

# CONSTRAINED HELLINGER–KANTOROVICH BARYCENTERS: LEAST-COST SOFT AND CONIC MULTI-MARGINAL FORMULATIONS\*

MACIEJ BUZE†

**Abstract.** We show that the problem of finding the barycenter in the Hellinger—Kantorovich setting admits a least-cost soft multi-marginal formulation, provided that a one-sided hard marginal constraint is introduced. The constrained approach is then shown to admit a conic multi-marginal reformulation based on defining a single joint *multi-marginal perspective cost function* in the conic multi-marginal formulation, as opposed to separate two-marginal perspective cost functions for each two-marginal problem in the coupled-two-marginal formulation, as was studied previously in literature. We further establish that, as in the Wasserstein metric, the recently introduced framework of unbalanced multi-marginal optimal transport can be reformulated using the notion of the least cost. Subsequently, we discuss an example when input measures are Dirac masses and numerically solve an example for Gaussian measures. Finally, we also explore why the constrained approach can be seen as a natural extension of a Wasserstein space barycenter to the unbalanced setting.

**Key words.** optimal transport, unbalanced optimal transport, barycenters, multi-marginal optimal transport

**MSC codes.** 49Q22, 49Q20

**1. Introduction.** The theory of optimal transportation, dating back to Gaspard Monge’s work in 1781 [20], continues to develop at pace as one of the fundamental mathematical theories with an ever-growing list of diverse applications in fields such as economics, computer vision, image processing and machine learning – see monographs in [28, 26, 25] for a broad overview.

A central challenge in many applications concerns finding a representative, or *barycentric* (probability) distribution, which provides some average description of a given set of distributions. The basic optimal transport approach to this problem is to find the barycenter by minimizing the sum of weighted two-marginal optimal transport costs between the barycenter and each input distributions. It was subsequently shown that an equivalent and computationally favourable approach is to instead solve a single *least-cost multi-marginal* optimal transport problem [9, 1].

If the input distributions do not all have equal mass, an unbalanced barycenter can be found via a recourse to the emerging theory of unbalanced optimal transportation [10, 17, 18, 19]. This however, can be done in a number of ways, depending on how one penalises mass deviations, what cost function is employed and whether one wishes to consider the conic formulation. Recent important contributions in this area concern:

**(HKB):** the study of coupled-two-marginal Hellinger-Kantorovich barycenter and its equivalence to a conic least-cost multi-marginal formulation, as well as the non-existence of an equivalent soft multi-marginal formulation [13, 11, 5];

**(UMOT<sub>ε</sub>):** the study of entropy-regularised unbalanced multi-marginal optimal transport and its equivalence to the entropy regularised coupled-two-marginal formulation where the second marginal is fixed [4] – notably, this is done in an extended space setting, as opposed to the least-cost

---

\*

**Funding:** This work was supported by research grant EP/V00204X/1.

†School of Mathematical Sciences, Lancaster University, Lancaster, LA1 4YW, UK (m.buze@lancaster.ac.uk, <https://mbuze.github.io>).

setting known from the balanced case.

In particular, apart from the explicitly different starting coupled-two-marginal formulation for **(HKB)** and **(UMOT $_\varepsilon$ )** (putting aside the question of entropic regularisation), the relation between the two frameworks, especially with regards to their multi-marginal reformulations, has not been thoroughly explored. At the same time, multi-marginal approaches to (unbalanced) optimal transport continue to attract considerable attention both from a theoretical point of view and for applications, c.f., among many others, [23, 16, 3, 14, 2, 27, 24], so it is vital to study the mathematical underpinning of its unbalanced variant.

**1.1. Outline of the paper and its contributions.** In this paper we introduce and discuss the notion of a *constrained Hellinger-Kantorovich* barycenter, which, in its basic form and similarly to **(UMOT $_\varepsilon$ )**, concerns a coupled-two-marginal formulation of the unbalanced barycenter problem where the cost function and the first marginal entropy functional are chosen as in the Hellinger-Kantorovich (HK) metric, but the second marginal entropy functional is replaced with a hard constraint — the marginal has to match the barycenter exactly.

Concretely, we seek to find a barycenter  $\bar{\nu} \in \mathcal{M}(X)$  (a set of finite positive Radon measures on  $X$ , which for simplicity we take to be a compact subset of  $\mathbb{R}^d$ ) of a set of  $N$  measures

$$\vec{\mu} = (\mu_1, \dots, \mu_N), \quad \mu_i \in \mathcal{M}(X)$$

with weights  $\vec{\lambda} = (\lambda_1, \dots, \lambda_N)$ ,  $\lambda_i \geq 0$ , adding to one, by solving a constrained coupled two-marginal problem (CC2M, see Table 1 for more details on the subscript notation being used)

$$\text{HK}_{\text{CC2M}}(\vec{\mu}) = \inf_{\nu \in \mathcal{M}(X)} \left\{ \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \nu) \right\},$$

where

$$\overline{\text{HK}}^2(\mu_i, \nu) = \inf_{\gamma \in \mathcal{M}(X \times X)} \left\{ \int_{X \times X} c(x_1, x_2) d\gamma(x_1, x_2) + \mathcal{F}(\gamma_1 | \mu_i) + \bar{\mathcal{F}}(\gamma_2 | \nu) \right\}.$$

Here  $\gamma_i := (\pi^i)_\# \gamma$  is the  $i$ th marginal, the cost function  $c$  and entropy functional  $\mathcal{F}$  are as is the HK metric (see Section 2.1 for details), but  $\bar{\mathcal{F}}(\gamma_2 | \nu)$  vanishes if  $\gamma_2 = \nu$  and is equal to  $+\infty$  otherwise.

Complimentary to **(UMOT $_\varepsilon$ )**, we consider the unregularised problem and furthermore show that a natural *least-cost* soft multi-marginal (SMM) reformulation exists and is given by

$$\text{HK}_{\text{SMM}}(\vec{\mu}) := \inf_{\gamma \in \mathcal{M}(X^N)} \left\{ \int_{X^N} \tilde{c}(\vec{x}) d\gamma(\vec{x}) + \sum_i \lambda_i \mathcal{F}(\gamma_i | \mu_i) \right\},$$

where  $\vec{x} := (x_1, \dots, x_N)$  and the least cost  $\tilde{c}$  is given by

$$(1.1) \quad \tilde{c}(\vec{x}) = \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x).$$

To deepen the connection with **(HKB)**, we in fact begin the analysis by defining and studying a new conic multi-marginal problem (CMM), given by

$$\overline{\text{HK}}_{\text{CMM}}(\vec{\mu}) = \inf_{\alpha \in \mathcal{M}(Y^N)} \left\{ \int_{Y^N} \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) \alpha(\vec{x}, \vec{s}) \mid (\pi^{x_i})_\#(s_i \alpha) = \mu_i \right\},$$

where  $Y := X \times \mathbb{R}_+$  is the cone space (up to an equivalence relation for all  $y = (x, s) \in Y$  with  $s = 0$ ), and  $\vec{s} = (s_1, \dots, s_N) \in \mathbb{R}_+^N$  are the mass variables. The new *multi-marginal perspective cost function*  $\overline{H}_{\text{CMM}} : Y^N \rightarrow \overline{\mathbb{R}}$  is given by

$$\overline{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{(x,s) \in Y} \inf_{t > 0} \sum_{i=1}^N \left\{ t \lambda_i c(x_i, x) + \lambda_i s_i F\left(\frac{t}{s_i}\right) + \lambda_i s F\left(\frac{t}{s}\right) \right\},$$

where  $F$  is the entropy function in the definition of the functional  $\mathcal{F}$ , (see Section 2.1) and, as will be discussed in Section 3.1, is fundamentally different from the two-marginal perspective cost function approach from **(HKB)**.

The central result of the paper is the establishing of a three-way equality

$$\text{HK}_{\text{CC2M}}(\vec{\mu}) = \text{HK}_{\text{SMM}}(\vec{\mu}) = \overline{\text{HK}}_{\text{CMM}}(\vec{\mu}).$$

While for simplicity we only consider  $X$  to be a compact subset of  $\mathbb{R}^d$  and, to connect with **(HKB)**, we only focus on the HK setting, it is expected that the same equality should hold for the class of cost functions and entropy functionals covered by the theory of optimal entropy-transport problems developed in [18, Part I], provided that in (1.1) a unique minimizer exists or the infimum is given by  $+\infty$ .

The paper is organised as follows. In Section 2 we discuss the preliminaries, which includes a reminder on the Hellinger-Kantorovich distance and its conic formulation, followed by a brief account of the theory of (unconstrained) HK barycenters **(HKB)**.

The main results of the paper are presented in Section 3. We begin with analysis of the new conic multi-marginal problem  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  in Section 3.1, followed by showing in Section 3.2 that it is equivalent to the soft multi-marginal formulation  $\text{HK}_{\text{SMM}}(\vec{\mu})$ , for which existence of solutions is established (and hence also for the CMM formulation). Finally, in Section 3.3 we discuss the constrained coupled-two-marginal problem  $\text{HK}_{\text{CC2M}}(\vec{\mu})$ , show that it admits a solution and establish its equivalence with the soft multi-marginal formulation.

Section 4 is devoted to a discussion about the obtained results. In Section 4.1 we discuss the relation of our results to **(UMOT $_\varepsilon$ )**. This includes a numerical example in which we reproduce an example from [4] using our least-cost approach. This is followed by a discussion about different ways of characterising the difference between the constrained approach and **(HKB)** and includes an example about Dirac masses. Finally, in Section 4.3, we formally discuss the general idea of trying to extend the notion of a Wasserstein barycenter to the unbalanced setting and argue that the constrained approach might be a more natural one.

The paper finishes with Section 6, where proofs are gathered.

**Notation.** The theory developed in this paper necessitates a heavy use of the subscript notation to concisely distinguish between different barycenter and multi-marginal problems being considered. Table 1 will hopefully help readers navigate this.

## 2. Preliminaries.

**2.1. The Hellinger-Kantorovich distance.** The Hellinger-Kantorovich (HK) distance between two finite positive Radon measures  $\mu_1, \mu_2 \in \mathcal{M}(X)$  on a Hausdorff topological space  $X$  endowed with an extended distance function  $\mathbf{d} : X \times X \rightarrow [0, +\infty]$ , introduced independently by three different groups [10, 17, 18], is given by

$$(2.1) \quad \text{HK}^2(\mu_1, \mu_2) = \inf_{\gamma \in \mathcal{M}(X \times X)} \left\{ (c, \gamma) + \sum_{i=1,2} \mathcal{F}(\gamma_i \mid \mu_i) \right\},$$

Table 1: The subscript notation used throughout the paper.

Notation	Full name	Defined in
$\text{HK}_{\text{C2M}}(\vec{\mu})$	The <b>coupled two-marginal (C2M)</b> Hellinger-Kantorovich barycenter problem	(2.14)
$\text{HK}_{\text{CMM}}(\vec{\mu})$	The <b>conic multi-marginal (CMM)</b> formulation of the Hellinger-Kantorovich barycenter problem	(2.15)
$\text{HK}_{\text{CC2M}}(\vec{\mu})$	The <b>constrained coupled two-marginal (CC2M)</b> Hellinger-Kantorovich barycenter problem	(3.6)
$\text{HK}_{\text{SMM}}(\vec{\mu})$	The <b>soft multi-marginal (SMM)</b> formulation of the constrained Hellinger-Kantorovich barycenter problem	(3.4)
$\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$	The <b>conic multi-marginal (CMM)</b> formulation of the constrained Hellinger-Kantorovich barycenter problem	(3.1)

where

$$(c, \gamma) := \int_{X \times X} c(x_1, x_2) d\gamma(x_1, x_2)$$

and the cost function  $c$  is given by

$$(2.2) \quad c(x_1, x_2) := \begin{cases} -\log(\cos^2(d(x_1, x_2))) & \text{if } d(x_1, x_2) < \pi/2 \\ +\infty & \text{otherwise.} \end{cases}$$

As will become apparent in (2.10), in the classical metric framework this choice of the cost function is special, as it gives rise to the conic distance over the cone space  $Y = X \times \mathbb{R}_+$ . The entropy functional  $\mathcal{F}$  is given by

$$(2.3) \quad \mathcal{F}(\gamma_i | \mu_i) := \int_X F(\sigma_i) d\mu_i + F'_\infty \gamma_i^\perp(X), \quad \gamma_i := (\pi^i)_\# \gamma, \pi^i(x_1, x_2) = x_i.$$

In this definition we have used the Lebesgue decomposition (see e.g. [18, Lemma 2.3])

$$(2.4) \quad \gamma_i = \sigma_i \mu_i + \gamma_i^\perp, \quad \mu_i = \rho_i \gamma_i + \mu_i^\perp.$$

and we further note that the function  $F : [0, +\infty) \rightarrow [0, +\infty)$  is the relative entropy function, also known as Kullback-Leibler (KL) divergence, and is given by

$$(2.5) \quad F(s) := s \log s - s + 1, \quad \left( \implies F'_\infty := \lim_{s \rightarrow \infty} \frac{F(s)}{s} = +\infty \right),$$

thus enforcing  $\gamma_i \ll \mu_i$  for any feasible plan  $\gamma$  in (2.1).

**2.2. Reverse formulation and lifting to the cone.** The HK distance admits several useful reformulations, with two of them particularly relevant to the present work. Given the entropy function  $F$  defined in (2.5), the corresponding reverse entropy function  $R : [0, \infty) \rightarrow [0, \infty]$ , defined as

$$(2.6) \quad R(s) := \begin{cases} sF(1/s) & \text{if } s > 0, \\ F'_\infty & \text{if } s = 0, \end{cases}$$

is in fact given by

$$(2.7) \quad R(s) = s - \log s - 1.$$

One can then show [18, Theorem 3.11] that

$$(2.8) \quad \text{HK}^2(\mu_1, \mu_2) = \inf_{\gamma \in \mathcal{M}(X \times X)} \left\{ \int_{X \times X} c(x_1, x_2) + R(\rho_1(x_1)) + R(\rho_2(x_2)) d\gamma + \sum_{i=1,2} R'_\infty(\mu_i - \rho_i \gamma_i)(X) \right\},$$

where  $R'_\infty = F(0) = 1$  and  $\rho_1, \rho_2$  are defined through the Lebesgue decomposition (2.4).

This sets the scene for the lifting to the cone space  $Y := X \times \mathbb{R}_+$  (up to an equivalence relation for all  $y = (x, s) \in Y$  with  $s = 0$ ), which is facilitated by the introduction of the *marginal perspective cost function*,

$$H : X \times \mathbb{R}_+ \times X \times \mathbb{R}_+ \rightarrow [0, +\infty],$$

given by

$$(2.9) \quad H(x_1, s_1, x_2, s_2) := \inf_{t>0} \left\{ tc(x_1, x_2) + tR\left(\frac{s_1}{t}\right) + tR\left(\frac{s_2}{t}\right) \right\}.$$

A direct computation shows that, for  $R$  given by (2.7),  $H$  admits an explicit formulation given by

$$(2.10) \quad H(x_1, s_1, x_2, s_2) = s_1 + s_2 - 2s_1^{1/2} s_2^{1/2} \exp\left(\frac{-c(x_1, x_2)}{2}\right).$$

It can then be shown [18, Theorem 5.8] that, after lifting to the cone space  $Y$ , we have

$$(2.11) \quad \text{HK}^2(\mu_1, \mu_2) = \inf_{\alpha \in S(\mu_1, \mu_2)} (H, \alpha),$$

where

$$(H, \alpha) = \int_{Y \times Y} H(x_1, s_1, x_2, s_2) d\alpha(x_1, s_1, x_2, s_2)$$

and

$$S(\mu_1, \mu_2) = \{ \alpha \in \mathcal{M}(Y \times Y) \mid \mathbf{h}_i \alpha = \mu_i \},$$

where the homogeneous marginal  $\mathbf{h}_i : \mathcal{M}(Y \times Y) \rightarrow \mathcal{M}(X)$  is defined as

$$(2.12) \quad \mathbf{h}_i \alpha := \pi_{\#}^{x_i}(s_i \alpha).$$

Note that  $\mathbf{h}_i \alpha = \mu_i$  ensures that the lifted counterpart to the singular part in (2.8) satisfies

$$(2.13) \quad \sum_{i=1,2} R'_\infty(\mu_i - \mathbf{h}_i \alpha)(X) = 0.$$

**2.3. Unconstrained barycenters for the Hellinger-Kantorovich distance.** In what follows, for simplicity, we focus on the case when  $X$  is a compact subset of  $\mathbb{R}^n$  with the distance  $\mathbf{d}$  given by the Euclidean distance. We further suppose we are given  $N$  measures

$$\vec{\mu} = (\mu_1, \dots, \mu_N), \quad \mu_i \in \mathcal{M}(X)$$

and  $N$  weights

$$\vec{\lambda} = (\lambda_1, \dots, \lambda_N), \quad \lambda_i \in \mathbb{R}_+, \quad \sum_{i=1}^N \lambda_i = 1.$$

We will now briefly recall relevant known results about the HK barycenter problem, as studied in [13, 11]. We will aim to follow the presentation in [13], with a few relevant notational changes.

**2.3.1. Coupled two-marginal formulation.** The *unconstrained* Hellinger-Kantorovich barycenter problem is given by

$$(2.14) \quad \text{HK}_{\text{C2M}}(\vec{\mu}) = \inf_{\nu \in \mathcal{M}(X)} \left\{ \sum_{i=1}^N \lambda_i \text{HK}^2(\mu_i, \nu) \right\},$$

where we have used the convention employed in [13] to use the subscript C2M to emphasise that this is the coupled two-marginal formulation. We refer to this formulation as *unconstrained* because, in contrast to the new approach to be introduced in Section 3, the second marginals of optimal plans minimising  $\text{HK}^2(\mu_i, \nu)$  are not constrained to be equal to  $\nu$ .

**2.3.2. Multi-marginal formulation.** In analogy with the Wasserstein distance barycenters [1], it is of considerable interest to study whether the HK barycenter problem defined in (2.14) admits a multi-marginal reformulation. As established in [13, 11] and further studied in [5], this is possible starting from the cone formulation described in Section 2.2, treating the marginal perspective cost function  $H$  defined in (2.9) as a given cost function. In particular, in [13], the authors consider a conic multi-marginal (CMM) formulation, which, using the notation

$$X^N := \underbrace{X \times \dots \times X}_{N \text{ times}}, \quad \vec{x} = (x_1, \dots, x_N) \in X^N, \quad \vec{s} = (s_1, \dots, s_N) \in \mathbb{R}_+^N,$$

is given by

$$(2.15) \quad \text{HK}_{\text{CMM}}(\vec{\mu}) = \inf_{\alpha \in \mathcal{S}(\vec{\mu})} (H_{\text{CMM}}, \alpha),$$

where

$$(H_{\text{CMM}}, \alpha) = \int_{Y^N} H_{\text{CMM}}(\vec{x}, \vec{s}) d\alpha(\vec{x}, \vec{s}),$$

and the cost function is given by

$$(2.16) \quad H_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{(x,s) \in Y} \sum_{i=1}^N \lambda_i H(x_i, s_i, x, s),$$

where  $H$  is given by (2.9). The infimum in (2.15) is taken over the set

$$(2.17) \quad \mathcal{S}(\vec{\mu}) := \{\alpha \in \mathcal{M}(Y^N) \mid \mathbf{h}_i \alpha = \mu_i\},$$

where the homogeneous marginal  $\mathbf{h}_i : \mathcal{M}(Y^N) \rightarrow \mathcal{M}(X)$  is as in (2.12), up to an obvious adjustment from  $Y \times Y$  to  $Y^N$ , and the space  $Y^N$  is defined as

$$Y^N := \underbrace{Y \times \dots \times Y}_{N \text{ times}}.$$

Among other results, the following is proven.

THEOREM 2.1 ([13, Theorem 5.13]). *It holds that*

$$\text{HK}_{\text{CMM}}(\vec{\mu}) = \text{HK}_{\text{C2M}}(\vec{\mu}).$$

The result is very difficult to obtain because the minimisation in (2.16) does not admit an explicit solution and thus a convex relaxation of the perspective cost function  $H_{\text{CMM}}$  needs to be introduced and the associated dual problems considered. Furthermore, it is shown in [13, Section 8], that the unconstrained Hellinger-Kantorovich barycenter problem  $\text{HK}_{\text{C2M}}(\vec{\mu})$  does not admit an original space -based multi-marginal formulation (the so-called soft multi-marginal formulation). This is a significant deviation from the Wasserstein barycenters (see Section 4.3 for an extended discussion on this). It is thus of considerable interest to explore what, ideally minimal, changes are needed to “simplify” the framework developed in [13, 11, 5] and e.g. relate it to the recently explored idea of unbalanced multi-marginal optimal transport [4] (see Section 4.1 for an extended discussion). This is what we will discuss now.

### 3. Main results.

**3.1. Conic multi-marginal formulation.** The starting point of the present work is a new conic multi-marginal formulation of, what will turn out to be, the *constrained* HK barycenter problem. In contrast to (2.15), we introduce it as

$$(3.1) \quad \overline{\text{HK}}_{\text{CMM}}(\vec{\mu}) := \inf_{\alpha \in S(\vec{\mu})} (\overline{H}_{\text{CMM}}, \alpha),$$

where the new *multi-marginal perspective cost function* is given by

$$(3.2) \quad \overline{H}_{\text{CMM}}(\vec{x}, \vec{s}) := \inf_{(x,s) \in Y} \inf_{t > 0} \sum_{i=1}^N \left\{ t \left( \lambda_i c(x_i, x) + \lambda_i R\left(\frac{s_i}{t}\right) + \lambda_i R\left(\frac{s}{t}\right) \right) \right\}$$

and the set  $S(\vec{\mu})$  is as in (2.17).

The initial idea behind defining such a cost was to look back at the two-marginal formulation and, using the reverse formulation from (2.8), rewrite it, as much as possible, as a collection of integrals with respect to two-marginal plans  $\gamma_i \in \mathcal{M}(X \times X)$ , namely

$$\begin{aligned} \text{HK}_{\text{C2M}}(\vec{\mu}) = \inf_{\nu \in \mathcal{M}(X)} \sum_{i=1}^N \inf_{\gamma_i \in \mathcal{M}(X \times X)} & \left( \lambda_i (\mu_i - \rho_{i,1} \gamma_{i,1})(X) + \lambda_i (\nu - \rho_{i,2} \gamma_{i,2})(X) \right. \\ & \left. + \int_{X \times X} \lambda_i c(x_1, x_2) + \lambda_i R(\rho_{i,1}(x_1) + \lambda_i R(\rho_{i,2}(x_2))) d\gamma_i \right) \end{aligned}$$

where, similarly to (2.4), we have used the Lebesgue decomposition

$$(3.3a) \quad \gamma_{i,1} = \sigma_{i,1} \mu_i + \gamma_{i,1}^\perp, \quad \mu_i = \rho_{i,1} \gamma_{i,1} + \mu_i^\perp$$

$$(3.3b) \quad \gamma_{i,2} = \sigma_{i,2} \nu + \gamma_{i,2}^\perp, \quad \nu = \rho_{i,2} \gamma_{i,2} + \nu^\perp.$$

*Remark 3.1* (Potential notational confusion). To avoid notation becoming too cumbersome, in the coupled-two-marginal formulations, we use  $\gamma_i \in \mathcal{M}(X \times X)$  to denote the  $i$ th two-marginal plan and  $\gamma_{i,j} \in \mathcal{M}(X)$  to denote the  $j$ th marginal of the  $i$ th plan. In contrast, in the multi-marginal formulation, we have a multi-marginal plan  $\gamma \in \mathcal{M}(X^N)$  and  $\gamma_i \in \mathcal{M}(X)$  denotes its  $i$ th marginal. This can be particularly confusing when  $N = 2$  and we will try to always indicate the space to which  $\gamma_i$  belongs to minimise potential confusion.

As first attempt, by trying to mimic the Wasserstein barycenters theory, a candidate for a multi-marginal cost function is given by

$$\inf_{(x,s) \in Y} \sum_{i=1}^N \left\{ (\lambda_i c(x_i, x) + \lambda_i R(s_i) + \lambda_i R(s)) \right\}.$$

Secondly, to mimic the lifting to the cone strategy developed in [18] and recalled in Section 2.2, to arrive at a multi-marginal perspective cost function, we further introduce an extra infimization over  $t > 0$ ,

$$\bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{(x,s) \in Y} \inf_{t > 0} \sum_{i=1}^N \left\{ t(\lambda_i c(x_i, x) + \lambda_i R\left(\frac{s_i}{t}\right) + \lambda_i R\left(\frac{s}{t}\right)) \right\}.$$

To directly compare with the cost function  $H_{\text{CMM}}$  considered in [13], we recall that it is given by

$$H_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{(x,s) \in Y} \sum_{i=1}^N \inf_{t_i > 0} \left\{ t_i(\lambda_i c(x_i, x) + \lambda_i R\left(\frac{s_i}{t_i}\right) + \lambda_i R\left(\frac{s}{t_i}\right)) \right\}.$$

The difference is thus that in the known multi-marginal formulation, the extra infimization over  $t_i > 0$  is introduced separately for each two-marginal problem. The key distinction is that in the new approach, we “exchange” the order of taking the infimum over  $t_i > 0$  and the summation in  $i$  and as a result only infimize over a joint multi-marginal  $t > 0$ . The idea of a multi-marginal perspective function is adapted from [8], where a similar notion was used to study the entropic regularisation of the unbalanced optimal transport problems.

We begin by proving the following.

**LEMMA 3.2.** *The multi-marginal perspective cost function  $\bar{H}_{\text{CMM}}$  defined in (3.2) can be equivalently written as*

$$\begin{aligned} \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) &= \inf_{(x,s) \in Y} \left( \sum_{i=1}^N \lambda_i s_i + s - 2s^{\frac{1}{2}} \left( \prod_{j=1}^N s_j^{\frac{\lambda_j}{2}} \right) \exp \left( \frac{-\sum_{k=1}^N \lambda_k c(x_k, x)}{2} \right) \right) \\ &= \inf_{s \in \mathbb{R}_+} \left( \sum_{i=1}^N \lambda_i s_i + s - 2s^{\frac{1}{2}} \left( \prod_{j=1}^N s_j^{\frac{\lambda_j}{2}} \right) \exp \left( -\frac{1}{2} \inf_{x \in X} \sum_{k=1}^N \lambda_k c(x_k, x) \right) \right) \\ &= \sum_{i=1}^N \lambda_i s_i - \prod_{j=1}^N s_j^{\lambda_j} \exp \left( -\inf_{x \in X} \sum_{k=1}^N \lambda_k c(x_k, x) \right). \end{aligned}$$

For proof, see Section 6.1. The key point here is that the minimizer in  $s$  has an explicit formula, hence allowing us to obtain the last equality.

**3.2. Soft multi-marginal formulation.** The result of Lemma 3.2 allows us to characterise the new conic multi-marginal problem as a soft multi-marginal (SMM) problem on the space  $X^N$ , which is given by

$$(3.4) \quad \text{HK}_{\text{SMM}}(\vec{\mu}) := \inf_{\gamma \in \mathcal{M}(X^N)} \left\{ (\check{c}, \gamma) + \sum_i \lambda_i \mathcal{F}(\gamma_i \mid \mu_i) \right\},$$



where the cost function  $\tilde{c} : X^N \rightarrow \overline{\mathbb{R}}$  is given by

$$(3.5) \quad \tilde{c}(\vec{x}) = \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x).$$

We first prove that the soft multi-marginal problem has a solution.

PROPOSITION 3.3. *There exists  $\bar{\gamma} \in \mathcal{M}(X^N)$  such that*

$$\text{HK}_{\text{SMM}}(\vec{\mu}) = (\tilde{c}, \bar{\gamma}) + \sum_i \lambda_i \mathcal{F}(\bar{\gamma}_i \mid \mu_i).$$

For proof, see Section 6.2. The result follows from simple adjustments to the argument in [18, Theorem 3.3], which concerns the case  $N = 2$ .

This brings us to the first main result of the paper.

THEOREM 3.4. *The conic multi-marginal formulation of the barycenter problem  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  defined in (3.1) admits a soft multi-marginal reformulation given by (3.4). In other words,*

$$\overline{\text{HK}}_{\text{CMM}}(\vec{\mu}) = \text{HK}_{\text{SMM}}(\vec{\mu}).$$

Furthermore, if  $\bar{\gamma} \in \mathcal{M}(X^N)$  is a solution to  $\text{HK}_{\text{SMM}}(\vec{\mu})$ , then  $\bar{\alpha} \in \mathcal{M}(Y^N)$  given by

$$\bar{\alpha} := (x_1, \bar{\rho}_1(x_1), \dots, x_N, \rho_N(x_N))_{\#} \bar{\gamma},$$

where

$$\mu_i = \bar{\rho}_i \bar{\gamma}_i + \mu_i^{\perp},$$

is a solution to  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$ .

For proof, see Section 6.3.

**3.3. Constrained coupled two-marginal formulation.** It is perhaps tempting to think that the new conic multi-marginal formulation  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  defined in (3.1) simply coincides with the unconstrained coupled two-marginal barycenter problem  $\text{HK}_{\text{C2M}}(\vec{\mu})$  introduced in (2.14). This cannot be the case, however, since, on the one hand, through Theorem 3.4 we know that  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  is equivalent to the soft multi-marginal reformulation  $\text{HK}_{\text{SMM}}(\vec{\mu})$ , and on the other, through [13, Section 8], we know that the unconstrained problem  $\text{HK}_{\text{C2M}}(\vec{\mu})$  is not equivalent to it. It turns out, however, that the difference between  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  and  $\text{HK}_{\text{C2M}}(\vec{\mu})$  can be described quite explicitly, namely it lies in introducing an extra constraint in the coupled two-marginal formulation. We define a *constrained coupled two-marginal (CC2M) HK barycenter problem*, given by

$$(3.6) \quad \text{HK}_{\text{CC2M}}(\vec{\mu}) := \inf_{\nu \in \mathcal{M}(X)} \left\{ \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \nu) \right\},$$

where, in contrast to the usual formulation of the HK distance defined in (2.1), we have

$$\overline{\text{HK}}^2(\mu_i, \nu) := \inf_{\gamma \in \mathcal{M}(X \times X)} \left\{ (c, \gamma) + \mathcal{F}(\gamma_1 \mid \mu_i) + \overline{\mathcal{F}}(\gamma_2 \mid \nu) \right\},$$

and, reusing the general entropy functional notation (2.3),

$$(3.7) \quad \overline{\mathcal{F}}(\gamma_i \mid \mu_i) := \int_X \overline{F}(\sigma_i) d\mu_i + \overline{F}'_{\infty} \gamma_i^{\perp}(X), \quad \overline{F}(s) = \begin{cases} 0, & \text{if } s = 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

In other words,  $\overline{F}$  is the convex indicator  $\iota_{\{1\}}$  and thus we impose a hard constraint on the second marginal of  $\gamma$  to exactly coincide with  $\nu$ . This of course implies that  $\overline{\text{HK}}$  is no longer a metric, but nonetheless retains some features of the HK distance, as we still use the same cost function and the entropy functional for the first marginal remains the same. In particular, the one-sided hard-constraint setup is covered by the general theory of optimal entropy-transport problems developed in [18, Part I] – see Example E.8 in Section 3.3 therein. As a result, we can show that  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  admits at least one solution.

PROPOSITION 3.5. *There exists  $\bar{\nu} \in \mathcal{M}(X)$  such that*

$$\text{HK}_{\text{CC2M}}(\vec{\mu}) = \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \bar{\nu}).$$

For proof, see Section 6.4.

We further note that the idea of a one-sided hard constraint was pioneered in [4], albeit in a slightly different context (see Section 4.1 for an extended discussion).

This brings us to the second main result of the paper.

THEOREM 3.6. *The constrained coupled two-marginal formulation of the barycenter problem  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  defined in (3.6) admits a soft multi-marginal reformulation given by (3.4). In other words,*

$$\text{HK}_{\text{CC2M}}(\vec{\mu}) = \text{HK}_{\text{SMM}}(\vec{\mu}).$$

Furthermore, if  $\bar{\gamma} \in \mathcal{M}(X^N)$  is a solution to  $\text{HK}_{\text{SMM}}(\vec{\mu})$ , then either

(i)  $\bar{\gamma}$  is the zero measure on  $X^N$ , in which case the zero measure on  $X$  is a solution to  $\text{HK}_{\text{CC2M}}(\vec{\mu})$ ; or

(ii)  $\bar{\gamma}$  is not the zero measure on  $X^N$ , in which case  $\bar{\nu} = T_{\#}\bar{\gamma}$  is a solution to  $\text{HK}_{\text{CC2M}}(\vec{\mu})$ . Here  $T : A \rightarrow X$  is a mapping from

$$A := \left\{ \vec{x} = (x_1, \dots, x_N) \in X^N \mid \bigcap_{i=1}^N B_{\frac{\sigma}{2}}(x_i) \neq \emptyset \right\}$$

to the unique solution to

$$\inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x),$$

which exists for all  $\vec{x} \in A$ .

For proof, see Section 6.5.

We note that by combining Theorem 3.4 and Theorem 3.6 we thus obtain a three-way equality

$$\text{HK}_{\text{CC2M}}(\vec{\mu}) = \overline{\text{HK}}_{\text{CMM}}(\vec{\mu}) = \text{HK}_{\text{SMM}}(\vec{\mu}),$$

as opposed to the two-way equality for the unconstrained HK barycenter framework for which we have

$$\text{HK}_{\text{C2M}}(\vec{\mu}) = \text{HK}_{\text{CMM}}(\vec{\mu}).$$

See Section 4.2 for an extended discussion on how the two approaches differ.

**4. Discussion.** Before we proceed to the details of proofs, we finish the main part of the paper by discussing how problems  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  and  $\text{HK}_{\text{SMM}}(\vec{\mu})$  relate to the recently studied theory of unbalanced multi-marginal optimal transport [4] and also how the constrained framework compares with the unconstrained one. We finally also informally examine whether the constrained setup can be seen as a more natural extension of the Wasserstein barycenter problem to the unbalanced setting.

**4.1. Relation to the unbalanced multi-marginal optimal transport theory.** A result complimentary to Theorem 3.6 is proven in [4, Theorem 5.2] in the case when both problems are additionally entropy-regularised and for a much wider class of cost functions and entropy functionals. The key difference is that (a non-regularised counterpart to) the soft multi-marginal formulation considered therein is given by

$$(4.1) \quad \widehat{\text{HK}}_{\text{SMM}}(\vec{\mu}) := \inf_{\gamma \in \mathcal{M}(X^N \times X)} (\hat{c}, \gamma) + \sum_{i=1}^N \lambda_i \mathcal{F}(\gamma_i \mid \mu_i),$$

where

$$\hat{c}(\vec{x}, x_{N+1}) := \sum_{i=1}^N \lambda_i c(x_i, x_{N+1}),$$

meaning that an extra  $X$  space is introduced and the cost function is not defined as an infimum. Putting aside entropic regularisation, formally the non-regularised version of their result is that in fact

$$\widehat{\text{HK}}_{\text{SMM}}(\vec{\mu}) = \text{HK}_{\text{CC2M}}(\vec{\mu}).$$

In [4, Remark 5.5] and closing remarks the authors identify a result as in Theorem 3.6 as something desirable, as it would establish sparsity of the optimal plan in (4.1) in its last component. More broadly, such a result fits with wider literature on studying the sparsity of optimal plans in multi-marginal problems [15, 21, 22, 7, 6]. We show that, at least in the HK-related setting, it can be done. Our proof, which will be presented in Section 6.5, relies merely on using the reverse formulation to obtain the inequality  $\text{HK}_{\text{CC2M}}(\vec{\mu}) \geq \text{HK}_{\text{SMM}}(\vec{\mu})$  and, at least on the face of it, should apply to a wide class of entropy functions and cost functions (provided that the infimization in (3.5) admits a unique minimizer, or the infimum is equal to  $+\infty$ ).

We further note that our new conic multi-marginal formulation provides an explicit link between the results of [4] and the work on HK barycenters in [13, 11], which, to the best of our knowledge, is new.

*Example 4.1.* To showcase that the least-cost approach works in practice, we reuse the setup of the example about computing the barycenters of 1D unbalanced Gaussians presented in [4, Section 6.1]. We consider two empirical measures  $\mu_1$  and  $\mu_2$  obtained from sampling truncated normal distributions  $\mathcal{N}(0.2, 0.05)$  and  $2\mathcal{N}(0.8, 0.08)$  on  $[0, 1]$  using a uniform grid with 200 points. To showcase the fact that the least-cost unbalanced approach also makes sense for other cost functions, we use both the quadratic cost and the HK cost in our tests.

To approximately solve the soft multi-marginal problem  $\text{HK}_{\text{SMM}}(\vec{\mu})$ , we use the unbalanced Sinkhorn algorithm [10], as implemented in the POT (Python Optimal Transport) toolbox [12]. We publish a Jupyter notebook detailing the computation<sup>1</sup>.

The results are presented in Figure 1 and clearly the barycenters computed are qualitatively comparable the ones shown in the right column of [4, Figure 2].

<sup>1</sup>[https://github.com/mbuze/CHK\\_barycenters](https://github.com/mbuze/CHK_barycenters)

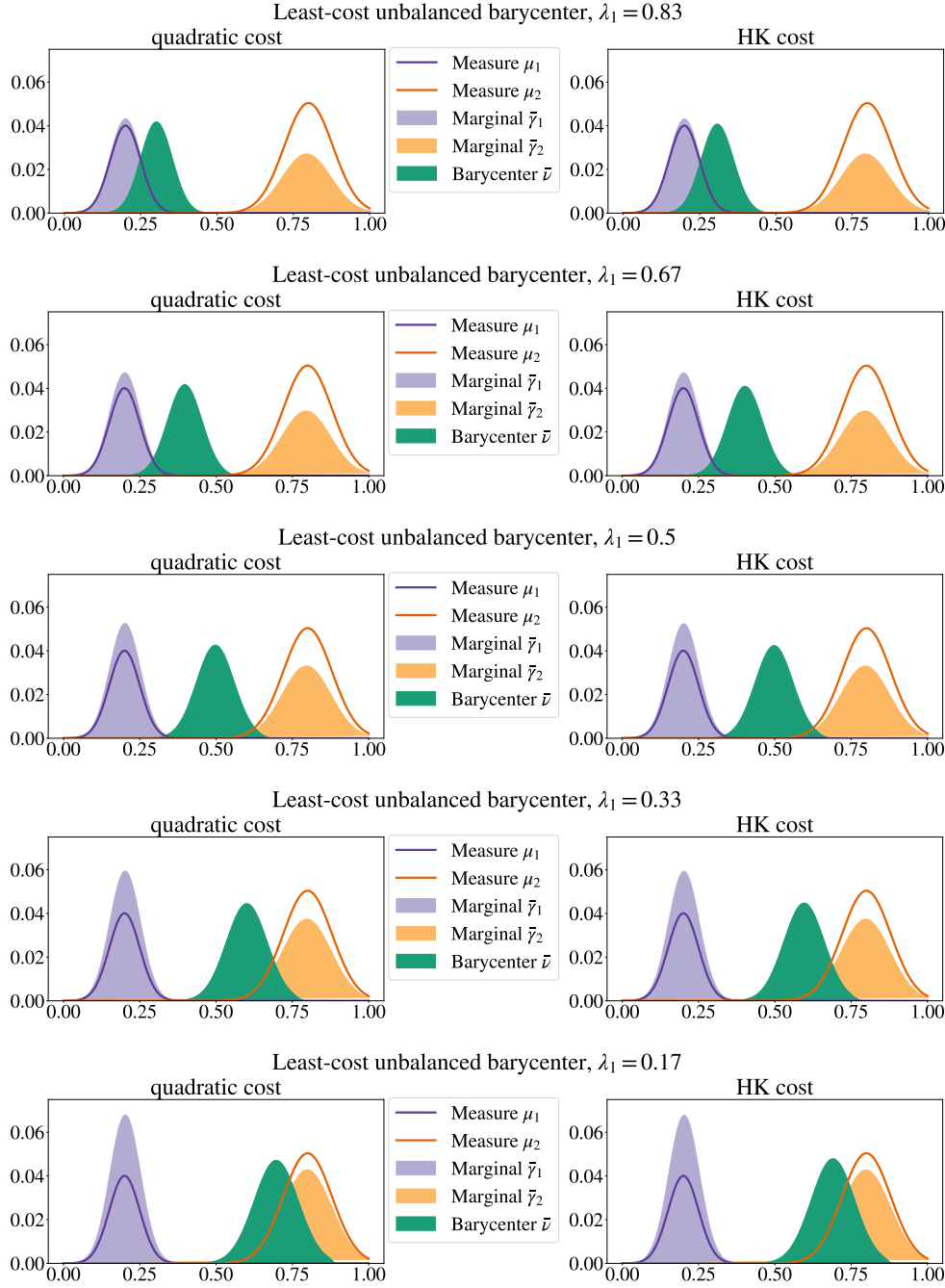


Figure 1: Computation of the least-cost unbalanced barycenter between measures  $\mu_1$  and  $\mu_2$  with weights  $\lambda_1$  and  $\lambda_2 = 1 - \lambda_1$ , obtained by approximately solving  $\text{HK}_{\text{SMM}}(\bar{\mu})$  defined in (3.4) using quadratic cost (left column) and HK cost (right column). We plot the input measures, the marginals of the optimal plan  $\bar{\gamma}$  and the resulting barycenter  $\bar{\nu}$ .

**4.2. Comparison with the unconstrained setting.** In this section we will gather several observations relating the constrained approach proposed in this paper with previous work on the unconstrained HK barycenters in [13, 11, 5].

We start with presenting two ways in which the difference between the two approaches can be described.

PROPOSITION 4.2. *It holds that*

$$\text{HK}_{\text{C2M}}(\vec{\mu}) \leq \text{HK}_{\text{CC2M}}(\vec{\mu}).$$

The equality can only be achieved if the measures  $\vec{\mu}$  are such there exists minimiser  $\bar{\nu}$  of  $\text{HK}_{\text{C2M}}(\vec{\mu})$  for which there exist minimizers  $\bar{\gamma}^i$  of  $\text{HK}^2(\mu_i, \bar{\nu})$  for which  $\bar{\gamma}_2^i = \bar{\nu}$  for each  $i$ .

*Proof.* This is an immediate consequence of the fact that any minimiser of  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  is a competitor in the unconstrained problem and further that  $F(s) \geq F(1) = 0$  for any  $s \in [0, +\infty)$ .  $\square$

A somewhat less trivial characterisation of the difference between the two approaches follows from the dual description of the conic cost functions.

LEMMA 4.3. *The joint multi-marginal perspective cost function  $\bar{H}_{\text{CMM}}$  defined (3.2) admits a dual description*

$$(4.2) \quad \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{x, s \in Y} \sup_{\phi_1, \dots, \phi_N, \phi_{N+1}} \left\{ \sum_{i=1}^N -s_i (\lambda_i F)^*(-\phi_i) - s F^*(-\phi_{N+1}) \right. \\ \left. \left| \sum_{i=1}^N \phi_i + \phi_{N+1} \leq \sum_{i=1}^N \lambda_i c(x_i, x) \right. \right\},$$

where  $\phi_i \in \mathbb{R}_+$ . The dual description of the unconstrained multi-marginal cost function  $H_{\text{CMM}}$  from (2.16) is on the other hand given by

$$H_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{x, s \in Y} \sum_{i=1}^N \sup_{\phi_{i,1}, \phi_{i,2}} \left\{ -s_i (\lambda_i F)^*(-\phi_{i,1}) - s (\lambda_i F)^*(-\phi_{i,2}) \right. \\ \left. \left| \phi_{i,1} + \phi_{i,2} \leq \lambda_i c(x_i, x) \right. \right\},$$

where  $\phi_{i,j} \in \mathbb{R}_+$ .

Based on this result, we can establish the following.

PROPOSITION 4.4. *It holds that*

$$H_{\text{CMM}}(\vec{x}, \vec{s}) \leq \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}),$$

which, as already established in Proposition 4.2, implies that

$$\text{HK}_{\text{CMM}}(\vec{\mu}) \leq \bar{\text{HK}}_{\text{CMM}}(\vec{\mu}).$$

*Proof.* Fix  $(x, s) \in Y$  and suppose  $(\bar{\phi}_{i,1}, \bar{\phi}_{i,2}) \in \mathbb{R}^2$  maximizes

$$\sup_{\phi_{i,1}, \phi_{i,2}} \left\{ -s_i (\lambda_i F)^*(-\phi_{i,1}) - s (\lambda_i F)^*(-\phi_{i,2}) \left| \phi_{i,1} + \phi_{i,2} \leq \lambda_i c(x_i, x) \right. \right\}.$$

We can then set  $\hat{\phi}_i := \bar{\phi}_{i,1}$  and  $\hat{\phi}_{N+1}^j := \frac{\bar{\phi}_{j,2}}{\lambda_j}$  for some  $j \in \{1, \dots, N\}$ , which defines a competitor in (4.2) and thus

$$\begin{aligned} \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) &\geq \inf_{(x,s) \in Y} \sum_{i=1}^N -s_i (\lambda_i F)^*(\hat{\phi}_i) - s F^*(-\hat{\phi}_{N+1}^j) \\ &= \inf_{(x,s) \in Y} \sum_{i=1}^N -s_i (\lambda_i F)^*(\bar{\phi}_{i,1}) - s F^*(-\bar{\phi}_{j,2}/\lambda_j) \end{aligned}$$

Since this holds for any particular choice of  $j$ , we can take a convex combination with weights  $\bar{\lambda}$  of the right-hand side and arrive at

$$\begin{aligned} \bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) &\geq \inf_{(x,s) \in Y} \sum_{j=1}^N \lambda_j \left( \sum_{i=1}^N -s_i (\lambda_i F)^*(\bar{\phi}_{i,1}) - s F^*(-\bar{\phi}_{j,2}/\lambda_j) \right) \\ &= \inf_{(x,s) \in Y} \sum_{i=1}^N -s_i (\lambda_i F)^*(\bar{\phi}_{i,1}) + \sum_{j=1}^N s \lambda_j F^*(-\bar{\phi}_{j,2}/\lambda_j) \\ &= H_{\text{CMM}}(\vec{x}, \vec{s}). \end{aligned}$$

The last equality follows from the fact that, for any  $j$ ,

$$-s (\lambda_j F)^*(-\bar{\phi}_{j,2}) = -s \lambda_j F^*\left(\frac{-\bar{\phi}_{j,2}}{\lambda_j}\right),$$

which is the standard scaling property of the Legendre transform.  $\square$

We further discuss an explicit example to emphasise that, at least for Dirac masses, the constrained approach differs from the unconstrained one quite considerably.

*Example 4.5* (Dirac masses). Consider  $N$  Dirac masses  $\mu_i = \delta_{z_i}$  for  $z_i \in X$  and we recall that we assume that  $X$  is (a compact subset of)  $\mathbb{R}^n$ . It readily follows from the analysis of the cost function  $\tilde{c}$  in Lemma 6.1 that if

$$(4.3) \quad \bigcap_{i=1}^N B_{\frac{\pi}{2}}(z_i) \neq \emptyset,$$

then the barycenter  $\bar{\nu}$  obtained by minimizing in  $\text{HK}_{\text{CC2M}}(\bar{\mu})$  as in Proposition 3.5, is given by a single Dirac mass

$$\bar{\nu} = M \delta_{T(\vec{z})},$$

where  $T(\vec{z})$  is the mapping

$$\vec{z} \mapsto \min_{z \in X} \sum_{i=1}^N \lambda_i c(z_i, z)$$

which is shown to exist in Lemma 6.1, provided that (4.3) holds. Furthermore, an explicit calculation reveals that the mass  $M$  is given by the formula

$$(4.4) \quad M = \prod_{i=1}^N m_i^{\lambda_i} \exp(-\tilde{c}(\vec{z})) = \prod_{i=1}^N m_i^{\lambda_i} \exp\left(-\sum_{j=1}^N \lambda_j c(z_j, T(\vec{z}))\right),$$

which can be seen as a variant of a weighted geometric mean for masses of input measures.

If, on the other hand, (4.3) does not hold, then the barycenter is given by the zero measure. On an informal level this is consistent with (4.4), since

$$M \rightarrow 0, \text{ as } \tilde{c}(\vec{z}) \rightarrow +\infty.$$

This example reveals that, for small distances between source points, the constrained HK barycenter behaves similarly to the Wasserstein barycenter and degenerates if the distance between the support of at least one of the input measures from the supports of other measures is too large. As thoroughly discussed in [5], a radically different, clustering behaviour can occur in the unconstrained setup. At the same time, it is feasible to imagine a (perhaps not very efficient) clustering algorithm based on the constrained approach, since we can interpret the barycenter being the zero measure as an indication that there are at least two clusters in the data set and we can tune what we consider a cluster by adjusting the length over which transport is preferred over mass creation/annihilation.

**4.3. Extending Wasserstein barycenters to the unbalanced setting.** Using the hard-constraint entropy functional defined in (3.7), the multi-marginal formulation of the Wasserstein barycenter problem, as studied in [1], can be formulated as

$$W_{\text{MM}}(\vec{\mu}) = \inf_{\gamma \in \mathcal{M}(X^N)} \left\{ (\tilde{c}, \gamma) + \sum_{i=1}^N \lambda_i \overline{\mathcal{F}}(\gamma_i \mid \mu_i) \right\}, \quad \tilde{c}(\vec{x}) := \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x),$$

where now the masses of each  $\mu_i$  have to coincide, and, canonically, we have  $c(x, y) = |x - y|^2$ . By comparing with the definition of  $\text{HK}_{\text{SMM}}(\vec{\mu})$  in (3.4), the soft multi-marginal formulation is a very intuitive extension of the Wasserstein barycenter problem to the unbalanced setting – we replace  $N$  hard marginal functionals  $\overline{\mathcal{F}}(\gamma_i \mid \mu_i)$  with  $N$  soft functionals  $\mathcal{F}(\gamma_i \mid \mu_i)$  defined in (2.3).

Similarly, the coupled-two-marginal formulation of the Wasserstein barycenter problem can be written as

$$W_{\text{C2M}}(\vec{\mu}) = \inf_{\nu \in \mathcal{M}(X)} \left\{ \sum_{i=1}^N \lambda_i \inf_{\gamma_i \in \mathcal{M}(X \times X)} \{ (c, \gamma_i) + \overline{\mathcal{F}}(\gamma_{i,1} \mid \mu_i) + \overline{\mathcal{F}}(\gamma_{i,2} \mid \nu) \} \right\}.$$

By directly comparing with the corresponding unconstrained setting, given by  $\text{HK}_{\text{C2M}}(\vec{\mu})$  and defined in (2.14), we see that the difference lies in replacing  $2N$  hard-constraint entropy functionals,  $\overline{\mathcal{F}}(\gamma_{i,1} \mid \mu_i)$  and  $\overline{\mathcal{F}}(\gamma_{i,2} \mid \nu)$ , with  $\mathcal{F}(\gamma_{i,1} \mid \mu_i)$  and  $\mathcal{F}(\gamma_{i,2} \mid \nu)$ .

On the other hand, in the constrained setting, given by  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  and defined in (3.6),  $N$  hard-constraint entropy functionals  $\overline{\mathcal{F}}(\gamma_{i,1} \mid \mu_i)$  get replaced with  $\mathcal{F}(\gamma_{i,1} \mid \mu_i)$ . In that sense the constrained setting can be seen as being half-way between the Wasserstein barycenter problem and the unconstrained HK barycenter problem. To a significant degree it is the hard constraint that  $\gamma_{i,2} = \nu$ , that ensures the link between the multi-marginal and coupled-two-marginal formulations for both the Wasserstein and constrained HK barycenter can be established. Our result indicate that we thus retain more of the Wasserstein barycenter framework by using the constrained approach.

**5. Conclusion.** We have introduced a new conic multi-marginal barycenter problem related to the Hellinger-Kantorovich metric, which admits a very natural soft multi-marginal formulation which in its spirit exactly matches its Wasserstein metric counterpart. Additionally, we have shown that it corresponds to a coupled-two-marginal formulation with a one-sided hard marginal constraint, which can be argued to be a more natural extension of the idea of a Wasserstein barycenter to the unbalanced setting. Furthermore, our analysis of the conic multi-marginal formulation has established an explicit link between the unbalanced multi-marginal optimal transport framework studied in [4] and conic formulation of the HK barycenter problem studied in [13, 11, 5]. We further have extended the results of [4] by showing that, as in the Wasserstein metric, the soft multi-marginal formulation can be posed over the space  $X^N$  with the cost function defined as "the least cost" [9].

In the future it would be very interesting to explore whether, based on the ideas developed here (see especially Section 4.3), the unconstrained HK barycenter problem admits a "generalised" soft-marginal formulation.

**Acknowledgments.** This work was inspired by a great talk on this topic given by Bernhard Schmitzer at the ICMS Workshop: Optimal Transport and Calculus of Variations in December 2023 in Edinburgh, whom I would like to sincerely thank for bringing my attention to this problem and for the discussion we have had on related topics.

## 6. Proofs.

### 6.1. Proof of Lemma 3.2.

*Proof.* Given the explicit form of  $R$  in (2.7), for a fixed  $x$ ,  $\vec{x}$ ,  $s$ , and  $\vec{s}$ , we can solve

$$\inf_{t>0} \sum_{i=1}^N \left\{ t \left( \lambda_i c(x_i, x) + \lambda_i R\left(\frac{s_i}{t}\right) + \lambda_i R\left(\frac{s}{t}\right) \right) \right\}$$

via a direct computation relying on differentiating in  $t$ . It readily follows that

$$\bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{(x,s) \in Y} \left( \sum_{i=1}^N \lambda_i s_i + s - 2s^{\frac{1}{2}} \left( \prod_{j=1}^N s_j^{\frac{\lambda_j}{2}} \right) \exp \left( \frac{-\sum_{k=1}^N \lambda_k c(x_k, x)}{2} \right) \right).$$

The next step is to recognise that since  $s_i \geq 0$  and  $s \geq 0$ , the monotonicity of the exponential function ensures that we also have

(6.1)

$$\bar{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \inf_{s \in \mathbb{R}_+} \left( \sum_{i=1}^N \lambda_i s_i + s - 2s^{\frac{1}{2}} \left( \prod_{j=1}^N s_j^{\frac{\lambda_j}{2}} \right) \exp \left( -\frac{1}{2} \sum_{k=1}^N \lambda_k \inf_{x \in X} c(x_k, x) \right) \right).$$

The infimum over  $s$  can then again be solved directly by differentiating in  $s$  and recognising that the optimal  $s$  is given by

$$s = \left( \left( \prod_{j=1}^N s_j^{\frac{\lambda_j}{2}} \right) \exp \left( \frac{-\inf_{x \in X} \sum_{k=1}^N \lambda_k c(x_k, x)}{2} \right) \right)^2.$$

By plugging this formula into (6.1) and simplifying, we arrive at the final reformula-



tion, given by

$$\overline{H}_{\text{CMM}}(\vec{x}, \vec{s}) = \sum_{i=1}^N \lambda_i s_i - \prod_{j=1}^N s_j^{\lambda_j} \exp \left( - \inf_{x \in X} \sum_{k=1}^N \lambda_k c(x_k, x) \right). \quad \square$$

### 6.2. Proof of Proposition 3.3.

*Proof.* Define  $\mathcal{E}(\cdot \mid \vec{\mu}) : \mathcal{M}(X^N) \rightarrow \overline{R}$  as

$$\mathcal{E}(\gamma \mid \vec{\mu}) := (\tilde{c}, \gamma) + \sum_{i=1}^N \lambda_i \mathcal{F}(\gamma_i \mid \mu_i).$$

The functional is feasible since when  $\gamma = 0$  (the zero measure), we have

$$\mathcal{E}(0 \mid \vec{\mu}) = \sum_{i=1}^N \mu_i(X) < \infty.$$

It is bounded below by zero, thus a minimising sequence  $(\gamma_k)_{k \in \mathbb{N}}$  exists. From the inequality

$$\mathcal{E}(\gamma \mid \vec{\mu}) \geq \gamma(X^N) \inf \tilde{c} + \sum_{i=1}^N \mu_i(X) F \left( \frac{\gamma(X^N)}{\mu_i(X)} \right) \geq \sum_{i=1}^N \mu_i(X) F \left( \frac{\gamma(X^N)}{\mu_i(X)} \right)$$

and the superlinearity of  $F$ , we can deduce that the masses of  $(\gamma_k)_{k \in \mathbb{N}}$  are bounded, thus the sequence has a weak-\* cluster point  $\bar{\gamma}$ , meaning that, up to a subsequence, for all continuous and bounded  $f$ ,

$$\int_{X^N} f(\vec{x}) d\gamma_k(\vec{x}) \rightarrow \int_{X^N} f(\vec{x}) d\bar{\gamma}(\vec{x}).$$

But it has to be a minimiser, since  $\tilde{c}$  and  $\mathcal{F}$  are both lower semi-continuous.  $\square$

### 6.3. Proof of Theorem 3.4.

*Proof.* The proof of this theorem closely resembles the results detailed in [8, Section 3], which themselves are inspired by the results presented in [18, Section 3]. Recall from (3.4) that the proposed soft multi-marginal (SMM) formulation is given by

$$\text{HK}_{\text{SMM}}(\vec{\mu}) = \inf_{\gamma \in \mathcal{M}(X^N)} \left\{ (\tilde{c}, \gamma) + \sum_i \lambda_i \mathcal{F}(\gamma_i \mid \mu_i) \right\}.$$

Our aim is to establish that  $\text{HK}_{\text{SMM}}(\vec{\mu}) = \overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$ . Using the reverse entropies  $R$  from (2.7) and the Lebesgue decomposition between  $\gamma_i$  and  $\mu_i$ ,

$$\gamma_i = \sigma_i \mu_i + \gamma_i^\perp, \quad \mu_i = \rho_i \gamma_i + \mu_i^\perp$$

it follows from an obvious adjustment of [18, Theorem 3.11] that the problem admits a reverse formulation, given by

$$\text{HK}_{\text{SMM}}(\vec{\mu}) = \inf_{\gamma \in \mathcal{M}(X^N)} \underbrace{\left\{ \int_{X^N} \left( \tilde{c}(\vec{x}) + \sum_{i=1}^N \lambda_i R(\rho_i(x_i)) \right) d\gamma + \sum_{i=1}^N \lambda_i (\mu_i - \rho_i \gamma_i)(X) \right\}}_{=:\mathcal{R}(\vec{\mu}, \gamma)},$$

where we have used the fact that  $R'_\infty = F(0) = 1$ .

Likewise, we can replicate the argument in [18, Proposition 4.3, Theorem 4.11] to conclude that the dual formulation is given by

$$\text{HK}_{\text{SMM}}(\vec{\mu}) = \sup_{\vec{\psi} \in \Psi} \left\{ \sum_{i=1}^N (\psi_i, \mu_i) \right\} = \sup_{\vec{\phi} \in \Phi} \left\{ \sum_{i=1}^N (-\lambda_i F)^*(-\phi_i), \mu_i \right\},$$

where

$$\vec{\psi} = (\psi_1, \dots, \psi_N), \quad \vec{\phi} = (\phi_1, \dots, \phi_N)$$

belong to

$$\Psi = \left\{ \vec{\psi} \mid \psi_i \in C_b(X), (\lambda_i R)^*(\psi_i) \in C_b(X), \bigoplus_{i=1}^N (\lambda_i R)^*(\psi_i) \leq \tilde{c} \right\},$$

$$\Phi = \left\{ \vec{\phi} \mid \phi_i \in C_b(X), -(\lambda_i F)^*(-\phi_i) \in C_b(X), \bigoplus_{i=1}^N \phi_i \leq \tilde{c} \right\}.$$

Note that  $(\lambda_i F)^*$  is the Legendre dual of  $\lambda_i F$ , which is defined, for a generic scalar function  $f$ , as

$$f^*(\phi) := \sup_{s>0} (s\phi - f(s)) \implies (\lambda_i f)^*(\phi) = \lambda_i f^* \left( \frac{\phi}{\lambda_i} \right).$$

In our specific case we have

$$F^*(\phi) = \exp(\phi) - 1, \quad R^*(\psi) = -\log(1 - \psi)$$

and we note the change of variables relation

$$\phi_i = (\lambda_i R)^*(\psi_i), \quad \psi_i = -(\lambda_i F)^*(-\phi_i).$$

Finally, we introduce the marginal perspective function

$$H_{\text{SMM}}(\vec{x}, \vec{s}) := \inf_{t>0} \left\{ t \left( \sum_{i=1}^N \lambda_i R \left( \frac{s_i}{t} \right) + \tilde{c}(\vec{x}) \right) \right\},$$

which, by a direct calculation, admits a formula

$$H_{\text{SMM}}(\vec{x}, \vec{s}) = \sum_{i=1}^N \lambda_i s_i - \prod_{j=1}^N s_j^{\lambda_j} \exp(-\tilde{c}(\vec{x})).$$

In other words, since  $\tilde{c}(\vec{x}) = \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x)$ , we have established that in fact  $H_{\text{SMM}} = \overline{H}_{\text{CMM}}$  from (3.2).

We also have a dual representation for  $H_{\text{SMM}}$ , namely

$$H_{\text{SMM}}(\vec{x}, \vec{s}) = \sup_{\vec{\psi}} \left\{ \sum_{i=1}^N s_i \psi_i \mid \sum_{i=1}^N (\lambda_i R)^*(\psi_i) \leq \tilde{c}(\vec{x}) \right\}$$

$$= \sup_{\vec{\phi}} \left\{ \sum_{i=1}^N -s_i (\lambda_i F)^*(-\phi_i) \mid \sum_{i=1}^N \phi_i \leq \tilde{c}(\vec{x}) \right\},$$

where now  $\phi_i$  and  $\psi_i$  are scalars.

We can thus introduce a homogeneous formulation, given by

$$(6.2) \quad \mathcal{H}(\vec{\mu} \mid \gamma) := \int_{X^N} H_{\text{SMM}}(\vec{x}, \rho_1(x_1), \dots, \rho_N(x_N)) d\gamma + \sum_{i=1}^N \lambda_i (\mu_i - \rho_i \gamma_i)(X)$$

and, using the strategy from [18, Theorem 5.5], it follows from the dual representation for  $H_{\text{SMM}}$  that

$$(6.3) \quad \underbrace{\sup_{\vec{\phi} \in \Phi} \left\{ \sum_{i=1}^N (-\lambda_i F)^*(-\phi_i, \mu_i) \right\}}_{=\text{HK}_{\text{SMM}}(\vec{\mu})} \leq \inf_{\gamma \in \mathcal{M}(X^N)} \mathcal{H}(\vec{\mu} \mid \gamma) \leq \underbrace{\inf_{\gamma \in \mathcal{M}(X^N)} \mathcal{R}(\vec{\mu} \mid \gamma)}_{=\text{HK}_{\text{SMM}}(\vec{\mu})},$$

which of course imply that the middle term is equal to  $\text{HK}_{\text{SMM}}(\vec{\mu})$  too.

The final step is to lift to the cone. This is done by defining

$$\inf_{\alpha \in S(\vec{\mu})} (H_{\text{SMM}}, \alpha),$$

where we recall from (2.17) and (2.12) that

$$S(\vec{\mu}) := \{\alpha \in \mathcal{M}(Y^N) \mid \mathbf{h}_i \alpha = \mu_i\}, \quad \mathbf{h}_i \alpha = \pi_{\#}^{x_i}(s_i \alpha).$$

This is the cone-equivalent of (6.2), especially since, as in (2.13), the singular part is zero due to the constraint on the homogeneous marginal in the set  $S(\vec{\mu})$ .

Obvious adjustments to the argument in [18, Theorem 5.8] allow us to conclude that in fact the sandwich inequality in (6.3) holds for the cone formulation too. Since we have already shown that  $H_{\text{SMM}} = \overline{H}_{\text{CMM}}$ , we thus obtain that

$$\text{HK}_{\text{SMM}}(\vec{\mu}) = \inf_{\alpha \in S(\vec{\mu})} (\overline{H}_{\text{CMM}}, \alpha) = \overline{\text{HK}}_{\text{CMM}}(\vec{\mu}),$$

which is what we set out to prove.

The fact that

$$\bar{\alpha} = (x_1, \bar{\rho}_1(x_1), \dots, x_N, \rho_N(x_N))_{\#} \bar{\gamma},$$

where

$$\mu_i = \bar{\rho}_i \bar{\gamma}_i + \mu_i^{\perp},$$

is a minimiser of  $\overline{\text{HK}}_{\text{CMM}}(\vec{\mu})$  is an immediate extension of [18, Theorem 5.8].  $\square$

#### 6.4. Proof of Proposition 3.5.

*Proof.* The argument is similar to the proof in [13, Proposition 4.2]. Define

$$J : \mathcal{M}(X) \rightarrow \overline{\mathbb{R}}$$

given by

$$J(\nu) := \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \nu).$$

The functional  $J$  is feasible, that is  $\inf_{\nu \in \mathcal{M}(X)} J(\nu) < \infty$ . This follows naturally from the fact that the zero measure is a competitor and

$$J(0) = \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, 0) = \sum_{i=1}^N \lambda_i \mu_i(X) < \infty.$$

The functional  $J$  is also bounded below. This follows from

$$J(\nu) = \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \nu) \geq \sum_{i=1}^N \lambda_i \text{HK}^2(\mu_i, \nu) \geq 0,$$

where the last inequality holds since  $\text{HK}$  is a distance. The first inequality follows from the fact that the minimizer of  $\overline{\text{HK}}(\nu_i, \mu)$  is a competitor in  $\text{HK}(\nu_i, \mu)$  and further that  $F(1) = 0$ .

Thus a minimising sequence  $(\nu_k)_{k \in \mathbb{N}}$  exists and from the lower bound

$$\overline{\text{HK}}^2(\mu, \nu) \geq \text{HK}^2(\mu, \nu) \geq (\sqrt{\mu(X)} - \sqrt{\nu(X)})^2,$$

we can conclude that  $\nu_k(X)$  is uniformly bounded, and hence the sequence has a weak-\* cluster point  $\bar{\nu}$ , meaning that, up to a subsequence, for all continuous and bounded  $f$ ,

$$\int_X f(x) d\nu_k(x) \rightarrow \int_X f(x) d\bar{\nu}(x).$$

At the same time, due to [18, Theorem 3.3], for each  $\nu_k$ , there exists a collection  $\{\gamma^{i,k}\}_{i=1}^N \subset \mathcal{M}(X \times X)$  of minimizers of  $\overline{\text{HK}}^2(\mu_i, \nu_k)$ . Using the notation

$$\mathcal{E}(\gamma \mid \nu, \mu) := (c, \gamma) + \mathcal{F}(\gamma_1 \mid \mu) + \overline{\mathcal{F}}(\gamma_2 \mid \nu),$$

we thus have

$$J(\nu_k) = \sum_{i=1}^N \lambda_i \mathcal{E}(\gamma^{i,k} \mid \mu_i, \nu_k).$$

In particular, the hard-constraint entropy functional  $\overline{\mathcal{F}}(\gamma_2^{i,k} \mid \nu_k)$  ensures that

$$\gamma^{i,k}(X \times X) = \gamma_2^{i,k}(X) = \nu_k(X),$$

implying that the masses of  $\{\gamma^{i,k}\}_{k \in \mathbb{N}}$  are uniformly bounded, thus ensuring that the sequence admits a weak-\* cluster point  $\bar{\gamma}^i$ , meaning that, up to a subsequence, for all continuous and bounded  $f$ ,

$$\int_{X \times X} f(x_1, x_2) d\gamma^{i,k}(x_1, x_2) \rightarrow \int_{X \times X} f(x_1, x_2) d\bar{\gamma}^i(x_1, x_2).$$

The final step is to realise that, up to choosing a subsequence, we have

$$\begin{aligned} \inf_{\nu \in \mathcal{M}(X)} J(\nu) &= \lim_{k \rightarrow \infty} J(\nu_k) = \lim_{k \rightarrow \infty} \sum_{i=1}^N \lambda_i \mathcal{E}(\gamma^{i,k} \mid \mu_i, \nu_k) \\ &\geq \sum_{i=1}^N \lambda_i \mathcal{E}(\bar{\gamma}^i \mid \mu_i, \bar{\nu}) \geq \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \bar{\nu}) \geq \inf_{\nu \in \mathcal{M}(X)} J(\nu). \end{aligned}$$

The first inequality follows from the fact that the cost function  $(c, \gamma^{i,k})$  and the entropy functional  $\mathcal{F}(\gamma_1^{i,k} \mid \mu_i)$  are both lower semi-continuous, whereas  $\overline{\mathcal{F}}(\gamma_2^{i,k} \mid \nu_k)$  is (trivially) jointly lower semi-continuous in both variables, (c.f. [18, Corollary 2.9]). The second inequality follows merely from the fact that  $\bar{\gamma}^i$  is a competitor. We have thus established that  $\bar{\nu}$  is a minimiser.  $\square$

**6.5. Proof of Theorem 3.6.** We begin by proving an auxiliary result concerning the characterisation of the cost function

$$\tilde{c}(\vec{x}) = \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x),$$

defined in (3.5).

LEMMA 6.1. *The function  $\tilde{c}$  attains a finite value only on the set*

$$A := \left\{ \vec{x} \in X^N \mid \bigcap_{i=1}^N B_{\frac{\pi}{2}}(x_i) \neq \emptyset \right\}$$

and is identically  $+\infty$  otherwise. Furthermore, if  $\vec{x} \in A$ , then there exists a unique minimizer of

$$\inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x)$$

and consequently there exists a map  $T : A \rightarrow X$  mapping  $\vec{x}$  to the minimizer.

*Proof.* The fact that  $\tilde{c}(\vec{x}) < \infty$  if and only if  $\vec{x} \in A$  follows directly from the definition of  $c$  and  $\tilde{c}$ . We note that  $A$  is convex. Now let us fix  $\vec{x} \in A$  and define  $f : B \rightarrow \mathbb{R}$ , where

$$B := \bigcap_{i=1}^N B_{\frac{\pi}{2}}(x_i), \quad f(x) := \sum_{i=1}^N \lambda_i c(x_i, x).$$

By design,  $f$  is finite on its domain and  $B$  is a non-empty convex set, since it is a finite intersection of balls. We can thus rewrite  $f$  as

$$f(x) = - \sum_{i=1}^N \lambda_i \log(\cos^2(|x_i - x|)) = - \log \left( \prod_{i=1}^N \cos^{2\lambda_i}(|x_i - x|) \right)$$

It can be readily shown that

$$g : B_{\frac{\pi}{2}}(0) \rightarrow \mathbb{R}, \quad g(x) = - \log(\cos^2(|x|))$$

is strictly convex, and thus so is  $f$ , since it is a sum of convex functions and  $B$  is convex. As a result, it has a unique minimiser, thus establishing existence of the mapping  $T$ .  $\square$

We can now proceed to proving the main theorem of this section.

*Proof of Theorem 3.6.* As in the corresponding proof in the Wasserstein metric [1], we will proceed in two steps, but first, in light of Lemma 6.1, we discuss the special case where

$$\bigcap_{i=1}^N B_{\frac{\pi}{2}}(\text{supp}\mu_i) = \emptyset,$$

where we use the notation

$$B_{\frac{\pi}{2}}(A) := \left\{ x \in X \mid d(x, A) < \frac{\pi}{2} \right\}, \quad d(x, A) := \inf_{y \in A} |x - y|.$$

This is the case where no transport of mass can occur. This is, on the one hand, because the minimizer of  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  is the zero measure, that is  $\bar{\nu} = 0$ , which is the consequence of the hard marginal constraint  $\bar{\gamma}_{i,2} = \bar{\nu}$ . As a result only the zero measure ensures that  $\overline{\text{HK}}^2(\mu_i, \nu)$  is finite for all  $i$ . Consequently, each two-marginal problem is minimized by the zero measure too, that is  $\bar{\gamma}_i = 0$ .

On the other hand, in the soft marginal formulation, the minimiser of  $\text{HK}_{\text{SMM}}(\vec{\mu})$  is given by  $\bar{\gamma} = 0$ , which follows from the fact that any feasible plan  $\gamma \in \mathcal{M}(X^N)$  has to satisfy  $\gamma_i \ll \mu_i$ , from which it follows that  $(\bar{c}, \gamma) = +\infty$  unless  $\gamma = 0$ . In this case the mapping  $T$  does not exist, but it does not need to be discussed since the pushforward of the zero measure is always the zero measure.

In what follows we thus focus on the other case, namely that

$$(6.4) \quad \bigcap_{i=1}^N B_{\frac{\pi}{2}}(\text{supp}\mu_i) \neq \emptyset.$$

**Step 1: showing that  $\text{HK}_{\text{CC2M}}(\vec{\mu}) \geq \text{HK}_{\text{SMM}}(\vec{\mu})$ .** Let  $\bar{\nu} \in \mathcal{M}(X)$  be a minimizer of  $\text{HK}_{\text{CC2M}}(\vec{\mu})$  (proven to exist in Proposition 3.5) and  $\bar{\gamma}_i \in \mathcal{M}(X \times X)$  be a minimizer of  $\overline{\text{HK}}^2(\mu_i, \bar{\nu})$ . By  $(\bar{\gamma}_i^{x_2})_{x_2 \in X}$  we denote the disintegration of  $\bar{\gamma}_i$  with respect to its second marginal, which, by construction, for each  $i$  is given by  $\bar{\gamma}_{i,2} = \bar{\nu}$ .

This allows us to introduce  $\hat{\gamma} \in \mathcal{M}(X^N)$  via

$$\int_{X^N} \phi(\vec{x}) d\hat{\gamma} := \int_{X^{N+1}} \phi(\vec{x}) d\bar{\gamma}_1^y(x_1), \dots, d\bar{\gamma}_N^y(x_N) d\bar{\nu}(y).$$

We note that the  $i$ th marginal of  $\hat{\gamma}$  is given by  $\hat{\gamma}_i = \bar{\gamma}_{i,1}$ , which follows from observing that

$$\begin{aligned} \int_X \phi(x) d\hat{\gamma}_i &= \int_{X^N} \phi(\pi^i(\vec{x})) d\hat{\gamma} = \int_{X \times X} \phi(x_i) d\bar{\gamma}_i^y(x_i) d\bar{\nu} = \int_{X \times X} \phi(x_1) d\bar{\gamma}_i \\ &= \int_X \phi(x) d\bar{\gamma}_{i,1}. \end{aligned}$$

This in particular naturally implies that in the Lebesgue decomposition between  $\hat{\gamma}_i$  and  $\mu_i$ ,

$$\mu_i = \hat{\rho}_i \hat{\gamma}_i + \mu_i^\perp,$$

we have  $\hat{\rho}_i = \bar{\rho}_{i,1}$ . As a result

$$\begin{aligned}
 \text{HK}_{\text{SMM}}(\vec{\mu}) &\leq (\tilde{c}, \hat{\gamma}) + \sum_{i=1}^N \lambda_i \mathcal{F}(\hat{\gamma}_i \mid \mu_i) \\
 &= \int_{X^N} \left( \tilde{c}(\vec{x}) + \sum_{i=1}^N \lambda_i R(\hat{\rho}_i(x_i)) \right) d\hat{\gamma} + \sum_{i=1}^N \lambda_i (\mu_i - \hat{\rho}_i \hat{\gamma}_i)(X) \\
 &= \int_{X^{N+1}} \left( \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x) + \sum_{i=1}^N \lambda_i R(\bar{\rho}_{i,1}(x_i)) \right) d\hat{\gamma} + \sum_{i=1}^N \lambda_i (\mu_i - \bar{\rho}_{i,1} \bar{\gamma}_{i,1})(X) \\
 &\leq \int_{X^{N+1}} \left( \sum_{i=1}^N \lambda_i c(x_i, y) + \sum_{i=1}^N \lambda_i R(\bar{\rho}_{i,1}(x_i)) \right) d\hat{\gamma} + \sum_{i=1}^N \lambda_i (\mu_i - \bar{\rho}_{i,1} \bar{\gamma}_{i,1})(X) \\
 &= \sum_{i=1}^N \left( \lambda_i \int_{X \times X} (c(x_i, y) + R(\bar{\rho}_{i,1}(x_i))) d\bar{\gamma}_i^y(x_i) d\bar{\nu}(y) \right) + \sum_{i=1}^N \lambda_i (\mu_i - \bar{\rho}_{i,1} \bar{\gamma}_{i,1})(X) \\
 &= \sum_{i=1}^N \left( \lambda_i \int_{X \times X} c(x_i, y) + R(\bar{\rho}_{i,1}(x_i)) d\bar{\gamma}_i + \lambda_i (\mu_i - \bar{\rho}_{i,1} \bar{\gamma}_{i,1})(X) \right) \\
 &= \sum_{i=1}^N \lambda_i (c, \bar{\gamma}_i) + \lambda_i \mathcal{F}(\bar{\gamma}_i \mid \mu_i) \\
 &= \sum_{i=1}^N \lambda_i \overline{\text{HK}}^2(\mu_i, \bar{\nu}) = \text{HK}_{\text{CC2M}}(\vec{\mu}).
 \end{aligned}$$

**Step 2: showing that  $\text{HK}_{\text{CC2M}}(\vec{\mu}) \leq \text{HK}_{\text{SMM}}(\vec{\mu})$ .** Let  $\bar{\gamma} \in \mathcal{M}(X^N)$  be the minimiser of  $\text{HK}_{\text{SMM}}(\vec{\mu})$  and  $T : A \rightarrow X$  be the mapping to the infimum in the cost function  $\tilde{c}$ , which exists due to the assumption on the support of measures  $\mu_1, \dots, \mu_N$  in (6.4), as detailed in Lemma 6.1. We thus have

$$\tilde{c}(\vec{x}) = \inf_{x \in X} \sum_{i=1}^N \lambda_i c(x_i, x) = \sum_{i=1}^N \lambda_i c(x_i, T(\vec{x})).$$

We can then set  $\hat{\gamma}_i := (\pi^i, T)_{\#} \bar{\gamma}$  and  $\hat{\nu} := T_{\#} \bar{\gamma}$ . It readily follows that

$$\hat{\gamma}_{i,1} = \bar{\gamma}_i, \quad \hat{\gamma}_{i,2} = \hat{\nu},$$

which allows us to argue that

$$\begin{aligned}
 \text{HK}_{\text{SMM}}(\vec{\mu}) &= \sum_{i=1}^N \lambda_i \left( \int_{X^N} c(x_i, T(\vec{x})) d\bar{\gamma} + \mathcal{F}(\bar{\gamma}_i \mid \mu_i) \right) \\
 &= \sum_{i=1}^N \lambda_i \left( (c, \hat{\gamma}_i) + \mathcal{F}(\hat{\gamma}_{i,1} \mid \mu_i) + \bar{\mathcal{F}}(\hat{\gamma}_{i,2} \mid \hat{\nu}) \right) \\
 &\geq \text{HK}_{\text{CC2M}}(\vec{\mu}),
 \end{aligned}$$

where the second equality follows from the fact that by construction  $\hat{\gamma}_{i,2} = \hat{\nu}$  and thus  $\bar{\mathcal{F}}(\hat{\gamma}_{i,2} \mid \hat{\nu}) = 0$ . The final inequality follows from the fact that  $\hat{\gamma}_i$  and  $\hat{\nu}$  are admissible competitors.  $\square$

## REFERENCES

- [1] M. AGUEH AND G. CARLIER, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43 (2011), pp. 904–924, <https://doi.org/10.1137/100805741>.
- [2] J. M. ALTSCHULER AND E. BOIX-ADSERA, *Polynomial-time algorithms for multimarginal optimal transport problems with structure*, Mathematical Programming, 199 (2023), pp. 1107–1178, <https://doi.org/10.1007/s10107-022-01868-7>.
- [3] F. A. BA AND M. QUELLMALZ, *Accelerating the Sinkhorn algorithm for sparse multi-marginal optimal transport via fast Fourier transforms*, Algorithms, 15 (2022), p. 311, <https://doi.org/10.3390/a15090311>.
- [4] F. BEIER, J. VON LINDHEIM, S. NEUMAYER, AND G. STEIDL, *Unbalanced multi-marginal optimal transport*, Journal of Mathematical Imaging and Vision, 65 (2023), pp. 394–413, <https://doi.org/10.1007/s10851-022-01126-7>.
- [5] M. BONAFINI, O. MINEVICH, AND B. SCHMITZER, *Hellinger–Kantorovich barycenter between Dirac measures*, ESAIM: Control, Optimisation and Calculus of Variations, 29 (2023), p. 19, <https://doi.org/10.1051/cocv/2022088>.
- [6] C. BRIZZI, G. FRIESECKE, AND T. RIED, *h-Wasserstein barycenters*, arXiv e-prints, (2024), <https://doi.org/10.48550/arXiv.2402.13176>.
- [7] C. BRIZZI, G. FRIESECKE, AND T. RIED, *p-Wasserstein barycenters*, arXiv e-prints, (2024), <https://doi.org/10.48550/arXiv.2405.09381>.
- [8] M. BUZE AND M. H. DUONG, *Entropic regularisation of unbalanced optimal transportation problems*, arXiv e-prints, (2023), <https://doi.org/10.48550/arXiv.2305.02410>.
- [9] G. CARLIER AND I. EKELAND, *Matching for teams*, Economic theory, 42 (2010), pp. 397–418, <https://doi.org/10.1007/s00199-008-0415-z>.
- [10] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, AND F.-X. VIALARD, *An interpolating distance between optimal transport and Fisher–Rao metrics*, Foundations of Computational Mathematics, 18 (2016), pp. 1–44, <https://doi.org/10.1007/s10208-016-9331-y>.
- [11] N.-P. CHUNG AND M.-N. PHUNG, *Barycenters in the Hellinger–Kantorovich space*, Applied Mathematics & Optimization, 84 (2021), pp. 1791–1820, <https://doi.org/10.1007/s00245-020-09695-y>.
- [12] R. FLAMARY, N. COURTY, A. GRAMFORT, M. Z. ALAYA, A. BOISBUNON, S. CHAMBON, L. CHAPEL, A. CORENFLOS, K. FATRAS, N. FOURNIER, L. GAUTHERON, N. T. GAYRAUD, H. JANATI, A. RAKOTOMAMONJY, I. REDKO, A. ROLET, A. SCHUTZ, V. SEGUY, D. J. SUTHERLAND, R. TAVENARD, A. TONG, AND T. VAYER, *Pot: Python optimal transport*, Journal of Machine Learning Research, 22 (2021), pp. 1–8, <http://jmlr.org/papers/v22/20-451.html>.
- [13] G. FRIESECKE, D. MATTHES, AND B. SCHMITZER, *Barycenters for the Hellinger–Kantorovich distance over  $\mathbb{R}^d$* , SIAM Journal on Mathematical Analysis, 53 (2021), pp. 62–110, <https://doi.org/10.1137/20M1315555>.
- [14] G. FRIESECKE AND M. PENKA, *The GenCol algorithm for high-dimensional optimal transport: general formulation and application to barycenters and Wasserstein splines*, SIAM Journal on Mathematics of Data Science, 5 (2023), pp. 899–919, <https://doi.org/10.1137/22M1524254>.
- [15] W. GANGBO AND A. ŚWIĘCH, *Optimal maps for the multidimensional monge-kantorovich problem*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 51 (1998), pp. 23–45.
- [16] I. HAASLER, A. RINGH, Y. CHEN, AND J. KARLSSON, *Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem*, SIAM Journal on Control and Optimization, 59 (2021), pp. 2428–2453, <https://doi.org/10.1137/20M1320195>.
- [17] S. KONDRATYEV, L. MONSAINGEON, AND D. VOROTNIKOV, *A new optimal transport distance on the space of finite Radon measures*, Advances in Differential Equations, 21 (2016), pp. 1117–1164, <https://doi.org/10.57262/ade/1476369298>.
- [18] M. LIERO, A. MIELKE, AND G. SAVARÉ, *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures*, Inventiones mathematicae, 211 (2018), pp. 969–1117, <https://doi.org/10.1007/s00222-017-0759-8>.
- [19] M. LIERO, A. MIELKE, AND G. SAVARÉ, *Fine properties of geodesics and geodesic  $\lambda$ -convexity for the Hellinger–Kantorovich distance*, Archive for Rational Mechanics and Analysis, 247 (2023), p. 112, <https://doi.org/10.1007/s00205-023-01941-1>.
- [20] G. MONGE, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l’Académie Royale des Sciences, (1781), pp. 666–704.
- [21] B. PASS, *On the local structure of optimal measures in the multi-marginal optimal transportation problem*, Calculus of Variations and Partial Differential Equations, 43 (2012),



- pp. 529–536, <https://doi.org/10.1007/s00526-011-0421-z>.
- [22] B. PASS, *Multi-marginal optimal transport and multi-agent matching problems: Uniqueness and structure of solutions*, Discrete and Continuous Dynamical Systems, 34 (2014), pp. 1623–1639, <https://doi.org/10.3934/dcds.2014.34.1623>.
- [23] B. PASS, *Multi-marginal optimal transport: theory and applications*, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 49 (2015), pp. 1771–1790, <https://doi.org/10.1051/m2an/2015020>.
- [24] B. PASS AND A. VARGAS-JIMÉNEZ, *A general framework for multi-marginal optimal transport*, Mathematical Programming, (2024), pp. 1–36, <https://doi.org/10.1007/s10107-023-02032-5>.
- [25] G. PEYRÉ AND M. CUTURI, *Computational optimal transport: with applications to data science*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607, <https://doi.org/10.1561/22000000073>.
- [26] F. SANTAMBROGIO, *Optimal transport for applied mathematicians*, Birkäuser, NY, 55 (2015), p. 94, <https://doi.org/10.1007/978-3-319-20828-2>.
- [27] N. G. TRILLOS, M. JACOBS, AND J. KIM, *The multimarginal optimal transport formulation of adversarial multiclass classification*, Journal of Machine Learning Research, 24 (2023), pp. 1–56, <https://www.jmlr.org/papers/v24/22-0698.html>.
- [28] C. VILLANI, *Optimal transport: old and new*, vol. 338, Springer, 2009, <https://doi.org/10.1007/978-3-540-71050-9>.