with $\mathsf{c}(x_1, x_2) := |x_1 - x_2|$ we recover the *generalized Wasserstein distance* $W_1^{1,1}$ introduced and studied by [38]; it provides an equivalent variational characterization of the flat metric [39].

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

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

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 [37] 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 in (1.3) along geodesics in $(\mathcal{P}_2(X), \mathsf{W}_{\mathsf{d}})$ [16,35,37] 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 [3–5,21,34,47,48].

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

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

the corresponding Logarithmic-Entropy Transport problem

$$\mathsf{LET}(\mu_1, \mu_2) := \min_{\boldsymbol{\gamma} \in \mathcal{M}(X)} \sum_i \int_X \big(\sigma_i \log \sigma_i - \sigma_i + 1\big) \, \mathrm{d}\mu_i$$

$$+ \int_{X^2} \ell\big(\mathsf{d}(x_1, x_2)\big) \, \mathrm{d}\boldsymbol{\gamma}, \quad \sigma_i = \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}, \qquad (1.20)$$

coincides with a (squared) distance in $\mathcal{M}(X)$ (which we will call *Hellinger–Kantorovich distance* and denote by HK) that can play the same fundamental role like the Kantorovich–Wasserstein distance for $\mathcal{P}_2(X)$.

Here is a (still non exhaustive) list of our main results of part II concerning the Hellinger–Kantorovich distance.

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

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

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

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

which is the so-called *cone distance* in the metric cone $\mathfrak{C}$ over X, cf. [10]. 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"

$$\mu = \mathfrak{h}^2 \alpha := \pi_{\sharp}^x(r^2 \alpha),$$
$$\int_X \zeta(x) \, \mathrm{d}\mu = \int_{\mathfrak{C}} \zeta(x) r^2 \, \mathrm{d}\alpha(x, r) \quad \text{with } \zeta \in C_b(X), \qquad (1.23)$$

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 $\alpha := \mu \otimes \delta_1$). The Hellinger–Kantorovich distance $\mathsf{HK}(\mu_1, \mu_2)$ can then be defined by taking the best Kantorovich–Wasserstein distance between all the possible lifts of μ_1, μ_2 in $\mathcal{P}_2(\mathfrak{C})$, i.e.

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

It turns out that (the square of) (1.24) yields an equivalent variational representation of the LET functional. In particular, (1.24) shows that in the case of concentrated measures

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

Notice that (1.24) resembles the very definition (1.18) of the Kantorovich–Wasserstein distance, where now the role of the marginals $\pi_\sharp^i$ 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 $\mathsf{LET} = \mathsf{HK}^2$.

(ii) 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 metric space, a property that is 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, thus it is topologically equivalent to the flat or Bounded Lipschitz distance [19, Sect. 11.3], see also [27, Thm. 3]. 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], [33]) 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.

(iii) 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

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

$$= \inf_{x' \in X} \frac{1}{t}\left(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x, x'))}{1 + 2t\xi(x')}\right).$$

(iv) 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

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

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

$$\frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) = \sup\left\{ \int_X \xi_1 \,\mathrm{d}\mu_1 - \int_X \xi_0 \,\mathrm{d}\mu_0 : \xi \in \mathrm{C}^k([0, 1]; \mathrm{Lip}_b(X)), \right.$$
$$\left. \partial_t\xi_t(x) + \frac{1}{2}|\mathrm{D}_X\xi_t|^2(x) + 2\xi_t^2(x) \le 0 \quad \text{in } X \times (0, 1)\right\},$$
$$(1.27)$$

which reproduces, at the level of HK, the nice link between W_d and Hamilton–Jacobi equations. One of the direct applications of (1.27) 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).

(v) Formula (1.27) clarifies that the HK distance can be interpreted as a sort of inf-convolution (see the Remark 8.19) 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

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

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

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

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

(vi) 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" [7,18] (see also the recent [27]) in $X = \mathbb{R}^d$:

$$\mathsf{HK}^2(\mu_0, \mu_1) = \min \Big\{ \int_0^1 \int \Big(|\boldsymbol{v}_t|^2 + \frac{1}{4}|w_t|^2 \Big) \mathrm{d}\mu_t \, \mathrm{d}t :$$

$$\mu \in \mathrm{C}([0, 1]; \mathcal{M}(\mathbb{R}^d)), \mu_{t=i} = \mu_i,$$

$$\partial_t \mu_t + \nabla \cdot (\boldsymbol{v}_t \mu_t) = w_t \mu_t \text{ in } \mathscr{D}'(\mathbb{R}^d \times (0, 1)) \Big\}. \tag{1.28}$$

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

$$\partial_t \mu_t + \nabla \cdot (\mathrm{D}_x \xi_t \mu_t) = 4\xi_t \mu_t, \quad \partial_t \xi_t + \frac{1}{2}|\mathrm{D}_x \xi^2|^2 + 2\xi_t^2 = 0 \tag{1.29}$$

holds for a suitable solution $\xi_t = \mathscr{P}_t \xi_0$. The representation (1.28) is the starting point for further investigations concerning the link to gradient systems and reaction-diffusion equations, the cone geometry, the representation of geodesics and of λ-convex integral functionals: we refer the interested reader to the examples collected in [30].

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 variations 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

$$\mathsf{GHK}^2(\mu_1, \mu_2) := \min \Big\{ \int \Big(r_1^2 + r_2^2 - 2r_1 r_2 \exp(-\mathsf{d}^2(x_1, x_2)/2) \Big) \mathrm{d}\boldsymbol{\alpha} \Big\} \tag{1.30}$$

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 [44]) 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 some examples and proving a general existence result for optimal plans. The "reverse" formulation of Theorem 3.11, though simple, justifies the importance of dealing with the largest class of entropies and will play a crucial role in Sect. 5.

Section 4 is devoted to finding the dual formulation, proving its equivalence with the primal problem (cf. Theorem 4.11), deriving sharp optimality conditions (cf. Theorem 4.6) and proving 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 in some detail.

Section 5 introduces the third formulation (1.12) based on the marginal perspective function (1.11) and its "homogeneous" version (Sect. 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 (Sect. 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 Sect. 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 review of the cone geometry (Sect. 7.1) we discuss in some detail the crucial notion of homogeneous marginals in Sect. 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 Sect. 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 Sect. 7.5. The equivalence with the LET formulation is the main achievement of Sect. 7.6 (Theorem 7.20), with applications to the duality formula (Theorem 7.21), to the comparisons with the classical Hellinger and Kantorovich distances (Sect. 7.7) and with the Gaussian Hellinger–Kantorovich distance (Sect. 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 Sect. 8.1, we derive the basic superposition principle in Theorem 8.4. This is the cornerstone to obtain a precise characterization of geodesics (Theorem 8.6), a sharp lower curvature bound in the Alexandrov sense (Theorem 8.8), and to prove the dynamic

characterization à la Benamou-Brenier of Sect. 8.5. The other powerful tool is provided by the duality with subsolutions to the Hamilton–Jacobi equation (Theorem 8.12), 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 Sect. 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 8.11 is interesting by itself and could find important applications in different contexts. From the point of view of Entropy-Transport problems, Theorem 8.11 is particularly relevant since it provides a dynamic interpretation of the dual characterization of the LET functional. In Sect. 8.6 we show that in the Euclidean case $X = \mathbb{R}^d$ all geodesic curves are characterized by the system (1.29). The last Sect. 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 developed 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, see Appendix A for some remarks concerning the chronological development of our theory.

In June 2015 the authors became aware of the parallel work [27], which mainly concerns the dynamical approach to the Hellinger–Kantorovich distance discussed in Sect. 8.5 and the metric-topological properties of Sect. 7.5 in the Euclidean case.

Moreover, in mid August 2015 they became aware of the works [13,14], which start from the dynamical formulation of the Hellinger–Kantorovich distance in the Euclidean case, prove existence of geodesics and sufficient optimality and uniqueness conditions (which we state in a stronger form in Sect. 8.6) with a precise characterization in the case of a pair of Dirac masses. Moreover, they provide a detailed discussion of curvature properties following Otto's formalism [36], and study more general dynamic costs on the cone space with their equivalent primal and dual static formulation (leading to characterizations analogous to (7.1) and (6.14) in the Hellinger–Kantorovich case).

Apart from the few above remarks, these independent works did not influence the first (cf. arXiv1508.07941v1) and the present version of this manuscript, which is essentially a minor modification and correction of the first version. In the final Appendix A we give a brief account of the chronological development of our theory.

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 [44], i.e. σ-additive set functions $\mu : \mathcal{B}(X) \to [0, \infty)$ such that

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

The restriction $B \mapsto \mu(B \cap A)$ of a Radon measure μ to a Borel set A will be denoted by $\mu \llcorner A$.

Radon measures have strong continuity properties with respect to monotone convergence. For this, denote by $\mathrm{LSC}(X)$ the space of all lower semicontinuous real-valued functions on X and consider a nondecreasing directed family $(f_\lambda)_{\lambda \in \mathbb{L}} \subset \mathrm{LSC}(X)$ (where $\mathbb{L}$ is a possibly uncountable directed set) of nonnegative and lower semicontinuous functions f_λ converging to f, we have (cf. [44, Prop. 5, p. 42])

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

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 [44, 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 [44, 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 $C_b(X)$. However, when (X, τ) is completely regular, i.e.

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

(in particular when τ is metrizable), narrow and weak topology coincide [44, p. 371]. Therefore when (X, τ) is a Polish space we recover the usual setting of Borel measures endowed with the weak topology.

We now turn to the compactness properties of subsets of $\mathcal{M}(X)$. Let us first recall that a set $\mathcal{K} \subset \mathcal{M}(X)$ is *bounded* if $\sup_{\mu \in \mathcal{K}} \mu(X) < \infty$; it is *equally tight* if

$$\forall \varepsilon > 0 \quad \exists K_\varepsilon \subset X \text{ compact} \quad \forall \mu \in \mathcal{K} : \mu(X \setminus K_\varepsilon) \leq \varepsilon. \qquad (2.4)$$

Compactness with respect to narrow topology is guaranteed by an extended version of *Prokhorov's Theorem* [44, Thm. 3, p. 379]. Tightness of weakly convergent *sequences* in metrizable spaces is due to LE CAM [29].

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* [44, Ch. I, Sect. 5] if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that $\mu(X \setminus K_\varepsilon) \leq \varepsilon$ and the restriction of T to K_ε is continuous. We denote by $T_\sharp \mu \in \mathcal{M}(Y)$ the push-forward measure defined by
$$T_\sharp \mu(B) := \mu(T^{-1}(B)) \text{ for every } B \in \mathcal{B}(Y). \qquad (2.5)$$

For $\mu \in \mathcal{M}(X)$ and a Lusin μ-measurable $T : X \to Y$, we have $T_\sharp \mu \in \mathcal{M}(Y)$. The linear space $B(X)$ (resp. $B_b(X)$) denotes the space of real Borel (resp. bounded Borel) functions. If $\mu \in \mathcal{M}(X)$, $p \in [1, \infty]$, we will denote by $L^p(X, \mu)$ 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 $\gamma(X) + \mu(X) > 0$ there exists Borel functions $\sigma, \varrho : X \to [0, \infty)$ and a Borel partition (A, A_γ, A_μ) of X with the following properties:*

$$A = \{x \in X : \sigma(x) > 0\} = \{x \in X : \varrho(x) > 0\}, \quad \sigma \cdot \varrho \equiv 1 \quad in \ A, \qquad (2.6)$$

$$\gamma = \sigma\mu + \gamma^\perp, \quad \sigma \in L^1_+(X, \mu), \quad \gamma^\perp \perp \mu, \quad \gamma^\perp(X \setminus A_\gamma) = \mu(A_\gamma) = 0,$$
(2.7)

$$\mu = \varrho\gamma + \mu^\perp, \quad \varrho \in L^1_+(X, \gamma), \quad \mu^\perp \perp \gamma, \quad \mu^\perp(X \setminus A_\mu) = \gamma(A_\mu) = 0.$$
(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 $\theta \in B(X; [0, 1])$ be the Lebesgue density of γ w.r.t. $\nu := \mu + \gamma$. Thus, θ is 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 mini-max properties of convex-concave functions in convex subsets of vector spaces and refer to [20, Prop. 1.2+3.2, Chap. VI] for a general exposition.

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

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

$$b \mapsto L(a, b) \quad \text{is concave in } B \text{ for every } a \in A.$$
(2.9b)

Notice that for arbitrary functions L one always has

$$\inf_{a\in A} \sup_{b\in B} L(a, b) \geq \sup_{b\in B} \inf_{a\in A} L(a, b);$$
(2.10)

so that equality holds in (2.10) 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.

The statement has the advantage of involving a minimal set of topological assumptions (we refer to [45, Thm. 3.1] for the proof; see also [9, Chapter 1, Prop. 1.1]).

Theorem 2.4 (Minimax duality) *Assume that* (2.9a) *and* (2.9b) *hold. If there exists* $b_\star \in B$ *and* $C > \sup_{b \in B} \inf_{a \in A} L(a, b)$ *such that*

$$\{a \in A : L(a, b_\star) \le C\} \quad \text{is compact in } A, \tag{2.11}$$

then

$$\inf_{a \in A} \sup_{b \in B} L(a, b) = \sup_{b \in B} \inf_{a \in A} L(a, b). \tag{2.12}$$

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

$$F \text{ is convex and lower semicontinuous and } D(F) \cap (0, \infty) \ne \emptyset, \tag{2.13}$$

where

$$D(F) := \{s \ge 0 : F(s) < \infty\}. \tag{2.14}$$

It is useful to recall that for every $x_0 \in D(F)$ the map $x \mapsto \frac{F(x)-F(x_0)}{x-x_0}$ is increasing in $D(F) \setminus \{x_0\}$, thanks to the convexity of F. The recession constant F'_∞, the right derivative F'_0 at 0, and the asymptotic affine coefficient $\mathrm{aff}\, F_\infty$ are defined by

$$F'_\infty := \lim_{s \to \infty} \frac{F(s)}{s} = \sup_{s > 0} \frac{F(s) - F(s_o)}{s - s_o}, \quad s_o \in D(F);$$

$$F'_0 := \begin{cases} -\infty & \text{if } F(0) = +\infty, \\ \lim_{s \downarrow 0} \frac{F(s)-F(0)}{s} & \text{otherwise}; \end{cases}$$

$$\mathrm{aff}\, F_\infty := \begin{cases} +\infty & \text{if } F'_\infty = +\infty, \\ \lim_{s \to \infty} (F'_\infty s - F(s)) & \text{otherwise}. \end{cases} \tag{2.15}$$

To avoid trivial cases, we assumed in (2.13) that the proper domain $D(F)$ contains at least one strictly positive real number. By convexity, $D(F)$ is a subinterval of $[0, \infty)$, and we will mainly focus on the case when $D(F)$ has nonempty interior and F has superlinear growth, i.e. $F'_\infty = +\infty$. Still it will be useful to deal with the general class defined by (2.13).

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

$$F^*(\phi) := \sup_{s \ge 0} (s\phi - F(s)), \tag{2.16}$$

with proper domain $D(F^*) := \{\phi \in \mathbb{R} : F^*(\phi) \in \mathbb{R}\}$; we will also denote by $\mathring{D}(F^*)$ the interior of $D(F^*)$. Strictly speaking, F^* is the conjugate of the convex function $\tilde{F} : \mathbb{R} \to (-\infty, +\infty]$, obtained by extending F to $+\infty$ for negative arguments, and it is related to the subdifferential $\partial F : \mathbb{R} \to 2^{\mathbb{R}}$ by

$$\phi \in \partial F(s) \quad \Leftrightarrow \quad s \in D(F), \quad \phi \in D(F^*), \quad F(s) + F^*(\phi) = s\phi. \quad (2.17)$$

Notice that

$$\inf D(F^*) = -\infty, \quad \sup D(F^*) = F'_\infty, \quad \mathring{D}(F^*) = (-\infty, F'_\infty), \quad (2.18)$$

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

$$\lim_{\phi \downarrow -\infty} F^*(\phi) = \inf F^* = -F(0), \quad \sup F^* = \lim_{\phi \uparrow +\infty} F^*(\phi) = +\infty. \quad (2.19)$$

Concerning the behavior of F^* at the boundary of its proper domain we can distinguish a few cases depending on the behavior of F at 0 and $+\infty$:

- If $F'_0 = -\infty$ (in particular if $F(0) = +\infty$) then F^* is strictly increasing in $D(F^*)$.
- 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 $\lim_{\phi \uparrow F'_\infty} F^*(\phi) = \text{aff} F_\infty$. Thus $F'_\infty \in D(F^*)$ only if $\text{aff} F_\infty < \infty$.
- The degenerate case when $F'_\infty = F'_0$ occurs only when F is linear.

If F is not linear, we always have

$$\left. \begin{array}{c} F^* \text{ is an increasing homeomorphism} \\ \text{between } (F'_0, F'_\infty) \text{ and } (-F(0), \text{aff} F_\infty) \end{array} \right\} \qquad (2.20)$$

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

We introduce the closed convex subset $\mathfrak{F}$ of $\mathbb{R}^2$ associated to the epigraph of F^*

$$\begin{aligned} \mathfrak{F} &:= \{(\phi, \psi) \in \mathbb{R}^2 : \psi \le -F^*(\phi)\} \\ &= \{(\phi, \psi) \in \mathbb{R}^2 : s\phi + \psi \le F(s) \, \forall s > 0\}; \end{aligned} \qquad (2.21)$$

since $D(F^*)$ has nonempty interior, $\mathfrak{F}$ has nonempty interior $\mathring{\mathfrak{F}}$ as well, with

$$\mathring{\mathfrak{F}} = \{(\phi, \psi) \in \mathbb{R}^2 : \phi \in \mathring{D}(F^*), \, \psi < -F^*(\phi)\}, \qquad (2.22)$$

and that $\mathfrak{F} = \overline{\overset{\circ}{\mathfrak{F}}}$. The function F can be recovered from F^* and from $\mathfrak{F}$ through the dual Fenchel–Moreau formula

$$F(s) = \sup_{\phi \in \mathbb{R}} \left(s\phi - F^*(\phi) \right) = \sup_{(\phi,\psi) \in \mathfrak{F}} s\phi + \psi = \sup_{(\phi,\psi) \in \overset{\circ}{\mathfrak{F}}} s\phi + \psi. \quad (2.23)$$

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

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

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

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

We will also use the duality formula

$$\left(\lambda F(\cdot) \right)^* = \lambda F^*(\cdot/\lambda) \quad \lambda > 0$$

and we adopt the notation ϕ_- and ϕ_+ to denote the negative and the positive part of a function ϕ, where $\phi_-(x) := \min\{\phi(x), 0\}$ and $\phi_+(x) := \max\{\phi(x), 0\}$.

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

$$U_p \in C^\infty(0, \infty), \quad U_p(1) = U'_p(1) = 0,$$
$$U''_p(s) = s^{p-2}, \quad U_p(0) = \lim_{s \downarrow 0} U_p(s). \quad (2.26)$$

Equivalently, we have the explicit formulas

$$U_p(s) = \begin{cases} \frac{1}{p(p-1)} \left(s^p - p(s-1) - 1 \right) & \text{if } p \neq 0, 1, \\ s \log s - s + 1 & \text{if } p = 1, \\ s - 1 - \log s & \text{if } p = 0, \end{cases} \quad \text{for } s > 0, \quad (2.27)$$

with $U_p(0) = 1/p$ if $p > 0$ and $U_p(0) = +\infty$ if $p \leq 0$.

Using the dual exponent $q = p/(p-1)$, the corresponding Legendre conjugates read

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

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

$$R(r) := \begin{cases} rF(1/r) & \text{if } r > 0, \\ F'_\infty & \text{if } r = 0. \end{cases} \tag{2.28}$$

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

$$R(0) = F'_\infty, \quad R'_\infty = F(0), \quad \text{aff } F_\infty = -R'_0, \quad \text{aff } R_\infty = -F'_0, \tag{2.29}$$

so that $R \in \Gamma(\mathbb{R}_+)$ and the map $F \mapsto R$ is an involution on $\Gamma(\mathbb{R}_+)$. 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 (2.21): it is easy to check that

$$(\phi, \psi) \in \mathfrak{F} \quad \Leftrightarrow \quad (\psi, \phi) \in \mathfrak{R}, \tag{2.30}$$

a relation that obviously holds for the interiors of $\mathfrak{F}$ and $\mathfrak{R}$ as well. It follows that the Legendre transform of R and F are related by

$$\psi \le -F^*(\phi) \quad \Leftrightarrow \quad \phi \le -R^*(\psi) \quad \Leftrightarrow \quad (\phi, \psi) \in \mathfrak{F} \quad \text{for every } \phi, \psi \in \mathbb{R}, \tag{2.31}$$

and, recalling (2.22),

$$\phi \in \mathring{D}(F^*), \ \psi < -F^*(\phi) \quad \Leftrightarrow \quad \psi \in \mathring{D}(R^*), \ \phi < -R^*(\psi). \tag{2.32}$$

Both the above conditions characterize the interior of $\mathfrak{F}$. As in (2.20) we have

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

with $\mathring{D}(R^*) = (-\infty, F(0))$. A last useful identity involves the subdifferentials of F and R: for every $s, r > 0$ with $sr = 1$, and $\phi, \psi \in \mathbb{R}$ we have

$$\Big(\phi \in \partial F(r) \text{ and } \psi = -F^*(\phi)\Big) \quad \Longleftrightarrow \quad \Big(\psi \in \partial R(s) \text{ and } \phi = -R^*(\psi)\Big).$$
(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

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

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

where $\gamma = \sigma\mu + \gamma^\perp$ is the Lebesgue decomposition of γ w.r.t. μ, see Lemma 2.3. Notice that

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

and, whenever η_0 is the null measure, we have

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

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

Because of our applications in Sect. 3, our next lemma deals with Borel functions $\varphi \in B(X; \bar{\mathbb{R}})$ taking values in the extended real line $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. By $\bar{\mathfrak{F}}$ we denote the closure of $\mathfrak{F}$ in $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$, i.e.

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

and, symmetrically by (2.29) and (2.30),

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

In particular, we have

$$(\phi, \psi) \in \bar{\mathfrak{F}} \quad \Longrightarrow \quad \Big(\phi \leq F'_\infty \text{ and } \psi \leq F(0)\Big).$$
(2.40)

Lemma 2.6 *If $\gamma, \mu \in \mathcal{M}(X)$ and $(\phi, \psi) \in B(X; \bar{\mathfrak{F}})$ satisfy*

$$\mathscr{F}(\gamma|\mu) < \infty, \quad \psi_- \in L^1(X, \mu) \quad (resp. \; \phi_- \in L^1(X, \gamma)),$$

then $\phi_+ \in L^1(X, \gamma)$ (resp. $\psi_+ \in L^1(X, \mu)$) and

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

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

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

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

Equation (2.42) *can equivalently be formulated as $\psi \in \partial R(\varrho)$ and $\phi = -R^*(\psi)$.*

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

If $\phi_- \in L^1(X, \gamma)$ (and thus $\phi > -\infty$ γ-a.e.) the analogous inequality $\psi_+ \leq F(s) + s\phi_-$ yields $\psi_+ \in L^1(X, \mu)$. Then, (2.41) follows from (2.21) and (2.40).

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

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

and by (2.21) and (2.40) the equality case immediately yields that each of the three integrals of the previous formula vanishes. Since (ϕ, ψ) lies in $\bar{\mathfrak{F}} \subset \mathbb{R}^2$ $(\mu + \gamma)$-a.e. in A, the vanishing of the first integrand yields $\psi = -F^*(\sigma)$ and $\phi \in \partial F(\sigma)$ by (2.17) for μ and $(\mu + \gamma)$ almost every point in A. The equivalence (2.34) provides the reversed identities $\psi \in \partial R(\varrho), \phi = -R^*(\psi)$.

The relations in (2.43) follow 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. $\qquad \square$

A simple application of (2.41) yields the following variant of Jensen's inequality

$$\mathscr{F}(\gamma|\mu) \geq \mu(X)F\big(\gamma(X)/\mu(X)\big) \quad \text{whenever } \mu(X) > 0. \tag{2.44}$$

In order to prove it, we first choose arbitrary $(\bar{\phi}, \bar{\psi}) \in \mathfrak{F}$ and constant functions $\phi(x) \equiv \bar{\phi}, \psi(x) \equiv \bar{\psi}$ in (2.41), obtaining

$$\mathscr{F}(\gamma|\mu) \geq \mu(X)\Big[\bar{\psi} + \frac{\gamma(X)}{\mu(X)}\bar{\phi}\Big];$$

we then take the supremum with respect to $(\bar{\phi}, \bar{\psi}) \in \mathfrak{F}$, recalling (2.23).

The next theorem gives a characterization of the relative entropy $\mathscr{F}$, which 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. Whenever $A \subset \mathbb{R}$, we denote by $\mathrm{LSC}_s(X; A)$ the class of lower semicontinuous simple real functions

$$\mathrm{LSC}_s(X) := \Big\{\varphi : X \to \mathbb{R} \ : \ \varphi \in \mathrm{LSC}(X), \ \varphi(X) \text{ is a finite subset of } A\Big\},$$

by omitting A when $A = \mathbb{R}$; we introduce the notation $\varphi = -\phi$ and the concave increasing function

$$F^\circ(\varphi) := -F^*(-\varphi), \quad F^\circ(\varphi) = \inf_{s \geq 0}\big(\varphi s + F(s)\big); \qquad (2.45)$$

by (2.18) and (2.19) the interior of the proper domain of F° is $\mathring{D}(F^\circ) = (-F'_\infty, +\infty)$ and $\lim_{\varphi \downarrow -F'_\infty} F^\circ(\varphi) = -\infty$, $\lim_{\varphi \uparrow +\infty} F^\circ(\varphi) = F(0)$.

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

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

$$(\phi(x), \psi(x)) \in \mathring{\mathfrak{F}} \ \forall \, x \in X\Big\} \quad (2.46)$$

$$= \sup\Big\{ \int_X \psi \, d\mu - \int_X R^*(\psi) \, d\gamma : \psi \in \mathrm{LSC}_s(X, \mathring{D}(R^*))\Big\} \qquad (2.47)$$

$$= \sup\Big\{ \int_X F^\circ(\varphi) \, d\mu - \int_X \varphi \, d\gamma : \varphi \in \mathrm{LSC}_s(X, \mathring{D}(F^\circ))\Big\}. \quad (2.48)$$

Moreover, the space $\mathrm{LSC}_s(X)$ in the supremum of (2.46), can also be replaced by the space $\mathrm{LSC}_b(X)$ of bounded l.s.c. functions or by the space $B_b(X)$ of bounded Borel functions and the constraint $(\phi(x), \psi(x)) \in \mathring{\mathfrak{F}}$ in (2.46) can also be relaxed to $(\phi(x), \psi(x)) \in \mathfrak{F}$ for every $x \in X$. Similarly, the

spaces $\mathrm{LSC}_s(X, \mathring{\mathrm{D}}(R^*))$ *(resp.* $\mathrm{LSC}_s(X, \mathring{\mathrm{D}}(F^\circ))$) *of* (2.47) *(resp.* (2.48)*) can be replaced by* $\mathrm{LSC}_b(X, \mathrm{D}(R^*))$ *or* $\mathrm{B}_b(X, \mathrm{D}(R^*))$ *(resp.* $\mathrm{LSC}_b(X, \mathrm{D}(F^\circ))$ *or* $\mathrm{B}_b(X, \mathrm{D}(F^\circ))$*).*

Remark 2.8 If (X, τ) is completely regular (recall (2.3)), then we can equivalently replace lower semicontinuous functions by continuous ones in (2.46), (2.47) and (2.48). E.g. in the case of (2.46) we have

$$\mathscr{F}(\gamma|\mu) = \sup \left\{ \int_X \psi \, \mathrm{d}\mu + \int_X \phi \, \mathrm{d}\gamma : (\phi, \psi) \in \mathrm{C}_b(X; \mathring{\mathfrak{F}}) \right\}, \qquad (2.49)$$

whereas (2.47) and (2.48) become

$$\mathscr{F}(\gamma|\mu) = \sup \left\{ \int_X \psi \, \mathrm{d}\mu - \int_X R^*(\psi) \, \mathrm{d}\gamma : \psi, R^*(\psi) \in \mathrm{C}_b(X) \right\}$$
$$= \sup \left\{ \int_X F^\circ(\varphi) \, \mathrm{d}\mu - \int_X \varphi \, \mathrm{d}\gamma : \varphi, F^\circ(\varphi) \in \mathrm{C}_b(X) \right\}.$$

In fact, considering first (2.46), by complete regularity it is possible to express every pair ϕ, ψ of bounded lower semicontinuous functions with values in $\mathring{\mathfrak{F}}$ 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 given by $\mathring{\mathfrak{F}}$ due to (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 (2.47) 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 (2.49).

Proof Since the statements are trivial in the case when $\mu = \gamma = \eta_0$ are the null measure, it is clearly not restrictive to assume $(\mu + \gamma)(X) > 0$. Let us prove (2.46): denoting by $\mathscr{F}'$ its right-hand side, Lemma 2.6 yields $\mathscr{F} \geq \mathscr{F}'$. In order to prove the opposite inequality we consider the Lebesgue decomposition given by Lemma 2.3: let $A_\gamma \in \mathcal{B}(X)$ be a μ-negligible Borel set where $\gamma^\perp$ is concentrated, let $\tilde{A} := X \setminus A_\gamma = A \cup A_\mu$ and let $\sigma : X \to [0, \infty)$ be a Borel density for γ w.r.t. μ. We consider a countable subset $(\phi_n, \psi_n)_{n=1}^\infty$ with $\psi_1 = \phi_1 = 0$, which is dense in $\mathring{\mathfrak{F}}$ 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) we have

$$F(\sigma(x)) = \lim_{N \uparrow \infty} F_N(x), \text{ where } \forall x \in X : F_N(x) := \sup_{1 \leq n \leq N} \bar{\psi}_n + \sigma(x)\bar{\phi}_n.$$

Hence, Beppo Levi's monotone convergence theorem (notice that $F_N \geq F_1 = 0$) implies $\mathscr{F}(\gamma|\mu) = \lim_{N \uparrow \infty} \mathscr{F}'_N(\gamma|\mu)$, where

$$\mathscr{F}'_N(\gamma|\mu) := \int_{\tilde{A}} F_N(x) \, d\mu(x) + \bar{\phi}_N \gamma(A_\gamma).$$

It is therefore sufficient to prove that

$$\mathscr{F}'(\gamma|\mu) \geq \mathscr{F}'_N(\gamma|\mu) \quad \text{for every } N \in \mathbb{N}. \tag{2.50}$$

We fix $N \in \mathbb{N}$, set $\phi_0 := \bar{\phi}_N$, $\psi_0 := \bar{\psi}_N$, and recursively define the Borel sets A_j, for $j = 0, \ldots, N$, with $A_0 := A_\gamma$ and

$$
\begin{aligned}
A_1 &:= \{x \in \tilde{A} : F_1(x) = F_N(x)\}, \\
A_j &:= \{x \in \tilde{A} : F_N(x) = F_j(x) > F_{j-1}(x)\} \quad \text{for } j = 2, \ldots, N.
\end{aligned}
\tag{2.51}
$$

Since $F_1 \leq F_2 \leq \cdots \leq F_N$, the sets A_i, $i = 1, \ldots, N$, form a Borel partition of $\tilde{A}$. As 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

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

where

$$
\begin{aligned}
S_N &:= \max_{0 \leq n \leq N} \left[(\phi_n - \phi_{\min}^N) + (\psi_n - \psi_{\min}^N) \right], \\
\phi_{\min}^N &:= \min_{0 \leq j \leq N} \phi_j, \quad \psi_{\min}^N := \min_{0 \leq j \leq N} \psi_j.
\end{aligned}
$$

Since $(\phi_n, \psi_n) \in \mathfrak{F}$ for every $n \in \mathbb{N}$ and $\mathring{\mathfrak{F}}$ satisfies the monotonicity property (2.24) $(\phi_{\min}^N, \psi_{\min}^N) \in \mathring{\mathfrak{F}}$; since the sets G_n are disjoint, the functions

$$\psi_N(x) := \begin{cases} \psi_n & \text{if } x \in G_n, \\ \psi_{\min}^N & \text{if } x \in X \setminus \left(\cup_{n=1}^N G_n \right) \end{cases}$$

$$\phi_N(x) := \begin{cases} \phi_n & \text{if } x \in G_n, \\ \phi_{\min}^N & \text{if } x \in X \setminus \left(\cup_{n=1}^N G_n \right) \end{cases}$$

take values in $\mathring{\mathfrak{F}}$ and are lower semicontinuous thanks to the representation formula

$$\psi_N(x) = \psi_{\min}^N + \sum_{n=0}^{N} \left(\psi_n - \psi_{\min}^N\right) \chi_{G_n}(x),$$

$$\phi_N(x) = \phi_{\min}^N + \sum_{n=0}^{N} \left(\phi_n - \phi_{\min}^N\right) \chi_{G_n}(x). \qquad (2.52)$$

Moreover, they satisfy

$$
\begin{aligned}
\mathscr{F}_N'(\gamma|\mu) &= \sum_{j=1}^{N} \int_{A_j} F_j \, \mathrm{d}\mu + \phi_0 \gamma(A_0) \\
&= \phi_{\min}^N \gamma(X) + \psi_{\min}^N \mu(X) \\
&\quad + \sum_{j=0}^{N} \left(\int_{A_j} (\phi_j - \phi_{\min}^N) \, \mathrm{d}\gamma + \int_{A_j} (\psi_j - \psi_{\min}^N) \, \mathrm{d}\mu \right) \\
&\leq \phi_{\min}^N \gamma(X) + \psi_{\min}^N \mu(X) \\
&\quad + \sum_{j=0}^{N} \left(\int_{K_j} (\phi_j - \phi_{\min}^N) \, \mathrm{d}\gamma + \int_{K_j} (\psi_j - \psi_{\min}^N) \, \mathrm{d}\mu \right) + \varepsilon \\
&\leq \int_X \phi_N \, \mathrm{d}\gamma + \int_X \psi_N \, \mathrm{d}\mu + \varepsilon.
\end{aligned}
$$

Since ε is arbitrary we obtain (2.50).

Equation (2.47) follows directly by (2.46) and the previous Lemma 2.6. In fact, denoting by $\mathscr{F}''$ the right-hand side of (2.47), Lemma 2.6 shows that $\mathscr{F}''(\gamma|\mu) \leq \mathscr{F}(\gamma|\mu) = \mathscr{F}'(\gamma|\mu)$. On the other hand, if $\phi, \psi \in \mathrm{LSC}_s(X)$ with $(\phi, \psi) \in \mathring{\mathfrak{F}}$ then ψ takes values in $\mathring{\mathrm{D}}(R^*)$ and $-R^*(\psi) \geq \phi$. Hence, the map $x \mapsto R^*(\psi(x))$ belongs to $\mathrm{LSC}_s(X)$ since R^* is real valued and nondecreasing in the interior of its domain, and it is bounded from above by $-\phi$. We thus get $\mathscr{F}''(\gamma|\mu) \geq \mathscr{F}'(\gamma|\mu)$.

In order to show (2.48), we observe that for every $\psi \in \mathrm{LSC}_s(X, \mathring{\mathrm{D}}(R^*))$ and $\varepsilon > 0$ we can set $\varphi := R^*(\psi) + \varepsilon \in \mathrm{LSC}_s(X, \mathring{\mathrm{D}}(F^\circ))$; since $(\psi, -R^*(\psi) - \varepsilon) \in \mathring{\mathfrak{R}}$, (2.31) yields $\psi < -F^*(-\varphi) = F^\circ(\varphi)$ so that $\int F^\circ(\varphi) \, \mathrm{d}\mu - \int \varphi \, \mathrm{d}\gamma \geq \int \psi \, \mathrm{d}\mu - \int R^*(\psi) \, \mathrm{d}\gamma - \varepsilon \gamma(X)$. By construction and (2.30) we also have

$(-\varphi, F^\circ(\varphi)) \in \mathring{\mathfrak{F}}$ so that $\int F^\circ(\varphi) \, d\mu - \int \varphi \, d\gamma \leq \mathscr{F}(\gamma|\mu)$ by Lemma 2.6. Passing to the limit as $\varepsilon \downarrow 0$ and recalling (2.47) we obtain (2.48).

When one replaces $\mathrm{LSC}_s(X)$ with $\mathrm{LSC}_b(X)$ or $\mathrm{B}_b(X)$ in (2.46) and the constraint $(\phi(x), \psi(x)) \in \mathring{\mathfrak{F}}$ with $(\phi(x), \psi(x)) \in \mathfrak{F}$ (or even $\bar{\mathfrak{F}}$), the supremum is taken on a larger set, so that the right-hand side of (2.46) 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 (2.47) or (2.48).   □

The following result provides lower semicontinuity of the relative entropy or of an increasing sequence of relative entropies.

Corollary 2.9 *The functional $\mathscr{F}$ is jointly convex and lower semicontinuous in $\mathcal{M}(X) \times \mathcal{M}(X)$. More generally, if $F \in \Gamma(\mathbb{R}_+)$ is the pointwise limit of an increasing net $(F_\lambda)_{\lambda \in \mathbb{L}} \subset \Gamma(\mathbb{R}_+)$ indexed by a directed set $\mathbb{L}$ and $(\mu, \gamma) \in \mathcal{M}(X) \times \mathcal{M}(X)$ is the narrow limit of a net $(\mu_\lambda, \gamma_\lambda)_{\lambda \in \mathbb{L}} \subset \mathcal{M}(X) \times \mathcal{M}(X)$, then the corresponding entropy functionals $\mathscr{F}_\lambda, \mathscr{F}$ satisfy*

$$\liminf_{\lambda \in \mathbb{L}} \mathscr{F}_\lambda(\gamma_\lambda|\mu_\lambda) \geq \mathscr{F}(\gamma|\mu). \tag{2.53}$$

Proof The lower semicontinuity of $\mathscr{F}$ follows by (2.46), which provides a representation of $\mathscr{F}$ as the supremum of a family of lower semicontinuous functionals for the narrow topology. Using $F_\alpha \leq F_\lambda$ for $\alpha \leq \lambda$ in $\mathbb{L}$, α fixed, we have

$$\liminf_{\lambda \in \mathbb{L}} \mathscr{F}_\lambda(\gamma_\lambda|\mu_\lambda) \geq \liminf_{\lambda \in \mathbb{L}} \mathscr{F}_\alpha(\gamma_\lambda|\mu_\lambda) \geq \mathscr{F}_\alpha(\gamma|\mu),$$

by the above lower semicontinuity. Hence, it suffices to check that

$$\lim_{\lambda \in \mathbb{L}} \mathscr{F}_\lambda(\gamma|\mu) = \mathscr{F}(\gamma|\mu) \quad \text{for every } \gamma, \mu \in \mathcal{M}(X). \tag{2.54}$$

This formula follows by the monotonicity of the convex sets $\mathfrak{F}_\lambda$ (associated to F_λ by (2.21)), i.e. $\mathfrak{F}_\alpha \subset \mathfrak{F}_\lambda$ if $\alpha \leq \lambda$ in $\mathbb{L}$, and by the fact that $\mathring{\mathfrak{F}} \subset \cup_{\lambda \in \mathbb{L}} \mathring{\mathfrak{F}}_\lambda$; in order to show the latter property, we argue by contradiction and we suppose that there exists $(\phi, \psi) \in \mathring{\mathfrak{F}}$ which does not belong to $\mathfrak{F}' := \cup_{\lambda \in \mathbb{L}} \mathring{\mathfrak{F}}_\lambda$. Notice that every $\mathfrak{F}_\lambda$ has nonempty interior, so that $\mathfrak{F}'$ is a nonempty convex and open set. We also notice that $\phi < F'_\infty$ and $\lim_{\lambda \in \mathbb{L}} (F_\lambda)'_\infty = F'_\infty$ so that there exists $\alpha \in \mathbb{L}$ with $F_\alpha^*(\phi) < \infty$; thus there exists $-\bar{\psi} > \psi$ such that $(\phi, \bar{\psi}) \in \mathfrak{F}'$. Applying the geometric form of the Hahn-Banach theorem, we can find a non vertical line separating (ϕ, ψ) from $\mathfrak{F}'$, i.e. there exists $\theta \in \mathbb{R}$ such that

$$-\psi' \geq -\psi + \theta(\phi' - \phi) \quad \text{for every } (\phi', \psi') \in \mathfrak{F}'.$$

Recalling that $\overline{\mathfrak{F}_\lambda^\circ} = \mathfrak{F}_\lambda$ we deduce

$$F_\lambda^*(\phi') \geq -\psi + \theta(\phi' - \phi) \quad \text{for every } \phi' \in D(F_\lambda^*), \ \lambda \in \mathbb{L};$$

taking the supremum w.r.t. ϕ' we obtain

$$\psi + \theta\phi \geq F_\lambda(\theta) \quad \text{for every } \lambda \in \mathbb{L}$$

and passing to the limit w.r.t. $\lambda \in \mathbb{L}$ we get

$$\psi + \theta\phi \geq F(\theta) \quad \text{so that } -\psi \leq \theta\phi - F(\phi) \leq F^*(\phi),$$

which contradicts the fact that $(\phi, \psi) \in \text{int } \mathfrak{F}$.

Thus for every pair of simple and lower semicontinuous functions (ϕ, ψ) taking values in $\text{int } \mathfrak{F}$ we have $(\psi(x), \phi(x)) \in \text{int } \mathfrak{F}_\alpha$ for every $x \in X$ and some $\alpha \in \mathbb{L}$ so that

$$\liminf_{\lambda \in \mathbb{L}} \mathscr{F}_\lambda(\gamma|\mu) \geq \mathscr{F}_\alpha(\gamma|\mu) \geq \int_X \psi \, d\mu + \int_X \phi \, d\gamma.$$

Since ϕ, ψ are arbitrary we conclude applying the duality formula (2.46). □

Next, we provide a compactness result for the sublevels of the relative entropy, which will be useful in Sect. 3.4 (see Theorem 3.3 and Lemma 3.9).

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

$$\Xi_C := \left\{ \gamma \in \mathcal{M}(X) : \mathscr{F}(\gamma|\mu) \leq C \text{ for some } \mu \in \mathcal{K} \right\}, \tag{2.55}$$

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

$$\lambda\gamma(B) \leq \mathscr{F}(\gamma|\mu) + F^*(\lambda)\mu(B) \quad \text{for every } \lambda \in (0, F_\infty'), \text{ and } B \in \mathcal{B}(X). \tag{2.56}$$

This follows easily by considering the decomposition $\gamma = \sigma\mu + \gamma^\perp$ and by integrating the Young inequality $\lambda\sigma \leq F(\sigma) + F^*(\lambda)$ for $\lambda > 0$ in B with respect to μ; notice that

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

Choosing first $B = X$ in (2.56) and an arbitrary λ in $(0, F'_\infty)$ (notice that $F^*(\lambda) < \infty$ thanks to (2.18)) 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 $\mu \in \mathcal{K}$. (2.56) shows that $\gamma(X \setminus K) \leq \varepsilon$ for every $\gamma \in \Xi$.   □

We conclude this section with a useful representation of $\mathscr{F}$ in terms of the reverse entropy R (2.28) and the corresponding functional $\mathscr{R}$. We will use the result in Sect. 3.5 for the reverse formulation of the primal entropy-transport problem.

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

$$\mathscr{R}(\mu|\gamma) = \int_X R(\varrho) \, d\gamma + R'_\infty \mu^\perp(X), \tag{2.57}$$

where $\mu = \varrho\gamma + \mu^\perp$ *is the reverse Lebesgue decomposition given by Lemma 2.3. Then*

$$\mathscr{F}(\gamma|\mu) = \mathscr{R}(\mu|\gamma). \tag{2.58}$$

Proof It is an immediate consequence of the dual characterization in (2.46) and the equivalence in (2.30).   □

3 Optimal Entropy-Transport problems

The major object of Part I is the entropy-transport functional, where two measures $\mu_1 \in \mathcal{M}(X_1)$ and $\mu_2 \in \mathcal{M}(X_2)$ are given, and one has to find a transport plan $\gamma \in \mathcal{M}(X_1 \times X_2)$ that minimizes the functional.

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), $i = 1, 2$, which define the Cartesian product $X := X_1 \times X_2$ and the canonical projections $\pi^i : X \to X_i$;
- two entropy functions $F_i \in \Gamma(\mathbb{R}_+)$, thus satisfying (2.13);
- a proper, lower semicontinuous cost function $c : X \to [0, \infty]$;
- a pair of nonnegative Radon measures $\mu_i \in \mathcal{M}(X_i)$ with finite mass $m_i := \mu_i(X_i)$, satisfying the compatibility condition

$$J := \left(m_1 \, D(F_1)\right) \cap \left(m_2 \, D(F_2)\right) \neq \emptyset. \tag{3.1}$$

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

$$F_1 \text{ and } F_2 \text{ are superlinear, i.e. } (F_i)'_\infty = +\infty; \tag{3.2a}$$

$$(F_1)'_\infty + (F_2)'_\infty + \inf \mathsf{c} > 0 \text{ and } \mathsf{c} \text{ has compact sublevels.} \tag{3.2b}$$

For every transport plan $\boldsymbol{\gamma} \in \mathcal{M}(X)$ we define the marginals $\gamma_i := \pi_\sharp^i \boldsymbol{\gamma}$ and, as in (2.35), we define the relative entropies

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

With this, we introduce the *Entropy-Transport functional* as

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

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 Sect. 3.1, we want to investigate the following problem.

Problem 3.1 (Entropy-Transport minimization) *Given* $\mu_i \in \mathcal{M}(X_i)$ *find* $\boldsymbol{\gamma} \in \mathcal{M}(X) = \mathcal{M}(X_1 \times X_2)$ *minimizing* $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$, *i.e.*

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

We denote by $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2) \subset \mathcal{M}(X)$ *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 $\boldsymbol{\eta}_0$, 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 nondegenerate case $m_1 m_2 \neq 0$ is that there exist functions B_1 and B_2 with

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

In fact, the plans

$$\boldsymbol{\gamma} = \frac{\theta}{m_1 m_2} \mu_1 \otimes \mu_2 \quad \text{with } \theta \in J \quad \text{given by (3.1)} \tag{3.9}$$

are Radon [44, Thm. 17, p. 63], have finite cost $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_1) < \infty$ and provide the estimate

$$\begin{aligned}
\mathsf{ET}(\mu_1, \mu_2) \leq \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_1) \leq{} & m_1 F_1(\theta/m_1) + m_2 F_2(\theta/m_2) \\
& + \theta \sum_i m_i^{-1} \|B_i\|_{\mathrm{L}^1(X_i, \mu_i)}, \quad \text{for every } \theta \in J.
\end{aligned}$$

Notice that (3.1) is also necessary for feasibility: in fact (2.44) yields

$$\mathscr{F}_i(\boldsymbol{\gamma}|\mu_i) \geq m_i F_i(m/m_i), \quad \text{where} \quad m := \gamma_i(X_i) = \boldsymbol{\gamma}(X).$$

Thus, whenever $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) < \infty$, we have

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geq m \inf \mathsf{c} + m_1 F_1(m/m_1) + m_2 F_2(m/m_2), \tag{3.10}$$

and therefore

$$m = \boldsymbol{\gamma}(X) \in \big(m_1 \, \mathrm{D}(F_1)\big) \cap \big(m_2 \, \mathrm{D}(F_2)\big) = J. \tag{3.11}$$

We will often strengthen (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:

$$\Big(\mathrm{int} \, \big(m_1 \mathrm{D}(F_1)\big) \cap m_2 \, \mathrm{D}(F_2) \Big) \cup \Big(m_1 \, \mathrm{D}(F_1) \cap \mathrm{int} \, \big(m_2 \mathrm{D}(F_2)\big) \Big) \neq \emptyset. \tag{3.12}$$

This condition is surely satisfied if J has nonempty interior, i.e. $\max(m_1 s_1^-, m_2 s_2^-) < \min(m_1 s_1^+, m_2 s_2^+)$, where $s_i^- = \inf \mathrm{D}(F_i)$, $s_i^+ := \sup \mathrm{D}(F_i)$.

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 $\mathrm{D}(F_i)$ (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 **Costless transport**: Consider the case $c \equiv 0$. 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

$$\mathsf{ET}(\mu_1, \mu_2) = H_0(m_1, m_2) := \min \left\{ m_1 F_1(\theta/m_1) + m_2 F_2(\theta/m_2) : \theta \geq 0 \right\}. \tag{3.13}$$

E.2 **Entropy-potential problems**: If $\mu_2 \equiv \eta_0$ is the null measure and, just to fix ideas, X_i are Polish spaces with X_2 compact and c is real valued, then setting $V(x_1) := \min_{x_2 \in X_2} c(x_1, x_2)$ we get

$$\mathsf{ET}(\mu, \eta_0) = \inf_{\gamma \in \mathcal{M}(X_1)} \mathscr{F}_1(\gamma | \mu) + \int_{X_1} V \, d\gamma + (F_2)'_\infty \gamma(X_1). \tag{3.14}$$

In fact for every $\gamma \in \mathcal{M}(X)$ we have $\mathscr{F}_2(\gamma_2 | \eta_0) = (F_2)'_\infty \gamma_2(X_2) = (F_2)'_\infty \gamma_1(X_1)$; moreover by applying the von Neumann measurable selection Theorem [44, Thm. 13, p. 127] it is not difficult to check that

$$\min \left\{ \int_X c \, d\gamma : \gamma \in \mathcal{M}(X), \pi^1_\sharp \gamma = \gamma \right\} = \int_{X_1} V(x_1) \, d\gamma(x_1).$$

E.3 **Pure transport problems:** We choose $F_i(r) = I_1(r) = \begin{cases} 0 & \text{if } r = 1 \\ +\infty & \text{otherwise.} \end{cases}$

In this case any feasible plan γ should have μ_1 and μ_2 as marginals and the functional just reduces to the pure transport part

$$\mathsf{T}(\mu_1, \mu_2) = \min \left\{ \int_{X_1 \times X_2} c \, d\gamma : \pi^i_\sharp \gamma = \mu_i \right\}. \tag{3.15}$$

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 (3.12) does not hold. In this case, the set J defined by (3.1) contains only one point θ which separates $m_1 D(F_1)$ and $m_2 D(F_2)$:

$$\theta = m_1 s_1^+ = m_2 s_2^- \quad \text{or} \quad \theta = m_1 s_1^- = m_2 s_2^+. \tag{3.16}$$

It is not difficult to check that in this case

$$\mathsf{ET}(\mu_1, \mu_2) = m_1 F_1(\theta/m_1) + m_2 F_2(\theta/m_2) + \mathsf{T}(\mu_1, \mu_2). \tag{3.17}$$

E.4 **Optimal transport with density constraints:** We realize density constraints by introducing characteristic functions of intervals $[a_i, b_i]$, viz. $F_i(r) := I_{[a_i, b_i]}(r)$, $a_i \le 1 \le b_i$. E.g. when $a_i = 1$, $b_i = +\infty$ we have

$$\mathsf{ET}(\mu_1, \mu_2) = \min \left\{ \int_{X_1 \times X_2} \mathsf{c}\, d\boldsymbol{\gamma} : \pi_\sharp^i \boldsymbol{\gamma} \ge \mu_i \right\}. \tag{3.18}$$

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

$$\mathsf{ET}(\mu_1, \mu_2) = \min \left\{ \int_{X_1 \times X_2} \mathsf{c}\, d\boldsymbol{\gamma} : \pi_\sharp^1 \boldsymbol{\gamma} \le \mu_1,\ \pi_\sharp^2 \boldsymbol{\gamma} \ge \mu_2 \right\}, \tag{3.19}$$

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

E.5 **Pure entropy problems:** These problems arise if $X_1 = X_2 = X$ and transport is forbidden, i.e. $(F_i)'_\infty = +\infty$, $\mathsf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2 \\ +\infty & \text{otherwise.} \end{cases}$
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 $\gamma = \vartheta \mu$ and then $\mu_i = \vartheta_i \mu$. Since $\gamma \ll \mu_i$ 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

$$\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \int_X \left(\vartheta_1 F_1(\vartheta/\vartheta_1) + \vartheta_2 F_2(\vartheta/\vartheta_2) \right) d\mu, \tag{3.20}$$

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 (3.13) we get

$$\mathsf{ET}(\mu_1, \mu_2) = \int_X H_0 \left(\frac{d\mu_1}{d\mu}, \frac{d\mu_2}{d\mu} \right) d\mu, \quad \text{whenever } \mu_i \ll \mu. \tag{3.21}$$

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

$$H_0(\theta_1, \theta_2) = \theta_1 + \theta_2 - 2\sqrt{\theta_1 \theta_2} = \left(\sqrt{\theta_1} - \sqrt{\theta_2} \right)^2. \tag{3.22}$$

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

$$H_0(\theta_1; \theta_2) = \theta_1 \log \left(\frac{2\theta_1}{\theta_1 + \theta_2} \right) + \theta_2 \log \left(\frac{2\theta_2}{\theta_1 + \theta_2} \right).$$

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 obtain

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

E.6 **Regular Entropy-Transport problems:** These problems correspond to the choice of a pair of differentiable entropies F_i with $D(F_i) \supset (0, \infty)$, as in the case of the power-like entropies U_p defined in (2.26). 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:** For a metric space (X, d), set $X_1 = X_2 = X$ and let τ be induced by d. Further, set $F_1(s) = F_2(s) := U_1(s) = s \log s - s + 1$ and

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

This case will be thoroughly studied in the second part of the present paper, see Sect. 6.

E.8 **Marginal Entropy-Transport problems:** In this case one of the two marginals of $\boldsymbol{\gamma}$ is fixed, say γ_1, by choosing $F_1(r) := I_1(r)$. 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, namely

$$\mathsf{ET}(\mu_1, \mu_2) = \min_{\gamma_2 \in \mathcal{M}(X_2)} \left\{ \mathsf{T}(\mu_1, \gamma_2) + \mathscr{F}_2(\gamma_2|\mu_2) \right\},$$

where T has been defined by (3.15). 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 [24].

E.9 **The Piccoli-Rossi "generalized Wasserstein distance"** [38,39]: For a metric space (X, d), set $X_1 = X_2 = X$, let τ be induced by d, and consider $F_1(s) = F_2(s) := V(s) = |s - 1|$ with $\mathsf{c}(x_1, x_2) := \mathsf{d}(x_1, x_2)$. This example can be considered as the natural extension of the L^1-Kantorovich–Wasserstein distance (corresponding to (3.15) with the distance cost) to measures with different masses, due to its dual representation in terms of the flat metric, see (7.47).

E.10 **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 $\mathsf{c}_{i,j} := \mathsf{c}(x_i, y_j)$. In the case of superlinear entropy functions F_i, the Entropy-Transport problem for this discrete model consists in finding coefficients $\gamma_{i,j} \geq 0$ which minimize

$$\mathcal{E}(\gamma_{i,j} | \alpha_i, \beta_j) := \sum_i \alpha_i F_1\left(\frac{\sum_j \gamma_{i,j}}{\alpha_i}\right) + \sum_j \beta_j F_2\left(\frac{\sum_i \gamma_{i,j}}{\beta_j}\right) + \sum_{i,j} \mathsf{c}_{i,j} \gamma_{i,j}.$$

(3.23)

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 Sect. 3.1.

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

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

(ii) c has compact sublevels in 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* $\mathcal{M}(X)$.

Proof We can apply the Direct Method of Calculus of Variations: since the map $\gamma \mapsto \mathcal{E}(\gamma | \mu_1, \mu_2)$ 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 the Prokhorov Theorem 2.2. In both cases boundedness follows by the coercivity assumptions and the estimate (3.10). In particular, by formula (2.15) defining $(F_i)'_\infty$, we can find $\bar{s} \geq 0$ such that $\frac{m_i}{m} F_i(\frac{m}{m_i}) \geq \frac{1}{2}(F_i)'_\infty$ whenever $m \geq \bar{s} m_i$; if $a := \inf c + \sum_i (F_i)'_\infty > 0$ the estimate (3.10) yields

$$\gamma(X) \leq \frac{2}{a} \mathcal{E}(\gamma | \mu_1, \mu_2) \quad \text{for every } \gamma \in \mathcal{M}(X) \text{ with } \gamma(X)$$

$$\geq \bar{s}\,\max(\mu_1(X_1), \mu_2(X_2)).$$

In case *(ii)* equal tightness is a consequence of the Markov inequality and the nonnegativity of F_i: by 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

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

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 and we thus conclude by [2, Lemma 5.2.2].   □

Remark 3.4 The assumptions *(i)* and *(ii)* in the previous theorem are almost optimal, and it is not hard to find examples violating them such that the statement of Theorem 3.3 does not hold. 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 $\nu_i \in \mathcal{M}(\mathbb{R})$ and $\nu_i(\{0\}) = 0$. Setting $n_i := \nu_i(\mathbb{R})$ we find

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

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

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

Moreover, $2\log 2$ is the infimum, which is reached 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. The first result only addresses the marginals $\gamma_i = \pi_\sharp^i \boldsymbol{\gamma}$.

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 $\gamma_i = \pi_\sharp^i \boldsymbol{\gamma}$ 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

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

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

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

Since F_i is strictly convex, the above identity implies $\sigma_i = \sigma'_i = \sigma''_i$ μ_i-a.e. in X. When F_i are superlinear then $\gamma_i = \sigma_i$ thanks to (2.36). $\qquad\square$

The next corollary reduces the 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 pair 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 Sect. 3.3) admits a unique solution. Then $\mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ contains at most one plan.*

Proof We can assume $m_i = \mu_i(X_i) > 0$ for $i = 1, 2$. 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 marginals γ_i. Since $\gamma_i \ll \mu_i$ by (2.36) and γ_1 and γ_2 are unique by Lemma 3.5, we conclude. $\qquad\square$

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 $\mu_1 \ll \mathcal{L}^d$, 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 (Nonuniqueness 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 $\mathsf{c}(x_1, x_2) = h(|x_1 - x_2|)$. For the measures $\mu_1 = \delta_{(-1,0)} + \delta_{(1,0)}$, and μ_2 with support in $\{0\} \times \mathbb{R}$ and containing at least two points, there are an infinite number of optimal plans. In fact, we shall see that the first marginal γ_1 of any optimal plan γ will have full support in $A := \{(-1, 0), (1, 0)\}$, 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 $B := \{0\} \times \mathbb{R}$ and will contain at least two points. Any plan σ with marginals γ_1, γ_2 will then be optimal, since it will be supported in $A \times B$ where the cost c just depends on the second variable, since $|(\pm 1, 0) - (0, y)| = \sqrt{1 + y^2}$ for every $y \in \mathbb{R}$. Therefore the cost contribution of σ to the total energy is

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

and we can choose σ of the form [2, Sect. 5.3]

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

with arbitrary nonnegative densities α, β satisfying $\alpha + \beta = 1$ and $\int \alpha \, \mathrm{d}\gamma_2(y) = a$, $\int \beta \, \mathrm{d}\gamma_2(y) = b$ will be admissible.

We conclude this section by proving a simple lower semicontinuity property for the energy-transport functional ET. Note that in metrizable spaces any weakly convergent sequence of Radon measures is equally 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*

$$\liminf_{\lambda \in \mathbb{L}} \mathsf{ET}^\lambda\left(\mu_1^\lambda, \mu_2^\lambda\right) \geq \mathsf{ET}(\mu_1, \mu_2). \tag{3.24}$$

Proof Let $\gamma^\lambda \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1^\lambda, \mu_2^\lambda) \subset \mathcal{M}(X)$ be a corresponding net of optimal plans. The statement follows if, assuming that $\mathscr{E}(\gamma^\lambda | \mu_1^\lambda, \mu_2^\lambda) = \mathsf{ET}(\mu_1^\lambda, \mu_2^\lambda) \leq C < \infty$, we can prove that $\mathsf{ET}(\mu_1, \mu_2) \leq C$. By applying Proposition 2.10, we obtain that the sequences of marginals $\pi_\sharp^i \gamma^\lambda$ are equally tight in $\mathcal{M}(X_i)$, so that the net γ^λ is also equally tight by [2, Lemma 5.2.2]. By extracting a suitable subnet (not relabeled) narrowly converging to γ in $\mathcal{M}(X)$, we can still apply Proposition 2.10 and Corollary 2.9 to obtain

$$\liminf_{\lambda \in \mathbb{L}} \sum_i \mathscr{F}_i^\lambda\left(\gamma^\lambda | \mu_i^\lambda\right) \geq \sum_i \mathscr{F}_i(\gamma | \mu_i).$$

A standard monotonicity argument and the lower semicontinuity of the cost functions c^λ show that for every $\alpha \in \mathbb{L}$

$$\liminf_{\lambda \in \mathbb{L}} \int c^\lambda \, d\gamma^\lambda \geq \liminf_{\lambda \in \mathbb{L}} \int c^\alpha \, d\gamma^\lambda \geq \int c^\alpha \, d\gamma.$$

Passing now to the limit with respect to $\alpha \in \mathbb{L}$ and recalling (2.2) we conclude. □

As a simple application we prove the extremality of the class of Optimal Transport problems (see Example E.3 in Sect. 3.3) in the set of entropy-transport problems.

Corollary 3.10 *Let $F_1, F_2 \in \Gamma(\mathbb{R}_+)$ satisfy $F_i(r) > F_i(1) = 0$ 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 pair 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*

$$\lim_{n \uparrow \infty} \mathsf{ET}^n(\mu_{1,n}, \mu_{2,n}) = \mathsf{T}(\mu_1, \mu_2). \tag{3.25}$$

3.5 The reverse formulation of the primal problem

Let us recall the definition (2.28) of the reverse entropy functions R_i associated to F_i by the formula

$$R_i(r) := \begin{cases} r F_i(1/r) & \text{if } r > 0, \\ (F_i)'_\infty & \text{if } r = 0, \end{cases} \tag{3.26}$$

and let $\mathcal{R}_i$ be the corresponding integral functionals as in (2.57).

Keeping the notation of Lemma 2.3

$$\gamma_i := \pi^i_\sharp \gamma \in \mathcal{M}(X_i), \quad \mu_i = \varrho_i \gamma_i + \mu_i^\perp, \tag{3.27}$$

we can thus define

$$\mathcal{R}(\mu_1, \mu_2 | \gamma) := \sum_i \mathcal{R}_i(\mu_i | \gamma_i) + \int_X c \, d\gamma$$

$$= \int_X \Big(R_1(\varrho_1(x_1)) + R_2(\varrho_2(x_2)) + c(x_1, x_2) \Big) \, d\gamma + \sum_i F_i(0) \mu_i^\perp(X_i). \tag{3.28}$$

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

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

$$\mathscr{E}(\gamma|\mu_1, \mu_2) = \mathscr{R}(\mu_1, \mu_2|\gamma). \tag{3.29}$$

In particular

$$\mathsf{ET}(\mu_1, \mu_2) = \inf_{\gamma \in \mathcal{M}(X)} \mathscr{R}(\mu_1, \mu_2|\gamma), \tag{3.30}$$

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

The functional $\mathscr{R}(\mu_1, \mu_2|, \cdot)$ is still a convex functional, and it will be useful in Sect. 5.

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 Sect. 3.1. The derivation of the dual problem will be carried out in Sect. 4.1 by writing a saddle-point formulation of the primal problem 3.1 based on the duality Theorem 2.7 for the entropy functionals $\mathscr{F}_i$. The subsequent sections will then perform a systematic study of the duality and of the related optimality conditions.

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 will use the reverse entropy functions R_i defined as in (2.28) or Sect. 3.5 and their conjugates $R_i^* : \mathbb{R} \to (-\infty, +\infty]$, which can be expressed by

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

The equivalences (2.31) yield, for all $(\phi, \psi) \in \mathbb{R}^2$, the equivalence

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

As a first step we use the dual formulation of the entropy functionals given by Theorem 2.7 (cf. (2.47)) and find

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

It is now natural to introduce the saddle function $\mathscr{L}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ depending on $\boldsymbol{\gamma} \in \mathcal{M}(X)$ and $\boldsymbol{\psi} = (\psi_1, \psi_2)$ with $\psi_i \in \mathrm{LSC}_s(X_i, \mathring{\mathrm{D}}(R_i^*))$ (omitting the dependence on the fixed measures $\mu_i \in \mathcal{M}(X_i)$) via

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

Notice that $R_i^*(\psi_i)$ are bounded, so $\mathscr{L}$ cannot take the value $-\infty$; in order to guarantee that $\mathscr{L} < +\infty$, we consider the convex set

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

We thus have

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

and the Entropy-Transport Problem can be written as

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

The dual problem can be obtained by interchanging the order of inf and sup as in Sect. 2.2. 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 $\psi_i \in \mathrm{LSC}_s(X_i, \mathring{\mathrm{D}}(R_i^*))$,

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

we obtain

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

Thus, (4.6) 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 Ψ be the convex set*

$$\Psi := \left\{ \psi \in \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(R_1^*)) \times \mathrm{LSC}_s(X_2, \mathring{\mathrm{D}}(R_2^*)) : R_1^*(\psi_1) \oplus R_2^*(\psi_2) \le \mathsf{c} \right\}. \tag{4.7}$$

The dual Entropy-Transport problem consists in finding a maximizer $\psi \in \Psi$ for

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

As usual, by the following change of variable

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

as in (2.45) for the duality Theorem 2.7 (recall the notation $\phi_i = -\varphi_i$ we used in Sect. 2.3), we can obtain an equivalent formulation of the dual functional D as the supremum of the concave functionals

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

on the simpler convex set

$$\Phi := \left\{ \varphi \in \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(F_1^\circ)) \times \mathrm{LSC}_s(X_2, \mathring{\mathrm{D}}(F_2^\circ)), \; \varphi_1 \oplus \varphi_2 \le \mathsf{c} \right\}. \tag{4.11}$$

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

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

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 Let us first notice that replacing ψ_i with $\psi_i - \varepsilon$, $\varepsilon > 0$, and using the strict monotonicity of R_i^* in $(\mathrm{aff}(F_i)_\infty, F_i(0))$, as well as the fact that

$R_i^* \equiv -(F_i)_\infty'$ in $(-\infty, \mathrm{aff}(F_i)_\infty)$ and $\inf \mathsf{c} > -(F_1)_\infty' - (F_2)_\infty'$, one can replace Ψ in (4.8) by the smaller set

$$\Psi^\circ := \left\{ \psi \in \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(R_1^*)) \times \mathrm{LSC}_s(X_2, \mathring{\mathrm{D}}(R_2^*)) : R_1^*(\psi_1) \oplus R_2^*(\psi_2) < \mathsf{c} \right\}.$$

Since R_i^* is nondecreasing, for every $\psi \in \Psi^\circ$ the functions $\varphi_i := R_i^*(\psi_i) + \delta$, where $\delta := \frac{1}{2} \inf \mathsf{c} - R_1^*(\psi_1) \oplus R_2^*(\psi_2) > 0$, belong to $\mathrm{LSC}_s(X_i, \mathring{\mathrm{D}}(R_i^*))$ and satisfy $\varphi_1 \oplus \varphi_2 < \mathsf{c}$, with $(-\varphi_i, \psi_i) \in \mathring{\mathfrak{F}}_i$. It then follows that $(\varphi_1, \varphi_2) \in \Phi$ and $\tilde{\psi}_i := -F_i^*(-\varphi_i) = F_i^\circ(\varphi_i) \geq \psi_i$ so that $\mathsf{D}' \geq \mathsf{D}$. An analogous argument shows the converse inequality.   □

Since "inf sup $\geq$ sup inf" (cf. (2.10)), our derivation via (4.5) yields

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

Using Theorem 2.4 we will show in Sect. 4.3 that (4.13) is in fact an equality. Before this, we first discuss for which class of functions ψ_i, φ_i the dual formulations are still meaningful. Moreover, we analyze the optimality conditions associated to the equality case in (4.13).

Extension to Borel functions. In some cases we will also consider larger classes of potentials ψ or φ by allowing Borel functions with extended real values, under suitable summability conditions. It is clear that in the formulation of a dual problem it can be useful to deal with a smaller set of "competitors" (as in Problem 4.1 where we consider simple and lower semicontinuous functions) to derive various properties by exploiting the specific features of the involved functions. On the other hand, when one aims to prove the *existence* of dual optimizers, it is natural to enlarge the set of competitors in order to gain better closure properties. This is one of the main motivation to extend the dual formulation to general Borel functions.

First of all, recalling (2.19) and (2.29), we extend R^* and F° to $\bar{\mathbb{R}}$ by setting

$$\begin{aligned} R^*(-\infty) &:= -F_\infty', & R^*(+\infty) &:= +\infty; \\ F^\circ(-\infty) &:= -\infty, & F^\circ(+\infty) &:= F(0), \end{aligned} \tag{4.14}$$

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

$$(-\varphi, F^\circ(\varphi)) \text{ and } (-R^*(\psi), \psi) \text{ lie in } \bar{\mathfrak{F}} \text{ if } \psi \leq F(0) \text{ and } \varphi \geq -F_\infty'. \tag{4.15}$$

We also set

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

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

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

The following result is the natural extension of Lemma 2.6, stating that $\mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$ for a larger class of $\boldsymbol{\gamma}$ and $\boldsymbol{\varphi}$ than before.

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

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

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

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

$\square$

Proof Let us consider (4.18) in the case that $(F_i^\circ \circ \varphi_i)_- \in L^1(X_i, \mu_i)$ (the calculations in the other cases, including (4.19), 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 $(\varphi_i)_- \in L^1(X_i, \gamma_i)$ and then

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

$$(4.20)$$

Notice that the integrability of the negative part of φ_i w.r.t. γ_i yields $\varphi_i(\pi^i(x_1, x_2)) > -\infty$ for $\boldsymbol{\gamma}$-a.e. $(x_1, x_2) \in X$ 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

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

□

Optimality conditions. If there exists a pair $\boldsymbol{\varphi}$ 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

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

$$\int_X \left(\mathsf{c}(x_1, x_2) - (\varphi_1(x_1) +_o \varphi_2(x_2)) \right) \mathrm{d}\boldsymbol{\gamma} = 0.$$

The second part of Lemma 2.6 then yields

$$\varphi_1(x_1) +_o \varphi_2(x_2) = \mathsf{c}(x_1, x_2) \quad \boldsymbol{\gamma}\text{-a.e. in } X, \tag{4.21a}$$

$$-\varphi_i \in \partial F_i(\sigma_i) \quad (\mu_i + \gamma_i)\text{-a.e. in } A_i \tag{4.21b}$$

$$\varphi_i = -(F_i)_\infty' \quad \gamma_i^\perp\text{-a.e. in } A_{\gamma_i}, \tag{4.21c}$$

$$F_i^\circ(\varphi_i) = F_i(0) \quad \mu_i^\perp\text{-a.e. in } A_{\mu_i}, \tag{4.21d}$$

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

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

and the joint optimality conditions 4.21 is also sufficient to prove that a feasible $\boldsymbol{\gamma} \in \mathcal{M}(X)$ is optimal. We emphasize that we do not need any integrability assumption on $\boldsymbol{\varphi}$.

Theorem 4.6 *Let us suppose that* Problem 3.1 *is feasible (see* Remark 3.2) *for $\mu_i \in \mathcal{M}(X_i)$ and let $\boldsymbol{\gamma} \in \mathcal{M}(X)$; if there exists a pair $\boldsymbol{\varphi}$ as in (4.22) that satisfies the joint optimality conditions (4.21), then $\boldsymbol{\gamma}$ is optimal.*

Proof We want to repeat the calculations in (4.20) of Proposition 4.4, but now taking care of the integrability issues. We use a clever truncation argument of [43], 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

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

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 $0 \geq T_n(\varphi_1) + T_n(\varphi_2) \downarrow \varphi_1 + \varphi_2$ (notice that the cases when $\varphi_1 = \pm\infty$, $\varphi_2 = \mp\infty$ correspond to $T_n(\varphi_1) + T_n(\varphi_2) = \varphi_1 +_o \varphi_2 = 0 \leq \mathsf{c}$).

In particular if $\boldsymbol{\varphi}$ satisfies (4.22) then $T_n(\varphi_i) \in \mathrm{B}_b(X_i)$, $T_n(\varphi_1) \oplus T_n(\varphi_2) \leq \mathsf{c}$, and $T_n(\varphi_i) \geq -(F_i)'_\infty$ due to $(F_i)'_\infty \geq 0$ and $\varphi_i \geq -(F_i)'_\infty$. The boundedness of $T_n(\varphi_i)$ and Proposition 4.4 yield for every $\tilde{\boldsymbol{\gamma}} \in \mathcal{M}(X)$

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

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

$$\int_{X_i} F_i^\circ(T_n(\varphi_i)) \, d\mu_i \overset{(4.21\text{b, d})}{=} \int_{X_i} \left(F_{i,n}(\sigma_i) + \sigma_i T_n(\varphi_i) \right) d\mu_i$$

$$\overset{(4.21\text{c})}{=} \int_{X_i} F_{i,n}(\sigma_i) \, d\mu_i + (F_i)'_\infty \gamma_i^\perp(X_i) + \int_{X_i} T_n(\varphi_i) \, d\gamma_i,$$

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

$$\mathscr{E}(\tilde{\boldsymbol{\gamma}}|\mu_1, \mu_2) \geq \sum_i \left(\int_{X_i} F_{i,n}(\sigma_i) \, d\mu_i + (F_i)'_\infty \gamma_i^\perp(X_i) \right)$$

$$+ \int_X \left(T_n(\varphi_1) \oplus T_n(\varphi_2) \right) d\boldsymbol{\gamma}.$$

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, obtaining the desired optimality $\mathscr{E}(\tilde{\boldsymbol{\gamma}}|\mu_1, \mu_2) \geq \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$. □

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

(i) *$F_{i,n}$ are Lipschitz, $F_{i,n}(s) \leq F_i(s)$, and $F_{i,n}(s) \uparrow F_i(s)$ as $n \uparrow \infty$.*
(ii) *For every $s \in \mathrm{D}(F_i)$ and $\varphi_i \in \mathbb{R} \cup \{+\infty\}$ we have*

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

In particular, both cases in (4.26) *give* $F_i^\circ(T_n(\varphi_i)) = F_{i,n}(s) + s T_n(\varphi_i)$.

Proof Property (i): By (2.23) and the definition in (4.24) 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 $D(F_i^*) \supset (-\infty, 0)$ (see (2.18)), 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): Let us set $F_{i,n}^* := F_i^* + I_{[-n,n]}$ and notice that $F_{i,n} = (F_{i,n}^*)^*$ so that $(F_{i,n})^* = F_{i,n}^*$. Recalling that $(\partial F_i)^{-1} = \partial F_i^*$, $(\partial F_{i,n})^{-1} = \partial F_{i,n}^*$, $\partial F_i^* + \partial I_{[-n,n]} \subset \partial F_{i,n}^*$ and

$$\partial I_{[-n,n]}(\phi) = \begin{cases} 0 & \text{if } \phi \in (-n, n), \\ [0, \infty) & \text{if } \phi = n, \\ (-\infty, 0] & \text{if } \phi = -n, \\ \emptyset & \text{if } |\phi| > n, \end{cases}$$

we can easily prove the first implication of (4.26). In fact $-\varphi_i \in \partial F_i(s)$ yields $s \in \partial F_i^*(-\varphi_i)$ and $s = T_n(s) \in \partial F_{i,n}^*(-\varphi_i)$ if $|\varphi_i| \le n$; when $-\varphi_i > n$ the monotonicity of the subdifferential and the fact that $n \in \mathring{D}(F^*)_i \subset D(\partial F_i^*)$ yields $s \ge s'$ for every $s' \in \partial F_i(n)$ so that $s \in \partial F_i^* + [0, \infty) \subset \partial F_{i,n}^*$. A similar argument holds when $-\varphi_i < -n$.

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

The last statement in (ii) is an immediate application of (4.26) and the link between subdifferential and Fenchel duality stated in (2.17). □

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 Sect. 3.3), using the strengthened feasibility condition in (3.12). First note that the pair $\varphi_i \equiv 0$ provides an obvious lower bound for $D(\mu_1, \mu_2)$, viz.

$$D(\mu_1, \mu_2) \ge \mathscr{D}(0, 0 | \mu_1, \mu_2) = \sum_i m_i F_i^\circ(0) = \sum_i m_i \inf F_i. \tag{4.27}$$

We derive an upper and lower bound for the potential φ_1 under the assumption that c is bounded.

Lemma 4.8 *Let* $m_i = \mu_i(X_i)$ *and assume* $\mathrm{int}\,\big(m_1\mathrm{D}(F_1)\big) \cap m_2\,\mathrm{D}(F_2) \neq \emptyset$, *so that*

$$\exists\, s_1^-, s_1^+ \in \mathrm{D}(F_1),\ s_2 \in \mathrm{D}(F_2): \quad m_1 s_1^- < m_2 s_2 < m_1 s_1^+, \qquad (4.28)$$

and $S := \sup \mathsf{c} < \infty$. *Then every pair* $\boldsymbol{\varphi} = (\varphi_1, \varphi_2) \in \boldsymbol{\Phi}$ *with* $\mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2) \geq \sum_i m_i \inf F_i$ *satisfies*

$$\Phi_1^- \leq \sup \varphi_1 \leq \Phi_1^+,$$
$$\Phi_1^\pm := \frac{m_1(F_1(s_1^\pm) - \inf F_1) + m_2(F_2(s_2) - \inf F_2) + m_2 s_2 S}{m_2 s_2 - m_1 s_1^\pm}. \ (4.29)$$

Proof Since $\boldsymbol{\varphi} = (\varphi_1, \varphi_2) \in \boldsymbol{\Phi}$ satisfies $\sup \varphi_1 + \sup \varphi_2 \leq S$, the definition of $\mathscr{D}$ in (4.10) and the monotonicity of F° yield

$$\sum_i m_i \inf F_i \leq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2) \leq m_1 F_1^\circ(\sup \varphi_1) + m_2 F_2^\circ(S - \sup \varphi_1)$$

Using the dual bound $F_i^\circ(\varphi_i) \leq \varphi_i s_i + F_i(s_i)$ for $s_i \in \mathrm{D}(F_i)$ (cf. (4.9)) now implies

$$\sum_i m_i \inf F_i \leq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$$
$$\leq (m_1 s_1 - m_2 s_2) \sup \varphi_1 + m_1 F_1(s_1) + m_2 F_2(s_2) + m_2 s_2 S.$$

Exploiting (4.28), the choice $s_1 := s_1^-$ shows the upper bound in (4.29); and $s_1 = s_1^+$ the lower bound. $\qquad\qquad\square$

We improve the previous result by showing that in the case of bounded cost functions it is sufficient to consider bounded potentials φ_i. This 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; it will also play an important role in the third step of the proof of Theorem 4.11, which contains the main result concerning the dual representation.

Lemma 4.9 *If* $\sup \mathsf{c} = S < \infty$, *then for every pair* $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$, *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*

$$\sup \tilde{\varphi}_i - \inf \tilde{\varphi}_i \leq S, \quad 0 \leq \sup \tilde{\varphi}_1 + \sup \tilde{\varphi}_2 \leq S. \qquad (4.30)$$

If moreover (3.12) *holds, than there exist a constant* $\varphi_{\max} \geq 0$ *only depending on* F_i, m_i, S *such that*

$$-\varphi_{\max} \leq \inf \tilde{\varphi}_i \leq \sup \tilde{\varphi}_i \leq \varphi_{\max}. \qquad (4.31)$$

Proof Since $c \geq 0$, possibly replacing φ_1 with $\tilde{\varphi}_1 := \varphi_1 \vee (-\sup \varphi_2)$ we obtain a new pair $(\tilde{\varphi}_1, \varphi_2)$ with

$$\tilde{\varphi}_1 \geq \varphi_1, \quad \tilde{\varphi}_1(x_1) + \varphi_2(x_2) \leq \big(\varphi_1(x_1) + \varphi_2(x_2)\big) \wedge 0 \leq c(x_1, x_2)$$

so that $(\tilde{\varphi}_1, \varphi_2) \in \Phi$ 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

$$\sup \varphi_1 + \sup \varphi_2 \leq S \tag{4.32}$$

we thus obtain a new pair $(\tilde{\varphi}_1, \tilde{\varphi}_2) \in \Sigma$ with

$$\mathscr{D}(\tilde{\varphi}_1, \tilde{\varphi}_2 | \mu_1, \mu_2) \geq \mathscr{D}(\varphi_1, \varphi_2 | \mu_1, \mu_2), \quad \sup \tilde{\varphi}_i - \inf \tilde{\varphi}_i \leq S. \tag{4.33}$$

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 $\mathscr{D}$ while still preserving the constraint Φ. Thus, (4.30) is established.

When (3.12) holds (e.g. in the case considered by (4.28)) the previous Lemma 4.8 provides constants $\varphi_1^\pm$ such that $\varphi_1^- \leq \sup \tilde{\varphi}_1 \leq \varphi_1^+$. Now, (4.30) shows that $\varphi_2^- \leq \sup \tilde{\varphi}_2 \leq \varphi_2^+$ with $\varphi_2^- := -\varphi_1^+$ and $\varphi_2^+ := S - \varphi_1^-$. Applying (4.30) once again, we obtain (4.31) with $\varphi_{\max} := S + \varphi_1^+ - \varphi_1^-$. □

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

$$\varphi_1^c(x_2) := \inf_{x \in X_1} \big(c(x, x_2) - \varphi_1(x)\big) \quad \text{and} \quad \varphi_2^c(x_1) := \inf_{x \in X_2} \big(c(x_1, x) - \varphi_2(x)\big). \tag{4.34}$$

It is not difficult to show (see e.g. [2, Sect. 6.1]) that if $\varphi_1 \oplus \varphi_2 \leq c$ with $\sup \varphi_i < \infty$ then

$$\varphi_1^c \text{ and } \varphi_2^c \text{ are bounded}, \quad \varphi_1^{cc} \oplus \varphi_1^c \leq c, \quad \varphi_1^{cc} \geq \varphi_1, \quad \text{and } \varphi_1^c \geq \varphi_2. \tag{4.35}$$

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

Since F_i° are nondecreasing, it is also clear that whenever $\varphi_1^{cc}, \varphi_1^c$ are μ_i-measurable we have the estimate

$$\forall \boldsymbol{\varphi} \in B(X_1) \times B(X_2), \quad \varphi_1 \oplus \varphi_2 \leq c :$$
$$\mathscr{D}((\varphi_1, \varphi_2) | \mu_1, \mu_2) \leq \mathscr{D}((\varphi_1^{cc}, \varphi_2^c) | \mu_1, \mu_2) \tag{4.36}$$

The next lemma concerns the lower semicontinuity of φ_i^c in the case when c has the particular form (cf. [26])

$$\mathsf{c} = \sum_{n=1}^{N} c_n \chi_{A_n^1 \times A_n^2}, \quad \text{with } c_n \geq 0 \text{ and } A_n^i \text{ open in } X_i. \tag{4.37}$$

Lemma 4.10 *Let us assume that* c *has the form* (4.37) *and that* $\boldsymbol{\varphi} \in \mathrm{B}_s(X_1) \times \mathrm{B}_s(X_2)$ *is a pair of simple functions taking values in* $\mathrm{D}(F_1^\circ) \times \mathrm{D}(F_2^\circ)$ *and satisfying* $\varphi_1 \oplus \varphi_2 \leq \mathsf{c}$. *Then* $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}) \in \boldsymbol{\Phi}$ *with* $\mathscr{D}((\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}})|\mu_1, \mu_2) \geq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$.

Proof It is easy to check that $\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}$ are simple, since the infima in (4.34) are taken on a finite number of possible values. By (4.35) it is thus sufficient to check that they are lower semicontinuous functions.

We do this for φ_1^{c}, the argument for $\varphi_1^{\mathsf{cc}} = (\varphi_1^{\mathsf{c}})^{\mathsf{c}}$ is completely analogous. For this, consider the sets

$$Z := \left\{z = (z_n)_{n=1}^N \in \{0, 1\}^N : \exists\, y \in X_1 \,\forall n = 1, \ldots, N : z_n = \chi_{A_n^1}(y)\right\},$$
$$Y_z := \{y \in X_1 : \forall n = 1, \ldots, N : \chi_{A_n^1}(y) = z_n\}.$$

Clearly, $(Y_z)_{z \in Z}$ defines a Borel partition of X_1; we define $\varphi_z := \sup\{\varphi_1(y) : y \in Y_z\}$.

By construction, for every $z \in Z$ and $y \in Y_z$ the map $f_z(x) := \mathsf{c}(y, x) - \varphi_z$ is independent of y in Y_z and it is lower semicontinuous w.r.t. $x \in X_2$ since c is lower semicontinuous. Since $\varphi_1^{\mathsf{c}}(x_2)$ is the minimum of a finite collection of lower semicontinuous functions, viz.

$$\varphi_1^{\mathsf{c}}(x_2) = \min\left\{f_z(x_2) : z \in Z\right\} \tag{4.38}$$

we obtain $\varphi_1^{\mathsf{c}} \in \mathrm{LSC}(X_1)$. $\qquad\square$

With all these auxiliary results at hand, we are now ready to prove our main result concerning the dual representation using Theorem 2.4.

Theorem 4.11 *In the basic coercive setting of rm Sect. 3.1 (i.e.* (3.2a) *or* (3.2b) *hold), the Entropy-Transport functional* (3.4) *and the dual functional* (4.10) *satisfy*

$$\inf_{\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)} \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \sup_{\boldsymbol{\varphi} \in \boldsymbol{\Phi}} \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2) \ \text{ for every } \mu_i \in \mathcal{M}(X_i),$$
$$\tag{4.39}$$

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

Proof Since $\mathsf{ET} \geq \mathsf{D}$ is obvious, it suffices to show $\mathsf{ET} \leq \mathsf{D}$. In particular, 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. Starting from Step 2 we will assume that $(F_i)'_\infty = +\infty$ (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 1. We show that if the cost c has compact sublevels then (4.39) *holds:* 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 $B = \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(R_1^*)) \times \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(R_1^*))$. Conditions (2.9a) and (2.9b) are clearly satisfied; in order to check (2.11) we make use of the coercivity assumption $(F_1)'_\infty + (F_2)'_\infty + \min \mathsf{c} > 0$ to find $\boldsymbol{\psi}_\star = (\bar{\psi}_1, \bar{\psi}_2) \in B$ with constant functions $\bar{\psi}_i \in \mathring{\mathrm{D}}(R_i^*)$ and $-R^*(\bar{\psi}_i) = -\bar{\varphi}_i = \bar{\phi}_i \in (-\infty, (F_i)'_\infty)$ such that

$$D = \min\left(\mathsf{c} - (\bar{\varphi}_1 \oplus \bar{\varphi}_2)\right) = \bar{\phi}_1 + \bar{\phi}_2 + \min \mathsf{c} > 0.$$

Since

$$\mathscr{L}(\boldsymbol{\gamma}, \boldsymbol{\psi}_\star) = \int_X \left(\mathsf{c} - \min \mathsf{c}\right) \mathrm{d}\boldsymbol{\gamma} + D\boldsymbol{\gamma}(X) + \sum_i \bar{\psi}_i \mu_i(X_i),$$

we immediately see that for C sufficiently big the sublevels $\{\boldsymbol{\gamma} \in \mathrm{M} : \mathscr{L}(\boldsymbol{\gamma}, \boldsymbol{\psi}_*) \leq C\}$ are closed, bounded (since $D > 0$) and equally tight (by the compactness of the sublevels of c), thus narrowly compact. Thus, (4.39), i.e. $\mathsf{ET} = \mathsf{D}$, follows from Theorem 2.4; this concludes the proof of Theorem 4.11 in the case when (3.2b) holds.

From now on we consider the case (3.2a), by assuming F_i superlinear, i.e. $(F_i)'_\infty = +\infty$.

Step 2. We show that if μ_i have compact support, if (3.12) *is satisfied, and if the cost c has the form* (4.37), *then* (4.39) *holds:* 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 constant $E < \mathsf{ET}(\mu_1, \mu_2)$ we find $\boldsymbol{\varphi} \in \mathrm{LSC}_s(\tilde{X}_1) \times \mathrm{LSC}_s(\tilde{X}_2)$ such that $\varphi_1 \oplus \varphi_2 \leq \mathsf{c}$ in $\tilde{X}_1 \times \tilde{X}_2$, that $F_i^\circ(\varphi_i)$ is finite, and that $\sum_i \int_{\tilde{X}_i} F_i^\circ(\varphi_i) \, \mathrm{d}\mu_i \geq E$.

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 pair of simple Borel functions with $\varphi_1 \oplus \varphi_2 \leq \mathsf{c}$ in X. We can eventually apply Lemma 4.10 to find $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}) \in \Phi$ with $\mathscr{D}(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}}|\mu_1, \mu_2) \geq E$. Since $E < \mathsf{ET}(\mu_1, \mu_2)$ was arbitrary, we conclude that (4.39) holds in this case as well.

In the next step we remove the assumption on the compactness of $\mathrm{supp}(\mu_i)$.

Step 3. We show that if (3.12) *is satisfied and if the cost c has the form* (4.37), *then* (4.39) *holds:* Since μ_i are Radon, we 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 $n \to \infty$, i.e. $\mu_{i,n} := \chi_{K_{i,n}} \cdot \mu_i$ converges narrowly 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

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

This yields

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

Using the lower semicontinuity of ET from Lemma 3.9 we obtain

$$\mathsf{D}(\mu_1, \mu_2) \geq \liminf_{n\to\infty} \mathscr{D}(\boldsymbol{\varphi}_n | \mu_1, \mu_2) \geq \lim_{n\to\infty} E'_n$$
$$= \liminf_{n\to\infty} \mathsf{ET}(\mu_{1,n}, \mu_{2,n}) \geq \mathsf{ET}(\mu_1, \mu_2).$$

Thus, (4.39) is established.

In the next step we remove the assumption (3.12) on F_i.

Step 4. We show that if the cost c has the form (4.37), *then* (4.39) *holds:* It is sufficient to approximate F_i by an increasing and pointwise converging sequence $F_i^n \in \Gamma(\mathbb{R}_+)$; we will denote by ET^n the corresponding optimal Entropy-Transport functional. The corresponding sequence $(F_i^n)^\circ : \varphi_i \mapsto \sup_{s\geq 0}(F_i^n(s) + s\varphi_i)$ of conjugate concave functions is also nondecreasing and pointwise converging to F_i°. By the previous step, if $E_n < \mathsf{ET}^n(\mu_1, \mu_2)$ with $\lim_{n\to\infty} E_n = \lim_{n\to\infty} \mathsf{ET}^n(\mu_1, \mu_2) = \mathsf{ET}(\mu_1, \mu_2)$ (the latter limit follows by Lemma 3.9) we can find $\boldsymbol{\varphi}_n \in \boldsymbol{\Phi}$ such that

$$E_n \leq \sum_i \int_{X_i} (F_i^n)^\circ(\varphi_i^n) \, d\mu_i \leq \sum_i \int_{X_i} F_i^\circ(\varphi_i^n) \, d\mu_i = \mathscr{D}(\boldsymbol{\varphi}_n | \mu_1, \mu_2).$$

Passing to the limit $n \to \infty$ we conclude $\mathsf{ET}(\mu_1, \mu_2) \leq \mathsf{D}(\mu_1, \mu_2)$ as desired.

Step 5, conclusion. We show that (4.39) *holds for a general cost* c: Let $\mathsf{c} : X \to [0, \infty]$ be an arbitrary proper l.s.c. cost and let us denote by $(\mathsf{c}^\alpha)_{\alpha \in \mathbb{A}}$ the class of costs characterized by (4.37) and majorized by c. Then, $\mathbb{A}$ is a directed set with the pointwise order $\leq$, since maxima of a finite number of cost functions in $\mathbb{A}$ can still be expressed as in (4.37). 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 pair $\boldsymbol{\varphi}^\alpha \in \mathrm{LSC}_s(X_1, \mathring{\mathrm{D}}(F_1^\circ)) \times \mathrm{LSC}_s(X_2, \mathring{\mathrm{D}}(F_2^\circ))$ with such that $\varphi_1^\alpha \oplus \varphi_2^\alpha \leq \mathsf{c}^\alpha$ in X and $\mathscr{D}(\boldsymbol{\varphi}^\alpha | \mu_1, \mu_2) \geq E$. Since $\mathsf{c}^\alpha \leq \mathsf{c}$ we have $\boldsymbol{\varphi}^\alpha \in \boldsymbol{\Phi}$ and $\mathsf{ET}(\mu_1, \mu_2) \leq \mathsf{D}(\mu_1, \mu_2)$ follows. $\square$

Arguing as in Remark 2.8 we can change the spaces of test potentials $\boldsymbol{\varphi} = (\varphi_1, \varphi_2) \in \boldsymbol{\Phi}$ introduced in (4.11).

Corollary 4.12 *The duality formula* (4.39) *[and the equivalence with* (4.8)*] still holds if we replace the spaces of simple lower semicontinuous functions* $\mathrm{LSC}_s(X_i, \mathring{\mathrm{D}}(F_i^\circ))$ *(resp.* $\mathrm{LSC}_s(X_i, \mathring{\mathrm{D}}(R_i^*))$*) in the definition of* $\boldsymbol{\Phi}$ *(resp.* $\boldsymbol{\Psi}$*) with the corresponding spaces of bounded lower semicontinuous functions* LSC_b *or with the spaces of bounded Borel functions* B_b.

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

$$\mathsf{ET}(\mu_1, \mu_2) = \sup\left\{ \sum_i \int_{X_i} F^\circ(\varphi_i)\, \mathrm{d}\mu_i : \varphi_i,\ F_i^\circ(\varphi_i) \in \mathrm{C}_b(X_i),\ \varphi_1 \oplus \varphi_2 \leq \mathsf{c} \right\}$$

$$= \sup\left\{ \sum_i \int_{X_i} \psi_i\, \mathrm{d}\mu_i : \psi_i,\ R_i^*(\psi_i) \in \mathrm{C}_b(X_i), \right.$$
$$\left. R_1^*(\psi_1) \oplus R_2^*(\psi_2) \leq \mathsf{c} \right\}. \tag{4.40}$$

Corollary 4.13 (Subadditivity of ET) *The functional* ET *is convex and positively* 1*-homogeneous. In particular it is subadditive, in the sense that for every* $\mu_i, \mu_i' \in \mathcal{M}(X)$ *and* $\lambda \geq 0$ *we have*

$$\mathsf{ET}(\lambda\mu_1, \lambda\mu_2) = \lambda\, \mathsf{ET}(\mu_1, \mu_2),$$
$$\mathsf{ET}(\mu_1 + \mu_1', \mu_2 + \mu_2') \leq \mathsf{ET}(\mu_1, \mu_2) + \mathsf{ET}(\mu_1', \mu_2'). \tag{4.41}$$

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. $\square$

4.4 Existence of optimal Entropy-Kantorovich potentials

In this section we will consider two cases, when the dual problem admits a pair of optimal Entropy-Kantorovich potentials $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$.

The first case is completely analogous to the transport setting.

Theorem 4.14 *In the basic coercive setting of* Sect. 3.1 *(i.e.* (3.2a) *or* (3.2b) *hold) let us suppose that* (X_i, d_i), $i = 1, 2$, *are complete metric spaces, that* (3.12)*holds, and that* 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* $X = X_1 \times X_2$. *Then there exists a pair of optimal Entropy-Kantorovich potentials* $\boldsymbol{\varphi} \in C_b(X_1) \times C_b(X_2)$ *satisfying*

$$\varphi_1 \oplus \varphi_2 \le \mathsf{c}, \quad \varphi_i \ge -(F_i)_\infty', \quad \mathsf{ET}(\mu_1, \mu_2) = \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2). \qquad (4.42)$$

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

$$\begin{aligned}
\left| \mathsf{c}(x_1', x_2) - \mathsf{c}(x_1, x_2) \right| &\le \omega(\mathsf{d}_1(x_1', x_1)), \\
\left| \mathsf{c}(x_1, x_2') - \mathsf{c}(x_1, x_2) \right| &\le \omega(\mathsf{d}_2(x_2', x_2)).
\end{aligned}$$

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 (4.34) of a bounded function is 1-Lipschitz (and in particular Borel).

We apply Corollary 4.12: let $\boldsymbol{\varphi}_n$ be a maximizing sequence in $\boldsymbol{\Phi}$. By Lemma 4.9 we can assume that $\boldsymbol{\varphi}_n$ is uniformly bounded; by (4.35) and (4.36) 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 μ_i measure in X_i, by applying the Ascoli-Arzelà theorem on each compact set $K_{i,n}$ and a standard diagonal argument, we can extract a subsequence (still denoted by $\boldsymbol{\varphi}_n$) pointwise convergent to $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ in $A_1 \times A_2$. By setting $\varphi_i := \liminf_{n \to \infty} \varphi_{i,n}$, $i = 1, 2$, we extend $\boldsymbol{\varphi}$ to X and we obtain a pair $\varphi_i \in B_b(X_i)$ satisfying $\varphi_1 \oplus \varphi_2 \le \mathsf{c}$, $\varphi_i \ge -(F_i)_\infty'$ and

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

thanks to the pointwise convergence in A_i, Fatou's Lemma and the fact that $F_i^\circ(\varphi_{i,n})$ are uniformly bounded from above since $\varphi_{i,n}$ are uniformly bounded. Eventually replacing (φ_1, φ_2) with $(\varphi_1^{\mathsf{cc}}, \varphi_1^{\mathsf{c}})$ we obtain a pair in $C_b(X_1) \times C_b(X_2)$ satisfying (4.42) thanks to Proposition 4.4. $\qquad \square$

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

Theorem 4.15 *In the basic coercive setting of* Sect. 3.1 *(i.e.* (3.2a) *or* (3.2b) *hold), let* $\mu_i \in \mathcal{M}(X_i)$ *with* $\mathsf{ET}(\mu_1, \mu_2) < \infty$ *(i.e.* Problem 3.1 *is feasible) and let us suppose that at least one of the following two conditions hold:*

(a) c *is everywhere finite and* (3.12) *holds*
(b) $F_i(0) < \infty$.

Then a plan $\boldsymbol{\gamma} \in \mathcal{M}(X)$ *is optimal if and only if there exists a pair* $\boldsymbol{\varphi}$ *as in* (4.22) *satisfying the optimality conditions* (4.21) *with respect to a Borel partition* $(A_i, A_{\mu_i}, A_{\gamma_i})$ *related to the Lebesgue decomposition of* (γ_i, μ_i) *as in Lemma* 2.3.

Our proof starts with an auxiliary result on subdifferentials, which will be used extensively.

Lemma 4.16 *Let* $F \in \Gamma(\mathbb{R}_+)$, $s \in D(F)$, *let* $\phi \in \mathbb{R} \cup \{\pm\infty\}$ *be an accumulation point of a sequence* $(\phi_n) \subset \mathbb{R}$ *satisfying*

$$\lim_{n \to \infty} \left(F(s) - s\phi_n + F^*(\phi_n) \right) = 0. \tag{4.43}$$

If $\phi \in \mathbb{R}$ *then* $\phi \in \partial F(s)$, *if* $\phi = +\infty$ *then* $s = \max D(F)$ *and if* $\phi = -\infty$ *then* $s = \min D(F)$. *In particular, if* $s \in \mathring{D}(F)$ *then* 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 $w \in D(F)$ the Young inequality $w\phi_n \le F(w) + F^*(\phi_n)$ yields

$$\limsup_{n \to \infty}(w - s)\phi_n \le \limsup_{n \to \infty} F(w) - F(s) + \left(F(s) - s\phi_n + F^*(\phi_n) \right) \tag{4.44}$$
$$= F(w) - F(s)$$

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

If $\phi \in \mathbb{R}$ then $(w - s)\phi \le F(w) - F(s)$ for every $w \in D(F)$, so that $\phi \in \partial F(s)$. Since the right-hand side of (4.44) is finite for every $w \in D(F)$, if $\phi = +\infty$ then $w \le s$ for every $w \in D(F)$, so that $s = \max D(F)$. An analogous argument holds when $\phi = -\infty$. □

Proof of Theorem 4.15 We already proved (Theorem 4.6) that the existence of a pair $\boldsymbol{\varphi}$ as in (4.22) satisfying (4.21) yields the optimality of $\boldsymbol{\gamma}$.

Let us now assume that $\boldsymbol{\gamma} \in \mathcal{M}(X)$ is optimal. If $\mu_i \equiv \eta_0$, then we also have $\boldsymbol{\gamma} = 0$ and (4.21) is always satisfied, since 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 pairs of measures γ_i, μ_i provided by Lemma 2.3 and observing that the vanishing difference $\mathsf{ET}(\mu_1, \mu_2) - \mathscr{D}(\boldsymbol{\varphi}_n | \mu_1, \mu_2)$ can be decomposed in the sum of the following three nonnegative contributions

$$\mathscr{E}(\boldsymbol{\gamma} | \mu_1, \mu_2) - \mathscr{D}(\boldsymbol{\varphi}_n | \mu_1, \mu_2) = \int_{X_1 \times X_2} \Big(\mathsf{c}(x_1, x_2) - \varphi_{1,n}(x_1) - \varphi_{2,n}(x_2) \Big) \, \mathrm{d}\boldsymbol{\gamma}$$
$$+ \int_{A_i \cup A_{\mu_i}} \Big(F_i(\sigma_i) + \sigma_i \varphi_{i,n} - F_i^\circ(\varphi_{i,n}) \Big) \, \mathrm{d}\mu_i$$
$$+ \int_{A_{\gamma_i}} \Big(\varphi_{i,n} + (F_i)'_\infty \Big) \, \mathrm{d}\gamma_i^\perp$$

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 $A' \subset X$ 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 $\boldsymbol{\gamma}(X \setminus A') = 0$ such that

$$\mathsf{c}(x_1, x_2) < \infty, \quad \lim_{n \to \infty} \mathsf{c}(x_1, x_2) - \varphi_{1,n}(x_1) - \varphi_{2,n}(x_2) = 0 \quad \text{in } A', \quad (4.45)$$

$$F_i(\sigma_i) < \infty, \quad \lim_{n \to \infty} F_i(\sigma_i) + \sigma_i \varphi_{i,n} - F_i^\circ(\varphi_{i,n}) = 0 \quad \text{in } A_i' \cup A_{\mu_i}',$$
$$(4.46)$$

$$\lim_{n \to \infty} \big(\varphi_{i,n} + (F_i)'_\infty \big) = 0 \quad \text{in } A_{\gamma_i}'.$$
$$(4.47)$$

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

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

If $\mu_1(X_1) = 0$ then $(F_1)'_\infty$ is finite (since $\gamma_1^\perp(X_1) = \gamma_1(X_1) = \boldsymbol{\gamma}(X) > 0$) 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 $\varphi_{2,n}(x_2)$, Lemma 4.16 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 $\boldsymbol{\gamma}(X) > 0$. In order to check (4.21a) and (4.21b) we distinguish two cases.

Case a: c *is everywhere finite and* (3.12) *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 $\bar{x}_2$: Equation (4.46) and Lemma 4.16 (with $F = F_2$, $s = \sigma_2(\bar{x}_2)$, $\phi_n := -\varphi_{2,n}(\bar{x}_2)$) yield $s_2^+ = \max \mathrm{D}(F_2) = \sigma_2(\bar{x}_2) < \infty$ and $\sigma_2 \equiv s_2^+$ in $A'_2 \cup A'_{\mu_2}$. We thus have $\mathrm{D}(F_2) \subset [0, s_2^+]$, $(F_2)'_\infty = +\infty$ and therefore $m_2 s_2^+ = \boldsymbol{\gamma}(X)$.

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 4.16 yields that $\sigma_1 \equiv s_1^- = \min \mathrm{D}(F_1)$ μ_1-a.e. It follows that $m_1 s_1^- = \gamma_1(X_1) = \boldsymbol{\gamma}(X) = m_2 s_2^+$, and this contradicts (3.12).

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 $\varphi_{i,n}(x_i)$, Lemma 4.16 yields $-\varphi_i(x_i) \in \partial F_i(\sigma_i(x_i))$ in A'_i so that (4.21b) is also satisfied.

Case b: $F_i(0) < \infty$. 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 $F_i^\circ(\varphi_i) \in \mathrm{L}^1(X_1, \mu_1)$ 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 Lemma 4.16, since $\sigma_1(x_1) > 0 = \min \mathrm{D}(F_1)$ 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 $\varphi_1 \circ \pi^1 \in \mathbb{R}$ holds $\boldsymbol{\gamma}$-a.e. in X. It follows that (4.21a) holds and $\varphi_2 \circ \pi^2 \in \mathrm{L}^1(X, \boldsymbol{\gamma})$ so that $\varphi_2 \in \mathbb{R}$ $(\mu_2 + \gamma_2)$-a.e. in A'_2. A further application of Lemma 4.16 yields (4.21b) for $i = 2$. $\qquad\square$

Corollary 4.17 *Let us suppose that* $\mathrm{D}(F_i) \supset (0, \infty)$ *and* F_i *are differentiable in* $(0, \infty)$ *and let* $\mu_i \in \mathcal{M}(X_i)$ *with* $\mathsf{ET}(\mu_1, \mu_2) < \infty$. *A plan* $\boldsymbol{\gamma} \in \mathcal{M}(X)$ *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

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

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 (where we will keep the same notation of Corollary 4.17) 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 the marginals γ_i of an optimal plan $\boldsymbol{\gamma}$ cannot be too small, since $\mathrm{supp}(\gamma_i) \supset \mathrm{supp}(\mu_i) \setminus \mathrm{supp}(\mu_i^\perp)$.

Corollary 4.18 (Spread of the support) *Let us suppose that*

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

and let $\mu_i \in \mathcal{M}(X_i)$ *with* $\mathsf{ET}(\mu_1, \mu_2) < \infty$ *and* $\boldsymbol{\gamma} \in \mathcal{M}(X)$. *Then* $\boldsymbol{\gamma}$ *is an optimal plan if and only if* $\gamma_i \ll \mu_i$, *for every* $(x_1, x_2) \in \mathrm{supp}(\mu_1) \times \mathrm{supp}(\mu_2)$ *we have* $\mathsf{c}(x_1, x_2) = +\infty$ *if* $x_1 \in \mathrm{supp}\,\mu_1^\perp$ *or* $x_2 \in \mathrm{supp}\,\mu_2^\perp$, *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*

$$\begin{aligned} F_1'(\sigma_1) \oplus F_2'(\sigma_2) &\geq -\mathsf{c} \text{ in } A_1 \times A_2, \\ F_1'(\sigma_1) \oplus F_2'(\sigma_2) &= -\mathsf{c} \ \boldsymbol{\gamma}\text{-a.e. in } A_1 \times A_2. \end{aligned} \tag{4.50}$$

Remark 4.19 Apart from the case of pure transport problems (Example E.3 of Sect. 3.3), where the existence of Kantorovich potentials is well known (see [50, Thm. 5.10]), Theorem 4.15 covers essentially all the interesting cases, at least when the cost c takes finite values if $0 \notin \mathrm{D}(F_i)$. In fact, if the strengthened feasibility condition (3.12) 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. $X_i = \mathbb{R}$, $\mathsf{c}(x_1, x_2) := \frac{1}{2}|x_1 - x_2|^2$, $\mu_1 := e^{-\sqrt{\pi}x_1^2}\mathscr{L}^1$, $\mu_2 := e^{-\sqrt{\pi}(x_2+1)^2}\mathscr{L}^1$, and entropy functions F_i satisfying $\mathrm{D}(F_1) = [a, 1]$, $\mathrm{D}(F_2) = [1, b]$ with arbitrary choice of $a \in [0, 1)$

and $b \in (1, \infty]$. Since $m_1 = m_2 = 1$ the weak feasibility condition (3.1) holds, but (3.12) is violated. We find $\gamma_i = \mu_i$, $\sigma_i \equiv 1$, 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. $F_1(r) = (1-r)\log(1-r) + r$ with $D(F_1) = [0, 1]$, $F_2(r) = (r-1)\log(r-1) - r + 2$ with $D(F_2) = [1, \infty)$) so that (4.21b) can never hold, independently of the cost c.

5 "Homogeneous" formulations of optimal Entropy-Transport problems

Starting from the reverse formulation of the Entropy-Transport problem of Sect. 3.5 via the functional $\mathscr{R}$, see (3.28), 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 R_i and c.

We will always tacitly assume the *coercive* setting of Sect. 3.1, see (3.2).

5.1 The homogeneous marginal perspective functional

First of all we introduce the marginal perspective function H_c depending on the parameter $c \geq \inf \mathsf{c}$ (see [23, Chap. IV, Sect. 2.2] for the definition and the basic properties of the perspective; we use the term "marginal perspective" since we are infimizing w.r.t. the perspective parameter):

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

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

For $c = \infty$ we set

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

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

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

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 Lemma 5.3 below.

The marginal perspective cost (5.3) has an important interpretation in terms of the optimal Entropy Transport problem 3.1 between two Dirac masses: at least in the superlinear case (3.2a), it is easy to see that for every $x_i \in X$ and $r_i > 0, i = 1, 2$, we have

$$H(x_1, r_1; x_2, r_2) = \mathsf{ET}(r_1\delta_{x_1}, r_2\delta_{x_2}), \quad \text{when } (F_i)'_\infty = +\infty. \tag{5.4}$$

It is in fact sufficient to minimize $\mathscr{E}(\boldsymbol{\gamma}|r_1\delta_{x_1}, r_2\delta_{x_2})$ among the plans $\boldsymbol{\gamma}$ of the form $\theta\delta_{(x_1, x_2)}$.

Example 5.2 Let us consider the symmetric cases associated to the entropies U_p introduced in (1.4) and Example 2.5 and $V(s) = |s - 1|$:

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

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

A direct computation shows

$$\begin{aligned}
\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) &= r_1 + r_2 - 2\sqrt{r_1 r_2}\, \mathrm{e}^{-c/2} \\
&= \left(\sqrt{r_1} - \sqrt{r_2}\right)^2 + 2\sqrt{r_1 r_2}\left(1 - \mathrm{e}^{-c/2}\right).
\end{aligned} \tag{5.5}$$

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

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = r_1 \log r_1 + r_2 \log r_2 - (r_1 + r_2) \log\left(\frac{r_1 + r_2}{2 + c}\right). \tag{5.6}$$

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

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

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

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2)$$

$$= \frac{1}{p}\left[(r_1 + r_2) - \frac{r_1 r_2}{\left(r_1^{p-1} + r_2^{p-1}\right)^{1/(p-1)}}\left(2 - (p-1)c\right)_+^q\right], \quad (5.7)$$

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

$$\theta\left(U_{1-p}(\tfrac{r_1}{\theta}) + U_{1-p}(\tfrac{r_2}{\theta}) + c\right) = \frac{r_1^{1-p} + r_2^{1-p}}{p(p-1)}\theta^p$$

$$+ \frac{1}{p}(r_1 + r_2) + \frac{1}{p-1}((p-1)c - 2)\theta)$$

$$= \frac{1}{p}(r_1 + r_2) + \frac{1}{p-1}\left[\frac{1}{p}\left((r_1^{1-p} + r_2^{1-p})^{1/p}\,\theta\right)^p - \left(2 - (p-1)c\right)\theta\right],$$

and (5.7) follows by minimizing w.r.t. θ. For example, when $p = q = 2$,

$$H_c(r_1, r_2) = \frac{1}{2}(r_1 + r_2) - \frac{1}{2}\frac{r_1 r_2}{r_1 + r_2}(2 - c)_+^2$$

$$= \frac{1}{2(r_1 + r_2)}\left((r_1 - r_2)^2 + h(c)r_1 r_2\right), \quad (5.8)$$

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 (5.7) yields

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = \sqrt{(r_1^2 + r_2^2)(2 + 2c)} - (r_1 + r_2). \quad (5.9)$$

E.4 In the case of the total variation entropy $R_i(s) = V(s) = |s - 1|$ we easily find

$$\tilde{H}_c(r_1, r_2) = H_c(r_1, r_2) = r_1 + r_2 - (2 - c)_+(r_1 \wedge r_2)$$

$$= |r_2 - r_1| + (c \wedge 2)(r_1 \wedge r_2).$$

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*

$$H_c(r_1, r_2)$$

$$= \sup\left\{r_1 \psi_1 + r_2 \psi_2 : \psi_i \in D(R_i^*),\ R_1^*(\psi_1) + R_2^*(\psi_2) \le c\right\} \quad (5.10)$$

$$= \sup\left\{r_1 F_1^\circ(\varphi_1) + r_2 F_2^\circ(\varphi_2) : \varphi_i \in D(F_i^\circ),\ \varphi_1 + \varphi_2 \le c\right\}. \quad (5.11)$$

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*

$$H_c(r_1, r_2) \leq H_\infty(r_1, r_2) = \sum_i F_i(0) r_i \quad \text{for every } c \geq 0, \ r_i \geq 0. \quad (5.12)$$

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* $H_c(r_1, r_2) = \tilde{H}_c(r_1, r_2)$ *whenever* $r_1 r_2 > 0$.

(b) *If* $(F_1)'_\infty + c \geq -(F_2)'_0$ *and* $(F_2)'_\infty + c \geq -(F_1)'_0$, *then*

$$H_c(r_1, r_2) = \sum_i F_i(0) r_i \quad \text{if } r_1 r_2 = 0. \quad (5.13)$$

Proof Since $\sup D(R_i^*) = F_i(0)$ by (2.33), one immediately gets (5.10) in the case $c = \infty$; we can thus assume $c < \infty$.

It is not difficult [23, Chap. IV, Sect. 2.2] to check that the perspective 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), and it is concave w.r.t c since it is the infimum of a family of affine functions in c.

By Legendre duality [41, Thm. 12.2], its lower semicontinuous envelope is given by

$$H_c(r_1, r_2) = \sup \Big\{ \sum_i \psi_i r_i : H_c^*(\psi_1, \psi_2) \leq 0 \Big\}, \quad (5.14)$$

where

$$H_c^*(\psi_1, \psi_2) = \sup \Big\{ \sum_i \psi_i r_i - \tilde{H}_c(r_1, r_2) : r_i \geq 0 \Big\}$$

$$= \sup_{r_i \geq 0, \theta > 0} \sum_i \Big(\psi_i r_i - \theta R_i(r_i/\theta) \Big) - c\theta$$

$$= \sup_{\theta > 0} \theta \Big(\sum_i R_i^*(\psi_i) - c \Big)$$

$$= \begin{cases} 0 & \text{if } R_i^*(\psi_i) < \infty, \quad \sum_i R_i^*(\psi_i) \leq c \\ +\infty & \text{otherwise.} \end{cases}$$

Now, (5.11) immediately follows from (5.10) by the usual change of variable $\varphi_i = R_i^*(\psi_i)$, recall (2.31) and $F_i^\circ(\varphi_i) = -F_i^*(-\varphi_i)$.

In order to prove point (a) it is sufficient to recall that convex functions are always continuous in the interior of their domain [41, Thm. 10.1]. In particular,

since $\lim_{\theta \downarrow 0} \theta \big(R_1(r_1/\theta) + R_2(r_2/\theta) + c \big) = \sum_i (R_i)'_\infty r_i = \sum_i F_i(0) r_i$ for every $r_1, r_2 > 0$, we have $\tilde{H}_c(r_1, r_2) \leq \sum_i F_i(0) r_i$, 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 $\sup D(R_i^*) = F_i(0)$, $\lim_{r \uparrow F_i(0)} R_i^*(r) = -(F_i)'_0$, and $\inf R_i^* = -(F_i)'_\infty$ (see (2.33)) yield

$$H_c(r_1, 0) = \sup \left\{ \psi_1 r_1 : R_1^*(\psi_1) \leq c - \inf R_2^* \right\} = F_1(0) r_1.$$

An analogous formula holds when $0 = r_1 < r_2$. $\qquad\square$

A simple consequence of Lemma 5.3 and (2.31) is the lower bound

$$\tilde{H}_c(r_1, r_2) \geq H_c(r_1, r_2) \geq \sum_i \psi_i r_i \quad \text{for } (-\varphi_i, \psi_i) \in \mathfrak{F}_i \text{ with } \varphi_1 + \varphi_2 \leq c. \tag{5.15}$$

We now introduce the integral functional associated with the marginal perspective cost (5.3), which is based on the Lebesgue decomposition $\mu_i = \varrho_i \gamma_i + \mu_i^\perp$ (see Lemma 2.3),

$$\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) := \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, d\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i^\perp(X_i) \tag{5.16}$$

where we adopted the same notation as in (3.27). Let us first show that $\mathscr{H}$ is always greater than $\mathscr{D}$.

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

$$\int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, d\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i'(X_i) \geq \mathscr{D}(\boldsymbol{\varphi} | \mu_1, \mu_2). \tag{5.17}$$

Proof Recalling that $F_i^\circ(\varphi_i) = -F^*(-\varphi_i) \leq F_i(0)$ by (2.19) and (2.45), and using (5.15) with $r_j = \rho_j$ and $\psi_j = F_j^\circ(\rho_j)$ we have

$$\int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, d\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i'(X_i)$$

$$\overset{(5.15)}{\geq} \int_X \left(F_1^\circ(\varphi_1(x_1)) \varrho_1(x_1) + F_2^\circ(\varphi_2(x_2)) \varrho_2(x_2) \right) d\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i'(X_i)$$

$$= \sum_i \int_{X_i} F_i^\circ(\varphi_i) \varrho_i(x_i) \, d\gamma_i + \sum_i F_i(0) \mu_i'(X_i)$$

$$\overset{(2.19)}{\geq} \sum_i \int_{X_i} F_i^\circ(\varphi_i) \varrho_i(x_i) \, d\gamma_i + \sum_i \int_{X_i} F_i^\circ(\varphi_i) \, d\mu_i'$$

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

An immediate consequence of the previous lemma is the following important result concerning the marginal perspective cost functional $\mathscr{H}$ defined by (5.16). It can be nicely compared to the Reverse Entropy-Transport functional $\mathscr{R}$ for which Theorem 3.11 stated $\mathscr{R}(\mu_1, \mu_2|\boldsymbol{\gamma}) = \mathscr{E}(\boldsymbol{\gamma}|\mu_1, \mu_2)$.

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

$$\mathscr{R}(\mu_1, \mu_2|\boldsymbol{\gamma}) \geq \mathscr{H}(\mu_1, \mu_2|\boldsymbol{\gamma}) \geq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2). \tag{5.18}$$

In particular

$$\mathsf{ET}(\mu_1, \mu_2) = \mathsf{H}(\mu_1, \mu_2) := \min_{\boldsymbol{\gamma} \in \mathcal{M}(X)} \mathscr{H}(\mu_1, \mu_2|\boldsymbol{\gamma}), \tag{5.19}$$

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 $\mathcal{M}(X)$ and satisfies

$$H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) = \sum_i R_i(\varrho_i(x_i)) + \mathsf{c}(x_1, x_2) \quad \boldsymbol{\gamma}\text{-a.e. in} X, \tag{5.20}$$

where ϱ_i is defined as in (2.8). If moreover the following conditions

$$F_1(0) = +\infty \text{ or there exists } \bar{x}_2 \in X_2 \text{ with } \mu_2(\{\bar{x}_2\}) = 0,$$
$$F_2(0) = +\infty \text{ or there exists } \bar{x}_1 \in X_1 \text{ with } \mu_1(\{\bar{x}_1\}) = 0, \tag{5.21}$$

are satisfied, then

$$\mathsf{ET}(\mu_1, \mu_2) = \min \left\{ \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, d\boldsymbol{\gamma} : \right.$$

$$\left. \boldsymbol{\gamma} \in \mathcal{M}(X), \ \mu_i = \varrho_i \gamma_i \right\}. \tag{5.22}$$

Proof The inequality $\mathscr{R}(\mu_1, \mu_2|\boldsymbol{\gamma}) \geq \mathscr{H}(\mu_1, \mu_2|\boldsymbol{\gamma})$ is an immediate consequence of the fact that $\sum_i R_i(r_i) + c \geq \tilde{H}_c(r_1, r_2) \geq H_c(r_1, r_2)$ for every $r_i, c \in [0, \infty]$, obtained by choosing $\theta = 1$ in (5.1). The estimate $\mathscr{H}(\mu_1, \mu_2|\boldsymbol{\gamma}) \geq \mathscr{D}(\boldsymbol{\varphi}|\mu_1, \mu_2)$ was shown in 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 (5.19) and the characterization (5.20).

To establish the identity (5.22) we note that the difference to (5.19) only lies in dropping the additional restriction $\mu_i^\perp = 0$. When both $F_1(0) = F_2(0) = +\infty$ the equivalence is obvious since the finiteness of the functional $\boldsymbol{\gamma} \mapsto \mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma})$ yields $\mu_1^\perp = \mu_2^\perp = 0$.

In the general case, one immediately see that the right-hand side E' of (5.22) (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 (5.12), we obtain

$$
\begin{aligned}
\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) &= \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\gamma} \\
&\quad + \frac{1}{\tilde{m}_1 \tilde{m}_2} \int_X H(x_1, \tilde{m}_1; x_2, \tilde{m}_2) \, \mathrm{d}\mu_1^\perp \otimes \mu_2^\perp \\
&\le \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\gamma} + \sum_i F_i(0) \tilde{m}_i = \mathscr{H}(\mu_1, \mu_2 | \bar{\gamma}),
\end{aligned}
$$

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: since in this case $\tilde{m}_1 = \mu_1^\perp(X_1) > 0$ and therefore $F_1(0) < \infty$, by applying (5.21) we can choose $\boldsymbol{\gamma} := \bar{\boldsymbol{\gamma}} + \frac{1}{\tilde{m}_1} \mu_1^\perp \otimes \delta_{\bar{x}_2}$, obtaining

$$
\begin{aligned}
&\mathscr{H}(\mu_1, \mu_2 | \boldsymbol{\gamma}) \\
&= \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\gamma} + \frac{1}{\tilde{m}_1} \int_{X_1} H(x_1, \tilde{m}_1; \bar{x}_2, 0) \, \mathrm{d}\mu_1^\perp \\
&\le \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\gamma} + F_1(0) \tilde{m}_1 = \mathscr{H}(\mu_1, \mu_2 | \bar{\gamma}).
\end{aligned}
$$

$\square$

Remark 5.6 Notice that (5.21) is always satisfied if the spaces X_i are uncountable. If X_i is countable, one can always add an isolated point $\bar{x}_i$ (sometimes called "cemetery") to X_i and 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 c to $+\infty$ on $\bar{X}_1 \times \bar{X}_2 \setminus (X_1 \times X_2)$. We can recover (5.22) 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 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.

In particular, the following construction (typical in the Young measure approach to variational problems) allows us to consider the entropy-transport problems in a setting of greater generality. We replace a pair (γ, ϱ), where γ and ϱ are a measure on X and a nonnegative Borel function, by a measure $\alpha \in \mathcal{M}(Y)$ on the extended space $Y = X \times [0, \infty)$. The original pair (γ, ϱ) corresponds to a measure $\alpha = (x, \varrho(x))_{\sharp}\gamma$ concentrated on the graph of ϱ in Y and whose first marginal is γ.

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

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

$$\int_Y |y|_p^p \, d\alpha < \infty. \tag{5.23}$$

If $\alpha \in \mathcal{M}_p(Y)$ the measures $r_i^p \alpha$ belong to $\mathcal{M}(Y)$, which allow us to define the "p-homogeneous" marginal $\mathsf{h}_i^p(\alpha)$ of $\alpha \in \mathcal{M}_p(Y)$ as the x_i-marginal of $r_i^p \alpha$, namely

$$\mathsf{h}_i^p(\alpha) := \pi_{\sharp}^{x_i}(r_i^p \alpha) \in \mathcal{M}(X_i). \tag{5.24}$$

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

$$\mathrm{dil}_{\vartheta, p}(\alpha) := (\mathrm{prd}_{\vartheta})_{\sharp}(\vartheta^p \alpha), \quad \text{i.e. for } \varphi \in \mathrm{B}_b(Y)$$
$$\int \varphi(y) \, d(\mathrm{dil}_{\vartheta, p}(\alpha)) = \int \varphi(x_1, r_1/\vartheta; x_2, r_2/\vartheta)\vartheta^p(y) \, d\alpha(y). \tag{5.25}$$

Using (5.24) we obviously have

$$h_i^p(\mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha})) = h_i^p(\boldsymbol{\alpha}). \tag{5.26}$$

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

$$\vartheta(\boldsymbol{y}) := \frac{1}{r_*} \begin{cases} |\boldsymbol{y}|_p & \text{if } |\boldsymbol{y}|_p \neq 0, \\ 1 & \text{if } |\boldsymbol{y}|_p = 0, \end{cases} \quad r_* := \left(\int_{\boldsymbol{Y}} |\boldsymbol{y}|_p^p \, \mathrm{d}\boldsymbol{\alpha} + \boldsymbol{\alpha}(\{|\boldsymbol{y}|_p = 0\}) \right)^{1/p} \tag{5.27a}$$

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

$$\tilde{\boldsymbol{\alpha}} = \mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha}) \in \mathcal{P}_p(\boldsymbol{Y}), \quad h_i^p(\tilde{\boldsymbol{\alpha}}) = h_i^p(\boldsymbol{\alpha}), \quad \tilde{\boldsymbol{\alpha}}(\boldsymbol{Y} \setminus \boldsymbol{Y}_{r_*,p}) = 0. \tag{5.27b}$$

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

$$\begin{aligned}
\mathcal{H}_{\leq}^p(\mu_1, \mu_2) &:= \left\{ \boldsymbol{\alpha} \in \mathcal{M}_p(\boldsymbol{Y}) : h_i^p(\boldsymbol{\alpha}) \leq \mu_i \right\}, \\
\mathcal{H}_{=}^p(\mu_1, \mu_2) &:= \left\{ \boldsymbol{\alpha} \in \mathcal{M}_p(\boldsymbol{Y}) : h_i^p(\boldsymbol{\alpha}) = \mu_i \right\}.
\end{aligned} \tag{5.28}$$

Clearly $\mathcal{H}_{=}^p(\mu_1, \mu_2) \subset \mathcal{H}_{\leq}^p(\mu_1, \mu_2)$ and they are nonempty since every plan of the form

$$\boldsymbol{\alpha} = \frac{1}{a_1^p a_2^p} \left(\mu_1 \otimes \delta_{a_1} \right) \otimes \left(\mu_2 \otimes \delta_{a_2} \right), \quad \text{with } a_1, a_2 > 0 \tag{5.29}$$

belongs to $\mathcal{H}_{=}^p(\mu_1, \mu_2)$. It is not difficult to check that $\mathcal{H}_{\leq}^p(\mu_1, \mu_2)$ is also narrowly closed, while, on the contrary, this property fails for $\mathcal{H}_{=}^p(\mu_1, \mu_2)$ if $\mu_1(X_1)\mu_2(X_2) \neq 0$. To see this, it is sufficient to consider any $\boldsymbol{\alpha} \in \mathcal{H}_{=}^p(\mu_1, \mu_2) \setminus \{0\}$ and look at the vanishing sequence $\mathrm{dil}_{n^{-1},p}(\boldsymbol{\alpha})$ for $n \to \infty$.

There is a natural correspondence between $\mathcal{H}_{\leq}^p(\mu_1, \mu_2)$ (resp. $\mathcal{H}_{=}^p(\mu_1, \mu_2)$) and $\mathcal{H}_{\leq}^1(\mu_1, \mu_2)$ (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 $\boldsymbol{\alpha} \in \mathcal{H}_{\leq}^1(\mu_1, \mu_2)$ we can prove a result similar to Lemma 5.4 but now we obtain a linear functional in $\boldsymbol{\alpha}$.

Lemma 5.7 *For $p \in (0, \infty)$, $\mu_i \in \mathcal{M}(X_i)$, $\boldsymbol{\varphi} \in \Phi$, and $\boldsymbol{\alpha} \in \mathcal{H}_{\leq}^p(\mu_1, \mu_2)$ we have*

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

where $\mu_i' := \mu_i - h_i^p(\boldsymbol{\alpha})$. $\tag{5.30}$

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

$$\int_Y H(x_1, r_1^p; x_2, r_2^p) \, d\boldsymbol{\alpha} + \sum_i F_i(0)\mu_i'(X_i)$$

$$\overset{(5.15)}{\geq} \int_Y \left(F_1^\circ(\varphi_1(x_1))r_1^p + F_2^\circ(\varphi_2(x_2))r_2^p \right) d\boldsymbol{\alpha} + \sum_i F_i(0)\mu_i'(X_i)$$

$$= \sum_i \int_{X_i} F_i^\circ(\varphi_i) \, d(\mathrm{h}_i^p(\boldsymbol{\alpha})) + \sum_i F_i(0)\mu_i'(X_i)$$

$$\overset{(2.19)}{\geq} \sum_i \int_{X_i} F_i^\circ(\varphi_i) \, d(\mathrm{h}_i^p(\boldsymbol{\alpha})) + \sum_i \int_{X_i} F_i^\circ(\varphi_i) \, d\mu_i'$$

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

$$\square$$

As a consequence, we can characterize the entropy-transport minimum via measures $\boldsymbol{\alpha} \in \mathcal{M}(Y)$.

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

$$\mathsf{ET}(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha} \in \mathcal{H}_{\leq}^p(\mu_1,\mu_2)} \int_Y \left(\sum_i R_i(r_i^p) + \mathsf{c}(x_1, x_2) \right) d\boldsymbol{\alpha}$$

$$+ \sum_i F_i(0)(\mu_i - \mathrm{h}_i^p(\boldsymbol{\alpha}))(X_i) \qquad (5.31)$$

$$= \min_{\boldsymbol{\alpha} \in \mathcal{H}_{\leq}^p(\mu_1,\mu_2)} \int_Y H(x_1, r_1^p; x_2, r_2^p) \, d\boldsymbol{\alpha} + \sum_i F_i(0)(\mu_i - \mathrm{h}_i^p(\boldsymbol{\alpha}))(X_i)$$

$$(5.32)$$

$$= \min_{\boldsymbol{\alpha} \in \mathcal{H}_{=}^p(\mu_1,\mu_2)} \int_Y H(x_1, r_1^p; x_2, r_2^p) \, d\boldsymbol{\alpha}. \qquad (5.33)$$

Moreover, for every plan $\boldsymbol{\gamma} \in \mathrm{Opt}_{\mathsf{ET}}(\mu_1, \mu_2)$ (resp. optimal for (5.19) or for (5.22)) 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 (5.31) (resp. (5.32) or (5.33)).

Remark 5.9 When $F_i(0) = +\infty$ (5.31) and (5.32) simply read as

$$
\mathsf{ET}(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha} \in \mathcal{H}^p_{\le}(\mu_1, \mu_2)} \int_Y \left(\sum_i R_i(r_i^p) + \mathsf{c}(x_1, x_2) \right) d\boldsymbol{\alpha}
$$

$$
= \min_{\boldsymbol{\alpha} \in \mathcal{H}^1_{\le}(\mu_1, \mu_2)} \int_Y H(x_1, r_1^p; x_2, r_2^p) \, d\boldsymbol{\alpha}.
$$

$\square$

Proof of Theorem 5.8 Let us denote by E' (resp. E'', E''') the right-hand side of (5.31) (resp. of (5.32), (5.33)), where "min" has been replaced by "inf". If $\boldsymbol{\gamma} \in \mathcal{M}(X)$ and $\mu_i = \varrho_i \gamma_i + \mu_i^\perp$ (in the case of (5.33) $\mu_i^\perp = 0$) is the usual Lebesgue decomposition as in Lemma 2.3 , 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 $(\varrho_1^{1/p}, \varrho_2^{1/p}) : X \to \mathbb{R}^2$ is Borel and takes values in a metrizable and separable space, it is Lusin $\boldsymbol{\gamma}$-measurable [44, Thm 5, p. 26], so that $\boldsymbol{\alpha}$ is a Radon measure in $\mathcal{M}(Y)$. For every nonnegative $\phi_i \in B_b(X_i)$ we easily get

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

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

$$
\mathscr{R}(\mu_1, \mu_2 | \boldsymbol{\gamma}) = \int_X \left(\sum_i R_i(\varrho_i(x_i)) + \mathsf{c}(x_1, x_2) \right) d\boldsymbol{\gamma} + \sum_i F_i(0) \mu_i^\perp(X_i)
$$

$$
= \int_Y \sum_i R_i(r_i^p) + \mathsf{c}(x_1, x_2) \Big) d\boldsymbol{\alpha}
$$

$$
+ \sum_i F_i(0)(\mu_i - \mathsf{h}_i^p(\boldsymbol{\alpha}))(X_i) \ge E'.
$$

Taking the infimum w.r.t. $\boldsymbol{\gamma}$ and recalling (3.30) 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 $\boldsymbol{\varphi} \in \Phi$. 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 (5.21) hold, by choosing $\boldsymbol{\alpha}$ induced by a minimizer of (5.22) we get the opposite inequality $E''' \le \mathsf{ET}(\mu_1, \mu_2)$.

If (5.21) 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

$$
\int_Y H(x_1, r_1^p; x_2, r_2^p) \, \mathrm{d}\boldsymbol{\alpha} = \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\boldsymbol{\gamma}}
$$

$$
+ \int_X H(x_1, 1; x_2, 0) \, \mathrm{d}\mu_1^\perp \otimes \nu
$$

$$
\overset{(5.12)}{\leq} \int_X H(x_1, \varrho_1(x_1); x_2, \varrho_2(x_2)) \, \mathrm{d}\bar{\boldsymbol{\gamma}} + F_1(0)\tilde{m}_1
$$

$$
= \mathcal{H}(\mu_1, \mu_2 | \bar{\boldsymbol{\gamma}}) = \mathsf{ET}(\mu_1, \mu_2),
$$

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

Remark 5.10 (Rescaling invariance) By recalling (5.27a,b) and exploiting the 1-homogeneity of H it is not restrictive to solve the minimum problem (5.32) in the smaller class of probability plans concentrated in

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

First, we note that it is not restrictive to assume that $\boldsymbol{\alpha}(\{y \in Y : |y|_p = 0\}) = 0$ in (5.32): we can always replace $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}' := \boldsymbol{\alpha} \llcorner \{y \in Y : |y|_p > 0\}$ since $H(x_1, 0; x_2, 0) = 0$ for every $x_i \in X_i$ and the homogeneous marginals of $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ coincide. As a second step, for every $\boldsymbol{\alpha} \in \mathcal{H}_{\leq}^p(\mu_1, \mu_2)$ with $\boldsymbol{\alpha}(Y) > 0$, the choice $\tilde{\boldsymbol{\alpha}}$ given by (5.27a,b) yields a new probability plan concentrated on $Y_{r_*,p} \subset Y_{r,p}$ with the same homogeneous marginals as $\boldsymbol{\alpha}$ and satisfying

$$
\int_Y H(x_1, r_1^p; x_2, r_2^p) \, \mathrm{d}\boldsymbol{\alpha} = \int_Y H(x_1, (r_1/\vartheta)^p; x_2, (r_2/\vartheta)^p)\vartheta^p \, \mathrm{d}\tilde{\boldsymbol{\alpha}}
$$

$$
= \int_Y H(x_1, r_1^p; x_2, r_2^p) \, \mathrm{d}\tilde{\boldsymbol{\alpha}},
$$

where ϑ is the function defined in (5.27a) and we used the 1-homogeneity of H w.r.t. the variables (r_1, r_2).

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 nonseparable extended distances play an important role, so that it will be useful to deal with an auxiliary topology, see e.g. [1,3].

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

$$F_i(s) = U_1(s) := s \log s - s + 1, \qquad (6.1)$$

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

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

so that

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

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

LE.1 F_i are superlinear, C^∞ in $(0, \infty)$, strictly convex, with $D(F_i) = [0, \infty)$, $F_i(0) = 1$, and $(F_i)'_0 = -\infty$. For $s > 0$ we have $\partial F_i(s) = \{\log s\}$.

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

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

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

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

$$\ell''(d) = 2 \exp(\ell(d)), \quad \ell(0) = \ell'(0) = 0, \qquad (6.4)$$

since it satisfies

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

so that

$$\ell(d) \ge d^2, \quad \ell'(d) = 2\tan d \ge 2d, \quad \ell''(d) \ge 2. \tag{6.6}$$

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 $2\ell\ell'' \ge (\ell')^2$ and a direct calculation shows

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

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

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

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

Since the function

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

will have an important geometric interpretation (see Sect. 7.1), in the following we will choose the exponent $p = 2$ in the setting of Sect. 5.2.

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

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 $\gamma \in \mathcal{M}(X)$ minimizing*

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

where $\sigma_i = \frac{\mathrm{d}\gamma_i}{\mathrm{d}\mu_i}$. We denote by $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ the set of all the minimizers γ in (6.10).

6.2 The Logarithmic Entropy-Transport problem: main results

In the next theorem we collect the main properties of the Logarithmic Entropy-Transport (LET) problem relying on the reverse function $\mathscr{R}$ from Sect. 3.5, cf. (3.28), and $\mathscr{H}$ from Sect. 5.1, cf. (5.16).

Theorem 6.2 (Direct formulation of the LET problem) *Let $\mu_i \in \mathcal{M}(X)$ be given and let ℓ, $\mathsf{d}_{\pi/2}$ 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 $\mathcal{M}(X)$, LET is a convex and positively 1-homogeneous functional (see* (4.41)*) satisfying $0 \leq \mathsf{LET}(\mu_1, \mu_2) \leq \sum_i \mu_i(X)$.*

(b) ***Reverse formulation*** *($\mathsf{LET} = \min \mathscr{R}_{\mathsf{LE}}$). The functional LET has the equivalent reverse formulation as*

$$\mathsf{LET}(\mu_1, \mu_2) = \min\left\{\mathscr{R}_{\mathsf{LE}}(\mu_1, \mu_2|\boldsymbol{\gamma}) \; : \; \boldsymbol{\gamma} \in \mathcal{M}(X), \; \mu_i = \varrho_i \gamma_i + \mu_i^{\perp}\right\},$$

where

$$\mathscr{R}_{\mathsf{LE}}(\mu_1, \mu_2|\boldsymbol{\gamma}) := \sum_i \left(\mu_i^{\perp}(X) + \int_X \left(\varrho_i - 1 - \log \varrho_i\right) \mathrm{d}\gamma_i\right)$$
$$+ \int_X \ell\big(\mathsf{d}(x_1, x_2)\big) \mathrm{d}\boldsymbol{\gamma}, \qquad (6.11)$$

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*** *($\mathsf{LET} = \min \mathscr{H}_{\mathsf{LE}}$). The functional $\mathsf{LET}(\mu_1, \mu_2)$ can be equivalently characterized as*

$$\mathsf{LET}(\mu_1, \mu_2) = \min\left\{\mathscr{H}_{\mathsf{LE}}(\mu_1, \mu_2|\boldsymbol{\gamma}) \; : \; \boldsymbol{\gamma} \in \mathcal{M}(X)\right\}, \; \text{where}$$

$$\mathscr{H}_{\mathsf{LE}}(\mu_1, \mu_2|\boldsymbol{\gamma}) := \sum_i \mu_i(X) - 2 \int_X \sqrt{\varrho_1(x_1)\varrho_2(x_2)} \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \mathrm{d}\boldsymbol{\gamma}$$

$$= \sum_i \mu_i^{\perp}(X) + \int_X \left(\varrho_1(x_1) + \varrho_2(x_2) - 2\sqrt{\varrho_1(x_1)\varrho_2(x_2)} \cos(\mathsf{d}_{\pi/2}(x_1, x_2))\right) \mathrm{d}\boldsymbol{\gamma}$$
$$(6.12)$$

and $\gamma_i = \varrho_i \mu_i + \mu_i^{\perp}$. Moreover, 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 Sect. 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 pair $\phi_i : X \to \mathbb{R}$ 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 X. In part a) we relate to Sect. 4.2, whereas b)–d) discusses the optimality conditions from Sect. 4.4.

Theorem 6.3 (Dual formulation and optimality conditions)
  *(a) **The dual problem** ($\mathsf{LET} = \mathsf{D}_{\mathsf{LE}} = \mathsf{D}'_{\mathsf{LE}}$). For all $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{LET}(\mu_1, \mu_2) = \sup \left\{ \mathscr{D}_{\mathsf{LE}}(\boldsymbol{\varphi}|\mu_1, \mu_2) \; : \; \varphi_i \in \mathrm{LSC}_s(X), \; \varphi_1 \oplus \varphi_2 \le \ell(\mathsf{d}) \right\}, \tag{6.13}$$

$$= \sup \left\{ \sum_i \int_X \psi_i \, \mathrm{d}\mu_i \; : \; \psi_i \in \mathrm{LSC}_s(X), \; \sup_X \psi_i < 1, \right.$$

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

where $\mathscr{D}_{\mathsf{LE}}(\boldsymbol{\varphi}|\mu_1, \mu) := \sum_i \int_X \left(1 - \mathrm{e}^{-\varphi_i}\right) \mathrm{d}\mu_i$. 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 $\boldsymbol{\gamma} \in \mathcal{M}(X)$ is optimal if and only if its marginals γ_i are absolutely continuous w.r.t. μ_i,*

$$\mathsf{d} \ge \pi/2 \quad \text{in} \quad \left(\mathrm{supp}\, \mu_1^\perp \times \mathrm{supp}\, \mu_2 \right) \bigcup \left(\mathrm{supp}\, \mu_1 \times \mathrm{supp}\, \mu_2^\perp \right), \tag{6.15}$$

and there exist Borel sets $A_i \subset \mathrm{supp}\, \gamma_i$ 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

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

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

or, equivalently, in terms of the densities $\varrho_i = \sigma_i^{-1}$ of μ_i w.r.t. γ_i

$$\varrho_1(x_1)\varrho_2(x_2) \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \le 1 \quad \text{in } A_1 \times A_2, \tag{6.18}$$

$$\varrho_1(x_1)\varrho_2(x_2) \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) = 1 \quad \boldsymbol{\gamma}\text{-a.e. in } A_1 \times A_2. \tag{6.19}$$

  *(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 T with cost $\ell(\mathsf{d})$ (see (3.15) of*

Sect. 3.3) *between its marginals γ_i. In particular it is $\ell(\mathsf{d})$-cyclically mono-
tone, i.e. it is concentrated on a Borel set $G \subset X$ ($G = \mathrm{supp}(\boldsymbol{\gamma})$ when d is
continuous) such that for every choice of $(x_1^n, x_2^n)_{n=1}^N \subset G$ and every permu-
tation $\kappa : \{1, \ldots, N\} \to \{1, \ldots, N\}$*

$$\Pi_{n=1}^N \cos^2 \left(\mathsf{d}_{\pi/2}(x_1^n, x_2^n) \right) \geq \Pi_{n=1}^N \cos^2 \left(\mathsf{d}_{\pi/2}(x_1^n, x_2^{\kappa(n)}) \right). \tag{6.20}$$

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

$$\varphi_i := \begin{cases} -\log \sigma_i & \text{in } A_i, \\ -\infty & \text{in } X \setminus \mathrm{supp}\, \mu_i, , \\ +\infty & \text{otherwise,} \end{cases} \qquad \psi_i := \begin{cases} 1 - \sigma_i & \text{in } A_i, \\ -\infty & \text{in } X \setminus \mathrm{supp}\, \mu_i, \\ 1 & \text{otherwise,} \end{cases} \tag{6.21}$$

*satisfy $\varphi_1 \oplus_o \varphi_2 \leq \ell(\mathsf{d})$, $\log(1-\psi_1) \oplus_o \log(1-\psi_2) \geq \ell(\mathsf{d})$, and the optimality
conditions corresponding to* (4.50)

$$\varphi_1(x_1) + \varphi_2(x_2) = \log(1-\psi_1(x_1)) + \log(1-\psi_2(x_2))$$
$$= \ell(\mathsf{d}(x_1, x_2)) \quad \boldsymbol{\gamma}\text{-a.e. in } A_1 \times A_2.$$

Moreover $\mathrm{e}^{-\varphi_i}, \psi_i \in \mathrm{L}^1(X, \mu_i)$ and

$$\mathsf{LET}(\mu_1, \mu_2) = \sum_i \int_X \left(1 - \mathrm{e}^{-\varphi_i}\right) \mathrm{d}\mu_i$$

$$= \sum_i \int_X \psi_i \, \mathrm{d}\mu_i = \sum_i \mu_i(X) - 2\boldsymbol{\gamma}(X). \tag{6.22}$$

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

Identity (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 Corol-
lary 4.12.

The statement of point (b) follows by Corollary 4.18; notice that the LET
problem is always feasible.

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

Point (d) can be easily deduced by (b) or by applying Corollaries 4.17
and 4.18, observing that the formula defining φ_i of (6.21) corresponds to
(4.48) with $(F_i)'_\infty = +\infty = -(F_i)'_0$ and the optimality condition corresponds
to (4.50). Finally, ψ_i are just related to φ_i by $\psi_i = F_i^\circ(\varphi_i) = 1 - \exp(-\varphi_i)$. □

In the one-dimensional case, the $\ell(\mathsf{d})$-cyclic monotonicity of part (c) of the previous theorem reduces to classical monotonicity.

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.*

$$(x_1, x_2), \ (x_1', x_2') \in \mathrm{supp}(\boldsymbol{\gamma}), \ x_1 < x_1' \quad \Rightarrow \quad x_2 \leq x_2'. \qquad (6.23)$$

Proof As the function ℓ is uniformly convex, (6.20) is equivalent to monotonicity. □

The next result provides a variant of the reverse formulation in Theorem 6.2, which expresses the LET problem as a supremum of the *linear* mass functional on $\boldsymbol{\gamma}$ on a convex set characterized by the marginals of $\boldsymbol{\gamma}$ and the cost.

Corollary 6.5 *For all $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{LET}(\mu_1, \mu_2) = \sum_i \mu_i(X) - 2 \max \Big\{ \boldsymbol{\gamma}(X) : \ \boldsymbol{\gamma} \in \mathcal{M}(X), \ \gamma_i = \sigma_i \mu_i,$$

$$\sigma_1(x_1)\sigma_2(x_2) \leq \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \ \boldsymbol{\gamma}\text{-a.e. in } X \Big\}. \qquad (6.24)$$

Proof Let us denote by M' the right-hand side and let $\boldsymbol{\gamma} \in \mathcal{M}(X)$ be a plan satisfying the conditions of (6.24). 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, the densities ϱ_i of μ_i w.r.t. γ_i satisfy $\varrho_i(x_i) = 1/\sigma_i(x_i)$ in A_i so that $\sigma_1(x_1)\sigma_2(x_2) \leq \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2))$ yields $\varrho_1(x_1)\varrho_2(x_2) \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \geq 1$. Since $(\log \varrho_i)_+ \in \mathrm{L}^1(X, \gamma_i)$ we have

$$\sum_i \Big(\mu_i^\perp(X) + \int_X \big(\varrho_i - 1 - \log \varrho_i \big) \, \mathrm{d}\gamma_i \Big) + \int_X \ell\big(\mathsf{d}(x_1, x_2)\big) \, \mathrm{d}\boldsymbol{\gamma}$$

$$= \sum_i \big(\mu_i(X) - \gamma_i(X) \big) - \int_X \log \big(\varrho_1(x_1)\varrho_2(x_2) \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \big) \, \mathrm{d}\boldsymbol{\gamma}$$

$$\leq \sum_i \mu_i(X) - 2\boldsymbol{\gamma}(X).$$

By (6.11) we get $M' \geq \mathsf{LET}(\mu_1, \mu_2)$. On the other hand, choosing any $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ the optimality condition (6.17) shows that $\bar{\boldsymbol{\gamma}}$ is an admissible competitor for (6.24) and (6.22) shows that $M' = \mathsf{LET}(\mu_1, \mu_2)$. □

Combining (6.12), (6.13), (6.14), and (6.24), we find that the nonnegative and concave functional $(\mu_1, \mu_2) \mapsto \sum_i \mu_i(X) - \mathsf{LET}(\mu_1, \mu_2)$ can be represented as in the following equivalent ways:

$$\sum_i \mu_i(X) - \mathsf{LET}(\mu_1, \mu_2) = 2 \max_{\gamma \in \mathcal{M}(X)} \int_X \sqrt{\varrho_1(x_1)\varrho_2(x_2)} \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \, \mathrm{d}\gamma \tag{6.25}$$

$$= \inf \left\{ \sum_i \int_X \mathrm{e}^{-\varphi_i} \, \mathrm{d}\mu_i : \varphi_i \in \mathrm{LSC}_s(X), \ \varphi_1 \oplus \varphi_2 \leq \ell(\mathsf{d}) \right\} \tag{6.26}$$

$$= \inf \left\{ \sum_i \int_X \tilde{\psi}_i \, \mathrm{d}\mu_i : \tilde{\psi}_i \in \mathrm{USC}_s(X), \ \inf_X \tilde{\psi}_i > 0, \right.$$

$$\left. \psi_1(x_1)\psi_2(x_2) \geq \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)) \text{ in } X \right\} \tag{6.27}$$

$$= 2 \max \left\{ \gamma(X) : \gamma \in \mathcal{M}(X), \ \gamma_i = \sigma_i \mu_i, \right.$$

$$\left. \sigma_1(x_1)\sigma_2(x_2) \leq \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2))\gamma \text{-a.e. in} X \right\}. \tag{6.28}$$

The next result concerns uniqueness of the optimal plan γ 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 6.6 (Uniqueness) *Let $\mu_i \in \mathcal{M}(X)$ and $\gamma \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$.*

(i) *The marginals $\gamma_i = \pi_\sharp^i \gamma$ are uniquely determined.*

(ii) *If $X = \mathbb{R}$ with the usual distance then γ 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 μ_1-a.e. point of A_1 and γ is the unique element of $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$. The transport plan γ is concentrated on the graph of a function $t : \mathbb{R}^d \to \mathbb{R}^d$ satisfying*

$$t(x_1) = x_1 - \frac{\arctan(|\boldsymbol{\xi}(x_1)|)}{|\boldsymbol{\xi}(x_1)|} \boldsymbol{\xi}(x_1),$$

$$\boldsymbol{\xi}(x_1) = -\frac{1}{2}\tilde{\mathrm{D}} \log \sigma_1(x_1) \tag{6.29}$$

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 $\mathbb{R}$ (see e.g. [42, Chap. 2]).

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); notice that since $\sigma_1 > 0$ in A_1 the classes of μ_1- and γ_1-negligible subsets of A_1 coincide.

Since $\mu_1 = u\mathscr{L}^d \ll \mathscr{L}^d$ with density $u \in L^1(\mathbb{R}^d)$, 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 μ_1-negligible subsets of A_1 coincide. For every $n \in \mathbb{N}$ we define

$$A_{2,n} := \{x_2 \in A_2 : \sigma_2(x_2) \geq 1/n\}, \quad s_n(x_1) := \sup_{x_2 \in A_{2,n}} \cos^2(|x_1 - x_2|)/\sigma_2(x_2).$$
(6.30)

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

By (6.16) we have $\sigma_1(x_1) \geq s_n(x_1)$ in A_1. By (6.17) we know 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 and hence the family $(B_n)_{n \in \mathbb{N}}$ of sets $B_n := \{x_1 \in A_1 : \sigma_1(x_1) > s_n(x_1)\}$ is 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' the set of all $x_1 \in A_1$ such that σ_1 is approximately differentiable at every $x_1 \in A_1'$ with approximate differential $\tilde{D}\sigma_1(x_1)$ equal to $Ds_n(x_1)$ 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 by (6.16) and (6.17) 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

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

so that x_2 is uniquely determined, and (6.29) follows. □

We conclude this section with the last representation formula for $\mathsf{LET}(\mu_1, \mu_2)$ given in terms of transport plans $\boldsymbol{\alpha}$ in $\boldsymbol{Y} := Y \times Y$ with $Y := X \times [0, \infty)$ with constraints on the homogeneous marginals, keeping the notation of Sect. 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, and it will play a crucial role in Sect. 7.6, where the link between the LET formulation and the Hellinger–Kantorovich distance will be studied. The interest of (6.34) relies in the particular form of its integrand, by recalling that by (5.4) and (6.9) we have

$$\mathsf{LET}(r_1^2 \delta_{x_1}, r_2^2 \delta(x_2)) = r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)).$$
(6.31)

Theorem 6.7 *For every $\mu_i \in \mathcal{M}(X)$ we have*

$$\mathsf{LET}(\mu_1, \mu_2) = \sum_i \mu_i(X) - 2 \max_{\boldsymbol{\alpha} \in \mathcal{H}^2_{\leq}(\mu_1, \mu_2)} \int_X r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \, \mathrm{d}\boldsymbol{\alpha} \tag{6.32}$$

$$= \min \left\{ \int_Y \left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \right) \mathrm{d}\boldsymbol{\alpha} \right.$$

$$\left. + \sum_i (\mu_i - \mathsf{h}_i^2(\boldsymbol{\alpha}))(X) : \boldsymbol{\alpha} \in \mathcal{M}(Y), \; \mathsf{h}_i^2(\boldsymbol{\alpha}) \leq \mu_i \right\} \tag{6.33}$$

$$= \min \left\{ \int_Y \left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \right) \mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}(Y), \; \mathsf{h}_i^2(\boldsymbol{\alpha}) = \mu_i \right\} \tag{6.34}$$

Moreover, for every plan $\bar{\boldsymbol{\gamma}} \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ and every pair 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 (6.33) and (6.32).

Proof Identity (6.33) (resp. (6.34)) follows directly by (5.32) (resp. (5.33)) of Theorem 5.8. Relation (6.32) is just a different form for (6.33). □

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

In this section we want to show that the functional

$$(\mu_1, \mu_2) \mapsto \sqrt{\mathsf{LET}(\mu_1, \mu_2)} \tag{7.1}$$

defines a distance in $\mathcal{M}(X)$, which is then called the Hellinger–Kantorovich distance and denoted HK.

In order to introduce this distance we will adopt a geometric point of view, which is strictly related to the characterization given in Theorem 6.7: it will mainly exploit the link with Optimal Transport in the so-called *geometric cone* $\mathfrak{C}$ constructed on X, cf. [10, Sect. 3.6]. This is possible since the function

$$(x_1, r_1; x_2, r_2) \mapsto r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}(x_1, x_2) \wedge a), \quad a > 0, \tag{7.2}$$

appearing in (6.31) and (6.34) with $a = \pi/2$, is a (possibly extended) squared semidistance in $Y = X \times [0, \infty)$, whenever $a \in (0, \pi]$.

In the next two sections we will briefly study this function and the associated metric for the particular choices of $a = \pi$ (the canonical one in metric geometry) and $a = \pi/2$ (related to the minimal cost between a pair of Dirac masses (6.31)): the role of these two values will be clarified in Remark 7.2 and in Sect. 7.6, we also refer to [30, Sect. 3] for more motivation and examples. The

induced metric space $\mathfrak{C}$, can be obtained by taking the quotient with respect to the equivalence classes of points with distance 0. Radon measures on $\mathfrak{C}$ can be projected to Radon measures on X by taking suitable *homogeneous marginals*, which will be studied in Sect. 7.2.

The definition and the basic properties of the Hellinger–Kantorovich distance will be given in Sect. 7.3; the main metric properties will be derived in Sects. 7.4 and 7.5: they rely on a refined gluing technique and on the flexibility of the notion of homogeneous marginals, which allow us to transfer many useful properties of the Kantorovich–Wasserstein distance on the cone $\mathfrak{C}$ to corresponding properties for HK. Section 7.6 will then show the equivalent characterization of HK in terms of the Logarithmic Entropy Transport problem LET and its dual formulation, thus providing a direct and robust formulation of HK as a convex minimization problem enjoying all the properties we recalled in the previous section.

7.1 The cone construction

Let us quickly recall a few basic facts concerning the cone construction, referring to [10, Sect. 3.6] for further details. In the extended metric-topological space (X, τ, d) of Sect. 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 $\mathsf{d}_{\mathfrak{C}} : Y \times Y \to [0, \infty)$

$$\mathsf{d}_{\mathfrak{C}}^2((x_1, r_1), (x_2, r_2)) := r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_\pi(x_1, x_2)) \qquad (7.3)$$

is nonnegative, symmetric, and satisfies the triangle inequality (see e.g. [10, Prop. 3.6.13]). We also notice that

$$\mathsf{d}_{\mathfrak{C}}^2(y_1, y_2) = |r_1 - r_2|^2 + 4r_1 r_2 \sin^2\left(\mathsf{d}_\pi(x_1, x_2)/2\right), \qquad (7.4)$$

which implies the useful estimates

$$\max\left(|r_1 - r_2|, \frac{2}{\pi}\sqrt{r_1 r_2}\, \mathsf{d}_\pi(x_1, x_2)\right) \le \mathsf{d}_{\mathfrak{C}}(y_1, y_2)$$
$$\le |r_1 - r_2| + \sqrt{r_1 r_2}\, \mathsf{d}_\pi(x_1, x_2). \qquad (7.5)$$

From this it follows that $\mathsf{d}_{\mathfrak{C}}$ induces a true distance in the quotient space $\mathfrak{C} = Y/\sim$ where

$$y_1 \sim y_2 \quad \Leftrightarrow \quad r_1 = r_2 = 0 \ \text{ or } \ r_1 = r_2, \ x_1 = x_2. \qquad (7.6)$$

Equivalence classes are usually denoted by $\eta = [y] = [x, r]$, where the vertex $[x, 0]$ plays a distinguished role. 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

$$\{[x, r] : 0 \leq r < \varepsilon\} = \{\eta \in \mathfrak{C} : \mathsf{d}_{\mathfrak{C}}(\eta, \mathfrak{o}) < \varepsilon\}, \quad \varepsilon > 0, \quad (7.7)$$

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 a 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 denote the canonical projection by

$$\mathfrak{p} : Y \to \mathfrak{C}, \quad \mathfrak{p}(x, r) = [x, r]. \quad (7.8)$$

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

$$\mathsf{r} : \mathfrak{C} \to [0, \infty), \quad \mathsf{r}[x, r] = r,$$

$$\mathsf{x} : \mathfrak{C} \to X, \quad \mathsf{x}[x, r] = \begin{cases} x & \text{if } r > 0, \\ \bar{x} & \text{if } r = 0, \end{cases} \quad \text{and} \quad \mathsf{y} := (\mathsf{x}, \mathsf{r}). \quad (7.9)$$

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

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

$$\eta \cdot \lambda := \begin{cases} \mathfrak{o} & \text{if } \eta = \mathfrak{o}, \\ [x, \lambda r] & \text{if } \eta = [x, r], \ s > 0. \end{cases} \quad (7.10)$$

We conclude this introductory section by a characterization of compact sets in $(\mathfrak{C}, \tau_{\mathfrak{C}})$.

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*

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

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 $\rho = \inf_K \mathsf{r}$. If $\rho > 0$ then K is compact since it is a closed subset of the compact set $\mathfrak{p}\big(K(\rho) \times [\rho, r_0]\big)$.

If $\rho = 0$ then $\mathfrak{o}$ is an accumulation point of K by (7.7) and therefore $\mathfrak{o} \in K$ since K is closed. If $\mathscr{U}$ is an open covering of K, we can pick $U_0 \in \mathscr{U}$ such that $\mathfrak{o} \in U_0$. By (7.7) there exists $\varepsilon > 0$ such that $K \setminus U_0 \subset \mathfrak{p}\big(K(\varepsilon) \times [\varepsilon, r_0]\big)$: since $\mathfrak{p}\big(K(\varepsilon) \times [\varepsilon, r_0]\big)$ is compact, we can thus find a finite subcover $\{U_1, \cdots, U_N\} \subset \mathscr{U}$ of $K \setminus U_0$. $\{U_n\}_{n=0}^N$ is therefore a finite subcover of K. $\square$

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 6.7 an analogous formula with $\mathsf{d}_{\pi/2}$ and threshold $\pi/2$ played a crucial role. We could then consider the distance

$$\mathsf{d}_{\pi/2,\mathfrak{C}}^2([x_1, r_1], [x_2, r_2]) := r_1^2 + r_2^2 - 2r_1 r_2 \cos(\mathsf{d}_{\pi/2}(x_1, x_2)) \quad (7.11a)$$

$$= |r_1 - r_2|^2 + 4r_1 r_2 \sin^2(\mathsf{d}_{\pi/2}(x_1, x_2)/2) \quad (7.11b)$$

on $\mathfrak{C}$, which satisfies

$$\mathsf{d}_{\pi/2,\mathfrak{C}} \le \mathsf{d}_{\mathfrak{C}} \le \sqrt{2}\,\mathsf{d}_{\pi/2,\mathfrak{C}}. \quad (7.12)$$

The notation (7.11a) 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 natural, since it preserves important metric properties concerning geodesics (see [10, Thm. 3.6.17] and the next Sect. 8.1) and curvature (see [10, Sect. 4.7] and the next Sect. 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 uses d_π instead of $\mathsf{d}_{\pi/2}$. Nevertheless, we will prove in Sect. 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 (7.9).

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

$$\int_{\mathfrak{C}} r^2 \, d\nu = \int_{\mathfrak{C}} d_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o}) \, d\nu = \int_Y r^2 \, d\bar{\nu} < \infty, \quad \bar{\nu} = y_\sharp \nu. \tag{7.13}$$

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 $d_{\mathfrak{C}}$, which justifies the index 2 in $\mathcal{M}_2(\mathfrak{C})$. Notice moreover that the measure $r^2 \bar{\nu}$ does not charge $X \times \{0\}$ and it is independent of the choice of the point $\bar{x}$ in (7.9).

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 $\mathfrak{y} = (\mathfrak{y}_i)_{i=1}^N = ([x_i, r_i])_{i=1}^N$ a point in $\mathfrak{C}^{\otimes N}$ and we will set $r_i(\mathfrak{y}) = r(\mathfrak{y}_i) = r_i$, $x_i(\mathfrak{y}) = x(\mathfrak{y}_i) \in X$. Projections on the i-coordinate from $\mathfrak{C}^{\otimes N}$ to $\mathfrak{C}$ are usually denoted by π^i or $\pi^{\mathfrak{y}_i}$, $\mathfrak{p} = \mathfrak{p}^{\otimes N} : (Y)^{\otimes N} \to \mathfrak{C}^{\otimes N}$, $\mathbf{y} = y^{\otimes N} : \mathfrak{C}^{\otimes N} \to (Y)^{\otimes N}$ are the Cartesian products of the projections and of the lifts.

Recall that the L^2-Kantorovich–Wasserstein (extended) distance $W_{d_{\mathfrak{C}}}$ in $\mathcal{M}_2(\mathfrak{C})$ induced by $d_{\mathfrak{C}}$ is defined by

$$W_{d_{\mathfrak{C}}}^2(\nu_1, \nu_2) := \min \left\{ \int d_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) \, d\alpha : \alpha \in \mathcal{M}(\mathfrak{C}), \ \pi_\sharp^{\mathfrak{y}_i} \alpha = \nu_i \right\}, \tag{7.14}$$

with the convention that $W_{d_{\mathfrak{C}}}(\nu_1, \nu_2) = +\infty$ if $\nu_1(\mathfrak{C}) \neq \nu_2(\mathfrak{C})$ and thus the minimum in (7.14) is taken on an empty set. We want to mimic the above definition, replacing the usual marginal conditions in (7.14) with the homogeneous marginals $\mathfrak{h}_i^2$ which we are going to define.

Let us consider now a plan α in $\mathcal{M}(\mathfrak{C}^{\otimes N})$ with $\bar{\alpha} = \mathbf{y}_\sharp \alpha \in \mathcal{M}(Y^{\otimes N})$: we say that α lies in $\mathcal{M}_2(\mathfrak{C}^{\otimes N})$ if

$$\int_{\mathfrak{C}^{\otimes N}} \sum_i r_i^2 \, d\alpha = \int_{Y^{\otimes N}} \sum_i r_i^2 \, d\bar{\alpha} < \infty. \tag{7.15}$$

Its "canonical" marginals in $\mathcal{M}(\mathfrak{C})$ are $\alpha_i = \pi_\sharp^{\mathfrak{y}_i} \alpha$, whereas the "homogeneous" marginals correspond to (5.24) with $p = 2$:

$$\mathfrak{h}_i^2(\alpha) := (x_i)_\sharp (r_i^2 \alpha) = \pi_\sharp^{x_i} (r_i^2 \bar{\alpha}) = h_i^2(\bar{\alpha}) \in \mathcal{M}(X), \quad \bar{\alpha} := \mathbf{y}_\sharp \alpha. \tag{7.16}$$

We will omit the index i when $N = 1$. Notice that $r_i^2 \alpha$ does not charge $(\pi^i)^{-1}(\mathfrak{o})$ (similarly, $r_i^2 \bar{\alpha}$ does not charge $Y^{\otimes i-1} \times \{(\bar{x}, 0)\} \times Y^{\otimes N-i}$) so that (7.16) is independent of the choice of the point $\bar{x}$ in (7.9).

As in (5.26), the homogeneous marginals on the cone are invariant with respect to dilations: if $\vartheta : \mathfrak{C}^{\otimes N} \to (0, \infty)$ is a Borel map in $L^2(\mathfrak{C}^{\otimes N}, \boldsymbol{\alpha})$ we set

$$\left(\mathrm{prd}_\vartheta(\mathfrak{y})\right)_i := \mathfrak{y}_i \cdot \left(\vartheta(\mathfrak{y})\right)^{-1} \quad \text{and} \quad \mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha}) := (\mathrm{prd}_\vartheta)_\sharp(\vartheta^2 \boldsymbol{\alpha}), \quad (7.17)$$

so that

$$\mathfrak{h}_i^2(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})) = \mathfrak{h}_i^2(\boldsymbol{\alpha}) \quad \text{for every } \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N}). \quad (7.18)$$

As for the canonical marginals, a uniform control of the homogeneous marginals is sufficient to get equal tightness, cf. (2.4) for the definition. We state this result for an arbitrary number of components, and we emphasize 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 $\mathcal{M}(X)$. Then, the set*

$$\left\{ \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^N) : \mathfrak{h}_i^2 \boldsymbol{\alpha} \in \mathcal{K}_i \text{ for } i = 1, \ldots, N \right\} \quad (7.19)$$

is equally tight in $\mathcal{M}(\mathfrak{C}^N)$.

Proof By applying [2, Lem. 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 $\mathcal{H} := \left\{ \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}) : \mathfrak{h}^2 \boldsymbol{\alpha} \in \mathcal{K} \right\}$ is equally tight. For $A \subset X$, $R \subset (0, \infty)$ we will use the short notation $A \times_{\mathfrak{C}} R$ for $\mathfrak{p}(A \times R) \subset \mathfrak{C}$. If A and R are compact, then $A \times_{\mathfrak{C}} R$ is compact in $\mathfrak{C}$.

Let $M := \sup_{\mu \in \mathcal{K}} \mu(X) < \infty$; since $\mathcal{K}$ is equally 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

$$\mathfrak{K}_m = \{\mathfrak{o}\} \cup (K_m \times_{\mathfrak{C}} [2^{-m}, 2^m]) \cup \left(\bigcup_{n=1}^{\infty} K_{n+m} \times_{\mathfrak{C}} [2^{-n}, 2^{-n+1}] \right). \quad (7.20)$$

Setting $K_\infty = \bigcup_{n=1}^{\infty} K_n$, we have $\mu(X \setminus K_\infty) = 0$ and

$$\mathfrak{C} \setminus \mathfrak{K}_m \subset \left(K_m \times_{\mathfrak{C}} (2^m, \infty) \right) \cup \left(\bigcup_{n=1}^{\infty} (K_{n+m} \setminus K_{n+m-1}) \times_{\mathfrak{C}} (2^{-n+1}, \infty) \right)$$
$$\cup \left((X \setminus K_\infty) \times_{\mathfrak{C}} (0, \infty) \right).$$

Since for every $\boldsymbol{\alpha} \in \mathcal{H}$ with $\mathfrak{h}^2 \boldsymbol{\alpha} = \mu$ and every $A \in \mathcal{B}(X)$ we have

$$\boldsymbol{\alpha}(A \times_{\mathfrak{C}} (s, \infty)) \le s^{-2} \mu(A) \le s^{-2} M \quad \text{and} \quad \boldsymbol{\alpha}\left((X \setminus K_\infty) \times_{\mathfrak{C}} (0, \infty) \right) = 0,$$

we conclude

$$
\begin{aligned}
\boldsymbol{\alpha}(\mathfrak{C} \setminus \mathfrak{K}_m) &\le M\,4^{-m} + \sum_{n=1}^{\infty} \boldsymbol{\alpha}\big((X \setminus K_{n+m-1}) \times_{\mathfrak{C}} (2^{-n+1}, \infty)\big) \\
&\le M\,4^{-m} + \sum_{n=1}^{\infty} 4^{n-1} 8^{1-n-m} \le 4^{-m}\Big(M + \sum_{n=1}^{\infty} 4^{-n}\Big) \\
&\le 4^{-m}\big(1 + M\big),
\end{aligned}
$$

for every $\boldsymbol{\alpha} \in \mathcal{H}$. Since all $\mathfrak{K}_m$ are compact, we obtain the desired equal tightness. $\qquad\square$

7.3 The Hellinger–Kantorovich problem

In this section we will always consider $N = 2$, keeping the shorter notation $Y = Y^{\otimes 2}$ and $\mathfrak{C} = \mathfrak{C}^{\otimes 2}$. As in (5.28), for every $\mu_1, \mu_2 \in \mathcal{M}_2(X)$ we define the sets

$$
\begin{aligned}
\mathfrak{H}^2_{\le}(\mu_1, \mu_2) &:= \Big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}) \ : \ \mathfrak{h}^2_i \boldsymbol{\alpha} \le \mu_i\Big\} \text{ and} \\
\mathfrak{H}^2_{=}(\mu_1, \mu_2) &:= \Big\{\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}) \ : \ \mathfrak{h}^2_i \boldsymbol{\alpha} = \mu_i\Big\}.
\end{aligned}
\tag{7.21}
$$

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 (5.29). The condition $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\le}(\mu_1, \mu_2)$ is equivalent to ask that

$$
\int r_i^2 \varphi(\mathsf{x}_i)\,\mathrm{d}\boldsymbol{\alpha} \le \int \varphi\,\mathrm{d}\mu_i \quad \text{for every nonnegative } \varphi \in \mathrm{B}_b(X). \tag{7.22}
$$

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 $\boldsymbol{\alpha}_{\mathrm{opt}} \in \mathcal{H}^2_{=}(\mu_1, \mu_2) \subset \mathcal{M}_2(\mathfrak{C})$ solving the minimum problem*

$$
\mathsf{HK}(\mu_1, \mu_2)^2 := \min\Big\{\int \mathrm{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}),\ \mathfrak{h}^2_i \boldsymbol{\alpha} = \mu_i\Big\}. \tag{7.23}
$$

We denote by $\mathrm{Opt}_{\mathsf{HK}}(\mu_1, \mu_2) \subset \mathcal{M}(\mathfrak{C})$ the collection of all the optimal plans $\boldsymbol{\alpha}$ realizing the minimum in (7.23) and by $\mathsf{HK}^2(\mu_1, \mu_2)$ the value of the minimum in (7.23) (whose existence is guaranteed by the next Theorem 7.6).

Remark 7.5 (Lifting of plans in Y) Since any plan $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ 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

$$\min\left\{\int d_{\mathfrak{C}}^2(y_1, y_2)\, d\bar{\boldsymbol{\alpha}} : \bar{\boldsymbol{\alpha}} \in \mathcal{M}(Y \times Y), \quad \mathsf{h}_i^2(\bar{\boldsymbol{\alpha}}) = \mu_i\right\}. \qquad (7.24)$$

The advantage to work in the quotient space $\mathfrak{C}$ is to gain compactness, as the next Theorem 7.6 will show.

An importance feature of the cone distance and the homogeneous marginals is an invariance under **rescaling**, which can be done by the dilations from (7.17). Let us set

$$\mathfrak{C}[R] := \left\{[x, r] \in \mathfrak{C} : r \le R\right\} \text{ and } \mathbf{\mathfrak{C}}[R] := \mathfrak{C}[R] \times \mathfrak{C}[R]. \qquad (7.25)$$

It is not restrictive to solve the previous problem 7.4 by also assuming that $\boldsymbol{\alpha}$ is a probability plan in $\mathcal{P}(\mathbf{\mathfrak{C}})$ concentrated on $\mathbf{\mathfrak{C}}[R]$ with $R^2 = \sum_i \mu_i(X)$, i.e.

$$\mathsf{HK}^2(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha} \in C} \int d_{\mathfrak{C}}^2\, d\boldsymbol{\alpha},$$

$$\text{where } C := \left\{\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{\mathfrak{C}}) : \mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i,\ \boldsymbol{\alpha}\left(\mathbf{\mathfrak{C}} \setminus \mathbf{\mathfrak{C}}[R]\right) = 0\right\}. \quad (7.26)$$

In fact the functional $d_{\mathfrak{C}}^2$ and the constraints have a natural scaling invariance induced by the dilation maps defined by (7.17). Since

$$\int d_{\mathfrak{C}}^2\, d(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})) = \int \vartheta^2 d_{\mathfrak{C}}^2([x_1, r_1/\vartheta]; [x_2, r_2/\vartheta])\, d\boldsymbol{\alpha} = \int d_{\mathfrak{C}}^2\, d\boldsymbol{\alpha}, \qquad (7.27)$$

restricting first $\boldsymbol{\alpha}$ to $\mathbf{\mathfrak{C}} \setminus \{(\mathfrak{o}, \mathfrak{o})\}$ and then choosing ϑ as in (5.27a) with $p = 2$ we obtain a probability plan $\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha} \llcorner \mathbf{\mathfrak{C}} \setminus \{(\mathfrak{o}, \mathfrak{o})\})$ in $\mathcal{H}_{\underline{=}}^2(\mu_1, \mu_2)$ concentrated in $\mathbf{\mathfrak{C}}[R] \setminus \{(\mathfrak{o}, \mathfrak{o})\}$ with the same cost $\int d_{\mathfrak{C}}^2\, d\boldsymbol{\alpha}$. In order to show that Problem 7.4 has a solution we can then use the formulation (7.26) and prove that the set C where the minimum will be found is narrowly compact in $\mathcal{P}(\mathbf{\mathfrak{C}})$. Notice that the analogous property would not be true in $\mathcal{P}(Y \times Y)$ (unless X is compact) since the collection of measures concentrated in $(X \times \{0\}) \times (X \times \{0\})$ would not be equally tight. 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 $\mathcal{P}(\mathbf{\mathfrak{C}})$, as in (7.23).

Theorem 7.6 (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* $\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{\mathfrak{C}})$ *concentrated on* $\mathbf{\mathfrak{C}}[R] \setminus \{(\mathfrak{o}, \mathfrak{o})\}$ *with* $R^2 = \sum_i \mu_i(X)$.

Proof By the rescaling (7.27) it is not restrictive to look for minimizers $\boldsymbol{\alpha}$ of (7.26). Since $\mathbf{\mathfrak{C}}[R]$ is closed in $\mathbf{\mathfrak{C}}$ and the maps r_i^2 are continuous and bounded in $\mathbf{\mathfrak{C}}[R]$, C is clearly narrowly closed. By Lemma 7.3, C is also equally tight in $\mathcal{P}(\mathbf{\mathfrak{C}})$, thus narrowly compact by Theorem 2.2. Since the $d_{\mathfrak{C}}^2$ is

lower semicontinuous in $\mathfrak{C}$, the existence of a minimizer of (7.26) then follows by the direct method of the calculus of variations. □

We can also prove an interesting characterization of HK in terms of the Kantorovich–Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$ given by (7.14). An even deeper connection will be discussed in the next section, see Corollary 7.13.

Corollary 7.7 (HK and the Kantorovich–Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$) *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have (recall the notation $\mathfrak{h}$ explained after (7.16))*

$$\mathsf{HK}(\mu_1, \mu_2) = \min \left\{ \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) : \alpha_i \in \mathcal{P}_2(\mathfrak{C}), \quad \mathfrak{h}^2 \alpha_i = \mu_i \right\}, \qquad (7.28)$$

and there exist optimal measures $\bar{\alpha}_i$ for (7.28) 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.

$$\mathsf{HK}(\mathfrak{h}^2 \alpha_1, \mathfrak{h}^2 \alpha_2) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) \quad \text{for every } \alpha_i \in \mathcal{P}_2(\mathfrak{C}). \qquad (7.29)$$

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 (7.14) 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 (7.23) and $\alpha_i := \pi^i_{\sharp} \boldsymbol{\alpha} \in \mathcal{P}_2(\mathfrak{C})$ 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 (7.28). □

We conclude this section with two simple properties of the HK functional. We denote by η_0 the null measure.

Lemma 7.8 (Subadditivity of HK^2) *The functional HK^2 satisfies*

$$\mathsf{HK}^2(\mu, \eta_0) = \mu(X), \quad \mathsf{HK}^2(\mu_1, \mu_2) \le \mu_1(X) + \mu_2(X) \qquad (7.30)$$

for every $\mu, \mu_i \in \mathcal{M}(X)$, and it is subadditive, i.e. for every $\mu_i, \mu'_i \in \mathcal{M}(X)$ we have

$$\mathsf{HK}^2(\mu_1 + \mu'_1, \mu_2 + \mu'_2) \le \mathsf{HK}^2(\mu_1, \mu_2) + \mathsf{HK}^2(\mu'_1, \mu'_2). \qquad (7.31)$$

Proof The relations in (7.30) are 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 (7.31). □

Subsequently we will use the symbol "$\llcorner$" for the restriction of measures.

Lemma 7.9 (A formulation with relaxed constraints) *For every $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}^2(\mu_1, \mu_2)$$

$$= \min_{\boldsymbol{\alpha} \in \mathfrak{H}^2_{\leq}(\mu_1, \mu_2)} \left\{ \int \mathsf{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) \, \mathrm{d}\boldsymbol{\alpha} + \sum_i (\mu_i - \mathfrak{h}^2_i \boldsymbol{\alpha})(X) \right\} \qquad (7.32a)$$

$$= \mu_1(X) + \mu_2(X) - \max_{\boldsymbol{\alpha} \in \mathfrak{H}^2_{\leq}(\mu_1, \mu_2)} \left\{ 2 \int \mathsf{r}_1 \, \mathsf{r}_2 \cos(\mathsf{d}_\pi (\mathsf{x}_1, \mathsf{x}_2)) \, \mathrm{d}\boldsymbol{\alpha} \right\}. \tag{7.32b}$$

Moreover,

(i) Equations (7.32a)–(7.32b) *share the same class of optimal plans.*

(ii) *A plan* $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\leq}(\mu_1, \mu_2)$ *is optimal for* (7.32a)–(7.32b) *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* (7.32a)–(7.32b) *with* $\mu'_i := \mu_i - \mathfrak{h}^2_i \boldsymbol{\alpha}$, *then* $\tilde{\boldsymbol{\alpha}} := \boldsymbol{\alpha} + \boldsymbol{\alpha}'$ *is an optimal plan in* $\mathrm{Opt}_{\mathsf{HK}}(\mu_1, \mu_2)$ *for all* $\boldsymbol{\alpha}' \in \mathfrak{H}^2_{\equiv}(\mu'_1, \mu'_2)$.

(iv) *A plan* $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\equiv}(\mu_1, \mu_2)$ *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* (7.32a)–(7.32b).

Proof The formulas (7.32a) and (7.32b) are just two different ways to write the same functional, since for every $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\leq}(\mu_1, \mu_2)$ we have

$$\int \mathsf{d}^2_{\mathfrak{C}} \, \mathrm{d}\boldsymbol{\alpha} + \sum_i (\mu_i - \mathfrak{h}^2_i \boldsymbol{\alpha})(X) = \sum_i \mu_i(X) - 2 \int \mathsf{r}_1 \, \mathsf{r}_2 \cos(\mathsf{d}_\pi (\mathsf{x}_1, \mathsf{x}_2)) \, \mathrm{d}\boldsymbol{\alpha}. \tag{7.33}$$

Thus, to prove (i) it is sufficient to show (7.32a). The inequality $\geq$ is obvious, since $\mathfrak{H}^2_{\leq}(\mu_1, \mu_2) \supset \mathfrak{H}^2_{\equiv}(\mu_1, \mu_2)$ and for every $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\equiv}(\mu_1, \mu_2)$ the term $\sum_i (\mu_i - \mathfrak{h}^2_i \boldsymbol{\alpha})(X)$ vanishes.

On the other hand, whenever $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\leq}(\mu_1, \mu_2)$, setting $\mu''_i := \mathfrak{h}^2_i \boldsymbol{\alpha} \in \mathcal{M}(X)$, $\mu'_i := \mu_i - \mu''_i$ and observing that $\boldsymbol{\alpha} \in \mathfrak{H}^2_{\equiv}(\mu''_1, \mu''_2)$ we get

$$\int \mathsf{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) \, \mathrm{d}\boldsymbol{\alpha} + \sum_i (\mu_i - \mathfrak{h}^2_i \boldsymbol{\alpha})(X) \geq \mathsf{HK}^2(\mu''_1, \mu''_2) + \mu'_1(X) + \mu'_2(X)$$

$$\overset{(7.30)}{\geq} \mathsf{HK}^2(\mu'_1, \mu'_2) + \mathsf{HK}^2(\mu''_1, \mu''_2) \overset{(7.31)}{\geq} \mathsf{HK}^2(\mu_1, \mu_2).$$

The same calculations also prove point (iii).

In order to check (ii) it is sufficient to observe that the integrand in (7.32b) 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 (7.23), then by the consideration above it is optimal for (7.32b) and therefore (ii) shows that $\boldsymbol{\alpha}_\mathfrak{o}$ is optimal as well. The converse implication follows by (iii). □

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 $\mathcal{M}(X)$. As in Optimal Transport (see e.g. [2, Sect. 7.1]), the triangle inequality can be obtained by a gluing technique that allows us to join a couple of optimal transport plans with a common marginal. Here we will deal with transport plans on the cone $\mathfrak{C}$ and homogeneous marginals. We will also consider a more general situation where a sequence of measures is involved: it will turn out to be extremely useful in various topological (Theorem 7.15) and metric (Theorems 7.17, 8.4, 8.6, 8.8) results.

The main technical step is provided by the following property for plans in $\mathcal{M}(\mathfrak{C}^{\otimes N})$ with given homogeneous marginals, which is a simple application of the rescaling invariance in (7.27). This property is nontrivial since homogeneous marginals are considerably less rigid than the usual marginals and therefore the gluing technique requires a preliminary normalization, which does not affect the computation of the HK distance.

Lemma 7.10 (Normalization of lifts) *Let* $\boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$, $N \geq 2$, *be a plan satisfying*

$$\mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i \in \mathcal{M}(X) \text{ for } i = 1, ..., N,$$
$$\text{and } a_i = \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i) \, \mathsf{d}\boldsymbol{\alpha} \text{ for } i = 2, ..., N, \tag{7.34}$$

and let $j \in \{1, \ldots, N\}$ *be fixed. Then, it is possible to find a new plan* $\bar{\boldsymbol{\alpha}} \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$ *which still satisfies* (7.34) *and additionally the normalization of the jth lift,*

$$\pi_\sharp^j(\bar{\boldsymbol{\alpha}}) = \delta_{\mathfrak{o}} + \mathfrak{p}_\sharp(\mu_j \otimes \delta_1). \tag{7.35}$$

Proof By possibly adding $\otimes^N \delta_{\mathfrak{o}}$ to $\boldsymbol{\alpha}$ (which does not modify (7.34)) we may suppose that

$$\omega_j := \boldsymbol{\alpha}\big(\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \pi^j(\mathfrak{y}) = \mathfrak{o}\}\big) \geq 1,$$

where j is fixed as in the lemma. In order to find $\bar{\boldsymbol{\alpha}}$ it is sufficient to rescale $\boldsymbol{\alpha}$ by the function

$$\vartheta(\mathfrak{y}) := \begin{cases} \mathsf{r}_j(\mathfrak{y}) & \text{if } \mathfrak{y}_j \neq \mathfrak{o}, \\ \omega_j^{-1/2} & \text{otherwise.} \end{cases} \tag{7.36}$$

With the notation of (7.17) 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

$$\int \zeta(\mathfrak{y}_j) \, d\bar{\boldsymbol{\alpha}} = \int \zeta(\mathfrak{y}_j \cdot \vartheta^{-1}(\mathfrak{y})) \vartheta^2(\mathfrak{y}) \, d\boldsymbol{\alpha}$$

$$= \int \zeta(\mathfrak{o}) \omega_j^{-1} \, d\boldsymbol{\alpha}' + \int \zeta([x_j, r_j/\vartheta(\mathfrak{y})]) \vartheta^2(\mathfrak{y}) \, d\boldsymbol{\alpha}''$$

$$= \zeta(\mathfrak{o}) + \int \zeta([x_j, 1]) r_j^2 d\boldsymbol{\alpha}'' = \zeta(\mathfrak{o}) + \int \zeta([x_j, 1]) r_j^2 d\boldsymbol{\alpha}$$

$$= \zeta(\mathfrak{o}) + \int \zeta \circ \mathfrak{p} \, d(\mu_j \otimes \delta_1)$$

which yields (7.35). $\qquad \square$

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 (see e.g. [2, Lem. 5.3.4]). Here its formulation is strongly related to the rescaling invariance of optimal plans given by Lemma 7.10.

Lemma 7.11 (Gluing lemma) *Let us consider a finite collection of measures $\mu_i \in \mathcal{M}(X)$ for $i = 1, \ldots, N$ with $N \geq 2$. Set*

$$\Theta := \sqrt{\mu_1(X)} + \sum_{i=2}^{N} \mathsf{HK}(\mu_{i-1}, \mu_i) \ \text{ and } \ M^2 := \sum_{i=1}^{N} \mu_i(X). \qquad (7.37)$$

Then there exist plans $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ such that

$$\mathfrak{h}_i^2 \boldsymbol{\alpha}_k = \mu_i \ \text{ for } i = 1, \ldots, N \ \text{ and}$$

$$\int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i) \, d\boldsymbol{\alpha}_k = \mathsf{HK}^2(\mu_{i-1}, \mu_i) \ \text{ for } i = 2, \ldots, N. \quad (7.38)$$

Moreover, the plans $\boldsymbol{\alpha}_k$ satisfy the following additional conditions:

$$\boldsymbol{\alpha}_1 \ \text{is concentrated on } \Big\{ \mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sum_i r_i^2(\mathfrak{y}) \leq M^2 \Big\}, \qquad (7.39)$$

$$\boldsymbol{\alpha}_2 \ \text{is concentrated on } \Big\{ \mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sup_i r_i(\mathfrak{y}) \leq \Theta \Big\} = \big(\mathfrak{C}[\Theta] \big)^{\otimes N}. \quad (7.40)$$

Proof We first construct a plan $\boldsymbol{\alpha}$ satisfying (7.38), then suitable rescalings will provide $\boldsymbol{\alpha}_k$ satisfying (7.39) or (7.40). 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 7.6. Let us now 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 (7.38)

hold, and another optimal plan $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_N, \mu_{N+1})$ for the pair μ_N, μ_{N+1}. Applying the normalization Lemma 7.10 to $\boldsymbol{\alpha}^N$ (with $j = N$) and to $\boldsymbol{\alpha}$ (with $j = 1$) we can assume that

$$\pi_{\sharp}^N(\boldsymbol{\alpha}^N) = \delta_{\mathfrak{o}} + \mathfrak{p}_{\sharp}(\mu_N \otimes \delta_1) = \pi_{\sharp}^1(\boldsymbol{\alpha}).$$

Therefore we can apply the standard gluing Lemma in $\left(\prod_{i=1}^{N-1} \mathfrak{C}_i\right), \mathfrak{C}_N, \mathfrak{C}_{N+1}$ (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 $\pi_{\sharp}^{1,2,\cdots,N} \boldsymbol{\alpha}^{N+1} = \boldsymbol{\alpha}^N$ and $\pi^{N,N+1} \boldsymbol{\alpha}^{N+1} = \boldsymbol{\alpha}$. In particular, $\boldsymbol{\alpha}^{N+1}$ satisfies (7.38).

A further application of the rescaling (7.27) with ϑ as in (5.27a) yields a plan $\boldsymbol{\alpha}_1$ satisfying also (7.39).

In order to obtain $\boldsymbol{\alpha}_2$, we can assume $\boldsymbol{\alpha}(\{|\mathfrak{y}| = 0\}) = 0$ and set $\boldsymbol{\alpha}_2 = \mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})$, where we use the rescaling function

$$\vartheta(\mathfrak{y}) := r^{-1}|\mathfrak{y}|_{\infty} = r^{-1} \sup_i r_i(\mathfrak{y}) \quad \text{with} \quad r^2 := \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_{\infty}^2 \, \mathrm{d}\boldsymbol{\alpha}.$$

To obtain (7.40) it remains to estimate r. We consider arbitrary coefficients $\theta_i > 0$ and use for $n = 2, \ldots, N$ the inequality

$$r_n \leq r_1 + \sum_{i=2}^n |r_i - r_{i-1}| \leq \left(\sum_{i=1}^n \theta_i^{-1}\right)^{1/2} \left(\theta_1 r_1^2 + \sum_{i=2}^n \theta_i |r_i - r_{i-1}|^2\right)^{1/2}$$
$$\leq \left(\sum_{i=1}^N \theta_i^{-1}\right)^{1/2} \left(\theta_1 r_1^2 + \sum_{i=2}^N \theta_i \mathrm{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_{i-1})\right)^{1/2},$$

which yields

$$\sup_i r_i(\mathfrak{y}) \leq \left(\sum_{i=1}^N \theta_i^{-1}\right)^{1/2} \left(\theta_1 r_1^2 + \sum_{i=2}^N \theta_i \mathrm{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_{i-1})\right)^{1/2},$$

and therefore

$$r^2 = \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_{\infty}^2 \, \mathrm{d}\boldsymbol{\alpha} \leq \left(\sum_{i=1}^N \theta_i^{-1}\right) \int_{\mathfrak{C}^{\otimes N}} \left(\theta_1 r_1^2 + \sum_{i=2}^N \theta_i \mathrm{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_{i-1})\right) \mathrm{d}\boldsymbol{\alpha}$$
$$= \left(\sum_{i=1}^N \theta_i^{-1}\right) \cdot \left(\theta_1 \mu_1(X) + \sum_{i=2}^N \theta_i \, \mathsf{HK}^2(\mu_{i-1}, \mu_i)\right);$$

optimizing with respect to $\theta_i > 0$ we obtain the value of Θ given by (7.37). $\square$

The next remark gives a similar rescaling result for probability couplings $\boldsymbol{\beta} \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$.

Remark 7.12 In a completely similar way (see [2, Lemma 5.3.4]), for every $N \geq 2$ and every finite collection of measures $\mu_i \in \mathcal{M}(X)$, there exists a plan $\boldsymbol{\beta} \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ concentrated on $\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : \sup_i r_i(\mathfrak{y}) \leq \Xi\}$ with

$$\Xi := \sqrt{\mu_1(X)} + \sum_{i=2}^{N} \mathsf{HK}(\mu_1, \mu_i), \qquad (7.41)$$

such that

$$\mathfrak{h}_i^2 \boldsymbol{\beta} = \mu_i \quad \text{and} \quad \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_i) \, \mathrm{d}\boldsymbol{\beta} = \mathsf{HK}^2(\mu_1, \mu_i) \qquad (7.42)$$

for $i = 1, \ldots, N$. $\square$

Arguing as in the proof of Corollary 7.7 one immediately obtains the following result, which will be needed for the proof of Theorem 8.8 and for the subsequent corollary.

Corollary 7.13 *For every finite collection of measures* $\mu_i \in \mathcal{M}(X)$, $i = 1, \ldots, N$, *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* (7.37) *and* β_i *concentrated in* $\mathfrak{C}[\Xi]$ *given by* (7.41) *such that*

$$\mathfrak{h}^2 \alpha_i = \mu_i \text{ and } \quad \mathfrak{h}^2 \beta_i = \mu_i \text{ for } i = 1, \ldots, N,$$
$$\mathsf{HK}(\mu_1, \mu_i) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\beta_1, \beta_i) \text{ and}$$
$$\mathsf{HK}(\mu_i, \mu_{i+1}) = \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_i, \alpha_{i+1}) \text{ for } i = 2, \ldots, N.$$

We are now in the position to show that the functional HK is a true distance on $\mathcal{M}(X)$, where we deduce the triangle inequality from that for $\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}$ by using normalized lifts.

Corollary 7.14 (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 the triangle inequality*

$$\mathsf{HK}(\mu_1, \mu_3) \leq \mathsf{HK}(\mu_1, \mu_2) + \mathsf{HK}(\mu_2, \mu_3). \qquad (7.43)$$

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 (7.43) it is sufficient to apply the previous corollary 7.13 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

$$\mathsf{HK}(\mu_1, \mu_3) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_3) \le \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_1, \alpha_2) + \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_2, \alpha_3)$$
$$= \mathsf{HK}(\mu_1, \mu_2) + \mathsf{HK}(\mu_2, \mu_3).$$

$\square$

As a consequence of the previous two 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 on X is induced by d and that (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, $\mathcal{M}(X)$ equipped with the weak topology 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.

$$\lim_{n \to \infty} \int_{\mathfrak{C}} \varphi \, \mathrm{d}\alpha_n = \int_{\mathfrak{C}} \varphi \, \mathrm{d}\alpha \quad \text{for every } \varphi \in \mathsf{C}(\mathfrak{C}) \text{ with } |\varphi(\mathfrak{y})| \le A + B \mathsf{r}^2(\mathfrak{y}),$$
(7.44)

for some constants $A, B \ge 0$ depending on φ. Recall that $\mathsf{r}^2(\mathfrak{y}) = \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o})$.

Theorem 7.15 (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 $\zeta \in \mathsf{C}_b(X)$ and a subsequence (still denoted by μ_n) such that

$$\inf_n \left| \int_X \zeta \, \mathrm{d}\mu_n - \int_X \zeta \, \mathrm{d}\mu \right| > 0.$$
(7.45)

The first estimate of (7.30) and the triangle inequality show that

$$\limsup_{n\to\infty} \mu_n(X) \le \limsup_{n\to\infty} \left(\mathsf{HK}(\mu_n, \mu) + \mathsf{HK}(\mu, \eta_0)\right)^2 = \mu(X),$$

so that $\sup_n \mu_n(X) = M^2 < \infty$. By Corollary 7.7 we can find measures $\alpha_n, \alpha'_n \in \mathcal{P}_2(\mathfrak{C})$ concentrated on $\mathfrak{C}[2M]$ 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 $\mathfrak{C}[2M]$ 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 η_0, then $\lim_{n\to\infty} \mu_n(X) = 0$ so that $\lim_{n\to\infty} \mathsf{HK}(\mu_n, \mu) = 0$ by (7.30).

So we can suppose that $m := \mu(X) > 0$ and have $m_n := \mu_n(X) \ge m/2 > 0$ for sufficiently large n. We now consider the measures $\alpha_n, \alpha \in \mathcal{P}(\mathfrak{C})$ given by

$$\alpha_n := \mathfrak{p}_\sharp\left(m_n^{-1}\mu_n \otimes \delta_{\sqrt{m_n}}\right) \text{ and } \alpha := \mathfrak{p}_\sharp\left(m^{-1}\mu \otimes \delta_{\sqrt{m}}\right).$$

Since $\mathfrak{h}^2\alpha_n = \mu_n$ and $\mathfrak{h}^2\alpha = \mu$, by (7.29) 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 $\mathfrak{p}$. Hence, in order to conclude that $\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_n, \alpha) \to 0$ it is now sufficient to prove the convergence of their quadratic moments with respect to the vertex $\mathfrak{o}$. However, this is is immediate because of

$$\lim_{n\to\infty} \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o})\, \mathrm{d}\alpha_n = \lim_{n\to\infty} \int \mathsf{r}^2\, \mathrm{d}\alpha_n = \lim_{n\to\infty} m_n = m = \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{o})\, \mathrm{d}\alpha.$$

$\square$

Corollary 7.16 (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 $\sup_{\mu \in \mathcal{C}} \mu(X) < \infty$. Then the classical weak sequential compactness of closed bounded sets in $\mathcal{M}(X)$ gives the result. □

The following completeness result for $(\mathcal{M}(X), \mathsf{HK})$ is obtained by suitable liftings of measures μ_i to probability measures $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$, supported in some $\mathfrak{C}[\Theta]$. Then the completeness of the Kantorovich–Wasserstein space $(\mathcal{P}_2(\mathfrak{C}), \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}})$ is exploited.

Theorem 7.17 (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 $\mathsf{HK}(\mu_m, \mu_{m'}) \leq 2^{-k}$ whenever $m, m' \geq n(k)$ and we consider the subsequence $\mu_i' := \mu_{n(i)}$, so that

$$\sqrt{\mu_1(X)} + \sum_{i=2}^{N} \mathsf{HK}(\mu_{n(i)}, \mu_{n(i-1)}) \leq \sqrt{\mu_1(X)} + 1.$$

By applying the Gluing Lemma 7.11, for every $N > 0$ we can find measures $\alpha_i^N \in \mathcal{P}_2(\mathfrak{C})$, $i = 1, \ldots, N$, concentrated on $\mathfrak{C}[\Theta]$ with $\Theta := \sqrt{\mu_1(X)} + 1$, 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 equally tight by Lemma 7.3 and 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

$$\mathfrak{h}^2 \alpha_i = \mu_i, \quad \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\alpha_i, \alpha_{i-1}) = \mathsf{HK}(\mu_i', \mu_{i-1}') \leq 2^{i-1}.$$

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 [2, Prop. 7.1.5] and therefore there exists $\alpha \in \mathcal{P}_2(\mathfrak{C})$ 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 for HK with the *Bounded Lipschitz* metric (cf. [19, Sect. 11.3]), see also [27, Thm. 3], and the flat metric. The Bounded Lipschitz metric is defined via

$$\mathsf{BL}(\mu_1, \mu_2) := \sup \Big\{ \int \xi \, \mathrm{d}(\mu_1 - \mu_2) : \xi \in \mathrm{Lip}_b(X), \ \sup_X |\xi| + \mathrm{Lip}(\xi, X) \le 1 \Big\},$$

(7.46)

and it is metrically equivalent to the flat metric

$$\widetilde{\mathsf{BL}}(\mu_1, \mu_2) := \sup \Big\{ \int \xi \, \mathrm{d}(\mu_1 - \mu_2) : \ \xi \in \mathrm{Lip}_b(X), \ \sup_X |\xi| \vee \mathrm{Lip}(\xi, X) \le 1 \Big\},$$

(7.47)

in the sense that $\mathsf{BL}(\mu_1, \mu_2) \le \widetilde{\mathsf{BL}}(\mu_1, \mu_2) \le 2\mathsf{BL}(\mu_1, \mu_2)$. In its turn, $\widetilde{\mathsf{BL}}(\mu_1, \mu_2)$ coincides with the Piccoli-Rossi distance we considered in Example E.9 of Sect. 3.3, see [39]. We do not claim that the constant C_* below is optimal.

Proposition 7.18 *For every* $\mu_1, \mu_2 \in \mathcal{M}(X)$ *we have*

$$\widetilde{\mathsf{BL}}(\mu_1, \mu_2) \le C_* \Big(\sum_i \mu_i(X) \Big)^{1/2} \mathsf{HK}(\mu_1, \mu_2), \ \ where \ C_* := \sqrt{2 + \pi^2/2}.$$

(7.48)

Proof Let $\xi \in \mathrm{Lip}_b(X)$ with $\sup_X |\xi| \le 1$ and $\mathrm{Lip}(\xi, X) \le 1$, and let $\boldsymbol{\alpha} \in \mathcal{P}(\mathfrak{C})$ optimal for (7.26) and concentrated on $\mathfrak{C}[R]$ with $R^2 := \mu_1(X_1) + \mu_2(X_2)$. Notice that

$$|\xi(x_1) - \xi(x_2)| \le \min(\mathsf{d}(x_1, x_2), 2) \le 2\mathsf{d}_2(x_1, x_2) \le 2\mathsf{d}_\pi(x_1, x_2)$$
$$\le 2\pi \sin(\mathsf{d}_\pi(x_1, x_2)/2).$$

We consider the function $\zeta : \mathfrak{C} \to \mathbb{R}$ defined by $\zeta(\mathfrak{y}) := \xi(\mathsf{x})\mathsf{r}^2$. Hence, ζ satisfies

$$\Big| \zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2) \Big| \le |\xi(\mathsf{x}_1) - \xi(\mathsf{x}_2)|\mathsf{r}_1\mathsf{r}_2 + \big(|\xi(\mathsf{x}_1)|\mathsf{r}_1 + |\xi(\mathsf{x}_2)|\mathsf{r}_2\big)|\mathsf{r}_1 - \mathsf{r}_2|$$
$$\le 2\pi \sin(\mathsf{d}_\pi(\mathsf{x}_1, \mathsf{x}_2)/2)\mathsf{r}_1\mathsf{r}_2 + (\mathsf{r}_1 + \mathsf{r}_2)|\mathsf{r}_1 - \mathsf{r}_2|$$
$$\overset{(7.4)}{\le} \sqrt{(\mathsf{r}_1 + \mathsf{r}_2)^2 + \pi^2\mathsf{r}_1\mathsf{r}_2} \ \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)$$
$$\le C_* \sqrt{\mathsf{r}_1^2 + \mathsf{r}_2^2} \ \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)$$

Since the optimal plan $\boldsymbol{\alpha}$ is concentrated on $\{\mathsf{r}_1^2 + \mathsf{r}_2^2 \le R^2\}$ we obtain

$$\left| \int_X \xi \, d(\mu_1 - \mu_2) \right| = \left| \int \zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2) \, d\boldsymbol{\alpha} \right| \le \int |\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)| \, d\boldsymbol{\alpha}$$

$$\le C_* R \int d_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) \, d\boldsymbol{\alpha} \le C_* R \, \mathsf{HK}(\mu_1, \mu_2).$$

$\square$

7.6 Hellinger–Kantorovich distance and Entropy-Transport functionals

In this section we will establish our main result connecting the Hellinger–Kantorovich Problem 7.4 defining HK with the Logarithmic Entropy-Transport Problem 6.1 defining LET.

It is clear that the definition of HK does not change if we replace the distance d on 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 $\mathfrak{C}$); in [30, Sect. 3.2] it is is called the *presence of a sufficiently large reservoir*, which shows that transport over distances larger than $\pi/2$ is never optimal, since it is cheaper to transport into or out of the reservoir in $\mathfrak{o}$. This will provide an essential piece of information to connect the HK and the LET functionals.

In order to prove that transport only occurs of distances $\le \pi/2$ we define the subset

$$\mathfrak{C}' := \left\{ \mathsf{d}_{\pi/2,\mathfrak{c}} < \mathsf{d}_{\mathfrak{C}} \right\} = \left\{ (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o} : \mathsf{d}(x_1, x_2) > \pi/2 \right\} \quad (7.49)$$

and consider the partition $(\mathfrak{C}', \mathfrak{C}'')$ of $\mathfrak{C} = \mathfrak{C} \times \mathfrak{C}$, where $\mathfrak{C}'' := \mathfrak{C} \setminus \mathfrak{C}' = \left\{ \mathsf{d}_{\pi/2,\mathfrak{c}} = \mathsf{d}_{\mathfrak{C}} \right\}$. Observe that

$$\mathfrak{C}''_\mathfrak{o} := \mathfrak{C}'' \cap (\mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o}) = \left\{ (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o} : \mathsf{d}(x_1, x_2) \le \pi/2 \right\}. \quad (7.50)$$

In the following lemma we show that minimizers $\boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_1, \mu_2)$ are concentrated on $\mathfrak{C}''$, i.e. $\boldsymbol{\alpha}(\mathfrak{C}') = 0$ which holds if and only if $\boldsymbol{\alpha}_\mathfrak{o} = \boldsymbol{\alpha} \llcorner (\mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o})$ is concentrated on $\mathfrak{C}''_\mathfrak{o}$. To handle the mass that is transported into or out of $\mathfrak{o}$, we use the continuous projections

$$\mathfrak{g}_i : \mathfrak{C} \to \mathfrak{C}, \quad \mathfrak{g}_1(\mathfrak{y}_1, \mathfrak{y}_2) := (\mathfrak{y}_1, \mathfrak{o}), \quad \mathfrak{g}_2(\mathfrak{y}_1, \mathfrak{y}_2) := (\mathfrak{o}, \mathfrak{y}_2). \quad (7.51)$$

Lemma 7.19 (Plan restriction) *For every* $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ *the plan*

$$\hat{\boldsymbol{\alpha}} := \boldsymbol{\alpha}'' + (\mathfrak{g}_1)_\sharp \boldsymbol{\alpha}' + (\mathfrak{g}_2)_\sharp \boldsymbol{\alpha}' \quad \text{with} \quad \boldsymbol{\alpha}' := \boldsymbol{\alpha} \llcorner \mathfrak{C}', \quad \boldsymbol{\alpha}'' := \boldsymbol{\alpha} \llcorner \mathfrak{C}'', \quad (7.52)$$

is concentrated on $\mathfrak{C}''$, has the same homogeneous marginals as $\boldsymbol{\alpha}$, i.e. $\mathfrak{h}_i^2 \hat{\boldsymbol{\alpha}} = \mathfrak{h}_i^2 \boldsymbol{\alpha}$, and

$$\int_{\mathfrak{C}} \mathsf{d}_{\mathfrak{C}}^2 \, d\hat{\boldsymbol{\alpha}} = \int_{\mathfrak{C}} \mathsf{d}_{\pi/2,\mathfrak{C}}^2 \, d\hat{\boldsymbol{\alpha}} \le \int_{\mathfrak{C}} \mathsf{d}_{\mathfrak{C}}^2 \, d\boldsymbol{\alpha}, \tag{7.53}$$

where the inequality is strict if $\boldsymbol{\alpha}(\mathfrak{C}') > 0$. In particular for every $\mu_1, \mu_2 \in \mathcal{M}(X)$

$$\mathsf{HK}^2(\mu_1, \mu_2) = \min \left\{ \int \mathsf{d}_{\pi/2,\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) \, d\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathcal{M}_2(\mathfrak{C}), \; \mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i \right\}. \tag{7.54}$$

Proof For every $\zeta \in B_b(X)$, since $\mathsf{r}_1 \circ \mathfrak{g}_2 = 0$ and $\mathsf{r}_1 \circ \mathfrak{g}_1 = \mathsf{r}_1$, we have

$$\int \zeta \, d(\mathfrak{h}_1^2 \hat{\boldsymbol{\alpha}}) = \int \zeta(\mathsf{x}_1) \mathsf{r}_1^2 \, d\hat{\boldsymbol{\alpha}} = \int \zeta(\mathsf{x}_1) \mathsf{r}_1^2 \, d\boldsymbol{\alpha}'' + \sum_k \int \zeta(\mathsf{x}_1(\mathfrak{g}_k)) \mathsf{r}_1(\mathfrak{g}_k)^2 \, d\boldsymbol{\alpha}'$$

$$= \int \zeta(\mathsf{x}_1) \mathsf{r}_1^2 \, d\boldsymbol{\alpha}'' + \int \zeta(\mathsf{x}_1) \mathsf{r}_1^2 \, d\boldsymbol{\alpha}' = \int \zeta(\mathsf{x}_1) \mathsf{r}_1^2 \, d\boldsymbol{\alpha} = \int \zeta \, d(\mathfrak{h}_1^2 \boldsymbol{\alpha}),$$

so that $\mathfrak{h}_1^2 \hat{\boldsymbol{\alpha}} = \mathfrak{h}_1^2 \boldsymbol{\alpha}$; a similar calculation holds for $\mathfrak{h}_2^2$ so that $\hat{\boldsymbol{\alpha}} \in \mathfrak{H}_{=}^2(\mu_1, \mu_2)$. 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

$$\int \mathsf{d}_{\mathfrak{C}}^2 \, d\hat{\boldsymbol{\alpha}} = \int \left(\mathsf{d}_{\mathfrak{C}}^2 \circ \mathfrak{g}_1 + \mathsf{d}_{\mathfrak{C}}^2 \circ \mathfrak{g}_2 \right) d\boldsymbol{\alpha}' + \int \mathsf{d}_{\mathfrak{C}}^2 \, d\boldsymbol{\alpha}''$$

$$< \int \mathsf{d}_{\mathfrak{C}}^2 \, d\boldsymbol{\alpha}' + \int \mathsf{d}_{\mathfrak{C}}^2 \, d\boldsymbol{\alpha}'' = \int \mathsf{d}_{\mathfrak{C}}^2 \, d\boldsymbol{\alpha},$$

which proves (7.53) and characterizes the equality case. (7.54) then follows by (7.53) and the fact that the homogeneous marginals of $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$ coincide. $\square$

In (7.54) we have established that $\boldsymbol{\alpha} \in \mathsf{Opt}_{\mathsf{HK}}(\mu_1, \mu_2)$ has support in $\mathfrak{C}''$. This allows us to prove the identity $\mathsf{LET} = \mathsf{HK}^2$. For this, we introduce the open set $\mathfrak{G} \subset \mathfrak{C}''$ via

$$\mathfrak{G} := \left\{ ([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o} : \mathsf{d}(x_1, x_2) < \pi/2 \right\}$$

and note that $\mathsf{r}_1 \mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) > 0$ in $\mathfrak{G}$. Recall also $\mathfrak{p} = \mathfrak{p} \otimes \mathfrak{p} : Y \to \mathfrak{C}$, where $\mathfrak{p}$ is defined in (7.8).

Theorem 7.20 ($\mathsf{HK}^2 = \mathsf{LET}$) *For all* $\mu_1, \mu_2 \in \mathcal{M}(X)$ *we have*

$$\mathsf{HK}^2(\mu_1, \mu_2) = \mathsf{LET}(\mu_1, \mu_2), \tag{7.55}$$

and $\boldsymbol{\alpha}(\mathfrak{C}') = 0$ *for optimal solution* $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ *of* Problem 7.4 *or of* (7.32a, b). *Moreover,*

(i) $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ *is an optimal plan for* (7.32a, b) *if and only if* $\boldsymbol{\alpha}(\mathfrak{C}') = 0$ *and* $\mathsf{y}_\sharp(\boldsymbol{\alpha} \llcorner \mathfrak{C}_\mathfrak{o} \times \mathfrak{C}_\mathfrak{o})$ *is an optimal plan for* (6.33)–(6.32).

(ii) $\bar{\boldsymbol{\alpha}} \in \mathcal{M}(Y)$ *is any optimal plan for* (6.34) *if and only if the plan* $\hat{\boldsymbol{\alpha}}$ *obtained from* $\boldsymbol{\alpha} := \mathfrak{p}_\sharp \bar{\boldsymbol{\alpha}}$ *as in* (7.52) *is an optimal plan for the Hellinger–Kantorovich Problem* 7.4.

(iii) *In the case that* $\boldsymbol{\gamma} \in \mathcal{M}(X \times X)$ *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* $\boldsymbol{\beta} := \left(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2))\right)_\sharp \boldsymbol{\gamma}$ *is an optimal plan for* (7.32a,b), *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* $\boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C})$ *is an optimal plan for Problem* 7.4 *then* $\tilde{\boldsymbol{\alpha}} := \boldsymbol{\alpha} \llcorner \mathfrak{G}$ *is an optimal plan for* (7.32a,b). *Moreover,*
  - *the plan* $\boldsymbol{\beta} := \mathrm{dil}_{\vartheta,2}(\tilde{\boldsymbol{\alpha}})$, *with* $\vartheta := \left(\mathsf{r}_1 \mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2))\right)^{1/2}$, *is an optimal plan for* (7.32a,b) *satisfying* $\mathsf{r}_1 \mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) = 1$ $\boldsymbol{\beta}$-*a.e.*
  - *If* (X, τ) *is separable and metrizable,* $\boldsymbol{\gamma} := (\mathsf{x}_1, \mathsf{x}_2)_\sharp \boldsymbol{\beta}$ *belongs to* $\mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$,
  - *If* (X, τ) *is separable and metrizable,* $\boldsymbol{\beta} = \left(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2))\right)_\sharp \boldsymbol{\gamma}$.

Proof Identity (7.55) and the first statement immediately follow by combining the previous Lemma 7.19 with Remark 7.5 and (6.34). Claim (ii) follows as well.

In order to prove (i), we observe that if $\boldsymbol{\alpha}$ is an optimal plan for the formulation (7.32a,b) we can apply Lemma 7.9(iii) to find $\tilde{\boldsymbol{\alpha}} \geq \boldsymbol{\alpha}$ optimal for (7.23), so that $\boldsymbol{\alpha}(\mathfrak{C}') \leq \tilde{\boldsymbol{\alpha}}(\mathfrak{C}') = 0$. Given this property, (7.32a,b) correspond to (6.33)–(6.32).

(iii) is a consequence of Theorem 6.7 and of the optimality conditions (6.19), which show that $\boldsymbol{\beta}$ is concentrated on $\mathfrak{G}$ and satisfies $\mathsf{r}_1 \mathsf{r}_2 \cos(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)) = 1$ $\boldsymbol{\beta}$-a.e. Therefore, $\boldsymbol{\beta}$ is optimal for (7.32a,b) thanks to claim (i).

Concerning (iv), the optimality of $\tilde{\boldsymbol{\alpha}}$ is obvious from the formulation (7.32b) and the optimality of $\boldsymbol{\beta} = \mathrm{dil}_{\vartheta,2}(\tilde{\boldsymbol{\alpha}})$ follows from the invariance of (7.32b) with respect to dilations. We notice that $\boldsymbol{\beta}$-almost everywhere in $\mathfrak{G}$ we have

$$\sum_i U_0(\mathsf{r}_i^2) + \mathsf{c}(\mathsf{x}_1, \mathsf{x}_2) = \sum_i \left(\mathsf{r}_i^2 - 1 - \log \mathsf{r}_i^2\right) - \log(\cos^2(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)))$$

$$= \sum_i r_i^2 - 2 - 2\log(r_1 r_2 \cos(d_{\pi/2}(x_1, x_2)))$$

$$= r_1^2 + r_2^2 - 2r_1 r_2 \cos(d_{\pi/2}(x_1, x_2)),$$

so that by (7.32a) we arrive at

$$\int \left(\sum_i U_0(r_i^2) + c(x_1, x_2) \right) d\boldsymbol{\beta} + \sum_i \left(\mu_i(X) - \mathfrak{h}_i^2 \boldsymbol{\beta}(X) \right) = \mathsf{HK}^2(\mu_1, \mu_2).$$
(7.56)

Let us now set $\boldsymbol{\gamma} := (x_1, x_2)_\sharp \boldsymbol{\beta} \in \mathcal{M}(X \times X)$ and $\beta_i := \pi_\sharp^i \boldsymbol{\beta} \in \mathcal{M}(\mathfrak{C})$, which yield $\gamma_i := \pi_\sharp^i \boldsymbol{\gamma} = (x_i)_\sharp \boldsymbol{\beta} = x_\sharp \beta_i \in \mathcal{M}(X)$ and $\tilde{\mu}_i := \mathfrak{h}_i^2 \boldsymbol{\beta} = (x_i)_\sharp (r_i^2 \boldsymbol{\gamma}) = x_\sharp (r^2 \beta_i)$. Denoting by $(\beta_{i,x_i})_{x_i \in X}$ the disintegration of β_i with respect to γ_i (here we need the metrizability and separability of (X, τ), see [2, Sect. 5.3]), we find

$$\int_X \zeta \, d\tilde{\mu}_i = \int_{\mathfrak{C}} \zeta(x) r^2 \, d\beta_i = \int_X \left(\int_{\mathfrak{C}} \zeta(x) r^2 \, d\beta_{i,x} \right) d\gamma_i$$

$$= \int_X \zeta(x) \left(\int_{\mathfrak{C}} r^2 \, d\beta_{i,x} \right) d\gamma_i$$

for all $\zeta \in B_b(X)$, so that

$$\tilde{\mu}_i = \tilde{\varrho}_i \gamma_i \le \mu_i \quad \text{with} \quad \tilde{\varrho}_i(x) := \int_{\mathfrak{C}} r^2 \, d\beta_{i,x}.$$

Applying Jensen's inequality we obtain

$$\int U_0(r_i^2) \, d\boldsymbol{\beta} = \int U_0(r_i^2) \, d\beta_i = \int \left(\int U_0(r_i^2) \, d\beta_{i,x_i}(r_i) \right) d\gamma_i$$

$$\ge \int U_0 \left(\int r_i^2 \, d\beta_{i,x_i}(r_i) \right) d\gamma_i = \int U_0(\tilde{\varrho}_i(x)) \, d\gamma_i.$$

Now $\int c(x_1, x_2) \, d\boldsymbol{\beta} = \int c(x_1, x_2) \, d\boldsymbol{\gamma}$ and (7.56) imply

$$\mathsf{HK}^2(\mu_1, \mu_2) \ge \sum_i \int_X U_0(\tilde{\varrho}_i) \, d\gamma_i + \int_{X \times X} c \, d\boldsymbol{\gamma} + \sum_i v_i(X)$$

with $v_i := \mu_i - \tilde{\mu}_i \in \mathcal{M}(X)$. Hence, since $\mu_i = \tilde{\varrho}_i \gamma_i + v_i$ and the standard decomposition $\mu_i = \varrho_i \gamma_i + \mu_i^\perp$ (cf. (2.8)) give $v_i = \mu_i^\perp + (\varrho_i - \tilde{\varrho}_i) \gamma_i \ge \mu_i^\perp$,

$U_0(s) = s - 1 - \log s$ and the monotonicity of the logarithm yield

$$
\begin{aligned}
\mathsf{HK}^2(\mu_1, \mu_2) &\geq \sum_i \left(\int_X U_0(\tilde{\varrho}_i) \, d\gamma_i + \nu_i(X) \right) + \int c \, d\gamma \\
&= \sum_i \left(\int_X \left(U_0(\tilde{\varrho}_i) + \varrho_i - \tilde{\varrho}_i \right) d\gamma_i + \mu_i^\perp(X) \right) + \int c \, d\gamma \\
&\geq \sum_i \left(\int_X U_0(\varrho_i) \, d\gamma_i + \mu_i^\perp(X) \right) + \int c \, d\gamma \geq \mathsf{LET}(\mu_1, \mu_2),
\end{aligned}
$$

where the last estimate follows from Theorem 6.2(b). Above, the first inequality is strict if $\nu_i \neq \mu_i^\perp$ so that $\varrho_i > \tilde{\varrho}_i$ on some set with positive γ_i-measure.

By the first statement of the Theorem it follows that $\gamma \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$. Hence, all the inequalities are in fact identities, and we conclude $\tilde{\varrho}_i \equiv \varrho_i$. Since U_0 is strictly convex, the disintegration measure β_{i,x_i} is a Dirac measure concentrated on $\sqrt{\varrho_i(x_i)}$, so that $\boldsymbol{\beta} = \left(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2)) \right)_\sharp \gamma$. $\square$

We observe that the system $(\gamma, \varrho_1, \varrho_2)$ provided by the previous Theorem enjoys a few remarkable properties, that are not obvious from the original Hellinger–Kantorovich formulation.

(a) First of all, thanks to (6.15), 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 $\gamma \in \mathrm{Opt}_{\mathsf{LET}}(\mu_1, \mu_2)$ provides an optimal plan $\boldsymbol{\alpha} = \left(\mathfrak{p} \circ (x_1, \varrho_1^{1/2}(x_1); x_2, \varrho_2^{1/2}(x_2)) \right)_\sharp \gamma$ 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 $\mathbb{R}_+ \times \mathbb{R}_+$, where the maps ϱ_i are independent, in the sense that ϱ_i only depends on x_i.

(c) A third important application of Theorem 7.20 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 Sect. 8.4. It is based on the inf-convolution formula

$$
\begin{aligned}
\mathscr{P}_1 \xi(x) &= \inf_{x' \in X} \left(\frac{\xi(x')}{1 + 2\xi(x')} + \frac{\sin^2(\mathsf{d}_{\pi/2}(x, x'))}{2(1 + 2\xi(x'))} \right) \\
&= \inf_{x' \in X} \frac{1}{2} \left(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x, x'))}{1 + 2\xi(x')} \right). \tag{7.57}
\end{aligned}
$$

where $\xi \in B(X)$ with $\xi > -1/2$.

Theorem 7.21 (Duality formula for HK)

(i) *If $\xi \in B_b(X)$ with $\inf_X \xi > -1/2$ then the function $\mathscr{P}_1\xi$ defined by (7.57) belongs to $\mathrm{Lip}_b(X)$, satisfies $\sup_X \mathscr{P}_1\xi < 1/2$, and admits the equivalent representation*

$$\mathscr{P}_1\xi(x) = \inf_{x' \in B_{\pi/2}(x)} \frac{1}{2}\left(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x, x'))}{1 + 2\xi(x')}\right). \qquad (7.58)$$

In particular, if ξ has bounded support then $\mathscr{P}_1\xi \in \mathrm{Lip}_{bs}(X)$, 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*

$$\frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) = \sup\left\{\int \mathscr{P}_1\xi \, \mathrm{d}\mu_1 - \int \xi \, \mathrm{d}\mu_0 : \right.$$
$$\left. \xi \in \mathrm{Lip}_{bs}(X), \; \inf_X \xi > -1/2\right\}. \qquad (7.59)$$

Proof Let us first observe that if

$$-\frac{1}{2} < a \le \xi \le b \text{ in } X \quad \Rightarrow \quad \frac{a}{1 + 2a} \le \mathscr{P}_1\xi \le \frac{b}{1 + 2b} \text{ in } X, \qquad (7.60)$$

where the upper bound follows using $x' = x$, while the lower bound is easily seen from the first form of $\mathscr{P}_1\xi$ in (7.57) and $\sin^2 \ge 0$. 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.

Moreover we have the estimate

$$\frac{1}{2}\left(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x, x'))}{1 + 2\xi(x')}\right) = \frac{1}{2} > \frac{b}{1 + 2b} \quad \text{if } \mathsf{d}(x, x') \ge \pi/2, \qquad (7.61)$$

which immediately gives (7.58). In particular, we have

$$\xi \equiv 0 \text{ in } X \setminus B \quad \Rightarrow \quad \mathscr{P}_1\xi \equiv 0 \text{ in } \{x \in X : \mathsf{d}(x, B) \ge \pi/2\}. \qquad (7.62)$$

Let us now prove statement (ii). We denote by E the the right-hand side of (7.59) and by E' the analogous expression where ξ runs in $C_b(X)$:

$$E' := \sup\left\{\int \mathscr{P}_1\xi \, \mathrm{d}\mu_1 - \int \xi \, \mathrm{d}\mu_0 : \; \xi \in C_b(X), \; \inf_X \xi > -1/2\right\}. \qquad (7.63)$$

It is clear that $E' \geq E$. If $\xi \in C_b(X)$ 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

$$\big(1 - \psi_2(x_2)\big)\big(1 - \psi_1(x_1)\big) \geq \cos^2(\mathsf{d}_{\pi/2}(x_1, x_2)).$$

Hence, the pair (ψ_1, ψ_2) is admissible for (6.14) (with $C_b(X)$ instead of $\mathrm{LSC}_s(X)$; note that τ is metrizable and thus completely regular), so that $\mathsf{HK}^2(\mu_0, \mu_1) = \mathsf{LET}(\mu_0, \mu_1) \geq E'$.

On the other hand, if $(\psi_1, \psi_2) \in C_b(X) \times C_b(X)$ 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 $2\tilde{\xi}_2 \geq \psi_2$ giving $E' \geq \mathsf{HK}^2(\mu_0, \mu_1)$, so that $\mathsf{HK}^2(\mu_0, \mu_1) = E'$.

It remains to show that $E \geq E'$. We first approximate $\xi \in C_b(X)$ with $\inf_X \xi > -1/2$ by a decreasing sequence of Lipschitz and bounded functions (e.g. by taking $\xi_n(x) := \sup_y \xi(y) - n\mathsf{d}_\pi(x, y)$) pointwise converging to ξ, observing that $\mathscr{P}_1\xi_n$ is also decreasing, uniformly bounded and pointwise converging to $\mathscr{P}_1\xi$. We deduce that the supremum in (7.63) does not change if we restrict it to $\mathrm{Lip}_b(X)$.

In the last step of the proof we want to show that we can eventually restrict the supremum in (7.63) to $\mathrm{Lip}_{bs}(X)$, by a further approximation argument. We fix a Lipschitz function ξ valued in $[a, b]$ with $-1/2 < a \leq 0 \leq b$ and we consider the increasing sequence of nonnegative cut-off functions $\zeta_n(x) := 0 \vee \big(n - \mathsf{d}(x, \bar{x})\big) \wedge 1$: they 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 $\frac{a}{1+2a} \leq \mathscr{P}_1\xi_n \leq \frac{b}{1+2b}$ for every $n \in \mathbb{N}$ by (7.60).

Since $\xi_n(x) = 0$ if $\mathsf{d}(x, \bar{x}) \geq n$ and $\xi_n(x) = \xi(x)$ if $\mathsf{d}(x, \bar{x}) \leq n - 1$, by (7.58) we get

$$\begin{aligned} \mathscr{P}_1\xi_n(x) &= 0 \quad \text{if } \mathsf{d}(x, \bar{x}) \geq n + \pi/2, \\ \mathscr{P}_1\xi_n(x) &= \mathscr{P}_1\xi(x) \quad \text{if } \mathsf{d}(x, \bar{x}) < n - 1 - \pi/2. \end{aligned} \tag{7.64}$$

Thus $\mathscr{P}_1\xi_n \in \mathrm{Lip}_{bs}(X)$ and $\mathscr{P}_1\xi_n(x) \to \mathscr{P}_1\xi(x)$ for every $x \in X$ as $n \to \infty$. Applying the Lebesgue Dominated Convergence theorem we conclude that

$$\lim_{n\to\infty} \int_X \mathscr{P}_1\xi_n \, \mathrm{d}\mu_1 - \int_X \xi_n \, \mathrm{d}\mu_0 = \int_X \mathscr{P}_1\xi \, \mathrm{d}\mu_1 - \int_X \xi \, \mathrm{d}\mu_0.$$

$\square$

7.7 Limiting cases: recovering the Hellinger—Kakutani distance 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 Sect. 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

$$\mathsf{d}_{\mathsf{He}}(x_1, x_2) := \begin{cases} a & \text{if } x_1 \neq x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases} \quad \text{with } a \in [\pi, \infty]. \tag{7.65}$$

The induced cone distance in this case is

$$\mathsf{d}_{\mathfrak{C}}^2([x_1, r_1], [x_2, r_2]) = \begin{cases} (r_1 - r_2)^2 & \text{if } x_1 = x_2, \\ (r_1 + r_2)^2 & \text{if } x_1 \neq x_2. \end{cases} \tag{7.66}$$

and the induced cost function for the Entropy-Transport formalism is given by

$$\mathsf{c}_{\mathsf{He}}(x_1, x_2) := \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases} \tag{7.67}$$

Recalling (3.21)–(3.22) we obtain

$$\mathsf{He}^2(\mu_1, \mu_2) = \mathsf{LET}_{\mathsf{He}}(\mu_1, \mu_2) = \int_X \left(\sqrt{\varrho_1} - \sqrt{\varrho_2} \right)^2 \, d\gamma$$
$$\text{with } \mu_i = \varrho_i \gamma \ll \gamma \in \mathcal{M}(X). \tag{7.68}$$

Since $\mathsf{c}_{\mathsf{He}} \geq \mathsf{c} = \ell(\mathsf{d})$ for every distance function on X, we always have the upper bound

$$\mathsf{HK}(\mu_1, \mu_2) \leq \mathsf{He}(\mu_1, \mu_2) \quad \text{for every } \mu_1, \mu_2 \in \mathcal{M}(X). \tag{7.69}$$

Applying Lemma 3.9 we easily get

Theorem 7.22 (Convergence of HK to He) *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 pair $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}_{\lambda \mathsf{d}}(\mu_1, \mu_2) \uparrow \mathsf{He}(\mu_1, \mu_2) \quad \text{as } \lambda \uparrow \infty. \tag{7.70}$$

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 $\mathsf{W_d}$ (the upper bound is trivial when $\mu_1(X) \neq \mu_2(X)$, since in this case $\mathsf{W_d}(\mu_1, \mu_2) = +\infty$).

Proposition 7.23 *For every pair $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\mathsf{HK}(\mu_1, \mu_2) \leq \mathsf{W}_{\mathsf{d}_{\pi/2}}(\mu_1, \mu_2) \leq \mathsf{W_d}(\mu_1, \mu_2). \tag{7.71}$$

Proof It is not restrictive to assume that $\mathsf{W}^2_{\mathsf{d}_{\pi/2}}(\mu_1, \mu_2) = \int \mathsf{d}^2_{\pi/2}\gamma < \infty$ for an optimal plan γ with marginals μ_i. We then define the plan $\alpha := \mathfrak{s}_\sharp \gamma \in \mathcal{M}(\mathfrak{C} \times \mathfrak{C})$ where $\mathfrak{s}(x_1, x_2) := ([x_1, 1], [x_2, 1])$, so that $\mathfrak{h}_i^2 \alpha = \mu_i$. By using (7.54) and the identity $2 - 2\cos^2(d) = 4\sin^2(d/2)$ we obtain

$$\mathsf{HK}^2(\mu_1, \mu_2) \leq 4 \int_{\mathfrak{C}} \sin^2(\mathsf{d}_{\pi/2}(\mathsf{x}_1, \mathsf{x}_2)/2) \, d\alpha \leq \int_X \mathsf{d}^2_{\pi/2}(x_1, x_2) \, d\gamma$$
$$\leq \mathsf{W}^2_{\mathsf{d}_{\pi/2}}(\mu_1, \mu_2).$$

$\square$

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 7.24 (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 $\lambda^{-1}\mathsf{d}$ for $\lambda > 0$. Then, for all $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have*

$$\lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1, \mu_2) \uparrow \mathsf{W_d}(\mu_1, \mu_2) \quad as \ \lambda \uparrow \infty. \tag{7.72}$$

Proof Let us denote by $\mathsf{LET}_\lambda = \mathsf{HK}^2_{\mathsf{d}/\lambda}$ the optimal value of the LET-problem associated to d/λ. Since the Kantorovich–Wasserstein distance is invariant by the rescaling $\lambda\mathsf{W}_{\mathsf{d}/\lambda} = \mathsf{W_d}$, estimate (7.71) shows that $\lambda\mathsf{HK}_{\mathsf{d}/\lambda} \leq \mathsf{W_d}$.

As $x \mapsto \sin(x \wedge \pi/2)$ is concave in $[0, \infty)$, the function $x \mapsto \sin(x \wedge \pi/2)/x$ is decreasing in $[0, \infty)$, so that $\alpha \sin((d/\alpha) \wedge \pi/2) \leq \lambda \sin((d/\lambda) \wedge \pi/2)$ for every $d \geq 0$ and $0 < \alpha < \lambda$. Combining (7.54) with (7.11b) we see that the map $\lambda \mapsto \lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1, \mu_2)$ is nondecreasing.

It remains to prove that $L := \lim_{\lambda \to \infty} \lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1, \mu_2) = \sup_{\lambda \geq 1} \lambda\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1, \mu_2) \geq \mathsf{W_d}(\mu_1, \mu_2)$. For this, it is not restrictive to assume that L is finite.

Let γ_λ be an optimal plan for $\mathsf{HK}_{\mathsf{d}/\lambda}(\mu_1, \mu_2)$ with marginals $\gamma_{\lambda,i} = \pi_\sharp^i \gamma_\lambda$. We denote by $\mathscr{F}$ the entropy functionals associated to logarithmic entropy $U_1(s)$ and by $\mathscr{G}$ the entropy functionals associated to $I_1(s)$ as in Example E.3

of Sect. 3.3. Since the transport part of the LET-functional is associated to the costs

$$c_\lambda(x_1, x_2) = \lambda^2 \ell(d(x_1, x_2)/\lambda) \overset{(6.6)}{\geq} d^2(x_1, x_2),$$

we obtain the estimate

$$L^2 \geq \lambda^2 \mathsf{LET}_\lambda(\mu_1, \mu_2) \geq \sum_i \lambda^2 \mathscr{F}(\gamma_{\lambda,i}|\mu_i) + \int_X d^2(x_1, x_2) \, d\boldsymbol{\gamma}_\lambda. \quad (7.73)$$

Proposition 2.10 shows that the family of plans $(\boldsymbol{\gamma}_\lambda)_{\lambda \geq 1}$ is relatively compact with respect to narrow convergence in $\mathcal{M}(X \times X)$. Since $\lambda^2 F(s) \uparrow I_1(s)$, 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

$$\sum_i \mathscr{G}(\gamma_i|\mu_i) \leq L^2, \quad \text{which implies} \quad \gamma_i = \mu_i.$$

In particular, we conclude that $\mu_1(X) = \boldsymbol{\gamma}(X \times X) = \mu_2(X)$. Since d is lower semicontinuous, narrow convergence of $\boldsymbol{\gamma}_{\lambda(\alpha)}$ and (7.73) also yield

$$L^2 \geq \liminf_{\alpha \in \mathbb{A}} \int_X d^2(x_1, x_2) \, d\boldsymbol{\gamma}_{\lambda(\alpha)} \geq \int_X d^2(x_1, x_2) \, d\boldsymbol{\gamma} \geq W_d^2(\mu_1, \mu_2).$$

$$\square$$

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 $\sqrt{\ell}$:

$$g(z) := \arccos(e^{-z^2/2}), \quad \text{giving } g(0) = 0, g'(0) = 1, \ell(g(d)) = d^2. \quad (7.74)$$

Since $\sqrt{\ell}$ is a convex function, g is a concave increasing function in $[0, \infty)$ with $g(z) \leq z$ and $\lim_{z \to \infty} g(z) = \pi/2$.

It follows that $\mathsf{g} := g \circ d$ is a distance in X, inducing the same topology as d. We can now introduce a distance HK_g associated to g. The corresponding distance on $\mathfrak{C}$ is given by

$$\mathsf{g}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) := r_1^2 + r_2^2 - 2r_1 r_2 \exp(-d^2(x_1, x_2)/2). \quad (7.75)$$

From $g(z) \leq z$ we have $\mathsf{g}_{\mathfrak{C}} \leq \mathsf{d}_{\mathfrak{C}}$.

We can then apply Corollary 7.14, Theorems 7.15, 7.17, 7.20, and 6.3 to obtain the following result.

Theorem 7.25 (The Gaussian Hellinger–Kantorovich distance) *The functional*

$$\mathsf{GHK}^2(\mu_1, \mu_2) := \mathsf{HK}_\mathsf{g}^2(\mu_1, \mu_2)$$

$$= \min \left\{ \int \mathsf{g}_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) \, \mathsf{d}\boldsymbol{\alpha} \; : \; \boldsymbol{\alpha} \in \mathcal{M}(\mathfrak{C}), \; \mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i \right\} \qquad (7.76)$$

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

$$\mathsf{GHK}^2(\mu_1, \mu_2) = \min \left\{ \sum_i \mathscr{F}(\gamma_i | \mu_i) + \int_X \mathsf{d}^2(x_1, x_2) \, \mathsf{d}\boldsymbol{\gamma} \; : \; \boldsymbol{\gamma} \in \mathcal{M}(X) \right\}$$

$$= \sup \left\{ \sum_i \int \left(1 - \mathrm{e}^{-\varphi_i} \right) \mathsf{d}\mu_i \; : \; \varphi_1 \oplus \varphi_2 \leq \mathsf{d}^2 \right\}. \qquad (7.77)$$

We shall see in the next Sect. 8.2 that HK is the length distance induced by GHK if d is a length distance on X.

8 Dynamic interpretation of the Hellinger–Kantorovich distance

In this section we collect our main results concerning the dynamic interpretation of the Hellinger–Kantorovich distance: it reveals another deep connection with Optimal Transport problems, in particular as a natural generalization of the Benamou-Brenier [7] characterization of the Kantorovich–Wasserstein distance, see the next Sect. 8.4 and [30, Sect. 4], where a more direct approach has been adopted for the case $X = \mathbb{R}^d$.

In order to deal with arbitrary geodesic spaces X and to obtain other important results concerning general representation formulae for geodesics and absolutely continuous curves (Sect. 8.2), lower curvature bounds (Sect. 8.3), duality relations with subsolutions to Hamilton–Jacobi equations (Sects. 8.4 and 8.6), and contraction properties for diffusion semigroups (Sect. 8.7), we adopted here the point of view of dynamic plans (i.e. probability measures on continuous paths), which provide a powerful tool in Optimal Transport, cf. [2, Chap. 8]. It is not difficult to imagine that the natural objects to deal with the Hellinger–Kantorovich distance are dynamic plans in the cone $\mathfrak{C}$, so we will devote the next section to recall the basic metric properties of curves in $\mathfrak{C}$.

As in Sect. 7.5, in all this section we will suppose that (X, d) is a complete and separable (possibly extended) metric space and τ coincides with the topology induced by d. All the results admit a natural generalization to the framework of extended metric-topological spaces [1, Sect. 4].

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 $\mathsf{z} : I \to Z$ is absolutely continuous if there exists $m \in \mathrm{L}^1(I)$ such that

$$\mathsf{d}_Z(\mathsf{z}(t_0), \mathsf{z}(t_1)) \leq \int_{t_0}^{t_1} m(t) \, \mathrm{d}t \quad \text{whenever } t_0, t_1 \in I, \ t_0 < t_1. \qquad (8.1)$$

Its metric derivative $|\mathsf{z}'|_{\mathsf{d}_Z}$ (we will omit the index d_Z when the choice of the metric is clear from the context) is the Borel function defined by

$$|\mathsf{z}'|_{\mathsf{d}_Z}(t) := \limsup_{h \to 0} \frac{\mathsf{d}_Z(\mathsf{z}(t + h), \mathsf{z}(t))}{|h|} \qquad (8.2)$$

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 $\mathsf{z} : I \to Z$ with $|\mathsf{z}'| \in \mathrm{L}^p(I)$; 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].)

A curve $\mathsf{z} : [0, 1] \to Z$ is a (minimal, constant speed) geodesic if

$$\mathsf{d}_Z(\mathsf{z}(t_0), \mathsf{z}(t_1)) = |t_1 - t_0| \mathsf{d}_Z(\mathsf{z}(0), \mathsf{z}(1)) \quad \text{for every } t_0, t_1 \in [0, 1]. \qquad (8.3)$$

In particular z is Lipschitz and $|\mathsf{z}'| \equiv \mathsf{d}_Z(\mathsf{z}(t_0), \mathsf{z}(t_1))$ in $[0, 1]$. We denote by $\mathrm{Geo}(Z) \subset \mathrm{C}([0, 1]; Z)$ the closed subset of all the geodesics.

By using the fact that $\int_0^1 f^2 \, \mathrm{d}t \geq \left(\int_0^1 f \, \mathrm{d}t \right)^2$ with equality if and only if f is constant a.e. in $(0, 1)$, it is easy to check that a curve

$$\mathsf{z} \in \mathrm{AC}^2([0, 1]; Z) \text{ is a geodesic if and only if } \int_0^1 |\mathsf{z}'|_{\mathsf{d}_Z}^2 \, \mathrm{d}t \leq \mathsf{d}_Z^2(\mathsf{z}(0), \mathsf{z}(1));$$

$$(8.4)$$

notice that the opposite inequality in (8.4) is satisfied along any curve.

A metric space (Z, d_Z) is called a length (or intrinsic) space if the distance between arbitrary pairs of points can be obtained as the infimum of the length of the absolutely continuous curves connecting them; by a simple reparametrization technique (see e.g. [2, Lem. 1.1.4]), this property is equivalent to assume that for every pair of points $z_0, z_1 \in Z$ at finite distance and every $\kappa > 1$ there exists a Lipschitz curve $z_\kappa : [0, 1] \to Z$ such that

$$z(i) = z_i, \ i = 0, 1, \quad |z'|_{\mathsf{d}_Z}(t) \le \kappa \, \mathsf{d}_Z(z_0, z_1) \quad \text{for every } t \in [0, 1]. \quad (8.5)$$

(Z, d_Z) is called a geodesic (or strictly intrinsic) space if every pair of points z_0, z_1 at finite distance can be joined by a geodesic (for which (8.5) holds with $\kappa = 1$).

Geodesics in $\mathfrak{C}$. If (X, d) is a geodesic (resp. length) space, then also $\mathfrak{C}$ is a geodesic (resp. length) space, cf. [10, Sect. 3.6]. The geodesic connecting a point $\mathfrak{y} = [x, r]$ with $\mathfrak{o}$ is

$$\mathfrak{y}(t) = [x, tr] = \mathfrak{y} \cdot t \ \text{ for } t \in [0, 1]. \quad (8.6)$$

If $x_1, x_2 \in X$ with $\mathsf{d}(x_1, x_2) \ge \pi$, 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 $[0, \mathsf{d}(x_1, x_2)]$. To find the radius $r(t)$, we use the complex plane $\mathbb{C}$ and write the curve connecting $z_1 = r_1 \in \mathbb{C}$ to $z_2 = r_2 \exp(i \, \mathsf{d}(x_1, x_2)) \in \mathbb{C}$ in polar coordinates, namely

$$\begin{aligned}
\mathsf{z}(t) &= r(t) \exp(i \, \theta(t)), \\
r^2(t) &= (1-t)^2 r_1^2 + t^2 r_2^2 + 2t(1-t) r_1 r_2 \cos(\mathsf{d}(x_1, x_2)), \\
\cos(\theta(t)) &= \frac{(1-t) r_1 + t r_2 \cos(\mathsf{d}(x_1, x_2))}{r(t)}, \quad \theta(t) \in [0, \pi],
\end{aligned} \quad (8.7)$$

and then the geodesic curve and the distance in $\mathfrak{C}$ take the form

$$\mathfrak{y}(t) = [\mathsf{x}(\theta(t)), r(t)], \quad \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) = |z_2 - z_1|. \quad (8.8)$$

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 $r(t) := r(\mathfrak{y}(t))$ is a continuous curve with values in $[0, \infty)$. We can then consider the open set $O_r = r^{-1}\big((0, \infty)\big)$ and the map $\mathsf{x} : [0, 1] \to X$ defined by $\mathsf{x}(t) := \mathsf{x}(\mathfrak{y}(t))$, whose restriction to O_r is also continuous. Thus any continuous curve $\mathfrak{y} : I \to \mathfrak{C}$ can be lifted to a pair of maps $\mathsf{y} = \mathsf{y} \circ \mathfrak{y} = (\mathsf{x}, \mathsf{r}) : [0, 1] \to Y$ with r continuous and x

continuous on O_r and constant on its complement. Conversely, it is clear that starting from a pair $y = (x, r)$ as above, then $\mathfrak{y} = \mathfrak{p} \circ y$ is continuous in $\mathfrak{C}$. We thus introduce the set

$$\widetilde{C}([0, 1]; Y) := \big\{ y = (x, r) : [0, 1] \to Y : $$
$$ r \in C([0, 1]; \mathbb{R}_+), \ x_{|O_r} \text{ is continuous} \big\} \qquad (8.9)$$

and for $p \geq 1$ the analogous spaces

$$\widetilde{AC}^p([0, 1]; Y) := \Big\{ y = (x, r) : r \in AC^p([0, 1]; \mathbb{R}_+), $$
$$ x_{|O_r} \in AC_{loc}^p(O_r; X), \ r|x'| \in L^p(O_r) \Big\}. \qquad (8.10)$$

If $y = (x, r) \in \widetilde{AC}^p([0, 1]; Y)$ we define the Borel map $|y'| : [0, 1] \to \mathbb{R}_+$ by

$$|y'|^2(t) := |r'(t)|^2 + r^2(t)|x'|_d^2(t) \quad \text{if } t \in O_r, \quad |y'|(t) = 0 \text{ otherwise.} \quad (8.11)$$

For absolutely continuous curves the following characterization holds:

Lemma 8.1 *Let $\mathfrak{y} \in C([0, 1]; \mathfrak{C})$ be lifted to $y = y \circ \mathfrak{y} \in \widetilde{C}([0, 1]; Y)$. Then $\mathfrak{y} \in AC^p(I; \mathfrak{C})$ if and only if $y = (x, r) \in \widetilde{AC}^p([0, 1]; Y)$ and*

$$|\mathfrak{y}'|_{d_{\mathfrak{C}}}(t) = |y'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in [0, 1]. \qquad (8.12)$$

Proof By (7.5) one immediately sees that if $\mathfrak{y} = \mathfrak{p} \circ y \in AC^p([0, 1]; \mathfrak{C})$ then r belongs to $AC^p([0, 1]; \mathbb{R})$ and $x \in AC_{loc}^p(O_r; X)$. Since $\mathfrak{y}$ is absolutely continuous, we can evaluate the metric derivative at a.e. $t \in O_r$ where also r′ and $|x'|$ exist: starting from (7.4) leads to the limit

$$\lim_{h \downarrow 0} \frac{d_{\mathfrak{C}}^2(\mathfrak{y}(t + h), \mathfrak{y}(t))}{h^2}$$
$$= \lim_{h \downarrow 0} \frac{|r(t + h) - r(t)|^2 + 4r(t + h)r(t) \sin^2(\frac{1}{2}d_\pi(x(t + h), x(t)))}{h^2}$$
$$= |r'(t)|^2 + r^2(t)|x'|_d^2(t)$$

which provides (8.12).

Moreover, the same calculations show that if the lifting y belongs to $\widetilde{AC}^p([0, 1]; Y)$ then the restriction of $\mathfrak{y}$ to each connected component of O_r is absolutely continuous with metric velocity given by (8.12) in $L^p(0, 1)$. Since $\mathfrak{y}$ is globally continuous and constant in $[0, 1] \setminus O_r$, we conclude that $\mathfrak{y} \in AC^p([0, 1]; \mathfrak{C})$. $\qquad \square$

As a consequence, in a length space, we get the variational representation formula

$$\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) = \inf \left\{ \int_{[0,1] \cap \{r>0\}} \left(r^2(t)|x'|_{\mathsf{d}}^2(t) + |r'(t)|^2 \right) \mathrm{d}t : \right.$$
$$\left. (x, r) \in \widetilde{\mathrm{AC}}^2([0,1]; Y), \ [x(i), r(i)] = \mathfrak{y}_i, \ i = 0, 1 \right\}.$$
$$(8.13)$$

Remark 8.2 (The Euclidean case) In the Euclidean case $X = \mathbb{R}^d$ with the usual Euclidean distance $\mathsf{d}(x_1, x_2) := |x_1 - x_2|$ we can give a more explicit interpretation of the metric velocity (8.12) and write a simple duality formula for the chain rule of smooth functions that will turn out to be useful in Sect. 8.5.

For $\mathfrak{y} = [x, r] \in \mathrm{AC}^2([0, 1]; \mathfrak{C})$, we can define a Borel vector field $\mathfrak{y}'_{\mathfrak{C}} : [0, 1] \to \mathbb{R}^{d+1}$ by

$$\mathfrak{y}'_{\mathfrak{C}}(t) := \begin{cases} (r(t)x'(t), r'(t)) & \text{whenever } r(t) \neq 0 \text{ and the derivatives exist,} \\ (0, 0) & \text{otherwise.} \end{cases}$$
$$(8.14)$$

Then, (8.12) yields $|\mathfrak{y}'|_{\mathsf{d}_{\mathfrak{C}}}(t) = |\mathfrak{y}'_{\mathfrak{C}}(t)|_{\mathbb{R}^{d+1}}$ for $\mathscr{L}^1$-a.e. $t \in (0, 1)$ and the Euclidean norm of $\mathfrak{y}'_{\mathfrak{C}}(t)$ corresponds to the Riemannian norm of $\mathfrak{y}'$ with respect to the metric tensor $g_{[x,r]}(\dot{x}, \dot{r}) := r^2|\dot{x}|^2 + \dot{r}^2$.

For $\xi \in \mathrm{C}^1(\mathbb{R}^d \times [0, 1])$ we set $\zeta([x, r], t) := \xi(x, t)r^2$ and obtain $\partial_t \zeta([x, r], t) := \partial_t \xi(x, t)r^2$. Now defining the Borel map $\mathrm{D}_{\mathfrak{C}}\zeta : \mathfrak{C} \to (\mathbb{R}^{d+1})^*$ via

$$\mathrm{D}_{\mathfrak{C}}\zeta(\mathfrak{y}, t) := \begin{cases} (r\mathrm{D}_x\xi(x, t), 2r\xi(x, t)) & \text{for } \mathfrak{y} \neq \mathfrak{o}, \\ (0, 0) & \text{otherwise,} \end{cases}$$
$$(8.15)$$

we see that the map $t \mapsto \zeta(\mathfrak{y}(t), t)$ is absolutely continuous and satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(\mathfrak{y}(t), t) = \partial_t \zeta(\mathfrak{y}(t), t) + \langle \mathrm{D}_{\mathfrak{C}}\zeta(\mathfrak{y}(t), t), \mathfrak{y}'_{\mathfrak{C}}(t) \rangle_{\mathbb{R}^{d+1}}$$
$$(8.16)$$

$\mathscr{L}^1$-a.e. in $(0, 1)$. □

Note that the first component of $\mathrm{D}_{\mathfrak{C}}\zeta$ contains the factor r rather than r^2, since $\mathfrak{y}'_{\mathfrak{C}}$ in (8.12) already has one factor r in its first component. The Euclidean norm of $\mathrm{D}_{\mathfrak{C}}\zeta$ corresponds to the dual Riemannian norm of the differential of ζ.

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 $\mathcal{P}(C(I; Z))$, and we say that $\boldsymbol{\pi}$ has finite 2-energy if it is concentrated on $AC^2(I; Z)$ and

$$\int \left(\int_0^1 |z'|_{d_Z}^2 (t) \, dt \right) d\boldsymbol{\pi}(z) < \infty. \tag{8.17}$$

We denote by e_t the evaluation map on $C(I; Z)$ given by $e_t(z) := z(t)$. If $\boldsymbol{\pi}$ is a dynamic plan, $\alpha_t = (e_t)_\sharp \boldsymbol{\pi} \in \mathcal{M}(Z)$ is its marginal at time $t \in I$ and the curve $t \mapsto \alpha_t$ belongs to $C(I; (\mathcal{M}(Z), W_{d_Z}))$. If moreover $\boldsymbol{\pi}$ is a dynamic plan with finite 2-energy, then $\alpha \in AC^2(I; (\mathcal{M}(Z), W_{d_Z}))$ (see [33, Thm. 4]).

We say that $\boldsymbol{\pi}$ is an *optimal geodesic plan* between $\alpha_0, \alpha_1 \in \mathcal{P}(Z)$ if $(e_i)_\sharp \boldsymbol{\pi} = \alpha_i$ for $i = 0, 1$, if it is a dynamic plan concentrated on $\mathrm{Geo}(Z)$, and if

$$\int d_Z^2(z(0), z(1)) \, d\boldsymbol{\pi}(z) = \iint_0^1 |z'|^2 \, dt \, d\boldsymbol{\pi}(z) = W_{d_Z}^2(\alpha_0, \alpha_1). \tag{8.18}$$

Recalling (8.4), one immediately sees that for every dynamic plan $\boldsymbol{\pi}$ concentrated on $AC^2([0, 1]; Z)$ with $(e_i)_\sharp \boldsymbol{\pi} = \alpha_i$ the condition

$$\iint_0^1 |z'|^2 \, dt \, d\boldsymbol{\pi}(z) \le W_{d_Z}^2(\alpha_0, \alpha_1) \tag{8.19}$$

is sufficient to conclude that $\boldsymbol{\pi}$ is an optimal geodesic plan for $\alpha_0, \alpha_1 \in \mathcal{P}(Z)$.

When $Z = \mathfrak{C}$ we will denote by $\mathfrak{h}_t^2 = \mathfrak{h}^2 \circ (e_t)_\sharp$ the homogeneous marginal at time $t \in I$. Since $\mathfrak{h}^2 : \mathcal{P}(\mathfrak{C}) \to \mathcal{M}(X)$ is 1-Lipschitz (cf. Corollary 7.13), it follows that for every dynamic plan $\boldsymbol{\pi}$ with finite 2-energy the curve $\mu_t := \mathfrak{h}^2 \alpha_t = \mathfrak{h}_t^2 \boldsymbol{\pi}$ belongs to $AC^2(I; (\mathcal{M}(X), \mathsf{HK}))$ and moreover

$$|\mu_t'|_{\mathsf{HK}}^2 \le \int |\mathfrak{y}'|_{d_\mathfrak{C}}^2 (t) \, d\boldsymbol{\pi}(\mathfrak{y}) \quad \text{for a.e. } t \in (0, 1). \tag{8.20}$$

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 7.7 we find $\alpha_i \in \mathcal{P}_2(\mathfrak{C})$ such that $\mathfrak{h}^2 \alpha_i = \mu_i$ and $\mathsf{HK}(\mu_1, \mu_2) = W_{d_\mathfrak{C}}(\alpha_1, \alpha_2)$. Since $\mathfrak{C}$ is a length space, it is well known that $\mathcal{P}_2(\mathfrak{C})$ is a length space (see [47]); recalling (8.5), for every $\kappa > 1$ there exists $\alpha \in \mathrm{Lip}([0, 1]; (\mathcal{P}_2(\mathfrak{C}), W_{d_\mathfrak{C}}))$ connecting α_1 to α_2 such that $|\alpha'|_{W_{d_\mathfrak{C}}} \le \kappa \, W_{d_\mathfrak{C}}(\alpha_1, \alpha_2)$. Setting $\mu_t := \mathfrak{h}^2 \alpha_t$ we obtain a Lipschitz curve connecting μ_1 to μ_2 with length $\le \kappa \, \mathsf{HK}(\mu_1, \mu_2)$.

The converse property is a consequence of the next representation Theorem 8.4 and the fact that if $(\mathcal{P}_2(\mathfrak{C}), W_d)$ is a length (resp. geodesic) space, then $\mathfrak{C}$ and thus X are length (resp. geodesic) spaces.

We want to prove the converse representation result that every absolutely continuous curve $\mu : [0, 1] \to (\mathcal{M}(X), \mathsf{HK})$ can be written via a dynamic plan $\boldsymbol{\pi}$ as $\mu_t = \mathfrak{h}_t^2 \boldsymbol{\pi}$. 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 a curve in $\mathrm{AC}^p([0, 1]; (\mathcal{M}(X), \mathsf{HK}))$, $p \in [1, \infty]$, with*

$$\Theta := \sqrt{\mu_0(X)} + \int_0^1 |\mu'|_{\mathsf{HK}} \, \mathrm{d}t. \tag{8.21}$$

Then there exists a curve $(\alpha_t)_{t \in [0,1]}$ in $\mathrm{AC}^p([0, 1]; (\mathcal{P}_2(\mathfrak{C}), W_{d_\mathfrak{C}}))$ such that α_t is concentrated on $\mathfrak{C}[\Theta]$ for every $t \in [0, 1]$ and

$$\mu_t = \mathfrak{h}^2 \alpha_t \text{ in } [0, 1], \quad |\mu_t'|_{\mathsf{HK}} = |\alpha_t'|_{W_{d_\mathfrak{C}}} \text{ for a.e. } t \in (0, 1). \tag{8.22}$$

Moreover, when $p = 2$, there exists a dynamic plan $\boldsymbol{\pi} \in \mathcal{P}(\mathrm{AC}^2([0, 1]; \mathfrak{C}))$ such that

$$\alpha_t = (\mathsf{e}_t)_\sharp \boldsymbol{\pi}, \quad \mu_t = \mathfrak{h}_t^2 \boldsymbol{\pi} = \mathfrak{h}^2 \alpha_t \text{ in } [0, 1],$$
$$|\mu_t'|_{\mathsf{HK}}^2 = |\alpha_t'|_{W_{d_\mathfrak{C}}}^2 = \int |\mathfrak{y}'|_{d_\mathfrak{C}}^2(t) \, \mathrm{d}\boldsymbol{\pi}(\mathfrak{y}) \text{ for a.e. } t \in (0, 1). \tag{8.23}$$

Proof By Lisini's lifting Theorem [33, Theorem 5] (8.23) is a consequence of the first part of the statement and (8.22) in the case $p = 2$. It is therefore sufficient to prove that for a given $\mu \in \mathrm{AC}([0, 1]; (\mathcal{M}(X), \mathsf{HK}))$ there exists a curve $\alpha \in \mathrm{AC}([0, 1]; (\mathcal{P}_2(\mathfrak{C}), W_{d_\mathfrak{C}}))$ such that $\mu_t = \mathfrak{h}^2(\alpha_t)$ and $|\mu_t'| = |\alpha_t'|$ a.e. in $(0, 1)$. By a standard reparametrization technique (see e.g. [2, Lem. 1.1.4]), we may assume that μ is Lipschitz continuous and $|\mu_t'| = L$.

We divide the interval $I = [0, 1]$ into 2^N-intervals of size 2^{-N}, namely $I_i^N := [t_{i-1}^N, t_i^N]$ with $t_i^N := i \, 2^{-N}$ for $i = 1, \ldots, 2^N$. Setting $\mu_i^N := \mu_{t_i^N}$ we can apply the Gluing Lemma 7.11 (starting from $i = 0$ to 2^N) to obtain measures $\alpha_i^N \in \mathcal{P}_2(\mathfrak{C})$ such that

$$\mathfrak{h}(\alpha_i^N) = \mu_i^N, \quad W_{d_\mathfrak{C}}(\alpha_i^N, \alpha_{i+1}^N) = \mathsf{HK}(\mu_i^N, \mu_{i+1}^N) \le L 2^{-N}, \tag{8.24}$$

and concentrated on $\mathfrak{C}[\Theta_N]$ where

$$\Theta_N = \sqrt{\mu_0(X)} + \sum_{i=1}^{2^N} \mathsf{HK}(\mu_{i-1}^N, \mu_i^N) \le \Theta.$$

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 $\mathfrak{h}^2(\alpha^N(t)) = \mu_t$ and such that $W_{\mathsf{d}_\mathfrak{C}}(\alpha^N(t), \alpha^N(s)) \leq L|t-s|$ if $s = m2^{-N}$ and $t = n2^{-N}$ 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 $(\mathcal{P}_2(\mathfrak{C}), W_{\mathsf{d}_\mathfrak{C}})$ for every dyadic point t. Since $W_{\mathsf{d}_\mathfrak{C}}(\alpha(s), \alpha(t)) \leq L|t-s|$ for every dyadic s, t, we can extend α to a L-Lipschitz curve, still denoted by α, which satisfies $\mathfrak{h}^2(\alpha(t)) = \mu_t$. Since $\mathfrak{h}^2$ 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 a curve in* $\mathrm{AC}^2([0, 1]; (\mathcal{M}(X), \mathsf{HK}))$ *and let* Θ *as in* (8.21). *Then there exists a dynamic plan* $\tilde{\pi}$ *in* $\mathcal{P}(\widetilde{\mathrm{C}}([0, 1]; Y))$ *concentrated on* $\widetilde{\mathrm{AC}}^2([0, 1]; Y)$ *such that* $\alpha_t = (\mathsf{e}_t)_\sharp \pi$ *is concentrated in* $X \times [0, \Theta]$, *that* $\mu_t = \mathsf{h}^2((\mathsf{e}_t)_\sharp \pi)$, *and that*

$$|\mu_t'|^2_{\mathsf{HK}} = \int |\mathsf{y}'|^2(t) \, \mathrm{d}\pi(\mathsf{y}) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in [0, 1], \qquad (8.25)$$

where $|\mathsf{y}'|$ *is defined in* (8.11).

Another important consequence of the previous representation result is a precise characterization of the geodesics in $(\mathcal{M}(X), \mathsf{HK})$.

Theorem 8.6 (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* π *in* $\mathcal{P}(\mathrm{Geo}(\mathfrak{C}))$ *(recall* (8.18)*) such that*
  (a) π*-a.e. curve* $\mathfrak{y}$ *is a geodesic in* $\mathfrak{C}$,
  (b) $[0, 1] \ni t \mapsto \alpha_t := (\mathsf{e}_t)_\sharp \pi$ *is a geodesic in* $(\mathcal{P}_2(\mathfrak{C}), W_{\mathsf{d}_\mathfrak{C}})$, *where all* α_t *are concentrated on* $\mathfrak{C}[\Theta]$ *with* $\Theta^2 = 2(\mu_0(X) + \mathsf{HK}^2(\mu_0, \mu_1))$,
  (c) $\mu_t = \mathfrak{h}_t^2 \pi = \mathfrak{h}^2 \alpha_t$ *for every* $t \in [0, 1]$, *and*
  (d) $(\mathsf{e}_s, \mathsf{e}_t)_\sharp \pi \in \mathrm{Opt}_{\mathsf{HK}}(\mu_s, \mu_t)$ *if* $0 \leq s < t \leq 1$.

(ii) *If* (X, d) *is a geodesic space, for every* $\mu_0, \mu_1 \in \mathcal{M}(X)$ *and every* $\alpha \in \mathrm{Opt}_{\mathsf{HK}}(\mu_0, \mu_1)$ *there exists an optimal geodesic plan* $\pi \in \mathcal{P}(\mathrm{Geo}(\mathfrak{C}))$ *such that* $(\mathsf{e}_0, \mathsf{e}_1)_\sharp \pi = \alpha$.

Proof The statement (i) is an immediate consequence of Theorem 8.4. Notice that $|\mu_t'|^2_{\mathsf{HK}} = \mathsf{HK}^2(\mu_0, \mu_1)$ in $(0, 1)$ since $(\mu_t)_{t \in [0,1]}$ is a geodesic, so that (8.23) yields

$$\int \int_0^1 |\mathfrak{y}'|^2_{\mathsf{d}_\mathfrak{C}}(t) \, \mathrm{d}t \, \mathrm{d}\pi(\mathfrak{y}) = \int_0^1 |\mu_t'|^2_{\mathsf{HK}} \, \mathrm{d}t = \mathsf{HK}^2(\mu_0, \mu_1) \leq W^2_{\mathsf{d}_\mathfrak{C}}(\alpha_0, \alpha_1),$$

so that π satisfies (8.19) in $\mathfrak{C}$ and we deduce that it is an optimal geodesic plan.

Statement (ii) is a well known property [33, Thm. 6] 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 HK and GHK introduced in Sect. 7.8.

Corollary 8.7 *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*

$$|\mu'|_{\mathsf{GHK}}(t) = |\mu'|_{\mathsf{HK}}(t) \ \ \textit{for } \mathscr{L}^1\textit{-a.e. } t \in [0, 1]. \tag{8.26}$$

In particular if (X, d) *is a length metric space then* HK *is the length distance generated by* GHK.

Proof Since $\mathsf{GHK} \leq \mathsf{HK}$ 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 (8.26) 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 assertion follows by (8.23) 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) spaces in the sense of Alexandrov, referring to [10] and to [11] for other equivalent definitions and for the more general case of spaces with curvature $\geq k$, $k \in \mathbb{R}$. In the case of a smooth Riemannian manifold (M, g) equipped with the Riemannian distance d_g all the local definitions are equivalent to assume that the sectional curvature of M is nonnegative (or bounded by $k\mathsf{g}$, in the case of curvature $\geq k$).

According to Sturm [46], a metric space (Z, d_Z) is a Positively Curved (PC) metric space in the large if for every choice of points $z_0, z_1, \ldots, z_N \in Z$ and coefficients $\lambda_1, \ldots, \lambda_N \in (0, \infty)$ we have

$$\sum_{i,j=1}^N \lambda_i \lambda_j \mathsf{d}_Z^2(z_i, z_j) \leq 2 \sum_{i,j=1}^N \lambda_i \lambda_j \mathsf{d}_Z^2(z_0, z_j). \tag{8.27}$$

If every point of Z has a neighborhood that 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, which is based on

triangle comparison: for every choice of $z_0, z_1, z_2 \in Z$, every $t \in [0, 1]$, and every point z_t such that $\mathsf{d}_Z(z_t, z_k) = |k-t|\mathsf{d}_Z(z_0, z_1)$ for $k = 0, 1$ we have

$$\mathsf{d}_Z^2(z_2, z_t) \geq (1-t)\,\mathsf{d}_Z^2(z_2, z_0) + t\,\mathsf{d}_Z^2(z_2, z_1) - 2t(1-t)\,\mathsf{d}_Z^2(z_0, z_1). \quad (8.28)$$

When Z is also complete, the local and the global definitions are equivalent [46, Corollary 1.5]. Next we provide conditions on (X, d) or $(\mathfrak{C}, \mathsf{d}_{\mathfrak{C}})$ that guarantee that $(\mathcal{M}(X), \mathsf{HK})$ is a PC space.

Theorem 8.8 *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, cf. (8.27), then $(\mathcal{M}(X), \mathsf{HK}(X))$ is a PC space.*

(iii) *If (X, d) is separable, complete and geodesic, then $(\mathcal{M}(X), \mathsf{HK})$ 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 $(\mathcal{M}(\Omega), \mathsf{HK})$ is not PC, see [30, 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 [47]. Applying Corollary 7.13, 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

$$\mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}(\beta_0, \beta_i) = \mathsf{HK}(\mu_0, \mu_i) \quad \text{for } i = 1, \ldots, N, \quad (8.29)$$

where it is crucial that β_0 is the same for every i. It then follows that

$$\sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{HK}^2(\mu_i, \mu_j) \leq \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}^2(\beta_i, \beta_j) \leq 2 \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{W}_{\mathsf{d}_{\mathfrak{C}}}^2(\beta_0, \beta_i)$$

$$= 2 \sum_{i,j=1}^{N} \lambda_i \lambda_j \mathsf{HK}^2(\mu_0, \mu_i).$$

Let us now consider (iii) "$\Rightarrow$": If $(\mathcal{M}(X), \mathsf{HK})$ is PC, we have to prove that (X, d) has locally curvature ≥ 1. By Theorem [10, Thm. 4.7.1] it is sufficient to prove that $\mathfrak{C} \setminus \{\mathfrak{o}\}$ is locally PC to conclude that (X, 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 $\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_t, \mathfrak{y}_i) = |i-t|\mathsf{d}(\mathfrak{y}_0, \mathfrak{y}_1)$ for $i = 0, 1$.

Setting $\mu_i := r_i \delta_{x_i}$, $\mu_t := s_t \delta_{x_t}$, it is easy to check (cf. [30, Sect. 3.3.1])
that
$$\begin{aligned} \mathsf{HK}(\mu_i, \mu_j) &= \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_j) \text{ for } i, j \in \{0, 1, 2\}, \\ \mathsf{HK}(\mu_t, \mu_k) &= |k{-}t|\, \mathsf{HK}(\mu_0, \mu_1) \text{ for } k \in \{0, 1\}. \end{aligned} \tag{8.30}$$

We can thus apply (8.28) 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) "$\Leftarrow$": In order to prove the converse property we apply Remark 7.12.
For μ_0, μ_1, μ_2, and $\mu_3 = \mu_t \in \mathcal{M}(X)$ with $t \in [0, 1]$ and $\mathsf{HK}(\mu_3, \mu_k) = |k - t|\mathsf{HK}(\mu_0, \mu_1)$, 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

$$\mathfrak{h}_i^2 \boldsymbol{\alpha} = \mu_i, \quad \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_j) \, \mathsf{d}\boldsymbol{\alpha} = \mathsf{HK}^2(\mu_i, \mu_j) \tag{8.31}$$

for $(i, j) \in A = \{(0, 3), (1, 3), (2, 3)\}$. The triangle inequality, the elemen-
tary inequality $t(1 - t)(a + b)^2 \le (1 - t)a^2 + tb^2$, and the very definition of
HK yield for $t \in (0, 1)$ the estimate

$$\begin{aligned} t(1{-}t)\mathsf{HK}^2(\mu_0, \mu_1) &\le t(1 - t) \int \mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \, \mathsf{d}\boldsymbol{\alpha} \\ &\le \int t(1 - t)\big((\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) + \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1)\big)^2 \, \mathsf{d}\boldsymbol{\alpha} \\ &\le \int (1{-}t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_3) + t\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_3, \mathfrak{y}_1) \, \mathsf{d}\boldsymbol{\alpha} \\ &= (1{-}t)\mathsf{HK}^2(\mu_0, \mu_3) + t\mathsf{HK}^2(\mu_3, \mu_1) = t(1 - t)\mathsf{HK}^2(\mu_0, \mu_1). \end{aligned}$$

This series of inequalities shows in particular that

$$\begin{aligned} (1 - t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_3) + t\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_3, \mathfrak{y}_1) &= t(1 - t)\big(\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) + \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1)\big)^2 \\ &= t(1 - t)\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \quad \boldsymbol{\alpha}\text{-a.e.} \end{aligned}$$

so that

$$\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_3) = t\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) \text{ and } \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_3, \mathfrak{y}_1) = (1 - t)\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) \quad \boldsymbol{\alpha}\text{-a.e.}$$

Moreover, $\pi_\sharp^{\mathfrak{y}_0, \mathfrak{y}_1} \boldsymbol{\alpha} \in \mathrm{Opt}_{\mathsf{HK}}(\mu_0, \mu_1)$, so that (8.31) holds for $(i, j) \in A' = A \cup \{(0, 1)\}$.

By Theorem 7.20 we deduce that

$$\mathsf{d}(\mathsf{x}_i, \mathsf{x}_j) \le \pi/2 \quad \boldsymbol{\alpha}\text{-a.e. for } (i, j) \in A'.$$

If one of the points $\mathfrak{y}_i$, $i = 0, 1, 2$, is the vertex $\mathfrak{o}$, then it is not difficult to check by a direct computation that

$$d_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_3) \geq (1-t)d_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_0) + td_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_1) - 2t(1-t)d_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1). \quad (8.32)$$

When $\mathfrak{y}_i \in \mathfrak{C} \setminus \{\mathfrak{o}\}$ for every $i = 0, 1, 2$, we use $d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_0) \leq \frac{3}{2}\pi < 2\pi$, and Theorem [10, Thm. 4.7.1] yields (8.32) because of the assumption that X is PC. Integrating (8.32) w.r.t. $\boldsymbol{\alpha}$, by taking into account (8.31), the fact that $(\pi^0, \pi^1)_{\sharp}\boldsymbol{\alpha} \in \mathsf{Opt}_{\mathsf{HK}}(\mu_0, \mu_1)$, and that

$$\int d_{\mathfrak{C}}^2(\mathfrak{y}_2, \mathfrak{y}_i) \, d\boldsymbol{\alpha} \geq \mathsf{HK}^2(\mu_2, \mu_i) \text{ for } i = 0, 1,$$

we obtain

$$\mathsf{HK}^2(\mu_2, \mu_3) \geq (1-t)\mathsf{HK}^2(\mu_2, \mu_0) + t\mathsf{HK}^2(\mu_2, \mu_1) - 2t(1-t)\mathsf{HK}^2(\mu_0, \mu_1).$$

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 intimate connections of the duality formula of Theorem 7.21 with Lipschitz subsolutions of the Hamilton–Jacobi equation in $X \times (0, 1)$ given by

$$\partial_t \xi_t + \frac{1}{2}|\mathsf{D}_X \xi_t|^2 + 2\xi_t^2 = 0 \quad (8.33)$$

and its counterpart in the cone space

$$\partial_t \zeta_t + \frac{1}{2}|\mathsf{D}_{\mathfrak{C}} \zeta_t|^2 = 0. \quad (8.34)$$

Indeed, the first derivation of HK via LET was obtained by solving (8.33) for $X = \mathbb{R}^d$, see the remarks on the chronological development in Section A.

At a formal level, it is not difficult to check that solutions to (8.33) corresponds to the special class of solutions to (8.34) of the form

$$\zeta_t([x, r]) := \xi_t(x)r^2. \quad (8.35)$$

Indeed, still on the formal level we have the formula

$$|\mathsf{D}_{\mathfrak{C}} \zeta|^2 = \frac{1}{r^2}|\mathsf{D}_X \zeta|^2 + |\partial_r \zeta|^2 = |\mathsf{D}_X \xi|^2 r^2 + 4\xi^2 r^2 \quad \text{if } \zeta = \xi \, r^2. \quad (8.36)$$

Since the Kantorovich–Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$ can be defined in duality with subsolutions to (8.34) via the Hopf–Lax formula (see e.g. [3,50]) and 2-homogeneous marginals are modeled on test functions as in (8.35), 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 (8.33) and (8.34), 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* $|\mathsf{D}_Z f|$ and the *asymptotic Lipschitz constant* $|\mathsf{D}_Z f|_a$ are defined by [2,3,12]

$$
\begin{aligned}
|\mathsf{D}_Z f|(z) &:= \limsup_{x \to z} \frac{|f(x) - f(z)|}{\mathsf{d}_Z(x, z)}, \\
|\mathsf{D}_Z f|_a(z) &:= \lim_{r \downarrow 0} \sup_{\substack{x, y \in B_r(z) \\ y \neq x}} \frac{|f(y) - f(x)|}{\mathsf{d}_Z(x, y)}
\end{aligned}
\tag{8.37}
$$

with the convention that $|\mathsf{D}_Z f|(z) = |\mathsf{D}_Z f|_a(z) = 0$ whenever z is an isolated point. It is not difficult to check that $|\mathsf{D}_Z f|_a$ can also be defined as the minimal constant $L \geq 0$ such that there exists a function $G_L : Z \times Z \to [0, \infty)$ satisfying

$$
|f(x) - f(y)| \leq G_L(x, y)\mathsf{d}_Z(x, y), \quad \limsup_{x, y \to z} G_L(x, y) \leq L. \tag{8.38}
$$

Note that $|\mathsf{D}_Z f|_a$ is always an upper semicontinuous function clearly satisfying $|\mathsf{D}_Z f|_a \geq |\mathsf{D}_Z f|$. When Z is a length space, (8.5) and the chain rule along Lipschitz curves easily yield

$$
|f(x) - f(y)| \leq \mathsf{d}_Z(x, y) \sup_{B_r(z)} |\mathsf{D}_Z f| \quad \text{for every } x, y \in B_{2r}(z),
$$

so that $|\mathsf{D}_Z f|_a$ is the upper semicontinuous envelope of the metric slope $|\mathsf{D}_Z f|$. We will often write $|\mathsf{D}f|$, $|\mathsf{D}f|_a$ whenever the space Z will be clear from the context.

Remark 8.9 The notion of locally Lipschitz function and the value $|\mathsf{D}_Z f|_a$ does not change if we replace the distance d_Z with a distance $\tilde{\mathsf{d}}_Z$ of the form

$$\tilde{\mathsf{d}}_Z(z_1, z_2) := h(\mathsf{d}_Z(z_1, z_2)) \ \text{ for } z_1, z_2 \in Z,$$

$$\text{with } h : [0, \infty) \to [0, \infty) \text{ concave and } \lim_{r \downarrow 0} \frac{h(r)}{r} = 1. \tag{8.39}$$

In particular, the truncated distances $\mathsf{d}_Z \wedge \kappa$ with $\kappa > 0$, the distances $a \sin((\mathsf{d}_Z \wedge \kappa)/a)$ with $a > 0$ and $\kappa \in (0, a\pi/2]$, and the distance $\mathsf{g} = g(\mathsf{d})$ given by (7.74) 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 $\mathfrak{o}$, relation (7.12) yields

$$|\mathrm{D}_{\mathfrak{C}} f|_a(\mathfrak{o}) \le |\mathrm{D}_{(\mathfrak{C}, \mathsf{d}_{\pi/2,\mathfrak{C}})} f|_a(\mathfrak{o}) \le \sqrt{2}\, |\mathrm{D}_{\mathfrak{C}} f|_a(\mathfrak{o}). \tag{8.40}$$

$\square$

The next result shows that the asymptotic Lipschitz constant satisfies formula (8.36) for $\zeta([x, r]) = \xi(x) r^2$.

Lemma 8.10 *For $\xi : X \to \mathbb{R}$ 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 $\xi \in \mathrm{Lip}_b(X)$ with*

$$\sup_X |\xi| \le \frac{1}{R^2} \sup_{\mathfrak{C}[R]} |\zeta| \le \frac{1}{R} \mathrm{Lip}(\zeta, \mathfrak{C}[R]) \ \text{and}$$

$$\mathrm{Lip}(\xi, X) \le \frac{1}{R} \mathrm{Lip}(\zeta, \mathfrak{C}[R]). \tag{8.41}$$

(ii) *If $\xi \in \mathrm{Lip}_b(X)$, then ζ is $\mathsf{d}_{\mathfrak{C}}$-Lipschitz in $\mathfrak{C}[R]$ for every $R > 0$ with*

$$\sup_{\mathfrak{C}[R]} |\zeta| \le R^2 \sup_X |\xi| \ \text{and}$$

$$\mathrm{Lip}^2(\zeta, \mathfrak{C}[R]) \le R^2 \left(\mathrm{Lip}^2(\xi, (X, \tilde{\mathsf{d}})) + 4 \sup_X |\xi|^2 \right), \tag{8.42}$$

where $\tilde{\mathsf{d}} := 2 \sin(\mathsf{d}_{\pi}/2)$.

(iii) *In the cases (i) or (ii) we have, for every $x \in X$ and $r \ge 0$, the relation*

$$|\mathrm{D}_{\mathfrak{C}} \zeta|_a^2([x, r]) = \begin{cases} \left(|\mathrm{D}_X \xi|_a^2(x) + 4\xi^2(x) \right) r^2 & \text{for } r > 0, \\ 0 & \text{for } r = 0. \end{cases} \tag{8.43}$$

The analogous formula holds for the metric slope $|\mathrm{D}_{\mathfrak{C}} \zeta|([x, r])$. Moreover, equation (8.43) remains true if $\mathsf{d}_{\mathfrak{C}}$ is replaced by the distance $\mathsf{d}_{\pi/2,\mathfrak{C}}$.

Proof As usual we set $\mathfrak{y}_i = [x_i, r_i]$ and $\mathfrak{y} = [x, r]$.

Let us first check statement (i). If ζ is locally Lipschitz then $|\xi(x)| = \frac{1}{R^2}|\zeta([x, R]) - \zeta([x, 0])| \leq \frac{1}{R} \text{Lip}(\zeta; \mathfrak{C}[R])$ for every R sufficiently small, so that ξ is uniformly bounded. Moreover, using (7.4) for every $R > 0$ we have

$$R^2|\xi(x_1) - \xi(x_2)| \leq |\zeta(x_1, R) - \zeta(x_2, R)| \leq \text{Lip}(\zeta; \mathfrak{C}[R]) R \tilde{\mathsf{d}}(x_1, x_2)$$
$$\leq \text{Lip}(\zeta; \mathfrak{C}[R]) R \mathsf{d}(x_1, x_2),$$

so that ξ is uniformly Lipschitz and (8.41) holds.

Concerning (ii), for $\xi \in \text{Lip}_b(X)$ we set $S := \sup |\xi|$ and $L := \text{Lip}(\xi, (X, \tilde{\mathsf{d}}))$ and use the identity

$$\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2) = (\xi(x_1) - \xi(x_2))r_1 r_2 + 2\xi(x)r(r_1 - r_2)$$
$$+ \omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y})(r_1 - r_2), \tag{8.44}$$

where $\omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y}) := r_1\xi(x_1) + r_2\xi(x_2) - 2r\xi(x)$ with $\lim_{\mathfrak{y}_1, \mathfrak{y}_2 \to \mathfrak{y}} \omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y}) = 0$. Since $|\omega(\mathfrak{y}_1, \mathfrak{y}_2; 0)| \leq 2RS$ if $\mathfrak{y}_i \in \mathfrak{C}[R]$, equation (8.44) with $r = 0$ yields

$$|\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)| \leq L \tilde{\mathsf{d}}(x_1, x_2)r_1 r_2 + 2RS|r_1 - r_2|$$
$$\leq \left(L^2 + 4S^2\right)^{1/2} R \, \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2).$$

Letting $R \downarrow 0$ the inequality above also proves (8.43) in the case $r = 0$.

In order to prove (8.43) when $r \neq 0$ let us set $L_{\mathfrak{C}} := |\mathsf{D}_{\mathfrak{C}}\zeta|^2_a([x, r])$, $L_X := |\mathsf{D}_X\xi|_a(x)$, and let G_L be a function satisfying (8.38) with respect to the distance d (see Remark 8.9). Equation (8.44) yields, for all $\mathfrak{y} = [x, r]$, the relation

$$|\zeta(\mathfrak{y}_1) - \zeta(\mathfrak{y}_2)| \leq G_L(x_1, x_2)\tilde{\mathsf{d}}(x_1, x_2)r_1 r_2 + \left(2|\xi(x)|r + |\omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y})|\right)|r_1 - r_2|$$
$$\leq \left(G_L^2(x_1, x_2)r_1 r_2 + \left(2|\xi(x)|r + |\omega(\mathfrak{y}_1, \mathfrak{y}_2; \mathfrak{y})|\right)^2\right)^{1/2} \mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2).$$

Passing to the limit $\mathfrak{y}_1, \mathfrak{y}_2 \to \mathfrak{y}$ and using the fact that $x_1, x_2 \to x$ due to $r \neq 0$, we obtain $L_{\mathfrak{C}} \leq r\left(L_X^2 + 4|\xi(x)|^2\right)^{1/2}$.

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 $\xi(x_{1,n}) - \xi(x_{2,n}) \geq L'\delta_n$ where $0 < \delta_n := \tilde{\mathsf{d}}(x_{1,n}, x_{2,n}) \to 0$. Choosing $r_{1,n} := r$ and $r_{2,n} = r(1 + \lambda\delta_n)$ for an arbitrary constant $\lambda \in \mathbb{R}$ with the same sign as $\xi(x)$, we can apply (8.44) and arrive at

$$L_{\mathfrak{C}} \geq \liminf_{n \to \infty} \frac{|\zeta(\mathfrak{y}_{1,n}) - \zeta(\mathfrak{y}_{2,n})|}{\mathsf{d}_{\mathfrak{C}}(\mathfrak{y}_{1,n}, \mathfrak{y}_{2,n})} \geq \liminf_{n \to \infty} \frac{L'\delta_n r^2 + 2|\xi(x)|r^2|\lambda|\delta_n + o(\delta_n)}{\sqrt{\lambda^2 r^2 \delta_n^2 + r^2 \delta_n^2 + o(\delta_n)}}$$

$$= r\frac{L' + 2|\xi(x)|\,|\lambda|}{\sqrt{\lambda^2 + 1}}.$$

Optimizing with respect to λ we obtain

$$L_{\mathfrak{C}}^2 \geq r^2\big((L')^2 + 4|\xi(x)|^2\big), \quad \text{where } L' \leq L_X \text{ is arbitrary.}$$

This proves (8.43) for the asymptotic Lipschitz constant $|\mathsf{D}_{\mathfrak{C}}\zeta|_a$. The arguments for proving (8.43) for metric slopes $|\mathsf{D}_{\mathfrak{C}}\zeta|$ are completely analogous. $\qquad\square$

Hopf–Lax formula and subsolutions to metric Hamilton–Jacobi equation in the cone $\mathfrak{C}$. Whenever $f \in \mathrm{Lip}_b(\mathfrak{C})$ the Hopf–Lax formula

$$\mathscr{Q}_t f(\mathfrak{y}) := \inf_{\mathfrak{y}' \in \mathfrak{C}} \Big(f(\mathfrak{y}') + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2(\mathfrak{y}, \mathfrak{y}') \Big) \quad \text{for } \mathfrak{y} \in \mathfrak{C} \text{ and } t > 0, \qquad (8.45)$$

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

$$\inf_{\mathfrak{C}} f \leq \mathscr{Q}_t f \leq \sup_{\mathfrak{C}} f, \quad \mathrm{Lip}(\mathscr{Q}_t f; \mathfrak{C}) \leq 2\,\mathrm{Lip}(f, \mathfrak{C}), \qquad (8.46)$$

and solves

$$\partial_t^+ \mathscr{Q}_t f(\mathfrak{z}) + \frac{1}{2}|\mathsf{D}_{\mathfrak{C}}\mathscr{Q}_t f|_a^2(\mathfrak{z}) \leq 0 \quad \text{for every } \mathfrak{z} \in \mathfrak{C}, \ t > 0, \qquad (8.47)$$

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 possibly countable exceptions and that (8.47) 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. This is stated in our main result:

Theorem 8.11 (Metric subsolution of Hamilton–Jacobi equation in X) *Let* $\xi \in \mathrm{Lip}_b(X)$ *satisfy the uniform lower bound* $P := 1 + 2\inf_X(\xi \wedge 0) > 0$ *and let us set* $\zeta([x, r]) := \xi(x)r^2$. *Then, for every* $t \in [0, 1]$ *we have*

$$\mathscr{Q}_t\zeta([x, r]) = \xi_t(x)r^2, \quad \text{where } \xi_t(x) := \mathscr{P}_t\xi(x) \text{ and}$$

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

$$= \inf_{x' \in X} \frac{1}{2t}\Big(1 - \frac{\cos^2(\mathsf{d}_{\pi/2}(x, x'))}{1 + 2t\xi(x')} \Big). \qquad (8.48)$$

Moreover, for every $R > 0$ *we have*

$$\xi_t(x)r^2 = \inf_{\mathfrak{y}'=[x',r']\in\mathfrak{C}[R]} \left(\xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r']) \right)$$
$$\text{for all } x \in X, \ r \leq PR. \tag{8.49}$$

The map $t \mapsto \xi_t$ *is Lipschitz from* $[0,1]$ *to* $\mathsf{C}_b(X)$ *with* $\xi_t \in \mathrm{Lip}_b(X)$ *for every* $t \in [0,1]$. *Moreover,* ξ_t *is a subsolution to the generalized Hamilton–Jacobi equation*

$$\partial_t^+ \xi_t(x) + \frac{1}{2}|\mathsf{D}_X\xi_t|_a^2(x) + 2\xi_t^2(x) \leq 0 \quad \text{for } x \in X \text{ and } t \in [0,1]. \tag{8.50a}$$

For every $x \in X$ *the map* $t \mapsto \xi_t(x)$ *is time differentiable with at most countable exceptions. If* (X,d) *is a length space,* (8.50a) *holds with equality and* $|\mathsf{D}_X\xi_t|_a(x) = |\mathsf{D}_X\xi_t|(x)$ *for every* $x \in X$ *and* $t \in [0,1]$:

$$\partial_t^+ \xi_t(x) + \frac{1}{2}|\mathsf{D}_X\xi_t|_a^2(x) + 2\xi_t^2(x) = 0, \quad |\mathsf{D}_X\xi_t|_a(x) = |\mathsf{D}_X\xi_t|(x). \tag{8.50b}$$

Notice that when $\xi(x) \equiv \xi$ is constant, (8.48) reduces to $\mathscr{P}_t\xi = \xi/(1+2t\xi)$ which is the solution to the elementary differential equation $\frac{\mathsf{d}}{\mathsf{d}t}\xi + 2\xi^2 = 0$.

Proof Let us observe that $\inf_{t\in[0,1],z\in X}(1 + 2t\xi(z)) = P > 0$. A simple calculation shows

$$\xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r'])$$
$$= \frac{1}{2t}\left((1+2t\xi(x'))(r')^2 + r^2 - 2r\,r'\cos(\mathsf{d}_\pi(x,x')) \right)$$
$$= \frac{1}{2t\,(1+2t\xi(x'))}\left[\left((1+2t\xi(x'))r' - \cos(\mathsf{d}_\pi(x,x'))r \right)^2 \right.$$
$$\left. + r^2\left(2t\xi(x') + \sin^2(\mathsf{d}_\pi(x,x')) \right) \right].$$

Hence, if we choose

$$r' := \begin{cases} r\cos(\mathsf{d}_\pi(x,x'))/(1+2t\xi(x')) & \text{if } \mathsf{d}(x,x') \leq \pi/2 \\ 0 & \text{otherwise,} \end{cases} \tag{8.51}$$

we find (notice the truncation at $\pi/2$ instead of π)

$$\inf_{r'\geq 0} \xi(x')(r')^2 + \frac{1}{2t}\mathsf{d}_{\mathfrak{C}}^2([x,r];[x',r'])$$
$$= \frac{r^2}{2t\,(1+2t\xi(x'))}\left(2t\xi(x') + \sin^2(\mathsf{d}_{\pi/2}(x,x')) \right), \tag{8.52}$$

which yields (8.48). Now (8.49) also follows, since $r' \le r/P$ in (8.51).

Equation (8.49) 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 8.10.

Equation (8.50a) and the other regularity properties then follow by (8.43) and the general properties of the Hopf–Lax formula in $\mathfrak{C}[R]$. $\qquad\square$

Duality between the Hellinger–Kantorovich distance and subsolutions to the generalized Hamilton–Jacobi equation. We conclude this section with the main application of the above results to the Hellinger–Kantorovich distance.

Theorem 8.12 *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 (8.50a), then the curve $t \mapsto \int \xi_t \, \mathrm{d}\mu_t$ is absolutely continuous and*

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_X \xi_t \, \mathrm{d}\mu_t \le \frac{1}{2} |\mu_t'|^2_{\mathsf{HK}} \tag{8.53}$$

(ii) *If (X, d) is a length space, then for every μ_0, μ_1 and $k \in \mathbb{N} \cup \{\infty\}$ we have*

$$\frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) = \sup \left\{ \int_X \xi_1 \, \mathrm{d}\mu_1 - \int_X \xi_0 \, \mathrm{d}\mu_0 \ : \ \xi \in \mathrm{C}^k([0, 1]; \mathrm{Lip}_b(X)), \right.$$
$$\left. \partial_t \xi_t(x) + \frac{1}{2} |\mathrm{D}_X \xi_t|^2(x) + 2\xi_t^2(x) \le 0 \ in \ X \times (0, 1) \right\}. \tag{8.54}$$

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 (8.50a) 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 8.10 the function ζ solves

$$\partial_t^+ \zeta_t + \frac{1}{2} |\mathrm{D}_{\mathfrak{C}} \zeta_t|^2_a \le 0 \quad \text{in } \mathfrak{C} \times (0, 1).$$

According to Theorem 8.4 we find $\theta > 0$ and a curve $\alpha \in \mathrm{AC}^2([0, 1]; (\mathcal{P}_2(\mathfrak{C}[\theta]),$ $\mathsf{W}_{\mathsf{d}_\mathfrak{C}}))$ satisfying (8.22). Applying the results of [6, Sect. 6], the map $t \mapsto \int_\mathfrak{C} \zeta_t \, \mathrm{d}\alpha_t$ is absolutely continuous with

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\mathfrak{C} \zeta_t \, \mathrm{d}\alpha_t \leq \frac{1}{2} |\alpha'_t|^2_{\mathsf{W}_{\mathsf{d}_\mathfrak{C}}} \quad \mathscr{L}^1\text{-a.e. in } (0, 1).$$

Since $\int_\mathfrak{C} \zeta_t \, \mathrm{d}\alpha_t = \int_X \xi_t \, \mathrm{d}\mu_t$ we obtain (8.53).

Let us now prove (ii). As a first step, denoting by S the right-hand side of (8.54), we prove that $\mathsf{HK}^2(\mu_0, \mu_1) \geq S$. If $\xi \in \mathrm{C}^1([0, 1]; \mathrm{Lip}_b(X))$ satisfies the pointwise inequality

$$\partial_t \xi_t(x) + \frac{1}{2} |\mathrm{D}_X \xi_t|^2(x) + 2\xi_t^2(x) \leq 0, \tag{8.55}$$

then it also satisfies (8.50a), because (8.55) provides the relation

$$\frac{1}{2} |\mathrm{D}_X \xi_t|^2(x) \leq -\left(\partial_t \xi_t(x) + 2\xi_t^2(x)\right) \quad \text{for every } (x, t) \in X \times (0, 1), \tag{8.56}$$

where the right hand side is bounded and continuous in X. Equation (8.56) 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 $|\mu'_t|_{\mathsf{HK}} \leq \lambda \mathsf{HK}(\mu_0, \mu_1)$, whose existence is guaranteed by the length property of X and a standard rescaling technique. Relation (8.53) yields

$$2\int_X \xi_1 \, \mathrm{d}\mu_1 - 2\int_X \xi_0 \, \mathrm{d}\mu_0 \leq \int_0^1 |\mu'_t|^2_{\mathsf{HK}} \, \mathrm{d}t \leq \lambda^2 \mathsf{HK}^2(\mu_0, \mu_1).$$

Since $\lambda > 1$ is arbitrary, we get $\mathsf{HK}^2(\mu_0, \mu_1) \geq S$.

In order to prove the converse inequality in (8.54) we fix $\eta > 0$ and apply the duality Theorem 7.21 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

$$2\int_X \mathscr{P}_1 \xi_0 \, \mathrm{d}\mu_1 - 2\int_X \xi_0 \, \mathrm{d}\mu_0 \geq \mathsf{HK}^2(\mu_0, \mu_1) - \eta. \tag{8.57}$$

Setting $\xi_t := \mathscr{P}_t \xi_0$ we find a solution to (8.50a) 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 $\int \xi_i^\varepsilon \, \mathrm{d}\mu_i \to \int \xi_i \, \mathrm{d}\mu_i$ as $\varepsilon \downarrow 0$ for $i = 0, 1$.

We use an argument of [1], which relies on the scaling invariance of the generalized Hamilton–Jacobi equation: If ξ solves (8.55) and $\lambda > 0$, then $\xi^\lambda_t(x) := \lambda\xi_{\lambda t+t_0}(x)$ solves (8.55) as well. Hence, by approximating ξ_t with $\lambda\xi(\lambda t+(1-\lambda)/2, x)$ with $0 < \lambda < 1$ and passing to the limit $\lambda \uparrow 1$, 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 $\kappa \in C^\infty_c(\mathbb{R})$ with integral 1 defined via

$$\xi^\varepsilon_t(x) := (\xi_{(\cdot)}(x) * \kappa_\varepsilon)_t = \int_\mathbb{R} \xi_w(x)\kappa_\varepsilon(t-w)\,\mathrm{d}w, \qquad (8.58)$$

where $\kappa_\varepsilon(t) := \varepsilon^{-1}\kappa(t/\varepsilon)$. It yields a curve $\xi^\varepsilon \in C^\infty([0, 1]; \mathrm{Lip}_b(X))$ satisfying

$$\partial_t\xi^\varepsilon_t + \frac{1}{2}\big(|D_X\xi_{(\cdot)}|^2\big) * \kappa_\varepsilon + 2\big(\xi^2_{(\cdot)}\big) * \kappa_\varepsilon \leq 0 \quad \text{in } X \times [0, 1].$$

By Jensen's inequality, we have the two estimates $\xi^2_{(\cdot)} * \kappa_\varepsilon \geq (\xi_{(\cdot)} * \kappa_\varepsilon)^2$ and $|D_X\xi_{(\cdot)}|^2 * \kappa_\varepsilon \geq (|D_X\xi_{(\cdot)}| * \kappa_\varepsilon)^2$. Moreover, applying the following Lemma 8.13 we also get $|D_X\xi_{(\cdot)}| * \kappa_\varepsilon \geq |D_X\xi^\varepsilon_{(\cdot)}|$, so that the smooth convolution ξ^ε_t satisfies (8.55). Since $\xi^\varepsilon_t \to \xi_t$ uniformly in X for every $t \in [0, 1]$, we easily get

$$S \geq \lim_{\varepsilon\downarrow 0} 2\Big(\int_X \xi^\varepsilon_1\,\mathrm{d}\mu_1 - \int_X \xi^\varepsilon_0\,\mathrm{d}\mu_0\Big) \geq \mathsf{HK}^2(\mu_0, \mu_1) - \eta.$$

Since $\eta > 0$ is arbitrary the proof of (ii) is complete. $\qquad\square$

The next result shows that averaging w.r.t. a probability measure $\pi \in \mathcal{P}(\Omega)$ does not increase the metric slope nor the asymptotic Lipschitz constant. This was used in the last proof for the temporal smoothing and will be used for spatial smoothing in Corollary 8.14.

Lemma 8.13 *Let (X, d) be a separable metric space, let $(\Omega, \mathcal{B}, \pi)$ be a probability space (i.e. $\pi(\Omega) = 1$) 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 $x \mapsto \xi(x) := \int_\Omega \xi_\omega(x)\,\mathrm{d}\pi(\omega)$ belongs to $\mathrm{Lip}_b(X)$ and for every $x \in X$ the maps $\omega \mapsto |D_X\xi_\omega|(x)$ and $\omega \mapsto |D_X\xi_\omega|_a(x)$ are $\mathcal{B}$-measurable and satisfy*

$$|D_X\xi|_a(x) \leq \int_X |D_X\xi_\omega|_a(x)\,\mathrm{d}\pi(\omega), \quad |D_X\xi|(x) \leq \int_X |D_X\xi_\omega|(x)\,\mathrm{d}\pi(\omega). \tag{8.59}$$

Proof The fact that $\xi \in \mathrm{Lip}_b(X)$ is obvious. To show measurability we fix $x \in X$ and use the expression (8.37) for $|D_X\xi|_a(x)$. 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 $B_r(x)$. Thus, the measurability follows, because the pointwise supremum of countably many measurable functions is measurable. An analogous argument holds for $|D_X \xi_\omega|$.

Using the definition $\xi := \int \xi_\omega d\pi$ we have

$$\frac{|\xi(y) - \xi(z)|}{\mathsf{d}(y, z)} \leq \int_\Omega \frac{|\xi_\omega(y) - \xi_\omega(z)|}{\mathsf{d}(y, z)} \, d\pi(\omega) \text{ for } y \neq z.$$

Taking the supremum with respect to $y, z \in \tilde{B}_r(x)$ and $y \neq z$, we obtain

$$\sup_{y \neq z \in B_r(x)} \frac{|\xi(y) - \xi(z)|}{\mathsf{d}(y, z)} \leq \int_\Omega s_{r,\omega}(x) \, d\pi(\omega).$$

A further limit as $r \downarrow 0$ and the application of the Lebesgue Dominated convergence Theorem yields the first inequality of (8.59). The argument to prove the second inequality is completely analogous. □

When $X = \mathbb{R}^d$ the characterization (8.54) of HK holds for an even smoother class of subsolutions ξ of the generalized Hamilton–Jacobi equation.

Corollary 8.14 *Let $X = \mathbb{R}^d$ be endowed with the Euclidean distance. Then*

$$\mathsf{HK}^2(\mu_0, \mu_1) = 2 \sup \left\{ \int_X \xi_1 \, d\mu_1 - \int_X \xi_0 \, d\mu_0 \, : \, \xi \in C_c^\infty(\mathbb{R}^d \times [0, 1]), \right.$$
$$\left. \partial_t \xi_t(x) + \frac{1}{2} |D_x \xi_t(x)|^2 + 2\xi_t^2(x) \leq 0 \text{ in } X \times (0, 1) \right\}.$$
$$\tag{8.60}$$

Proof We just have to check that the supremum of (8.54) does not change if we substitute $C^\infty([0, 1]; \text{Lip}_{bs}(\mathbb{R}^d))$ with $C_c^\infty(\mathbb{R}^d \times [0, 1])$. This can be achieved by approximating any subsolution $\xi \in C^\infty([0, 1]; \text{Lip}_{bs}(\mathbb{R}^d))$ via convolution in space with a smooth kernel with compact support, which still provides a subsolution thanks to Lemma 8.13. □

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 8.12 with subsolutions of the Hamilton-Jacobi equation to quickly derive a dynamic formulation "à la Benamou-Brenier" [7,37], [2,

Sect. 8] of the Hellinger–Kantorovich distance, which has also been considered in the recent [27]. 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 our companion paper [30].

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

$$\int \xi \, \mathrm{d}\mu_I := \int_0^1 \int_{\mathbb{R}^d} \xi_t(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \qquad (8.61)$$

induced by a curve $\mu \in C^0([0, 1]; \mathcal{M}(\mathbb{R}^d))$.

Definition 8.15 Let $\mu \in C^0([0, 1]; \mathcal{M}(\mathbb{R}^d))$, let $(\boldsymbol{v}, w) : \mathbb{R}^d \times (0, 1) \to \mathbb{R}^{d+1}$ be a Borel vector field in $L^2(\mathbb{R}^d \times (0, 1), \mu_I; \mathbb{R}^{d+1})$, thus satisfying

$$\int_0^1 \int_{\mathbb{R}^d} \left(|\boldsymbol{v}_t(x)|^2 + w_t^2(x) \right) \mathrm{d}\mu_t(x) \, \mathrm{d}t = \int |(\boldsymbol{v}, w)|^2 \, \mathrm{d}\mu_I < \infty. \quad (8.62)$$

We say that μ satisfies the *continuity equation with reaction* governed by $(\boldsymbol{v}, w)$ if

$$\partial_t \mu_t + \nabla \cdot (\boldsymbol{v}_t \mu_t) = w_t \mu_t \quad \text{holds in } \mathscr{D}'(\mathbb{R}^d \times (0, 1)), \qquad (8.63)$$

i.e. for every test function $\xi \in C_c^\infty(\mathbb{R}^d \times (0, 1))$

$$\int_0^1 \int_{\mathbb{R}^d} \left(\partial_t \xi_t(x) + \mathrm{D}_x \xi_t(x) \boldsymbol{v}_t(x) + \xi_t(x) w_t(x) \right) \mathrm{d}\mu_t \, \mathrm{d}t = 0. \qquad (8.64)$$

An equivalent formulation [2, Sect. 8.1] of (8.63) is

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \xi(x) \, \mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \left(\mathrm{D}_x \xi(x) \boldsymbol{v}_t(x) + \xi(x) w_t(x) \right) \mathrm{d}\mu_t \quad \text{in } \mathscr{D}'(0, 1),$$
$$(8.65)$$

for every $\xi \in C_c^\infty(\mathbb{R}^d)$. We have a first representation result for absolutely continuous curves $t \mapsto \mu_t$, which relies in Theorem 8.4, where we constructed suitable lifted plans $\boldsymbol{\pi} \in \mathcal{P}(\mathrm{AC}^2([0, 1]; \mathfrak{C}))$, i.e. $\mu_t = \mathfrak{h}_t^2 \boldsymbol{\pi}$, where $\mathfrak{C}$ is now the cone over $\mathbb{R}^d$.

Theorem 8.16 *Let $(\mu_t)_{t \in [0, 1]}$ be a curve in $\mathrm{AC}^2([0, 1]; (\mathcal{M}(\mathbb{R}^d), \mathsf{HK}))$. Then μ satisfies the continuity equation with reaction* (8.63) *with a Borel vector field $(\boldsymbol{v}, w) \in L^2(\mathbb{R}^d \times (0, 1), \mu_I; \mathbb{R}^{d+1})$ satisfying*

$$(\boldsymbol{v}_t, w_t) \in L^2(\mathbb{R}^d; \mu_t), \quad \int \left(|\boldsymbol{v}_t|^2 + \frac{1}{4} |w_t|^2 \right) \mathrm{d}\mu_t \leq |\mu_t'|^2 \qquad (8.66)$$

for $\mathscr{L}^1$-a.e. $t \in (0, 1)$.

Proof We will denote by I the interval $[0, 1]$ endowed with the Lebesgue measure $\lambda = \mathscr{L}^1 \llcorner [0, 1]$. Recalling the map $(\mathsf{x}, \mathsf{r}) : \mathfrak{C} \to \mathbb{R}^d \times [0, \infty)$ we define the maps $\mathsf{x}_I : C(I; \mathfrak{C}) \times I \to \mathbb{R}^d \times I$ and $\mathsf{R} : C(I; \mathfrak{C}) \times I \to \mathbb{R}_+$ via $\mathsf{x}_I(z, t) := (\mathsf{x}(z(t)), t)$ and $\mathsf{R}(z, t) := \mathsf{r}(z(t))$.

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 $\hat{\mu}_I := (\mathsf{x}_I)_\sharp \boldsymbol{\pi}_I$ and the disintegration $(\tilde{\boldsymbol{\pi}}_{x,t})_{(x,t) \in \mathbb{R}^d \times I}$ of $\boldsymbol{\pi}_I$ with respect to $\hat{\mu}_I$. Since $\boldsymbol{\pi}$ is in fact a dynamic plan on $\mathfrak{C}[\Theta]$, where Θ is given by (8.21), we notice that $\boldsymbol{\pi}_I \leq \Theta^2(\boldsymbol{\pi} \otimes \lambda)$ so that $\boldsymbol{\pi}_I$ has finite mass and

$$\hat{\mu}_I = \int_0^1 (\mu_t \otimes \delta_t) \, \mathrm{d}\lambda(t), \tag{8.67}$$

coincides with μ_I in (8.61), because for every $\xi \in \mathrm{B}_b(\mathbb{R}^d \times I)$ we have

$$\int \xi \, \mathrm{d}\hat{\mu}_I = \int \xi(\mathsf{x}(z(t)), t) \mathsf{r}^2(z(t)) \, \mathrm{d}(\boldsymbol{\pi} \otimes \lambda)(z, t)$$

$$= \int_0^1 \int_{\mathbb{R}^d} \xi_t(x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t = \int \xi \, \mathrm{d}\mu_I.$$

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 (8.14). 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 $\boldsymbol{u}_I = (\boldsymbol{v}, \hat{w}) \in \mathrm{L}^2(\mathbb{R}^d \times I; \mu_I; \mathbb{R}^{d+1})$ which satisfies

$$\boldsymbol{u}_I(x, t) = \int \boldsymbol{u} \, \mathrm{d}\boldsymbol{\pi}_{x,t} \quad \text{for } \mu_I\text{-a.e. } (x, t) \in \mathbb{R}^d \times I \text{ and}$$

$$\int \left(|\boldsymbol{v}_t|^2 + \hat{w}_t^2 \right) \mathrm{d}\mu_t \leq |\mu'_t|^2. \tag{8.68}$$

Choosing a test function $\zeta([x, r], t) := \xi(x)\eta(t)r^2$ with $\xi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$ and $\eta \in \mathrm{C}_c^\infty(I)$ we can exploit the chain rule (8.16) in $\mathbb{R}^d$ and find

$$-\int_0^1 \eta' \int_{\mathbb{R}^d} \xi \, \mathrm{d}\mu_t \, \mathrm{d}t = -\int_{\mathbb{R}^d \times I} \eta'(t)\xi(x) \, \mathrm{d}\mu_I$$

$$= -\int \xi(\mathsf{x}(\mathfrak{y}(t)) \, \mathsf{r}^2(\mathfrak{y}(t)))\eta'(t) \, \mathrm{d}(\boldsymbol{\pi} \otimes \lambda)$$

$$= -\int \partial_t \zeta(\mathfrak{y}(t), t) \, \mathrm{d}(\boldsymbol{\pi} \otimes \lambda)$$

$$= \int \left(-\frac{\mathrm{d}}{\mathrm{d}t} \zeta(\mathfrak{y}(t), t) + \langle \mathrm{D}_{\mathfrak{C}}\zeta(\mathfrak{y}(t), t), \mathfrak{y}'_{\mathfrak{C}}(t) \rangle \right) \mathrm{d}(\boldsymbol{\pi} \otimes \lambda)$$

$$= \int \left(\int_0^1 -\frac{d}{dt} \zeta(\mathfrak{y}(t), t) \, dt \right) d\boldsymbol{\pi} + \int \langle (D_x \xi(\mathbf{x}_I), 2\xi(\mathbf{x}_I)), \boldsymbol{u} \rangle \mathsf{R}^2 \, d(\boldsymbol{\pi} \otimes \lambda)$$

$$= \int \eta(t) \langle (D_x \xi(x), 2\xi(x)), \boldsymbol{u}_I \rangle \, d\mu_I$$

$$= \int_0^1 \eta(t) \int_{\mathbb{R}^d} \left(\langle D_x \xi(x), \boldsymbol{v}_t(x) \rangle + 2\xi(x) \hat{w}_t(x) \right) d\mu_t \, dt.$$

Setting $w_t = 2\hat{w}_t$ the continuity equation with reaction (8.65) holds. □

The next result provides the opposite inequality, which will be deduced from the duality between the solutions of the generalized Hamilton–Jacobi equation and HK developed in Theorem 8.12.

Theorem 8.17 *Let* $(\mu_t)_{t \in [0,1]}$ *be a continuous curve in* $\mathcal{M}(\mathbb{R}^d)$ *that solves the continuity equation with reaction* (8.63) *governed by the Borel vector field* $(\boldsymbol{v}, w) \in L^2(\mathbb{R}^d \times [0, 1], \mu_I; \mathbb{R}^{d+1})$ *with* μ_I *given by* (8.61). *Then* $\mu \in$ $\mathrm{AC}^2([0, 1]; (\mathcal{M}(\mathbb{R}^d), \mathsf{HK}))$ *and*

$$|\mu_t'|^2 \le \int_{\mathbb{R}^d} \left(|\boldsymbol{v}_t|^2 + \frac{1}{4} |w_t|^2 \right) d\mu_t \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, 1). \tag{8.69}$$

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, Corollary 8.14 yields

$$\mathsf{HK}^2(\mu_0, \mu_1) = 2(b-a) \sup \left\{ \int_{\mathbb{R}^d} \xi_b \, d\mu_1 - \int_{\mathbb{R}^d} \xi_a \, d\mu_0 : \xi \in \mathrm{C}_c^\infty(\mathbb{R}^d \times [a, b]), \right.$$
$$\left. \partial_t \xi_t(x) + \frac{1}{2} \left| D_x \xi_t(x) \right|^2 + 2\xi_t^2(x) \le 0 \text{ in } \mathbb{R}^d \times (a, b) \right\}. \tag{8.70}$$

Let $\xi \in \mathrm{C}_c^\infty(\mathbb{R}^d \times [0, 1])$ be a subsolution to the Hamilton–Jacobi equation $\partial_t \xi + \frac{1}{2} |D\xi|^2 + 2\xi^2 \le 0$ in $\mathbb{R}^d \times [0, 1]$. By a standard argument (see [2, Lem. 8.1.2]), the integrability (8.62), the weak continuity of $t \mapsto \mu_t$ and (8.64) yield

$$2 \int_{\mathbb{R}^d} \xi_{t_1} \, d\mu_{t_1} - 2 \int_{\mathbb{R}^d} \xi_{t_0} \, d\mu_{t_0} = 2 \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left(\partial_t \xi_t + \langle D_x \xi_t, \boldsymbol{v}_t \rangle + \xi_t w_t \right) d\mu_t \, dt$$

$$\le 2 \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left(-\frac{1}{2} |D_x \xi_t|^2 - 2\xi_t^2 + \langle D_x \xi_t, \boldsymbol{v}_t \rangle + \xi_t w_t \right) d\mu_t \, dt$$

$$\le \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left(|\boldsymbol{v}_t|^2 + \frac{1}{4} |w_t|^2 \right) d\mu_t \, dt.$$

Applying Corollary 8.14 and (8.70) we find

$$\mathsf{HK}^2(\mu_{t_0}, \mu_{t_1}) \leq (t_1 - t_0) \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left(|\boldsymbol{v}_t|^2 + \frac{1}{4} |w_t|^2 \right) \mathrm{d}\mu_t \, \mathrm{d}t$$

for every $0 \leq t_0 < t_1 \leq 1$, which yields (8.69). □

Combining Theorems 8.16 and 8.17 with Theorem 8.4 and the geodesic property of $(\mathcal{M}(\mathbb{R}^d), \mathsf{HK})$ we immediately have the desired dynamic representation.

Theorem 8.18 (Representation of HK à la Benamou-Brenier) *For every $\mu_0, \mu_1 \in \mathcal{M}(\mathbb{R}^d)$ we have*

$$\mathsf{HK}^2(\mu_0, \mu_1) = \min \Big\{ \int_0^1 \int_{\mathbb{R}^d} \left(|\boldsymbol{v}_t|^2 + \frac{1}{4} |w_t|^2 \right) \mathrm{d}\mu_t \, \mathrm{d}t :$$
$$\mu \in C([0, 1]; \mathcal{M}(\mathbb{R}^d)), \mu_{t=i} = \mu_i,$$
$$\partial_t \mu_t + \nabla \cdot (\boldsymbol{v}_t \mu_t) = w_t \mu_t \text{ in } \mathscr{D}'(\mathbb{R}^d \times (0, 1)) \Big\}. \quad (8.71)$$

The Borel vector field $(\boldsymbol{v}, w)$ *realizing the minimum in* (8.71) *is uniquely determined* μ_I*-a.e. in* $\mathbb{R}^d \times (0, 1)$.

The discussion in [30] reveals however that there may be many geodesic curves, so in general μ_I is not unique. Indeed, the set of all geodesics connecting $\mu_0 = a_0 \delta_{x_0}$ and $\mu_1 = a_1 \delta_{x_1}$ with $a_0, a_1 > 0$ and $|x_1 - x_0| = \pi/2$ is infinite dimensional, see [30, Sect. 5.2].

Remark 8.19 (Inf-convolution of length distances) Here we want to explain why we may interpret the characterization (8.71) of HK as an infimal convolution (shortly inf-convolution) of the Kantorovich–Wasserstein distance $\mathsf{W_d}$ and the Hellinger–Kakutani distance He.

Let us first recall that if $\| \cdot \|_i$, $i = 1, 2$, are Hilbert norms on a linear space V, the classical inf-convolution for convex functionals induces the inf-convolution Hilbertian norm $\| \cdot \|_{\triangledown}$ defined by

$$\|v\|_{\triangledown}^2 := \inf \Big\{ \|v_1\|_1^2 + \|v_2\|_2^2 \ : \ v = v_1 + v_2 \Big\}.$$

When a finite dimensional manifold M is endowed with two Riemannian tensors g_1 and g_2, we can define the inf-convolution distance by computing the inf-convolution of the metric tensors in each tangent space. This leads to

the formula

$$d_\nabla^2(x_0, x_1) = \inf\left\{ \int_0^1 \left(|v_1(s)|_{g_1(x(s))}^2 + |v_2(s)|_{g_2(x(s))}^2\right) ds : \right.$$

$$x \in C^1([0, 1]; M), \, x(i) = x_i,$$

$$\left. \dot{x}(s) = v_1(s) + v_2(s) \text{ in } T_{x(s)}(M) \text{ for a.e. } s \in (0, 1) \right\}.$$

By optimizing the decomposition $\dot{x} = v_1 + v_2$ we easily find that the inf-convolution distance is generated by the metric tensor g_∇ whose dual g_∇^* is given by $g_\nabla^* = g_1^* + g_2^*$. This formula reflects the fact that the Legendre transform of an inf-convolution is the sum of the two Legendre transforms of the convoluted functionals. One can think that (8.71) exhibits a non-smooth, infinite dimensional example sharing the same structure. For another infinite-dimensional application we refer to [8, Eq. (16)].

When d_i, $i = 1, 2$, are length metrics on a given set Z, a purely metric inf-convolution $d_\nabla = d_1 \underset{\text{inf}}{\nabla} d_2$ respecting the local Hilbert-space structure reads

$$d_\nabla^2(z_1, z_2) := \liminf_{N \to \infty} \inf\left\{ N \sum_{i=1}^N \left(d_1^2(x_{i-1}, y_i) + d_2^2(y_i, x_i)\right) : \right.$$

$$\left. x_i, y_i \in Z, \, x_0 = z_1, \, y_N = z_2 \right\}.$$

One can expect that this inf-convolution applied to W_d and He exactly generates HK, namely $HK = W_d \underset{\text{inf}}{\nabla} He$.

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 $(\mathcal{M}(\mathbb{R}^d), HK)$ can be characterized by the system (cf. [30, Sect. 5])

$$\partial_t \mu_t + \nabla \cdot (\mu_t D_x \xi_t) = 4\xi_t \mu_t, \quad \partial_t \xi_t + \frac{1}{2}|D_x \xi_t|^2 + 2\xi_t^2 = 0. \qquad (8.72)$$

In order to give a precise meaning to (8.72) we first have to select an appropriate regularity for ξ_t. On the one hand we cannot expect C^1 smoothness for solutions of the Hamilton–Jacobi equation (8.72) (in contrast with subsolutions, that can be regularized as in Corollary 8.14) 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 (8.61). 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

$$\lim_{\substack{x',x''\to x \\ x'\neq x''}} \frac{f(x') - f(x'') - \mathrm{D}f(x)(x' - x'')}{|x' - x''|} = 0. \tag{8.73}$$

According to [15, Prop. 2.2.4] a locally Lipschitz function f is strictly differentiable at x if and only if the Clarke subgradient [15, Sect. 2.1] of f at x reduces to the singleton $\{\mathrm{D}f(x)\}$. In particular, denoting by $\boldsymbol{D} \subset \mathbb{R}^d$ the set where f is differentiable and denoting by κ_ε a smooth convolution kernel as in (8.58), Rademacher's theorem and [15, Thm. 2.5.1] yield

$$\lim_{\substack{x'\to x \\ x'\in \boldsymbol{D}}} \mathrm{D}f(x') = \mathrm{D}f(x), \quad \lim_{\varepsilon\downarrow 0} \mathrm{D}(f * \kappa_\varepsilon)(x) = \mathrm{D}f(x) \text{ for all } x \in \boldsymbol{D}. \tag{8.74}$$

In the proofs we will also need to deal with pointwise representatives of the time derivative of a locally Lipschitz function $\xi : \mathbb{R}^d \times (0, 1) \to \mathbb{R}$: if $D(\partial_t\xi)$ will denote the set (of full $\mathscr{L}^{d+1}$ measure) where ξ is differentiable w.r.t. time and $\widetilde{\partial_t\xi}$ the extension of $\partial_t\xi$ to 0 outside $D(\partial_t\xi)$, we set

$$(\partial_t\xi_t)_-(x) := \liminf_{\varepsilon\to 0} \left(\widetilde{\partial_t\xi_t} * \kappa_\varepsilon\right)(x), \quad (\partial_t\xi_t)^+(x) := \limsup_{\varepsilon\to 0} \left(\widetilde{\partial_t\xi_t} * \kappa_\varepsilon\right)(x). \tag{8.75}$$

It is not difficult to check that such functions are Borel; even if they depend on the specific choice of κ_ε, they will still be sufficient for our aims (a more robust definition would require the use of approximate limits).

We are now ready to characterize the set of all geodesic curves by giving a precise meaning to (8.72). The proof that the conditions (i)–(iv) below are sufficient for geodesic follows directly with the subsequent Lemma 8.21, whereas the proof of necessity is more involved and relies on the existence of optimal potentials ψ_1 for $\mathsf{LET} = \mathsf{HK}^2$ in Theorem 6.3(d), on the characterization of subsolutions of the generalized Hamilton–Jacobi equation in Theorem 8.11, and on the characterization of curves $t \mapsto \mu_t$ in $\mathrm{AC}^2\big([0, 1]; (\mathcal{M}(\mathbb{R}^d), \mathsf{HK})\big)$.

Theorem 8.20 *Let $\mu \in \mathrm{C}^0([0, 1]; \mathcal{M}(\mathbb{R}^d))$ be a weakly continuous curve. 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 $(0, 1)$ (equivalently, the map $(x, t) \mapsto \xi_t(x)$ is bounded and Lipschitz in $\mathbb{R}^d \times [a, b]$ for every compact subinterval $[a, b] \subset (0, 1)$),*
(ii) *ξ is strictly differentiable w.r.t. x at μ_I-a.e. $(x, t) \in \mathbb{R}^d \times (0, 1)$,*

(iii) ξ *satisfies*

$$\partial_t \xi_t + \frac{1}{2}\big|D_x\xi_t(x)\big|^2 + 2\xi_t^2(x) = 0 \quad \mathscr{L}^{d+1}\text{-}a.e. \text{ in } \mathbb{R}^d \times (0,1), \quad (8.76)$$

(iv) *and the curve* $(\mu_t)_{t\in[0,1]}$ *solves the continuity equation with reaction with the vector field* $(D_x\xi, 4\xi)$ *in every compact subinterval of* $(0,1)$, *i.e.*

$$\partial_t\mu_t + \nabla\cdot(\mu_t D_x\xi_t) = 4\xi_t\mu_t \quad \text{in } \mathscr{D}'(\mathbb{R}^d\times(0,1)), \quad (8.77)$$

then μ *is a geodesic w.r.t. the* HK *distance. Conversely, if* μ *is a geodesic then it is possible to find* $\xi \in \mathrm{Lip}_{\mathrm{loc}}((0,1);\mathrm{C}_b(\mathbb{R}^d))$ *that satisfies the properties* (i)–(iv) *above, is right differentiable w.r.t. t in* $\mathbb{R}^d\times(0,1)$, *and fulfills* (8.50b) *everywhere in* $\mathbb{R}^d\times(0,1)$.

Notice that (8.76) seems the weakest natural formulation of the Hamilton–Jacobi equation, in view of Rademacher's Theorem. The assumption of strict differentiability of ξ at μ_I-a.e. point provides an admissible vector field $D_X\xi$ for (8.77).

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 $(i), \ldots, (iv)$.

Since $D(\partial_t\xi)$ has full $\mathscr{L}^{d+1}$-measure in $\mathbb{R}^d\times(0,1)$, Fubini's Theorem shows that $N := \{t \in (0,1) : \mathscr{L}^d(\{x \in \mathbb{R}^d : (x,t) \notin D(\partial_t\xi)\}) > 0\}$ is $\mathscr{L}^1$-negligible. By (8.76) we get

$$(\partial_t\xi)_-(x) = -\limsup_{\varepsilon\downarrow 0}\left(\left(\frac{1}{2}|D_x\xi_t|^2 + 2\xi_t^2\right)*\kappa_\varepsilon\right)(x) \geq -\frac{1}{2}|D\xi_t|_a^2(x) - 2\xi_t^2(x) \tag{8.78}$$

for every $x \in \mathbb{R}^d$ and $t \in (0,1)\setminus N$.

We apply Lemma 8.21 below with $\boldsymbol{v} = D_x\xi$ and $w = 4\xi$: observing that $|D\xi_t|_a(x) = |D_x\xi_t(x)|$ at every point x of strict differentiability of ξ_t, we get, for all $0 < a < b < 1$,

$$2\int_{\mathbb{R}^d}\xi_b\,d\mu_b - 2\int_{\mathbb{R}^d}\xi_a\,d\mu_a \geq 2\int_{\mathbb{R}^d\times(a,b)}\left((\partial_t\xi)_- + |D_x\xi_t(x)|^2 + 4\xi_t^2(x)\right)d\mu_I$$

$$\overset{(8.78)}{=} 2\int_{\mathbb{R}^d\times(a,b)}\left(\frac{1}{2}|D_x\xi_t(x)|^2 + 2\xi_t^2(x)\right)d\mu_I$$

$$\overset{(8.69)}{\geq} \int_a^b |\mu_t'|^2\,dt \;\geq\; \frac{1}{b-a}\mathsf{HK}^2(\mu_a, \mu_b).$$

On the other hand, since $\mathbb{R}^d$ is a length space, Theorem 8.12 yields

$$\frac{1}{b-a}\mathsf{HK}^2(\mu_a, \mu_b) \geq 2 \int_{\mathbb{R}^d} \xi_b \, d\mu_b - 2 \int_{\mathbb{R}^d} \xi_a \, d\mu_a,$$

so that all the above inequalities are in fact identities and, hence,

$$\mathsf{HK}(\mu_a, \mu_b) = (b-a) \, |\mu'_t| \quad \mathscr{L}^1\text{-a.e. in } [a, b].$$

This 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 8.16 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 (8.63) and (8.66) 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 $\xi := -\frac{1}{2}\psi_1$ and $\xi_t := \mathscr{P}_t \xi$ for $t \in (0, 1)$. Even if we are considering more general initial data $\xi \in \mathrm{B}(\mathbb{R}^d; [-1/2, \infty])$ in (8.48), it is not difficult to check that the same statement of Theorem 8.11 holds in every subinterval $[a, b]$ with $0 < a < b < 1$ and

$$\lim_{t \downarrow 0} \mathscr{P}_t \xi(x) = \sup_{t>0} \mathscr{P}_t \xi(x) = \xi_*(x), \quad \text{where} \quad \xi_*(x) := \lim_{r \downarrow 0} \inf_{x' \in B_r(x)} \xi(x') \tag{8.79}$$

is the lower semicontinuous envelope of ξ. Moreover, setting

$$\xi_1(x) = \mathscr{P}_1 \xi(x) := \lim_{t \uparrow 1} \xi_t(x) = \inf_{0 < t < 1} \xi_t(x), \tag{8.80}$$

the function ξ_1 is upper semicontinuous with values in $[-\infty, 1/2]$ and the optimality properties stated in Theorem 6.3 d) yield

$$\frac{1}{2}\psi_2 \leq \xi_1 \quad \text{in } \mathbb{R}^d, \qquad \frac{1}{2}\psi_2 = \xi_1 \quad \mu_1\text{-a.e. in } \mathbb{R}^d. \tag{8.81}$$

By introducing the semigroup $\bar{\mathscr{P}}_t \xi := -\mathscr{P}_t(-\xi)$ and reversing time, we can define

$$\bar{\xi}_t := \bar{\mathscr{P}}_{1-t}\left(\tfrac{1}{2}\psi_2\right). \tag{8.82}$$

By using the link with the Hopf–Lax semigroup in $\mathfrak{C}$ given by Theorem 8.11, the optimality of (ψ_1, ψ_2), and arguing as in [50, Thm. 7.36] it is not difficult to check that

$$\bar{\xi}_t \leq \xi_t \quad \text{in } \mathbb{R}^d, \quad \bar{\xi}_0 = \xi_0 = -\frac{1}{2}\psi_1 \quad \mu_0\text{-a.e. in } \mathbb{R}^d. \tag{8.83}$$

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 $t \mapsto \int \xi_t \, d\mu_t$ and $t \mapsto \int \bar{\xi}_t \, d\mu_t$ are absolutely continuous in $(0, 1)$, Theorem 8.12(i) yields

$$\frac{d}{dt} \int \xi_t \, d\mu_t \leq \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1), \tag{8.84}$$

so that

$$\int \xi_b \, d\mu_b - \int \xi_a \, d\mu_a \leq \frac{b - a}{2}\mathsf{HK}^2(\mu_0, \mu_1).$$

Passing to the limit first as $a \downarrow 0$ and then as $b \uparrow 1$ by monotone convergence (notice that $\xi_t \leq 1/2$) and using optimality once again, we obtain

$$\mathsf{HK}^2(\mu_0, \mu_1) = \int \psi_1 \, d\mu_0 + \int \psi_2 \, d\mu_1 = 2\int \xi_1 \, d\mu_1 - 2\int \xi_0 \, d\mu_0$$

$$= \lim_{a \downarrow 0, b \uparrow 1} 2\left(\int \xi_b \, d\mu_b - \int \xi_a \, d\mu_a\right). \tag{8.85}$$

By (8.84) it follows that

$$\frac{d}{dt} \int \xi_t \, d\mu_t = \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) \quad \text{in } (0, 1). \tag{8.86}$$

Reversing time, the analogous argument yields

$$\frac{d}{dt} \int \bar{\xi}_t \, d\mu_t = \frac{1}{2}|\mu_t'|^2 = \frac{1}{2}\mathsf{HK}^2(\mu_0, \mu_1) \quad \text{in } (0, 1). \tag{8.87}$$

Hence, we have proved that the maps $t \mapsto \int \xi_t \, d\mu_t$ and $t \mapsto \int \bar{\xi}_t \, d\mu_t$ are affine in $[0, 1]$ and coincide at $t = 0$ and $t = 1$, which implies that

$$\int \xi_t \, d\mu_t = \int \bar{\xi}_t \, d\mu_t \quad \text{for every } t \in [0, 1]. \tag{8.88}$$

Recalling (8.83), 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 ξ_t is Lipschitz and semiconcave (thus everywhere superdifferentiable) for $t \in (0, 1)$ and since $\bar{\xi}_t$ is Lipschitz and semiconvex (thus everywhere subdifferentiable), we conclude that ξ_t is strictly differentiable in Z_t, and thus it satisfies conditions (i) and (ii).

Since (iii) is guaranteed by Theorem 8.11 ($\mathbb{R}^d$ is a length space), it remains to check (8.77). We apply the following Lemma 8.21 by observing that [3, Prop. 3.2, 3.3] and Theorem 8.11 yield

$$\limsup_{x' \to x} \partial_t^+ \xi_t(x') \leq \limsup_{x' \to x} \partial_t^- \xi_t(x') \leq \partial_t^- \xi_t(x), \quad \liminf_{x' \to x} \partial_t^+ \xi_t(x') \geq \partial_t^+ \xi_t(x);$$

since $\partial_t^- \xi_t(x) = \partial_t^+ \xi_t(x)$ μ_I-a.e. we get $(\partial_t \xi)^+ = (\partial_t \xi)_- = \partial_t^+ \xi$ μ_I-a.e. and therefore (8.89) holds with equality.

Recalling that $|D\xi_t|_a^2(x) = |D_x \xi_t(x)|^2$ at every point of Z_t, for every $0 < a < b < 1$ we have

$$\frac{b-a}{2} \mathsf{HK}^2(\mu_0, \mu_1) = \int_{\mathbb{R}^d} \xi_b \, d\mu_b - \int_{\mathbb{R}^d} \xi_a \, d\mu_a$$

$$\stackrel{(8.89)}{=} \int_{\mathbb{R}^d \times (a,b)} \left(\partial_t^+ \xi + D_x \xi \, \boldsymbol{v} + \xi w \right) d\mu_I$$

$$= \int_{\mathbb{R}^d \times (a,b)} \left(-\frac{1}{2} |D_x \xi_t|^2 - 2\xi_t^2 + D_x \xi \, \boldsymbol{v} + \xi w \right) d\mu_I$$

$$= \int_{\mathbb{R}^d \times (a,b)} \left(-\frac{1}{2} |D_x \xi_t - \boldsymbol{v}|^2 - 2(\xi_t - \frac{1}{4} w)^2 + \frac{1}{2} |\boldsymbol{v}|^2 + \frac{1}{8} w^2 \right) d\mu_I$$

$$\stackrel{(8.66)}{\leq} -\int_{\mathbb{R}^d \times (a,b)} \left(\frac{1}{2} |D_x \xi_t - \boldsymbol{v}|^2 + 2(\xi_t - \frac{1}{4} w)^2 \right) d\mu_I + \frac{1}{2} \int_a^b |\mu_t'|^2 \, dt$$

We deduce that $\boldsymbol{v} = D_x \xi$ and $w = 4\xi$ holds μ_I-a.e. □

The following lemma provides the "integration by parts" formulas that where used in the sufficiency and necessity part of the previous proof of Theorem 8.20. It is established by a suitable temporal and spatial smoothing, involving a smooth kernel κ_ε as in (8.58).

Lemma 8.21 *Let* $\mu \in \mathrm{AC}^2_{\mathrm{loc}}((0,1); (\mathcal{M}(\mathbb{R}^d), \mathsf{HK}))$ *satisfy the continuity equation with reaction* (8.63) *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) *of Theorem* 8.20, *then for all* $0 < a \leq b < 1$ *we have*

$$\int_{\mathbb{R}^d \times (a,b)} \left((\partial_t \xi)^+ + D_x \xi \, \boldsymbol{v} + \xi w \right) d\mu_I \geq \int_{\mathbb{R}^d} \xi_b \, d\mu_b - \int_{\mathbb{R}^d} \xi_a \, d\mu_a$$

$$\geq \int_{\mathbb{R}^d \times (a,b)} \left((\partial_t \xi)_- + D_x \xi \, \boldsymbol{v} + \xi w \right) d\mu_I, \tag{8.89}$$

where $(\partial_t \xi)^+, (\partial_t \xi)_-$ *are defined in terms of a space convolution kernel* κ_ε *as in* (8.75).

Proof We fix a compact subinterval $[a, b] \subset (0, 1)$, $b' \in (b, 1)$, and 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 (8.58) by setting $\xi^\varepsilon := \xi * \kappa_\varepsilon$ and perform a further regularization in time, viz.

$$\xi_t^{\varepsilon,\tau}(x) := \frac{1}{\tau} \int_0^\tau \xi_{t+r}^\varepsilon(x)\, dr, \quad 0 < \tau < b' - b. \tag{8.90}$$

Since $\xi^{\varepsilon,\tau} \in \mathrm{C}_b^1(\mathbb{R}^d \times [a, b])$ and μ is a weakly continuous solution to (8.63), we can argue as in [2, Lem. 8.1.2] and obtain, for every $\varepsilon > 0$ and $\tau \in (0, b'-b)$, the identity

$$\int_{\mathbb{R}^d} \xi_b^{\varepsilon,\tau}\, d\mu_b - \int_{\mathbb{R}^d} \xi_a^{\varepsilon,\tau}\, d\mu_a = \int_{\mathbb{R}^d \times (a,b)} \left(\partial_t \xi^{\varepsilon,\tau} + \mathrm{D}_x \xi^{\varepsilon,\tau}\, \boldsymbol{v} + \xi^{\varepsilon,\tau} w \right) d\mu_I. \tag{8.91}$$

We first pass to the limit as $\tau \downarrow 0$, observing that $\xi^{\varepsilon,\tau} \to \xi^\varepsilon$ uniformly because 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, using

$$\partial_t \xi_t^{\varepsilon,\tau}(x) = \frac{1}{\tau}(\xi_{t+\tau}^\varepsilon(x) - \xi_t^\varepsilon(x)) = \int_{\mathbb{R}^d} \frac{1}{\tau}(\xi_{t+\tau}^\varepsilon(x') - \xi_t^\varepsilon(x'))\kappa_\varepsilon(x - x')\, dx',$$

and the fact that $N := \{t \in (0, 1) : \mathscr{L}^d(\{x \in \mathbb{R}^d : (x, t) \notin D(\partial_t \xi)\}) > 0\}$ is $\mathscr{L}^1$-negligible by the theorems of Rademacher and Fubini, an application of Lebesgue's Dominated Convergence Theorem yields

$$\lim_{\tau \downarrow 0} \partial_t \xi_t^{\varepsilon,\tau}(x) = \partial_t \xi_t^\varepsilon(x) = ((\partial_t \xi) * \kappa_\varepsilon)(x) \quad \text{for every } x \in \mathbb{R}^d, \quad t \in (a, b) \setminus N. \tag{8.92}$$

Since $\mathbb{R}^d \times N$ is also μ_I-negligible, a further application of Lebesgue's Dominated Convergence Theorem yields

$$\int_{\mathbb{R}^d} \xi_b^\varepsilon\, d\mu_b - \int_{\mathbb{R}^d} \xi_a^\varepsilon\, d\mu_a = \int_{\mathbb{R}^d \times (a,b)} \left(\partial_t \xi^\varepsilon + \mathrm{D}_x \xi^\varepsilon\, \boldsymbol{v} + \xi^\varepsilon w \right) d\mu_I. \tag{8.93}$$

Now, (8.89) will be deduced by passing to the limit $\varepsilon \downarrow 0$ in (8.93) as follows. We observe that ξ^ε converges uniformly to ξ because ξ is bounded and Lipschitz. Moreover, since $\lim_{\varepsilon \downarrow 0} \mathrm{D}_x \xi_t^\varepsilon(x) = \mathrm{D}_x \xi_t(x)$ at every point $x \in \mathbb{R}^d$ where ξ_t is strictly differentiable, we obtain

$$|\mathrm{D}_x \xi^\varepsilon\, \boldsymbol{v}| \le L |\boldsymbol{v}| \in \mathrm{L}^1(\mathbb{R}^d \times (a, b); \mu_I) \quad \text{and}$$
$$\lim_{\varepsilon \downarrow 0} \mathrm{D}_x \xi^\varepsilon = \mathrm{D}_x \xi \quad \mu_I\text{-a.e. in } \mathbb{R}^d \times [a, b],$$

so that

$$\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}^d} \xi^\varepsilon_{a,b} \, \mathrm{d}\mu_{a,b} = \int_{\mathbb{R}^d} \xi_{a,b} \, \mathrm{d}\mu_{a,b},$$

$$\int_{\mathbb{R}^d \times (a,b)} \left(\mathrm{D}_x \xi^\varepsilon \, \boldsymbol{v} + \xi^\varepsilon w \right) \mathrm{d}\mu_I = \int_{\mathbb{R}^d \times (a,b)} \left(\mathrm{D}_x \xi \, \boldsymbol{v} + \xi w \right) \mathrm{d}\mu_I.$$

Finally, since $\partial_t \xi^\varepsilon_t$ is also uniformly bounded, Fatou's Lemma yields

$$\limsup_{\varepsilon \downarrow 0} \int_{\mathbb{R}^d \times (a,b)} \partial_t \xi^\varepsilon_t \, \mathrm{d}\mu_I \leq \int_{\mathbb{R}^d \times (a,b)} (\partial_t \xi_t)^+ \, \mathrm{d}\mu_I,$$

$$\liminf_{\varepsilon \downarrow 0} \int_{\mathbb{R}^d \times (a,b)} \partial_t \xi^\varepsilon_t \, \mathrm{d}\mu_I \geq \int_{\mathbb{R}^d \times (a,b)} (\partial_t \xi_t)_- \, \mathrm{d}\mu_I.$$

Thus, (8.89) follows from (8.93). □

8.7 Contraction properties: convolution and Heat equation in RCD(0, ∞) 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:

$$\mathsf{HK}(f_\sharp \mu_1, f_\sharp \mu_2) \leq \mathsf{HK}(\mu_1, \mu_2). \tag{8.94}$$

Proof It is sufficient to observe that the map $\mathfrak{f} : \mathfrak{C}_X \mapsto \mathfrak{C}_Y$ defined by $\mathfrak{f}([x, r]) := [f(x), r]$ satisfies $\mathsf{d}_{\mathfrak{C}_Y}(\mathfrak{f}([x_1, r_1]), \mathfrak{f}([x_2, r_2])) \leq \mathsf{d}_{\mathfrak{C}_X}([x_1, r_1], [x_2, r_2])$ for every $[x_i, r_i] \in \mathfrak{C}_X$. Thus $\mathfrak{f}_\sharp$ is 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 hence $f_\sharp$ satisfies (8.94). □

A second application concerns convolutions in $\mathbb{R}^d$.

Theorem 8.23 *Let* $X = \mathbb{R}^d$ *with the Euclidean distance and let* $\nu \in \mathcal{M}(\mathbb{R}^d)$. *Then the map* $\mu \mapsto \mu * \nu$ *is contractive w.r.t.* HK *if* $\nu(\mathbb{R}^d) = 1$ *and, more generally,*

$$\mathsf{HK}^2(\mu_1 * \nu, \mu_2 * \nu) \leq \nu(\mathbb{R}^d)\mathsf{HK}^2(\mu_1, \mu_2) \quad \text{for } \mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d). \quad (8.95)$$

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 (7.31), it follows that if $\nu = \sum_k a_k \delta_{x_k}$ for some $a_k \geq 0$, then

$$\begin{aligned}
\mathsf{HK}^2(\mu_1 * \nu, \mu_2 * \nu) &= \mathsf{HK}^2\left(\sum_k a_k \mu_1 * \delta_{x_k}, \sum_k a_k \mu_2 * \delta_{x_k}\right) \\
&\leq \sum_k a_k \mathsf{HK}^2(\mu_1 * \delta_{x_k}, \mu_2 * \delta_{x_k}) \\
&= \sum_k a_k \mathsf{HK}^2(\mu_1, \mu_2) = \nu(\mathbb{R}^d)\mathsf{HK}^2(\mu_1, \mu_2).
\end{aligned}$$

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 by Theorem 7.15, we obtain (8.95). $\square$

An easy application of the previous result is the contraction property of the (adjoint) Heat semigroup $(P_t^*)_{t \geq 0}$ 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 $(P_t)_{t \geq 0}$ generated by
(A) the Heat equation on a open convex domain $\Omega \subset \mathbb{R}^d$ with homogeneous Neumann conditions

$$\partial_t u = \Delta u \quad \text{in } \Omega \times (0, \infty), \qquad \partial_n u = 0 \quad \text{on } \partial\Omega \times (0, \infty),$$

(B) the Heat equation on a complete Riemannian manifold $(\mathbb{M}^d, g)$ with non-negative 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
(C) the Fokker-Planck equation in $\mathbb{R}^d$ generated by the gradient of a convex potentials $V : \mathbb{R}^d \to \mathbb{R}$, viz.

$$\partial_t u = \Delta u - \nabla \cdot (u \, \mathrm{D}V) \quad \text{in } \mathbb{R}^d \times (0, \infty).$$

Theorem 8.24 *Let* (X, d, m) *be a complete and separable metric-measure space with nonnegative Riemannian Ricci Curvature, i.e. satisfying the* RCD$(0, \infty)$ *condition, and let* $(P_t^*)_{t \geq 0} : \mathcal{M}(X) \to \mathcal{M}(X)$ *be the Heat semigroup in the measure setting. Then*

$$\mathsf{HK}(P_t^* \mu_1, P_t^* \mu_2) \leq \mathsf{HK}(\mu_1, \mu_2) \quad \text{for all } \mu_1, \mu_2 \in \mathcal{M}(X) \text{ and } t > 0. \quad (8.96)$$

Proof Recall that in RCD$(0, \infty)$ metric measure spaces the L^2-gradient flow of the Cheeger energy induces a symmetric Markov semigroup $(P_t)_{t \geq 0}$ in $L^2(X, m)$ [4, Sect. 6.1], 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 (cf. [4, Thm. 6.2] or [5, Cor. 4.18])

$$|\mathrm{D}_X P_t f|^2(x) \leq P_t(|\mathrm{D}_X f|^2)(x) \quad \text{for } f \in \mathrm{Lip}_b(X), \ x \in X, \ t \geq 0. \quad (8.97)$$

Its adjoint $(P_t^*)_{t \geq 0}$ 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 $F(s) = U_1(s) = s \log s - s + 1$ and defines a semigroup in $\mathcal{M}(X)$ by the formula

$$\int_X f \, \mathrm{d}(P_t^* \mu) = \int_X P_t f \, \mathrm{d}\mu \quad \text{for every } f \in \mathrm{B}_b(X) \text{ and } \mu \in \mathcal{M}(X). \quad (8.98)$$

In order to prove (8.96) we use (8.54) (RCD-spaces satisfy the length property, [4, Thm. 5.1]) 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

$$\partial_\theta \psi_\theta + \frac{1}{4} |\mathrm{D}_X \psi_\theta|^2 + \psi_\theta^2 \leq 0 \quad \text{in } X \times (0, 1). \quad (8.99)$$

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 $\mathrm{C}^1([0, 1]; \mathrm{Lip}_b(X))$. Now, (8.97) and the Markov property yield

$$|\mathrm{D}_X P_t \psi_\theta|^2(x) \leq P_t(|\mathrm{D}_X \psi_\theta|^2)(x),$$
$$(P_t \psi_\theta)^2(x) \leq P_t(\psi_\theta^2)(x) \quad \text{for } x \in X, \ \theta \in [0, 1], \ t \geq 0.$$

Thus, for every $t \geq 0$ we obtain

$$\partial_\theta \psi_{\theta,t} + \frac{1}{4} |\mathrm{D}_X \psi_{\theta,t}|^2 + \psi_{\theta,t}^2 \leq 0 \quad \text{in } X \times (0, 1),$$

and therefore

$$\int_X \psi_1 \, d(P_t^*)\mu_1 - \int_X \psi_0 \, d(P_t^*)\mu_0$$

$$= \int_X P_t \psi_1 \, d\mu_1 - \int_X P_t \psi_0 \, d\mu_0 \leq \mathsf{HK}^2(\mu_1, \mu_0).$$

We conclude by taking the supremum with respect to all the subsolutions of (8.99) in $C^1([0, 1]; \mathrm{Lip}_b(X))$ and applying (8.54). □

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 [30] 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 $\partial_t \boldsymbol{u}_t + \nabla \cdot \left(\mathbb{M}_{\mathrm{diff}}(\boldsymbol{u}_t)\Xi_t\right) = \mathbb{K}_{\mathrm{react}}(\boldsymbol{u}_t)\boldsymbol{\xi}_t$. The central question was and still is the understanding of diffusion equations with reactions in the gradient-flow form $\partial_t \boldsymbol{u} = \nabla \cdot \left(\mathbb{M}_{\mathrm{diff}}(\boldsymbol{u})\nabla\delta\mathcal{F}(\boldsymbol{u})\right) - \mathbb{K}_{\mathrm{react}}(\boldsymbol{u})\delta\mathcal{F}(\boldsymbol{u})$, see [30, 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 (1.29) 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

$$\partial_t \xi_t + \frac{\alpha}{2}|\mathsf{D}_x \xi_t|^2 + \frac{\beta}{2}\xi_t^2 = 0.$$

As in the other parts of this paper, we restrict to the case $\alpha = 1$ and $\beta = 4$ subsequently, but refer to [30] for the general case. Thus, the conjectured characterization (8.54) 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. (8.48)) 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 $\Phi(z, R) = \frac{1}{2}\left(1 - \frac{A(R)}{1+2z}\right)$, where $A(R) = \cos^2\left(R \wedge (\pi/2)\right)$. 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

$$c(x_0, x_1) = -\log A(|x_0-x_1|) = \begin{cases} -2\log\left(\cos|x_0-x_1|\right) & \text{for } |x_0-x_1| < \pi/2, \\ \infty & \text{otherwise,} \end{cases}$$

for the first time and obtained the characterization (1.7), namely

$$HK(\mu_0, \mu_1)^2 = D(\mu_0, \mu_1) = \sup\left\{\mathscr{D}(\varphi_0, \varphi_1|\mu_0, \mu_1) : \varphi_0 \oplus \varphi_1 \leq c\right\}.$$

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 $LET(\mu_0, \mu_1) = \min \mathscr{E}(\gamma|\mu_0, \mu_1)$ was easily obtained, it was not clear at all, why and how HK defined via $HK^2(\mu_0, \mu_1) = \min \mathscr{E}(\cdot|\mu_0, \mu_1)$ 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. Sect. 8.4) and the derivation of the geodesic equations in $\mathbb{R}^d$ (cf. Sect. 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. Ambrosio, L., Erbar, M., Savaré, G.: Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces. Nonlinear Anal. **137**, 77–134 (2016)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd ed. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel (2008)
3. Ambrosio, L., Gigli, N., Savaré, G.: Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. Invent. Math. **195**, 289–391 (2014)
4. Ambrosio, L., Gigli, N., Savaré, G.: Metric measure spaces with Riemannian Ricci curvature bounded from below. Duke Math. J. **163**, 1405–1490 (2014)
5. Ambrosio, L., Gigli, N., Savaré, G.: Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. Ann. Probab. **43**, 339–404 (2015)
6. Ambrosio, L., Mondino, A., Savaré, G.: Nonlinear diffusion equations and curvature conditions in metric measure spaces, to appear on Memoirs of the A.M.S. (2015). arXiv:1509.07273
7. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**, 375–393 (2000)
8. Brenier, Y., Otto, F., Seis, C.: Upper bounds on coarsening rates in demixing binary viscous liquids. SIAM J. Math. Anal. **43**, 114–134 (2011)
9. Brézis, H.: 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)
10. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry, vol 33 of Graduate Studies in Mathematics. American Mathematical Society, Providence (2001)
11. Burago, Y., Gromov, M., Perel'man, G.: A. D. Aleksandrov spaces with curvatures bounded below. Uspekhi Mat. Nauk **47**, 3–51, 222 (1992)
12. Cheeger, J.: Differentiability of Lipschitz functions on metric measure spaces. Geom. Funct. Anal. **9**, 428–517 (1999)
13. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: An interpolating distance between optimal transport and Fisher–Rao metrics. Found Comput Math (2016). https://doi.org/10.1007/s10208-016-9331-y
14. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Unbalanced optimal transport: geometry and Kantorovich formulation (2015). arXiv:1508.05216v1
15. Clarke, F .H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983). A Wiley-Interscience Publication
16. Cordero-Erausquin, D., McCann, R.J., Schmuckenschläger, M.: A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. Invent. Math. **146**, 219–257 (2001)
17. Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. **2**, 299–318 (1967)
18. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Part. Differ. Equ. **34**, 193–231 (2009)

19. Dudley, R .M.: Real Analysis and Probability, vol. 74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2002). Revised reprint of the 1989 original
20. Ekeland, I., Temam, R.: Analyse Convexe et Problèmes Variationnels. Dunod, Gauthier-Villars (1974)
21. Erbar, M., Kuwada, K., Sturm, K.-T.: On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. Invent. Math. **201**, 993–1071 (2015)
22. Hellinger, E.: Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. J. Reine Angew. Math. **136**, 210–271 (1909)
23. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex analysis and minimization algorithms. I, vol. 305 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Berlin (1993)
24. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM J. Math. Anal **29**, 1–17 (1998). (electronic)
25. Kakutani, S.: On equivalence of infinite product measures. Ann. Math. (2) **49**, 214–224 (1948)
26. Kellerer, H.G.: Duality theorems for marginal problems. Z. Wahrsch. Verw. Gebiete **67**, 399–432 (1984)
27. Kondratyev, S., Monsaingeon, L., Vorotnikov, D.: A new optimal transport distance on the space of finite Radon measures. Adv. Differ. Equ. **21**, 1117–1164 (2016)
28. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
29. LeCam, L.: Convergence in distribution of stochastic processes. Univ. Calif. Publ. Stat. **2**, 207–236 (1957)
30. Liero, M., Mielke, A., Savaré, G.: Optimal transport in competition with reaction—the Hellinger-Kantorovich distance and geodesic curves. SIAM J. Math. Anal. **48**, 2869–2911 (2016)
31. Liese, F., Vajda, I.: On divergences and informations in statistics and information theory. IEEE Trans. Inf. Theory **52**, 4394–4412 (2006)
32. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory **37**, 145–151 (1991)
33. Lisini, S.: Characterization of absolutely continuous curves in Wasserstein spaces. Calc. Var. Part. Differ. Equ. **28**, 85–120 (2007)
34. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. Ann. Math. (2) **169**, 903–991 (2009)
35. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**, 153–179 (1997)
36. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Part. Differ. Equ. **26**, 101–174 (2001)
37. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**, 361–400 (2000)
38. Piccoli, B., Rossi, F.: Generalized Wasserstein distance and its application to transport equations with source. Arch. Ration. Mech. Anal. **211**, 335–358 (2014)
39. Piccoli, B., Rossi, F.: On properties of the Generalized Wasserstein distance. Arch. Ration. Mech. Anal. **222**, 1339–1365 (2016)
40. Rachev, S .T., Rüschendorf, L.: Mass Transportation Problems. Vol. I–II, Probability and Its Applications. Springer, New York (1998)
41. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
42. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Springer, New York (2015)
43. Schachermayer, W., Teichmann, J.: Characterization of optimal transport plans for the Monge–Kantorovich problem. Proc. Am. Math. Soc. **137**, 519–529 (2009)

44. Schwartz, L.: 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
45. Simons, S.: Minimax and Monotonicity, vol. 1693 of Lecture Notes in Mathematics. Springer, Berlin (1998)
46. Sturm, K.-T.: Metric spaces of lower bounded curvature. Expos. Math. **17**, 35–47 (1999)
47. Sturm, K.-T.: On the geometry of metric measure spaces. I. Acta Math. **196**, 65–131 (2006)
48. Sturm, K.-T.: On the geometry of metric measure spaces. II. Acta Math. **196**, 133–177 (2006)
49. Villani, C.: Topics in Optimal Transportation, vol. 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence (2003)
50. Villani, C.: Optimal Transport. Old and New, vol. 338 of Grundlehren der Mathematischen Wissenschaften. Springer, Berlin (2009)