A RELAXATION VIEWPOINT TO UNBALANCED OPTIMAL TRANSPORT: DUALITY, OPTIMALITY AND MONGE FORMULATION

GIUSEPPE SAVARÉ AND GIACOMO ENRICO SODINI

ABSTRACT. We present a general convex relaxation approach to study a wide class of Unbalanced Optimal Transport problems for finite non-negative measures with possibly different masses. These are obtained as the lower semicontinuous and convex envelope of a cost for non-negative Dirac masses. New general primal-dual formulations, optimality conditions, and metric-topological properties are carefully studied and discussed.

CONTENTS

1. INTRODUCTION

The problem of extending Optimal Transport methods to pairs of unbalanced non-negative measures has been considered in a large number of works with different techniques and different aims.

For what concerns dynamical formulations, many models inspired by the Benamou-Brenier's [BB00] fluid dynamic formulation of the classical Kantorovich-Rubinstein-Wasserstein Optimal Transport metric have been proposed, see for example [KMV16; Maa+15; LM15; PR14; PR16; Chi+18a; Chi+18b; LMS18; LM19]. In such works, the authors consider source terms in the continuity equation, thus leading to gain/loss of mass during the evolution. The models proposed differ in the kind of source or penalization. We refer to [Chi+18a] where a more detailed description of these models is given.

Static formulations of the Unbalanced Optimal Transport problem were proposed already by Kantorovich and Rubinstein [KR58] and subsequently extended by Hanin [Han99] (see also the dual norm in [Han92]). These approaches can be thought as a classical Optimal Transport problem where a fraction of the mass is allowed to go (or come from) a point at infinity (see also [FG10; Gui02]). More recent approaches are given by the so called optimal partial transport [CM10; Fig10], which was previously related to image retrieval [PW08; RGT97].

Optimal partial transport (see [Chi+18a]) is in turn also related to [PR14; PR16], since this latter works also provide a dynamic formulation of optimal partial transport. We also mention

---

2020 *Mathematics Subject Classification.* Primary: 49Q22, 28A33 ; Secondary: 49K27.
Key words and phrases. Unbalanced Optimal Transport, Kantorovich-Monge problem, Optimality conditions.

that [PR14; PR16] are also connected to [Ben03] where it was proposed to change the marginal constraints and to add a penalization term.

The important case of the Hellinger-Kantorovich metric (a.k.a. Wasserstein-Fisher-Rao), somehow interpolating the Wasserstein and the Hellinger metrics, has been proposed independently in [KMV16; Chi+18a; Chi+18b; LMS18], with different equivalent characterizations.

A very useful one [LMS18] involves an entropic relaxation of the marginal constraints in the classical static formulation of Optimal Transport, it provides a paradigmatic example of the wide class of Optimal Entropy-Transport problems, and suggests the crucial role of the so-called cone geometry.

[Chi+18b] introduces a general class of Unbalanced Optimal Transport problems in compact subsets of some Euclidean space induced by a sublinear cost function depending on positions and masses, using a suitable static semi-coupling formulation and proving a general duality result, with interesting applications to various dynamic problems.

It turns out that all these different viewpoints are specific examples of a general construction based on convex relaxation, which we want to exploit in the present paper.

A convex relaxation viewpoint. Adopting the same approach of [SS22] used for the classical Optimal Transport problem and extending some ideas already contained in [LMS18], the aim of this work is to define a general class of Unbalanced Optimal Transport problems between non-negative measures as the *convex and lower semicontinuous envelope* of a cost initially defined only between weighted Dirac masses, and then to study the corresponding duality formulas, optimality conditions and metric-topologial properties.

Let us briefly recall the result of [SS22] which is the starting point for the present work; given a pair of completely regular spaces X_1 and X_2, every proper and lower semicontinuous cost function $\mathsf{c} : \mathsf{X}_1 \times \mathsf{X}_2 \to [0, +\infty]$ can be naturally lifted to a singular cost functional $\mathcal{F}_\mathsf{c} : \mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2) \to [0, +\infty]$, where $\mathcal{M}(\mathsf{X}_i)$ denote the space of signed and finite Radon measures in X_i. $\mathcal{F}_\mathsf{c}$ is finite only between balanced pairs of Dirac masses and can be defined as

$$\mathcal{F}_\mathsf{c}(\mu_1, \mu_2) := \begin{cases} r\mathsf{c}(x_1, x_2) & \text{if} \quad \begin{array}{l} \mu_1 = r\delta_{x_1},\, \mu_2 = r\delta_{x_2}, \\ x_1 \in \mathsf{X}_1, x_2 \in \mathsf{X}_2,\, r \geq 0, \end{array} \\ +\infty & \text{elsewhere.} \end{cases} \tag{1.1}$$

By [SS22, Theorem 4.4] the convex and lower semicontinuous (w.r.t. the product weak topology) envelope of $\mathcal{F}_\mathsf{c}$ coincides with the Optimal Transport functional

$$\mathsf{OT}_\mathsf{c}(\mu_1, \mu_2) := \begin{cases} \inf \left\{ \displaystyle\int \mathsf{c} \,\mathrm{d}\gamma : \gamma \in \Gamma(\mu_1, \mu_2) \right\} & \text{if } \mu_i \in \mathcal{M}_+(X_i),\ \mu_1(\mathsf{X}_1) = \mu_2(\mathsf{X}_2), \\ +\infty & \text{elsewhere,} \end{cases} \tag{1.2}$$

where $\mathcal{M}_+(\mathsf{X}_i)$ denotes the cone of nonnegative (finite Radon) measures in $\mathcal{M}(\mathsf{X}_i)$ and $\Gamma(\mu_1, \mu_2)$ is the subset of couplings between μ_1 and μ_2, i.e. measures in $\mathcal{M}_+(\mathsf{X}_1 \times \mathsf{X}_2)$ with marginals μ_1 and μ_2 respectively. Clearly $\Gamma(\mu_1, \mu_2)$ is empty if $\mu(\mathsf{X}_1) \neq \mu_2(\mathsf{X}_2)$.

In order to extend this construction to the case of non-negative measures with possibly different masses, we represent weighted Dirac masses in X of the form $r\delta_x$, $r \geq 0$, as points (x, r) of the so called geometric cone $\mathfrak{C}[\mathsf{X}] := \big(\mathsf{X} \times [0, +\infty)\big)/\sim$ (see Section 2.1), where the equivalence relation $\sim$ identifies all the points of the form $(x, 0)$, $x \in \mathsf{X}$, which correspond to the null measure in X. A cost on weighted Dirac masses can thus be expressed by a function

$$\mathsf{H} : \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2] \to [0, +\infty],$$

which we will assume to be proper, lower semicontinuous and *radially* 1-*homogeneous*, in the sense that the map

$$\mathbb{R}_2^+ \ni (r_1, r_2) \mapsto \mathsf{H}([x_1, r_1], [x_2, r_2])$$

is 1-homogeneous for every fixed $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$. While the properness and the lower semicontinuity assumptions are natural, the 1-homogeneity assumption deserves a comment: from a modeling point of view we are saying that moving $mr_1\delta_{x_1}$ to $mr_2\delta_{x_2}$ costs exactly m times moving $r_1\delta_{x_1}$ to $r_2\delta_{x_2}$. In analogy with (1.1) we can define the unbalanced singular cost $\mathscr{S}_\mathsf{H} : \mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2) \to [0, +\infty]$ as

$$\mathscr{S}_\mathsf{H}(\mu_1, \mu_2) := \begin{cases} \mathsf{H}([x_1, r_1], [x_2, r_2]) & \text{if} \quad \begin{array}{l} \mu_1 = r_1\delta_{x_1},\ \mu_2 = r_2\delta_{x_2}, \\ x_1 \in \mathsf{X}_1, x_2 \in \mathsf{X}_2,\ r_1, r_2 \geq 0, \end{array} \\ +\infty & \text{elsewhere,} \end{cases}$$

and look for the largest convex and lower semicontinuous functional $\mathscr{U}_\mathsf{H}$ on $\mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2)$ dominated by $\mathscr{S}_\mathsf{H}$. It turns out that such a functional admits an explicit characterization which resembles (1.2) via the formula

$$\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) := \inf \left\{ \int_{\mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2]} \mathsf{H} \, \mathrm{d}\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \right\}, \tag{1.3}$$

where $\mathfrak{H}^1(\mu_1, \mu_2)$ is the set of *homogeneous couplings*, i.e. measures $\boldsymbol{\alpha}$ on $\mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2]$ with 1-*homogeneous marginals* μ_1 and μ_2 (see (2.12)): $\boldsymbol{\alpha}$ belongs to $\mathfrak{H}^1(\mu_1, \mu_2)$ if

$$\begin{cases} \mu_1(A_1) = \displaystyle\int_{A_1 \times [0,+\infty) \times \mathfrak{C}[\mathsf{X}_2]} r_1 \, \mathrm{d}\boldsymbol{\alpha}(x_1, r_1; x_2, r_2), \\ \mu_2(A_2) = \displaystyle\int_{\mathfrak{C}[\mathsf{X}_1] \times A_2 \times [0,+\infty)} r_2 \, \mathrm{d}\boldsymbol{\alpha}(x_1, r_1; x_2, r_2), \end{cases} \quad \text{for every Borel subsets } A_i \subset \mathsf{X}_i.$$

This choice has been inspired by [LMS18], where the authors prove that the class of Optimal Entropy-Transport problems can be formulated precisely as in (1.3) for a suitable choice of the function H (see in particular [LMS18, Definition 5.1]).

Structural properties of Unbalanced Optimal Transport problems. The following result collects some of the fundamental properties of the unbalanced framework (see Theorems 3.4, 3.5 and 3.15), in parallel to similar results of the classical Optimal Transport theory.

Theorem 1.1. *Let $\mathsf{X}_1, \mathsf{X}_2$ be completely regular spaces and let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2] \to [0, +\infty]$ be a proper, radially 1-homogeneous and lower semicontinuous function.*

(1) For every $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ such that $\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) < +\infty$, there exists an optimal homogeneous coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ such that

$$\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) = \int_{\mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2]} \mathsf{H} \, \mathrm{d}\boldsymbol{\alpha}.$$

(2) $\mathscr{U}_\mathsf{H}$ is a lower semicontinuous convex function and satisfies

$$\mathscr{U}_\mathsf{H}(r_1\delta_{x_1}, r_2\delta_{x_2}) = \overline{\mathrm{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) \leq \mathsf{H}([x_1, r_1], [x_2, r_2])$$

for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ and every $(r_1, r_2) \in \mathbb{R}^2_+$, where $\overline{\mathrm{co}}\,(\mathsf{H})$ is the l.s.c. convex envelope of $\mathsf{H}([x_1, \cdot], [x_2, \cdot])$ with respect to the variables $(r_1, r_2) \in \mathbb{R}^2_+$. If, in addition, H is also radially convex (i.e. the map $(r_1, r_2) \mapsto \mathsf{H}([x_1, r_1], [x_2, r_2])$ is convex for every fixed $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$), then the above inequality is an equality.

(3) $\mathscr{U}_\mathsf{H}$ is the convex l.s.c. envelope of $\mathscr{S}_\mathsf{H}$ and admits the dual formulation

$$\begin{aligned} \mathscr{U}_\mathsf{H}(\mu_1, \mu_2) &= \mathscr{U}_{\overline{\mathrm{co}}(\mathsf{H})}(\mu_1, \mu_2) = \overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H})\,(\mu_1, \mu_2) \\ &= \sup \left\{ \int_{\mathsf{X}_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \, \mathrm{d}\mu_2 \mid (\varphi_1, \varphi_2) \in \Phi_\mathsf{H} \right\} \end{aligned} \tag{1.4}$$

for every $(\mu_1,\mu_2)\in\mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$, *where*

$$\begin{aligned}\Phi_{\mathsf{H}}:=\Big\{(\varphi_1,\varphi_2)\in \mathrm{C}_b(\mathsf{X}_1)\times\mathrm{C}_b(\mathsf{X}_2):\varphi_1(x_1)r_1+\varphi_2(x_2)r_2\le\mathsf{H}([x_1,r_1],[x_2,r_2])\\ \textit{for every } (x_1,x_2)\in\mathsf{X}_1\times\mathsf{X}_2,\ r_1,r_2\ge 0\Big\}.\end{aligned}\tag{1.5}$$

Notice that, if the cost function H is given by

$$\mathsf{H}([x_1,r_1],[x_2,r_2]):=\begin{cases} r\mathsf{c}(x_1,x_2) & \text{if } r_1=r_2=r\ge 0,\\ +\infty & \text{elsewhere,}\end{cases}$$

for some proper and lower semicontinuous function $\mathsf{c}:\mathsf{X}_1\times\mathsf{X}_2\to[0,+\infty]$, then $\mathscr{U}_{\mathsf{H}}=\mathsf{OT}_{\mathsf{c}}$, $\mathscr{S}_{\mathsf{H}}=\mathcal{F}_{\mathsf{c}}$, and we recover precisely the analogous results for the classical Optimal Transport theory (see in particular [SS22, Theorems 3.5 and 4.4]). The relaxation formula (1.4) extends the corresponding one of [Chi+18b, Proposition 3.11], showing in particular that the primal formulation (1.3) in terms of homogeneous couplings is equivalent to the semi-coupling approach of [Chi+18b]. (1.4) has also nice applications to the structure of the Hellinger-Kantorovich metric, see Remark 3.18.

Along the same lines of the classical Optimal Transport theory, it is natural to investigate the *existence of optimal potentials maximizing* (1.4) *and thus solving the dual problem.* We first focus on the class of continuous functions, when $\mathsf{X}_1,\mathsf{X}_2$ are compact and metrizable and H is finite and continuous on the whole product cone $\mathfrak{C}[\mathsf{X}_1]\times\mathfrak{C}[\mathsf{X}_2]$, radially 1-homogeneous and convex. In this case a sufficient condition relies on the singular behaviour of the normal derivative of H at the boundary of the product cone as in the fundamental example

$$\mathsf{H}([x_1,r_1],[x_2,r_2]):=r_1+r_2-2\sqrt{r_1r_2}\,\mathrm{e}^{-|x_1-x_2|^2/2},\quad x_i\in\mathbb{R}^d,\ r_i\ge 0.\tag{1.6}$$

We refer to Section 4 for a detailed discussion of these hypotheses and Appendix B for an alternative set of assumptions (i.e. when H is finite and continuous on an open conical domain). In both these situations it is possible to define the analogue of the celebrated c-transform for a pair $(\varphi_1,\varphi_2)\in\Phi_{\mathsf{H}}$ as

$$\varphi_1^{\mathsf{H}}(x_2):=\inf_{x_1\in\mathsf{X}_1}\inf_{\alpha\ge 0}\Big\{\mathsf{H}([x_1,\alpha],[x_2,1])-\alpha\varphi_1(x_1)\Big\},\quad x_2\in\mathsf{X}_2,$$

$$\varphi_1^{\mathsf{HH}}(x_1):=\inf_{x_2\in\mathsf{X}_2}\inf_{\alpha\ge 0}\Big\{\mathsf{H}([x_1,1],[x_2,\alpha])-\alpha\varphi_1^{\mathsf{H}}(x_2)\Big\},\quad x_1\in\mathsf{X}_1,$$

and prove that the transformed potentials are equicontinuous so that the use of a compactness argument in the space $\mathrm{C}(\mathsf{X}_1)\times\mathrm{C}(\mathsf{X}_2)$ is possible. This will produce an optimal pair of potentials, see Theorems 4.7 and B.3.

Optimality conditions and relaxed duality. In this unbalanced setting, general optimality conditions still involve the notion of cyclical monotonicity: a subset $\Gamma\subset\mathfrak{C}[\mathsf{X}_1]\times\mathfrak{C}[\mathsf{X}_2]$ is H-cyclically monotone if for every finite family of points $\{(\mathfrak{y}_1^i,\mathfrak{y}_2^i)\}_{i=1}^N\subset\Gamma$ and every permutation σ of $\{1,\dots,N\}$ it holds

$$\sum_{i=1}^N\mathsf{H}(\mathfrak{y}_1^i,\mathfrak{y}_2^i)\le\sum_{i=1}^N\mathsf{H}(\mathfrak{y}_1^i,\mathfrak{y}_2^{\sigma(i)}).$$

Similarly to the classical Optimal Transport case, *every optimal homogeneous coupling is concentrated on a* H*-cyclically monotone set* Γ (see Proposition 5.2) which in addition is *a radial convex cone* in the sense that

$$([x_1,r_1^i],[x_2,r_2^i])\in\Gamma,\ \lambda_i\ge 0,\ i=1,2\ \Rightarrow\Big([x_1,\sum_{i=1}^2\lambda_ir_1^i],[x_2,\sum_{i=1}^2\lambda_ir_2^i]\Big)\in\Gamma.$$

This further property comes from the radial homogeneity and convexity assumption on H.

Formulating a converse statement to the above proposition, thus involving sufficient optimality conditions, requires additional assumptions on the compatibility between the radial cone Γ, on which an admissible homogeneous coupling $\boldsymbol{\alpha}$ between measures $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ is concentrated, and the cost function H: in Section 5.1 we study the natural directed graph structures induced by H and Γ, similarly to what has been done in the classical Optimal Transport theory [Bei+09; BC10] for possibly infinite costs. This leads to the notion of H-connectedness: Γ is H-connected if, whenever $\mathfrak{y}_1, \mathfrak{y}_1' \in \pi^1(\Gamma)$ (the projection of Γ on the first cone $\mathfrak{C}[\mathsf{X}_1]$), we can find a sequence of points $\{\mathfrak{y}_2^1, \mathfrak{y}_1^2, \mathfrak{y}_2^2, \dots, \mathfrak{y}_1^N, \mathfrak{y}_2^N\}$ such that

$$(\mathfrak{y}_1, \mathfrak{y}_2^1), (\mathfrak{y}_1^i, \mathfrak{y}_2^i) \in \Gamma,\ i = 2, \dots, N, \quad \mathsf{H}(\mathfrak{y}_1', \mathfrak{y}_2^N), \mathsf{H}(\mathfrak{y}_1^{i+1}, \mathfrak{y}_2^i) < +\infty,\ i = 1, \dots, N-1.$$

In other words, the sequence $\{\mathfrak{y}_2^1, \mathfrak{y}_1^2, \mathfrak{y}_2^2, \dots, \mathfrak{y}_1^N, \mathfrak{y}_2^N\}$ "connects" $\mathfrak{y}_1$ to $\mathfrak{y}_1'$ keeping the cost finite when moving from a point in $\mathfrak{C}[\mathsf{X}_2]$ to a point in $\mathfrak{C}[\mathsf{X}_1]$ and imposing that the considered pair is in Γ, when moving from a point in $\mathfrak{C}[\mathsf{X}_1]$ to a point in $\mathfrak{C}[\mathsf{X}_2]$. For example, if H is everywhere finite, then any Γ is H-connected (see Theorem 5.8 for simple conditions implying H-connectedness).

Upon assuming that Γ is a H-cyclically monotone and H-connected subset of the effective domain of H also intersecting the interior of such domain, it is possible to prove the existence of *relaxed optimal potentials*: these are Borel functions $\varphi_i : \mathsf{X}_i \to \mathbb{R} \cup \{\pm\infty\}$ satisfying

$$\varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \text{for every } x_i \in \mathsf{X}_i,\ r_i \ge 0, \tag{1.7}$$

$$\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2 = \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \text{if } ([x_1, r_1], [x_2, r_2]) \in \Gamma, \tag{1.8}$$

where the inequality in (1.7) corresponds to the constraint as in (1.5) and the notation $+_o$ means that, whenever an ambiguity $\pm\infty \mp \infty$ arises, the sum is set equal to 0. This existence result, together with some finiteness conditions relating μ_i and H (see (5.16)), imply that $\varphi_i \in \mathcal{L}^1(\mathsf{X}_i, \mu_i)$ solve the dual problem

$$\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \,\mathrm{d}\mu_2 = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2), \tag{1.9}$$

and yields optimality for the homogeneous plan $\boldsymbol{\alpha}$. We refer to Theorem 5.10 for the detailed sufficiency result.

A Monge-like formulation via transport-growth maps. As in classical Optimal Transport problems, Unbalanced Optimal Transport admits a more restrictive Monge-like formulation in terms of *transport-growth* pairs [LMS23]: they are maps (T, g) from X_1 to $\mathsf{X}_2 \times [0, +\infty)$ acting on measures $\mu_1 \in \mathcal{M}(\mathsf{X}_1)$ via the formula

$$(\mathsf{T}, \mathsf{g})_\star \mu_1 = \mathsf{T}_\sharp(\mathsf{g}\mu_1), \quad \mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1 \ \Leftrightarrow\ \mu_2(B) = \int_{\mathsf{T}^{-1}(B)} \mathsf{g} \,\mathrm{d}\mu_1 \quad \text{for every Borel set } B \subset \mathsf{X}_2. \tag{1.10}$$

Whenever $\mathsf{g} \in L^1(\mathsf{X}_1, \mu_1)$, a transport-growth pair induces a homogeneous coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$, $\mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1$, via the formula

$$\boldsymbol{\alpha} = ([\mathrm{id}_{\mathsf{X}_1}, 1], [\mathsf{T}, \mathsf{g}])_\sharp \mu_1, \qquad \int \mathsf{H} \,\mathrm{d}\boldsymbol{\alpha} = \int \mathsf{H}([x, 1], [\mathsf{T}(x), \mathsf{g}(x)]) \,\mathrm{d}\mu_1(x), \tag{1.11}$$

and one can study the corresponding Monge formulation of (1.3):

$$\mathscr{M}_{\mathsf{H}}(\mu_1, \mu_2) := \inf \Big\{ \int_{\mathsf{X}_1} \mathsf{H}([x, 1], [\mathsf{T}(x), \mathsf{g}(x)]) \,\mathrm{d}\mu_1(x) : \mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1 \Big\}. \tag{1.12}$$

Clearly, $\mathscr{U}_{\mathsf{H}} \le \mathscr{M}_{\mathsf{H}}$; notice also that, whenever an optimal plan $\boldsymbol{\alpha}$ between μ_1 and μ_2 charges the set $\{\mathfrak{y}_1 = \mathfrak{o}_1\}$, then it is impossible that such $\boldsymbol{\alpha}$ can be induced by a transport-growth map (T, g) as in (1.11). This corresponds to the fact that some of the mass of μ_2 does not come from μ_1 but it is "created". Under the same hypotheses of the classical balanced case [Pra07] (i.e. the cost function is continuous and μ_1 is atomless),

Theorem 1.2. *Let X_i be Polish spaces, let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2] \to [0, +\infty]$ be a proper, radially 1-homogeneous and continuous function, $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ with μ_1 diffuse (i.e. $\mu_1(\{x\}) = 0$ for every $x \in \mathsf{X}_1$) and such that $\mu_1(\mathsf{X}_1) > 0$. Then*

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) = \mathscr{M}_{\mathsf{H}}(\mu_1, \mu_2).$$

In general, the infimum in the definition of $\mathscr{M}_{\mathsf{H}}(\mu_1, \mu_2)$ is not attained; however, as a consequence of the existence of relaxed optimal potentials as in (1.9) and (1.7), when $\mathsf{X}_1 = \mathsf{X}_2 = \mathbb{R}^d$, μ_1 is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$ and the differential of H satisfies suitable assumptions, we also obtain the existence of an optimal transport-growth map attaining the minimum in (1.12). This is the case, for example, of the cost function in (1.6) associated with the Gaussian Hellinger-Kantorovich metric.

Theorem 1.3. *Let $\mathsf{H} : \mathfrak{C}[\mathbb{R}^d] \times \mathfrak{C}[\mathbb{R}^d] \to \mathbb{R}$ be given by*

$$\mathsf{H}([x_1, r_1], [x_2, r_2]) := r_1 + r_2 - 2\sqrt{r_1 r_2}\, \mathrm{e}^{-|x_1 - x_2|^2/2}, \quad x_i \in \mathbb{R}^d,\ r_i \geq 0.$$

If $\mu_i \in \mathcal{M}_+(\mathbb{R}^d)$ with $\mu_i(\mathbb{R}^d) > 0$ for $i = 1, 2$ and $\mu_1 \ll \mathcal{L}^d$, then there exists a transport-growth map $(\mathsf{T}, \mathsf{g}) : \mathbb{R}^d \to \mathbb{R}^d \times [0, +\infty)$ such that

$$\mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1 = \mathsf{T}_\sharp(\mathsf{g}\mu_1), \qquad \int_{\mathbb{R}^d} \mathsf{H}([x, 1], [\mathsf{T}(x), \mathsf{g}(x)])\, \mathrm{d}\mu_1(x) = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2).$$

The above theorem holds also for large class of cost functions H, we refer to Theorem 5.13 for the general result.

Metric and topological properties. Finally we frame in our general setting some of the results of [LMS18; Chi+18b; De 20; DM22] concerning the metric properties of Unbalanced Optimal Transport functionals. In particular, we prove that, in case H is (the p-th power of) a distance on $\mathfrak{C}[\mathsf{X}]$, then the resulting cost $\mathscr{U}_{\mathsf{H}}$ is itself (the p-th power of) a distance on an appropriate subset of $\mathcal{M}_+(\mathsf{X})$ metrizing the weak convergence of measures (see Theorems A.4 and A.5), precisely as it is for the standard Optimal Transport problem [AGS08, Proposition 7.1.5].

Plan of the paper. Section 2 is devoted to establish the general setting and a few technical tools that will be used in the sequel.

Section 3 contains the core of our results: the convexification approach, some structural properties and the comparison with the semi-coupling approach of [Chi+18b], the duality, and the Monge formulation.

Section 4 treats the dual attainment in spaces of continuous potentials under additional regularity assumptions on H and on the spaces X_i.

In Section 5 we present the general optimality conditions, the relaxed duality result and, as a consequence of the latter, the existence of optimal transport-growth maps.

Appendix A contains a few remarks on the metric and topological properties of $\mathscr{U}_{\mathsf{H}}$ in case H is (the p-th power of) a distance. Finally Appendix B reproduces the results of Section 4 under a different set of additional assumptions on H.

Acknowledgments. The authors gratefully acknowledge the support of the Institute for Advanced Study of the Technical University of Munich, funded by the German Excellence Initiative. G.S. has also been supported by IMATI-CNR, Pavia and by the MIUR-PRIN 202244A7YL project *Gradient Flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning.*

2. Preliminaries, measures and functions on the cone

In this section, X is a completely regular space (i.e. it is Hausdorff and for every closed set C and point $x \in \mathsf{X}\setminus C$ there exists a continuous function $f : \mathsf{X} \to [0, 1]$ s.t. $f(x) = 0$ and $f(C) = \{1\}$). We denote by $\mathscr{B}(\mathsf{X})$ the Borel σ-algebra on X and by $\mathcal{M}(\mathsf{X})$ (resp. $\mathcal{M}_+(\mathsf{X})$) the vector space of real valued (resp. the cone of non-negative) Radon measures on X i.e. the countably additive set functions $\mu : \mathscr{B}(\mathsf{X}) \to \mathbb{R}$ (resp. $\mu : \mathscr{B}(\mathsf{X}) \to [0, +\infty)$) s.t.

$$\text{for every } B \in \mathscr{B}(\mathsf{X}) \text{ and every } \varepsilon > 0 \text{ there exists } K \subset B \text{ compact s.t. } |\mu|(\mathsf{X} \setminus K) < \varepsilon,$$

where $|\mu|$ denotes the total variation measure of μ. We say that $\mu \in \mathcal{M}_+(\mathsf{X})$ is concentrated on a set $\Gamma \subset \mathsf{X}$ if for every $\varepsilon > 0$ there exists a Borel set $\Gamma_\varepsilon \subset \Gamma$ such that $\mu(\mathsf{X} \setminus \Gamma_\varepsilon) < \varepsilon$. The Radon property ensures that if μ is concentrated on Γ then it is concentrated on a σ-compact subset $\Gamma' \subset \Gamma$.

We denote by $\mathcal{P}(\mathsf{X})$ the set of probability measures on X i.e. the elements μ of $\mathcal{M}_+(\mathsf{X})$ s.t. $\mu(\mathsf{X}) = 1$. When X is a Polish space (i.e. its topology is induced by a metric d such that (X, d) is complete and separable), then every finite Borel measure is Radon so that $\mathcal{M}(\mathsf{X})$ coincides with the space of real valued Borel measures.

The zero measure in X is denoted by $\mathbf{0}_\mathsf{X}$. We define the sets of non-negative discrete measures and non-negative weighted Dirac masses respectively as

$$\mathcal{M}_{+,\mathrm{f}}(\mathsf{X}) := \left\{ \sum_{j=1}^m c_i \delta_{x_j} : \{c_j\}_{j=1}^m \subset [0, +\infty)\,, \{x_j\}_{j=1}^m \subset \mathsf{X}\,, m \in \mathbb{N} \right\},$$
$$\Delta_+(\mathsf{X}) := \{ r\delta_x : r \in [0, +\infty)\,, x \in \mathsf{X} \}\,.$$

Notice that $\mathcal{M}_{+,\mathrm{f}}(\mathsf{X}), \Delta_+(\mathsf{X}) \subset \mathcal{M}_+(\mathsf{X})$. If X, Y are two completely regular spaces and $f : \mathsf{X} \to \mathsf{Y}$ is a Borel function we denote by $f_\sharp : \mathcal{M}(\mathsf{X}) \to \mathcal{M}(\mathsf{Y})$ the push forward operator defined, for every $\mu \in \mathcal{M}(\mathsf{X})$, as

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

We denote by $\mathrm{C}_b(\mathsf{X})$ the vector space of continuous and bounded real functions. There is a natural duality pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{M}(\mathsf{X})$ and $\mathrm{C}_b(\mathsf{X})$

$$\langle \mu, \varphi \rangle := \int_\mathsf{X} \varphi \,\mathrm{d}\mu \quad \text{for every } \mu \in \mathcal{M}(\mathsf{X}),\ \varphi \in \mathrm{C}_b(\mathsf{X}). \tag{2.1}$$

(2.1) defines a real nondegenerate bilinear form in $\mathcal{M}(\mathsf{X}) \times \mathrm{C}_b(\mathsf{X})$, for if a Radon measure $\mu \in \mathcal{M}(\mathsf{X})$ satisfies $\int_\mathsf{X} \varphi \,\mathrm{d}\mu = 0$ for every $\varphi \in \mathrm{C}_b(\mathsf{X})$, then $|\mu|(B) = 0$ for every $B \in \mathscr{B}(\mathsf{X})$ (e.g. by the approximation result [Bog06, Lemma 7.2.8]) so that μ is the null measure. Hence we can endow $\mathcal{M}(\mathsf{X})$ with the weak Hausdorff topology $\sigma(\mathcal{M}(\mathsf{X}), \mathrm{C}_b(\mathsf{X}))$: the coarsest topology on $\mathcal{M}(\mathsf{X})$ for which the maps $\mu \mapsto \int_\mathsf{X} \varphi \,\mathrm{d}\mu$ are continuous for every $\varphi \in \mathrm{C}_b(\mathsf{X})$. Since in general $\mathcal{M}(\mathsf{X})$ is not first-countable (but $\mathcal{M}_+(\mathsf{X})$ is Polish, thus metrizable, if X is Polish), we will mostly deal with general nets $(\mu_\lambda)_{\lambda\in\mathbb{L}}$, i.e. maps $\lambda \to \mu_\lambda$ defined in a directed set $\mathbb{L}$ with values in $\mathcal{M}(\mathsf{X})$, see e.g. [Fol99, §4.3]. By definition, a net $(\mu_\lambda)_{\lambda\in\mathbb{L}} \subset \mathcal{M}(\mathsf{X})$ converges to $\mu \in \mathcal{M}(\mathsf{X})$ in the weak topology if

$$\lim_{\lambda\in\mathbb{L}} \int_\mathsf{X} \varphi \,\mathrm{d}\mu_\lambda = \int_\mathsf{X} \varphi \,\mathrm{d}\mu \quad \text{for every } \ \varphi \in \mathrm{C}_b(\mathsf{X}).$$

If $f : \mathsf{X} \to (-\infty, +\infty]$ is a function, we denote by $D(f)$ its effective domain, defined as

$$D(f) := \{ x \in \mathsf{X} \mid f(x) < +\infty \}$$

and by $\bar{f}$ its lower semicontinuous envelope i.e.

$$\bar{f}(x) := \inf \left\{ \liminf_\lambda f(x_\lambda) : (x_\lambda)_{\lambda\in\mathbb{L}} \subset \mathsf{X},\ x_\lambda \to x \right\}, \quad x \in \mathsf{X}; \tag{2.2}$$

$\bar{f}$ is the largest lower semicontinuous function below f. If X is also a topological vector space and $A \subset \mathsf{X}$, we denote by $\mathrm{co}\,(A)$ (resp. by $\overline{\mathrm{co}}\,(A)$) the convex (resp. closed and convex) envelope of A. If $g : \mathsf{X} \to (-\infty, +\infty]$ is a function, we denote its convex envelope and its closed convex envelope by $\mathrm{co}\,(g)$ and $\overline{\mathrm{co}}\,(g)$, defined by

$$\mathrm{co}\,(g)\,(x) := \inf\left\{\sum_{i=1}^n \alpha_i g(x_i) : \{x_i\}_{i=1}^n \subset \mathsf{X},\ \{\alpha_i\}_{i=1}^n \in S_n,\ \sum_{i=1}^n \alpha_i x_i = x,\ n \geq 1\right\},\ x \in \mathsf{X} \tag{2.3}$$

$$\overline{\mathrm{co}}\,(g) := \overline{\mathrm{co}\,(g)}, \tag{2.4}$$

where

$$S_n := \left\{\{\alpha_i\}_{i=1}^n : \alpha_i \in [0,1],\ \sum_{i=1}^n \alpha_i = 1\right\}.$$

$\mathrm{co}\,(g)$ is the largest convex function below g and $\overline{\mathrm{co}}\,(g)$ is the largest lower semicontinuous and convex function below g. The following proposition is a simple density result whose proof easily follows adapting [Bog06, Example 8.1.6].

Proposition 2.1. *Let* X *be a completely regular space. Then for every* $c \geq 0$ *we have*

$$\overline{\{\mu \in \mathcal{M}_{+,\mathrm{f}}(\mathsf{X}) : \mu(\mathsf{X}) = c\}} = \{\mu \in \mathcal{M}_+(\mathsf{X}) : \mu(\mathsf{X}) = c\}, \quad \overline{\mathcal{M}_{+,\mathrm{f}}(\mathsf{X})} = \mathcal{M}_+(\mathsf{X}).$$

Remark 2.2. It holds

$$\overline{\mathrm{co}}\,(\Delta_+(\mathsf{X})) = \mathcal{M}_+(\mathsf{X}).$$

This is an immediate consequence of the fact that $\mathrm{co}\,(\Delta_+(\mathsf{X})) = \mathcal{M}_{+,\mathrm{f}}(\mathsf{X})$ and Proposition 2.1.

The following Lemma is a refinement of Proposition 2.1 showing that, given a Borel function f and a non-negative measure α, we can construct an approximating sequence of discrete measures for which we have convergence also of the integral of f.

Lemma 2.3. *Let* X *be a completely regular space and let* $\alpha \in \mathcal{M}_+(\mathsf{X})$. *Let* $f : \mathsf{X} \to [0, +\infty]$ *be a Borel function. Then there exists a net* $(\gamma_\lambda)_{\lambda \in \mathbb{L}} \subset \{\mu \in \mathcal{M}_{+,\mathrm{f}}(\mathsf{X}) : \mu(\mathsf{X}) = \alpha(\mathsf{X})\}$ *s.t.*

$$\lim_{\lambda \in \mathbb{L}} \gamma_\lambda = \alpha, \quad \lim_{\lambda \in \mathbb{L}} \int_{\mathsf{X}} f \,\mathrm{d}\gamma_\lambda = \int_{\mathsf{X}} f \,\mathrm{d}\alpha.$$

Proof. By Lusin's theorem, we can find an increasing sequence of compact sets such that

$$\mathsf{X}_k \subset \mathsf{X}_{k+1} \quad \text{for every } k \geq 1, \quad \alpha(\mathsf{X} \setminus \mathsf{X}_k) \leq \frac{1}{k}, \quad f|_{\mathsf{X}_k} \text{ is bounded and continuous.}$$

Consider now the family of measures $\{\alpha_k\}_{k\geq 1} \subset \{\mu \in \mathcal{M}_+(\mathsf{X}) : \mu(\mathsf{X}) = \alpha(\mathsf{X})\}$ defined as

$$\alpha_k := \frac{\alpha(\mathsf{X})}{\alpha(\mathsf{X}_k)} \alpha|_{\mathsf{X}_k} \quad \text{for every } k \geq 1.$$

We can easily observe that

$$\lim_{k \to +\infty} \alpha_k = \alpha$$

indeed, if $\varphi \in \mathrm{C}_b(\mathsf{X})$, we have

$$\lim_{k \to +\infty} \int_{\mathsf{X}} \varphi \,\mathrm{d}\alpha_k = \lim_{k \to +\infty} \frac{\alpha(\mathsf{X})}{\alpha(\mathsf{X}_k)} \int_{\mathsf{X}} \varphi \chi_{\mathsf{X}_k} \,\mathrm{d}\alpha = \int_{\mathsf{X}} \varphi \,\mathrm{d}\alpha$$

by monotone convergence. The same argument shows that we also have

$$\lim_{k \to +\infty} \int_{\mathsf{X}} f \,\mathrm{d}\alpha_k = \int_{\mathsf{X}} f \,\mathrm{d}\alpha. \tag{2.5}$$

By Proposition 2.1, for every $k \geq 1$, we can find a net $\{\gamma_\lambda^k\}_{\lambda \in \mathbb{L}_k} \subset \mathcal{M}_{+,\mathrm{f}}(\mathsf{X}_k) \cap \{\mu \in \mathcal{M}_+(\mathsf{X}) : \mu(\mathsf{X}) = \alpha(\mathsf{X})\}$, such that

$$\lim_{\lambda \in \mathbb{L}_k} \gamma_\lambda^k = \alpha_k.$$

Moreover, since $f|_{\mathsf{X}_k}$ is bounded and continuous, it holds

$$\lim_{\lambda\in\mathbb{L}_k}\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}^k_\lambda=\lim_{\lambda\in\mathbb{L}_k}\int_{\mathsf{X}_k} f\,\mathrm{d}\boldsymbol{\gamma}^k_\lambda=\int_{\mathsf{X}_k} f\,\mathrm{d}\alpha_k=\int_{\mathsf{X}} f\,\mathrm{d}\alpha_k.$$

This allows us to find, for every $k\geq 1$, some $\bar m(k)\in\mathbb{L}_k$ s.t.

$$\left|\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}^k_\lambda-\int_{\mathsf{X}} f\,\mathrm{d}\alpha_k\right|\leq\frac{1}{k}\quad\text{ for every }\ \lambda\geq\bar m(k).$$

Hence we can consider, for every $k\geq 1$, the directed sets $\mathbb{E}_k:=\{\lambda\in\mathbb{L}_k:\lambda_k\geq\bar m(k)\}$ and the corresponding new sequence of nets $\{\boldsymbol{\gamma}^k_\lambda\}_{\lambda\in\mathbb{E}_k}$. Obviously it holds

$$\lim_{\lambda\in\mathbb{E}_k}\boldsymbol{\gamma}^k_\lambda=\alpha_k,\quad\left|\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}^k_\lambda-\int_{\mathsf{X}} f\,\mathrm{d}\alpha_k\right|\leq\frac{1}{k}\quad\text{ for every }\ \lambda\in\mathbb{E}_k.\tag{2.6}$$

Define now the directed set

$$\mathbb{N}\otimes\mathbb{E}_k:=\{(k,\lambda):\lambda\in\mathbb{E}_k\}\text{ with order }(k,\lambda)\leq(k',\lambda')\iff k<k'\text{ or }(k=k'\wedge\lambda\leq\lambda').$$

By the diagonal principle for nets, we can find a directed set $\mathbb{B}$ and a monotone final function

$$h:\mathbb{B}\to\mathbb{N}\otimes\mathbb{E}_k,\quad h(\beta)=(h_1(\beta),h_2(\beta))\text{ with }h_2(\beta)\in\mathbb{E}_{h_1(\beta)}\quad\text{ for every }\ \beta\in\mathbb{B}$$

such that the diagonal net $\{\boldsymbol{\gamma}_\beta\}_{\beta\in\mathbb{B}}:=\{\boldsymbol{\gamma}^{h_1(\beta)}_{h_2(\beta)}\}_{\beta\in\mathbb{B}}\subset\mathcal{M}_{+,\mathrm{f}}(\mathsf{X})\cap\{\mu\in\mathcal{M}_+(\mathsf{X}):\mu(\mathsf{X})=\alpha(\mathsf{X})\}$ converges to α. We only need to prove that also the integral of f converges:

$$\begin{aligned}\left|\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}_\beta-\int_{\mathsf{X}} f\,\mathrm{d}\alpha\right|&\leq\left|\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}_\beta-\int_{\mathsf{X}} f\,\mathrm{d}\alpha_{h_1(\beta)}\right|+\left|\int_{\mathsf{X}} f\,\mathrm{d}\alpha_{h_1(\beta)}-\int_{\mathsf{X}} f\,\mathrm{d}\alpha\right|\\&=\left|\int_{\mathsf{X}} f\,\mathrm{d}\boldsymbol{\gamma}^{h_1(\beta)}_{h_2(\beta)}-\int_{\mathsf{X}} f\,\mathrm{d}\alpha_{h_1(\beta)}\right|+\left|\int_{\mathsf{X}} f\,\mathrm{d}\alpha_{h_1(\beta)}-\int_{\mathsf{X}} f\,\mathrm{d}\alpha\right|\\&\leq\frac{1}{h_1(\beta)}+\left|\int_{\mathsf{X}} f\,\mathrm{d}\alpha_{h_1(\beta)}-\int_{\mathsf{X}} f\,\mathrm{d}\alpha\right|,\end{aligned}$$

where we have used (2.6). Now it is enough to observe that $h_1:\mathbb{B}\to\mathbb{N}$ is a final monotone function i.e. it is an increasing monotone sequence converging to $+\infty$. Passing to $\lim_{\beta\in\mathbb{B}}$ and using (2.5), we conclude. □

2.1. **The cone construction.** It will be natural to state some definitions and results in the context of the so called geometric cone: we introduce on $\mathsf{X}\times\mathbb{R}_+$ the equivalence relation

$$(x,r)\sim(y,s)\quad\overset{\text{def}}{\Leftrightarrow}\quad[x=y,r=s\neq 0\quad\vee\quad r=s=0]\tag{2.7}$$

and the corresponding geometric cone $\mathfrak{C}[\mathsf{X}]:=(\mathsf{X}\times\mathbb{R}_+)/\sim$, whose points are denoted by gothic letters as $\mathfrak{y}$. We denote by $\mathfrak{p}$ the quotient map $\mathfrak{p}:\mathsf{X}\times\mathbb{R}_+\to\mathfrak{C}[\mathsf{X}]$ sending a point (x,r) to its equivalence class $[x,r]$. Notice that $\mathfrak{p}$ is just the identity map except for those points with $r=0$, which are all sent to the same equivalence class, the so called vertex of the cone, that we denote with $\mathfrak{o}$. We will consider the natural product operation on the cone given by

$$\lambda[x,r]:=[x,\lambda r]\quad\text{ for every }\lambda,r\geq 0,\,x\in\mathsf{X}.$$

In certain cases it is useful to add a further isolated point o (which plays the role of a source or a sink) to X and to consider the spaces $\mathsf{X}_o:=\mathsf{X}\sqcup\{o\}$ and $\mathsf{Y}_o:=(\mathsf{X}\times\mathbb{R}_+)\sqcup\{(o,0)\}\subset\mathsf{X}_o\times\mathbb{R}_+$. We can identify X (resp. $\mathsf{X}\times\mathbb{R}_+$) with a (closed) subset of X_o (resp. Y_o) and $\mathcal{M}_+(\mathsf{X})$ (resp. $\mathcal{M}_+(\mathsf{X}\times\mathbb{R}_+)$) with the (closed) subset of $\mathcal{M}_+(\mathsf{X}_o)$ (resp. $\mathcal{M}_+(\mathsf{Y}_o)$) given by all the measures not charging $\{o\}$ (resp. $\{(o,0)\}$). Notice that $\mathfrak{C}[\mathsf{X}]=\mathsf{Y}_o/\sim$ and we can trivially extend $\mathfrak{p}$ to Y_o.

On the cone we introduce the projections on $\mathbb{R}_+$ and X_o simply defined as $\mathsf{r}([x,r])=r$ and $\mathsf{x}([x,r])=x$ if $r>0$ and $\mathsf{x}([x,r])=\mathsf{x}(\mathfrak{o})=o$ if $r=0$.

We can define a right inverse of $\mathfrak{p} : \mathsf{Y}_o \to \mathfrak{C}[\mathsf{X}]$ as

$$\mathfrak{q}([x,r]) = (\mathsf{x}([x,r]), \mathsf{r}([x,r])).$$

Notice that also that $\mathfrak{q}(\mathfrak{C}[\mathsf{X}] \setminus \{\mathfrak{o}\}) = \mathsf{X} \times (0,+\infty)$.

On $\mathfrak{C}[\mathsf{X}]$ we consider the following topology, weaker then the quotient one: a local system of neighbourhoods of a point $[x,r]$ is just the image trough $\mathfrak{p}$ of the local system of neighbourhoods given by the product topology at $(x,r) \in \mathsf{X} \times \mathbb{R}_+$, if $r > 0$. A local system of neighbourhoods at $\mathfrak{o}$ is given by

$$\left\{\{[x,r] \in \mathfrak{C}[\mathsf{X}] : 0 \le r < \varepsilon\}\right\}_{\varepsilon > 0}. \tag{2.8}$$

If the topology of X is induced by a metric d, then the topology of $\mathfrak{C}[\mathsf{X}]$ is induced by the metric $\mathsf{d}_{\mathfrak{C}} : \mathfrak{C}[\mathsf{X}] \times \mathfrak{C}[\mathsf{X}] \to [0,+\infty)$ defined as

$$\mathsf{d}_{\mathfrak{C}}([x,r],[y,s]) := \big(r^2 + s^2 - 2rs\cos(\mathsf{d}(x,y) \wedge \pi)\big)^{\frac{1}{2}}, \quad [x,r],[y,s] \in \mathfrak{C}[\mathsf{X}]. \tag{2.9}$$

With the above topology, $\mathfrak{C}[\mathsf{X}]$ is completely regular and it is the right object to consider when one wants to represent elements in $\Delta_+(\mathsf{X})$; in particular we have the following result.

Lemma 2.4. *Let X be a completely regular space. Then $\Delta_+(\mathsf{X})$ is homeomorphic to $\mathfrak{C}[\mathsf{X}]$.*

Proof. The homeomorphism is given by the map $\varphi : \mathfrak{C}[\mathsf{X}] \to \Delta_+(\mathsf{X})$ defined as

$$\varphi([x,r]) := \begin{cases} r\delta_x & \text{if } r > 0, \\ \mathbf{0}_{\mathsf{X}} & \text{if } r = 0. \end{cases}$$

□

Remark 2.5. Let us point out an obvious but useful fact that we will extensively use in the following without having to recall it further: if a measure $\boldsymbol{\beta} \in \mathcal{M}_+\big(\mathfrak{C}[\mathsf{X}]\big)$ does not charge $\{\mathfrak{o}\}$ then $\mathsf{x}_\sharp\boldsymbol{\beta}$ does not charge $\{o\}$ and therefore we will identify it with a measure in $\mathcal{M}_+(\mathsf{X})$. In particular, this construction does not depend on the choice of the point o and the map x takes value in X $\boldsymbol{\beta}$-a.e.

If $R > 0$, we define

$$\mathfrak{C}_R[\mathsf{X}] := \{[x,r] \in \mathfrak{C}[\mathsf{X}] : 0 \le r \le R\} \tag{2.10}$$

and we will often identify measures on $\mathfrak{C}[\mathsf{X}]$ with support contained in $\mathfrak{C}_R[\mathsf{X}]$ with elements of $\mathcal{M}(\mathfrak{C}_R[\mathsf{X}])$. For every $p \ge 1$, we introduce moreover the set

$$\mathfrak{M}^p_+(\mathfrak{C}[\mathsf{X}]) := \left\{\alpha \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}]) : \int_{\mathfrak{C}[\mathsf{X}]} \mathsf{r}^p \, \mathrm{d}\alpha < +\infty\right\},$$

and the map (see the above Remark 2.5)

$$\mathfrak{h}^p : \mathfrak{M}^p_+(\mathfrak{C}[\mathsf{X}]) \to \mathcal{M}_+(\mathsf{X}), \quad \mathfrak{h}(\alpha) = \mathsf{x}_\sharp(\mathsf{r}^p\alpha).$$

We stress the fact that the map $\mathfrak{h}^p$ does not depend on the point o occuring in the definition of x.

We introduce now the product cone: given X_1 and X_2 completely regular spaces, we define $\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] := \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2]$ endowed with the product topology. Points in the product cone are denoted by bold gothic letters as $\mathfrak{y} = (\mathfrak{y}_1,\mathfrak{y}_2) = ([x_1,r_1],[x_2,r_2])$. We denote with $\mathfrak{o}_i$ (resp. o_i) the vertex of $\mathfrak{C}[\mathsf{X}_i]$ (resp. isolated points added to X_i and forming the disjoint union $\mathsf{X}_{i,o} := \mathsf{X}_i \sqcup \{o_i\}$), $i = 1,2$ and we set $\boldsymbol{\mathfrak{o}} := (\mathfrak{o}_1,\mathfrak{o}_2)$, $\boldsymbol{o} := (o_1,o_2)$, $\mathfrak{C}_{\boldsymbol{\mathfrak{o}}}[\mathsf{X}_1,\mathsf{X}_2] := \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \setminus \{\boldsymbol{\mathfrak{o}}\}$.

On the product cone we can consider the projections on the two components $\pi^i : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to \mathfrak{C}[\mathsf{X}_i]$ sending $([x_1,r_1],[x_2,r_2])$ to $[x_i,r_i]$ and the projections on $\mathbb{R}_+$ and X_i simply defined as $\mathsf{r}_i := \mathsf{r} \circ \pi^i$ and $\mathsf{x}_i := \mathsf{x} \circ \pi^i$ (x_i maps $\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]$ into $\mathsf{X}_{i,o}$). In analogy with (2.9), if the topologies of X_1 and X_2 are induced by distances d_1 and d_2 respectively, the topology of the product cone is induced by the distance

$$(\mathsf{d}_1 \otimes_{\mathfrak{C}} \mathsf{d}_2)((\mathfrak{y}_1, \mathfrak{y}_2), (\mathfrak{w}_1, \mathfrak{w}_2)) := \left(\mathsf{d}^2_{1,\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{w}_1) + \mathsf{d}^2_{2,\mathfrak{C}}(\mathfrak{y}_2, \mathfrak{w}_2)\right)^{\frac{1}{2}}, \quad (\mathfrak{y}_1, \mathfrak{y}_2), (\mathfrak{w}_1, \mathfrak{w}_2) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]. \tag{2.11}$$

As in (2.10), given $R > 0$, we define

$$\mathfrak{C}_R[\mathsf{X}_1, \mathsf{X}_2] := \mathfrak{C}_R[\mathsf{X}_1] \times \mathfrak{C}_R[\mathsf{X}_2] = \{\mathfrak{y} \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] : 0 \le \mathsf{r}_i(\mathfrak{y}) \le R,\ i = 1, 2\}$$

and we will identify measures on $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ with support contained in $\mathfrak{C}_R[\mathsf{X}_1, \mathsf{X}_2]$ with elements of $\mathcal{M}(\mathfrak{C}_R[\mathsf{X}_1, \mathsf{X}_2])$. For every $p \ge 1$, we introduce the set

$$\mathfrak{M}^p_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) := \left\{ \boldsymbol{\alpha} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) : \int_{\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]} (\mathsf{r}_1^p + \mathsf{r}_2^p)\, \mathrm{d}\boldsymbol{\alpha} < +\infty \right\},$$

and the maps (see the above Remark 2.5)

$$\mathfrak{h}^p_i : \mathfrak{M}^p_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) \to \mathcal{M}_+(\mathsf{X}_i), \quad \mathfrak{h}^p_i(\boldsymbol{\alpha}) = (\mathsf{x}_i)_\sharp(\mathsf{r}_i^p \boldsymbol{\alpha}) \quad i = 1, 2.$$

Finally we define, for every $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ and every $p \ge 1$, the set

$$\mathfrak{H}^p(\mu_1, \mu_2) := \left\{ \boldsymbol{\alpha} \in \mathfrak{M}^p_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) : \mathfrak{h}^p_i(\boldsymbol{\alpha}) = \mu_i,\ i = 1, 2 \right\}. \tag{2.12}$$

If $\boldsymbol{\alpha} \in \mathfrak{H}^p(\mu_1, \mu_2)$, we say that μ_1 and μ_2 are the p-homogeneous marginals of $\boldsymbol{\alpha}$ and that $\boldsymbol{\alpha}$ is a p-homogeneous plan/coupling between μ_1 and μ_2. Notice that if $\boldsymbol{\alpha} \in \mathfrak{H}^p(\mu_1, \mu_2)$ then also the restriction of $\boldsymbol{\alpha}$ to $\mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1, \mathsf{X}_2]$, $\boldsymbol{\alpha}_{\mathfrak{o}} := \boldsymbol{\alpha} \llcorner \mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1, \mathsf{X}_2]$, belongs to $\mathfrak{H}^p(\mu_1, \mu_2)$.

Remark 2.6 (Trivial cases). If $\mu_i(\mathsf{X}_i) = 0$ for $i = 1, 2$ (μ_i are the null measures in X_i), then $\mathfrak{H}^p(\mu_1, \mu_2) = \{\lambda\delta_{\mathfrak{o}} : \lambda \ge 0\}$.

If $\mu_1(\mathsf{X}_1) > \mu_2(\mathsf{X}_2) = 0$ then $\mathfrak{H}^p(\mu_1, \mu_2) = \{\alpha \otimes \delta_{\mathfrak{o}_2} : \alpha \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1]),\ \mathfrak{h}^p(\alpha) = \mu_1\}$. A similar characterization holds when $0 = \mu_1(\mathsf{X}_1) < \mu_2(\mathsf{X}_2)$.

We say that (μ_1, μ_2) is a non-trivial pair if $\mu_i(\mathsf{X}_i) > 0$ for $i = 1, 2$.

Remark 2.7. The class $\mathfrak{H}^p(\mu_1, \mu_2)$ satisfies a natural invariance property with respect to inclusion of the ambient spaces X_i. First of all, if X_i are subsets of completely regular spaces $\tilde{\mathsf{X}}_i$ (and the topology of X_i coincides with the relative topology induced by the inclusion in $\tilde{\mathsf{X}}_i$) there is a natural identification between (pair of) Radon measures $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ and (pair of) Radon measures $\tilde{\mu}_i \in \mathcal{M}_+(\tilde{\mathsf{X}}_i)$ concentrated on X_i (i.e. such that $\tilde{\mathsf{X}}_i \setminus \mathsf{X}_i$ is μ_i-negligible). Since $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ can be considered as a subset of $\mathfrak{C}[\tilde{\mathsf{X}}_1, \tilde{\mathsf{X}}_2]$ (with identification of the corresponding vertexes $\mathfrak{o}_i$ and $\tilde{\mathfrak{o}}_i$) every homogeneous coupling $\boldsymbol{\alpha} \in \mathfrak{H}^p(\mu_1, \mu_2)$ can be considered as a measure in $\mathcal{M}_+(\mathfrak{C}[\tilde{\mathsf{X}}_1, \tilde{\mathsf{X}}_2])$ which belongs to $\mathfrak{H}^p(\tilde{\mu}_1, \tilde{\mu}_2)$. Conversely, if $\tilde{\mu}_i \in \mathcal{M}_+(\tilde{\mathsf{X}}_i)$ are concentrated in X_i then every plan $\tilde{\boldsymbol{\alpha}} \in \mathfrak{H}^p(\tilde{\mu}_1, \tilde{\mu}_2)$ is concentrated in $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ and its restriction to $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ belongs to $\mathfrak{H}^p(\mu_1, \mu_2)$. In particular, recalling the construction of the spaces $\mathsf{X}_{i,o}$ at the beginning of Section 2.1, for every pair of measures $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ the set $\mathfrak{H}^p(\mu_1, \mu_2)$ can also be considered as a subset of $\mathcal{M}_+\big(\mathfrak{C}[\mathsf{X}_{1,o}, \mathsf{X}_{2,o}]\big)$.

The following renormalization result comes from [LMS18].

Lemma 2.8. *Let X_i for $i = 1, 2$ be completely regular spaces and let $p \ge 1$. Given $\boldsymbol{\alpha} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2])$ and $\vartheta : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to (0, +\infty)$ Borel measurable in $L^p(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2], \boldsymbol{\alpha})$ we can define*

$$\mathrm{prd}_\vartheta(\mathfrak{y}) := \left(\vartheta(\mathfrak{y})^{-1}\mathfrak{y}_1, \vartheta(\mathfrak{y})^{-1}\mathfrak{y}_2\right), \quad \mathfrak{y} \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2], \qquad \mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha}) := (\mathrm{prd}_\vartheta)_\sharp(\vartheta^p \boldsymbol{\alpha}).$$

Then we have

$$\mathfrak{h}_i(\mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha})) = \mathfrak{h}^p_i(\boldsymbol{\alpha}), \quad i = 1, 2.$$

In particular, if we define

$$\vartheta_{\boldsymbol{\alpha},p}(\mathfrak{y}) := \frac{1}{r^*(\boldsymbol{\alpha})}\begin{cases}\mathsf{r}_1^p(\mathfrak{y})+\mathsf{r}_2^p(\mathfrak{y}) & \text{if } \mathfrak{y}\neq(\mathfrak{o}_1,\mathfrak{o}_2)\\ 1 & \text{if } \mathfrak{y}=(\mathfrak{o}_1,\mathfrak{o}_2)\end{cases},$$

where $r^*(\boldsymbol{\alpha})$ is a normalization constant s.t. $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]}\vartheta^p_{\boldsymbol{\alpha},p}\,\mathrm{d}\boldsymbol{\alpha}=1$ given by

$$r^*(\boldsymbol{\alpha}) := \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]}(\mathsf{r}_1^p+\mathsf{r}_2^p)\,\mathrm{d}\boldsymbol{\alpha}+\boldsymbol{\alpha}(\{(\mathfrak{o}_1,\mathfrak{o}_2)\}),$$

we have that $\mathrm{dil}_{\vartheta_{\boldsymbol{\alpha},p},p}(\boldsymbol{\alpha})\in\mathcal{P}(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$, it has the same p-homogeneous marginals of $\boldsymbol{\alpha}$ and its support is contained in $\mathfrak{C}_{r^*(\boldsymbol{\alpha})}[\mathsf{X}_1,\mathsf{X}_2]$.

The following Lemma is a compactness result for homogeneous marginals (the corresponding statement for classical marginals is [SS22, Theorem 3.1]).

Lemma 2.9. *Let $p\geq 1$, $R>0$ and let X,X_i, $i=1,2$ be completely regular spaces. Then:*

(1) *if $(\alpha_\lambda)_{\lambda\in\mathbb{L}}$ is a net in $\mathcal{P}(\mathfrak{C}_R[\mathsf{X}])$ such that $\mu_\lambda:=\mathfrak{h}^p(\alpha_\lambda)\in\mathcal{M}_+(\mathsf{X})$, $\lambda\in\mathbb{L}$, converges to some $\mu\in\mathcal{M}_+(\mathsf{X})$, then there exists a subnet $(\alpha'_\beta)_{\beta\in\mathbb{B}}$ of $(\alpha_\lambda)_{\lambda\in\mathbb{L}}$ convergent to some $\alpha\in\mathcal{P}(\mathfrak{C}_R[\mathsf{X}])$ with $\mathfrak{h}^p(\alpha)=\mu$;*

(2) *if $(\boldsymbol{\alpha}_\lambda)_{\lambda\in\mathbb{L}}$ is a net in $\mathcal{P}(\mathfrak{C}_R[\mathsf{X}_1,\mathsf{X}_2])$ such that $\mu_{i,\lambda}:=\mathfrak{h}_i^p(\boldsymbol{\alpha}_\lambda)\in\mathcal{M}_+(\mathsf{X}_i)$, $i=1,2$, $\lambda\in\mathbb{L}$, converge to some μ_i in $\mathcal{M}(\mathsf{X}_i)$, then there exists a subnet $(\boldsymbol{\alpha}'_\beta)_{\beta\in\mathbb{B}}$ of $(\boldsymbol{\alpha}_\lambda)_{\lambda\in\mathbb{L}}$ convergent to some $\boldsymbol{\alpha}\in\mathfrak{H}^p(\mu_1,\mu_2)$.*

Proof. Thanks to [SS22, Theorem 3.1], it is enough to prove only the first claim. Define, for every $\lambda\in\mathbb{L}$,

$$\vartheta_\lambda := \mathfrak{q}_\sharp(\mathsf{r}^p\alpha_\lambda)\in\mathcal{M}_+(\mathsf{X}\times[0,R]).$$

Notice that this definition does not depend on the point $\bar{x}$ w.r.t. $\mathfrak{q}$ is defined. Observe that $\pi^{[0,R]}_\sharp\vartheta_\lambda\in\mathcal{M}_+([0,R])$ with mass bounded by R^p and $\pi^{\mathsf{X}}_\sharp\vartheta_\lambda=\mu_\lambda$. Then we can apply [SS22, Theorem 3.1] to $(\vartheta_\lambda)_{\lambda\in\mathbb{L}}$ and obtain that, up to passing to a subnet, there exists $\vartheta\in\mathcal{M}_+(\mathsf{X}\times[0,R])$ s.t. $\lim_{\lambda\in\mathbb{L}}\vartheta_\lambda=\vartheta$. Now we define

$$O_n := \left\{[x,r]\in\mathfrak{C}[\mathsf{X}] : 0\leq r\leq\frac{1}{n}\right\},\quad n\geq 1$$

and, for every $n\geq 1$, the nets of real numbers

$$m_{\lambda,n} := \boldsymbol{\alpha}_\lambda(O_n).$$

Observe that $0\leq m_{\lambda,n}\leq 1$ for every $n\geq 1$ and $\lambda\in\mathbb{L}$ then, up to passing to a subnet (the same for every $n\in\mathbb{N}$), they converge in $\lambda\in\mathbb{L}$ to some $m_n\in[0,1]$. Define then $m:=\inf_{n\geq 1}m_n$. We claim then that

$$\lim_{\lambda\in\mathbb{L}}\boldsymbol{\alpha}_\lambda = \frac{1}{\mathsf{r}^p}\mathfrak{p}_\sharp\vartheta+m\delta_{\mathfrak{o}} =: \boldsymbol{\alpha}.$$

Take any $\Omega\subset\mathfrak{C}[\mathsf{X}]$ open; if $\mathfrak{o}\notin\Omega$, we have

$$\begin{aligned}\liminf_{\lambda\in\mathbb{L}}\alpha_\lambda(\Omega) &= \liminf_{\lambda\in\mathbb{L}}\int_{\mathsf{X}\times[0,R]}(\chi_\Omega\circ\mathfrak{q})(x,r)\frac{1}{(\mathsf{r}^p\circ\mathfrak{q})(x,r)}\,\mathrm{d}\vartheta_\lambda(x,r)\\ &\geq \int_{\mathsf{X}\times[0,R]}(\chi_\Omega\circ\mathfrak{q})(x,r)\frac{1}{(\mathsf{r}^p\circ\mathfrak{q})(x,r)}\,\mathrm{d}\vartheta(x,r)\\ &= \boldsymbol{\alpha}(\Omega).\end{aligned}$$

If, on the other hand, $\mathfrak{o} \in \Omega$, we have that $O_N \subset \Omega$ for some $N \geq 1$; calling $\Omega_n := \Omega \setminus O_n$ (which is an open set), we have, for every $n \geq N$, that

$$\begin{aligned}\liminf_{\lambda \in \mathbb{L}} \boldsymbol{\alpha}_\lambda(\Omega) &\geq \liminf_{\lambda \in \mathbb{L}} \boldsymbol{\alpha}_\lambda(O_n) + \liminf_{\lambda \in \mathbb{L}} \boldsymbol{\alpha}_\lambda(\Omega_n) \\ &\geq \boldsymbol{\alpha}(\Omega_n) + m_n.\end{aligned}$$

Now we pass to the limit as $n \to +\infty$ and, using the monotone convergence theorem and the fact that $\Omega_n \uparrow \Omega \setminus \{\mathfrak{o}\}$, we obtain

$$\liminf_{\lambda \in \mathbb{L}} \boldsymbol{\alpha}_\lambda(\Omega) \geq \boldsymbol{\alpha}(\Omega \setminus \{\mathfrak{o}\}) + m = \boldsymbol{\alpha}(\Omega),$$

and this concludes the proof thanks to Portmanteau theorem (see e.g. [Bog06, Corollary 8.2.10]). □

2.2. **Functions on the product cone.** We describe a few properties of functions defined on the product cone that will be useful in the sequel. In this subsection X_1 and X_2 are completely regular spaces. Recall that the indicator function $\mathsf{I}_G : X \to \mathbb{R} \cup \{+\infty\}$ associated with a subset G of a set X is defined by

$$\mathsf{I}_G(x) := \begin{cases} 0 & \text{if } x \in G \\ +\infty & \text{otherwise} \end{cases}, \quad x \in X.$$

Definition 2.10. Consider a function $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ and a subset $\Gamma \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$. For every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ we define

$$\begin{aligned}\mathsf{H}_{x_1,x_2} :\ & \mathbb{R}^2_+ \to [0, +\infty] : \qquad (r_1, r_2) \mapsto \mathsf{H}([x_1, r_1], [x_2, r_2]), \\ \Gamma_{x_1,x_2} :=& \Big\{(r_1, r_2) \in \mathbb{R}^2_+ : ([x_1, r_1], [x_2, r_2]) \in \Gamma\Big\}.\end{aligned}$$

We say that

- H is radially p-homogeneous, $p \in [1, +\infty)$, if H_{x_1,x_2} is positively p-homogeneous for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ i.e.
$$\mathsf{H}_{x_1,x_2}(\lambda r_1, \lambda r_2) = \lambda^p \mathsf{H}_{x_1,x_2}(r_1, r_2) \quad \text{for every } \lambda > 0,\ (r_1, r_2) \in \mathbb{R}^2_+;$$
- H is radially convex if H_{x_1,x_2} is convex for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$.

We say that $\Gamma \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ is a radial cone (resp. radial convex set) if its indicator function is radially 1-homogeneous (resp. radially convex). Notice that we do not assume that a radially p-homogenoeus function vanishes at $(\mathfrak{o}_1, \mathfrak{o}_2)$ (or, similarly, that a radial cone contains $(\mathfrak{o}_1, \mathfrak{o}_2)$); however such a property follows immediately under lower semicontinuity and properness conditions, see Remark 2.11 below.

We define the (radially) 1-homogeneous, convex, and closed convex envelopes $\mathrm{hom}(\mathsf{H}), \mathrm{co}(\mathsf{H}), \overline{\mathrm{co}}(\mathsf{H}) : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ of H as (recall (2.2,2.3,2.4))

$$\begin{aligned}\mathrm{hom}(\mathsf{H})([x_1, r_1], [x_2, r_2]) &:= \inf_{\lambda \geq 0} \mathsf{H}([x_1, \lambda r_1], [x_2, \lambda r_2]), \\ \mathrm{co}(\mathsf{H})([x_1, r_1], [x_2, r_2]) &:= \mathrm{co}(\mathsf{H}_{x_1,x_2})(r_1, r_2) \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ (r_1, r_2) \in \mathbb{R}^2_+, \\ \overline{\mathrm{co}}(\mathsf{H}) &:= \overline{\mathrm{co}(\mathsf{H})}.\end{aligned}$$

Similarly, in the case of $\Gamma \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$, we define $\mathrm{hom}(\Gamma), \mathrm{co}(\Gamma)$, and $\overline{\mathrm{co}}(\Gamma)$ so that $\mathsf{I}_{\mathrm{hom}(\Gamma)} := \mathrm{hom}(\mathsf{I}_\Gamma)$, $\mathsf{I}_{\mathrm{co}(\Gamma)} := \mathrm{co}(\mathsf{I}_\Gamma)$ and $\mathsf{I}_{\overline{\mathrm{co}}(\Gamma)} := \overline{\mathrm{co}}(\mathsf{I}_\Gamma)$ respectively.

Remark 2.11. We added the terms *radial and radially* just to avoid ambiguities in the case when $\mathsf{X}_1, \mathsf{X}_2$ are linear spaces and the notions of 1-homogeneity and convexity could also refer to the joint behaviour of H w.r.t. all the variables x_i, r_i. In this paper, we will always interpret convexity and 1-homogeneity w.r.t. the radial variables r_i.

Notice that $\mathrm{hom}(\Gamma)$ (resp. $\mathrm{co}(\Gamma), \overline{\mathrm{co}}(\Gamma)$) is the set whose sections $\mathrm{hom}(\Gamma)_{x_1,x_2}$ are the cone (resp. convex, closed convex) envelopes of the corresponding sections Γ_{x_1,x_2} of Γ. In particular, racalling Carathéodory Theorem in $\mathbb{R}^2$, we have

$$\mathrm{hom}\,(\Gamma) := \bigcup_{\lambda>0} \Big\{([x_1,\lambda r_1],[x_2,\lambda r_2]) : ([x_1,r_1],[x_2,r_2]) \in \Gamma\Big\},$$

$$\mathrm{co}\,(\Gamma) := \bigcup \Big\{\Big([x_1,\sum_{i=0}^{2}\alpha_i r_1^i],[x_2,\sum_{i=0}^{2}\alpha_i r_2^i]\Big) : ([x_1,r_1^i],[x_2,r_2^i]) \in \Gamma,\ \alpha_i \geq 0,\ \sum_{i=0}^{2}\alpha^i = 1\Big\}.$$

We also note that if H is proper (i.e. not identically $+\infty$), lower semicontinuous, and radially 1-homogeneous, then $\mathsf{H}(\mathfrak{o}_1,\mathfrak{o}_2) = 0$.

The following result is a simple consequence of the 1-homogeneity property.

Lemma 2.12. *Let X_i, $i = 1,2$ be completely regular spaces, let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to [0,+\infty]$ be a radially 1-homogeneous Borel function and let $(\mu_1,\mu_2) \in \mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$. Then*

$$\inf\left\{\int \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)\right\} = \inf\left\{\int \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)\cap\mathcal{P}(\mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1,\mathsf{X}_2])\right\},$$

where

$$R(\mu_1,\mu_2) := \mu_1(\mathsf{X}_1) + \mu_2(\mathsf{X}_2). \tag{2.13}$$

Proof. It is of course enough to prove the $\geq$ inequality. If $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ we can assume that $\boldsymbol{\alpha}(\{(\mathfrak{o}_1,\mathfrak{o}_2)\}) = 0$ (if not, we can replace $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}_{\mathfrak{o}} = \boldsymbol{\alpha} \llcorner \mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1,\mathsf{X}_2]$ which has the same homogeneous marginals and a lower H-cost, since $\mathsf{H}(\mathfrak{o}_1,\mathfrak{o}_2) \geq 0$). By Lemma 2.8, we have that

$$\tilde{\boldsymbol{\alpha}} := \mathrm{dil}_{\vartheta_{\boldsymbol{\alpha}},1,1}(\boldsymbol{\alpha}) \in \mathfrak{H}^1(\mu_1,\mu_2)\cap\mathcal{P}(\mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1,\mathsf{X}_2])$$

and

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\tilde{\boldsymbol{\alpha}}.$$

This concludes the proof. □

Remark 2.13 (1-homogeneity and q-homogeneity). Given $q \in [1,+\infty)$, we define the map $\mathsf{T}_q : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]$ as

$$\mathsf{T}_q([x_1,r_1],[x_2,r_2]) := ([x_1,r_1^{1/q}],[x_2,r_2^{1/q}]). \tag{2.14}$$

It is easy to check that $(\mathsf{T}_q)_\sharp$ is a bijective transformation from $\mathfrak{H}^1(\mu_1,\mu_2)$ to $\mathfrak{H}^q(\mu_1,\mu_2)$ for any pair $(\mu_1,\mu_2) \in \mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$. Moreover, if $\mathsf{H}_q : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to [0,\infty]$ is radially q-homogeneous, then $\mathsf{H} := \mathsf{H}_q \circ \mathsf{T}_q$ is radially 1-homogeneous and

$$\inf\left\{\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}_q\,\mathrm{d}\boldsymbol{\alpha}_q : \boldsymbol{\alpha}_q \in \mathfrak{H}^q(\mu_1,\mu_2)\right\} = \inf\left\{\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)\right\}. \tag{2.15}$$

As an immediate consequence of the above formula, many of the results of this work can also be stated in the case of a q-homogeneous cost function H_q, provided that the unbalanced optimal transport functional is defined by using the q-homogeneous marginals $\mathfrak{h}^q$ and the corresponding set $\mathfrak{H}^q$ as in (2.15). For this reason, in order to keep a simpler notation we will limit our analysis to the 1-homogeneous case.

2.3. **Examples of cost functions.** In this subsection we present some examples of functions $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$, where X_1 and X_2 are completely regular spaces, satisfying (some of) the hypotheses we will assume throughout the paper. See also the examples of [Chi+18b, Section5].

Notice that given two radially 1-homogeneous, convex, lower semicontinuous and proper functions $\mathsf{H}_1, \mathsf{H}_2 : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty)$, and two lower semicontinuous and proper functions $\mathsf{c}_1, \mathsf{c}_2 : \mathsf{X}_1 \times \mathsf{X}_2 \to [0, +\infty)$ also the function

$$\mathsf{H}([x_1, r_1], [x_2, r_2]) = \mathsf{c}_1(x_1, x_2)\mathsf{H}_1([x_1, r_1], [x_2, r_2]) + \mathsf{c}_2(x_1, x_2)\mathsf{H}_2([x_1, r_1], [x_2, r_2])$$

is radially 1-homogeneous, convex, lower semicontinuous and proper. In this way, many other examples can be obtained starting from the ones presented below. Notice moreover that the radial behavior of a 1-homogeneous convex function $\mathsf{H}_{x_1,x_2} : \mathbb{R}_+ \times \mathbb{R}_+ \to [0, +\infty]$ can be characterized by the reduced function and its recession slope

$$\mathsf{h}(x_1, x_2; g) = \mathsf{h}_{x_1,x_2}(g) := \mathsf{H}_{x_1,x_2}(1, g), \quad h^\infty_{x_1,x_2} := \lim_{g\to+\infty} g^{-1}\mathsf{h}_{x_1,x_2}(g), \tag{2.16}$$

since

$$\mathsf{H}_{x_1,x_2}(r_1, r_2) = \begin{cases} r_1 \mathsf{h}_{x_1,x_2}(r_2/r_1) & \text{if } r_1 > 0, \\ r_2 \mathsf{h}^\infty_{x_1,x_2} & \text{if } r_1 = 0,\ r_2 > 0, \\ 0 & \text{if } r_1 = r_2 = 0. \end{cases} \tag{2.17}$$

2.3.1. *Mass-space product costs.* We consider cost functions of the form

$$\mathsf{H}([x_1, r_1], [x_2, r_2]) := \mathsf{H}_+(r_1, r_2) + \mathsf{H}_-(r_1, r_2)\mathsf{c}(x_1, x_2), \quad ([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2],$$

where $\mathsf{H}_+, \mathsf{H}_- : \mathbb{R}^2_+ \to \mathbb{R}$ are convex, 1-homogeneous and lower semicontinuous and $\mathsf{c} : \mathsf{X}_1 \times \mathsf{X}_2 \to [0, +\infty]$ is a lower semicontinuous function satisfying

$$\mathsf{H}_+(r_1, r_2) + \mathsf{H}_-(r_1, r_2)\mathsf{c}(x_1, x_2) \geq 0 \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ r_1, r_2 \geq 0.$$

Possible choices of H_+ and H_- (to be then coupled with a suitable cost c) are given by e.g.

(1) $m_p(r_1, r_2) := (r_1^p + r_2^p)^{\frac{1}{p}}, \quad p \in [1, +\infty)$,
(2) $m_p(r_1, r_2) := -m_p(r_1, r_2), \quad p \in (-\infty, 0) \cup (0, 1)$,
(3) $m_\infty(r_1, r_2) = r_1 \vee r_2, \quad m_{-\infty}(r_1, r_2) = -(r_1 \wedge r_2), \quad m_0 = -\sqrt{r_1 r_2}$,
(4) $n_\alpha(r_1, r_2) := \dfrac{|r_1 - r_2|^\alpha}{(r_1 + r_2)^{\alpha-1}}, \quad \alpha \geq 1$,
(5) $|r_1^\alpha - r_2^\alpha|^{1/\alpha}, \quad -|r_1^\alpha + r_2^\alpha|^{1/\alpha}, \quad 0 < \alpha \leq 1$.

In the various examples we are adopting standard conventions when the expressions are not defined. In particular, for $p < 0$ we set $m_p(r_1, r_2) = 0$ if $r_1 r_2 = 0$, $n_\alpha(0, 0) = 0$.

2.3.2. *Homogeneous marginal perspective functional.* Following [LMS18, Section 5] we can build H starting from two entropy functions $F_i : \mathsf{X}_i \to [0, +\infty]$, $i = 1, 2$ and a proper and lower semicontinuous cost function $\mathsf{c} : \mathsf{X}_1 \times \mathsf{X}_2 \to [0, +\infty]$. Assuming that each F_i, $i = 1, 2$ is convex, lower semicontinuous and finite in at least one positive point, we can define, for every number $c \in [0, +\infty]$, the function $\mathsf{H}_c : \mathbb{R}^2_+ \to [0, +\infty]$, as the lower semicontinuous envelope of

$$\tilde{\mathsf{H}}_c(r_1, r_2) := \begin{cases} \inf_{\theta > 0} \left\{ r_1 F_1(\theta/r_1) + r_2 F_2(\theta/r_2) + \theta c \right\}, & \text{if } c \in [0, +\infty) \\ F_1(0) r_1 + F_2(0) r_2 & \text{if } c = \infty, \end{cases} \qquad r_1, r_2 \in \mathbb{R}^2_+.$$

The function $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ is then defined as

$$\mathsf{H}([x_1, r_1], [x_2, r_2]) := \mathsf{H}_{\mathsf{c}(x_1,x_2)}(r_1, r_2), \quad ([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2].$$

Such function H is radially convex and 1-homogeneous (see [LMS18, Lemma 5.3]). Possible choices (see e.g. [LMS16; De 20]) for F_i are given by:

(1) power like entropies: for $p \in \mathbb{R}$ we define

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

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

(2) indicator functions: for numbers $0 \leq a \leq 1 \leq b \leq +\infty$ we define

$$I_{[a,b]}(s) := \begin{cases} 0 & \text{if } s \in [a,b], \\ +\infty & \text{if } s \notin [a,b]; \end{cases}$$

(3) χ^α divergences: for a parameter $\alpha \geq 1$ we define

$$\chi^\alpha(s) := |s-1|^\alpha, \quad s \in \mathbb{R}.$$

Some of the corresponding expression for H are for example

(1) In case of power like entropies with $F_1 = F_2 = U_p$:

$$\mathsf{H}([x_1,r_1],[x_2,r_2]) = \begin{cases} \dfrac{1}{p}\left[(r_1+r_2) - \dfrac{r_1 r_2}{(r_1^{p-1}+r_2^{p-1})^{\frac{1}{p-1}}}(2-(p-1)\mathsf{c}(x_1,x_2))_+^{\frac{p}{p-1}}\right], & \text{if } p \neq 0,1, \\ (\sqrt{r_1}-\sqrt{r_2})^2 + 2\sqrt{r_1 r_2}(1-e^{-\mathsf{c}(x_1,x_2)/2}), & \text{if } p = 1, \\ r_1\log(r_1) + r_2\log(r_2) - (r_1+r_2)\log\left(\dfrac{r_1+r_2}{2+\mathsf{c}(x_1,x_2)}\right), & \text{if } p = 0. \end{cases}$$

In particular, in case $p = 1$, for $\mathsf{X}_1 = \mathsf{X}_2 = \mathbb{R}^d$ we can chose as cost functions c

$$\mathsf{c}_{\mathsf{GHK}}(x_1,x_2) := |x_1-x_2|^2, \quad \mathsf{c}_{\mathsf{HK}}(x_1,x_2) := \begin{cases} -\log(\cos^2(|x_1-x_2|)) & \text{if } |x_1-x_2| < \pi/2, \\ +\infty & \text{else.} \end{cases}$$

The resulting cost functions H are thus given respectively by

$$\mathsf{H}_{\mathsf{GHK}}([x_1,r_1],[x_2,r_2]) = r_1 + r_2 - 2\sqrt{r_1 r_2}e^{-|x_1-x_2|^2/2}, \tag{2.18}$$

$$\mathsf{H}_{\mathsf{HK}}([x_1,r_1],[x_2,r_2]) = r_1 + r_2 - 2\sqrt{r_1 r_2}\cos(|x_1-x_2| \wedge \pi/2). \tag{2.19}$$

(2.18) and (2.19) are metrics on $\mathfrak{C}[\mathbb{R}^d]$, inducing the same canonical cone topology. In particular, H_{HK} is related to (2.9) via the transformation in Remark 2.13 with $q = 2$ (apart from the specific value of the truncation constant). Both functions are considered in [LMS18] and they generate the Gaussian Hellinger-Kantorovich and the Hellinger-Kantorovich metrics on non-negative measures, respectively.

(2) In case of indicator functions with $F_1 = F_2 = I_{[a,b]}$:

$$\mathsf{H}([x_1,r_1],[x_2,r_2]) = \begin{cases} 0 & \text{if } \frac{a}{b} \leq \frac{r_1}{r_2} \leq \frac{b}{a}, \\ +\infty & \text{else,} \end{cases}$$

where $\frac{b}{a} = +\infty$ if $a = 0$ and $\frac{a}{b} = 0$ if $b = +\infty$.

(3) In case of the χ^1 divergence with $F_1 = F_2 = \chi^1$:

$$\mathsf{H}([x_1,r_1],[x_2,r_2]) = |r_2 - r_1| + (\mathsf{c}(x_1,x_2) \wedge 2)(r_1 \wedge r_2).$$

3. Convexification and duality

This section presents the main convexification and duality results. In this section X_1 and X_2 are completely regular spaces.

3.1. Unbalanced Optimal Transport as convex relaxation of singular cost on Dirac masses.

Definition 3.1 (Homogeneous conical formulation of Unbalanced Optimal Transport problems)**.** Let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ be a proper (i.e. not identically equal to $+\infty$) Borel function. We define the singular cost and the Unbalanced Optimal Transport cost $\mathscr{S}_\mathsf{H}, \mathscr{U}_\mathsf{H} : \mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2) \to [0, +\infty]$ respectively as:

$$\mathscr{S}_\mathsf{H}(\mu_1, \mu_2) := \begin{cases} \mathsf{H}([x_1, r_1], [x_2, r_2]) & \text{if } \mu_1 = r_1\delta_{x_1},\ \mu_2 = r_2\delta_{x_2},\ (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ (r_1, r_2) \in \mathbb{R}^2_+, \\ +\infty & \text{elsewhere.} \end{cases}$$

$$\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) := \begin{cases} \inf\left\{ \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \right\} & \text{if } (\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2), \\ +\infty & \text{elsewhere.} \end{cases}$$

The aim of this section is to study the relation between $\mathscr{S}_\mathsf{H}$ and $\mathscr{U}_\mathsf{H}$; in particular we are interested in studying the lower semicontinuous and convex envelope of $\mathscr{S}_\mathsf{H}$.

Let us first recall two simple properties in the following remarks.

Remark 3.2. The value of $\mathscr{U}_\mathsf{H}(\mu_1, \mu_2)$ does not depend on the ambient spaces X_i (recall the discussion of Remark 2.7): if $\mathsf{X}_i \subset \tilde{\mathsf{X}}_i$, H is the restriction to $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ of a function $\tilde{\mathsf{H}}$ defined in $\mathfrak{C}[\tilde{\mathsf{X}}_1, \tilde{\mathsf{X}}_2]$, and $\tilde{\mu}_i \in \mathcal{M}_+(\tilde{\mathsf{X}}_i)$ are the canonical extensions of μ_i, then Remark 2.7 yields

$$\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) = \mathscr{U}_{\tilde{\mathsf{H}}}(\tilde{\mu}_1, \tilde{\mu}_2). \tag{3.1}$$

In particular, we can always "embed" Unbalanced Optimal Transport problems in $\mathsf{X}_{i,o}$ or, equivalently, suppose that there is at least one μ_i-negligible point o_i in each space X_i.

Remark 3.3. If $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ is a proper Borel function, then

$$\overline{\mathscr{S}_\mathsf{H}} = \mathscr{S}_{\overline{\mathsf{H}}}.$$

Indeed, both are equal to $+\infty$ outside the closed set $\Delta_+(\mathsf{X}_1) \times \Delta_+(\mathsf{X}_2)$ and the equality on $\Delta_+(\mathsf{X}_1) \times \Delta_+(\mathsf{X}_2)$ follows by Lemma 2.4.

For this reason and to exploit Lemma 2.12, we will usually assume that

$$\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty] \text{ is a proper, radially 1-homogeneous and l.s.c. function.} \tag{3.2}$$

In the following result we prove that the Unbalanced Optimal Transport cost $\mathscr{U}_\mathsf{H}$ is a lower semicontinuous convex functional in $\mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ and the infimum in the Definition of $\mathscr{U}_\mathsf{H}$ is attained, when the problem is feasible.

Theorem 3.4 (Existence of solutions to the Unbalanced Optimal Transport problem)**.** *Let* H *be as in* (3.2) *and let* $\mathscr{U}_\mathsf{H}$ *be as in Definition 3.1.*

(1) For every $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ *such that* $\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) < +\infty$*, there exists an optimal 1-homogeneous coupling* $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \cap \mathcal{P}(\mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1, \mathsf{X}_2])$ *not charging* $\mathfrak{o} = (\mathfrak{o}_1, \mathfrak{o}_2)$ *such that*

$$\mathscr{U}_\mathsf{H}(\mu_1, \mu_2) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha},$$

where $R(\mu_1, \mu_2)$ *is as in* (2.13)*.*

(2) $\mathscr{U}_\mathsf{H}$ *is a lower semicontinuous convex functional in* $\mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ *and satisfies*

$$\mathscr{U}_\mathsf{H}(r_1\delta_{x_1}, r_2\delta_{x_2}) \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \tag{3.3}$$

for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ *and every* $(r_1, r_2) \in \mathbb{R}^2_+$*.*

(3) If, in addition, H *is also radially convex, then* (3.3) *is an equality:*

$$\mathscr{U}_\mathsf{H}(r_1\delta_{x_1}, r_2\delta_{x_2}) = \mathsf{H}([x_1, r_1], [x_2, r_2]).$$

Proof. (1) Let $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$; by Lemma 2.12, it holds

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) = \inf \left\{ \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \cap \mathcal{P}(\mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1, \mathsf{X}_2]) \right\}.$$

Thanks to Lemma 2.9, we have that $\mathfrak{H}^1(\mu_1, \mu_2) \cap \mathcal{P}(\mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1, \mathsf{X}_2])$ is compact and the lower semicontinuity of H gives that the functional

$$\boldsymbol{\alpha} \mapsto \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} \tag{3.4}$$

is lower semicontinuous. We can thus conclude that a minimizer exists by the direct method in Calculus of Variations. Possibly replacing $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}_{\mathfrak{o}} = \boldsymbol{\alpha} \llcorner \mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1, \mathsf{X}_2]$ we obtain an optimal homogeneous coupling not charging $(\mathfrak{o}_1, \mathfrak{o}_2)$.

(2) The convexity of $\mathscr{U}_{\mathsf{H}}$ follows by the convexity of the constraints and the linearity of the objective function characterizing $\mathscr{U}_{\mathsf{H}}$: if $(\mu_1^1, \mu_2^1), (\mu_2^1, \mu_2^2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ and $t \in [0, 1]$, we can take, thanks to point (1), $\boldsymbol{\alpha}_1 \in \mathfrak{H}^1(\mu_1^1, \mu_2^1)$ and $\boldsymbol{\alpha}_2 \in \mathfrak{H}^1(\mu_1^2, \mu_2^2)$ such that

$$\mathscr{U}_{\mathsf{H}}(\mu_1^1, \mu_2^1) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_1, \quad \mathscr{U}_{\mathsf{H}}(\mu_1^2, \mu_2^2) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_2.$$

It is then enough to observe that $\boldsymbol{\alpha} := (1-t)\boldsymbol{\alpha}_1 + t\boldsymbol{\alpha}_2 \in \mathfrak{H}^1((1-t)\mu_1^1 + t\mu_1^2, (1-t)\mu_2^1 + t\mu_2^2)$.

The lower semicontinuity of $\mathscr{U}_{\mathsf{H}}$ is a consequence of Lemma 2.12: if $\{(\mu_1^\lambda, \mu_2^\lambda)\}_{\lambda\in\mathbb{L}} \subset \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$ is a net converging to $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$, we can set $R := \sup_\lambda R(\mu_1^\lambda, \mu_2^\lambda)$ (see (2.13)) and consider, for every $\lambda \in \mathbb{L}$, some $\boldsymbol{\alpha}_\lambda \in \mathfrak{H}^1(\mu_1^\lambda, \mu_2^\lambda) \cap \mathcal{P}(\mathfrak{C}_R[\mathsf{X}_1, \mathsf{X}_2])$ such that

$$\mathscr{U}_{\mathsf{H}}(\mu_1^\lambda, \mu_2^\lambda) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_\lambda.$$

We can thus use Lemma 2.9 to extract a convergent subnet of $(\boldsymbol{\alpha}_\lambda)_{\lambda\in\mathbb{L}}$ with limit $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$. Using again the lower semicontinuity of the functional in (3.4), we can conclude that $\mathscr{U}_{\mathsf{H}}$ is lower semicontinuous.

(3.3) follows by the fact that

$$\boldsymbol{\alpha} = \delta_{([x_1,r_1],[x_2,r_2])}$$

is an element of $\mathfrak{H}^1(r_1\delta_{x_1}, r_2\delta_{x_2})$.

(3) Let us assume that H is radially convex and let $\boldsymbol{\alpha} \in \mathfrak{H}^1(r_1\delta_{x_1}, r_2\delta_{x_2})$ be such that

$$\mathscr{U}_{\mathsf{H}}(r_1\delta_{x_1}, r_2\delta_{x_2}) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha};$$

we observe that $\boldsymbol{\alpha}$ is concentrated on

$$\big\{\lambda_1[x_1, 1], \lambda_2[x_2, 1] : \lambda_1, \lambda_2 \geq 0\big\}$$

with

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{r}_1\,\mathrm{d}\boldsymbol{\alpha} = r_1, \quad \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{r}_2\,\mathrm{d}\boldsymbol{\alpha} = r_2.$$

Hence, using Jensen's inequality and the convexity of H_{x_1,x_2}, we have

$$\begin{aligned}\mathscr{U}_{\mathsf{H}}(r_1\delta_{x_1}, r_2\delta_{x_2}) &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} \\ &= \int_{\mathbb{R}^2_+} \mathsf{H}_{x_1,x_2}\,\mathrm{d}(\mathsf{r}_1,\mathsf{r}_2)_\sharp\boldsymbol{\alpha} \\ &\geq \mathsf{H}_{x_1,x_2}\left(\int_{\mathbb{R}^2_+}(\mathsf{r}_1,\mathsf{r}_2)\,\mathrm{d}\boldsymbol{\alpha}\right) \\ &= \mathsf{H}_{x_1,x_2}(r_1,r_2) \\ &= \mathsf{H}([x_1,r_1],[x_2,r_2]). \end{aligned}$$

□

Thanks to Theorem 3.4, given $(\mu_1,\mu_2) \in \mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$ such that $\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) < +\infty$, the set

$$\mathfrak{H}^1_o(\mu_1,\mu_2) := \left\{\boldsymbol{\alpha}\in\mathfrak{H}^1(\mu_1,\mu_2) : \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]}\mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} = \mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2)\right\} \tag{3.5}$$

is not empty.

In the following we give a direct proof that the Unbalanced Optimal Transport cost $\mathscr{U}_{\mathsf{H}}$ is the lower semicontinuous convex envelope of the singular cost $\mathscr{S}_{\mathsf{H}}$. This result will also be obtained in Theorem 3.15 as a simple consequence of the dual characterization of $\mathscr{U}_{\mathsf{H}}$. However, the following proof highlights the role played by the discrete 1-homogeneous marginals, which is not evident in the proof of Theorem 3.15.

Theorem 3.5 (Convex l.s.c. envelope)**.** *Let H be as in (3.2) and let $\mathscr{S}_{\mathsf{H}}$ and $\mathscr{U}_{\mathsf{H}}$ be as in Definition 3.1. Then*

$$\overline{\mathrm{co}}\,(\mathscr{S}_{\mathsf{H}}) = \mathscr{U}_{\mathsf{H}} \ \textit{in}\ \mathcal{M}(\mathsf{X}_1)\times\mathcal{M}(\mathsf{X}_2).$$

Proof. We only need to prove equality on $\mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$, being both functions equal to $+\infty$ outside it. First of all let us compute $\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})$ on $\mathrm{co}\,(\Delta_+(\mathsf{X}_1)\times\Delta_+(\mathsf{X}_2)) = \mathrm{co}\,(\Delta_+(\mathsf{X}_1))\times\mathrm{co}\,(\Delta_+(\mathsf{X}_2))$ (outside this set $\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})$ is equal to $+\infty$). We take $(\mu_1,\mu_2)\in\mathrm{co}\,(\Delta_+(\mathsf{X}_1))\times\mathrm{co}\,(\Delta_+(\mathsf{X}_2))$ and we observe that any finite set in $\Delta_+(\mathsf{X}_1)\times\Delta_+(\mathsf{X}_2)$ is always contained in the cartesian product of two finite sets $M_1 = \{\mu_1^i = r_1^i\delta_{x_1^i} : i = 1,\cdots,I\}$ and $M_2 = \{\mu_2^j = r_1^j\delta_{x_2^j} : j = 1,\cdots,J\}$, so that

$$\begin{aligned}\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1,\mu_2) &= \inf\left\{\sum_{ij}\gamma_{ij}\mathscr{S}_{\mathsf{H}}(r_1^i\delta_{x_1^i}, r_2^j\delta_{x_2^j}) : \left(r_1^i,x_1^i,r_2^j,x_2^j,\gamma_{ij}\right)_{ij}\in A(\mu_1,\mu_2)\right\}, \\ A(\mu_1,\mu_2) &:= \left\{\left(r_1^i,x_1^i,r_2^j,x_2^j,\gamma_{ij}\right)_{ij} : (\mu_1,\mu_2) = \sum_{ij}\gamma_{ij}(r_1^i\delta_{x_1^i}, r_2^j\delta_{x_2^j}),\ \sum_{ij}\gamma_{ij} = 1, \gamma_{ij}\geq 0\right\}.\end{aligned}$$

In particular it holds that

$$\mu_1 = \sum_{ij}\gamma_{ij}r_1^i\delta_{x_1^i},\quad \mu_2 = \sum_{ij}\gamma_{ij}r_2^j\delta_{x_2^j}$$

for some $r_1^i, r_2^j\in\mathbb{R}_+$ and some $x_1^i\in\mathsf{X}_1$, $x_2^j\in\mathsf{X}_2$ not necessarily distinct points. Setting

$$\boldsymbol{\alpha} := \sum_{i,j}\gamma_{ij}\delta_{([x_1^i,r_1^i],[x_2^j,r_2^j])}\in\mathcal{P}_{\mathrm{f}}(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]) \tag{3.6}$$

one immediately checks that

$$\boldsymbol{\alpha}\in\mathfrak{H}^1(\mu_1,\mu_2),\quad \sum_{ij}\gamma_{ij}\mathscr{S}_{\mathsf{H}}(r_1^i\delta_{x_1^i}, r_2^j\delta_{x_2^j}) = \int\mathsf{H}\ \mathrm{d}\boldsymbol{\alpha}. \tag{3.7}$$

Conversely, every $\boldsymbol{\alpha} \in \mathcal{P}_{\mathrm{f}}(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) \cap \mathfrak{H}^1(\mu_1, \mu_2)$ can be written as in (3.6) with coefficients $(r_1^i, x_1^i, r_2^j, x_2^j, \gamma_{ij})_{ij}$ in $A(\mu_1, \mu_2)$; the integral identity of (3.7) eventually shows that $\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})$ can be written as

$$\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1, \mu_2) = \inf\left\{ \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \cap \mathcal{P}_{\mathrm{f}}(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) \right\}.$$

Reasoning as in the proof of Lemma 2.12, we have that

$$\mathrm{co}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1, \mu_2) = \inf\left\{ \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2) \cap \mathcal{P}_{\mathrm{f}}(\mathfrak{C}_*) \right\},$$

where $\mathfrak{C}_* := \mathfrak{C}_{R(\mu_1,\mu_2)}[\mathsf{X}_1, \mathsf{X}_2]$, with $R(\mu_1, \mu_2)$ as in (2.13). Thus

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) \le \mathrm{co}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1, \mu_2) \quad \text{ for every } \ (\mu_1, \mu_2) \in \mathrm{co}\,(\Delta_+(\mathsf{X}_1)) \times \mathrm{co}\,(\Delta_+(\mathsf{X}_2))\,.$$

Moreover $\mathscr{U}_{\mathsf{H}}$ is lower semicontinuous and convex hence, by definition of $\overline{\mathrm{co}}\,(\mathscr{S}_{\mathsf{H}})$, it must hold

$$\overline{\mathrm{co}}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1, \mu_2) \ge \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) \quad \text{ for every } \ (\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_1).$$

Then, in order to prove equality, we only need to prove the other inequality. To do so, fixed $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$, we prove that there exists a net $\{(\mu_1^\eta, \mu_2^\eta)\}_{\eta \in \mathbb{E}} \subset \mathrm{co}\,(\Delta_+(\mathsf{X}_1)) \times \mathrm{co}\,(\Delta_+(\mathsf{X}_2))$ s.t. $\lim_\eta (\mu_1^\eta, \mu_2^\eta) = (\mu_1, \mu_2)$ and a net

$$\{\boldsymbol{\gamma}_\eta\}_{\eta \in \mathbb{E}} \subset \mathcal{M}_{+,\mathrm{f}}(\mathfrak{C}_*) \cap \mathcal{P}(\mathfrak{C}_*)$$

s.t. $\boldsymbol{\gamma}_\eta \in \mathfrak{H}^1(\mu_1^\eta, \mu_2^\eta)$ for every $\eta \in \mathbb{E}$ satisfying

$$\lim_{\eta \in \mathbb{E}} \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma}_\eta = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^*$$

where $\boldsymbol{\alpha}^* \in \mathfrak{H}^1_o(\mu_1, \mu_2) \cap \mathcal{P}(\mathfrak{C}_*)$ (see (3.5)). To do so, we use Lemma 2.3 with $\mathsf{X} := \mathfrak{C}_*$, $f := \mathsf{H}$, $\alpha := \boldsymbol{\alpha}^*$ and we find $\{\boldsymbol{\gamma}_\eta\}_{\eta \in \mathbb{E}} \subset \mathcal{M}_{+,\mathrm{f}}(\mathfrak{C}_*) \cap \mathcal{P}(\mathfrak{C}_*)$ s.t.

$$\lim_{\eta \in \mathbb{E}} \boldsymbol{\gamma}_\eta = \boldsymbol{\alpha}^*, \quad \lim_{\eta \in \mathbb{E}} \int_{\mathfrak{C}_*} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma}_\eta = \int_{\mathfrak{C}_*} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^*.$$

Finally we can define

$$\mu_1^\eta := \mathfrak{h}_1^1(\boldsymbol{\gamma}_\eta), \quad \mu_2^\eta := \mathfrak{h}_2^1(\boldsymbol{\gamma}_\eta) \quad \text{ for every } \ \eta \in \mathbb{E}.$$

Obviously $\boldsymbol{\gamma}_\eta \in \mathfrak{H}^1(\mu_1^\eta, \mu_2^\eta)$ and $\mu_i = \lim_{\eta \in \mathbb{E}} \mu_i^\eta$, indeed if $\varphi_i \in \mathrm{C}_b(\mathsf{X}_i)$, then

$$\begin{aligned}
\lim_{\eta \in \mathbb{E}} \int_{\mathsf{X}_i} \varphi_i\,\mathrm{d}\mu_i^\eta &= \lim_{\eta \in \mathbb{E}} \int_{\mathsf{X}_i} \varphi_i\,\mathrm{d}\mathfrak{h}_i^1(\boldsymbol{\gamma}_\eta) = \lim_{\eta \in \mathbb{E}} \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\varphi \circ \mathsf{x}_i)\mathsf{r}_i\,\mathrm{d}\boldsymbol{\gamma}_\eta \\
&= \lim_{\eta \in \mathbb{E}} \int_{\mathfrak{C}_*} (\varphi_i \circ \mathsf{x}_i)\mathsf{r}_i\,\mathrm{d}\boldsymbol{\gamma}_\eta = \int_{\mathfrak{C}_*} (\varphi_i \circ \mathsf{x}_i)\mathsf{r}_i\,\mathrm{d}\boldsymbol{\alpha}^* \\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\varphi_i \circ \mathsf{x}_i)\mathsf{r}_i\,\mathrm{d}\boldsymbol{\alpha}^* = \int_{\mathsf{X}_i} \varphi_i\,\mathrm{d}\mathfrak{h}_i^1(\boldsymbol{\alpha}^*) \\
&= \int_{\mathsf{X}_i} \varphi_i\,\mathrm{d}\mu_i,
\end{aligned}$$

where we have used that $(\varphi \circ \mathsf{x}_i)\mathsf{r}_i \in \mathrm{C}_b(\mathfrak{C}_*)$ and the convergence of $\boldsymbol{\gamma}_\eta$ to $\boldsymbol{\alpha}^*$ in $\mathcal{P}(\mathfrak{C}_*)$. Notice that, in general, it is not true that $(\varphi \circ \mathsf{x}_i)\mathsf{r}_i \in \mathrm{C}_b(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2])$. Finally

$$\begin{aligned}\overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H})\,(\mu_1,\mu_2) &= \inf\Big\{\liminf_\lambda \mathrm{co}\,(\mathscr{S}_\mathsf{H})\,(\mu_1^\lambda,\mu_2^\lambda) : \{(\mu_1^\lambda,\mu_2^\lambda)\}_{\lambda\in\mathbb{L}} \subset \mathrm{co}\,(\Delta_+(\mathsf{X}_1)\times\Delta_+(\mathsf{X}_2))\,,\\ &\qquad (\mu_1,\mu_2) = \lim_\lambda(\mu_1^\lambda,\mu_2^\lambda)\Big\}\\ &\le \liminf_\eta \mathrm{co}\,(\mathscr{S}_\mathsf{H})\,(\mu_1^\eta,\mu_2^\eta) \le \liminf_\eta \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma}_\eta\\ &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^* = \mathscr{U}_\mathsf{H}(\mu_1,\mu_2).\end{aligned}$$ □

Theorem 3.5 immediately yields the following useful property.

Theorem 3.6 (Sublinearity of $\mathscr{U}_\mathsf{H}$)**.** *The functional $\mathscr{U}_\mathsf{H}$ is sublinear (i.e. convex and positively 1-homogeneous) in $\mathcal{M}_+(\mathsf{X}_1)\times\mathcal{M}_+(\mathsf{X}_2)$: for every $\mu_i', \mu_i'' \in \mathcal{M}_+(\mathsf{X}_i)$ and every $\lambda', \lambda'' \in [0,+\infty)$*

$$\mathscr{U}_\mathsf{H}(\lambda'\mu_1' + \lambda''\mu_1'', \lambda'\mu_2' + \lambda''\mu_2'') \le \lambda'\mathscr{U}_\mathsf{H}(\mu_1',\mu_2') + \lambda''\mathscr{U}_\mathsf{H}(\mu_1'',\mu_2''). \tag{3.8}$$

3.2. **Disintegration, barycentric projection and decomposition of homogeneous couplings.** Every plan $\boldsymbol{\alpha} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$ can be expressed as a superposition of a Borel family of probability measures $(\alpha_{\boldsymbol{x}})_{\boldsymbol{x}}$ in $\mathbb{R}^2_+$, $\boldsymbol{x} = (x_1,x_2) \in \mathsf{X}_{1,o}\times\mathsf{X}_{2,o}$. It is sufficient to consider the map $\mathbf{x} := (\mathsf{x}_1,\mathsf{x}_2) : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to \mathsf{X}_{1,o}\times\mathsf{X}_{2,o}$, where $\mathsf{x}_i : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to \mathsf{X}_{i,o}$ are defined in Section 2.1, and the plan $\boldsymbol{\gamma} := (\mathsf{x}_1,\mathsf{x}_2)_\sharp\boldsymbol{\alpha} \in \mathcal{M}_+(\mathsf{X}_{1,o}\times\mathsf{X}_{2,o})$. The disintegration of $\boldsymbol{\alpha}$ w.r.t. γ yields the Borel family $(\alpha_{\boldsymbol{x}})_{\boldsymbol{x}} \subset \mathcal{P}(\mathbb{R}^2_+)$ satisfying

$$\boldsymbol{\alpha} = \int_{\mathsf{X}_{1,o}\times\mathsf{X}_{2,o}} \alpha_{x_1,x_2}\,\mathrm{d}\boldsymbol{\gamma}(x_1,x_2), \quad \boldsymbol{\gamma} = \mathbf{x}_\sharp\boldsymbol{\alpha}. \tag{3.9}$$

Notice that if $\boldsymbol{\alpha}(\{\mathfrak{o}\}) = 0$ then $\boldsymbol{\gamma}$ is concentrated on $\mathsf{X}_{1,o}\times\mathsf{X}_{2,o}\setminus\{\boldsymbol{o}\}$. If moreover $\boldsymbol{\alpha} \in \mathcal{M}^1_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$ then we can define

$$\boldsymbol{\varrho} := (\varrho_1,\varrho_2), \quad \varrho_i(x_1,x_2) := \int_{\mathbb{R}^2_+} r_i\,\mathrm{d}\alpha_{x_1,x_2}(r_1,r_2), \quad \varrho_i \in L^1(\mathsf{X}_{1,o}\times\mathsf{X}_{2,o},\boldsymbol{\gamma}),\ \varrho_i \ge 0. \tag{3.10}$$

Definition 3.7 (Barycentric projection and reduced couplings)**.** The barycentric projection of a nontrivial plan $\boldsymbol{\alpha} \in \mathcal{M}^1_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$ is the plan $\boldsymbol{\alpha}_b \in \mathcal{M}^1_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$ defined by

$$\boldsymbol{\alpha}_b := \mathsf{T}[\boldsymbol{\varrho}]_\sharp\gamma, \quad \text{where} \quad \mathsf{T}[\boldsymbol{\varrho}](\boldsymbol{x}) := \big([x_1,\varrho_1(\boldsymbol{x})],[x_2,\varrho_2(\boldsymbol{x})]\big), \quad \boldsymbol{x} \in \mathsf{X}_{1,o}\times\mathsf{X}_{2,o}. \tag{3.11}$$

When $\boldsymbol{\alpha}(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]) = 0$ (i.e. $\boldsymbol{\alpha}$ is the null measure) we set $\boldsymbol{\alpha}_b := \boldsymbol{\alpha}$.
We say that $\boldsymbol{\alpha}$ is a reduced plan if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_b$ or, equivalently, if $\alpha_{\boldsymbol{x}}$ is a Dirac mass $\delta_{\boldsymbol{\varrho}(\boldsymbol{x})}$ for $\boldsymbol{\gamma}$-a.e. $\boldsymbol{x} \in \mathsf{X}_{1,o}\times\mathsf{X}_{2,o}$.

The interest in the barycentric projection and in reduced couplings is justified by the following result.

Proposition 3.8 (Reduced homogeneous couplings and minimizers)**.** *If $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$, $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ then also $\boldsymbol{\alpha}_b \in \mathfrak{H}^1(\mu_1,\mu_2)$. If H satisfies (3.2) and it is radially convex then*

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_b \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}. \tag{3.12}$$

In particular, if $\mathscr{U}_\mathsf{H}(\mu_1,\mu_2) < +\infty$, then the minimum of the Unbalanced Optimal Transport problem 3.1 is attained at a reduced coupling.

Proof. We can assume that $\boldsymbol{\alpha}$ is nontrivial. Let us first check that $\boldsymbol{\alpha}_b$ preserves homogeneous marginals. For every $\zeta_i \in \mathrm{C}_b(\mathsf{X}_i)$, we have

$$\int (\zeta_i \circ \mathsf{x}_i)\mathsf{r}_i \,\mathrm{d}\boldsymbol{\alpha}_b = \int \zeta_i(x_i)\varrho_i(x_1,x_2)\,\mathrm{d}\boldsymbol{\gamma}(x_1,x_2) = \int \zeta_i(x_i)\Big(\int r_i \,\mathrm{d}\alpha_{x_1,x_2}(r_1,r_2)\Big)\,\mathrm{d}\boldsymbol{\gamma}(x_1,x_2)$$
$$= \int (\zeta_i \circ \mathsf{x}_i)\mathsf{r}_i \,\mathrm{d}\boldsymbol{\alpha} = \int \zeta_i \,\mathrm{d}\mu_i.$$

By radial convexity, using Jensen's inequality, we have

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} = \int_{\mathsf{X}_{1,o}\times\mathsf{X}_{2,o}} \Big(\int_{\mathbb{R}^2_+} \mathsf{H}\,\mathrm{d}\alpha_{x_1,x_2}(r_1,r_2)\Big)\,\mathrm{d}\gamma(x_1,x_2)$$
$$\geq \int_{\mathsf{X}_{1,o}\times\mathsf{X}_{2,o}} \mathsf{H}([x_1,\varrho_1(x_1,x_2)],[x_2,\varrho_2(x_1,x_2)])\,\mathrm{d}\gamma(x_1,x_2) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_b. \qquad \square$$

We can now show that it is possible to improve the previous representation by decomposing the measures μ_i in two parts, corresponding to complete or partial distruction/creation of mass. We introduce the partition $\mathfrak{C}', \mathfrak{C}_1'', \mathfrak{C}_2''$ of $\mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1,\mathsf{X}_2]$, where

$$\mathfrak{C}' := (\mathfrak{C}[\mathsf{X}_1]\setminus\{\mathfrak{o}_1\})\times(\mathfrak{C}[\mathsf{X}_2]\setminus\{\mathfrak{o}_2\}),\quad \mathfrak{C}_1'' := (\mathfrak{C}[\mathsf{X}_1]\setminus\{\mathfrak{o}_1\})\times\{\mathfrak{o}_2\},\quad \mathfrak{C}_2'' := \{\mathfrak{o}_1\}\times(\mathfrak{C}[\mathsf{X}_2]\setminus\{\mathfrak{o}_2\}).$$

Every homogeneous plan $\boldsymbol{\alpha} \in \mathcal{M}^1_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2])$ not charging $\mathfrak{o} = (\mathfrak{o}_1,\mathfrak{o}_2)$ can be decomposed into the corresponding sum $\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \boldsymbol{\alpha}_1'' + \boldsymbol{\alpha}_2''$, where $\boldsymbol{\alpha}' = \boldsymbol{\alpha}\llcorner\mathfrak{C}'$ and $\boldsymbol{\alpha}_i'' := \boldsymbol{\alpha}\llcorner\mathfrak{C}_i''$. Correspondingly we can write $\boldsymbol{\gamma} = \boldsymbol{\gamma}' + \boldsymbol{\gamma}_1'' + \boldsymbol{\gamma}_2''$

$$\boldsymbol{\gamma}' := \mathsf{x}_\sharp\boldsymbol{\alpha}' \in \mathcal{M}_+(\mathsf{X}_1\times\mathsf{X}_2),\quad \boldsymbol{\gamma}_1'' = \mathsf{x}_\sharp\boldsymbol{\alpha}_1'' \in \mathcal{M}_+(\mathsf{X}_1\times\{o_2\}),\quad \boldsymbol{\gamma}_2'' = \mathsf{x}_\sharp\boldsymbol{\alpha}_2'' \in \mathcal{M}_+(\{o_1\}\times\mathsf{X}_2). \tag{3.13}$$

Theorem 3.9 (Distinguished optimal couplings)**.** *Let us suppose that H satisfies (3.2) and it is radially convex and let $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ non trivial such that $\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_1) < +\infty$.*

There exist Borel partitions $\{S_i', S_i''\}$ of X_i and an optimal homogeneous and reduced coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ not charging $\mathfrak{o}$ such that setting

$$\mu_i' := \mu_i \llcorner S_i',\quad \mu_i(S_i') > 0,\quad \mu_i'' := \mu_i\llcorner S_i'', \tag{3.14}$$

and keeping the above notation for $\varrho_i, \boldsymbol{\alpha}', \boldsymbol{\alpha}_i'', \boldsymbol{\gamma}, \boldsymbol{\gamma}', \boldsymbol{\gamma}_i''$ we have

$$\boldsymbol{\alpha}_1'' = \mathfrak{p}_\sharp(\mu_1''\otimes\delta_1)\otimes\delta_{\mathfrak{o}_2},\qquad \boldsymbol{\alpha}_2'' = \delta_{\mathfrak{o}_1}\times\mathfrak{p}_\sharp(\mu_2''\otimes\delta_1), \tag{3.15}$$

$$\boldsymbol{\gamma}_1'' = \mu_1''\otimes\delta_{o_2},\qquad \boldsymbol{\gamma}_2'' = \delta_{o_1}\otimes\mu_2'', \tag{3.16}$$

$$\mathfrak{h}_i^1(\boldsymbol{\alpha}') = \mu_i',\quad \boldsymbol{\gamma}' = \boldsymbol{\gamma}\llcorner(S_1'\times S_2'), \tag{3.17}$$

$$\boldsymbol{\alpha} = \mathsf{T}[\boldsymbol{\varrho}]_\sharp\boldsymbol{\gamma},\quad \boldsymbol{\alpha}' = \mathsf{T}[\boldsymbol{\varrho}]_\sharp\boldsymbol{\gamma}',\quad \mathsf{T}[\boldsymbol{\varrho}](x_1,x_2) := \big([x_1,\varrho_1(x_1,x_2)],[x_2,\varrho_2(x_1,x_2)]\big) \tag{3.18}$$

and

$$\varrho_i > 0 \text{ on } S_1'\times S_2',\quad \varrho_1 = 0 \text{ on } S_1''\times\mathsf{X}_2,\quad \varrho_2 = 0 \text{ on } \mathsf{X}_1\times S_2'', \tag{3.19}$$

$$\varrho_1(x_1,o_2) = 1 \text{ on } S_1'',\quad \varrho_2(o_1,x_2) = 1 \text{ on } S_2''\quad \boldsymbol{\gamma}\text{-a.e.}. \tag{3.20}$$

Proof. Let $\tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}' + \tilde{\boldsymbol{\alpha}}_1'' + \tilde{\boldsymbol{\alpha}}_2''$ be an optimal reduced coupling in $\mathfrak{H}^1(\mu_1,\mu_2)$ not charging $\mathfrak{o}$.

In the degenerate case when $\tilde{\boldsymbol{\alpha}}'$ vanishes, then setting $\mu_i = m_i\nu_i$, $\nu_i \in \mathcal{P}(\mathsf{X}_i)$, $m_i = \mu_i(\mathsf{X}_i) > 0$, we have

$$\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) = \int_{\mathsf{X}_1} \mathsf{H}([x_1,m_1],\mathfrak{o}_2)\,\mathrm{d}\nu_1 + \int_{\mathsf{X}_2} \mathsf{H}(\mathfrak{o}_1,[x_2,m_2])\,\mathrm{d}\nu_2 \geq \int_{\mathsf{X}_1\times\mathsf{X}_2} \mathsf{H}([x_1,m_1],[x_2,m_2])\,\mathrm{d}(\nu_1\otimes\nu_2)$$

so that the reduced coupling $\boldsymbol{\alpha} = ([x_1,m_1],[x_2,m_2])_\sharp(\nu_1\otimes\nu_2)$ satisfies all the assumptions with $S_i' = \mathsf{X}_i$.

We can therefore suppose that $\boldsymbol{\alpha}'$ is nontrivial and we set

$$\tilde{\mu}_i' := \mathfrak{h}_i^1(\tilde{\boldsymbol{\alpha}}'),\quad \tilde{\mu}_i'' := \mathfrak{h}_i^1(\tilde{\boldsymbol{\alpha}}_i''),\quad \tilde{\boldsymbol{\gamma}} := \mathsf{x}_\sharp\tilde{\boldsymbol{\alpha}},\quad \tilde{\boldsymbol{\gamma}}' := \mathsf{x}_\sharp\tilde{\boldsymbol{\alpha}}' := \tilde{\boldsymbol{\gamma}}\llcorner(\mathsf{X}_1\times\mathsf{X}_2),$$

and we can write $\tilde{\boldsymbol{\alpha}} = \mathsf{T}[\tilde{\boldsymbol{\varrho}}]_\sharp \tilde{\boldsymbol{\gamma}}$ as in (3.11).

The optimal cost $\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2)$ can be decomposed as

$$\int \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}} = \int_{\mathsf{X}_1 \times \mathsf{X}_2} \mathsf{H}([x_1, \tilde{\varrho}_1], [x_2, \tilde{\varrho}_2])\, \mathrm{d}\tilde{\boldsymbol{\gamma}}' + \int_{\mathsf{X}_1} \mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}\mu_1'' + \int_{\mathsf{X}_2} \mathsf{H}(\mathfrak{o}_2, [x_2, 1])\, \mathrm{d}\mu_2''. \tag{3.21}$$

Since $\mathfrak{h}_1^1(\tilde{\boldsymbol{\alpha}}_2'') = \boldsymbol{0}_{\mathsf{X}_1}$ and $\mathfrak{h}_2^1(\tilde{\boldsymbol{\alpha}}_1'') = \boldsymbol{0}_{\mathsf{X}_2}$ we have $\mu_i = \tilde{\mu}_i' + \tilde{\mu}_i''$: we call θ_i', θ_i'' Borel representatives of the Lebesgue densities of μ_i', μ_i'' w.r.t. μ_i and we call

$$S_i'' := \Big\{ x_i \in \mathsf{X}_i : \theta_i''(x_i) = 1 \Big\}, \quad S_i' := \mathsf{X}_i \setminus S_i''.$$

Since $\theta_i' + \theta_i'' = 1$ μ_i-a.e. in X_i, we clearly have $\theta_i' = 0$ μ_i-a.e. in S_i'' and $\theta_i' > 0$ μ_i-a.e. in S_i'. We also have $\mu_1'(S_1'') = \mathfrak{h}_1^1(\tilde{\boldsymbol{\alpha}}')(S_1'') = 0$ and since $\mathfrak{h}_1^1(\tilde{\boldsymbol{\alpha}}') = (\mathsf{x}_1)_\sharp(\tilde{\varrho}_1 \tilde{\boldsymbol{\gamma}}')$ we deduce that $\tilde{\varrho}_1 = 0$ $\tilde{\boldsymbol{\gamma}}'$-a.e. in $S_1'' \times \mathsf{X}_2$. With a similar argument, we deduce that $\tilde{\varrho}_2 = 0$ $\tilde{\boldsymbol{\gamma}}'$-a.e. in $\mathsf{X}_1 \times S_2''$.

We define μ_i', μ_i'' according to (3.14). Notice that $\mu_i' = (\theta_i')^{-1}\tilde{\mu}_i' \ll \tilde{\mu}_i'$, whereas $\mu_i'' = \tilde{\mu}_i'' \llcorner S_i''$. The quotients $q_i(x_i) := \theta_i''/\theta_i'$ are well defined μ_i-a.e. in S_i' and we have

$$\mu_i' = (1 + q_i)\tilde{\mu}_i', \quad \tilde{\mu}_i'' = \mu_i'' + \theta_i''\mu_i' = \mu_i'' + q_i\tilde{\mu}_i'.$$

We set

$$\varrho_1 := \begin{cases} \tilde{\varrho}_1(1+q_1) & \text{in } S_1' \times \mathsf{X}_2, \\ 0 & \text{in } S_1'' \times \mathsf{X}_2, \end{cases} \qquad \varrho_2 := \begin{cases} \tilde{\varrho}_2(1+q_2) & \text{in } \mathsf{X}_1 \times S_2', \\ 0 & \text{in } \mathsf{X}_1 \times S_2'', \end{cases}$$

and we define $\boldsymbol{\alpha}_i''$ as in (3.15). Finally

$$\boldsymbol{\gamma}' := \tilde{\boldsymbol{\gamma}}' \llcorner (S_1' \times S_2'), \quad \boldsymbol{\alpha}' := \mathsf{T}[\boldsymbol{\varrho}]_\sharp \tilde{\boldsymbol{\gamma}}' = \mathsf{T}[\boldsymbol{\varrho}]_\sharp \boldsymbol{\gamma}', \quad \boldsymbol{\alpha} := \boldsymbol{\alpha}' + \boldsymbol{\alpha}_1'' + \boldsymbol{\alpha}_2''.$$

Let us check (3.17); since $\boldsymbol{\gamma} = \mathbf{x}_\sharp \boldsymbol{\alpha}$, the equality $\boldsymbol{\gamma}' = \boldsymbol{\gamma} \llcorner (S_1 \times S_2')$ is immediate, so that we have to prove that $\mathfrak{h}_i^1(\boldsymbol{\alpha}') = \mu_i'$. We just consider the case $i = 1$ since the calculations in the case $i = 2$ are completely analogous. For every test function $\zeta \in \mathrm{C}_b(\mathsf{X}_1)$ we get

$$\begin{aligned} \int (\zeta \circ \mathsf{x}_1)\mathsf{r}_1\, \mathrm{d}\boldsymbol{\alpha}' &= \int_{\mathsf{X}_1 \times \mathsf{X}_2} \zeta(x_1)\varrho_1(x_1, x_2)\, \mathrm{d}\tilde{\boldsymbol{\gamma}}'(x_1, x_2) \\ &= \int_{S_1' \times \mathsf{X}_2} \zeta(x_1)\tilde{\varrho}_1(x_1, x_2)\Big(1 + q_1(x_1)\Big)\, \mathrm{d}\tilde{\boldsymbol{\gamma}}'(x_1, x_2) \\ &= \int_{S_1'} \zeta(x_1)(1 + q_1(x_1))\, \mathrm{d}\tilde{\mu}_1(x_1) = \int_{\mathsf{X}_1} \zeta\, \mathrm{d}\mu_1'. \end{aligned}$$

Let us now compute the H-cost of $\boldsymbol{\alpha}$:

$$\begin{aligned} \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha}' &= \int \mathsf{H}([x_1, \tilde{\varrho}_1 + q_1\tilde{\varrho}_1], [x_2, \tilde{\varrho}_2 + q_2\tilde{\varrho}_2])\, \mathrm{d}\tilde{\boldsymbol{\gamma}}' \\ &\le \int \mathsf{H}([x_1, \tilde{\varrho}_1], [x_2, \tilde{\varrho}_2])\, \mathrm{d}\tilde{\boldsymbol{\gamma}}' + \int q_1\tilde{\varrho}_1\mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}\tilde{\boldsymbol{\gamma}}' + \int q_2\tilde{\varrho}_2\mathsf{H}(\mathfrak{o}_1, [x_2, 1])\, \mathrm{d}\tilde{\boldsymbol{\gamma}}' \\ &= \int \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}}' + \int q_1\mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}\tilde{\mu}_1' + \int q_2\mathsf{H}(\mathfrak{o}_1, [x_2, 1])\, \mathrm{d}\tilde{\mu}_2', \\ \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha} &= \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha}' + \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha}_1'' + \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha}_2'' \\ &= \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha}' + \int \mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}\mu_1'' + \int \mathsf{H}(\mathfrak{o}_1, [x_2, 1])\, \mathrm{d}\mu_2'' \\ &\le \int \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}}' + \int \mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}(q_1\tilde{\mu}_1' + \mu_1'') \int \mathsf{H}(\mathfrak{o}_1, [x_2, 1])\, \mathrm{d}(q_2\tilde{\mu}_2' + \mu_2'') \\ &= \int \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}}' + \int \mathsf{H}([x_1, 1], \mathfrak{o}_2)\, \mathrm{d}\tilde{\mu}_1'' + \int \mathsf{H}(\mathfrak{o}_1, [x_2, 1])\, \mathrm{d}\tilde{\mu}_2'' = \int \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}} = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2). \end{aligned}$$

□

As an application of the previous properties, we compare the formulation 3.1 of Unbalanced Optimal Transport via homogeneous couplings with the formulation based on the notion of semi-couplings introduced by [Chi+18b] in a compact Euclidean setting. We obtain a considerable extension of the semi-coupling approach to a general, possibly non-compact, setting.

Given $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$, $i = 1, 2$, we can consider the set

$$\hat\Gamma(\mu_1, \mu_2) := \Big\{(\gamma_1, \gamma_2) \in (\mathcal{M}_+(\mathsf{X}_1 \times \mathsf{X}_2))^2 : \pi^i_\sharp(\gamma_i) = \mu_i,\ i = 1, 2\Big\}. \tag{3.22}$$

Every radially 1-homogeneous cost function H as in (3.2) induces a functional on $\hat\Gamma(\mu_1, \mu_2)$ given by

$$J_{\mathsf{H}}(\gamma_1, \gamma_2) := \int_{\mathsf{X}_1\times\mathsf{X}_2} \mathsf{H}\Big(\Big[x_1, \frac{\mathrm{d}\gamma_1}{\mathrm{d}\gamma}\Big], \Big[x_2, \frac{\mathrm{d}\gamma_2}{\mathrm{d}\gamma}\Big]\Big)\,\mathrm{d}\gamma, \quad \gamma \in \mathcal{M}_+(\mathsf{X}_1 \times \mathsf{X}_2),\ \gamma_i \ll \gamma. \tag{3.23}$$

Since H is radially 1-homogeneous, (3.23) does not depend on the choice of the dominating measure γ and one can consider the problem of minimizing J_{H} in $\hat\Gamma(\mu_1, \mu_2)$:

$$\widehat{\mathscr{U}_{\mathsf{H}}}(\mu_1, \mu_2) := \inf\Big\{J_{\mathsf{H}}(\gamma_1, \gamma_2) : (\gamma_1, \gamma_2) \in \hat\Gamma(\mu_1, \mu_2)\Big\}. \tag{3.24}$$

When $\widehat{\mathscr{U}_{\mathsf{H}}}(\mu_1, \mu_2) < \infty$, H satisfies (3.2) and it is also radially convex, and $\mathsf{X}_1, \mathsf{X}_2$ are compact, it is not difficult to prove the existence of an optimal pair $(\gamma_1, \gamma_2) \in \hat\Gamma(\mu_1, \mu_2)$ attaining the infimum in (3.24) (see [Chi+18b, Prop. 3.4]). The direct approach when X_i are not compact is less clear, but existence of an optimal pair can be obtained as a consequence of the argument we are going to detail below in Theorem 3.11.

Remark 3.10. The formulation (3.24) also allows for more general (l.s.c., radially 1-homogeneous and convex) functions $\hat{\mathsf{H}}$ defined in $(\mathsf{X}_1 \times [0, +\infty)) \times (\mathsf{X}_2 \times [0, +\infty))$ which are not compatible with the equivalence relation (2.7) inducing the topological cone $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$. This definition, however, may not satisfy the invariance property stated in Remark 3.2: consider, e.g., the case when $\mathsf{X}_1 = \mathsf{X}_2 = [0, a]$, $a > 0$, $\mu_1 = \delta_0$, $\mu_2 = \mathbf{0}$, and $\hat{\mathsf{H}}((x_1, r_1), (x_2, 0)) = r_1 \mathrm{e}^{-|x_1 - x_2|}$. We clearly have $\widehat{\mathscr{U}_{\hat{\mathsf{H}}}}(\mu_1, \mu_2) = \mathrm{e}^{-a}$. In order to have an intrinsic formulation, we will assume that H is compatible with the cone structure.

Theorem 3.11. *If H satisfies (3.2) then for every $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ non trivial we have*

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) \le \widehat{\mathscr{U}_{\mathsf{H}}}(\mu_1, \mu_2). \tag{3.25}$$

If, in addition to (3.2), H is also radially convex, then

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) = \widehat{\mathscr{U}_{\mathsf{H}}}(\mu_1, \mu_2). \tag{3.26}$$

and the infimum in (3.24) is attained.

Proof. We first observe that every triple $(\gamma, \gamma_1, \gamma_2) \in \mathcal{M}_+(\mathsf{X}_1 \times \mathsf{X}_2) \times \hat\Gamma(\mu_1, \mu_2)$ with $\gamma_i \ll \gamma$ induces a homogeneous coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ via the formula

$$\boldsymbol{\alpha} := G_\sharp \gamma, \quad G(x_1, x_2) := \Big(\big[x_1, \varrho_1(x_1, x_2)\big], \big[x_2, \varrho_2(x_1, x_2)\big]\Big), \quad \varrho_i := \frac{\mathrm{d}\gamma_i}{\mathrm{d}\gamma}, \tag{3.27}$$

such that

$$\int_{\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} = J_{\mathsf{H}}(\gamma_1, \gamma_2). \tag{3.28}$$

Therefore we immediately get (3.25).

Let us now assume that H is also radially convex and let $\boldsymbol{\alpha}$ be a distinguished optimal coupling as in Theorem 3.9. Keeping the same notation of that Theorem let us set $\mu_i(S_i') = m_i > 0$ and let us set $\nu_i := m_i^{-1}\mu_i'$. We define $\gamma := \boldsymbol{\gamma}' + \mu_1'' \otimes \nu_2 + \nu_1 \otimes \mu_2''$, $\gamma_1 := \varrho_1 \gamma' + \mu_1'' \otimes \nu_2$, $\gamma_2 := \varrho_2 \gamma' + \nu_1 \otimes \mu_2''$, obtaining $\pi^i_\sharp \gamma_i = \mu_i' + \mu_i'' = \mu_i$ so that $(\gamma_1, \gamma_2) \in \hat{\Gamma}(\mu_1, \mu_2)$. Moreover

$$\begin{aligned} J_{\mathsf{H}}(\gamma_1, \gamma_2) &= \int \mathsf{H}([x_1, \varrho_1], [x_2, \varrho_2])\, \mathrm{d}\boldsymbol{\gamma}' \\ &\quad + \int \mathsf{H}([x_1, 1], [x_2, 0])\, \mathrm{d}(\mu_1'' \otimes \nu_2) + \int \mathsf{H}([x_1, 0], [x_2, 1])\, \mathrm{d}(\nu_1 \otimes \mu_2'') \\ &= \int \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha} = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2). \end{aligned}$$

□

3.3. **The dual problem.** In the next results we study the dual definition of the Unbalanced Optimal Transport cost $\mathscr{U}_{\mathsf{H}}$, for which we need the following definitions.

Definition 3.12. Let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ be a function; we define the set of continuous functions

$$\Phi_{\mathsf{H}} := \left\{ (\varphi_1, \varphi_2) \in \mathrm{C}_b(\mathsf{X}_1) \times \mathrm{C}_b(\mathsf{X}_2) \ : \ \begin{array}{l} \varphi_1(x_1) r_1 + \varphi_2(x_2) r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \\ \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ (r_1, r_2) \in \mathbb{R}^2_+ \end{array} \right\} \tag{3.29}$$

and, for every $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$, the functional $\mathcal{D}(\cdot; \mu_1, \mu_2) : \mathrm{C}_b(\mathsf{X}_1) \times \mathrm{C}_b(\mathsf{X}_2) \to \mathbb{R}$ given by

$$\mathcal{D}(\varphi_1, \varphi_2; \mu_1, \mu_2) := \int_{\mathsf{X}_1} \varphi_1\, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\, \mathrm{d}\mu_2, \quad (\varphi_1, \varphi_2) \in \mathrm{C}_b(\mathsf{X}_1) \times \mathrm{C}_b(\mathsf{X}_2).$$

Before stating the main duality result, let us briefly recall the Fenchel-Moreau Theorem in the framework of a pair of vector spaces E, F placed in duality by a nondegenerate bilinear map $\langle \cdot, \cdot, \rangle$, see e.g. [ET87]. We endow E with the weak topology $\sigma(E, F)$, the coarsest topology for which all the functions $e \mapsto \langle e, f \rangle$, $f \in F$, are continuous.

Definition 3.13. Let $\mathscr{F} : E \to (-\infty, +\infty]$ be not identically $+\infty$ and satisfying

$$\mathscr{F}(e) \ge \langle e, f \rangle - c \quad \text{for some } f \in F,\ c \in \mathbb{R} \text{ and every } e \in E. \tag{3.30}$$

The polar (or conjugate) function of $\mathscr{F}$ is the function $\mathscr{F}^* : F \to (-\infty, +\infty]$ defined by

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

Theorem 3.14 (Fenchel-Moreau). *Let E and F be vector spaces placed in duality by a nondegenerate bilinear map $\langle \cdot, \cdot, \rangle$ and let $\mathscr{F} : E \to (-\infty, +\infty]$ be satisfying (3.30) and not identically $+\infty$. Then the lower semicontinuous (w.r.t. the topology $\sigma(E, F)$) and convex envelope of $\mathscr{F}$ is given by the dual formula*

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

In particular,

$$\textit{if } \mathscr{F} \textit{ is convex and lower semicontinuous then} \quad \mathscr{F} = \mathscr{F}^{**}.$$

The following is the main duality result and it gives also an independent proof of Theorem 3.5.

Theorem 3.15 (Duality). *Let H be as in (3.2) and let $\mathscr{S}_{\mathsf{H}}$ and $\mathscr{U}_{\mathsf{H}}$ be as in Definition 3.1. Then*

$$\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) = \overline{\mathrm{co}}\,(\mathscr{S}_{\mathsf{H}})\,(\mu_1, \mu_2) = \sup \left\{ \mathcal{D}(\varphi_1, \varphi_2; \mu_1, \mu_2) : (\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}} \right\} \tag{3.31}$$

for every $(\mu_1, \mu_2) \in \mathcal{M}_+(\mathsf{X}_1) \times \mathcal{M}_+(\mathsf{X}_2)$.

Proof. Set $E := \mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2)$ and $F := \mathrm{C}_b(\mathsf{X}_1) \times \mathrm{C}_b(\mathsf{X}_2)$ with the bilinear form

$$\langle \cdot, \cdot \rangle : E \times F \to \mathbb{R}, \qquad ((\mu_1, \mu_2), (\varphi_1, \varphi_2)) \mapsto \mathcal{D}(\varphi_1, \varphi_2; \mu_1, \mu_2).$$

This is a well defined nondegenerate bilinear form. We endow then E with the topology $\sigma(E, F)$ which coincides exactly with the product weak topology.
Consider then the function $\mathscr{S}_\mathsf{H} : E \mapsto (-\infty, +\infty]$ defined as in Definition 3.1. Then we have

$$\begin{aligned}
\mathscr{S}^*_\mathsf{H}(\varphi_1, \varphi_2) &= \sup_{(\mu_1,\mu_2)\in E} \{\langle (\mu_1, \mu_2), (\varphi_1, \varphi_2)\rangle - \mathscr{S}_\mathsf{H}(\mu_1, \mu_2)\} \\
&= \sup_{(\mu_1,\mu_2)\in \Delta_+(\mathsf{X}_1)\times\Delta_+(\mathsf{X}_2)} \{\langle (\mu_1, \mu_2), (\varphi_1, \varphi_2)\rangle - \mathscr{S}_\mathsf{H}(\mu_1, \mu_2)\} \\
&= \sup_{x_1,r_1,x_2,r_2} \{\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2 - \mathsf{H}([x_1, r_1], [x_2, r_2])\} \\
&= \begin{cases} 0 & \text{if } (\varphi_1, \varphi_2) \in \Phi_\mathsf{H} \\ +\infty & \text{elsewhere} \end{cases}.
\end{aligned}$$

Hence by Theorem 3.14 we have

$$\overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H})\,(\mu_1, \mu_2) = \mathscr{S}^{**}_\mathsf{H}(\mu_1, \mu_2) = \sup\left\{\mathcal{D}(\varphi_1, \varphi_2; \mu_1, \mu_2) \mid (\varphi_1, \varphi_2) \in \Phi_\mathsf{H}\right\}. \tag{3.32}$$

By Theorem 3.4 we know that $\mathscr{U}_\mathsf{H}$ is convex and lower semicontinuous and stays below $\mathscr{S}_\mathsf{H}$ so that we clearly have $\mathscr{U}_\mathsf{H} \le \overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H})$. The other inequality is immediate: take any $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ and any $(\varphi_1, \varphi_2) \in \Phi_\mathsf{H}$; then

$$\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \,\mathrm{d}\mu_2 = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} ((\varphi_1 \circ \mathsf{x}_1)\mathsf{r}_1 + (\varphi_2 \circ \mathsf{x}_2)\mathsf{r}_2) \,\mathrm{d}\boldsymbol{\alpha} \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H} \,\mathrm{d}\boldsymbol{\alpha}.$$

Passing to the supremum in Φ_H and to the infimum in $\mathfrak{H}^1(\mu_1, \mu_2)$ and using (3.32), we conclude that $\mathscr{U}_\mathsf{H} \ge \overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H})$. □

Remark 3.16. If H has the form

$$\mathsf{H}([x_1, r_1], [x_2, r_2]) := \begin{cases} r_1 \mathsf{c}(x_1, x_2) & \text{if } r_1 = r_2 \ge 0 \\ +\infty & \text{elsewhere} \end{cases},$$

for some $\mathsf{c} : \mathsf{X}_1 \times \mathsf{X}_2 \to [0, +\infty]$ lower semicontinuous function, we have that $\mathscr{U}_\mathsf{H}|_{\mathcal{P}(\mathsf{X}_1)\times\mathcal{P}(\mathsf{X}_2)}$ is exactly the classical Optimal Transport cost induced by c. This was indeed exploited in [SS22].

In Definition 2.10 we have introduced the closed and convex envelope of H which is obtained first convexifying H "slice by slice" and then closing its graph globally in $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$. One could also compute first the closed convex envelope "slice by slice" and then glue together the resulting functions (see the definition of $\widehat{\mathrm{co}}\,(\mathsf{H})$ below). In the following result we show, making use of Theorem 3.15, that the two procedures give raise to same object.

We also show that, even if H is not convex, taking the closed and convex envelope of the singular cost $\mathscr{S}_\mathsf{H}$ is sufficient to also recover the Unbalanced Optimal Transport cost induced by the closed and convex envelope of H.

Corollary 3.17. *Let H be as in (3.2) and let $\overline{\mathrm{co}}\,(\mathsf{H})$ be as in Definition 2.10. Let us define $\widehat{\mathrm{co}}\,(\mathsf{H}) : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ as*

$$\widehat{\mathrm{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) := \overline{\mathrm{co}}\,(\mathsf{H}_{x_1,x_2})\,(r_1, r_2) \quad \textit{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ (r_1, r_2) \in \mathbb{R}^2_+.$$

Then

$$\overline{\mathrm{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) = \widehat{\mathrm{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) = \sup\left\{\varphi(x_1)r_1 + \varphi_2(x_2)r_2 : (\varphi_1, \varphi_2) \in \Phi_\mathsf{H}\right\}$$

for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ and every $(r_1, r_2) \in \mathbb{R}^2_+$. Moreover

$$\overline{\mathrm{co}}\,(\mathscr{S}_\mathsf{H}) = \overline{\mathrm{co}}\left(\mathscr{S}_{\overline{\mathrm{co}}(\mathsf{H})}\right) = \mathscr{U}_\mathsf{H} = \mathscr{U}_{\overline{\mathrm{co}}(\mathsf{H})} \ \textit{in } \mathcal{M}(\mathsf{X}_1) \times \mathcal{M}(\mathsf{X}_2). \tag{3.33}$$

Proof. We denote by $\mathsf{U}_\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ the restriction of $\mathscr{U}_\mathsf{H}$ to $\Delta_+(\mathsf{X}_1) \times \Delta_+(\mathsf{X}_2) \cong \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ (see Lemma 2.4).
It is clear that $\overline{\text{co}}\,(\mathsf{H}) \le \widehat{\text{co}}\,(\mathsf{H}) \le \mathsf{H}$ so that

$$\mathscr{U}_{\overline{\text{co}}(\mathsf{H})} \le \mathscr{U}_{\widehat{\text{co}}(\mathsf{H})} \le \mathscr{U}_\mathsf{H}$$

and

$$\overline{\text{co}}\,(\mathsf{H}) = \mathsf{U}_{\overline{\text{co}}(\mathsf{H})} \le \widehat{\text{co}}\,(\mathsf{H}) = \mathsf{U}_{\widehat{\text{co}}(\mathsf{H})} \le \mathsf{U}_\mathsf{H} \le \mathsf{H},$$

where we used Theorem 3.4 and the radial convexity of $\overline{\text{co}}\,(\mathsf{H})$ and $\widehat{\text{co}}\,(\mathsf{H})$. Moreover, since U_H is, lower semicontinuous, convex and stays below H, we have that $\mathsf{U}_\mathsf{H} \le \overline{\text{co}}\,(\mathsf{H})$. This gives that

$$\overline{\text{co}}\,(\mathsf{H}) = \mathsf{U}_{\overline{\text{co}}(\mathsf{H})} = \widehat{\text{co}}\,(\mathsf{H}) = \mathsf{U}_{\widehat{\text{co}}(\mathsf{H})} = \mathsf{U}_\mathsf{H}$$

and in particular, using Theorem 3.15, that

$$\overline{\text{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) = \widehat{\text{co}}\,(\mathsf{H})\,([x_1, r_1], [x_2, r_2]) = \sup\left\{\varphi(x_1) r_1 + \varphi_2(x_2) r_2 : (\varphi_1, \varphi_2) \in \Phi_\mathsf{H}\right\}.$$

The fact that $\overline{\text{co}}\,(\mathsf{H}) = \mathsf{U}_\mathsf{H}$ gives that $\mathscr{U}_\mathsf{H} \le \mathscr{S}_{\overline{\text{co}}(\mathsf{H})}$ so that $\mathscr{U}_\mathsf{H} \le \overline{\text{co}}\left(\mathscr{S}_{\overline{\text{co}}(\mathsf{H})}\right)$. However, by Theorem 3.4, we know that $\mathscr{U}_\mathsf{H} = \overline{\text{co}}\,(\mathscr{S}_\mathsf{H})$ so that $\overline{\text{co}}\,(\mathscr{S}_\mathsf{H}) \le \overline{\text{co}}\left(\mathscr{S}_{\overline{\text{co}}(\mathsf{H})}\right)$. Since, obviously, the other inequality holds, we have $\overline{\text{co}}\,(\mathscr{S}_\mathsf{H}) = \overline{\text{co}}\left(\mathscr{S}_{\overline{\text{co}}(\mathsf{H})}\right)$. Applying again Theorem 3.4 to $\overline{\text{co}}\,(\mathsf{H})$ we conclude that

$$\overline{\text{co}}\,(\mathscr{S}_\mathsf{H}) = \overline{\text{co}}\left(\mathscr{S}_{\overline{\text{co}}(\mathsf{H})}\right) = \mathscr{U}_\mathsf{H} = \mathscr{U}_{\overline{\text{co}}(\mathsf{H})}.$$

□

Remark 3.18. As an interesting application of the previous Corollary, we obtain a clarifying justification of two equivalent characterizations of the Hellinger-Kantorovich metric (see (7.23) and (7.54) of [LMS18]) in a (complete, separable) metric space (X, d), namely the formula

$$\mathsf{HK}^2(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha}_2 \in \mathfrak{H}^2(\mu_1, \mu_2)} \int \mathsf{d}_\mathfrak{C}^2 \,\mathrm{d}\boldsymbol{\alpha}_2 = \min_{\boldsymbol{\beta}_2 \in \mathfrak{H}^2(\mu_1, \mu_2)} \int \mathsf{d}_{\pi/2, \mathfrak{C}}^2 \,\mathrm{d}\boldsymbol{\beta}_2, \tag{3.34}$$

where $\mathsf{d}_\mathfrak{C}$ is the canonical cone metric introduced in (2.9) and $\mathsf{d}_{\pi/2,\mathfrak{C}}$ is the cone metric obtained by trucating the argument of the cos function at $\pi/2$:

$$\mathsf{d}_{\pi/2,\mathfrak{C}}([x, r], [y, s]) := \left(r^2 + s^2 - 2rs\cos(\mathsf{d}(x, y) \wedge \pi/2)\right)^{\frac{1}{2}}, \quad [x, r], [y, s] \in \mathfrak{C}[\mathsf{X}]. \tag{3.35}$$

By Remark 2.13, the identity (3.34) is equivalent to

$$\mathsf{HK}^2(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)} \int \mathsf{d}_\mathfrak{C}^2 \circ \mathsf{T}_2 \,\mathrm{d}\boldsymbol{\alpha} = \min_{\boldsymbol{\beta} \in \mathfrak{H}^1(\mu_1, \mu_2)} \int \mathsf{d}_{\pi/2,\mathfrak{C}}^2 \circ \mathsf{T}_2 \,\mathrm{d}\boldsymbol{\beta}, \tag{3.36}$$

where T_2 is defined in (2.14); it is then not difficult to check that

$$\mathsf{d}_{\pi/2,\mathfrak{C}}^2 \circ \mathsf{T}_2 = \overline{\text{co}}\left(\mathsf{d}_\mathfrak{C}^2 \circ \mathsf{T}_2\right). \tag{3.37}$$

Therefore, the constant $\pi/2$ in (3.34) appears as a natural effect of the convexification of the cost function $\mathsf{d}_\mathfrak{C}^2 \circ \mathsf{T}_2$ and the identity (3.33).

3.4. **The Monge formulation.** We define the Monge formulation of the Unbalanced Optimal Transport problem.

Definition 3.19 (Transport-growth maps and Monge formulation)**.** Given $\mu_1 \in \mathcal{M}_+(\mathsf{X}_1)$ and Borel maps $(\mathsf{T}, \mathsf{g}) : \mathsf{X}_1 \to \mathsf{X}_2 \times [0, +\infty)$ with $\mathsf{g} \in L^1(\mathsf{X}_1, \mu_1)$, we denote by $(\mathsf{T}, \mathsf{g})_\star \mu_1$ the measure

$$(\mathsf{T}, \mathsf{g})_\star \mu_1 := \mathsf{T}_\sharp(\mathsf{g}\mu_1) \in \mathcal{M}_+(\mathsf{X}_2).$$

We say that (T, g) is a *transport-growth map* connecting μ_1 to μ_2. If $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ is a proper Borel function, we define the Monge formulation of the Unbalanced Optimal Transport problem as

$$\mathscr{M}_\mathsf{H}(\mu_1, \mu_2) := \inf\left\{\int_{\mathsf{X}_1} \mathsf{H}([x, 1], [\mathsf{T}(x), \mathsf{g}(x)]) \,\mathrm{d}\mu_1(x) \mid \mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1\right\}.$$

While in this section we only study the relation between the Kantorovich (i.e. the primal) and the Monge fofmulation of the Unbalanced Optimal Transport problem, in Theorem 5.13 we will provide sufficient conditions for the existence of an optimal transport-growth map realizing the infimum above. Clearly we have

$$\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) \le \mathscr{M}_{\mathsf{H}}(\mu_1,\mu_2), \quad \mu_i \in \mathcal{M}_+(\mathsf{X}_i)$$

as a consequence of the fact that any transport-growth map (T,g) induces an admissible $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ via the formula

$$\boldsymbol{\alpha} = ([\mathrm{id}_{\mathsf{X}_1},1],[\mathsf{T},\mathsf{g}])_\sharp \mu_1.$$

Recalling the definition (2.16) of the reduced cost function h, we also have

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} = \int_{\mathsf{X}_1} \mathsf{H}([x,1],[\mathsf{T}(x),\mathsf{g}(x)])\,\mathrm{d}\mu_1(x) = \int_{\mathsf{X}_1} \mathsf{h}(x,\mathsf{T}(x);\mathsf{g}(x))\,\mathrm{d}\mu_1(x).$$

We derive the main result of this subsection from the analogous one [Pra07] in the classical Optimal Transport theory. To do that we need the following definition.

Definition 3.20. Let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to [0.+\infty]$ be a proper and lower semicontinuous function. We define

$$\mathsf{OT}_{\mathsf{H}}(\alpha_1,\alpha_2) := \min\left\{\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma(\alpha_1,\alpha_2)\right\}, \quad \alpha_i \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_i]),$$

where $\Gamma(\alpha_1,\alpha_2)$ is the set of transport plans from α_1 to α_2 defined as

$$\Gamma(\alpha_1,\alpha_2) := \left\{\boldsymbol{\gamma} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]) : \pi^1_\sharp\boldsymbol{\gamma} = \alpha_1,\ \pi^2_\sharp\boldsymbol{\gamma} = \alpha_2\right\}.$$

Notice that $\Gamma(\alpha_1,\alpha_2)$ is not-empty if and only if $\alpha_1(\mathfrak{C}[\mathsf{X}_1]) = \alpha_2(\mathfrak{C}[\mathsf{X}_2])$. The following is an immediate consequence of [Pra07, Proof of Theorem B].

Theorem 3.21. *Let X_i be Polish spaces, let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to [0,+\infty]$ be continuous and let $\alpha_i \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_i])$ be such that α_1 is diffuse and $\alpha_1(\mathfrak{C}[\mathsf{X}_1]) > 0$. Then for every $\boldsymbol{\gamma} \in \Gamma(\alpha_1,\alpha_2)$ such that $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma} < +\infty$ and every $\varepsilon > 0$ there exists a Borel map $\mathsf{F} : \mathfrak{C}[\mathsf{X}_1] \to \mathfrak{C}[\mathsf{X}_2]$ such that*

$$\mathsf{F}_\sharp\alpha_1 = \alpha_2, \quad \int_{\mathfrak{C}[\mathsf{X}_1]} \mathsf{H}(\mathfrak{y}_1,\mathsf{F}(\mathfrak{y}_2))\,\mathrm{d}\alpha_1(\mathfrak{y}_1) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\gamma} + \varepsilon.$$

Choosing in particular an optimal $\boldsymbol{\gamma}$ realizing the minimum in the definition of OT_{H} one can prove the equivalence between the Monge and Kantorovich formulations of the Optimal Transport problem. The connection between balanced and unbalanced Optimal Transport is of course given by the fact that if $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ for measures $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$, then $\boldsymbol{\alpha} \in \Gamma(\alpha_1,\alpha_2)$, where $\alpha_i := \pi^i_\sharp\boldsymbol{\alpha}$. A stronger connection actually holds, as reported in the following result which is the analogue of [LMS18, Corollary 7.7, Corollary 7.13]; the proof of the second statement is identical and thus omitted, the proof of the first statement follows immediately by Lemma 2.12.

Proposition 3.22. *Let H be as in (3.2). Then for every $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ with $R = R(\mu_1,\mu_2)$ as in (2.13), it holds*

$$\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) = \min\left\{\mathsf{OT}_{\mathsf{H}}(\alpha_1,\alpha_2) : \alpha_i \in \mathcal{P}(\mathfrak{C}_R[\mathsf{X}_i]),\ \mathfrak{h}^1(\alpha_i) = \mu_i,\ i = 1,2\right\}.$$

In particular we have $\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) = \mathsf{OT}_{\mathsf{H}}(\pi^1_\sharp\boldsymbol{\alpha},\pi^2_\sharp\boldsymbol{\alpha})$ for every optimal $\boldsymbol{\alpha} \in \mathfrak{H}^1_o(\mu_1,\mu_2)$ (cf. (3.5)). If $\mathsf{X} = \mathsf{X}_1 = \mathsf{X}_2$ and $(\mu_i)_{i=1}^N \subset \mathcal{M}_+(\mathsf{X})$ with $N \ge 2$, then there exist $(\alpha_i)_{i=1}^N \subset \mathcal{P}(\mathfrak{C}[\mathsf{X}])$ such that

$$\mathfrak{h}^1(\alpha_i) = \mu_i, \quad \mathscr{U}_{\mathsf{H}}(\mu_{i-1},\mu_i) = \mathsf{OT}_{\mathsf{H}}(\alpha_{i-1},\alpha_i) \quad \textit{for every } i \in \{2,\dots,N\}.$$

We first show that the cost of any coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$, $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$, can be approximated by the cost of couplings with the same homogeneous marginals not charging the vertex of the first cone.

Proposition 3.23. *Let* $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ *be a proper, radially* 1*-homogeneous and continuous function, and let* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *be such that* $\mu_1(\mathsf{X}_1) > 0$*. Then for every* $\varepsilon > 0$ *and every* $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ *such that* $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} < +\infty$ *there exists* $\boldsymbol{\alpha}_\varepsilon \in \mathfrak{H}^1(\mu_1, \mu_2)$ *with* $\boldsymbol{\alpha}_\varepsilon(\{\mathfrak{y}_1 = \mathfrak{o}_1\}) = 0$ *such that*

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_\varepsilon \leq \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} + \varepsilon.$$

Proof. We define the measures

$$\boldsymbol{\alpha}^{tg} := \boldsymbol{\alpha} \llcorner \{\mathfrak{y}_1 \neq \mathfrak{o}_1\}, \quad \boldsymbol{\alpha}^{o} := \boldsymbol{\alpha} \llcorner \{\mathfrak{y}_1 = \mathfrak{o}_1\}, \quad \mu_2^{tg} := \mathfrak{h}_2^1(\boldsymbol{\alpha}^{tg}), \quad \mu_2^{o} := \mathfrak{h}_2^1(\boldsymbol{\alpha}^{o});$$

it is not restrictive to assume that $\mu_2^o(\mathsf{X}_2) > 0$ so that $\boldsymbol{\alpha}^o(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]) > 0$. For a bounded Borel map $\vartheta : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to (0, \infty)$ and a measure $\boldsymbol{\beta} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2])$ we set

$$\begin{aligned} \mathrm{prd}_{\vartheta,2}(\mathfrak{y}_1, \mathfrak{y}_2) &:= (\mathfrak{y}_1, \mathfrak{y}_2/\vartheta(\mathfrak{y}_1, \mathfrak{y}_2)), \quad \mathfrak{y}_i \in \mathfrak{C}[\mathsf{X}_i], \\ \mathrm{dil}_{\vartheta,2}(\boldsymbol{\beta}) &:= (\mathrm{prd}_{\vartheta,2})_\sharp(\vartheta\,\boldsymbol{\beta}). \end{aligned}$$

Notice that $\mathfrak{h}_2^1(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\beta})) = \mathfrak{h}_2^1(\boldsymbol{\beta})$ and

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\ \mathrm{d}\big(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\beta})\big) = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}(\vartheta(\mathfrak{y}_1, \mathfrak{y}_2)\mathfrak{y}_1, \mathfrak{y}_2)\ \mathrm{d}\boldsymbol{\beta}(\mathfrak{y}_1, \mathfrak{y}_2). \tag{3.38}$$

We fix $\varepsilon > 0$ and we set

$$\begin{aligned} \lambda'_\varepsilon(\mathfrak{y}_1, \mathfrak{y}_2) &:= \sup\Big\{\lambda \in [0, 1/2] : \mathsf{H}(r\mathfrak{y}_1, \mathfrak{y}_2) \leq \mathsf{H}(\mathfrak{o}_1, \mathfrak{y}_2) + \varepsilon_0 \text{ for every } r \in [0, \lambda]\Big\}, \\ \lambda''_\varepsilon(\mathfrak{y}_1, \mathfrak{y}_2) &:= \sup\Big\{\lambda \in [0, 1/2] : \mathsf{H}(r\mathfrak{y}_1, \mathfrak{y}_2) \leq \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) + \varepsilon_0 \text{ for every } r \in [1-\lambda, 1]\Big\}, \end{aligned}$$

where ε_0 is given by

$$\varepsilon_0 := \frac{1}{2} \frac{\varepsilon}{\mu_1(\mathsf{X}_1) + \boldsymbol{\alpha}^{tg}(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2])}.$$

Since H is continuous, it is not difficult to check that $\lambda'_\varepsilon, \lambda''_\varepsilon$ define strictly positive upper semicontinuous maps (thus Borel) from $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ to $(0, 1/2]$. Now we set $c := \mu_2^o(\mathsf{X}_2)/\mu_1(\mathsf{X}_1)$ and we consider any $\boldsymbol{\gamma} \in \Gamma(\mu_1, c^{-1}\mu_2^o)$. We "lift" $\boldsymbol{\gamma}$ to a plan $\boldsymbol{\beta} := \ell_\sharp\boldsymbol{\gamma} \in \mathfrak{H}^1(\mu_1, \mu_2^o)$ using the map $\ell(x_1, x_2) := ([x_1, 1], [x_2, c])$ and we consider the rescaled plan $\boldsymbol{\beta}'_\varepsilon$ and its first homogeneous marginal

$$\boldsymbol{\beta}'_\varepsilon = \mathrm{dil}_{\lambda'_\varepsilon,2}(\boldsymbol{\beta}), \quad \mu'_{1,\varepsilon} := \mathfrak{h}_1^1(\boldsymbol{\beta}'_\varepsilon).$$

Since $\lambda'_\varepsilon \leq 1/2$ we easily get $\mu'_{1,\varepsilon} \leq \frac{1}{2}\mu_1 \ll \mu_1$ and we call $\varrho'_{1,\varepsilon}$ the Lebesgue density of $\mu'_{1,\varepsilon}$ with respect to μ_1, i.e.

$$\mu'_{1,\varepsilon} = \varrho'_{1,\varepsilon}\mu_1, \quad \varrho'_{1,\varepsilon} \quad \text{is a Borel map with values in } (0, 1/2]. \tag{3.39}$$

In a similar way we define the homogeneous plan $\boldsymbol{\beta}''_\varepsilon$ and its first homogeneous marginal

$$\boldsymbol{\beta}''_\varepsilon := \mathrm{dil}_{(1-\lambda''_\varepsilon),2}(\boldsymbol{\alpha}^{tg}), \quad \mu''_{1,\varepsilon} := \mathfrak{h}_1^1(\boldsymbol{\beta}''_\varepsilon), \quad \mu''_{1,\varepsilon} = \varrho''_{1,\varepsilon}\mu_1.$$

We can select a Borel representative of $\varrho''_{1,\varepsilon}$ such that $(1 - \varrho''_{1,\varepsilon})$ takes values in $(0, 1/2]$. We eventually consider

$$\varrho_\varepsilon := \varrho'_{1,\varepsilon} \wedge (1 - \varrho''_{1,\varepsilon}), \quad \vartheta'_\varepsilon := \frac{\varrho_\varepsilon}{\varrho'_{1,\varepsilon}}, \qquad \vartheta''_\varepsilon := \frac{\varrho_\varepsilon}{1 - \varrho''_{1,\varepsilon}}, \tag{3.40}$$

$$\sigma'_\varepsilon := \vartheta'_\varepsilon\lambda'_\varepsilon, \qquad \sigma''_\varepsilon := \vartheta''_\varepsilon(1 - \lambda''_\varepsilon) + (1 - \vartheta''_\varepsilon), \tag{3.41}$$

$$\boldsymbol{\alpha}^o_\varepsilon := \mathrm{dil}_{\sigma'_\varepsilon,2}(\boldsymbol{\beta}), \qquad \boldsymbol{\alpha}^{tg}_\varepsilon := \mathrm{dil}_{\sigma''_\varepsilon,2}(\boldsymbol{\alpha}^{tg}). \tag{3.42}$$

Since the $\mathrm{dil}_{\cdot,2}$ dialations preserve the second homogeneous marginals, we easily get

$$\mathfrak{h}_2^1(\boldsymbol{\alpha}^o_\varepsilon) = \mu_2^o, \quad \mathfrak{h}_2^1(\boldsymbol{\alpha}^{tg}_\varepsilon) = \mu_2^{tg}. \tag{3.43}$$

Concerning the first marginals, for every test function $\zeta \in \mathrm{C}_b(\mathsf{X}_1)$ we have

$$\begin{aligned}
\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\,\mathrm{d}\boldsymbol{\alpha}^o_\varepsilon &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\vartheta'_\varepsilon\lambda'_\varepsilon\,\mathrm{d}\boldsymbol{\beta} = \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\vartheta'_\varepsilon\,\mathrm{d}\boldsymbol{\beta}'_\varepsilon\\
&= \int_{\mathsf{X}_1} \zeta\vartheta'_\varepsilon\,\mathrm{d}\mu'_{1,\varepsilon} = \int_{\mathsf{X}_1} \zeta\vartheta'_\varepsilon\varrho'_\varepsilon\,\mathrm{d}\mu_1 = \int_{\mathsf{X}_1} \zeta\rho_\varepsilon\,\mathrm{d}\mu_1,\\
\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\,\mathrm{d}\boldsymbol{\alpha}^{tg}_\varepsilon &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\sigma''_\varepsilon\,\mathrm{d}\boldsymbol{\alpha}^{tg}\\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\vartheta''_\varepsilon(1-\lambda''_\varepsilon)\,\mathrm{d}\boldsymbol{\alpha}^{tg} + \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1(1-\vartheta''_\varepsilon)\,\mathrm{d}\boldsymbol{\alpha}^{tg}\\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\zeta\circ\mathsf{x}_1)\mathsf{r}_1\vartheta''_\varepsilon\,\mathrm{d}\boldsymbol{\beta}''_\varepsilon + \int_{\mathsf{X}_1} \zeta(1-\vartheta''_\varepsilon)\,\mathrm{d}\mu_1\\
&= \int_{\mathsf{X}_1} \zeta\vartheta''_\varepsilon\,\mathrm{d}\mu''_{1,\varepsilon} + \int_{\mathsf{X}_1} \zeta(1-\vartheta''_\varepsilon)\,\mathrm{d}\mu_1\\
&= \int_{\mathsf{X}_1} \zeta\Big(\vartheta''_\varepsilon\varrho''_{1,\varepsilon} + (1-\vartheta''_\varepsilon)\Big)\,\mathrm{d}\mu_1\\
&= \int_{\mathsf{X}_1} \zeta(1-\varrho_\varepsilon)\,\mathrm{d}\mu_1
\end{aligned}$$

so that

$$\mathfrak{h}^1_1(\boldsymbol{\alpha}^o_\varepsilon) = \varrho_\varepsilon\mu_1,\quad \mathfrak{h}^1_1(\boldsymbol{\alpha}^{tg}_\varepsilon) = (1-\varrho_\varepsilon)\mu_1.$$

We deduce that the plan

$$\boldsymbol{\alpha}_\varepsilon := \boldsymbol{\alpha}^{tg}_\varepsilon + \boldsymbol{\alpha}^o_\varepsilon \quad\text{belongs to}\quad \mathfrak{H}^1(\mu_1,\mu_2).$$

By linearity, the H-cost associated with $\boldsymbol{\alpha}_\varepsilon$ is the sum of the corresponding costs associated with $\boldsymbol{\alpha}^{tg}_\varepsilon$ and $\boldsymbol{\alpha}^o_\varepsilon$; using (3.38) and observing that $0 < \sigma'_\varepsilon \le \lambda'_\varepsilon$, $1-\lambda''_\varepsilon \le \sigma''_\varepsilon \le 1$ we obtain

$$\begin{aligned}
\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^o_\varepsilon &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\big(\sigma'_\varepsilon\mathfrak{y}_1,\mathfrak{y}_2\big)\,\mathrm{d}\boldsymbol{\beta}(\mathfrak{y}_1,\mathfrak{y}_2) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \Big(\mathsf{H}\big(\mathfrak{o}_1,\mathfrak{y}_2\big)+\varepsilon_0\Big)\,\mathrm{d}\boldsymbol{\beta}(\mathfrak{y}_1,\mathfrak{y}_2)\\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^o + \varepsilon_0\mu_1(\mathsf{X}_1) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^o + \varepsilon/2,\\
\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^{tg}_\varepsilon &= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}(\sigma''_\varepsilon\mathfrak{y}_1,\mathfrak{y}_2)\,\mathrm{d}\boldsymbol{\alpha}^{tg}(\mathfrak{y}_1,\mathfrak{y}_2) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\mathsf{H}+\varepsilon)\,\mathrm{d}\boldsymbol{\alpha}^{tg}\\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^{tg} + \varepsilon\boldsymbol{\alpha}^{tg}(\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}^{tg} + \varepsilon/2.
\end{aligned}$$

□

Theorem 3.24. *Let X_i be Polish spaces, let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2] \to [0,+\infty]$ be a proper, radially 1-homogeneous and continuous function, and let $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ be such that μ_1 is a diffuse measure and $\mu_1(\mathsf{X}_1) > 0$. Then for every $\varepsilon > 0$ and every $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2)$ such that $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} < +\infty$ there exists a transport-growth map (T,g) connecting μ_1 to μ_2 such that*

$$\int_{\mathsf{X}_1} \mathsf{H}([x_1,1],[\mathsf{T}(x_1),\mathsf{g}(x_1)])\,\mathrm{d}\mu_1(x_1) \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} + \varepsilon.$$

In particular

$$\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2) = \mathscr{M}_{\mathsf{H}}(\mu_1,\mu_2).$$

Proof. Let $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ with $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} < +\infty$ and $\varepsilon > 0$ be fixed. By Proposition 3.23 we can find $\boldsymbol{\alpha}_\varepsilon \in \mathfrak{H}^1(\mu_1, \mu_2)$ with $\boldsymbol{\alpha}_\varepsilon(\{\mathfrak{y}_1 = \mathfrak{o}_1\}) = 0$ such that

$$\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_\varepsilon \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} + \varepsilon/2.$$

Using Lemma 2.8 applied to $\boldsymbol{\alpha}_\varepsilon$ and ϑ defined as

$$\vartheta([x_1, r_1], [x_2, r_2]) = \begin{cases} r_1 & \text{if } r_1 > 0, \\ 1 & \text{if } r_1 = 0, \end{cases}$$

we can assume that $\boldsymbol{\alpha}_\varepsilon$ is concentrated on $\{\mathsf{r}_1 = 1\}$. Since μ_1 is non atomic, we deduce that also $\alpha^1_\varepsilon := \pi^1_\sharp \boldsymbol{\alpha}_\varepsilon$ is non atomic so that, also observing that H is continuous, we can find thanks to Theorem 3.21 a Borel map $\mathsf{F} : \mathfrak{C}[\mathsf{X}_1] \to \mathfrak{C}[\mathsf{X}_2]$ such that $\mathsf{F}_\sharp \alpha^1_\varepsilon = \alpha^2_\varepsilon := \pi^2_\sharp \boldsymbol{\alpha}_\varepsilon$ and

$$\int_{\mathfrak{C}[\mathsf{X}_1]} \mathsf{H}(\mathfrak{y}_1, \mathsf{F}(\mathfrak{y}_1))\,\mathrm{d}\alpha^1_\varepsilon(\mathfrak{y}_1) - \varepsilon/2 \le \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}_\varepsilon. \tag{3.44}$$

We define the Borel maps

$$\mathsf{T}(x_1) := (\mathsf{x} \circ \mathsf{F})([x_1, 1]), \quad \mathsf{g}(x_1) := (\mathsf{r} \circ \mathsf{F})([x_1, 1]), \quad x_1 \in \mathsf{X}_1,$$

and we notice that $(\mathsf{T}, \mathsf{g})_\star \mu_1 = \mu_2$ and

$$\int_{\mathfrak{C}[\mathsf{X}_1]} \mathsf{H}(\mathfrak{y}_1, \mathsf{F}(\mathfrak{y}_1))\,\mathrm{d}\alpha^1_\varepsilon(\mathfrak{y}_1) = \int_{\mathsf{X}_1} \mathsf{H}([x_1, 1], [\mathsf{T}(x_1), \mathsf{g}(x_1)])\,\mathrm{d}\mu_1(x_1).$$

This concludes the proof. □

4. Existence of a maximazing pair for the dual problem

In this section we provide sufficent conditions for the existence of a maximizing pair $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ (see Definition 3.12) i.e. such that $(\varphi_1, \varphi_2) \in \mathrm{C}_b(\mathsf{X}_1) \times \mathrm{C}_b(\mathsf{X}_2)$ and

$$\varphi_1(x_1) r_1 + \varphi_2(x_2) r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \text{ for every } ([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2],$$
$$\int_{\mathsf{X}_1} \varphi_1\,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\,\mathrm{d}\mu_2 = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2).$$

A somehow complementary set of assumptions for which the same conclusion holds is presented in Appendix B. The following result is a simple consequence of compactness.

Lemma 4.1. *Let $(\mathsf{X}_i, \mathsf{d}_i)$ be compact metric spaces and assume that $\Omega \subset \mathsf{X}_1 \times \mathsf{X}_2$ is an open set such that $\pi^i(\Omega) = \mathsf{X}_i$, $i = 1, 2$. Then there exists a finite set $\mathcal{U} := \{x^n_1, x^n_2, r_n\}_{n=1}^N \subset \mathsf{X}_1 \times \mathsf{X}_2 \times (0, +\infty)$ such that*

$$\bigcup_n B_{r_n}(x^n_i) = \mathsf{X}_i, \quad i = 1, 2 \qquad \bigcup_n \overline{B_{r_n}(x^n_1)} \times \overline{B_{r_n}(x^n_2)} \subset \Omega. \tag{4.1}$$

If $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ is such that $\mathsf{supp}\,(\mu_i) = \mathsf{X}_i$, then

$$m(\mathcal{U}, \mu_1, \mu_2) := \min_{i=1,2} \min_{n=1,\dots,N} \mu_i\,(B_{r_n}(x^n_i)) > 0. \tag{4.2}$$

In order to use this result, in this section, we are going to assume without mentioning it again that X_1 and X_2 are compact and metrizable spaces. We also assume that

$$\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty) \text{ is continuous, radially 1-homogeneous and convex}$$

and that there exists an open set $\Omega_{\mathsf{H}} \subset \mathsf{X}_1 \times \mathsf{X}_2$ with $\pi^i(\Omega_{\mathsf{H}}) = \mathsf{X}_i$ such that

$$\lim_{r_1 \downarrow 0} \frac{\mathsf{H}([x_1, r_1], [x_2, 1]) - \mathsf{H}([\mathfrak{o}_1, [x_2, 1])}{r_1} = -\infty \quad \text{for every } (x_1, x_2) \in \Omega_{\mathsf{H}}, \tag{4.3}$$
$$\lim_{r_2 \downarrow 0} \frac{\mathsf{H}([x_1, 1], [x_2, r_2]) - \mathsf{H}([x_1, 1], \mathfrak{o}_2)}{r_2} = -\infty \quad \text{for every } (x_1, x_2) \in \Omega_{\mathsf{H}}.$$

To simplify the notation, we set

$$\mathsf{H}_1(x_1) := \mathsf{H}([x_1,1],\mathfrak{o}_2), \quad \mathsf{H}_2(x_2) := \mathsf{H}(\mathfrak{o}_1,[x_2,1]) \quad x_1 \in \mathsf{X}_1,\ x_2 \in \mathsf{X}_2$$

and

$$\kappa_1 := \int_{\mathsf{X}_1} \mathsf{H}_1 \,\mathrm{d}\mu_1 < +\infty, \quad \kappa_2 := \int_{\mathsf{X}_2} \mathsf{H}_2 \,\mathrm{d}\mu_2 < +\infty. \tag{4.4}$$

Clearly $\mathsf{H} \le \mathsf{H}_1 + \mathsf{H}_2$; we remark that the meaning of (4.3) is to impose a control on the derivatives of H at the boundary of $\mathbb{R}^2_+$ for a sufficiently large set of points (x_1, x_2).

Example 4.2. Both the functions in (2.19) and (2.18) satisfy (4.3) with $\Omega_{\mathsf{H}_{\mathsf{GHK}}} = \mathbb{R}^d \times \mathbb{R}^d$ and $\Omega_{\mathsf{H}_{\mathsf{HK}}} = \{(x_1,x_2) \in \mathbb{R}^d \times \mathbb{R}^d : |x_1 - x_2| < \pi/2\}$.

We start with a few preliminary lemmas that provide bounds on pairs in Φ_{H}.

Lemma 4.3. *Assume that* H *is as in* (4.3) *and let* $\mathcal{U} = \{x_1^n, x_2^n, r_n\}_{n=1}^N$ *be as in Lemma 4.1 for distances* d_i *metrizing* X_i, $i = 1,2$ *and* $\Omega = \Omega_{\mathsf{H}}$. *If* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that* $\mathsf{supp}\,(\mu_i) = \mathsf{X}_i$, *then any pair* $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ *such that*

$$\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \,\mathrm{d}\mu_2 \ge 0$$

satisfies also

$$\max_{B_{r_n}(x_i^n)} \varphi_i \ge -\frac{\kappa_1 + \kappa_2 + 1}{m(\mathcal{U}, \mu_1, \mu_2)} \quad \text{for every } n \in \{1, \dots, N\} \text{ and } i = 1, 2,$$

where κ_i *are as in* (4.4) *and* $m(\mathcal{U}, \mu_1, \mu_2)$ *is as in* (4.2).

Proof. We claim that

$$\int_{B_{r_i}(x_1^n)} \varphi_1 \,\mathrm{d}\mu_1 \ge -(\kappa_1 + \kappa_2 + 1), \quad \int_{B_{r_i}(x_2^n)} \varphi_2 \,\mathrm{d}\mu_2 \ge -(\kappa_1 + \kappa_2 + 1) \quad n = 1, \dots, N. \tag{4.5}$$

Indeed, if there exists $i \in \{1,2\}$ (say $i = 1$) and $n \in \{1, \dots, N\}$ such that

$$\int_{B_{r_n}(x_1^n)} \varphi_1 \,\mathrm{d}\mu_1 < -(\kappa_1 + \kappa_2 + 1),$$

then

$$\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 = \int_{B_{r_n}(x_1^n)} \varphi_1 \,\mathrm{d}\mu_1 + \int_{\mathsf{X}_1 \setminus B_{r_n}(x_1^n)} \varphi_1 \,\mathrm{d}\mu_1 < -(\kappa_1 + \kappa_1 + 1) + \kappa_1 = -(\kappa_2 + 1).$$

Thus

$$\kappa_2 \ge \int_{\mathsf{X}_2} \varphi_2 \,\mathrm{d}\mu_2 \ge -\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 \ge \kappa_2 + 1,$$

a contradiction. By (4.5) we have, for every $i = 1, 2$ and $n = 1, \dots, N$, that

$$\mu_i\left(B_{r_n}(x_i^n)\right) \sup_{B_{r_n}(x_i^n)} \varphi_i \ge \int_{B_{r_n}(x_i^n)} \varphi_i \,\mathrm{d}\mu_i \ge -(\kappa_1 + \kappa_2 + 1),$$

hence

$$\sup_{B_{r_n}(x_i^n)} \varphi_i \ge -\frac{\kappa_1 + \kappa_2 + 1}{m(\mathcal{U}, \mu_1, \mu_2)}.$$

□

Lemma 4.4. *Assume that* H *is as in* (4.3) *and that* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that* $\mathsf{supp}\,(\mu_i) = \mathsf{X}_i$. *Then there exists a constant* $\varepsilon > 0$ *such that any pair* $(\varphi_1, \varphi_2) \in \Phi_\mathsf{H}$ *such that*

$$\int_{\mathsf{X}_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \, \mathrm{d}\mu_2 \geq 0$$

satisfies also

$$\varphi_i(x_i) \leq \mathsf{H}_i(x_i) - \varepsilon \quad \textit{for every } x_i \in \mathsf{X}_i, \quad i = 1, 2.$$

Proof. We prove the statement for $i = 1$, being the other case completely analogous. Suppose by contradiction that there exists $(\varphi_1^j, \varphi_2^j)_j \subset \Phi_\mathsf{H}$ with $\int_{\mathsf{X}_1} \varphi_1^j \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2^j \, \mathrm{d}\mu_2 \geq 0$ and $(z_j)_j \subset \mathsf{X}_1$ such that

$$\mathsf{H}_1(z_j) - \varphi_1^j(z_j) \to 0$$

as $j \to +\infty$. Up to passing to a subsequence, we can assume that

$$0 \leq \mathsf{H}_1(z_j) - \varphi_1^j(z_j) \leq \frac{1}{j} \quad \text{for every } j \in \mathbb{N}$$

and, by compactness of X_1, the existence of $z \in \mathsf{X}_1$ such that $z_j \to z$. Let $\mathcal{U} = \{x_1^n, x_2^n, r_n\}_{n=1}^N$ be as in Lemma 4.1 for distances d_i metrizing X_i, $i = 1, 2$ and $\Omega = \Omega_\mathsf{H}$; let

$$C := \frac{\kappa_1 + \kappa_2 + 1}{m(\mathcal{U}, \mu_1, \mu_2)}.$$

Since $z \in B_{r_n}(x_1^n)$ (see (4.1)) for some $n \in \{1, \dots, N\}$, we can assume, up to passing again to a subsequence, that $z_j \in B_{r_n}(x_1^n)$ for every $j \in \mathbb{N}$. By Lemma 4.3, we can find, $y_j \in \overline{B_{r_n}(x_2^n)}$ such that $\varphi_2^j(y_j) \geq -C$. By compactness of X_2, we can assume that $y_j \to y \in \overline{B_{r_n}(x_2^n)}$. We have thus proven the existence of $(z_j, y_j) \in \Omega$ such that $(z_j, y_j) \to (z, y) \in \Omega$ with

$$0 \leq \mathsf{H}_1(z_j) - \varphi_1^j(z_j) \leq \frac{1}{j}, \quad \varphi_2^j(y_j) \geq -C \quad \text{for every } j \in \mathbb{N}.$$

We have

$$r_1 \left(\mathsf{H}_1(z_j) - \frac{1}{j} \right) - C r_2 \leq \varphi_1^j(z_j) r_1 + \varphi_2^j(y_j) r_2 \leq \mathsf{H}([z_j, r_1], [y_j, r_2]) \quad \text{for every } r_1, r_2 \geq 0.$$

Choosing $r_1 = 1$, we get

$$\frac{\mathsf{H}([z_j, 1], [y_j, r_2] - \mathsf{H}_1(z_j)}{r_2} \geq -C - \frac{1}{j r_2} \quad \text{for every } j \in \mathbb{N},\ r_2 > 0.$$

Passing first to the limit as $j \to +\infty$ and then to the limit as $r_2 \downarrow 0$, we obtain

$$\lim_{r_2 \downarrow 0} \frac{\mathsf{H}([z, 1], [y, r_2] - \mathsf{H}_1(x_\infty)}{r_2} \geq -C > -\infty,$$

a contradiction with (4.3). □

The following definition is simply the analogue of the classical c-transform (see e.g. [AGS08, Definition 6.1.2]).

Definition 4.5. Let $(\varphi_1, \varphi_2) \in \Phi_\mathsf{H}$. We define the Borel functions $\varphi_1^\mathsf{H} : \mathsf{X}_2 \to \mathbb{R}$, $\varphi_1^{\mathsf{HH}} : \mathsf{X}_1 \to \mathbb{R}$ as

$$\varphi_1^\mathsf{H}(x_2) := \inf_{x_1 \in \mathsf{X}_1} \inf_{\alpha \geq 0} \Big\{ \mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1) \Big\}, \quad x_2 \in \mathsf{X}_2,$$

$$\varphi_1^{\mathsf{HH}}(x_1) := \inf_{x_2 \in \mathsf{X}_2} \inf_{\alpha \geq 0} \Big\{ \mathsf{H}([x_1, 1], [x_2, \alpha]) - \alpha \varphi_1^\mathsf{H}(x_2) \Big\}, \quad x_1 \in \mathsf{X}_1.$$

Using the previous lemmas, in the following proposition we prove that the H-transform of a pair in Φ_{H} can be computed restricting the minimization to a compact set. As a consequence, we obtain uniform bounds and uniform continuity for the admissible pair, as it happens in the classical case (see e.g. the discussion after [San15, Definition 1.10]).

Proposition 4.6. *Assume that* H *is as in* (4.3) *and that* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that* $\mathsf{supp}\,(\mu_i) = \mathsf{X}_i$. *Then there exists constants* $R > 1$ *and* $M > 0$ *such that, if* $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ *are such that*

$$\int_{\mathsf{X}_1} \varphi_1 \,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \,\mathrm{d}\mu_2 \geq 0,$$

then $\|\varphi_1^{\mathsf{H}}\|_\infty, \|\varphi_1^{\mathsf{HH}}\|_\infty \leq M$ *and*

$$\varphi_1^{\mathsf{H}}(x_2) = \inf_{x_1 \in \mathsf{X}_1} \inf_{0 \leq \alpha \leq R} \Big\{ \mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1) \Big\}, \quad x_2 \in \mathsf{X}_2, \tag{4.6}$$

$$\varphi_1^{\mathsf{HH}}(x_1) = \inf_{x_2 \in \mathsf{X}_2} \inf_{0 \leq \alpha \leq R} \Big\{ \mathsf{H}([x_1, 1], [x_2, \alpha]) - \alpha \varphi_1^{\mathsf{H}}(x_2) \Big\}, \quad x_1 \in \mathsf{X}_1. \tag{4.7}$$

In particular, $(\varphi_1^{\mathsf{HH}}, \varphi_1^{\mathsf{H}}) \in \Phi_{\mathsf{H}}$, $\varphi_1^{\mathsf{HH}} \geq \varphi_1$, $\varphi_1^{\mathsf{H}} \geq \varphi_2$. *Finally, if* d_i *are distances metrizing* X_i, $i = 1, 2$, *then* φ_1^{HH} *is* d_1*-uniformly continuous and* φ_1^{H} *is* d_2*-uniformly continuous, both with the same (uniform)* $\mathsf{d}_1 \otimes_{\mathfrak{C}} \mathsf{d}_2$*-modulus of continuity of* H *on* $\mathfrak{C}_R[\mathsf{X}_1, \mathsf{X}_2]$ *(cf.* (2.10)*).*

Proof. Let (φ_1, φ_2) be as in the statement. By Lemma 4.4 we know that there exists $\varepsilon > 0$ (not depending on the pair) such that

$$\varphi_1(x_1) \leq \mathsf{H}_1(x_1) - \varepsilon \quad \text{for every } x_1 \in \mathsf{X}_1.$$

Then, by uniform continuity of H on $\mathfrak{C}_1[\mathsf{X}_1, \mathsf{X}_2]$, we can find $0 < \delta < 1$ such that

$$|\mathsf{H}([x_1, 1], [x_2, r_2]) - \mathsf{H}_1(x_1)| \leq \frac{\varepsilon}{2} \quad \text{for every } 0 \leq r_2 \leq \delta.$$

If we define $R := 1 + \frac{1}{\delta} + \frac{2}{\varepsilon}\left(\max_{\mathsf{X}_2} \mathsf{H}_2 + \max_{\mathsf{X}_1} \mathsf{H}_1 + 1\right)$, then, for every $\alpha > R$, we have

$$\begin{aligned}
\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1) &= \mathsf{H}([x_1, \alpha], [x_2, 1]) - \mathsf{H}_1(x_1)\alpha + \alpha\left(\mathsf{H}_1(x_1) - \varphi_1(x_1)\right) \\
&= \alpha\left(\mathsf{H}([x_1, 1], [x_2, 1/\alpha]) - \mathsf{H}_1(x_1)\right) + \alpha\left(\mathsf{H}_1(x_1) - \varphi_1(x_1)\right) \\
&\geq \alpha \frac{\varepsilon}{2} \\
&\geq \mathsf{H}_2(x_2) + 1.
\end{aligned}$$

Thus, for every $x_2 \in \mathsf{X}_2$, we get

$$\begin{aligned}
\inf_{x_1 \in \mathsf{X}_1} \inf_{\alpha > R} \{\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1)\} &\geq \mathsf{H}_2(x_2) + 1 \\
&> \inf_{x_1 \in \mathsf{X}_1} \inf_{0 \leq \alpha \leq R} \{\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1)\}
\end{aligned}$$

and this proves (4.6). The proof of (4.7) is analogous.
The fact that $\varphi_1^{\mathsf{H}} \geq \varphi_2$, $\varphi_1^{\mathsf{HH}} \geq \varphi_1$ and

$$\varphi_1^{\mathsf{HH}}(x_1) r_1 + \varphi_1^{\mathsf{H}}(x_2) r_2 \leq \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ r_1, r_2 \geq 0$$

follows by the definition of φ_1^{H} and φ_1^{HH}. It is then clear that φ_1^{H} (resp. φ_1^{HH}) is bounded from below by $\min_{x_2 \in \mathsf{X}_2} \varphi_2$ (resp. $\min_{x_1 \in \mathsf{X}_1} \varphi_1$) and by above by $\max_{x_2 \in \mathsf{X}_2} \mathsf{H}_2$ (resp. $\max_{x_1 \in \mathsf{X}_2} \mathsf{H}_1$).

Let d_i be distances metrizing X_i, $i=1,2$; let now $x_2, x_2' \in \mathsf{X}_2$, then, recalling (2.11), we have

$$\begin{aligned}\left|\varphi_1^{\mathsf{H}}(x_2)-\varphi_1^{\mathsf{H}}(x_2')\right| &\le \sup_{x_1\in\mathsf{X}_1}\sup_{0\le\alpha\le R}\left|\mathsf{H}([x_1,\alpha],[x_2,1])-\mathsf{H}([x_1,\alpha],[x_2',1])\right| \\ &\le \omega_{\mathsf{H}}^R\left((\mathsf{d}_1\otimes_{\mathfrak{C}}\mathsf{d}_2)\left(([x_1,\alpha],[x_2,1]),([x_1,\alpha],[x_2',1])\right)\right) \\ &= \omega_{\mathsf{H}}^R(\mathsf{d}_{2,\mathfrak{C}}([x_2,1],[x_2',1])) \\ &\le \omega_{\mathsf{H}}^R(\mathsf{d}_2(x_2,x_2')),\end{aligned} \tag{4.8}$$

where ω_{H}^R is the (uniform) modulus of continuity of H on $\mathfrak{C}_R[\mathsf{X}_1,\mathsf{X}_2]$ and we have used that $\mathsf{d}_{2,\mathfrak{C}}([x_2,1],[x_2',1]))\le\mathsf{d}_2(x_2,x_2')$ (see (2.9), (2.11) and formula (7.5) in [LMS18]). The analogous statement for φ_1^{HH} follows by the same strategy. This proves that φ_1^{H} and φ_1^{HH} are uniformly continuous with the same (uniform) modulus of continuity of H on $\mathfrak{C}_R[\mathsf{X}_1,\mathsf{X}_2]$ and concludes the proof that $(\varphi_1^{\mathsf{HH}},\varphi_1^{\mathsf{H}})\in\Phi_{\mathsf{H}}$. Let $\mathcal{U}=\{x_1^n,x_2^n,r_n\}_{n=1}^N$ be as in Lemma 4.1 for the distances d_i and $\Omega=\Omega_{\mathsf{H}}$; we define (recalling Lemma 4.3)

$$M:=\frac{\kappa_1+\kappa_2+1}{m(\mathcal{U},\mu_1,\mu_2)}+\omega_{\mathsf{H}}^R(\operatorname{diam}\mathsf{X}_1)+\omega_{\mathsf{H}}^R(\operatorname{diam}\mathsf{X}_2)+\max_{x_1\in\mathsf{X}_1}\mathsf{H}_1+\max_{x_2\in\mathsf{X}_2}\mathsf{H}_2,$$

we have that $\varphi_1^{\mathsf{H}}\le\mathsf{H}_2\le M$ and, by (4.8), we get

$$\varphi_1^{\mathsf{H}}(x_2)\ge\varphi_1^{\mathsf{H}}(x_2')-\omega_{\mathsf{H}}^R(\mathsf{d}_2(x_2,x_2'))\ge-\frac{\kappa_1+\kappa_2+1}{m(\mathcal{U},\mu_1,\mu_2)}-\omega_{\mathsf{H}}^R(\operatorname{diam}\mathsf{X}_2)\ge -M \quad \text{for every } x_2\in\mathsf{X}_2,$$

where $x_2'\in\mathsf{X}_2$ is some point where φ_1^{H} is larger than $-\frac{\kappa_1+\kappa_2+1}{m(\mathcal{U},\mu_1,\mu_2)}$ (whose existence is given by Lemma 4.3). The proof for φ_1^{HH} is the same. □

With the result of Proposition 4.6 it is straightforward to obtain the existence of a maximizing pair.

Theorem 4.7 (Existence of optimal continuous potentials)**.** *Assume that* H *is as in* (4.3) *and that* $\mu_i\in\mathcal{M}_+(\mathsf{X}_i)$ *are such that* $\operatorname{supp}(\mu_i)=\mathsf{X}_i$. *Then there exists* $(\varphi_1,\varphi_2)\in\Phi_{\mathsf{H}}$ *such that*

$$\int_{\mathsf{X}_1}\varphi_1\,\mathrm{d}\mu_1+\int_{\mathsf{X}_2}\varphi_2\,\mathrm{d}\mu_2=\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2).$$

Proof. If $\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2)=0$, we can take φ_1 and φ_2 to be the null functions. We thus assume that $\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2)>0$. If this is the case, we can find a maximizing sequence $(\varphi_1^j,\varphi_2^j)_j\subset\Phi_{\mathsf{H}}$ for the dual problem (3.31) with

$$\int_{\mathsf{X}_1}\varphi_1^j\,\mathrm{d}\mu_1+\int_{\mathsf{X}_2}\varphi_2^j\,\mathrm{d}\mu_2\ge 0 \quad \text{for every } j\in\mathbb{N}.$$

If we consider distances d_i metrizing X_i, by Proposition 4.6 we have that $(\varphi_1^{j,\mathsf{HH}},\varphi_1^{j,\mathsf{H}})_j\subset\Phi_{\mathsf{H}}$ is a maximizing sequence of equi-uniformly continuous and equi-bounded functions. By Arzelà–Ascoli theorem, we can assume, up to passing to a subsequence, that there exists a pair $(\varphi_1,\varphi_2)\in\Phi_{\mathsf{H}}$ such that $(\varphi_1^{j,\mathsf{HH}},\varphi_1^{j,\mathsf{H}})\to(\varphi_1,\varphi_2)$ uniformly on the compact space $\mathsf{X}_1\times\mathsf{X}_2$. By dominated convergence, we have

$$\int_{\mathsf{X}_1}\varphi_1\,\mathrm{d}\mu_1+\int_{\mathsf{X}_2}\varphi_2\,\mathrm{d}\mu_2=\lim_j\left(\int_{\mathsf{X}_1}\varphi_1^{j,\mathsf{HH}}\,\mathrm{d}\mu_1+\int_{\mathsf{X}_2}\varphi_1^{j,\mathsf{H}}\,\mathrm{d}\mu_2\right)=\mathscr{U}_{\mathsf{H}}(\mu_1,\mu_2). \qquad □$$

5. Optimality conditions

In this section we provide sufficient and necessary conditions for a plan $\boldsymbol{\alpha}\in\mathfrak{H}^1(\mu_1,\mu_2)$ to be optimal. In the following X_1 and X_2 are completely regular spaces and we will often assume that

$$\mathsf{H}:\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]\to[0,+\infty] \text{ is a proper, radially 1-homogeneous, convex and l.s.c. function.} \tag{5.1}$$

The cyclical monotonicity of the support of an admissible plan plays a crucial role also in the unbalanced setting. We recall here the important definition of cyclical monotonicity of a given set with respect to a cost H.

Definition 5.1 (H-cyclical monotonicity)**.** Let $\Gamma \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ and let $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$; we say that Γ is H-cyclically monotone if for every finite family of points $\{(\mathfrak{y}_1^i, \mathfrak{y}_2^i)\}_{i=1}^N \subset \Gamma$ and every permutation σ of $\{1, \ldots, N\}$ it holds

$$\sum_{i=1}^N \mathsf{H}(\mathfrak{y}_1^i, \mathfrak{y}_2^i) \le \sum_{i=1}^N \mathsf{H}(\mathfrak{y}_1^i, \mathfrak{y}_2^{\sigma(i)}).$$

The following result shows that optimal 1-homogeneous couplings are concentrated on a H-*cyclically monotone* set Γ which is also *a radial convex cone*. It extends to the unbalanced optimal transport setting the classical [AGS08, Necessity part of Theorem 6.14].

Proposition 5.2 (Necessity of cyclical monotonicity)**.** *Let H be as in (5.1), let $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ for $i = 1, 2$, let $\boldsymbol{\alpha} \in \mathfrak{H}_o^1(\mu_1, \mu_2)$ be optimal and suppose that $\int_{\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]} \mathsf{H} \, \mathrm{d}\boldsymbol{\alpha} < +\infty$. Then $\boldsymbol{\alpha}$ is concentrated on a σ-compact (thus Borel) radial convex cone $\Gamma \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ which is H-cyclically monotone.*

Proof. Let $\{(\varphi_1^k, \varphi_2^k)\}_{k \ge 1} \subset \Phi_{\mathsf{H}}$ be a maximizing sequence for the dual problem (3.31) and let us define

$$\mathsf{H}_k([x_1, r_1], [x_2, r_2]) := \mathsf{H}([x_1, r_1], [x_2, r_2]) - \varphi_1^k(x_1) r_1 - \varphi_2^k(x_2) r_2, \quad ([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2].$$

Since

$$\lim_{k \to \infty} \int \mathsf{H}_k([x_1, r_1], [x_2, r_2]) \, \mathrm{d}\boldsymbol{\alpha} = 0,$$

there exist a subsequence $m \mapsto k(m)$ and a σ-compact (thus Borel) subset $G \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ on which $\boldsymbol{\alpha}$ is concentrated s.t. $\mathsf{H}_{k(m)} \to 0$ on G as $m \to +\infty$. Since $\mathsf{H}_{k(m)}$ is a radially convex and 1-homogeneous function, this convergence takes place also on the radial convex cone $\Gamma := \hom(\mathrm{co}(G))$ generated by G:

$$\Gamma := \left\{ \left([x_1, \sum_{i=0}^2 \lambda_i r_1^i], [x_2, \sum_{i=0}^2 \lambda_i r_2^i] \right) \mid ([x_1, r_1^i], [x_2, r_2^i]) \in \Gamma, \ \lambda_i \ge 0 \right) \right\}.$$

Writing G as a countable union of an increasing sequence of compact sets $K_n \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$, it is easy to see that $\Gamma = \cup_{n \in \mathbb{N}} \hat{K}_n$, where $\hat{K}_n := \hom(\mathrm{co}(K_n))$ and it is not restrictive to assume that $(\mathfrak{o}_1, \mathfrak{o}_2) \in K_n$. Since each $\mathrm{co}(K_n)$ is clearly compact in $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ and $\hom(\mathrm{co}(K_n)) = \cup_{m \in \mathbb{N}} \{ (m\mathfrak{y}_1, m\mathfrak{y}_2) : (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathrm{co}(K_n) \}$ is the union of a countable family of compact sets, we conclude that Γ is σ-compact as well. Let now $\{(\mathfrak{y}_1^i, \mathfrak{y}_2^i)\}_{i=1}^N \subset \Gamma$ be a finite family of points and let σ be a permutation of $\{1, \ldots, N\}$. Then

$$\begin{aligned}
\sum_{i=1}^N \mathsf{H}(\mathfrak{y}_1^i, \mathfrak{y}_2^{\sigma(i)}) &\ge \sum_{i=1}^N \left(r_1^i \varphi_1^{k(m)}(x_1^i) + r_2^{\sigma(i)} \varphi_2^{k(m)}(x_2^{\sigma(i)}) \right) \\
&= \sum_{i=1}^N \left(r_1^i \varphi_1^{k(m)}(x_1^i) + r_2^i \varphi_2^{k(m)}(x_2^i) \right) \\
&= \sum_{i=1}^N \left(\mathsf{H}(\mathfrak{y}_1^i, \mathfrak{y}_2^i) - \mathsf{H}_{k(m)}(\mathfrak{y}_1^i, \mathfrak{y}_2^i) \right).
\end{aligned}$$

Letting $m \to +\infty$, we obtain the sought H-cyclical monotonicity of Γ. □

Remark 5.3. The previous proof shows that the radial convex cone $\hom(\mathrm{co}(G))$ generated by a σ-compact set $G \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$ is σ-compact. In particular, for a finite Radon measure $\boldsymbol{\alpha} \in \mathcal{M}_+(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2])$ the following properties are equivalent:

(i) $\boldsymbol{\alpha}$ is concentrated on a Borel set G such that the generated radial convex cone $\hom(\mathrm{co}(G))$ is H-cyclically monotone;

(ii) $\boldsymbol{\alpha}$ is concentrated on a σ-compact radial convex cone Γ which is H-cyclically monotone.

We devote the remaining part of this section to formulate a converse statement to Proposition 5.2. We first introduce a few notions related to the natural directed graph structures induced by H and subsets Γ of $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$. A similar approach has already been considered when dealing with optimality conditions for possibly infinite costs in the classical Optimal Transport theory [Bei+09; BC10].

5.1. **Simple directed graphs and oriented walks.** Recall that a directed graph $\mathcal{G}$ in a set Z is an ordered pair $(V, \mathcal{A})$ where $V \subset \mathsf{Z}$ and $\mathcal{A}$ is a set of ordered pairs in $V \times V$. A (oriented) $\mathcal{A}$-walk P in V is just a sequence of elements $(\mathfrak{y}^0, \cdots, \mathfrak{y}^N)$ in V^{N+1}, $N \in \mathbb{N}_+$, such that each pair of consecutive elements $(\mathfrak{y}^{h-1}, \mathfrak{y}^h)$ belongs to $\mathcal{A}$, $h = 1, \cdots, N$. We denote by $\mathcal{P}(\mathfrak{y}' \to \mathfrak{y}''|V, \mathcal{A})$ the collections of $\mathcal{A}$-walks in V whose first and last elements are $\mathfrak{y}'$ and $\mathfrak{y}''$ respectively. We will omit to write V when $V = \mathsf{Z}$ and we will also omit to write $\mathcal{A}$ if the set of arcs is clear from the context. When $\mathfrak{y}' = \mathfrak{y}''$ then we say that $\mathcal{P}$ is a cycle.

If $P' \in \mathcal{P}(\mathfrak{y}^1 \to \mathfrak{y}^2)$ and $P'' \in \mathcal{P}(\mathfrak{y}^2 \to \mathfrak{y}^3)$ we can construct a new walk $P = P' + P'' \in \mathcal{P}(\mathfrak{y}^1 \to \mathfrak{y}^3)$ by joining P' with P''.

We say that a (ordered) pair of points $(\mathfrak{y}', \mathfrak{y}'') \in \mathsf{Z}$ is $\mathcal{A}$-connected if $\mathcal{P}(\mathfrak{y}' \to \mathfrak{y}'')$ is not empty. A directed graph $(V, \mathcal{A})$ is connected if every pair of points in V is $\mathcal{A}$-connected.

Given a function $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ and a set $\Gamma \subset \mathrm{D}(\mathsf{H}) \subset \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2]$, we construct a directed (bipartite) graph $\mathcal{G}_{\mathsf{H},\Gamma}$ whose vertices belong to

$$\mathsf{Z} = \mathsf{Z}[\mathsf{X}_1, \mathsf{X}_2] := \mathfrak{C}[\mathsf{X}_1] \sqcup \mathfrak{C}[\mathsf{X}_2], \quad \text{the disjoint union of } \mathfrak{C}[\mathsf{X}_1] \text{ and } \mathfrak{C}[\mathsf{X}_2].$$

We first consider the set of arcs $\mathcal{A}_{\mathsf{H}}$ consisting of all the (ordered) pairs $(\mathfrak{y}_2, \mathfrak{y}_1) \in \mathsf{Z} \times \mathsf{Z}$ such that $\mathfrak{y}_i \in \mathfrak{C}[\mathsf{X}_i]$ and $\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) < +\infty$. We can also easily identify Γ with a set of arcs, i.e. all the (ordered) pairs $(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathsf{Z} \times \mathsf{Z}$ such that $(\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$ (notice that Γ can be canonically identified with a subset of $\mathsf{Z} \times \mathsf{Z}$).

We eventually set

$$\mathcal{A}_{\mathsf{H},\Gamma} := \mathcal{A}_{\mathsf{H}} \cup \Gamma.$$

When H is the null function 0 (or any finite function) we have $\mathcal{A}_0 = \mathfrak{C}[\mathsf{X}_2] \times \mathfrak{C}[\mathsf{X}_1]$ and

$$\mathcal{A}_{0,\Gamma} = \big(\mathfrak{C}[\mathsf{X}_2] \times \mathfrak{C}[\mathsf{X}_1]\big) \cup \Gamma.$$

Recalling that $\pi^i : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to \mathfrak{C}[\mathsf{X}_i]$ are the canonical projections, we also consider the subsets

$$\Gamma_i := \pi^i(\Gamma) \subset \mathfrak{C}[\mathsf{X}_i], \quad V_\Gamma := \Gamma_1 \sqcup \Gamma_2 \subset \mathsf{Z}; \tag{5.2}$$

V_Γ is the collection of all the vertices obtained by applying to Γ the two projections on $\mathfrak{C}[\mathsf{X}_i]$, $i = 1, 2$, canonically identified with the corresponding subsets of Z. We set

$$\overline{\mathcal{A}}_{\mathsf{H},\Gamma} := \mathcal{A}_{\mathsf{H},\Gamma} \cap (V_\Gamma \times V_\Gamma).$$

Notice that if the initial and final points $\mathfrak{y}^0, \mathfrak{y}^N$ of a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk P belong to $\mathfrak{C}[\mathsf{X}_1]$ we will have

$$\begin{gathered} N = 2n \text{ even}, \quad \mathfrak{y}_1^k := \mathfrak{y}^{2k} \in \mathfrak{C}[\mathsf{X}_1] \quad \text{for } 0 \le k \le n, \quad \mathfrak{y}_2^k := \mathfrak{y}^{2k+1} \in \mathfrak{C}[\mathsf{X}_2] \quad \text{for } 0 \le k < n, \\ (\mathfrak{y}^{2k}, \mathfrak{y}^{2k+1}) = (\mathfrak{y}_1^k, \mathfrak{y}_2^k) \in \Gamma, \quad \mathsf{H}(\mathfrak{y}^{2k+2}, \mathfrak{y}^{2k+1}) = \mathsf{H}(\mathfrak{y}_1^{k+1}, \mathfrak{y}_2^k) < +\infty, \\ \mathfrak{y}^{2k} = \mathfrak{y}_1^k \in \Gamma_1, \ \mathfrak{y}^{2k+1} = \mathfrak{y}_2^k \in \Gamma_2 \quad \text{for every } k = 0, \cdots, n-1. \end{gathered} \tag{5.3}$$

In particular, if also $\mathfrak{y}^N$ belongs to Γ_1 then $P \in \mathcal{P}(\mathfrak{y}^0 \to \mathfrak{y}^N | V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$. More generally, if the initial and final points of a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk P belong to V_Γ then all the points of P belong to V_Γ.

Definition 5.4 (H-connectedness)**.** We say that a set $\Gamma \subset \mathrm{D}(\mathsf{H})$ is H-connected if the graph $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$ is connected or, equivalently, if every pair of points of Γ_1 is $\mathcal{A}_{\mathsf{H},\Gamma}$-connected.

Let us now define the "oriented" cost function $\check{\mathsf{H}} : \mathsf{Z} \times \mathsf{Z} \to \bar{\mathbb{R}}$ starting from a cost function H as in (5.1):

$$\check{\mathsf{H}}(\mathfrak{y}', \mathfrak{y}'') := \begin{cases} -\mathsf{H}(\mathfrak{y}', \mathfrak{y}'') & \text{if } \mathfrak{y}' \in \mathfrak{C}[\mathsf{X}_1],\ \mathfrak{y}'' \in \mathfrak{C}[\mathsf{X}_2], \\ +\mathsf{H}(\mathfrak{y}'', \mathfrak{y}') & \text{if } \mathfrak{y}'' \in \mathfrak{C}[\mathsf{X}_1],\ \mathfrak{y}' \in \mathfrak{C}[\mathsf{X}_2], \\ 0 & \text{otherwise.} \end{cases}$$

The cost $\Theta(P)$ of a walk $P = (\mathfrak{y}^0, \cdots, \mathfrak{y}^N)$ in $(\mathsf{Z}, \mathcal{A}_{0,\Gamma})$ with $\Gamma \subset \mathrm{D}(\mathsf{H})$ is defined by

$$\Theta(P) := \sum_{h=1}^{N} \check{\mathsf{H}}(\mathfrak{y}^{h-1}, \mathfrak{y}^h);$$

notice that Θ is well defined and takes values in $\mathbb{R} \cup \{+\infty\}$ since $\Gamma \subset \mathrm{D}(\mathsf{H})$, so that the negative contributions of the arcs to $\Theta(P)$ are always finite. It is also easy to check that

$$\Theta(P) \text{ is finite if and only if } P \text{ is a } \mathcal{A}_{\mathsf{H},\Gamma}\text{-walk.} \tag{5.4}$$

In the case of (5.3) we can equivalently write

$$\Theta(P) = \sum_{k=0}^{n-1} \Big(\mathsf{H}(\mathfrak{y}_1^{k+1}, \mathfrak{y}_2^k) - \mathsf{H}(\mathfrak{y}_1^k, \mathfrak{y}_2^k) \Big).$$

Clearly, $\Gamma \subset \mathrm{D}(\mathsf{H})$ is H-cyclically monotone according to Definition 5.1 if and only if

$$\text{for every cycle } P \in \mathcal{P}(\mathfrak{y} \to \mathfrak{y}) \text{ in } (\mathsf{Z}, \mathcal{A}_{0,\Gamma}) \text{ we have } \quad \Theta(P) \geq 0. \tag{5.5}$$

Recalling (5.4) it is immediate to see that it is sufficient to check condition (5.5) on $\mathcal{A}_{\mathsf{H},\Gamma}$-cycles in V_Γ. We are going to use the following notation: if $P = (\mathfrak{y}^0, \cdots, \mathfrak{y}^N)$ and $A = (\mathfrak{y}^k, \mathfrak{y}^{k+1})$ is an internal arc for some $0 < k < N-1$, we can "remove" the arc A from P obtaining a new walk P^{-k} by setting $P^{-k} := (\mathfrak{y}^0, \cdots, \mathfrak{y}^{k-1}, \mathfrak{y}^{k+1}, \mathfrak{y}^N)$.

Lemma 5.5 (The effect of removing the vertex $(\mathfrak{o}_1, \mathfrak{o}_2)$)**.** *Let H be satisfying the standard assumption (5.1) and let $\Gamma \subset \mathrm{D}(\mathsf{H})$ be a radial cone.*

If P is a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk and $P_\mathfrak{o}$ is the walk obtained removing from P all the internal arcs of the form $(\mathfrak{o}_1, \mathfrak{o}_2)$, then

$$P_\mathfrak{o} \textit{ is a } \mathcal{A}_{\mathsf{H},\Gamma}\textit{-walk and} \quad \Theta(P_\mathfrak{o}) \leq \Theta(P). \tag{5.6}$$

Proof. Let $P = (\mathfrak{y}^0, \cdots, \mathfrak{y}^N)$ be a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk. The proof follows by a simple induction argument if we show that Θ decreases if we remove an internal arc of the form $(\mathfrak{o}_1, \mathfrak{o}_2)$. We can thus assume that $\mathfrak{y}^k = \mathfrak{o}_1, \mathfrak{y}^{k+1} = \mathfrak{o}_2$ for some $0 < k < N$ and we set $P' := P^{-k}$. We notice that $\mathfrak{y}^{k-1} = [x^{k-1}, r^{k-1}] \in \mathfrak{C}[\mathsf{X}_2]$ and $\mathfrak{y}^{k+2} = [x^{k+2}, r^{k+2}] \in \mathfrak{C}[\mathsf{X}_1]$ so that

$$\begin{aligned} \Theta(P) - \Theta(P') &= \check{\mathsf{H}}(\mathfrak{y}^{k-1}, \mathfrak{o}_1) + \check{\mathsf{H}}(\mathfrak{o}_2, \mathfrak{y}^{k+2}) - \check{\mathsf{H}}(\mathfrak{y}^{k-1}, \mathfrak{y}^{k+2}) \\ &= \mathsf{H}(\mathfrak{o}_1, \mathfrak{y}^{k-1}) + \mathsf{H}(\mathfrak{y}^{k+2}, \mathfrak{o}_2) - \mathsf{H}(\mathfrak{y}^{k+2}, \mathfrak{y}^{k-1}) \\ &= \mathsf{H}_{x^{k+2},x^{k-1}}(0, r^{k-1}) + \mathsf{H}_{x^{k+2},x^{k-1}}(r^{k+2}, 0) - \mathsf{H}_{x^{k+2},x^{k-1}}(r^{k+2}, r^{k-1}) \geq 0 \end{aligned}$$

thanks to the subadditivity of $\mathsf{H}_{x^{k+2},x^{k-1}}$. The same argument shows that $(\mathfrak{y}^{k-1}, \mathfrak{y}^{k+2}) \in \mathcal{A}_\mathsf{H}$ so that P' is an $\mathcal{A}_{\mathsf{H},\Gamma}$-walk.

□

Corollary 5.6. *Under the same assumptions of Lemma 5.5, if Γ is H-connected then also $\Gamma_\mathfrak{o} := \Gamma \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}$ is H-connected and if $\Gamma_\mathfrak{o}$ is H-cyclically monotone then also Γ is H-cyclically monotone (i.e. the arc $(\mathfrak{o}_1, \mathfrak{o}_2)$ is irrelevant for H-connectedness and H cyclical monotonicity).*

Proof. Let $\Gamma_{\mathfrak{o},i} := \pi^i(\Gamma_{\mathfrak{o}})$ and let $\mathfrak{y}', \mathfrak{y}'' \in \Gamma_{\mathfrak{o},1}$. If both are different from $\mathfrak{o}_1$, since Γ is H-connected and $\mathfrak{y}', \mathfrak{y}'' \in \Gamma_1$, we can find a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk $P \in \mathcal{P}(\mathfrak{y}' \to \mathfrak{y}'')$ and, by the previous Lemma 5.5 we can remove from P all the internal arcs of the form $(\mathfrak{o}_1, \mathfrak{o}_2)$ obtaining a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk $P_{\mathfrak{o}} \in \mathcal{P}(\mathfrak{y}' \to \mathfrak{y}'')$.

If $\mathfrak{y}' = \mathfrak{o}_1 \neq \mathfrak{y}''$, since $\mathfrak{o}_1 \in \pi^1(\Gamma_{\mathfrak{o}})$, we can find $\mathfrak{y}_2' \in \mathfrak{C}[\mathsf{X}_2] \setminus \{\mathfrak{o}_2\}$ such that $(\mathfrak{o}_1, \mathfrak{y}_2') \in \Gamma$ and, by the same argument above, a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk $P' \in \mathcal{P}(\mathfrak{y}_2' \to \mathfrak{y}'')$. It follows that $P := (\mathfrak{o}_1, \mathfrak{y}_2') + P'$ is a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk connecting $\mathfrak{o}_1$ to $\mathfrak{y}''$.

If $\mathfrak{y}' \neq \mathfrak{o}_1 = \mathfrak{y}''$, we can find $\mathfrak{y}_2'' \in \mathfrak{C}[\mathsf{X}_2] \setminus \{\mathfrak{o}_2\}$ such that $(\mathfrak{o}_1, \mathfrak{y}_2'') \in \Gamma \subset \mathrm{D}(\mathsf{H})$, so that in particular $\mathsf{H}(\mathfrak{o}_1, \mathfrak{y}_2'') < \infty$ and $(\mathfrak{y}_2'', \mathfrak{o}_1) \in \mathcal{A}_{\mathsf{H}}$. By the same argument above, we can find $P'' \in \mathcal{P}(\mathfrak{y}' \to \mathfrak{y}_2'')$ which is a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk so that $P := P'' + (\mathfrak{y}_2'', \mathfrak{o}_1) \in \mathcal{P}(\mathfrak{y}' \to \mathfrak{o}_1)$ is a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk as well.

If $\mathfrak{y}' = \mathfrak{y}'' = \mathfrak{o}_1$, the two arguments above show that we can find $\mathfrak{y}_2', \mathfrak{y}_2'' \in \mathfrak{C}[\mathsf{X}_2] \setminus \{\mathfrak{o}_2\}$ such that $(\mathfrak{o}_1, \mathfrak{y}_2') \in \Gamma$ and $(\mathfrak{y}_2'', \mathfrak{o}_1) \in \mathcal{A}_{\mathsf{H}}$. Since both $\mathfrak{y}_2', \mathfrak{y}_2'' \in \Gamma_2$ are different from $\mathfrak{o}_2$, we can apply again Lemma 5.5 and obtain a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk $P''' \in \mathcal{P}(\mathfrak{y}_2' \to \mathfrak{y}_2'')$. We can thus define $P := (\mathfrak{o}_1, \mathfrak{y}_2') + P''' + (\mathfrak{y}_2'', \mathfrak{o}_1) \in \mathcal{P}(\mathfrak{y}' \to \mathfrak{y}'')$ which is a $\mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$-walk.

Let us now suppose that $\Gamma_{\mathfrak{o}}$ is H-cyclically monotone and let P be a $\mathsf{A}_{\mathsf{H},\Gamma}$-cycle in V_Γ. It is not restrictive to suppose that P contains at least an element $\mathfrak{y} \in V_\Gamma \setminus \{\mathfrak{o}_1, \mathfrak{o}_2\}$ (otherwise $\Theta(P) = 0$). By a cyclic permutation of the arcs of P we can assume that the initial and final point of P is $\mathfrak{y}$. By Lemma 5.5 we can construct a cycle $P_{\mathfrak{o}}$ in $(V_{\Gamma_{\mathfrak{o}}}, \mathcal{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}})$ obtained by removing all the arcs $(\mathfrak{o}_1, \mathfrak{o}_2)$ from P with the same initial and final point. Since $\Gamma_{\mathfrak{o}}$ is H-cyclically monotone, we have $\Theta(P_{\mathfrak{o}}) \geq 0$. The inequality (5.6) shows that $\Theta(P) \geq 0$ as well. □

The next Proposition shows how H-cyclical monotonicity reflects on walks that are not precisely a cycle (cf. (5.5)) but are of the form $P \in \mathcal{P}(\mathfrak{y} \to \lambda\mathfrak{y})$ for some $\lambda > 0$.

Given $(\mathfrak{y}_1, \mathfrak{y}_2) = ([x_1, r_1], [x_2, r_2]) \in \mathrm{D}(\mathsf{H})$, we denote by $\partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2)$ the subdifferential of H_{x_1,x_2} at the point $(r_1, r_2) \in \mathbb{R}^2_+$, defined as

$$\partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) := \left\{ (a, b) \in \mathbb{R}^2 \ : \left| \begin{array}{l} \mathsf{H}_{x_1,x_2}(s_1, s_2) - \mathsf{H}_{x_1,x_2}(r_1, r_2) \geq a(s_1 - r_1) + b(s_2 - r_2) \\ \text{for every } (s_1, s_2) \in \mathbb{R}^2_+ \end{array} \right. \right\}.$$

Notice that $\partial\mathsf{H}(\lambda\mathfrak{y}_1, \lambda\mathfrak{y}_2) = \partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2)$ for every $\lambda > 0$. As usual, the proper domain of the subdifferential is denoted by $\mathrm{D}(\partial\mathsf{H}) := \{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[X_1, X_2] \mid \partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \neq \emptyset\}$. It is well known that $\mathrm{D}(\partial\mathsf{H})$ contains the *radial interior*

$$\text{rad-int}\,(\mathrm{D}(\mathsf{H})) := \Big\{ (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[X_1, X_2] : (r_1, r_2) \in \mathrm{int}\,(\mathrm{D}(\mathsf{H}_{x_1,x_2})) \Big\},$$

where the interior of $\mathrm{D}(\mathsf{H}_{x_1,x_2})$ refers to the usual topology of $\mathbb{R}^2$; in particular, if $(\mathfrak{y}_1, \mathfrak{y}_2) \in \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$, then $r_1 > 0, r_2 > 0$.

For every $(\mathfrak{y}_1, \mathfrak{y}_2) \in \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$ we can consider the quantity

$$\mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2) := \sup \Big\{ |a| : (a, b) \in \partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \Big\}; \tag{5.7}$$

notice that

$$\begin{aligned} \mathsf{a}(\lambda\mathfrak{y}_1, \lambda\mathfrak{y}_2) &= \mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2) && \text{for every } \lambda > 0, \\ -\mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|b| \leq ab &\leq \mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|b| && \text{for every } (a, b) \in \partial\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2). \end{aligned}$$

If $\mathfrak{y}_2 = \mathfrak{o}_2$ and $(\mathfrak{y}_1, \mathfrak{o}_2) \in \mathrm{D}(\mathsf{H})$ we also set

$$\mathsf{a}(\mathfrak{y}_1, \mathfrak{o}_2) := \mathsf{H}([x_1, 1], \mathfrak{o}_2) = r_1^{-1}\mathsf{H}(\mathfrak{y}_1, \mathfrak{o}_2) \quad \text{where } \mathfrak{y}_1 = [x_1, r_1].$$

Proposition 5.7. *Let* H *be as in* (5.1)*, let* $\Gamma \subset \mathrm{D}(\mathsf{H})$ *be a* H*-cyclically monotone radial cone, and let* $(\mathfrak{y}_1, \mathfrak{y}_2) = ([x_1, r_1], [x_2, r_2]) \in \Gamma_{\mathfrak{o}} = \Gamma \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}$*, such that* $(\mathfrak{y}_1, \mathfrak{y}_2) \in \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$ *(and therefore* $r_i > 0$*) or* $\mathfrak{y}_2 = \mathfrak{o}_2$ *(and therefore* $r_1 > 0$*). Let* $\mathfrak{y}_1' = [x_1, r_1'] \in \mathfrak{C}[\mathsf{X}_1]$ *with* $r_1' > 0$*; then:*

(1) for every $\varepsilon > 0$ *there exists a* $\mathsf{A}_{\mathsf{H},\Gamma_{\mathfrak{o}}}$*-walk* $P_\varepsilon \in \mathcal{P}(\mathfrak{y}_1' \to \mathfrak{y}_1)$ *with*

$$\Theta(P_\varepsilon) \leq \mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|r_1' - r_1| + \varepsilon; \tag{5.8}$$

(2) if $P \in \mathcal{P}(\mathfrak{y}_1 \to \mathfrak{y}_1')$ is any $\mathcal{A}_{\mathsf{H},\Gamma}$-walk, then

$$\Theta(P) \geq -\mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|r_1' - r_1|. \tag{5.9}$$

Proof. Notice that $\mathfrak{y}_1, \mathfrak{y}_1' \in \mathfrak{C}[\mathsf{X}_1] \setminus \{\mathfrak{o}_1\}$; if $\mathfrak{y}_2 \neq \mathfrak{o}_2$ we set $\bar{\mathsf{H}} := \mathsf{H}_{x_1,x_2}$, $q := \frac{r_2}{r_1} > 0$;
Claim (1). The case $\mathfrak{y}_2 = \mathfrak{o}_2$ is simple: since $(\mathfrak{y}_1, \mathfrak{o}_2) \in \Gamma \subset \mathrm{D}(\mathsf{H})$ we have $\mathsf{a}(\mathfrak{y}_1, \mathfrak{o}_2) = r_1^{-1}\mathsf{H}(\mathfrak{y}_1, \mathfrak{o}_2) < \infty$; since Γ is a radial cone and $(\mathfrak{y}_1', \mathfrak{o}_2) = r_1'/r_1(\mathfrak{y}_1, \mathfrak{o}_2)$ we have $(\mathfrak{y}_1', \mathfrak{o}_2) \in \Gamma$ with $\mathsf{H}(\mathfrak{y}_1', \mathfrak{o}_2) = r_1'\mathsf{a}(\mathfrak{y}_1, \mathfrak{o}_2)$. The walk $P = (\mathfrak{y}_1', \mathfrak{o}_2, \mathfrak{y}_1)$ belongs to $(V_{\Gamma_\mathfrak{o}}, \overline{\mathcal{A}}_{\mathsf{H},\Gamma_\mathfrak{o}})$ and

$$\Theta(P) = \mathsf{H}(\mathfrak{y}_1, \mathfrak{o}_2) - \mathsf{H}(\mathfrak{y}_1', \mathfrak{o}_2) = (r_1 - r_1')\mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2) \leq \mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|r_1 - r_1'|.$$

Let us now consider the case $\mathfrak{y}_2 \neq \mathfrak{o}_2$. For every $n \in \mathbb{N}$, we define $\vartheta_n := \left(\frac{r_1}{r_1'}\right)^{1/n}$ and we consider the points

$$\mathfrak{y}_1^k := \left[x_1, r_1' (\vartheta_n)^k\right], \quad \mathfrak{y}_2^k := \left[x_2, r_1' q (\vartheta_n)^k\right], \quad k = 0, \dots, n-1, \quad \mathfrak{y}_1^n := [x_1, r_1] = \mathfrak{y}_1,$$

inducing a walk P_n connecting $\mathfrak{y}_1'$ to $\mathfrak{y}_1$ according to (5.3), since

$$\mathfrak{y}_1^0 = [x_1, r_1'] = \mathfrak{y}_1', \quad (\mathfrak{y}_1^k, \mathfrak{y}_2^k) = (\rho_{n,k}[x_1, r_1], \rho_{n,k}[x_2, r_2]) \in \Gamma, \quad k = 0, \dots, n-1,$$

where $\rho_{n,k} = \frac{r_1'}{r_1} (\vartheta_n)^k$. We have

$$\begin{aligned}
\Theta(P_n) &= \sum_{k=0}^{n-1} \left(\mathsf{H}(\mathfrak{y}_1^{k+1}, \mathfrak{y}_2^k) - \mathsf{H}(\mathfrak{y}_1^k, \mathfrak{y}_2^k)\right) \\
&= \sum_{k=0}^{n-1} \left(\bar{\mathsf{H}}\left(r_1' (\vartheta_n)^{k+1}, r_1' q (\vartheta_n)^k\right) - \bar{\mathsf{H}}\left(r_1' (\vartheta_n)^k, r_1' q (\vartheta_n)^k\right)\right) \\
&= \sum_{k=0}^{n-1} r_1' (\vartheta_n)^k \left(\bar{\mathsf{H}}(\vartheta_n, q) - \bar{\mathsf{H}}(1, q)\right) \\
&\leq \sum_{k=1}^{n} r_1' (\vartheta_n)^k (1 - \vartheta_n) a_n = r_1' a_n (1 - (\vartheta_n)^n) = a_n(r_1' - r_1)
\end{aligned}$$

where $(a_n, b_n) \in \partial \bar{H}(\theta_n, q) \neq \emptyset$ for n sufficiently large, since $(\theta_n, q) \to (1, q)$ as $n \to +\infty$ and $\partial\bar{\mathsf{H}}(1, q) = \partial\bar{\mathsf{H}}(r_1, r_2)$ with $(r_1, r_2) \in \mathrm{int}\left(\mathrm{D}(\partial\bar{\mathsf{H}})\right) = \mathrm{int}\left(\mathrm{D}(\bar{\mathsf{H}})\right)$. This proves that P_n is a walk in $\mathcal{A}_{\mathsf{H},\Gamma_\mathfrak{o}}$ and, passing to the lim sup as $n \to \infty$ we have

$$\limsup_{n \to \infty} a_n(r_1' - r_1) \leq \mathsf{a}(\mathfrak{y}_1, \mathfrak{y}_2)|r_1' - r_1|$$

which yields (5.8).
Claim (2). If $P \in \mathcal{P}(\mathfrak{y}_1 \to \mathfrak{y}_1')$ is any $\mathcal{A}_{\mathsf{H},\Gamma}$-walk connecting $\mathfrak{y}_1$ to $\mathfrak{y}_1'$, we can consider the cycle $\tilde{P}_\varepsilon = P + P_\varepsilon$ starting and ending in $\mathfrak{y}_1$ obtained by joining P with the walk P_ε given by the previous claim. (5.5) and the H-cyclycal monotonicity of Γ yield

$$0 \leq \Theta(\tilde{P}_\varepsilon) = \Theta(P) + \Theta(P_\varepsilon)$$

which immediately gives (5.9). □

We can now give sufficient conditions for the H-connectedness of a set $\Gamma \subset \mathrm{D}(\mathsf{H})$, with $\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty]$ as in (5.1). We set

$$\mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{y}_2) := \inf\left\{\mathsf{H}(\lambda_1\mathfrak{y}_1, \lambda_2\mathfrak{y}_2) : \lambda_1, \lambda_2 > 0\right\}, \quad (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2].$$

Notice that we have

$$\text{if } \mathfrak{y}_i = [x_i, r_i] \neq \mathfrak{o}_i \text{ then } \mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{y}_2) < \infty \quad \Leftrightarrow \quad \exists\, s_i > 0 : \mathsf{H}([x_1, s_1], [x_2, s_2]) < \infty \tag{5.10}$$

whereas

$$\begin{aligned} \mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{o}_2) < \infty \quad &\Leftrightarrow \quad \mathsf{H}(\mathfrak{y}_1, \mathfrak{o}_2) < \infty, \\ \mathsf{H}_{\inf}(\mathfrak{o}_1, \mathfrak{y}_2) < \infty \quad &\Leftrightarrow \quad \mathsf{H}(\mathfrak{o}_1, \mathfrak{y}_2) < \infty. \end{aligned} \tag{5.11}$$

We can also easily see that $\mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{y}_2) < \infty$ if and only if it is possibile to connect $\mathfrak{y}_1$ (resp. $\mathfrak{y}_2$) to a multiple of $\mathfrak{y}_2$ (resp. of $\mathfrak{y}_1$) with finite H-cost:

$$\begin{aligned} \mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{y}_2) < \infty \quad &\Leftrightarrow \quad \exists\, \lambda_2 > 0 : \mathsf{H}(\mathfrak{y}_1, \lambda_2 \mathfrak{y}_2) < \infty \\ &\Leftrightarrow \quad \exists\, \lambda_1 > 0 : \mathsf{H}(\lambda_1 \mathfrak{y}_1, \mathfrak{y}_2) < \infty. \end{aligned}$$

Theorem 5.8 (Sufficient conditions for H-connectedness)**.** *If* H *is as in* (5.1) *and* $\Gamma \subset \mathrm{D}(\mathsf{H})$ *is a* H*-cyclically monotone radial cone satisfying at least one of the following conditions:*

(1) Γ *is* $\mathsf{H}_{\inf}$*-connected and* $\Gamma \setminus \Big(\{\mathfrak{o}_1\} \times \mathfrak{C}[\mathsf{X}_2] \cup \mathfrak{C}[\mathsf{X}_1] \times \{\mathfrak{o}_2\}\Big) \subset \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$,

(2) $\mathsf{H}_{\inf}$ *is finite on* $\Gamma_1 \times \Gamma_2$ *(recall* (5.2)*) and*

$$\Gamma \cap \text{rad-int}\,(\mathrm{D}(\mathsf{H})) \neq \emptyset, \tag{5.12}$$

(3) $\mathsf{H}_{\inf}$ *is finite on* $\Gamma_1 \times \Gamma_2$ *and*

$$\text{there exist points } \bar{\mathfrak{y}}_i \in \mathfrak{C}[\mathsf{X}_i] \setminus \{\mathfrak{o}_i\} \text{ such that } (\bar{\mathfrak{y}}_1, \mathfrak{o}_2) \in \Gamma \text{ and } (\mathfrak{o}_1, \bar{\mathfrak{y}}_2) \in \Gamma, \tag{5.13}$$

then Γ *is* H*-connected.*

Proof. We divide the proof in claims.

Claim (1). It is sufficient to prove that if a pair of points $(\mathfrak{y}_2, \mathfrak{y}_1) \in \Gamma_2 \times \Gamma_1$ satisfies $\mathsf{H}_{\inf}(\mathfrak{y}_1, \mathfrak{y}_2) < +\infty$ then we can find a $\mathcal{A}_{\mathsf{H},\Gamma}$ walk $P \in \mathcal{P}(\mathfrak{y}_2 \to \mathfrak{y}_1)$. If $\mathfrak{y}_1 = \mathfrak{o}_1$ or $\mathfrak{y}_2 = \mathfrak{o}_2$ we have $\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) < +\infty$ thanks to (5.11), so that the arc $(\mathfrak{y}_2, \mathfrak{y}_1)$ belongs to $\mathcal{A}_{\mathsf{H}}$.

We can thus suppose that $\mathfrak{y}_i = [x_i, r_i]$ with $r_i > 0$. Recalling (5.10) and using the 1-homogeneity of H we can find $r_1' > 0$ such that, setting $\mathfrak{y}_1' = [x_1, r_1']$, we get $\mathsf{H}(\mathfrak{y}_1', \mathfrak{y}_2) < +\infty$. On the other hand, since $\mathfrak{y}_1 \in \Gamma_1$ we can also find $\mathfrak{y}_2'$ such that $(\mathfrak{y}_1, \mathfrak{y}_2') \in \Gamma$.

If $\mathfrak{y}_2' = \mathfrak{o}_2$ then $(\mathfrak{y}_1', \mathfrak{o}_2) \in \Gamma$ as well, since Γ is a radial cone, and $\mathsf{H}(\mathfrak{y}_1, \mathfrak{o}_2) < +\infty$ since $\Gamma \subset \mathrm{D}(\mathsf{H})$. We conclude that $P = (\mathfrak{y}_2, \mathfrak{y}_1', \mathfrak{o}_2, \mathfrak{y}_1)$ is an admissible walk in $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$.

If $\mathfrak{y}_2' \neq \mathfrak{o}_2$ then $(\mathfrak{y}_1, \mathfrak{y}_2') \in \Gamma \cap \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$ and we can apply the first claim of Proposition 5.7 to find a walk $P \in \mathcal{P}(\mathfrak{y}_1' \to \mathfrak{y}_1)$ in $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$. Joining $(\mathfrak{y}_2, \mathfrak{y}_1')$ with P we obtain a connection from $\mathfrak{y}_2$ to $\mathfrak{y}_1$.

Claim (2). Let us pick $(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2) \in \Gamma \cap \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$, $\bar{\mathfrak{y}}_i = [\bar{x}_i, \bar{r}_i]$ with $\bar{r}_i > 0$ and $\bar{q} := \bar{r}_2/\bar{r}_1$. It is sufficient to show that for every $\mathfrak{y}_1 \in \Gamma_1$ we can find a walk $P'' \in \mathcal{P}(\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1)$ and a walk $P' \in \mathcal{P}(\mathfrak{y}_1 \to \bar{\mathfrak{y}}_1)$ in $(V_\Gamma, \mathcal{A}_{\mathsf{H},\Gamma})$.

Since $\mathsf{H}_{\inf}(\mathfrak{y}_1, \bar{\mathfrak{y}}_2) < \infty$, we can find $\mathfrak{y}_2 = [\bar{x}_2, r_2]$ with $r_2 > 0$ such that $\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) < +\infty$, so that $(\mathfrak{y}_2, \mathfrak{y}_1) \in \mathcal{A}_{\mathsf{H},\Gamma}$. Since Γ is a radial cone, setting $\mathfrak{y}_1' := [\bar{x}_1, r_2/\bar{q}]$ we have $(\mathfrak{y}_1', \mathfrak{y}_2) \in \Gamma$; moreover $(r_2/\bar{q}, r_2) = r_2/\bar{r}_2(\bar{r}_1, \bar{r}_2) \in \mathrm{int}\,(\mathrm{D}(\mathsf{H}_{\bar{x}_1, \bar{x}_2}))$ so that $(\mathfrak{y}_1', \mathfrak{y}_2) \in \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$. By Claim (1) of Proposition 5.7 we can eventually find a walk $\tilde{P} \in \mathcal{P}(\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1')$ in $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$. Thus joining $\tilde{P}$ with $(\mathfrak{y}_1', \mathfrak{y}_2, \mathfrak{y}_1)$ we obtain a walk P'' connecting $\bar{\mathfrak{y}}_1$ to $\mathfrak{y}_1$.

In order to find the second walk connecting $\mathfrak{y}_1$ to $\bar{\mathfrak{y}}_1$, we first select $\mathfrak{y}_2 = [x_2, r_2]$ so that $(\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$ (using the fact that $\mathfrak{y}_1 \in \Gamma_1$). Since $\mathsf{H}_{\inf}(\bar{\mathfrak{y}}_1, \mathfrak{y}_2) < \infty$ we find $\lambda_1 > 0$ such that $\mathsf{H}(\lambda_1 \bar{\mathfrak{y}}_1, \mathfrak{y}_2) < \infty$. By Proposition 5.7 we can eventually join $\lambda_1 \bar{\mathfrak{y}}_1$ to $\bar{\mathfrak{y}}_1$ by a walk $\tilde{P}$ in $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$. The walk P' obtained by joining $(\mathfrak{y}_1, \mathfrak{y}_2, \lambda_1 \bar{\mathfrak{y}}_1)$ with $\tilde{P}$ provides the requested connection from $\mathfrak{y}_1$ to $\bar{\mathfrak{y}}_1$ in $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$.

Claim (3). The proof is quite similar to the previous claim. We show that, for every $\mathfrak{y}_1 \in \Gamma_1$, we can connect $\mathfrak{y}_1$ to $\mathfrak{o}_1$ and $\mathfrak{o}_1$ to $\mathfrak{y}_1$ with a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk. Let $\mathfrak{y}_1 \in \Gamma_1$ be fixed. Since $\mathsf{H}_{\inf}(\mathfrak{y}_1, \bar{\mathfrak{y}}_2) < \infty$, we can find $\mathfrak{y}_2 = \lambda_2 \bar{\mathfrak{y}}_2$, $\lambda_2 > 0$, such that $\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) < +\infty$, so that $(\mathfrak{y}_2, \mathfrak{y}_1) \in \mathcal{A}_{\mathsf{H},\Gamma}$. The $\mathcal{A}_{\mathsf{H},\Gamma}$-walk $(\mathfrak{o}_1, \mathfrak{y}_2, \mathfrak{y}_1)$ connects $\mathfrak{o}_1$ to $\mathfrak{y}_1$.

In order to find the second walk connecting $\mathfrak{y}_1$ to $\mathfrak{o}_1$ we first select $\mathfrak{y}_2 = [x_2, r_2]$ so that $(\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$ (using the fact that $\mathfrak{y}_1 \in \Gamma_1$). Since $\mathsf{H}_{\inf}(\bar{\mathfrak{y}}_1, \mathfrak{y}_2) < \infty$ we find $\lambda > 0$ such that $\mathsf{H}(\lambda_1\bar{\mathfrak{y}}_1, \mathfrak{y}_2) < \infty$. Since Γ is a radial cone, $(\lambda\bar{\mathfrak{y}}_1, \mathfrak{o}_2) \in \Gamma$ so that the walk $P' = (\mathfrak{y}_1, \mathfrak{y}_2, \lambda\bar{\mathfrak{y}}_1, \mathfrak{o}_2, \mathfrak{o}_1)$ connects $\mathfrak{y}_1$ to $\mathfrak{o}_1$ and belongs to $(V_\Gamma, \overline{\mathcal{A}}_{\mathsf{H},\Gamma})$. □

Remark 5.9. If there is a common vector $(r_1, r_2) \neq (0,0)$ in the intersection $\bigcap_{(x_1,x_2)\in\mathsf{X}_1\times\mathsf{X}_2} \mathrm{D}(\mathsf{H}_{x_1,x_2})$ of all the domains of the functions H_{x_1,x_2}, then $\mathsf{H}_{\inf}$ is finite in $(\mathfrak{C}[\mathsf{X}_1] \setminus \{\mathfrak{o}_1\}) \times (\mathfrak{C}[\mathsf{X}_2] \setminus \{\mathfrak{o}_2\})$. Then, if $\Gamma \subset (\mathfrak{C}[\mathsf{X}_1] \setminus \{\mathfrak{o}_1\}) \times (\mathfrak{C}[\mathsf{X}_2] \setminus \{\mathfrak{o}_2\})$ and $\Gamma \subset \text{rad-int}\,(\mathrm{D}(\mathsf{H}))$, condition (1) of Theorem 5.8 is satisfied and Γ is H-connected. If H is finite then it is immediate to check that any $\Gamma \subset \mathfrak{C}[\mathsf{X}_1] \times \mathfrak{C}[\mathsf{X}_2]$ is H-connected.

5.2. **Sufficient conditions for optimality.** The next result provides sufficient conditions in order to guarantee the optimality of a 1-homogeneous coupling concentrated on a H-cyclically monotone radial cone Γ. The main assumptions concern H-connectedness of Γ (see the previous Theorem 5.8 for simple conditions guaranteeing this property), the fact that Γ has nonempty intersection with the radial interior rad-int $(\mathrm{D}(\mathsf{H}))$ of $\mathrm{D}(\mathsf{H})$ (or, alternatively, that (5.13) holds) and an integrability condition (5.16) which minimic the usual condition stated in the framework of balanced OT. To avoid trivial cases, we will assume that the measures μ_i have strictly positive mass. We also mention that the result below provides the existence of a relaxed solution for the dual problem, i.e. a pair of optimal potentials that are only Borel measurable.

In order to treat potentials that may take infinite values, we use the notation

$$\zeta_1(x_1)r_1 +_o \zeta_2(x_2)r_2 := \lim_{n\to+\infty} (-n \vee \zeta_1(x_1)r_1 \wedge n) + (-n \vee \zeta_1(x_1)r_2 \wedge n)$$

for functions $\zeta_i : \mathsf{X}_i \to \mathbb{R} \cup \{\pm\infty\}$, with the convention that $\pm\infty \cdot 0 = 0$. In particular $\zeta_1(x_1)r_1 +_o \zeta_2(x_2)r_2 = 0$ in case $\zeta_1(x_1) = \pm\infty$, $\zeta_2(x_2) = \mp\infty$ and $r_1, r_2 > 0$.

Theorem 5.10 (Sufficiency of H-cyclical monotonicity)**.** *Let H be as in (5.1), let $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ with $\mu_i(\mathsf{X}_i) > 0$ for $i = 1, 2$, and let $\boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1, \mu_2)$ be an admissible 1-homogeneous coupling concentrated on a σ-compact* radial cone *$\Gamma \subset \mathrm{D}(\mathsf{H})$ such that Γ is H-cyclically monotone and H-connected. We set*

$$\Gamma_i := \pi^i(\Gamma), \quad S_i := \mathsf{x}(\Gamma_i \setminus \{\mathfrak{o}_i\}).$$

(1) If Γ satisfies one of the two conditions (5.12) or (5.13) then the measures μ_i are concentrated on the sets S_i (in particular $S_i \neq \emptyset$) and there exist Borel functions $\varphi_i : \mathsf{X}_i \to \mathbb{R} \cup \{\pm\infty\}$ which are real valued on S_i and such that

$$\varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \textit{for every } x_i \in \mathsf{X}_i,\ r_i \ge 0, \tag{5.14}$$

$$\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2 = \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \textit{if } ([x_1, r_1], [x_2, r_2]) \in \Gamma. \tag{5.15}$$

(2) If moreover there exist nonnegative Borel functions $\varrho_i \in L^1_+(\mathsf{X}_i, \mu_i)$ such that

$$\begin{aligned} &\mu_1\left(\left\{x_1 \in \mathsf{X}_1 \mid \int_{\mathsf{X}_2} \mathsf{H}([x_1, \varrho_2(x_2)], [x_2, 1])\,\mathrm{d}\mu_2(x_2) < +\infty\right\}\right) > 0, \\ &\mu_2\left(\left\{x_2 \in \mathsf{X}_2 \mid \int_{\mathsf{X}_1} \mathsf{H}([x_1, 1], [x_2, \varrho_1(x_1)])\,\mathrm{d}\mu_1(x_1) < +\infty\right\}\right) > 0, \end{aligned} \tag{5.16}$$

then $\boldsymbol{\alpha}$ is optimal, $\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} < +\infty$, the functions φ_i belong to $\mathcal{L}^1(\mathsf{X}_i, \mu_i)$ and provide a relaxed solution for the dual problem (3.31) i.e.

$$\varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \textit{for every } x_i \in \mathsf{X}_i,\ r_i \ge 0,$$

$$\int_{\mathsf{X}_1} \varphi_1\,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\,\mathrm{d}\mu_2 = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2).$$

Proof. We set

$$\Gamma_{\mathfrak{o}} := \Gamma \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}, \quad \mathfrak{C}_{\mathfrak{o}}[\mathsf{X}_1, \mathsf{X}_2] := \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}.$$

We first construct a candidate function $\Phi : \mathfrak{C}[\mathsf{X}_1] \to \mathbb{R}$ inducing a potential ζ_1 via the identity $\Phi([x_1, r_1]) = \zeta_1(x_1)r_1$.

Step 1 (Definition of Φ in Γ_1). We pick a point $(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2) = ([\bar{x}_1, \bar{r}_1], [\bar{x}_2, \bar{r}_2]) \in \Gamma_{\mathfrak{o}}$ such that $(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2) \in$ rad-int $(\mathrm{D}(\mathsf{H}))$ (condition (5.12)) or $\bar{\mathfrak{y}}_2 = \mathfrak{o}_2$ (condition (5.13)). We set (see (5.7))

$$\bar{\mathsf{a}} := \mathsf{a}(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2).$$

We define $\Phi : \mathfrak{C}[\mathsf{X}_1] \to [-\infty, +\infty]$ as

$$\Phi(\mathfrak{y}_1) := \inf\Big\{\Theta(P) + \bar{\mathsf{a}}|r_1 - \lambda\bar{r}_1| : P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1), \lambda > 0\Big\}, \quad \mathfrak{y}_1 = [x_1, r_1] \in \mathfrak{C}[\mathsf{X}_1]. \tag{5.17}$$

Since Γ is H-connected and $\lambda\bar{\mathfrak{y}}_1 \in \Gamma_1$, for every $\mathfrak{y}_1 \in \Gamma_1$ we can find $P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1)$ with finite cost $\Theta(P)$ so that

$$\Phi(\mathfrak{y}_1) < +\infty \quad \text{for every } \mathfrak{y}_1 \in \Gamma_1.$$

In particular, choosing $\mathfrak{y}_1 := [\bar{x}_1, r] = \alpha\bar{\mathfrak{y}}_1$, $r > 0$, with $\alpha := r/\bar{r}_1$, $\lambda = \alpha$, and $P = (\alpha\bar{\mathfrak{y}}_1, \alpha\bar{\mathfrak{y}}_2, \alpha\bar{\mathfrak{y}}_1)$ we have $\Theta(P) = 0$ and we immediately get

$$\Phi([\bar{x}_1, r]) \le 0 \quad \text{for every } r > 0. \tag{5.18}$$

By Proposition 5.7, every walk $P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \bar{\mathfrak{y}}_1)$, $\lambda > 0$ satisfies

$$\Theta(P) \ge -\mathsf{a}(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2)|\lambda\bar{r}_1 - \bar{r}_1| \quad \text{for every } \lambda > 0,$$

so that (5.17) yields $\Phi(\bar{\mathfrak{y}}_1) \ge 0$; combining with (5.18) we obtain

$$\Phi(\bar{\mathfrak{y}}_1) = 0.$$

Since for every $\mathfrak{y}_1 \in \mathfrak{C}[\mathsf{X}_1]$, $\lambda > 0$, $P = (\lambda\bar{\mathfrak{y}}_1, \mathfrak{y}^1, \cdots, \mathfrak{y}^{N-1}, \mathfrak{y}_1) \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1)$ and every $\alpha > 0$ we have $P_\alpha := (\alpha\lambda\bar{\mathfrak{y}}_1, \alpha\mathfrak{y}^1, \cdots, \alpha\mathfrak{y}^{N-1}, \alpha\mathfrak{y}_1) \in \mathcal{P}(\alpha\lambda\bar{\mathfrak{y}}_1, \alpha\mathfrak{y}_1)$ with

$$\Theta(P_\alpha) + \bar{\mathsf{a}}|\alpha r_1 - \alpha\lambda\bar{r}_1| = \alpha\Big(\Theta(P) + \bar{\mathsf{a}}|r_1 - \lambda\bar{r}_1|\Big)$$

(since Γ is a radial cone and H is positively 1-homogeneous) we get that

$$\Phi(\alpha\mathfrak{y}_1) = \alpha\Phi(\mathfrak{y}_1) \quad \text{for every } \mathfrak{y}_1 \in \mathfrak{C}[\mathsf{X}_1],\ \alpha > 0.$$

If now $\mathfrak{y}_1 \in \Gamma_1$ and $\lambda > 0$, we use the H-connectedness of Γ to find a $\mathcal{A}_{\mathsf{H},\Gamma}$-walk $\bar{P} \in \mathcal{P}(\mathfrak{y}_1 \to \bar{\mathfrak{y}}_1)$. If $P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1)$, $P + \bar{P}$ is a walk in $\mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \bar{\mathfrak{y}}_1)$ so that (5.9) and the identity $\mathsf{a}(\lambda\bar{\mathfrak{y}}_1, \lambda\bar{\mathfrak{y}}_2) = \mathsf{a}(\bar{\mathfrak{y}}_1, \bar{\mathfrak{y}}_2) = \bar{\mathsf{a}}$ yield

$$-\bar{\mathsf{a}}|\lambda - 1|\bar{r}_1 \le \Theta(P + \bar{P}) = \Theta(P) + \Theta(\bar{P})$$

so that

$$\Phi(\mathfrak{y}_1) \ge -\Theta(\bar{P}) - \mathsf{a}|r_1 - \bar{r}_1| > -\infty.$$

Arguing as in [AGS08, Step 1 of Therem 6.14] and using the fact that Γ is a σ-compact set, we can see that Φ is a Borel function.

Let us now consider $\mathfrak{y}_1, \mathfrak{y}_1' \in \Gamma_1$, $\mathfrak{y}_2 \in \mathfrak{C}[\mathsf{X}_2]$ such that $(\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$, and any walk $P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1 \to \mathfrak{y}_1)$; if we consider the walk $P' = P + (\mathfrak{y}_1, \mathfrak{y}_2, \mathfrak{y}_1')$, P' is an admissible walk joining $\lambda\bar{\mathfrak{y}}_1$ to $\mathfrak{y}_1'$ so that

$$\Phi(\mathfrak{y}_1') \le \mathsf{H}(\mathfrak{y}_1', \mathfrak{y}_2) - \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) + \Theta(P) + \bar{\mathsf{a}}|r_1 - \lambda\bar{r}_1|.$$

Passing then to the infimum among all the walks $P \in \mathcal{P}(\lambda\bar{\mathfrak{y}}_1, \mathfrak{y}_1)$, $\lambda > 0$, we deduce that

$$\Phi(\mathfrak{y}_1') \le \Phi(\mathfrak{y}_1) + \mathsf{H}(\mathfrak{y}_1', \mathfrak{y}_2) - \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \quad \text{for every } \mathfrak{y}_1' \in \Gamma_1,\ (\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma.$$

Restricting Φ to Γ_1 (notice that Γ_i is Borel since Γ is σ-compact) we have proven that there exists a Borel function $\Phi : \Gamma_1 \to \mathbb{R}$ such that

$$\Phi([\bar{x}_1, r_1]) = 0 \qquad \text{for every } r_1 > 0, \tag{5.19}$$

$$\Phi(\lambda\mathfrak{y}_1) = \lambda\Phi(\mathfrak{y}_1) \qquad \text{for every } \mathfrak{y}_1 \in \Gamma_1,\ \lambda > 0, \tag{5.20}$$

$$\Phi(\mathfrak{y}_1') \le \Phi(\mathfrak{y}_1) + \mathsf{H}(\mathfrak{y}_1', \mathfrak{y}_2) - \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \qquad \text{for every } \mathfrak{y}_1' \in \Gamma_1,\ (\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma. \tag{5.21}$$

Step 2 (Definition of Ψ). We define $\Psi : \Gamma_2 \to [-\infty, +\infty]$ as

$$\Psi(\mathfrak{y}_2) := \inf_{\mathfrak{y}_1 \in \Gamma_1} \{\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) - \Phi(\mathfrak{y}_1)\}, \quad \mathfrak{y}_2 \in \Gamma_2.$$

It is clear from the definition that

$$\Phi(\mathfrak{y}_1) + \Psi(\mathfrak{y}_2) \le \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \quad \text{for every } (\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma_1 \times \Gamma_2.$$

Since Φ is real valued on Γ_1 and for every $\mathfrak{y}_2 \in \Gamma_2$ there exists $\mathfrak{y}_1 \in \Gamma_1$ such that $(\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$ so that $\mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) < \infty$, we immediately get $\Psi(\mathfrak{y}_2) < \infty$ for every $\mathfrak{y}_2 \in \Gamma_2$. By the definition of Ψ, the fact that Γ is a radial cone, the radial 1-homogeneity of H and of Φ given by (5.20), it also easily follows that

$$\Psi(\lambda\mathfrak{y}_2) = \lambda\Psi(\mathfrak{y}_2) \quad \text{for every } \mathfrak{y}_2 \in \Gamma_2,\ \lambda > 0.$$

The inequality in (5.21) immediately yields

$$\Psi(\mathfrak{y}_2) + \Phi(\mathfrak{y}_1) = \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \quad \text{for every } (\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma$$

and also gives that $\Psi(\mathfrak{y}_2) \in \mathbb{R}$ for every $\mathfrak{y}_2 \in \Gamma_2$. The Borel measurability of Ψ can be checked as in [AGS08, Step 2 of Theorem 6.14].

Summarizing the second step, we have proven that there exists a Borel function $\Psi : \Gamma_2 \to \mathbb{R}$ such that

$$\Phi(\mathfrak{y}_1) + \Psi(\mathfrak{y}_2) \le \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \qquad \text{on } \Gamma_1 \times \Gamma_2, \tag{5.22}$$

$$\Psi(\lambda\mathfrak{y}_2) = \lambda\Psi(\mathfrak{y}_2) \qquad \text{for every } \mathfrak{y}_2 \in \Gamma_2,\ \lambda > 0, \tag{5.23}$$

$$\Phi(\mathfrak{y}_1) + \Psi(\mathfrak{y}_2) = \mathsf{H}(\mathfrak{y}_1, \mathfrak{y}_2) \qquad \text{for every } (\mathfrak{y}_1, \mathfrak{y}_2) \in \Gamma. \tag{5.24}$$

Step 3 (Definition of φ_1 and φ_2). First of all notice that S_i is Borel since Γ is σ-compact. The following chain of inequalities shows that μ_i is concentrated on S_i:

$$\begin{aligned}
\mu_i(\mathsf{X}_i \setminus S_i) &= \mathfrak{h}_i^1(\boldsymbol{\alpha})(\mathsf{X}_i \setminus \mathsf{x}_i(\Gamma \setminus \{\mathfrak{y}_i = \mathfrak{o}_i\})) \\
&= (\mathsf{r}_i\boldsymbol{\alpha})\left(\mathsf{x}_i^{-1}(\mathsf{X}_i \setminus \mathsf{x}_i(\Gamma \setminus \{\mathfrak{y}_i = \mathfrak{o}_i\}))\right) \\
&= (\mathsf{r}_i\boldsymbol{\alpha})\left(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus \mathsf{x}_i^{-1}(\mathsf{x}_i(\Gamma \setminus \{\mathfrak{y}_i = \mathfrak{o}_i\}))\right) \\
&\le (\mathsf{r}_i\boldsymbol{\alpha})(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus (\Gamma \setminus \{\mathfrak{y}_i = \mathfrak{o}_i\})) \\
&\le (\mathsf{r}_i\boldsymbol{\alpha})(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus \Gamma) + (\mathsf{r}_i\boldsymbol{\alpha})(\Gamma \cap \{\mathfrak{y}_i = \mathfrak{o}_i\}) \\
&\le \int_{\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus \Gamma} \mathsf{r}_i \,\mathrm{d}\boldsymbol{\alpha} + \int_{\{\mathsf{r}_i = 0\}} \mathsf{r}_i \,\mathrm{d}\boldsymbol{\alpha} \\
&= 0.
\end{aligned}$$

Since we have assumed that the two measures μ_i are positive, this also shows that $S_i \neq \emptyset$, $i = 1, 2$. Let us define $\zeta_i : S_i \to \mathbb{R}$ as

$$\zeta_1(x_1) := \Phi([x_1, 1]), \quad \zeta_2(x_2) := \Psi([x_2, 1]), \quad x_i \in S_i = \mathsf{x}(\Gamma_i \setminus \{\mathfrak{o}_i\}).$$

Notice that ζ_1, ζ_2 are Borel functions. We claim that

$$\zeta_1(x_1)r_1 + \zeta_2(x_2)r_2 \le \mathsf{H}([x_1,r_1],[x_2,r_2]) \quad \text{for every } x_i \in S_i, r_i \ge 0, \tag{5.25}$$
$$\zeta_1(x_1)r_1 + \zeta_2(x_2)r_2 = \mathsf{H}([x_1,r_1],[x_2,r_2]) \quad \text{if } ([x_1,r_1],[x_2,r_2]) \in \Gamma. \tag{5.26}$$

Notice that the product $\zeta(x_i)r_i$ in (5.26) has to be intended equal to 0 in case $x_i \notin S_i$ and $r_i = 0$. We distinguish four cases:

(i) if $r_1 = r_2 = 0$, both sides in (5.25) and (5.26) are equal to 0;
(ii) if $r_1, r_2 \neq 0$, then $\Phi([x_1,r_1]) = r_1\zeta_1(x_1)$ and $\Psi([x_2,r_2]) = r_2\zeta_2(x_2)$ by (5.20) and (5.23), so that (5.25) corresponds to (5.22) and (5.26) corresponds to (5.24);
(iii) if $r_1 > 0 = r_2$, then $[x_2,r_2] = \mathfrak{o}_2$ and $\zeta_2(x_2)r_2 = 0$: for (5.25) we can just pass to the limit as $r_2 \downarrow 0$ in the same inequality, using the fact that $r_2 \mapsto \mathsf{H}([x_1,r_1],[x_2,r_2])$ is continuous at 0. Regarding (5.26), observe that saying that $([x_1,r_1],\mathfrak{o}_2) \in \Gamma$ gives in particular that $\mathfrak{o}_2 \in \Gamma_2$ so that (5.23) forces $\Psi(\mathfrak{o}_2) = 0$. Then (5.26), is exactly (5.24) with $\mathfrak{y}_2 = \mathfrak{o}_2$;
(iv) the case $r_1 = 0 < r_2$ is completely analogous to the previous one also using (5.19).

Now we can use the H-transform (compare also with Definition 4.5) to extend the potentials (ζ_1,ζ_2) to the whole spaces X_i while keeping intact the relations in (5.25) and (5.26): we define $\varphi_i : \mathsf{X}_i \to \mathbb{R} \cup \{\pm\infty\}$ as

$$\varphi_2(x_2) := \inf_{x_1 \in S_1, r_1 \ge 0} \{\mathsf{H}([x_1,r_1],[x_2,1]) - \zeta_1(x_1)r_1\}, \quad x_2 \in \mathsf{X}_2, \tag{5.27}$$
$$\varphi_1(x_1) := \inf_{x_2 \in \mathsf{X}_2, r_2 \ge 0} \{\mathsf{H}([x_1,1],[x_2,r_2]) - \varphi_2(x_2)r_2\}, \quad x_1 \in \mathsf{X}_1, \tag{5.28}$$

with the convention that $\pm\infty \cdot 0 = 0$ and $\mathsf{H}([x_1,1],[x_2,r_2]) - \varphi_2(x_2)r_2 = +\infty$ if $\mathsf{H}([x_1,1],[x_2,r_2]) = +\infty$, $\varphi_2(x_2)r_2 = +\infty$. It is not difficult to check that φ_i are Borel function, $\varphi_i(x_i) \in \mathbb{R}$ if $x_i \in S_i$ and

$$\varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2 \le \mathsf{H}([x_1,r_1],[x_2,r_2]) \quad \text{for every } x_i \in \mathsf{X}_i,\ r_i \ge 0, \tag{5.29}$$
$$\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2 = \mathsf{H}([x_1,r_1],[x_2,r_2]) \quad \text{if } ([x_1,r_1],[x_2,r_2]) \in \Gamma. \tag{5.30}$$

Notice that on Γ there is no need to use $+_o$ since either $x_i \in S_i$ (hence $\varphi(x_i) \in \mathbb{R}$) or $r_i = 0$ (hence $\varphi(x_i)r_i = 0$).

Step 4 (Conclusion). Since μ_1 is concentrated on S_1, using (5.16), we can find some $x_1 \in S_1$ such that $\int_{S_2} \mathsf{H}([x_1,\varrho_2(x_2)x],[x_2,1])\,\mathrm{d}\mu_2(x_2) < +\infty$; by (5.29) we get that

$$\varphi_1(x_1)\varrho_2(x_2) + \varphi_2(x_2) \le \mathsf{H}([x_1,\varrho_2(x_2)],[x_2,1]) \quad \text{for every } x_2 \in S_2$$

so that

$$\varphi_2^+(x_2) \le \mathsf{H}([x_1,\varrho_2(x_2)],[x_2,1]) + \varphi_1(x_1)^-\varrho_2(x_2) \quad \text{for every } x_2 \in S_2,$$

where we denoted by u^+ and u^- the positive and negative part respectively of a real number u. This gives that $\varphi_2^+ \in \mathcal{L}^1(\mathsf{X}_2,\mu_2)$. The argument for φ_1 is the same. We can thus conclude that

$$\begin{aligned}\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha} &= \int_\Gamma \mathsf{H}\,\mathrm{d}\boldsymbol{\alpha}\\ &= \int_\Gamma (\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2)\,\mathrm{d}\boldsymbol{\alpha}([x_1,r_1],[x_2,r_2])\\ &= \int_{\mathsf{X}_1} \varphi_1\,\mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\,\mathrm{d}\mu_2,\end{aligned}$$

showing that $\varphi_i \in \mathcal{L}^1(\mathsf{X}_i,\mu_i)$. Moreover, if $\tilde{\boldsymbol{\alpha}} \in \mathfrak{H}^1(\mu_1,\mu_2)$, setting $\tilde\alpha_i := \pi^i_\sharp\tilde{\boldsymbol{\alpha}}$, we have

$$\int_{\mathfrak{C}[\mathsf{X}_i]} (\varphi_i \mathsf{r})^\pm\,\mathrm{d}\tilde\alpha_i = \int_{\mathfrak{C}[\mathsf{X}_i]} \varphi_i^\pm \mathsf{r}\,\mathrm{d}\tilde\alpha_i = \int_{\mathsf{X}_i} \varphi_i^\pm\,\mathrm{d}\mu_i \in \mathbb{R} \Rightarrow \int_{\mathfrak{C}[\mathsf{X}_i]} \varphi_i \mathsf{r}\,\mathrm{d}\tilde\alpha_i = \int_{\mathsf{X}_i} \varphi_i\,\mathrm{d}\mu_i$$

so that $(\varphi_i \circ \mathsf{x}_i)\mathsf{r}_i \in L^1(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2], \tilde{\boldsymbol{\alpha}})$ and thus $(\varphi_1 \circ \mathsf{x}_1)\mathsf{r}_1 + (\varphi_2 \circ \mathsf{x}_2)\mathsf{r}_2 \in L^1(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2], \tilde{\boldsymbol{\alpha}})$. We deduce that the everywhere defined function

$$([x_1, r_1], [x_2, r_2]) \mapsto \varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2$$

is a representative of the $L^1(\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2], \tilde{\boldsymbol{\alpha}})$-equivalence class $(\varphi_1 \circ \mathsf{x}_1)\mathsf{r}_1 + (\varphi_2 \circ \mathsf{x}_2)\mathsf{r}_2$ (notice that the set where the sum would be undefined has null $\tilde{\boldsymbol{\alpha}}$-measure). Hence, for every $\tilde{\boldsymbol{\alpha}} \in \mathfrak{H}^1(\mu_1, \mu_2)$, we have

$$\begin{aligned}
\int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\, \mathrm{d}\tilde{\boldsymbol{\alpha}} &\geq \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} (\varphi_1(x_1)r_1 +_o \varphi_2(x_2)r_2)\, \mathrm{d}\tilde{\boldsymbol{\alpha}}([x_1, r_1], [x_2, r_2]) \\
&= \int ((\varphi_1 \circ \mathsf{x}_1)\mathsf{r}_1 + (\varphi_2 \circ \mathsf{x}_2)\mathsf{r}_2)\, \mathrm{d}\tilde{\boldsymbol{\alpha}} \\
&= \int_{\mathfrak{C}[\mathsf{X}_1]} \varphi_1 \mathsf{r}\, \mathrm{d}\tilde{\alpha}_1 + \int_{\mathfrak{C}[\mathsf{X}_1]} \varphi_2 \mathsf{r}\, \mathrm{d}\tilde{\alpha}_2 \\
&= \int_{\mathsf{X}_1} \varphi_1\, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\, \mathrm{d}\mu_2 \\
&= \int_\Gamma (\varphi_1(x_1)r_1 + \varphi_2(x_2)r_2)\, \mathrm{d}\boldsymbol{\alpha}([x_1, r_1], [x_2, r_2]) \\
&= \int_{\mathfrak{C}[\mathsf{X}_1,\mathsf{X}_2]} \mathsf{H}\, \mathrm{d}\boldsymbol{\alpha},
\end{aligned}$$

showing both that $\boldsymbol{\alpha}$ is optimal and that

$$\int_{\mathsf{X}_1} \varphi_1\, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2\, \mathrm{d}\mu_2 = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2). \qquad \square$$

Remark 5.11 ($\mathsf{supp}\,(\mu_i)$ and S_i). In the same setting of Theorem 5.10, since μ_i is concentrated on S_i, we clearly have that $\mathsf{supp}\,(\mu_i) \subset \overline{S_i}$. Moreover, we can assume that $S_i \subset \mathsf{supp}\,(\mu_i)$ since, up to defining

$$\Gamma' := \Gamma \cap \hom\,(\mathsf{supp}\,(\boldsymbol{\alpha}))\,,$$

we see that $\boldsymbol{\alpha}$ is still concentrated on Γ' and Γ' is also a σ-compact radial cone contained in $\mathrm{D}(\mathsf{H})$ which is H-cyclically monotone and H-connected. It is also easy to check that $\mathsf{x}(\tilde{\Gamma}_i \setminus \{\mathfrak{o}_i\}) \subset \mathsf{supp}\,(\mu_i)$. In this case we thus have $S_i \subset \mathsf{supp}\,(\mu_i) = \overline{S_i}$.

Remark 5.12 (The case of H finite). In the same setting of Theorem 5.10, in case the cost function H does not attain the value $+\infty$, it is easy to check that the potentials (φ_1, φ_2) defined via the H-transform of the pair (ζ_1, ζ_2) as in (5.27) and (5.28) cannot attain the value $+\infty$ so that the use of $+_o$ is not needed in (5.14).

We conclude this section showing that, under suitable additional assumptions on H, an optimal plan $\boldsymbol{\alpha}$ must me induced by a transport-growth map as in Definition 3.19 in the sense that

$$\boldsymbol{\alpha} = ([\mathrm{id}_{\mathbb{R}^d}, 1], [\mathsf{T}, \mathsf{g}])_\sharp \mu_1, \quad \mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1.$$

The proof is based on the classical approach to the existence of optimal transport maps (see e.g. [AGS08, Theorem 6.2.4]): one shows via the existence of optimal potentials as in Theorem 5.10 that the set on which $\boldsymbol{\alpha}$ is concentrated is a graph. Condition (1) below is used to guarantee that the optimal potential φ_1 is approximately differentiable at a.e. point, while condition (3) is the analogue of the classical twist condition (see e.g. [CD14]). On the other hand condition (2) (cp. with the condition in (4.3)) is used to prevent that $\mathfrak{o}_1$ belongs to the first projection of the support of an optimal plan $\boldsymbol{\alpha}$: this corresponds to the fact that some of the mass of μ_2 does not come from μ_1 but it is created: clearly in this case there cannot be any transport-growth map connecting μ_1 to μ_2.

Theorem 5.13 (Existence of an optimal transport-growth map)**.** *Let* $\mathsf{X}_1 = \mathsf{X}_2 = \mathbb{R}^d$, H *be as in* (5.1) *and finite,* $\mu_i \in \mathcal{M}_+(\mathbb{R}^d)$ *with* $\mu_i(\mathbb{R}^d) > 0$ *for* $i = 1, 2$*. Assume that condition* (5.16) *is satisfied,* $\mu_1 \ll \mathcal{L}^d$ *and that* $\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) < +\infty$*. Assume in addition that:*

(1) H *is partially differentiable w.r.t.* x_1, r_1 *in* $(\mathfrak{C}(\mathbb{R}^d) \setminus \{\mathfrak{o}_1\}) \times (\mathfrak{C}(\mathbb{R}^d) \setminus \{\mathfrak{o}_2\})$ *with continuous partial derivatives;*

(2) for every $x_2 \in \mathsf{supp}\,(\mu_2)$ *there exist* $x_1 \in \mathsf{supp}\,(\mu_1)$ *and* $\varepsilon > 0$ *such that* $\mathrm{B}(x_1, \varepsilon) \times \{x_2\} \subset A_{\mathsf{H}}$ *where*

$$A_{\mathsf{H}} := \left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \partial_{r_1} \mathsf{H}(x, 0; y, 1) = \lim_{r_1 \downarrow 0} \frac{\mathsf{H}([x, r_1], [y, 1]) - \mathsf{H}(\mathfrak{o}_1, [y, 1])}{r_1} = -\infty \right\};$$

(3) for every $x_1 \in \mathbb{R}^d$ *the map*

$$[y, q] \mapsto \begin{pmatrix} \partial_{x_1} \mathsf{H}([x_1, 1], [y, q]) \\ \partial_{r_1} \mathsf{H}([x_1, 1], [y, q]) \end{pmatrix}$$

is injective.

Then there exist Borel maps $(\mathsf{T}, \mathsf{g}) : \mathbb{R}^d \to \mathbb{R}^d \times [0, +\infty)$ *with* $\mathsf{g} \in L^1(\mathsf{X}_1, \mu_1)$ *such that*

$$\mu_2 = (\mathsf{T}, \mathsf{g})_\star \mu_1 = \mathsf{T}_\sharp(\mathsf{g}\mu_1), \qquad \int_{\mathbb{R}^d} \mathsf{H}([x_1, 1], [\mathsf{T}(x_1), \mathsf{g}(x_1)])\,\mathrm{d}\mu_1(x_1) = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2)$$

given by the formula

$$(\mathsf{T}, \mathsf{g})(x_1) = (\partial_{x_1, r_1} \mathsf{H})^{-1}(\tilde{\nabla}\varphi_1(x_1), \varphi_1(x_1)) \quad \textit{for } \mu_1\textit{-a.e. } x_1 \in \mathbb{R}^d,$$

where $(\partial_{x_1, r_1} \mathsf{H})^{-1}$ *denotes the inverse of the map in item 3 and* $\tilde{\nabla}$ *stands for the approximate differential.*

Proof. Since $\mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2) < +\infty$ by assumption, Theorem 3.4 gives the existence of an optimal 1-homogeneous coupling $\boldsymbol{\alpha} \in \mathfrak{H}^1_o(\mu_1, \mu_2)$ concentrated on $\mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}$. On the other hand, using Proposition 5.2 we get that $\boldsymbol{\alpha}$ is concentrated on a σ-compact and H-cyclically monotone radial convex cone $\tilde{\Gamma}$. As in Remark 5.11, we can define

$$\Gamma := (\mathrm{hom}\,(\mathsf{supp}\,(\boldsymbol{\alpha})) \cap \mathrm{D}(\mathsf{H}) \cap \tilde{\Gamma}) \setminus \{(\mathfrak{o}_1, \mathfrak{o}_2)\}$$

and we obtain that Γ is still a σ-compact radial cone which is H-cyclically monotone, contained in $\mathrm{D}(\mathsf{H})$ and such that $(\mathfrak{o}_1, \mathfrak{o}_2) \notin \Gamma$. Clearly $\boldsymbol{\alpha}$ is concentrated on Γ and by Remark 5.11 we also have that $S_i := \mathsf{x}(\Gamma_i \setminus \{\mathfrak{o}_i\})$ is such that $S_i \subset \mathsf{supp}\,(\mu_i) = \overline{S_i}$. We want to apply Theorem 5.10 to deduce the existence of optimal potentials. To do so, we need to check that Γ is H-connected and that at least one of the two conditions (5.12) or (5.13) is satisfied: since H is everywhere finite and radially convex, then

$$\text{rad-int}\,(\mathrm{D}(\mathsf{H})) = \mathfrak{C}_\mathfrak{o}[\mathbb{R}^d] \times \mathfrak{C}_\mathfrak{o}[\mathbb{R}^d] = \{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[\mathbb{R}^d] \times \mathfrak{C}[\mathbb{R}^d] : \mathfrak{y}_i \neq \mathfrak{o}_i\}.$$

This in particular gives that at least one of the conditions (5.12) or (5.13) must be satisfied by Γ (otherwise at least one between μ_1 and μ_2 must be the null measure, which is not allowed). Moreover, Γ is clearly H-connected since H is everywhere finite.

By Theorem 5.10 and Remark 5.12, we can find Borel functions $\varphi_i : \mathsf{X}_i \to [-\infty, +\infty)$ such that $\varphi_i \in \mathcal{L}^1(\mathbb{R}^d, \mu_i)$, $\varphi_i(x_i) \in \mathbb{R}$ if $x_i \in S_i$ and

$$\varphi_1(x_1) r_1 + \varphi_2(x_2) r_2 \le \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \text{for every } x_i \in \mathbb{R}^d,\ r_i \ge 0, \tag{5.31}$$

$$\varphi_1(x_1) r_1 + \varphi_2(x_2) r_2 = \mathsf{H}([x_1, r_1], [x_2, r_2]) \quad \text{if } ([x_1, r_1], [x_2, r_2]) \in \Gamma. \tag{5.32}$$

We want to show that φ_1 defined as above is approximately differential μ_1-a.e. in $\mathbb{R}^d$. Let us define, for every $R > 0$, the functions

$$\varphi_1^R(x_1) := \inf_{(x_2,r_2)\in \mathrm{B}(0,R)\times[0,R]} \left\{\mathsf{H}([x_1,1],[x_2,r_2]) - \varphi_2(x_2)r_2\right\}, \quad x_1 \in \mathbb{R}^d.$$

By assumption 1 we deduce that φ_1^R is locally Lipschitz and therefore differentiable $\mathcal{L}^d$-a.e. for R sufficiently large. Moreover, by definition of S_1 and since Γ is a radial cone, for every $x_1 \in S_1$, we can find $[x_2, r_2] \in \mathfrak{C}[\mathbb{R}^d]$ such that $([x_1, 1], [x_2, r_2]) \in \Gamma$ so that (5.32) holds for $([x_1, 1], [x_2, r_2])$. This, together with (5.31), implies that for μ_1-a.e. $x_1 \in S_1$ the decreasing family of sets $\{\varphi < \varphi_1^R\}$ has a μ_1-negligible intersection, i.e. μ_1-a.e. $x_1 \in S_1$ belongs to $\{\varphi_1 = \varphi_1^R\}$ for R large enough. It follows that for μ_1-a.e. $x_1 \in S_1$ the following two conditions are satisfied: x_1 is a point of density 1 of $\{\varphi_1 = \varphi_1^R\}$ for some R and φ_1^R is differentiable at x_1. We deduce that φ_1 is approximately differentiable at x_1 and $\tilde\nabla\varphi_1(x_1) = \nabla\varphi_1^R(x_1)$. Let us denote by $A_1 \subset S_1$ the full μ_1-measure set where this happens.

Let now $([x_1, r_1], [x_2, r_2]) \in \Gamma$ with $x_1 \in A_1$. Notice that the map

$$(x_1', r_1') \mapsto \mathsf{H}([x_1', 1], [x_2, r_2]) - \varphi_1(x_1')r_1'$$

attains its minimum at (x_1, r_1). Let us show that it must be that $r_1 > 0$: if, by contradiction $r_1 = 0$, then we get

$$\mathsf{H}([x_1', r_1'], [x_2, r_2]) - \varphi_1(x_1')r_1' \geq \mathsf{H}(\mathfrak{o}_1, [x_2, r_2]) \text{ for every } (x_1', r_1') \in \mathbb{R}^d \times [0, +\infty).$$

Since $[x_2, r_2] \neq \mathfrak{o}_2$ then $x_2 \in S_2$; by assumption 2 and since $\mathsf{supp}\,(\mu_1)$ cannot contain isolated points, we can find some $x_1' \in S_1$ such that $(x_1', x_2) \in A_{\mathsf{H}}$. We can divide this expression by $r_2 > 0$ so that

$$-\infty = \lim_{r_1'\downarrow 0} \frac{\mathsf{H}([x_1', r_1'/r_2], [x_2, 1]) - \mathsf{H}(\mathfrak{o}_1, [x_2, 1])}{r_1'} \geq \frac{1}{r_2}\varphi_1(x_1'),$$

a contradiction with the fact that $\varphi_1(x_1') \in \mathbb{R}$.

By differentiation we get that

$$\begin{aligned}\tilde\nabla\varphi_1(x_1)r_1 &= \partial_{x_1}\mathsf{H}([x_1, r_1], [x_2, r_2]),\\ \varphi_1(x_1) &= \partial_{r_1}\mathsf{H}([x_1, r_1], [x_2, r_2]).\end{aligned}$$

Using the radial 1-homogeneity of H and the radial 0-homogeneity of its subdifferential in r_1, we deduce that

$$\begin{aligned}\tilde\nabla\varphi_1(x_1) &= \partial_{x_1}\mathsf{H}([x_1, 1], [x_2, r_2/r_1]),\\ \varphi_1(x_1) &= \partial_{r_1}\mathsf{H}([x_1, 1], [x_2, r_2/r_1]).\end{aligned}$$

By the invertibility assumption 3, we deduce that

$$[x_2, r_2/r_1] = (\partial_{x_1,r_1}\mathsf{H})^{-1}(\tilde\nabla\varphi_1(x_1), \varphi_1(x_1)).$$

Since the set $\Gamma \cap \mathsf{x}_1^{-1}(A_1)$ has full $\boldsymbol{\alpha}$ measure (this follows by the fact that $0 \notin \mathsf{r}_1(\Gamma)$), this concludes the proof of the theorem. □

Remark 5.14. Theorem 5.13 could also be proven under a different set of hypotheses, namely dropping the assumption that H is finite everywhere and substituting condition 2 by imposing that H is finite on an open cone (cp. with (B.1)): there exists numbers $q_i \geq 0$ such that, setting

$$U_{q_1q_2} := \left\{([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \mid r_2 > r_1q_1,\ r_1 > r_2q_2\right\},$$

we have $\mathrm{D}(\mathsf{H}) = U_{q_1q_2}$. Clearly in this case the partial differentiability condition in 1 and the map in 3 should be restricted to $\{x_1 \in \mathbb{R}^d : \mathsf{H}([x_1, 1], [x_2, r_2]) < +\infty\}$ and $\{[y, q] \in \mathfrak{C}[\mathbb{R}^d] :$

$([x_1,1],[y,q]) \in \mathsf{D}(\mathsf{H})\}$, respectively. The proof stays unchanged and we only have to observe that in this case:

(1) $\text{rad-int}\,(\mathsf{D}(\mathsf{H})) = \mathsf{D}(\mathsf{H}) = U_{q_1q_2}$ so that $\Gamma \subset \text{rad-int}\,(\mathsf{D}(\mathsf{H}))$, hence condition (5.12) is satisfied. By Remark 5.9 Γ is also H-connected;
(2) whenever $([x_1,r_1],[x_2,r_2]) \in \Gamma$, since $\Gamma \subset U_{q_1q_2}$, we have $r_1 > 0$.

Example 5.15. The function $\mathsf{H}_{\mathsf{GHK}}$ in (2.18) satisfies the hypotheses of Theorem 5.13: $\mathsf{H}_{\mathsf{GHK}}$ is a finite, radially 1-homogeneous, convex and continuous function and condition (5.16) is satisfied with $\varrho_i = 1$. Condition (1) is clearly satisfied. Condition (3) follows by

$$\frac{\mathsf{H}_{\mathsf{GHK}}([x_1,r_1],[x_2,1]) - \mathsf{H}_{\mathsf{GHK}}(\mathfrak{o}_1,[x_2,1])}{r_1} = 1 - \frac{2}{\sqrt{r_1}} e^{-|x_1-x_2|^2/2} \to -\infty \text{ as } r_1 \downarrow 0,$$

for every $x_1, x_2 \in \mathbb{R}^d$. Condition (2) is easily seen to be satisfied since, if

$$\begin{pmatrix} \partial_{x_1}\mathsf{H}_{\mathsf{GHK}}([x_1,1],[y,q]) \\ \partial_{r_1}\mathsf{H}_{\mathsf{GHK}}([x_1,1],[y,q]) \end{pmatrix} = \begin{pmatrix} \sqrt{q}e^{-|x_1-y|^2/2}(x_1-y) \\ 1-\sqrt{q}e^{-|x_1-y|^2/2} \end{pmatrix} = \begin{pmatrix} v_0 \\ c_0 \end{pmatrix}$$

for some $(v_0,c_0) \in \mathbb{R}^d \times \mathbb{R}$, then

$$[y,q] = \left[x_1 - v_0(1-c_0)^{-1}, (1-c_0)^2 e^{|v_0|^2(1-c_0)^{-2}}\right] \text{ if } c_0 \neq 1, \quad [y,q] = \mathfrak{o}_2 \text{ if } c_0 = 1.$$

Appendix A. Metric and topological properties

In this section we study a few metric and topological properties of the Unbalanced Optimal Transport functional assuming that the cost function is related to a metric on the cone. As usual we fix a completely regular space X and an exponent $p \in [1,+\infty)$. We consider a function $\varrho : \mathfrak{C}[\mathsf{X},\mathsf{X}] \to [0,+\infty]$ such that

$$\varrho \text{ is an extended metric on } \mathfrak{C}[\mathsf{X}] \text{ such that } \varrho^p \text{ is radially 1-homogeneous, proper and lsc.} \tag{A.1}$$

We consider the Unbalanced Optimal Transport functional induced by ϱ^p on $\mathcal{M}_+(\mathsf{X})$ and the (extended) Wasserstein p-metric [Vil09; AGS08] induced by ϱ on $\mathcal{P}(\mathfrak{C}[\mathsf{X}])$.

Definition A.1. Let ϱ be as in (A.1). We define $\mathcal{D}_{\varrho,p} : \mathcal{M}_+(\mathsf{X}) \times \mathcal{M}_+(\mathsf{X}) \to [0,+\infty]$ and $W_{\varrho,p} : \mathcal{P}(\mathfrak{C}[\mathsf{X}]) \times \mathcal{P}(\mathfrak{C}[\mathsf{X}]) \to [0,+\infty]$ as

$$\mathcal{D}_{\varrho,p}(\mu_1,\mu_2) := \mathscr{U}_{\varrho^p}^{1/p}(\mu_1,\mu_2) = \left(\min\left\{ \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \,\mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^1(\mu_1,\mu_2) \right\} \right)^{1/p}, \quad \mu_1,\mu_2 \in \mathcal{M}_+(\mathsf{X}),$$

$$W_{\varrho,p}(\alpha_1,\alpha_2) := \mathsf{OT}_{\varrho^p}^{1/p}(\mu_1,\mu_2) = \left(\min\left\{ \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \,\mathrm{d}\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma(\alpha_1,\alpha_2) \right\} \right)^{1/p}, \quad \alpha_1,\alpha_2 \in \mathcal{P}(\mathfrak{C}[\mathsf{X}]),$$

where OT_{ϱ^p} is as in Definition 3.20. Finally we set

$$\mathcal{P}_{\varrho,p}(\mathfrak{C}[\mathsf{X}]) := \left\{ \alpha \in \mathcal{P}(\mathfrak{C}[\mathsf{X}] : \int_{\mathfrak{C}[\mathsf{X}]} \varrho^p(\mathfrak{y},\mathfrak{o}) \,\mathrm{d}\alpha(\mathfrak{y}) < +\infty \right\},$$
$$\mathcal{M}_{\varrho,p}(\mathsf{X}) := \left\{ \mu \in \mathcal{M}_+(\mathsf{X}) : \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o}) \,\mathrm{d}\mu(x) < +\infty \right\}.$$

Remark A.2. If $\mu \in \mathcal{M}_{\varrho,p}(\mathsf{X})$ then every $\alpha \in \mathcal{P}(\mathfrak{C}[\mathsf{X}])$ such that $\mathfrak{h}^1(\alpha) = \mu$ is an element of $\mathcal{P}_{\varrho,p}(\mathfrak{C}[\mathsf{X}])$.

Remark A.3. As explained in Remark 2.13, we limit our analysis to the case in which ϱ^p is radially 1-homogeneous. In case ϱ^p is radially q-homogeneous, $q \in (1, +\infty)$, we can argue as in Remark 2.13 and consider its composition with T_q, which is a radially 1-homogeneous metric, obtaining

$$\mathscr{U}_{\varrho^p \circ \mathsf{T}_q}(\mu_1, \mu_2) = \inf \left\{ \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \, \mathrm{d}\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathfrak{H}^q(\mu_1, \mu_2) \right\}.$$

As a relevant example, the case $\varrho = \mathsf{d}_{\mathfrak{C}}$ (see (2.9)) and $p = q = 2$ fits into this setting and the resulting functional $\mathcal{D}_{\varrho,p}$ is the Helliger-Kantorovich metric on non-negative measures introduced in [LMS18].

The next two theorems are generalizations of [LMS18, Corollary 7.14, Theorem 7.15] and, although the proofs are similar, there are some modifications to be taken into account so that we report them. Similar results are also treated in [De 20; DM22]. First of all we show that $\mathcal{D}_{\varrho,p}$ is indeed a metric.

Theorem A.4. *Let ϱ be as in (A.1). Then $(\mathcal{M}_+(\mathsf{X}), \mathcal{D}_{\varrho,p})$ is an extended metric space and $(\mathcal{M}_{\varrho,p}(\mathsf{X}), \mathcal{D}_{\varrho,p})$ is a metric space.*

Proof. Define $\mathsf{T} : \mathfrak{C}[\mathsf{X},\mathsf{X}] \to \mathfrak{C}[\mathsf{X},\mathsf{X}]$ as

$$\mathsf{T}(\mathfrak{y}_1, \mathfrak{y}_2) = (\mathfrak{y}_2, \mathfrak{y}_1), \quad (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[\mathsf{X},\mathsf{X}].$$

Then, for every $\mu_1, \mu_2 \in \mathcal{M}_+(\mathsf{X})$, $\mathsf{T}_\sharp : \mathfrak{H}^1(\mu_1, \mu_2) \to \mathfrak{H}^1(\mu_2, \mu_1)$ is a bijection satisfying

$$\int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \, \mathrm{d}\boldsymbol{\alpha} = \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \, \mathrm{d}\mathsf{T}_\sharp \boldsymbol{\alpha}$$

by the symmetry of ϱ. This gives that $\mathcal{D}_{\varrho,p}(\mu_1, \mu_2) = \mathcal{D}_{\varrho,p}(\mu_2, \mu_1)$ for every $\mu_1, \mu_2 \in \mathcal{M}_+(\mathsf{X})$.
If $\mu \in \mathcal{M}_+(\mathsf{X})$ and we define

$$\boldsymbol{\alpha} = ((\mathrm{id}_{\mathfrak{C}[\mathsf{X}]}, \mathrm{id}_{\mathfrak{C}[\mathsf{X}]}) \circ \mathfrak{p})_\sharp(\mu \otimes \delta_1) \in \mathfrak{H}^1(\mu, \mu),$$

we obtain that

$$\mathcal{D}^p_{\varrho,p}(\mu, \mu) \le \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p \, \mathrm{d}\boldsymbol{\alpha} = \int_{\mathfrak{C}[\mathsf{X},\mathsf{X}]} \varrho^p(\mathfrak{y}, \mathfrak{y}) \, \mathrm{d}(\mathfrak{p}_\sharp(\mu \otimes \delta_1))(\mathfrak{y}) = 0.$$

If, on the other hand, $\mu_1, \mu_2 \in \mathcal{M}_+(\mathsf{X})$ are s.t. $\mathcal{D}_{\varrho,p}(\mu_1, \mu_2) = 0$ and $\boldsymbol{\alpha} \in \mathfrak{H}^1_o(\mu_1, \mu_2)$ is optimal, we get that $\boldsymbol{\alpha}$ is concentrated on the diagonal $\{(\mathfrak{y}, \mathfrak{y}) : \mathfrak{y} \in \mathfrak{C}[\mathsf{X}]\}$, so that $\mu_1 = \mathfrak{h}^1_1(\boldsymbol{\alpha}) = \mathfrak{h}^1_2(\boldsymbol{\alpha}) = \mu_2$. This proves that $\mathcal{D}_{\varrho,p}(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$.
Finally if $\mu_1, \mu_2, \mu_3 \in \mathcal{M}_+(\mathsf{X})$, we can find, thanks to Proposition 3.22, $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(\mathfrak{C}[\mathsf{X}])$ such that

$$\mathfrak{h}^1(\alpha_i) = \mu_i, \quad \mathcal{D}_{\varrho,p}(\mu_{i-1}, \mu_i) = W_{\varrho,p}(\alpha_{i-1}, \alpha_i) \quad i = 2, 3.$$

Using again Proposition 3.22, we have

$$\mathcal{D}_{\varrho,p}(\mu_1, \mu_3) \le W_{\varrho,p}(\alpha_1, \alpha_3) \le W_{\varrho,p}(\alpha_1, \alpha_2) + W_{\varrho,p}(\alpha_2, \alpha_3) = \mathcal{D}_{\varrho,p}(\mu_1, \mu_2) + \mathcal{D}_{\varrho,p}(\mu_2, \mu_3).$$

This proves that $\mathcal{D}_{\varrho,p}$ satisfies the triangle inequality and concludes the proof that $(\mathcal{M}_+(\mathsf{X}), \mathcal{D}_{\varrho,p})$ is an extended metric space.
If $\mu \in \mathcal{M}_{\varrho,p}(\mathsf{X})$ and $\alpha \in \mathcal{P}(\mathfrak{C}[\mathsf{X}])$ is s.t. $\mathfrak{h}^1(\alpha) = \mu$, then

$$\int_{\mathfrak{C}[\mathsf{X}]} \varrho^p(\mathfrak{y}, \mathfrak{o}) \, \mathrm{d}\alpha(\mathfrak{y}) = \int_{\mathfrak{C}[\mathsf{X}]} r \varrho^p([x, 1], \mathfrak{o}) \, \mathrm{d}\alpha([x, r]) = \int_{\mathsf{X}} \varrho^p([x, 1], \mathfrak{o}) \, \mathrm{d}\mu(x) < +\infty,$$

so that $\alpha \in \mathcal{P}_{\varrho,p}(\mathfrak{C}[\mathsf{X}])$. Then, again from Proposition 3.22, if $\mu_1, \mu_2 \in \mathcal{M}_{\varrho,p}(\mathsf{X})$ we can find $\alpha_1, \alpha_2 \in \mathcal{P}_{\varrho,p}(\mathfrak{C}[\mathsf{X}])$ s.t. $\mathfrak{h}^1(\alpha_i) = \mu_i$ for $i = 1, 2$ so that

$$\mathcal{D}_{\varrho,p}(\mu_1, \mu_2) \le W_{\varrho,p}(\alpha_1, \alpha_2) < +\infty.$$

□

The following theorem extends to the unbalanced setting the well known result for the Wasserstein distance (see e.g. [AGS08, Remark 7.1.11]).

Theorem A.5. *Let ϱ be a metric in $\mathfrak{C}[\mathsf{X}]$ inducing the topology of $\mathfrak{C}[\mathsf{X}]$ such that ϱ^p is radially 1-homogeneous. If $(\mu_n)_n \subset \mathcal{M}_{\varrho,p}(\mathsf{X})$ and $\mu \in \mathcal{M}_{\varrho,p}(\mathsf{X})$, then*

$$\lim_{n\to+\infty} \mathcal{D}_{\varrho,p}(\mu_n,\mu) = 0 \quad \iff \quad \begin{cases} \mu_n \rightharpoonup \mu \\ \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu_n(x) \to \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu(x) \end{cases}.$$

In particular if X is separable, also $(\mathcal{M}_{\varrho,p}(\mathsf{X}),\mathcal{D}_{\varrho,p})$ is separable.

Proof. Let us first observe that (2.8) yields

$$\text{there exists } a>0 \text{ such that } \varrho^p([x,1],\mathfrak{o}) \ge a \quad \text{for every } x\in\mathsf{X}. \tag{A.2}$$

We first prove the $\Rightarrow$ implication. Notice that, denoting by $\boldsymbol{0}_{\mathsf{X}}$ the null measure in X, we have

$$\mathcal{D}_{\varrho,p}(\nu,\boldsymbol{0}_{\mathsf{X}}) = \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\nu(x) \quad \text{for every } \nu\in\mathcal{M}_{\varrho,p}(\mathsf{X}) \tag{A.3}$$

so that, by triangle inequality, we get

$$\int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu_n(x) = \mathcal{D}^p_{\varrho,p}(\mu_n,\boldsymbol{0}_{\mathsf{X}}) \to \mathcal{D}^p_{\varrho,p}(\mu,\boldsymbol{0}_{\mathsf{X}}) = \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu(x).$$

We show that $\mu_n \rightharpoonup \mu$ by contradiction: assume that there exist $\xi\in \mathrm{C}_b(\mathsf{X})$ and a (unrelabeled) subsequence s.t.

$$\inf_n \left| \int_{\mathsf{X}} \xi\,\mathrm{d}\mu_n - \int_{\mathsf{X}} \xi\,\mathrm{d}\mu \right| > 0. \tag{A.4}$$

Observe that

$$\mu_n(\mathsf{X}) \le \frac{1}{a}\int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu_n(x) \to \frac{1}{a}\int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu(x) < +\infty,$$

so that $R := (\sup_n \mu_n(\mathsf{X}) + \mu(\mathsf{X})) < +\infty$. By Theorem 3.4 and Lemma 2.12, we can find $(\boldsymbol{\alpha}_n)_n \subset \mathcal{P}(\mathfrak{C}_R[\mathsf{X},\mathsf{X}])$ such that $\boldsymbol{\alpha}_n \in \mathfrak{H}^1_o(\mu_n,\mu)$ is optimal for every $n\in\mathbb{N}$. Let us define $\alpha^1_n := \pi^1_\sharp \boldsymbol{\alpha}_n$, $\alpha^2_n := \pi^2_\sharp \boldsymbol{\alpha}_n$, $n\in\mathbb{N}$. Since $\mathfrak{h}^1(\alpha^2_n)=\mu$ for every $n\in\mathbb{N}$, we obtain by Lemma 2.9 the existence of a subsequence $k\mapsto n(k)$ and $\alpha_2\in\mathcal{P}(\mathfrak{C}_R[\mathsf{X}])$ with $\mathfrak{h}^1(\alpha_2)=\mu$ such that $\alpha^2_{n(k)} \rightharpoonup \alpha_2$. Moreover

$$\int_{\mathfrak{C}[\mathsf{X}]} \varrho^p(\mathfrak{y},\mathfrak{o})\,\mathrm{d}\alpha^2_n = \int_{\mathsf{X}} \varrho^p([x,1],\mathfrak{o})\,\mathrm{d}\mu(x) \quad \text{for every } n\in\mathbb{N},$$

giving that (see e.g. [AGS08, Proposition 7.1.5]) $W_{\varrho,p}(\alpha^2_{n(k)},\alpha_2)\to 0$. Then

$$W_{\varrho,p}(\alpha^1_{n(k)},\alpha_2) \le W_{\varrho,p}(\alpha^1_{n(k)},\alpha^2_{n(k)}) + W_{\varrho,p}(\alpha^2_{n(k)},\alpha_2) = \mathcal{D}_{\varrho,p}(\mu_{n(k)},\mu) + W_{\varrho,p}(\alpha^2_n,\alpha_2) \to 0,$$

where we used Proposition 3.22. Thus $W_{\varrho,p}(\alpha^1_{n(k)},\alpha_2)\to 0$ and, in particular, $\alpha^1_{n(k)} \rightharpoonup \alpha_2$ so that

$$\int_{\mathsf{X}} \xi\,\mathrm{d}\mu_{n(k)} = \int_{\mathfrak{C}[\mathsf{X}]} \xi(x) r\,\mathrm{d}\alpha^1_{n(k)}([x,r]) \to \int_{\mathfrak{C}[\mathsf{X}]} \xi(x) r\,\mathrm{d}\alpha_2([x,r]) = \int_{\mathsf{X}} \xi\,\mathrm{d}\mu, \tag{A.5}$$

where we used that the map

$$[x,r]\mapsto r\xi(x)$$

belongs to $\mathrm{C}_b(\mathfrak{C}_R[\mathsf{X}])$ and $\alpha^1_{n(k)}$ is concentrated on $\mathfrak{C}_R[\mathsf{X}]$ for every $k\in\mathbb{N}$. Since (A.5) is a contradiction with (A.4), this concludes the proof of the $\Rightarrow$ implication.
Let us prove the $\Leftarrow$ implication. If $\mu=\boldsymbol{0}_{\mathsf{X}}$, we have already by (A.3) that $\mathcal{D}_{\varrho,p}(\mu_n,\mu)\to 0$. Let us then assume that $m:=\mu(\mathsf{X})>0$. Up to passing to a (unrelabeled) subsequence, we can assume that $m_n := \mu_n(\mathsf{X}) \ge m/2 > 0$ for every $n\in\mathbb{N}$. Let us define $\alpha_n,\alpha\in\mathcal{P}(\mathfrak{C}[\mathsf{X}])$ as

$$\alpha := \mathfrak{p}_\sharp\left(m^{-1}\mu\otimes\delta_m\right), \quad \alpha_n := \mathfrak{p}_\sharp\left(m_n^{-1}\mu_n\otimes\delta_{m_n}\right) \quad n\in\mathbb{N}.$$

It is easy to check that $\mathfrak{h}^1(\alpha_n) = \mu_n$, $n \in \mathbb{N}$, $\mathfrak{h}^1(\alpha) = \mu$ and $\alpha_n \rightharpoonup \alpha$. To conclude is then sufficient to show that $W_{\varrho,p}(\alpha_n, \alpha) \to 0$ and then apply Lemma 3.22. The 1-homogeneity of ϱ^p yields

$$\begin{aligned}\int_{\mathfrak{C}[\mathsf{X}]} \varrho^p(\mathfrak{y}, \mathfrak{o}) \, \mathrm{d}\alpha_n(\mathfrak{y}) &= \int_{\mathsf{X}} m_n^{-1} \varrho^p([x, m_n], \mathfrak{o}) \, \mathrm{d}\mu_n(x) \\ &= \int_{\mathsf{X}} \varrho^p([x, 1], \mathfrak{o}) \, \mathrm{d}\mu_n(x) \overset{n\to\infty}{\to} \int_{\mathsf{X}} \varrho^p([x, 1], \mathfrak{o}) \, \mathrm{d}\mu(x) \\ &= \int_{\mathsf{X}} m^{-1} \varrho^p([x, m], \mathfrak{o}) \, \mathrm{d}\mu(x) = \int_{\mathfrak{C}[\mathsf{X}]} \varrho^p(\mathfrak{y}, \mathfrak{o}) \, \mathrm{d}\alpha(\mathfrak{y}),\end{aligned}$$

and we get that $W_{\varrho,p}(\alpha_n, \alpha) \to 0$ applying [AGS08, Proposition 7.1.5]. □

Appendix B. The case of a cost function finite on an open cone

In this appendix we repeat the constructions of Section 4 under different assumptions on H. As in Section 4, we assume that X_1 and X_2 are compact and metrizable spaces. Given non negative numbers q_i, $i = 1, 2$, we define the open cone

$$U_{q_1 q_2} := \{([x_1, r_1], [x_2, r_2]) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \mid r_2 > r_1 q_1,\ r_1 > r_2 q_2\}.$$

Using this notation, we assume that

$$\begin{gathered}\mathsf{H} : \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \to [0, +\infty] \text{ is continuous, radially 1-homogeneous and convex and} \\ \text{that there exists non-negative numbers } q_1, q_2 \text{ such that } \mathrm{D}(\mathsf{H}) = U_{q_1 q_2} \text{ and} \\ \lim_{r_1 \downarrow q_2} \inf_{(x_1,x_2) \in \mathsf{X}_1 \times \mathsf{X}_2} \mathsf{H}([x_1, r_1], [x_2, 1]) = \lim_{r_2 \downarrow q_1} \inf_{(x_1,x_2) \in \mathsf{X}_1 \times \mathsf{X}_2} \mathsf{H}([x_1, 1], [x_2, r_2]) = +\infty.\end{gathered} \tag{B.1}$$

Notice that, if $q_1 = q_2 = 0$, we are simply assuming that H is finite on the whole open cone.

Proposition B.1. *Assume that* H *is as in* (B.1) *and that* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that*

$$\mathsf{supp}\,(\mu_i) = \mathsf{X}_i,\ i = 1, 2, \quad q_1 < \frac{\mu_2(\mathsf{X}_2)}{\mu_1(\mathsf{X}_1)}, \quad q_2 < \frac{\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)},$$

where q_i *are as in* (B.1). *Then there exists a constant* $C > 0$ *such that, for every* $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ *with*

$$\int_{\mathsf{X}_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \, \mathrm{d}\mu_2 \geq 0,$$

it holds

$$\varphi_1(x_1) \leq C, \quad \varphi_2(x_2) \leq C \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$$

and there exists $(\bar{x}_1, \bar{x}_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ *such that* $\varphi_1(\bar{x}_1) \geq -C$, $\varphi_2(\bar{x}_2) \geq -C$.

Proof. We start from the last claim for φ_1. Assume by contradiction that there exists a sequence $(\varphi_1^j, \varphi_2^j)_j \subset \Phi_{\mathsf{H}}$ with $\int_{\mathsf{X}_1} \varphi_1^j \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2^j \, \mathrm{d}\mu_2 \geq 0$ such that $\max_{x_1 \in \mathsf{X}_1} \varphi_1^j(x_1) \to -\infty$. Let $(x_1^j)_j \subset \mathsf{X}_1$ be the sequence of points where the maxima are attained. We thus have

$$\varphi_1^j(x_1^j)\mu_1(\mathsf{X}_1) + \int_{\mathsf{X}_2} \varphi_2^j \, \mathrm{d}\mu_2 \geq 0 \quad \text{for every } j \in \mathbb{N}$$

so that we can find $(x_2^j)_j \subset \mathsf{X}_2$ such that

$$\varphi_2^j(x_2^j) \geq -\frac{\varphi_1^j(x_1^j)\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)} \quad \text{for every } j \in \mathbb{N}.$$

Since $(\varphi_1^j, \varphi_2^j) \in \Phi_{\mathsf{H}}$, we have

$$\varphi_1^j(x_1^j)\left(r_1 - \frac{\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)} r_2\right) \leq \varphi_1^j(x_1^j) r_1 + \varphi_2^j(x_2^j) \leq \mathsf{H}([x_1^j, r_1], [x_2^j, r_2]) \quad \text{for every } r_1, r_2 \geq 0.$$

We can assume, up to passing to a subsequence, that $(x_1^j, x_2^j) \to (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$. Thanks to (B.1) and the assumptions on μ_i, we can find $\bar r_1, \bar r_2 > 0$ such that

$$\mathsf{H}([x_1, \bar r_1], [x_2, \bar r_2]) < +\infty, \quad \bar r_1 - \frac{\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)} \bar r_2 < 0.$$

We thus have that

$$+\infty \le \mathsf{H}([x_1, \bar r_1], [x_2, \bar r_2],$$

a contradiction with (B.1). The proof for φ_2 is the same; we have thus proven that there exists a constant $D > 0$ independent of (φ_1, φ_2) and a point $(\bar x_1, \bar x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$ such that

$$\varphi_1(\bar x_1) \ge -D, \quad \varphi_2(\bar x_2) \ge -D.$$

Thus, if we set

$$C := D + \max_{(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2} \mathsf{H}([x_1, \bar r_1], [x_2, \bar r_2]),$$

where $\bar r_1$ and $\bar r_2$ are as above, we get that

$$\varphi_1(x_1) \le \frac{1}{\bar r_1} \left(\mathsf{H}([x_1, \bar r_1], [\bar x_2, \bar r_1]) - \bar r_2 \varphi_2(\bar x_2) \right) \le C \quad \text{for every } x_1 \in \mathsf{X}_1$$

and the corresponding statement for φ_2. □

In the following result, which is the analogue of Proposition 4.6, we use the notation of Definition 4.5.

Proposition B.2. *Assume that* H *is as in* (B.1) *and that* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that*

$$\mathsf{supp}\,(\mu_i) = \mathsf{X}_i,\ i = 1, 2, \quad q_1 < \frac{\mu_2(\mathsf{X}_2)}{\mu_1(\mathsf{X}_1)}, \quad q_2 < \frac{\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)},$$

where q_i *are as in* (B.1)*. Then there exist constants* $a_i, a_s, b_i, b_s, M > 0$ *such that, if* $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ *are such that*

$$\int_{\mathsf{X}_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \, \mathrm{d}\mu_2 \ge 0,$$

then $\|\varphi_1^{\mathsf{H}}\|_\infty, \|\varphi_1^{\mathsf{HH}}\|_\infty \le M$ *and*

$$\varphi_1^{\mathsf{H}}(x_2) = \inf_{x_1 \in \mathsf{X}_1} \inf_{a_i \le \alpha \le a_s} \Big\{ \mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha \varphi_1(x_1) \Big\}, \quad x_2 \in \mathsf{X}_2, \tag{B.2}$$

$$\varphi_1^{\mathsf{HH}}(x_1) = \inf_{x_2 \in \mathsf{X}_2} \inf_{b_i \le \alpha \le b_s} \Big\{ \mathsf{H}([x_1, 1], [x_2, \alpha]) - \alpha \varphi_1^{\mathsf{H}}(x_2) \Big\}, \quad x_1 \in \mathsf{X}_1. \tag{B.3}$$

Moreover the sets

$$\mathfrak{C}_0^1 := \{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \mid a_i \le \mathsf{r}(\mathfrak{y}_1) \le a_s,\ \mathsf{r}(\mathfrak{y}_2) = 1\},$$
$$\mathfrak{C}_0^2 := \{(\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}[\mathsf{X}_1, \mathsf{X}_2] \mid b_i \le \mathsf{r}(\mathfrak{y}_2) \le b_s,\ \mathsf{r}(\mathfrak{y}_1) = 1\}$$

are compact subset of $U_{q_1 q_2}$ *and* $(\varphi_1^{\mathsf{HH}}, \varphi_1^{\mathsf{H}}) \in \Phi_{\mathsf{H}}$*,* $\varphi_1^{\mathsf{HH}} \ge \varphi_1$*,* $\varphi_1^{\mathsf{H}} \ge \varphi_2$*. Finally, if* d_i *are distances metrizing* X_i*,* $i = 1, 2$*, then* φ_1^{HH} *is* d_1*-uniformly continuous with the same (uniform)* $\mathsf{d}_1 \otimes_{\mathfrak{C}} \mathsf{d}_2$*-modulus of continuity of* H *on* $\mathfrak{C}_0^2$ *and* φ_1^{H} *is* d_2*-uniformly continuous with the same (uniform)* $\mathsf{d}_1 \otimes_{\mathfrak{C}} \mathsf{d}_2$*-modulus of continuity of* H *on* $\mathfrak{C}_0^1$*.*

Proof. Let (φ_1, φ_2) be as in the statement. Let us set

$$\delta := \frac{1}{2} \frac{1 - q_1 q_2}{1 + q_1 + q_2}$$

so that, for every $0 < \varepsilon \le \delta$, we have $q_2 + \varepsilon < (q_1 + \varepsilon)^{-1}$. Let us fix a point $\bar\alpha \in (q_2 + \delta, \frac{1}{q_1 + \delta})$ and let us define

$$m := \max_{(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2} \mathsf{H}([x_1, \bar\alpha], [x_2, 1]) < +\infty,$$

since $([x_1, \bar\alpha], [x_2, 1]) \in U_{q_1 q_2}$ for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$.
By (B.1), we know that for every $L > 0$, there exists $\varepsilon_L > 0$ such that

$$\mathsf{H}([x_1, r_1], [x_2, 1] \geq L \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ 0 \leq r_1 < q_2 + \varepsilon_L,$$
$$\mathsf{H}([x_1, 1], [x_2, r_2] \geq L \quad \text{for every } (x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2,\ 0 \leq r_2 < q_1 + \varepsilon_L.$$

Let

$$L := \max\{m + C(\bar\alpha + q_2 + \delta), (q_1 + \delta)(m + \bar\alpha C) + C\},$$

where C comes from Proposition B.1, and let us take any $a_i, a_s > 0$ such that

$$q_2 < a_i < q_2 + \varepsilon_L \wedge \delta, \quad \frac{1}{q_1 + \varepsilon_L \wedge \delta} < a_s < \frac{1}{q_1},$$

so that $0 < a_i < a_s$, $\mathfrak{C}^1_0 \subset U_{q_1 q_2}$ and $a_i \leq \bar\alpha \leq a_s$.
If $\alpha > a_s$, then, for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$, we have

$$\begin{aligned}
\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha\varphi_1(x_1) &= \alpha\left(\mathsf{H}([x_1, 1], [x_2, 1/\alpha]) - \varphi_1(x_1)\right) \\
&\geq \alpha(L - C) \\
&\geq a_s(L - C) \\
&\geq m + \bar\alpha C \\
&\geq \mathsf{H}([\bar x_1, \bar\alpha]), [x_2, 1] - \bar\alpha\varphi_1(\bar x_1),
\end{aligned}$$

where $\bar x_1$ comes from Proposition B.1. If $\alpha < a_i$, then, for every $(x_1, x_2) \in \mathsf{X}_1 \times \mathsf{X}_2$, we have

$$\begin{aligned}
\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha\varphi_1(x_1) &\geq L - \alpha C \\
&\geq L - C a_i \\
&\geq m + \bar\alpha C \\
&\geq \mathsf{H}([\bar x_1, \bar\alpha], [x_2, 1]) - \bar\alpha\varphi_1(\bar x_1).
\end{aligned}$$

Thus, for every $x_2 \in \mathsf{X}_2$, we get

$$\begin{aligned}
\inf_{x_1 \in \mathsf{X}_1} \inf_{0 \leq \alpha < a_i \vee \alpha > a_s} \{\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha\varphi_1(x_1)\} &\geq \mathsf{H}([\bar x_1, \bar\alpha], [x_2, 1]) - \bar\alpha\varphi_1(\bar x_1) \\
&> \inf_{x_1 \in \mathsf{X}_1} \inf_{a_i \leq \alpha \leq a_s} \{\mathsf{H}([x_1, \alpha], [x_2, 1]) - \alpha\varphi_1(x_1)\}
\end{aligned}$$

and this proves (B.2). The proof of (B.3) is analogous.
The remaining part of the proof is identical to the one of Proposition 4.6. □

By Proposition B.2 we obtain the analogue of Theorem 4.7 also in this setting with exactly the same proof.

Theorem B.3. *Assume that* H *is as in* (B.1) *and that* $\mu_i \in \mathcal{M}_+(\mathsf{X}_i)$ *are such that*

$$\mathsf{supp}(\mu_i) = \mathsf{X}_i,\ i = 1, 2, \quad q_1 < \frac{\mu_2(\mathsf{X}_2)}{\mu_1(\mathsf{X}_1)}, \quad q_2 < \frac{\mu_1(\mathsf{X}_1)}{\mu_2(\mathsf{X}_2)},$$

where q_i *are as in* (B.1). *Then there exists* $(\varphi_1, \varphi_2) \in \Phi_{\mathsf{H}}$ *such that*

$$\int_{\mathsf{X}_1} \varphi_1 \, \mathrm{d}\mu_1 + \int_{\mathsf{X}_2} \varphi_2 \, \mathrm{d}\mu_2 = \mathscr{U}_{\mathsf{H}}(\mu_1, \mu_2).$$

References

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Second edition. Lectures in Mathematics ETH Zürich. Basel: Birkhäuser Verlag, 2008, pp. x+334.

[BB00] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numer. Math.* 84.3 (2000), pp. 375–393.

[BC10] Stefano Bianchini and Laura Caravenna. "On optimality of c-cyclically monotone transference plans". In: *C. R. Math. Acad. Sci. Paris* 348.11-12 (2010), pp. 613–618.

[Bei+09] Mathias Beiglböck, Martin Goldstern, Gabriel Maresch, and Walter Schachermayer. "Optimal and better transport plans". In: *J. Funct. Anal.* 256.6 (2009), pp. 1907–1927.

[Ben03] Jean-David Benamou. "Numerical resolution of an "unbalanced" mass transport problem". In: *M2AN Math. Model. Numer. Anal.* 37.5 (2003), pp. 851–868.

[Bog06] Vladimir Igorevich Bogachev. *Measure Theory.* Vol. II. Springer Berlin Heidelberg, 2006.

[CD14] Thierry Champion and Luigi De Pascale. "On the twist condition and c-monotone transport plans". In: *Discrete Contin. Dyn. Syst.* 34.4 (2014), pp. 1339–1353.

[Chi+18a] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. "An interpolating distance between optimal transport and Fisher-Rao metrics". In: *Found. Comput. Math.* 18.1 (2018), pp. 1–44.

[Chi+18b] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. "Unbalanced optimal transport: dynamic and Kantorovich formulations". In: *J. Funct. Anal.* 274.11 (2018), pp. 3090–3123.

[CM10] Luis A. Caffarelli and Robert J. McCann. "Free boundaries in optimal transport and Monge-Ampère obstacle problems". In: *Ann. of Math. (2)* 171.2 (2010), pp. 673–730.

[De 20] Nicolò De Ponti. "Metric properties of homogeneous and spatially inhomogeneous F-divergences". In: *IEEE Trans. Inform. Theory* 66.5 (2020), pp. 2872–2890.

[DM22] Nicolò De Ponti and Andrea Mondino. "Entropy-transport distances between unbalanced metric measure spaces". In: *Probab. Theory Related Fields* 184.1-2 (2022), pp. 159–208.

[ET87] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems.* 1st ed. Classics in Applied Mathematics. Society for Industrial Mathematics, 1987.

[FG10] Alessio Figalli and Nicola Gigli. "A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions". In: *J. Math. Pures Appl. (9)* 94.2 (2010), pp. 107–130.

[Fig10] Alessio Figalli. "The optimal partial transport problem". In: *Arch. Ration. Mech. Anal.* 195.2 (2010), pp. 533–560.

[Fol99] Gerald B. Folland. *Real analysis.* Second edition. Pure and Applied Mathematics (New York). Modern techniques and their applications, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. xvi+386.

[Gui02] Kevin Guittet. "Extended Kantorovich norms: a tool for optimization". In: *Tech. Rep. 4402, INRIA* (2002).

[Han92] Leonid G. Hanin. "Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces". In: *Proc. Amer. Math. Soc.* 115.2 (1992), pp. 345–352.

[Han99] Leonid G. Hanin. "An extension of the Kantorovich norm". In: *Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997).* Vol. 226. Contemp. Math. Amer. Math. Soc., Providence, RI, 1999, pp. 113–130.

[KMV16] Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. "A new optimal transport distance on the space of finite Radon measures". In: *Adv. Differential Equations* 21.11-12 (2016), pp. 1117–1164.

[KR58] Leonid Kantorovich and Gennady S. Rubinstein. "On a space of totally additive functions". In: *Vestnik Leningrad. Univ* 13 (1958), pp. 52–59.

[LM15] Damiano Lombardi and Emmanuel Maitre. "Eulerian models and algorithms for unbalanced optimal transport". In: *ESAIM Math. Model. Numer. Anal.* 49.6 (2015), pp. 1717–1744.

[LM19] Vaios Laschos and Alexander Mielke. "Geometric properties of cones with applications on the Hellinger-Kantorovich space, and a new distance on the space of probability measures". In: *J. Funct. Anal.* 276.11 (2019), pp. 3529–3576.

[LMS16] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves". In: *SIAM J. Math. Anal.* 48.4 (2016), pp. 2869–2911.

[LMS18] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures". In: *Invent. Math.* 211.3 (2018), pp. 969–1117.

[LMS23] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Fine Properties of Geodesics and Geodesic λ-Convexity for the Hellinger–Kantorovich Distance". In: *Arch. Ration. Mech. Anal.* 247.6 (2023), p. 112.

[Maa+15] Jan Maas, Martin Rumpf, Carola Schönlieb, and Stefan Simon. "A generalized model for optimal transport of images including dissipation and density modulation". In: *ESAIM Math. Model. Numer. Anal.* 49.6 (2015), pp. 1745–1769.

[PR14] Benedetto Piccoli and Francesco Rossi. "Generalized Wasserstein distance and its application to transport equations with source". In: *Arch. Ration. Mech. Anal.* 211.1 (2014), pp. 335–358.

[PR16] Benedetto Piccoli and Francesco Rossi. "On properties of the generalized Wasserstein distance". In: *Arch. Ration. Mech. Anal.* 222.3 (2016), pp. 1339–1365.

[Pra07] Aldo Pratelli. "On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation". In: *Ann. Inst. H. Poincaré Probab. Statist.* 43.1 (2007), pp. 1–13.

[PW08] Ofir Pele and Michael Werman. "A Linear Time Histogram Metric for Improved SIFT Matching". In: Oct. 2008, pp. 495–508.

[RGT97] Yossi Rubner, Leonidas Guibas, and Carlo Tomasi. "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval". In: *Proceedings of the ARPA Image Understanding Workshop* (1997).

[San15] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Vol. 87. Progress in Nonlinear Differential Equations and their Applications. Calculus of variations, PDEs, and modeling. Birkhäuser/Springer, Cham, 2015, pp. xxvii+353.

[SS22] Giuseppe Savaré and Giacomo E. Sodini. "A simple relaxation approach to duality for Optimal Transport problems in completely regular spaces". In: *Journal of Convex Analysis* 29.1 (2022), pp. 1–12.

[Vil09] Cédric Villani. *Optimal transport.* Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973.

Giuseppe Savaré: Department of Decision Sciences, Bocconi University. Via Roentgen 1, 20136 Milan (Italy)
Email address: giuseppe.savare@unibocconi.it

Giacomo Enrico Sodini: Institut für Mathematik - Fakultät für Mathematik - Universität Wien, Oskar-Morgenstern-Platz 1, 1090 Wien (Austria)
Email address: giacomo.sodini@univie.ac.at