Applied Mathematics & Optimization (2025) 92:68
https://doi.org/10.1007/s00245-025-10351-6

ORIGINAL PAPER

Generalized Wasserstein Barycenters

Francesco Tornabene[1] · Marco Veneroni[2] · Giuseppe Savaré[3]

Received: 7 November 2024 / Accepted: 16 October 2025 / Published online: 7 November 2025

Abstract

We study the existence and uniqueness of the barycenter of a signed distribution of probability measures on a Hilbert space. The barycenter is found, as usual, as a minimum of a functional. In the case where the positive part of the signed measure is a singleton, we can show also uniqueness. In the one-dimensional case, we characterize the quantile function of the unique minimum as the orthogonal projection of the L^2-barycenter of the quantiles on the cone of nonincreasing functions in $L^2(0,1)$. Further, we provide a stability estimate in dimension one and a counterexample to uniqueness in $\mathbb{R}^2$. Finally, we address the consistency of the barycenters and we prove that barycenters of a sequence of approximating measures converge (up to subsequences) to a barycenter of the limit measure.

Keywords Optimal transport · Wasserstein distance · Barycenter · Extrapolation

Mathematics Subject Classification MSC Classification: 49J27 · 49J45 · 49Q22

✉ Marco Veneroni
marco.veneroni@unipv.it

Francesco Tornabene
francesco.tornabene@gssi.it

Giuseppe Savaré
giuseppe.savare@unibocconi.it

1 Gran Sasso Science Institute, L'Aquila, Italy

2 Department of Mathematics, University of Pavia, Pavia, Italy

3 Department of Decision Sciences and BIDSA, Bocconi University, Milano, Italy

1 Introduction and Main Results

In this paper, we introduce an extension of the barycenter problem studied by Agueh and Carlier [1] in the Wasserstein space. Let H be a separable Hilbert space and $\mathcal{P}_2(H)$ be the space of Borel probability measures on H with finite second moment, endowed with the 2-Wasserstein distance, denoted W_2. Let λ be a signed measure on the Borel σ-algebra of $\mathcal{P}_2(H)$, such that $\lambda(\mathcal{P}_2(H)) = 1$ (see, e.g., [2] for a comprehensive treatise of measure theory). We study the minimization of the functional

$$\mathcal{E}(\mu) := \int_{\mathcal{P}_2(H)} W_2^2(\mu, \nu)\, \mathrm{d}\lambda(\nu), \tag{1}$$

which we call the *generalized* population barycenter problem of the signed measure λ. When λ is concentrated on a discrete set $\{\nu_1, \dots, \nu_n\} \subset \mathcal{P}_2(H)$ and it is nonnegative, i.e.

$$\lambda_i \geq 0, \text{ for } i = 1, \dots, n, \qquad \sum_{i=1}^{n} \lambda_i = 1, \qquad \lambda = \sum_{i=1}^{n} \lambda_i \delta_{\nu_i},$$

we recover the usual barycenter functional introduced in [1]:

$$\mathcal{E}(\mu) = \sum_{i=1}^{n} \lambda_i W_2^2(\mu, \nu_i).$$

The extension of the barycenter, from n discrete measures to a continuous distribution of probability measures (i.e., a probability distribution, or ‘population‘, of probability distributions) was studied in [3–5]. In our contribution, we allow for a signed measure, i.e., for negative weights, as long as $\lambda(\mathcal{P}_2(H)) > 0$ (or, without loss of generality, equal to one). This condition is necessary since, even in the discrete case, if $x_i \in H$ and $\sum_{i=1}^{n} \lambda_i < 0$, then

$$\inf_{x \in H} \sum_{i=1}^{n} \lambda_i |x - x_i|^2 = -\infty.$$

We denote by $\lambda = \lambda^+ - \lambda^-$ the (unique) Hahn-Jordan decomposition of λ, by $|\lambda| = \lambda^+ + \lambda^-$ the total variation measure of λ, and by $m_2^2(\nu)$ the second moment of a measure $\nu \in \mathcal{P}_2(H)$:

$$m_2^2(\nu) := \int_H |x|^2 \mathrm{d}\nu(x) < \infty.$$

We prove the existence of a minimizer of (1) and, in the particular case where λ^+ is concentrated on a single measure in $\mathcal{P}_2(H)$, we are able to show uniqueness of the solution.

Theorem 1 *Let H be a real separable Hilbert space and let λ be a signed measure on $\mathcal{P}_2(H)$ with*

$$\lambda(\mathcal{P}_2(H)) = 1 \qquad \textit{and} \qquad \int_{\mathcal{P}_2(H)} m_2^2(\nu)\mathrm{d}|\lambda|(\nu) < \infty. \tag{2}$$

Let $\mathscr{E} : \mathcal{P}_2(H) \to \mathbb{R}$ be the functional defined in (1). Then:
(i) *There exists a solution $\bar{\mu} \in \mathcal{P}_2(H)$ of the problem*

$$\min_{\mu \in \mathcal{P}_2(H)} \mathscr{E}(\mu).$$

(ii) *If, in particular, supp(λ^+) is a singleton, then there exists a unique minimizer $\bar{\mu} \in \mathcal{P}_2(H)$ of $\mathscr{E}$.*

It should be noticed that uniqueness is false, in general, in the classical barycenter problem with positive weights in $H = \mathbb{R}^d$. Moreover, the bound on the quadratic moments of $|\lambda|$ in (2) is the minimal assumption that ensures that $\mathcal{E}$ is finite on $\mathcal{P}_2(H)$, but finite-energy sequences are not relatively compact in $\mathcal{P}_2(H)$ if H is infinite-dimensional. In order to recover the compactness of a minimizing sequence, which we need for the direct method of the calculus of variations, we make use of the weak topology on $\mathcal{P}_2(H)$, introduced in [6], which we denote $\mathcal{P}_2^w(H)$ - see Sect. 2.2. The same framework is also crucial in the proof of the lower-semicontinuity of $\mathcal{E}$.

Let $(\lambda_k)_{k\in\mathbb{N}}$ be a sequence of signed measures on $\mathcal{P}_2(H)$, such that

$$\lambda_k(\mathcal{P}_2(H)) = 1, \qquad \sup_{k\in\mathbb{N}} \int_{\mathcal{P}_2(H)} m_2^2(\nu)\,\mathrm{d}|\lambda_k|(\nu) < +\infty. \tag{3}$$

Since Theorem 1 asserts that each λ_k admits a generalized Wasserstein barycenter, it is natural to ask whether a sequence of barycenters of λ_k converges to a barycenter of λ, if $\lambda_k \to \lambda$, in a suitable sense. We can provide an answer to this question in the framework of Γ-convergence (see, e.g., [7]) on $\mathcal{P}_2^w(H)$.

Theorem 2 *Let $(\lambda_k)_{k\in\mathbb{N}}$ and λ be signed measures on $\mathcal{P}_2(H)$, satisfying (3). Assume that*

$$\lim_{k\to\infty} \int_{\mathcal{P}_2(H)} \varphi(\nu)\,\mathrm{d}\lambda_k^{\pm}(\nu) = \int_{\mathcal{P}_2(H)} \varphi(\nu)\,\mathrm{d}\lambda^{\pm}(\nu) \qquad \forall\varphi \in C_b^0(\mathcal{P}_2(H)), \tag{4}$$

$$\lim_{k\to\infty} \int_{\mathcal{P}_2(H)} m_2^2(\nu)\,\mathrm{d}|\lambda_k|(\nu) = \int_{\mathcal{P}_2(H)} m_2^2(\nu)\,\mathrm{d}|\lambda|(\nu). \tag{5}$$

Let $\mathcal{E}_k : \mathcal{P}_2(H) \to \mathbb{R}$ be defined as

$$\mathcal{E}_k(\mu) := \int_{\mathcal{P}_2(H)} W_2^2(\nu, \mu) \mathrm{d}\lambda_k(\nu).$$

Then

(i) $\mathcal{E}_k$ Γ*-converges to* $\mathcal{E}$ *in* $\mathcal{P}_2(H)$*, with respect to the convergence induced by the 2-Wasserstein distance, and also with respect to the weak topology of* $\mathcal{P}_2^w(H)$*;*

(ii) *If* μ_k *is a minimizer of* $\mathcal{E}_k$ *for all* $k \in \mathbb{N}$*, the sequence* $(\mu_k)_{k\in\mathbb{N}}$ *is precompact with respect to the weak topology of* $\mathcal{P}_2^w(H)$*;*

(iii) *As a consequence of (i) and (ii), if* $\overline{\mu} \in \mathcal{P}_2(H)$ *is a limit point in* $\mathcal{P}_2^w(H)$ *of a subsequence of minimizers* (μ_{k_j})*, then*

$$\lim_{j\to\infty} \mathcal{E}_{k_j}(\mu_{k_j}) = \mathcal{E}(\overline{\mu}) = \min_{\mu\in\mathcal{P}_2(H)} \mathcal{E}(\mu).$$

We remark that Γ-convergence with respect to the weak and the strong topologies is also known as Mosco-convergence (see [8]). When the minimizer μ of $\mathscr{E}$ is unique, Theorem 2-(ii) and -(iii) imply that if μ_k is a minimizer of $\mathcal{E}_k$, then $\mu_k \to \mu$ with respect to the weak topology of $\mathcal{P}_2^w(H)$.

In the case $H = \mathbb{R}$, we provide an explicit characterization of the minimizer. Indeed, in the 1-dimensional case any probability measure $\mu \in \mathcal{P}_2(\mathbb{R})$ can be represented by its quantile function (see, e.g., [9, Theorem 2.18]), i.e., the pseudo-inverse X_μ of its distribution function F_μ:

$$\begin{aligned} F_\mu(x) &:= \mu((-\infty, x]), & \forall\, x \in \mathbb{R}, \\ X_\mu(w) &:= \inf\{x : F_\mu(x) > w\}, & \forall\, w \in (0,1). \end{aligned}$$

The map $\mu \mapsto X_\mu$ is an isometry between $\mathcal{P}_2(\mathbb{R})$, endowed with the 2-Wasserstein distance, and the convex cone $\mathcal{K}$ of nondecreasing functions in the Hilbert space $L^2(0,1)$. We denote by $P_{\mathcal{K}}$ the orthogonal projection operator, $P_{\mathcal{K}} : L^2(0,1) \to \mathcal{K}$. Our main result, in the 1-dimensional case, is the following.

Theorem 3 *Let* λ *be a signed measure on* $\mathcal{P}_2(\mathbb{R})$*, satisfying*

$$\lambda(\mathcal{P}_2(\mathbb{R})) = 1 \qquad \text{and} \qquad \int_{\mathcal{P}_2(\mathbb{R})} m_2^2(\nu) \mathrm{d}|\lambda|(\nu) < \infty. \tag{6}$$

Then there exists a unique solution $\bar{\mu} \in \mathcal{P}_2(\mathbb{R})$ *of*

$$\inf_{\mu\in\mathcal{P}_2(\mathbb{R})} \int_{\mathcal{P}_2(\mathbb{R})} W_2^2(\mu, \nu) \mathrm{d}\lambda(\nu),$$

and it is characterized by

$$X_{\bar{\mu}} = P_{\mathcal{K}} \left(\int_{\mathcal{P}_2(\mathbb{R})} X_\nu \mathrm{d}\lambda(\nu) \right), \tag{7}$$

where $X_\nu \in \mathcal{K}$ *is the pseudo-inverse function of (the distribution function of)* ν*.*

1.1 Literature Review and Related Problems

1.1.1 Wasserstein Barycenters

The barycenter of two probability measures, with positive weights adding to 1, is well-known as McCann's interpolation [10], while a complete study of the Wasserstein barycenters of *n* measures in $\mathbb{R}^d$ was done in [1]. In the latter, in particular, it was shown the existence of a minimizer and its uniqueness, in the case where one of the ν_i's vanishes on small sets. In the present paper we do not require the absolute continuity of one of the measures, but in order to obtain uniqueness we need the positive part of λ to be concentrated on a single measure. In the one-dimensional case, if λ is concentrated on a discrete set, if all weights λ_i are positive, and all measures ν_i are nonatomic, our characterization (7) reduces to the one-dimensional formula in [1, Sect. 6.1].

The barycenter of a continuous distribution of probability measures on $\mathbb{R}^d$ (also called a *population* barycenter) was first studied in [3] (published as [4]), where the authors, exploiting a duality argument, prove existence, uniqueness, and characterize the barycenter for compactly supported measures, and then study the convergence of the empirical barycenter to its population counterpart as the number of measures tends to infinity. In [5] existence and consistence of the population barycenter is studied in a general geodesic space. In [11] several applications of Wasserstein barycenters with possibly negative weights were given, in the context of Fréchet regression, for the analysis of data from demography and neuroimaging.

1.1.2 Metric Extrapolation

The barycenter of two measures $\nu_0, \nu_1 \in \mathcal{P}_2(H)$ with weights $\alpha > 1$ and $\beta = 1 - \alpha < 0$, that is

$$\min_{\mu \in \mathcal{P}_2(H)} \alpha W_2^2(\mu, \nu_0) + \beta W_2^2(\mu, \nu_1), \tag{8}$$

was studied in [12] and [13] in the context of metric extrapolation for the Backward Differentiation Formula of order 2 (BDF2). This connection requires some explanation. Given a metric space (*X*, *d*) and a functional $\Phi : X \to \mathbb{R} \cup \{+\infty\}$, the gradient flow of Φ with respect to the metric *d* is formally given by the solutions $u : [0, T] \to X$ of

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = -\nabla \Phi(u(t)),$$

or by its time-discrete equivalent $(u_\tau^k) \in X$

$$\frac{u_\tau^k - u_\tau^{k-1}}{\tau} \approx -\nabla \Phi(u_\tau^k), \tag{9}$$

for a small time-step $\tau > 0$. A rigorous metric framework for the derivatives that appear in these equations was established in [14]. The minimizing movement scheme (also referred to as Implicit Euler method or Jordan-Kinderlehrer-Otto (JKO) stepping) provides, under suitable assumptions, a discrete approximation of the gradient flow equation, by solving a sequence of minimization problems defined by

$$u^k = \arg\min_{w \in X} \left\{ \frac{1}{2\tau} d^2(u_\tau^{k-1}, w) + \Phi(w) \right\}.$$

The BDF2 scheme is a second-order discretization scheme, well-known for ODEs in $\mathbb{R}^d$ since the 1950's (see, e.g., [15]), which was recently proposed to approximate gradient flows in metric spaces [12]. In the Euclidean setting, the time-discrete approximation (9) is substituted by

$$\frac{3u_\tau^k - 4u_\tau^{k-1} + u_\tau^{k-2}}{2\tau} \approx -\nabla\Phi(u_\tau^k),$$

which leads, in a JKO step, to the minimization of

$$\Psi(\tau, u_\tau^{k-2}, u_\tau^{k-1}; w) := \frac{1}{\tau} d^2(u_\tau^{k-1}, w) - \frac{1}{4\tau} d^2(u_\tau^{k-2}, w) + \Phi(w).$$

In [12] several examples of metric spaces (X, d) are given, such that a sequence of piecewise-constant interpolations of the discrete solutions (u_τ^k) converges, locally uniformly in time, to a solution $u \in \mathrm{AC}^2([0,\infty), X)$ of the gradient flow of Φ, in the sense of the following Evolutionary Variational Inequality (EVI)

$$\frac{1}{2} d^2(u(t), w) - \frac{1}{2} d^2(u(s), w) \leq \int_s^t \left[\Phi(w) - \Phi(u(r)) - \frac{\lambda}{2} d^2(u(r), w) \right] \mathrm{d}r,$$

for all $0 \leq s < t$, where $\lambda \in \mathbb{R}$ is the λ-convexity modulus of Ψ ([12, Theorem 5.1]). Among examples given in [12], we mention the following ones.

(a) In the case where X is a Hilbert space, with the distance induced by its norm, the distance part of the JKO functional Ψ is convex on straight segments and it is minimized by the weighted average of u_τ^{k-1} and u_τ^{k-2}, as in a classical barycenter problem with positive coefficients.
(b) In the case $(X, d) = (\mathcal{P}_2(H), W_2)$, minimizing the distance part of the JKO functional Ψ is a Wasserstein barycenter problem like (8), with two assigned measures (corresponding to u_τ^{k-2} and u_τ^{k-1}) and real coefficients. It should be remarked that the negative sign of one of the weights is crucial for the λ-convexity of Ψ and thus for the existence and uniqueness of a minimizer. Indeed, we are going to exploit exactly [12, Theorem 3.4] for the uniqueness part of our Theorem 1.

Finally, we mention that two different convex formulations for (8) were studied in [16], together with an efficient numerical scheme to compute the minimizers.

1.1.3 Sticky Particles

In [13, Remarks 4.6 and 4.13], an interesting connection was pointed out, between metric extrapolation and the one-dimensional pressureless Euler system

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0 \\ \partial_t(\rho v) + \partial_x(\rho v^2) = 0 \end{cases} \tag{10}$$

in $\mathbb{R} \times (0, +\infty)$, with initial density and velocity conditions $\rho_{|t=0} = \rho_0$, $v_{|t=0} = v_0$, for the evolution of a system of particles that share their trajectories after a collision (also called *"sticky particle system"* (SPS)) (see also [17]). Here, we exploit the characterization of the solutions to the (SPS) given in [18], in order to show a direct connection with the minimizer of a generalized Wasserstein barycenter functional. Let $(\rho_t, \rho_t v_t)$ be the solution of the (SPS) at time $t > 0$, corresponding to a discrete initial datum

$$\rho_0 = \sum_{i=1}^N m_i \delta_{x_i}, \qquad \rho_0 v_0 = \sum_{i=1}^N m_i v_i \delta_{x_i}, \qquad m_i > 0,\ \sum_{i=1}^N m_i = 1,\ x_i, v_i \in \mathbb{R}.$$

Denoting X_0 (resp. X_t) the pseudo-inverse function of ρ_0 (resp. ρ_t), by [18, Theorem 2.6 - II] it holds

$$X_t = P_{\mathcal{K}}(X_0 + tV_0), \tag{11}$$

where $\mathcal{K}$, as above, is the convex cone of nondecreasing functions in $L^2(0,1)$, $P_{\mathcal{K}}$ is the orthogonal projection operator on $\mathcal{K}$, and V_0 is the piecewise-constant function such that $V_0(x) = v_i$ if $X_0(x) = x_i$. Let $\delta > 0$ be the first collision time of the evolution, then, for all $0 < s \leq \delta$

$$X_s = P_{\mathcal{K}}(X_0 + sV_0) = X_0 + sV_0,$$

and owing to (11), for all $t > s$

$$\begin{aligned} X_t &= P_{\mathcal{K}}(X_0 + tV_0) \\ &= P_{\mathcal{K}}\left(X_0 + t\frac{X_s - X_0}{s}\right) \\ &= P_{\mathcal{K}}\left(\left(1 - \frac{t}{s}\right) X_0 + \frac{t}{s} X_s\right). \end{aligned}$$

By (7), we immediately see that the solution $\rho_t = (X_t)_\sharp \mathcal{L}^1 \llcorner_{(0,1)}$ of the sticky particle system (10) at time t is also the minimizer of the generalized barycenter functional

$$\mathscr{E}(\rho) := \left(1 - \frac{t}{s}\right) W_2^2(\rho_0, \rho) + \frac{t}{s} W_2^2(\rho_s, \rho),$$

where the fixed measures are the initial density ρ_0 of the (SPS), with negative weight $1 - t/s$ and the solution of the system at time s, that is, ρ_s, with positive weight t/s.

1.2 Plan of the Paper

In Sect. 2, we recall the definitions and the main results we need concerning the Wasserstein distance, the weak topologies in spaces of probability measures, and signed measures. In Sect. 3 we study the one-dimensional case. After recalling Hoeffding's Fréchet's isometry, we state and prove the result concerning existence, uniqueness, and characterization of generalized barycenters for probability measures on the real line. In Subsect. 3.1, we provide a stability result and in Subsect. 3.2 an example where the generalized barycenter between one-dimensional Gaussian distributions is not a Gaussian. In Sect. 4 we prove the existence part of Theorem 1, for measures on a separable Hilbert space, while in Sect. 5 we prove uniqueness, in the case where λ^+ is concentrated on one single measure. In Subsect. 5.2, we provide an example of non-uniqueness. Finally, in Sect. 6, we prove Theorem 2.

2 Setting and Notation

We collect here the minimal definitions needed to state our problem. For an in-depth treatise on optimal transport and Wasserstein distances we refer to the textbooks [14] and [19].

2.1 Wasserstein Distance

We denote by H a real separable Hilbert space, endowed with the scalar product ' ' and the norm $|u| = \sqrt{u \cdot u}$. $\mathcal{P}(H)$ is the space of Borel probability measures μ on H and $\mathcal{P}_2(H)$ the subspace of measures μ with finite second moments:

$$m_2^2(\mu) := \int_H |x|^2 \mathrm{d}\mu(x) < +\infty. \tag{12}$$

Given metric spaces (X_1, d_1) and (X_2, d_2), for a general Borel map $f : X_1 \to X_2$ and a Borel measure μ on X_1, the push-forward measure of μ through f is defined as

$$f_\sharp\mu(A) := \mu(f^{-1}(A)), \quad \text{for every Borel set } A \subseteq X_2.$$

For $i = 1, 2$, let $\pi^i : H \times H \to H$ denote the projection operator on the i^{th} variable, $\pi^i(x_1, x_2) = x_i$. Given two measures $\mu, \nu \in \mathcal{P}_2(H)$, the set of admissible transport plans between μ and ν is

$$\Gamma(\mu,\nu) := \left\{\gamma \in \mathcal{P}(H\times H) : \pi^1_\sharp\gamma = \mu,\ \pi^2_\sharp\gamma = \nu\right\}.$$

The Wasserstein-Rubinstein-Kantorovich distance of order 2 between two measures $\mu, \nu \in \mathcal{P}_2(H)$ is defined as

$$W_2(\mu,\nu) := \min\left\{\left(\int_{H\times H} |x-y|^2 \mathrm{d}\gamma(x,y)\right)^{\frac{1}{2}} : \gamma \in \Gamma(\mu,\nu)\right\}.$$

We remark that W_2 satisfies the axioms of a distance on $\mathcal{P}_2(H)$: $W_2(\mu,\nu) = 0$ if and only if $\mu = \nu$, and W_2 satisfies the triangle inequality (symmetry and nonnegativity follow from the other two properties).

We denote by $\Gamma_{\mathrm{o}}(\mu,\nu)$ the set of *optimal transport plans* between $\mu, \nu \in \mathcal{P}_2(H)$. That is, if $\gamma \in \Gamma_{\mathrm{o}}(\mu,\nu)$, then $\gamma \in \Gamma(\mu,\nu)$ and

$$W_2^2(\mu,\nu) = \int_{H\times H} |x-y|^2 \mathrm{d}\gamma(x,y).$$

Definition 1 *(Narrow convergence)* We say that a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(H)$ is narrowly convergent to $\mu \in \mathcal{P}(H)$ if

$$\lim_{n\to+\infty} \int_H f(x)\,\mathrm{d}\mu_n(x) = \int_H f(x)\,\mathrm{d}\mu(x) \qquad \forall f \in C_b(H).$$

We will need the following Gluing Lemma ([14], Lemma 5.3.2)

Lemma 1 Let $\nu_0, \nu_1, \nu_2 \in \mathcal{P}_2(H)$ and $\gamma^{0\,i} \in \Gamma(\nu_0, \nu_i)$ for $i = 1, 2$. Then there exists $\boldsymbol{\gamma} \in \mathcal{P}(H\times H\times H)$ such that

$$(\pi^0, \pi^i)_\sharp \boldsymbol{\gamma} = \gamma^{0\,i} \quad \text{for } i = 1, 2.$$

An example of such a $\boldsymbol{\gamma}$ is given by the measure whose disintegration with respect to ν is

$$\boldsymbol{\gamma} := \int_X \left(\gamma^{0\,1}_{\nu_0} \times \gamma^{0\,2}_{\nu_0}\right) \mathrm{d}\nu_0,$$

where $\gamma^{0\,i} := \int_{\mathbb{R}^d} \gamma^{0\,i}_{\nu_0} \mathrm{d}\nu_0$ are the disintegrations of the transport plans $\gamma^{0\,i}$ with respect to ν_0.

2.2 Weak Topologies

In the proof of the existence of a minimizer for $\mathcal{E}$ we are going to use the framework of weak and strong-weak topologies of measures described in [6]. Let $C_2^w(H)$ be the Banach space of sequentially weakly continuous functions $\zeta : H \to \mathbb{R}$ satisfying

$$\lim_{|x|\to\infty} \frac{\zeta(x)}{1+|x|^2} = 0,$$

endowed with the norm

$$\|\zeta\|_{C_2^w(H)} := \sup_{x\in H} \frac{\zeta(x)}{1+|x|^2}.$$

We define $\mathcal{P}_2^w(H)$ to be the topological space of measures in $\mathcal{P}_2(H)$, endowed with the initial topology $\sigma(\mathcal{P}_2(H), C_2^w(H))$, i.e., the coarsest topology such that all functions

$$F_\zeta : \mathcal{P}_2(H) \to \mathbb{R}, \qquad F_\zeta(\mu) := \int_H \zeta(x)\mathrm{d}\mu(x),$$

with $\zeta \in C_2^w(H)$, are continuous.

In order to highlight the role of the different topologies, here we denote by H_s the space H, in order to emphasize that we are using the strong topology, and H_w the same space endowed with the weak topology. Let $Z = H_s \times H_w$ and

$$\mathcal{P}_2(Z) := \left\{\gamma \in \mathcal{P}(Z) : \int_Z \left(|x|^2 + |y|^2\right) \mathrm{d}\gamma(x,y) < +\infty\right\}.$$

We consider the space $C_2^{sw}(Z)$ of test functions $\zeta : Z \to \mathbb{R}$ such that:

- ζ is sequentially continuous with respect to the strong-weak product topology on Z;
- $\forall\varepsilon > 0\ \exists A_\epsilon \geq 0\ \ |\zeta(x,y)| \leq A_\varepsilon \left(1+|x|^2\right) + \varepsilon|y|^2$.

We endow $C_2^{sw}(Z)$ with the norm

$$\|\zeta\|_{C_2^{sw}(Z)} := \sup_{(x,y)} \frac{|\zeta(x,y)|}{1+|x|^2+|y|^2}.$$

We endow $\mathcal{P}_2(Z)$ with the initial topology induced by the functions

$$\gamma \mapsto \int_Z \zeta(x,y)\,\mathrm{d}\gamma(x,y), \qquad \zeta \in C_2^{sw}(Z);$$

we call $\mathcal{P}_2^{sw}(Z)$ the topological space $(\mathcal{P}_2(Z), \sigma(\mathcal{P}_2(Z), C_2^{sw}(Z)))$. Proposition 3.4 (a) in [6], adapted to our simplified setting, reads as follows.

Proposition 4 A sequence $(\gamma_n)_{n\in\mathbb{N}} \subset \mathcal{P}_2^{sw}(Z)$ and a measure $\gamma \in \mathcal{P}_2^{sw}(Z)$ satisfy

(i) $\gamma_n \rightharpoonup \gamma$ narrowly in $\mathcal{P}(Z)$;
(ii) $\lim_{n\to\infty} \int_Z |x|^2\,\mathrm{d}\gamma_n(x,y) \to \int_Z |x|^2\,\mathrm{d}\gamma(x,y)$;

(iii) $$\sup_{n\in\mathbb{N}} \int_Z |y|^2 \,\mathrm{d}\gamma_n(x,y) < +\infty,$$

if and only if

$\gamma_n \to \gamma \quad \text{in} \quad \mathcal{P}_2^{sw}(Z)$. Proposition 3.4 (b), in the particular case $Z = H_w$, yields the following result.

Proposition 5 For every constant $c < \infty$ the sets

$$\mathcal{K}_c := \left\{ \mu \in \mathcal{P}_2^w(H) : \int_H |x|^2 \mathrm{d}\mu(x) \le c \right\}$$

are compact and metrizable in $\mathcal{P}_2^w(H)$. In particular, they are sequentially compact. The following result [6, Corollary 3.6] is an immediate consequence of Propositions 4 and 5.

Corollary 1 It holds

(i) A sequence $(\mu_n)_{n\in\mathbb{N}}$ converges to μ in $\mathcal{P}_2^w(H)$ if and only if

$$(\mu_n)_{n\in\mathbb{N}} \text{ converges narrowly in } \mathcal{P}(H_w) \quad \text{and} \quad \sup_{n\in\mathbb{N}} \int_H |x|^2 \mathrm{d}\mu_n(x) < \infty.$$

(ii) A set $\mathcal{K} \subset \mathcal{P}_2(H)$ is relatively sequentially compact in $\mathcal{P}_2^w(H)$ if and only if

$$\sup_{\mu\in\mathcal{K}} \int_H |x|^2 \mathrm{d}\mu(x) < \infty.$$

2.3 Signed Measures

We refer to [2] for a comprehensive treatise of measure theory. Here, we will only use the following notions. Let $\mathcal{B}(X)$ be the Borel σ-algebra on the metric space (X, d), we denote by $\mathcal{M}^s(X)$ the space of finite countably additive signed measures on $\mathcal{B}(X)$. By Hahn's decomposition, there exist disjoint sets $\mathcal{A}^+, \mathcal{A}^- \in \mathcal{B}(X)$ such that $\mathcal{A}^+ \cup \mathcal{A}^- = X$ and for all $B \in \mathcal{B}(X)$

$$\lambda^+(B) := \lambda(\mathcal{A}^+ \cap B) \ge 0, \qquad \lambda^-(B) := -\lambda(\mathcal{A}^- \cap B) \ge 0.$$

Then $\lambda = \lambda^+ - \lambda^-$ is the (unique) Hahn-Jordan decomposition of λ and the total variation measure is defined as $|\lambda| = \lambda^+ + \lambda^-$. For $f \in L^1(|\lambda|)$, we set

$$\int_X f(x)\mathrm{d}\lambda(x) := \int_X f(x)\mathrm{d}\lambda^+(x) - \int_X f(x)\mathrm{d}\lambda^-(x).$$

It is immediate to check that

$$\left|\int_X f(x)\mathrm{d}\lambda(x)\right| \le \int_X |f(x)|\mathrm{d}\lambda^+(x) + \int_X |f(x)|\mathrm{d}\lambda^-(x) = \int_X |f(x)|\,\mathrm{d}|\lambda|(x). \quad (13)$$

We denote the variation of λ by

$$M(|\lambda|) := \lambda^+(X) + \lambda^-(X) = |\lambda|(X). \quad (14)$$

Finally, for $X = \mathcal{P}_2(H)$, the squared quadratic moment of $|\lambda|$ is

$$M_2^2(|\lambda|) := \int_{\mathcal{P}_2(H)} m_2^2(\nu)\mathrm{d}|\lambda|(\nu). \quad (15)$$

3 The One-Dimensional Case

In this section, we restrict to the one-dimensional setting because in this case there is a standard isometry between $\mathcal{P}_2(\mathbb{R})$ and the subset $\mathcal{K}$ of (essentially) nondecreasing functions in the Hilbert space of square-integrable functions $L^2(0,1)$. This relation allows us to exploit the properties of scalar products and projections in Hilbert spaces, which are crucial in our proof. We denote by $\|\cdot\|$ the standard norm in $L^2(0,1)$:

$$\|f\| := \left(\int_0^1 f^2(x)\mathrm{d}x\right)^{\frac{1}{2}}.$$

Let $\mu \in \mathcal{P}_2(\mathbb{R})$, the cumulative distribution function of μ is the function $F_\mu : \mathbb{R} \to [0,1]$

$$F_\mu(x) := \mu((-\infty, x]) \qquad \forall\, x \in \mathbb{R}.$$

Its pseudo-inverse (also called: 'monotone arrangement' or 'quantile function') is the nondecreasing function $X_\mu : (0,1) \to \mathbb{R}$

$$X_\mu(w) := \inf\{x : F_\mu(x) > w\} \qquad \forall\, w \in (0,1).$$

Observe that $X_\mu(F_\mu(x)) \ge x$ and that, if $m > F_\mu(x)$, then $X_\mu(m) > x$. Therefore, denoting by $\mathcal{L}^1$ the Lebesgue measure on (0, 1):

$$(X_\mu)_\#\mathcal{L}^1((-\infty, x]) = \mathcal{L}^1(X_\mu^{-1}((-\infty, x])) = \mathcal{L}^1((0, F_\mu(x)]) = F_\mu(x).$$

In other words, $(X_\mu)_\#\mathcal{L}^1 = \mu$.

Moreover, given two measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, $\gamma = (X_\mu, X_\nu)_\# \mathcal{L}^1$ is the unique monotone transport in $\Gamma(\mu, \nu)$; it follows that $\gamma \in \Gamma_o(\mu, \nu)$ ([20, Theorem 2.9]). Therefore

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}^2} |x-y|^2 \mathrm{d}(X_\mu, X_\nu)_\# \mathcal{L}^1 = \int_0^1 |X_\mu(x) - X_\nu(x)|^2 \mathrm{d}x = \|X_\mu - X_\nu\|^2. \tag{16}$$

We have shown that the mapping $\phi : \mathcal{P}_2(\mathbb{R}) \to L^2(0,1)$, $\phi(\mu) = X_\mu$ satisfies the following properties (see also the *Hoeffding-Fréchet* characterization of distributions with given marginals [21, Sect. 3.1])):

(i) for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ $\|\phi(\mu) - \phi(\nu)\| = W_2(\mu, \nu)$,
(ii) for all $f \in \mathcal{K}$ the measure $\mu = f_\sharp \mathcal{L}^1 \in \mathcal{P}_2(\mathbb{R})$ satisfies $\phi(\mu) = f$.

Therefore, the metric spaces $(\mathcal{P}_2(\mathbb{R}), W_2)$ and $(\mathcal{K}, \|\cdot\|)$ are isometric. We recall that the orthogonal projection operator $P_\mathcal{K} : L^2(0,1) \to \mathcal{K}$ can be characterized by the following property ([22], Theorem 5.2): for all $f \in L^2(0,1)$, $P_\mathcal{K}(f)$ is the unique element in $\mathcal{K}$ such that

$$\|P_\mathcal{K}(f) - f\| \le \|g - f\|, \qquad \forall g \in \mathcal{K}. \tag{17}$$

Moreover, the orthogonal projection does not increase distance ([22, Proposition 5.3]):

$$\|P_\mathcal{K}(f) - P_\mathcal{K}(g)\| \le \|f - g\|, \qquad \forall f, g \in L^2(0,1). \tag{18}$$

If $\lambda \in \mathcal{M}^s(\mathcal{P}_2(\mathbb{R}))$, $f \in \mathcal{K}$, and $\phi : \mathcal{P}_2(\mathbb{R}) \to L^2(0,1)$ is the Hoeffding-Fréchet isometry, then $\eta := \phi_\sharp \lambda \in \mathcal{M}^s(L^2(0,1))$ and

$$\int_{L^2(0,1)} \|f - g\|^2 \mathrm{d}\eta(g) = \int_{\mathcal{P}_2(\mathbb{R})} \|f - \phi(\nu)\|^2 \mathrm{d}\lambda(\nu) = \int_{\mathcal{P}_2(\mathbb{R})} W_2^2(\phi^{-1}(f), \nu)^2 \mathrm{d}\lambda(\nu).$$

In particular,

$$M_2^2(|\lambda|) < +\infty \quad \text{if and only if} \quad M_2^2(|\eta|) = \int_{L^2(0,1)} \|g\|^2 \mathrm{d}|\eta|(g) < +\infty.$$

We can therefore give the following equivalent statement of Theorem 3.

Theorem 6 *Let η be a signed measure on $L^2(0,1)$, such that supp$(\eta) \subset \mathcal{K}$,*

$$\eta(L^2(0,1)) = 1 \qquad \text{and} \qquad M_2(|\eta|) < \infty.$$

Then, there exists a unique minimizer $f^* \in \mathcal{K}$ of the functional

$$E : L^2(0,1) \to \mathbb{R}, \qquad E(f) := \int_{L^2(0,1)} \|f-g\|^2 \mathrm{d}\eta(g)$$

and it is characterized by

$$f^* = P_{\mathcal{K}}\left(\int_{L^2(0,1)} g \, \mathrm{d}\eta(g)\right).$$

The solution of Theorem 3 in the Wasserstein space can then be recovered by

$$\mu^* = \phi^{-1}(f^*) = f^*_\sharp \mathcal{L}^1.$$

Remark 1 Since the proof of Theorem 6 only relies on the Hilbert structure of the space $L^2(0,1)$ and on the closedness and convexity properties of $\mathcal{K}$, the same characterization of the solution holds in a more general setting. Precisely: let $(X, |\cdot|)$ be a real Hilbert space and let $K \subset X$ be a closed and convex subset. Let η be a signed measure on X, such that

$$\eta(X) = 1 \qquad \text{and} \qquad M_2(|\eta|) < \infty \tag{19}$$

(we do not need that supp$(\eta) \subset K$). Then there exists a unique minimizer $x^* \in K$ of the functional

$$G : X \to \mathbb{R}, \qquad G(x) := \int_X |x-y|^2 \mathrm{d}\eta(y)$$

and it is characterized by

$$x^* = P_K\left(\int_X y \, \mathrm{d}\eta(y)\right).$$

Remark 2 Theorem 3.1 in [18] gives a useful characterization of the orthogonal projection $P_{\mathcal{K}} : L^2(0,1) \to \mathcal{K}$, which may be employed for explicit computations: given $f \in L^2(0,1)$, let $F(t) = \int_0^t f(s)\mathrm{d}s$ be its integral function; let F^{**} be the lower semi-continuous convex envelope of F, i.e., the greatest lower semi-continuous convex function which is lower or equal to F. F^{**} may also be defined by

$$F^{**}(t) := \sup\{at + b : a, b \in \mathbb{R}, \ as + b \leq F(s) \text{ for a.e. } s \in (0,1)\}.$$

Being convex, in every point F^{**} admits a left derivative $\frac{d^-}{\mathrm{d}t}F^{**}$ and a right derivative $\frac{\mathrm{d}^+}{\mathrm{d}t}F^{**}$. The projection may then be characterized by

$$P_{\mathcal{K}}(f)(t) = \frac{\mathrm{d}^+}{\mathrm{d}t}F^{**}(t).$$

Proof of Theorem 6 We adopt the more general setting introduced in Remark 1: here $(X, |\cdot|)$ is a real Hilbert space and $K \subset X$ is a closed and convex subset. Let

$$\bar{x} := \int_X y \,\mathrm{d}\eta(y) \qquad \text{and} \qquad \mathcal{C}(\eta) := \int_X |y|^2 - |\bar{x}|^2 \mathrm{d}\eta(y).$$

These quantities are finite and only depend on η since, by (19),

$$|\bar{x}| \le \int_X |y| \,\mathrm{d}|\eta|(y) \le \frac{1}{2}\int_X \left(1 + |y|^2\right) \mathrm{d}|\eta|(y) = \frac{1}{2}\left(M(|\eta| + M_2^2(|\eta|))\right),$$
$$|\mathcal{C}(\eta)| \le M_2^2(|\eta|) + |\bar{x}|^2.$$

We compute

$$\begin{aligned} G(x) &= \int_X |x - y|^2 \mathrm{d}\eta(y) \\ &= \int_X |x|^2 - 2x \cdot y + |y|^2 \mathrm{d}\eta(y) \\ &= |x|^2 - 2x \cdot \bar{x} + \int_X |y|^2 \mathrm{d}\eta(y) \\ &= |x|^2 - 2x \cdot \bar{x} + |\bar{x}|^2 + \int_X |y|^2 \mathrm{d}\eta(y) - |\bar{x}|^2 \\ &= |x - \bar{x}|^2 + \mathcal{C}(\eta). \end{aligned}$$

Therefore, using the projection property (17) and exploiting the independence of $\mathcal{C}(\eta)$ from x, for all $z \in K$

$$\begin{aligned} G\left(P_K(\bar{x})\right) &= |P_K(\bar{x}) - \bar{x}|^2 + \mathcal{C}(\eta) \\ &\le |z - \bar{x}|^2 + \mathcal{C}(\eta) \\ &= G(z). \end{aligned}$$

We conclude that $P_K(\bar{x})$ is the unique minimum of G in K. □

3.1 The Discrete Case and Stability

The barycenter of a finite number of measures with positive and negative weights is a particular case of Theorem 3. Precisely, let $n \ge 2$ be a given integer. For $i = 1, \ldots, n$, let $\nu_i \in \mathcal{P}_2(\mathbb{R})$ be given and let λ_i be real numbers such that $\sum_{i=1}^n \lambda_i = 1$. We consider the signed measure:

$$\lambda := \sum_{i=1}^n \lambda_i \delta_{\nu_i},$$

where δ_{ν_i} is the Dirac measure concentrated in ν_i. Then, by Theorem 3 there exists a unique solution $\bar{\mu} \in \mathcal{P}_2(\mathbb{R})$ of

$$\inf_{\mu \in \mathcal{P}_2(\mathbb{R})} \sum_{i=1}^{n} \lambda_i W_2^2(\nu_i, \mu),$$

and it is characterized by

$$X_{\bar{\mu}} = P_{\mathcal{K}} \left(\sum_{i=1}^{n} \lambda_i X_{\nu_i} \right), \tag{20}$$

where $X_{\nu_i} \in \mathcal{K}$ is the pseudo-inverse function of (the distribution function of) ν_i.

The explicit characterization of the discrete Wasserstein barycenter in one dimension entails a stability result, with respect to perturbations of the fixed measures:

Lemma 2 Let $\nu_1, \dots, \nu_n$ and $\tilde{\nu}_1, \dots, \tilde{\nu}_n$ be probability measures in $\mathcal{P}_2(\mathbb{R})$; let $\lambda_1, \dots, \lambda_n$ be weights such that $\sum_{i=1}^{n} \lambda_i = 1$ and consider the barycenters μ, $\tilde{\mu}$ of the measures $\nu_1, \dots, \nu_n$ and $\tilde{\nu}_1, \dots, \tilde{\nu}_n$, with weights $\lambda_1, \dots, \lambda_n$. Then

$$W_2(\mu, \tilde{\mu}) \le \sum_{i=1}^{n} |\lambda_i| W_2(\nu_i, \tilde{\nu}_i).$$

Proof For ρ in $\mathcal{P}_2(\mathbb{R})$, let X_ρ be its monotone rearrangement. Then, using (16), (20), and the non-expanding property of the projection (18), we obtain

$$\begin{aligned} W_2(\mu, \tilde{\mu}) &= \left\| P_{\mathcal{K}} \left(\sum_{i=1}^{n} \lambda_i X_{\nu_i} \right) - P_{\mathcal{K}} \left(\sum_{i=1}^{n} \lambda_i X_{\tilde{\nu}_i} \right) \right\| \\ &\le \left\| \sum_{i=1}^{n} \lambda_i X_{\nu_i} - \sum_{i=1}^{n} \lambda_i X_{\tilde{\nu}_i} \right\| \\ &\le \sum_{i=1}^{n} |\lambda_i| \, \| X_{\nu_i} - X_{\tilde{\nu}_i} \| \\ &= \sum_{i=1}^{n} |\lambda_i| W_2(\nu_i, \tilde{\nu}_i). \end{aligned}$$

□

3.2 An Example in the Case of Two Gaussian Measures

While the (standard) barycenter of two Gaussian measures is always a Gaussian, it is possible to choose the parameters of the measures and a negative weight so that the generalized barycenter of two Gaussians is a Dirac delta.

Remark 3 If μ_1, μ_2, μ_3 are three probability measures on $\mathbb{R}$ and μ_2 is the barycenter of μ_1 and μ_3 with parameters $0 < \lambda < 1$ and $1 - \lambda$, then μ_3 is the generalized barycenter for μ_1 and μ_2 with parameters $(\lambda - 1)/\lambda$, $1/\lambda$ (and, likewise, μ_1 is a generalized barycenter for μ_2, μ_3 with parameters $1/(1 - \lambda)$ and $-\lambda/(1 - \lambda)$). This remark follows directly from the equation

$$X_{\mu_2} = (1 - \lambda)X_{\mu_1} + \lambda X_{\mu_3}$$

since then

$$X_{\mu_3} = \frac{1}{\lambda}X_{\mu_2} + \frac{\lambda - 1}{\lambda}X_{\mu_1}$$

and no projection on $\mathcal{K}$ is needed (despite the fact that $(\lambda - 1)/\lambda < 0$).

Let $\mu_1 \sim \mathcal{N}(m_1, \sigma_1^2)$, $\mu_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ be two Gaussian measures, and suppose that $\sigma_1 > \sigma_2$: then it is possible to choose $\bar{z} \in \mathbb{R}$ such that μ_2 is the barycenter between μ_1 and the Dirac delta $\delta_{\bar{z}}$. Owing to Remark 3, $\delta_{\bar{z}}$ is then the generalized barycenter between the Gaussians μ_1 and μ_2. In order to choose $\bar{z}$, we proceed as follows. For $z \in \mathbb{R}$, the monotone rearrangement of the Wasserstein barycenter $\bar{\mu}$ between μ_1 and an atomic measure δ_z satisfies

$$X_{\bar{\mu}} = (1 - \lambda)X_{\mu_1} + \lambda z.$$

Using the relation $(aX + b)^{-1}(x) = X^{-1}((x - b)/a)$, valid for all invertible functions $X : (0, 1) \to \mathbb{R}$ and all a, b in $\mathbb{R}$ with $a \neq 0$, we obtain

$$F_{\bar{\mu}}(x) = F_{\mu_1}((x - \lambda z)/(1 - \lambda)),$$

so that the density of $\bar{\mu}$ is given by

$$F'_{\bar{\mu}} = \frac{1}{1 - \lambda} f_{\mu_1}((x - \lambda z)/(1 - \lambda)) = K \exp\left(-\frac{(x - ((1 - \lambda)m_1 + \lambda z))^2}{2((1 - \lambda)\sigma_1)^2}\right)$$

where K doesn't depend on x, so $\bar{\mu} \sim \mathcal{N}((1 - \lambda)m_1 + \lambda z, ((1 - \lambda)\sigma_1)^2)$. Now we look for particular values $\bar{\lambda}$ and $\bar{z}$ for which $\bar{\mu} = \mu_2$: setting

$$\begin{cases} (1 - \bar{\lambda})m_1 + \bar{\lambda}\bar{z} = m_2, \\ (1 - \bar{\lambda})\sigma_1 = \sigma_2, \end{cases}$$

yields

$$\bar{\lambda} = \frac{\sigma_1 - \sigma_2}{\sigma_1}, \quad \bar{z} = \frac{\sigma_1 m_2 - \sigma_2 m_1}{(\sigma_1 - \sigma_2)}.$$

With this choice of parameters, we have

$$X_{\mu_2} = (1 - \bar{\lambda}) X_{\mu_1} + \bar{\lambda}\bar{z},$$

and therefore, owing to Remark 3, the Dirac measure $\delta_{\bar{z}}$ is the generalized barycenter of the Gaussian measures μ_1 and μ_2, with weights $1/\bar{\lambda} > 0$ and $(\bar{\lambda} - 1)/\bar{\lambda} < 0$.

4 Proof of Theorem 1-(i)

In this section we employ the direct method of the calculus of variations to prove the existence part of Theorem 1. The crucial point is that the functional $\mathcal{E}$ has a negative term, due to λ^-, so that lower-semicontinuity of the integrand is not sufficient. In order to overcome this difficulty, in Lemma 4 we prove continuity of the coupling term in $W_2(\nu, \mu)$. The main tool is the framework of strong-weak topologies in $\mathcal{P}_2(H)$, introduced in [6] and recalled in Sect. 2.2.

Lemma 3 (Lower bound)

Let m_2, M, and M_2 be defined as in (12), (14), and (15). For all $\mu \in \mathcal{P}_2(H)$

$$\mathcal{E}(\mu) \geq \frac{1}{2} m_2^2(\mu) - (1 + 2M(|\lambda|)) M_2^2(|\lambda|). \tag{21}$$

Proof For $\mu, \nu \in \mathcal{P}_2(H)$ and $\gamma \in \Gamma_o(\mu, \nu)$

$$\begin{aligned} W_2^2(\mu, \nu) &= \int_{H \times H} |x - y|^2 \mathrm{d}\gamma(x, y) \\ &= \int_H |x|^2 \mathrm{d}\mu(x) + \int_H |y|^2 \mathrm{d}\nu(y) - 2 \int_{H \times H} x \cdot y \, \mathrm{d}\gamma(x, y). \end{aligned}$$

Therefore, for all $\delta > 0$, by Young's inequality

$$\begin{aligned} \left| W_2^2(\mu, \nu) - m_2^2(\mu) \right| &\leq m_2^2(\nu) + 2 \left| \int_{H \times H} x \cdot y \, \mathrm{d}\gamma(x, y) \right| \\ &\leq m_2^2(\nu) + \int_{H \times H} \frac{|x|^2}{\delta} + \delta |y|^2 \, \mathrm{d}\gamma(x, y) \\ &= \frac{m_2^2(\mu)}{\delta} + (1 + \delta) m_2^2(\nu). \end{aligned}$$

Using $\lambda(\mathcal{P}_2(H)) = 1$, (13), and the last inequality, we obtain

$$
\begin{aligned}
\mathcal{E}(\mu) &= \int_{\mathcal{P}_2(H)} W_2^2(\mu,\nu)\mathrm{d}\lambda(\nu) \\
&= m_2^2(\mu) + \int_{\mathcal{P}_2(H)} \left(W_2^2(\mu,\nu) - m_2^2(\mu)\right)\mathrm{d}\lambda(\nu) \\
&\geq m_2^2(\mu) - \int_{\mathcal{P}_2(H)} \left(\frac{m_2^2(\mu)}{\delta} + (1+\delta)m_2^2(\nu)\right)\mathrm{d}|\lambda|(\nu) \\
&= \left(1 - \frac{M(|\lambda|)}{\delta}\right) m_2^2(\mu) - (1+\delta)M_2^2(|\lambda|).
\end{aligned}
$$

Choosing $\delta = 2M(|\lambda|)$ we obtain the thesis. $\square$
Let

$$
\begin{aligned}
\mathcal{R}(\nu,\mu) :&= W_2^2(\nu,\mu) - \int_H |x|^2 \mathrm{d}\nu(x) - \int_H |y|^2 \mathrm{d}\mu(y) \\
&= \min_{\gamma \in \Gamma(\nu,\mu)} \left\{ \int_{H\times H} -2x\cdot y \,\mathrm{d}\gamma(x,y) \right\}.
\end{aligned}
$$

In the next Lemma, we prove the continuity of $\mathcal{R}$ with respect to the strong-weak convergence of measure.

Lemma 4 (Continuity of $\mathcal{R}$) Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_2(H)$ be a sequence converging to $\mu \in \mathcal{P}_2(H)$ with respect to the (weak) topology of $\mathcal{P}_2^w(H)$ and let $(\nu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_2(H)$ be a sequence converging to $\nu \in \mathcal{P}_2(H)$ with respect to the (strong) topology of $\mathcal{P}_2(H)$. Then,

$$
\lim_{n\to\infty} \mathcal{R}(\nu_n,\mu_n) = \mathcal{R}(\nu,\mu). \tag{22}
$$

Proof For each couple (ν_n,μ_n), there exists $\gamma_n \in \Gamma_o(\mu_n,\nu_n)$ such that

$$
\mathcal{R}(\nu_n,\mu_n) = \int_{H\times H} -2x\cdot y \,\mathrm{d}\gamma_n(x,y).
$$

We are going to show that $\gamma_n \to \gamma \in \Gamma_o(\nu,\mu)$ narrowly in $\mathcal{P}(Z)$, in order to apply Proposition 4 to $\mathcal{R}(\nu_n,\mu_n)$. First, we notice that by Corollary 1-(i),

$$
\sup_{n\in\mathbb{N}} m_2^2(\mu_n) < +\infty. \tag{23}
$$

Since $(\pi^1_\sharp\gamma_n) = (\nu_n)$ is strongly converging in $\mathcal{P}_2(H)$ (and thus tight in $\mathcal{P}(H_s)$) and $(\pi^2_\sharp\gamma_n) = (\mu_n)$ is tight in $\mathcal{P}(H_w)$ by (23), then (γ_n) is tight in $\mathcal{P}(H_s \times H_w)$. Indeed, since (ν_n) is tight in $\mathcal{P}(H_s)$, there exists a lower-semicontinuous function $\psi : H \to [0,+\infty]$, with strongly compact sublevels, such that

$$\sup_{n\in\mathbb{N}}\int_H \psi(x)\mathrm{d}\nu_n(x) = S < +\infty.$$

Thus

$$\sup_{n\in\mathbb{N}}\int_{H\times H} \left(\psi(x)+|y|^2\right)\mathrm{d}\gamma_n(x,y) \le S + \sup_{n\in\mathbb{N}} m_2^2(\mu_n) < +\infty.$$

Since the sublevel sets of $\psi(x)+|y|^2$ are compact in $H_s\times H_w$, (γ_n) is tight in $\mathcal{P}(H_s\times H_w)$. Let γ be any narrow limit point of (γ_n). Then for all $f\in C_b(H_s)$ and for all $g\in C_b(H_w)$

$$\int_{H\times H} f(x)\mathrm{d}\gamma(x,y) = \lim_{n\to\infty}\int_{H\times H} f(x)\mathrm{d}\gamma_n(x,y) = \lim_{n\to\infty}\int_H f(x)\mathrm{d}\nu_n(x) = \int_H f(x)\mathrm{d}\nu(x)$$

and, by Corollary 1-(i),

$$\int_{H\times H} g(y)\mathrm{d}\gamma(x,y) = \lim_{n\to\infty}\int_{H\times H} g(y)\mathrm{d}\gamma_n(x,y) = \lim_{n\to\infty}\int_H g(y)\mathrm{d}\mu_n(y) = \int_H g(y)\mathrm{d}\mu(y).$$

Therefore, $\gamma\in\Gamma(\nu,\mu)$. Furthermore $\gamma\in\Gamma_{\mathrm{o}}(\mu,\nu)$ by (([6], Theorem 3.8). Since

$$\lim_{n\to\infty}\int_Z |x|^2\,\mathrm{d}\gamma_n(x,y) = \lim_{n\to\infty}\int_H |x|^2\,\mathrm{d}\nu_n(x) = \int_H |x|^2\,\mathrm{d}\nu(x) = \int_Z |x|^2\,\mathrm{d}\gamma(x,y)$$

and, by (23),

$$\int_Z |y|^2\,\mathrm{d}\gamma_n(x,y) = \int_H |y|^2\,\mathrm{d}\mu_n(y) = m_2^2(\mu_n) < +\infty,$$

by Proposition 4 we conclude that $\gamma_n\to\gamma$ in $\mathcal{P}_2^{sw}(Z)$. The function

$$\zeta: Z\to\mathbb{R},\qquad \zeta(x,y) := -2x\cdot y$$

is sequentially continuous with respect to the strong-weak product topology on Z and by Young's inequality $\forall\varepsilon>0$

$$|\zeta(x,y)| \le A_\varepsilon\left(1+|x|^2\right)+\varepsilon|y|^2,$$

with $A_\varepsilon = 1/\varepsilon$. Therefore, exploiting the optimality of γ_n and γ,

$$\lim_{n\to\infty}\mathcal{R}(\nu_n,\mu_n)=\lim_{n\to\infty}\left(W_2^2(\nu_n,\mu_n)-\int_H|x|^2\mathrm{d}\nu_n(x)-\int_H|y|^2\mathrm{d}\mu_n(y)\right)$$
$$=\lim_{n\to\infty}\int_{H\times H}-2x\cdot y\,\mathrm{d}\gamma_n(x,y)$$
$$=\int_{H\times H}-2x\cdot y\,\mathrm{d}\gamma(x,y)$$
$$=W_2^2(\nu,\mu)-\int_H|x|^2\mathrm{d}\nu(x)-\int_H|y|^2\mathrm{d}\mu(y)$$
$$=\mathcal{R}(\mu,\nu).$$

□

We have now all the elements to prove existence of a minimizer for $\mathscr{E}$.

Proof of Theorem 1-(i) By (21)

$$\underline{\mathcal{E}}:=\inf_{\mu\in\mathcal{P}_2(H)}\mathcal{E}(\mu)\geq\inf_{\mu\in\mathcal{P}_2(H)}\frac{1}{2}m_2^2(\mu)-(1+2\,M)M_2^2(|\lambda|)>-\infty.$$

Let $(\mu_n)_{n\in\mathbb{N}}$ be a minimizing sequence, that is

$$\lim_{n\to\infty}\mathcal{E}(\mu_n)=\underline{\mathcal{E}}.$$

By (21)

$$\sup_{n\in\mathbb{N}}m_2^2(\mu_n)=:S<+\infty$$

and by Corollary 1-(ii) (μ_n) is relatively sequentially compact in $\mathcal{P}_2^w(H)$. Then, there exists a subsequence (which we do not relabel) and a limit point $\mu\in\mathcal{P}_2(H)$ such that $\mu_n\to\mu$ with respect to the topology of $\mathcal{P}_2^w(H)$. Define the sequence of functions $f_n:\mathcal{P}_2(H)\to\mathbb{R}$

$$f_n(\nu):=\mathcal{R}(\nu,\mu_n)=W_2^2(\nu,\mu_n)-m_2^2(\nu)-m_2^2(\mu_n)$$

By (22), for all $\nu\in\mathcal{P}_2(H)$

$$\lim_{n\to\infty}f_n(\nu)=\mathcal{R}(\nu,\mu)=:f(\nu).$$

For any $\gamma_n\in\Gamma_o(\nu,\mu_n)$, by Young's inequality

$$|f_n(\nu)|=\left|\int_{H\times H}2x\cdot y\mathrm{d}\gamma_n(x,y)\right|\leq m_2^2(\nu)+m_2^2(\mu_n)\leq m_2^2(\nu)+S=:g(\nu).$$

By hypothesis (2), $g\in L^1(\mathcal{P}_2(H),|\lambda|)$ and therefore, by Lebesgue's dominated convergence theorem

$$
\begin{aligned}
\lim_{n\to\infty}\int_{\mathcal{P}_2(H)} f_n(\nu)\mathrm{d}\lambda(\nu) &= \lim_{n\to\infty}\int_{\mathcal{P}_2(H)} f_n(\nu)\mathrm{d}\lambda^+(\nu) - \lim_{n\to\infty}\int_{\mathcal{P}_2(H)} f_n(\nu)\mathrm{d}\lambda^-(\nu) \\
&= \int_{\mathcal{P}_2(H)} f(\nu)\mathrm{d}\lambda^+(\nu) - \int_{\mathcal{P}_2(H)} f(\nu)\mathrm{d}\lambda^-(\nu) \\
&= \int_{\mathcal{P}_2(H)} f(\nu)\mathrm{d}\lambda(\nu).
\end{aligned} \tag{24}
$$

The lower semicontinuity of $\mu \mapsto \int_H |y|^2 \mathrm{d}\mu(y) = m_2^2(\mu)$ with respect to the narrow convergence of $\mathcal{P}(H)$ is a standard result. The lower semicontinuity with respect to the weak topology of $\mathcal{P}_2^w(H)$ can be checked, e.g., by [6, Theorem 5.1], noticing that $m_2^2(\mu) = W_2^2(\mu, \delta_0)$. Finally, since

$$
\begin{aligned}
\mathcal{E}(\mu) &= \int_{\mathcal{P}_2(H)} W_2^2(\nu,\mu)\mathrm{d}\lambda(\nu) \\
&= \int_{\mathcal{P}_2(H)} \left(\int_H |x|^2\mathrm{d}\nu(x) + \int_H |y|^2\mathrm{d}\mu(y) + \mathcal{R}(\nu,\mu)\right)\mathrm{d}\lambda(\nu) \\
&= m_2^2(\mu) + \int_{\mathcal{P}_2(H)} \left(f(\nu) + m_2^2(\nu)\right)\mathrm{d}\lambda(\nu),
\end{aligned}
$$

owing to the lower semicontinuity of m_2^2 with respect to the topology of $\mathcal{P}_2^w(H)$ and to the convergence in (24),

$$
\begin{aligned}
\underline{\mathcal{E}} = \lim_{n\to\infty}\mathcal{E}(\mu_n) &\geq \liminf_{n\to\infty} m_2^2(\mu_n) + \lim_{n\to\infty}\int_{\mathcal{P}_2(H)} \left(f_n(\nu) + m_2^2(\nu)\right)\mathrm{d}\lambda(\nu) \\
&\geq m_2^2(\mu) + \int_{\mathcal{P}_2(H)} \left(f(\nu) + m_2^2(\nu)\right)\mathrm{d}\lambda(\nu) \\
&= \mathcal{E}(\mu),
\end{aligned}
$$

which shows that any weak limit point μ of (μ_n) is a minimizer of $\mathcal{E}$. $\square$

5 Uniqueness in Hilbert Spaces

In this section we specialize to the case where λ^+ is concentrated on a single measure $\nu_0 \in \mathcal{P}_2(H)$. In this case, uniqueness of the solution holds, even in contrast to the case of regular barycenters.

In order to highlight the different role of the positive and negative weights, we use the following notation. Let $\alpha > 0$ and $\nu_0 \in \mathcal{P}_2(H)$; let σ be a positive measure on $\mathcal{P}_2(H)$ such that

$$
\sigma(\nu_0) = 0, \qquad \sigma(\mathcal{P}_2(H)) = \alpha.
$$

We study

$$\mathscr{E} : \mathcal{P}_2(X) \to \mathbb{R}, \qquad \mathscr{E}(\mu) := (1+\alpha)W_2^2(\mu, \nu_0) - \int_{\mathcal{P}_2(H)} W_2^2(\mu, \nu)\mathrm{d}\sigma(\nu). \tag{25}$$

With respect to the notation in the previous section, we are setting

$$\lambda = \lambda^+ - \lambda^-, \qquad \lambda^+ = (1+\alpha)\delta_{\nu_0}, \quad \lambda^- = \sigma.$$

5.1 λ-Convexity

We will use the concept of $\lambda-$convexity along geodesics, as defined in [14, Definition 9.2.4]; the argument, which we briefly expose below, is essentially the same as [12, Theorem 4.1]: we adapt it to our functional. Let (*M*, *d*) be a complete metric space, and suppose we have an appropriate concept of connecting curve $\gamma(s) = \gamma_s : [0, 1] \to M$ between two points γ_0, γ_1 in *M*, that we call *generalized geodesic*. We say that a functional $F : M \to \mathbb{R} \cup \{+\infty\}$ is $\lambda-$convex along generalized geodesics (with $\lambda > 0$) if for every γ_0, γ_1 in *M*, for every generalized geodesic γ_s connecting γ_0 and γ_1, and for every *s* in [0, 1], the inequality

$$F(\gamma_s) \leq (1-s)F(\gamma_0) + sF(\gamma_1) - \lambda s(1-s)d^2(\gamma_0, \gamma_1)$$

holds. If *F* is $\lambda-$convex for some $\lambda > 0$ and some class of generalized geodesics, if it is bounded from below, and it is lower-semicontinuous with respect to the metric topology of *M*, then *F* admits a minimum in *M*. This is true because, taken a minimizing sequence $(\mu_i)_{i\in\mathbb{N}}$ in *M*, we have

$$d^2(\mu_m, \mu_n) \leq 4/\lambda \left[\frac{1}{2}(F(\mu_m) + F(\mu_n)) - F(\gamma_{1/2})\right]$$

where γ_s is a generalized geodesic connecting μ_m and μ_n and we have taken $s = 1/2$. Since the right-hand side term tends to 0 as $m, n \to \infty$, the sequence is Cauchy; by completeness of *M* and lower semi-continuity of *F* we conclude that $\lim \mu_n$ is a minimum point for *F*.

Note that $\lambda-$convexity is stronger than convexity: if we have a minimum for *F*, it must also be unique.

If we specialize to the case $(M, d) = (\mathcal{P}_2(H), W_2)$, the right concept of generalized geodesic has been introduced in [14, Definition 9.2.2]: we report here the definition and the basic properties.

Definition 2 Let γ_0, γ_1 and ν_0 be in $\mathcal{P}_2(H)$. Choose optimal transports $\boldsymbol{\gamma}_{02}$ in $\Gamma_o(\gamma_0, \nu_0)$ and $\boldsymbol{\gamma}_{12}$ in $\Gamma_o(\gamma_1, \nu_0)$. By Lemma 1, we can glue $\boldsymbol{\gamma}_{02}$ and $\boldsymbol{\gamma}_{12}$ along the common marginal ν_0, obtaining a measure $\boldsymbol{\gamma}$ in $\mathcal{P}(H \times H \times H)$ such that $(\pi^{0,2})_\#\boldsymbol{\gamma} = \boldsymbol{\gamma}_{02}$ and $(\pi^{1,2})_\#\boldsymbol{\gamma} = \boldsymbol{\gamma}_{12}$. For $t \in (0, 1)$, denote by $\pi_t^{0\to 1}$ the interpolating projection $(1-t)\pi^0 + t\pi^1 : H^3 \to H$. A generalized geodesic γ_t connecting γ_0 to γ_1, with basepoint ν_0, is the curve $t \mapsto (\pi_t^{0\to 1})_\#\boldsymbol{\gamma} \in \mathcal{P}_2(H)$ (see Fig. 1).

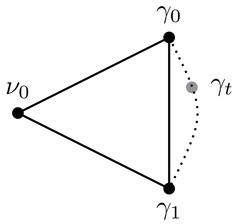

Fig. 1 Generalized geodesic with basepoint ν_0 (dotted). The continuous lines denote regular geodesics induced by the optimal transports $(\pi^{0,2})_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}_{02}$, $(\pi^{1,2})_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}_{12}$, and an optimal transport between γ_0 and γ_1; the dotted line is the curve induced by the (not necessarily optimal) transport $(\pi_t^{0\to 1})_{\#}\boldsymbol{\gamma}$

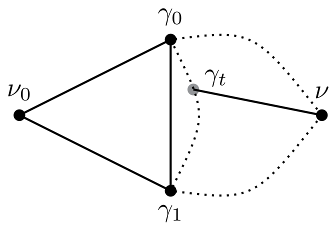

Fig. 2 Here's how we will use the lemma: with the same notation we used to define the generalized geodesic γ_t, γ will be $(\pi^{0,1})_{\#}\boldsymbol{\gamma}$, ν_0, is the fixed measure associated with the positive coefficient $(1+\alpha)$, ν will be integrated with respect to $\sigma = \lambda^-$ and t will be a fixed time in (0, 1). This lemma yields a measure $\boldsymbol{\nu}_t$ in $\mathcal{P}(H \times H \times H)$ such that $(\pi^{0,1})_{\#}\boldsymbol{\nu}_t = (\pi^{0,1})_{\#}\boldsymbol{\gamma}$ and $(\pi_t^{0\to 1}, \pi^2)_{\#}\boldsymbol{\nu}_t \in \Gamma_{\mathrm{o}}(\gamma_t, \nu)$

Notation: Following [14], we set

$$\pi_t^{i\to j} := (1-t)\pi^i + t\pi^j, \qquad \gamma_t^{i\to j} := (\pi_t^{i\to j})_{\#}\boldsymbol{\gamma}.$$

We will show that the functional $\mathscr{E}$ defined in (25) is $1-$convex; from this fact it will follow immediately that $\mathscr{E}$ has exactly one minimizer in $\mathcal{P}_2(X)$. Our proof follows the λ-convexity proof given in [12, Theorem 3.4] in the case of metric extrapolation, i.e., with one positive and one negative coefficient. Nonetheless, it should be noted that extending that theorem to a higher number of measures should be done carefully, since there are precise constraints on the conditions that can be imposed on the marginals, in order for a common product measure to exist (see also Remark 4 below).

We need the following lemma, where we adapt the notation to our case (see Fig. 2).

Lemma 5 (Curve extension lemma; [14], Proposition 7.3.1) Given $\gamma \in \mathcal{P}_2(H \times H)$, $\nu \in \mathcal{P}_2(H)$, and $t \in [0,1]$, there exists $\boldsymbol{\nu}_t \in \mathcal{P}_2(H \times H \times H)$ such that $(\pi^{0,1})_{\#}\boldsymbol{\nu}_t = \gamma$ and $(\pi_t^{0\to 1}, \pi^2)_{\#}\boldsymbol{\nu}_t \in \Gamma_{\mathrm{o}}((\pi_t^{0\to 1})_{\#}\gamma, \nu)$.

Theorem 7 *The functional $\mathscr{E}$ defined in (25) is $1-$convex along generalized geodesics with basepoint $\nu_0 \in \mathcal{P}_2(H)$.*

Proof We need the equality, valid in Hilbert spaces:

$$|(1-t)x_0 + tx_1 - x_2|^2 = (1-t)|x_0 - x_2|^2 + t|x_1 - x_2|^2 - t(1-t)|x_0 - x_1|^2. \quad (26)$$

Let γ_0, γ_1 be in $\mathcal{P}_2(H)$; let $\boldsymbol{\gamma}_{02}$ be in $\Gamma_0(\gamma_0, \nu_0)$ and $\boldsymbol{\gamma}_{12}$ be in $\Gamma_{\mathrm{o}}(\gamma_1, \nu_0)$. Let $\boldsymbol{\gamma}$ be a measure in $\mathcal{P}_2(H \times H \times H)$ obtained by gluing the optimal transports along the common marginal ν_0 and let $\gamma_t = (\pi_t^{0\to 1})_{\#}\boldsymbol{\gamma}$ be the associated generalized geodesic. Using (26) and the optimality of the marginals of $\boldsymbol{\gamma}$ we have

$$
\begin{aligned}
W_2^2(\gamma_t, \nu_0) &\leq \int_{H\times H} |y_1 - y_2|^2 \mathrm{d}(\pi_t^{0\to 1}, \pi^2)_{\#}\boldsymbol{\gamma}(y_1, y_2) \\
&= \int_{H\times H\times H} |(1-t)x_0 + tx_1 - x_2|^2 \mathrm{d}\boldsymbol{\gamma}(x_0, x_1, x_2) \\
&= (1-t)W_2^2(\gamma_0, \nu_0) + tW_2^2(\gamma_1, \nu_0) \\
&\quad - (1-t)t \int_{H\times H\times H} |x_0 - x_1|^2 \mathrm{d}\boldsymbol{\gamma}(x_0, x_1, x_2).
\end{aligned}
$$

Now let $\nu \in \mathcal{P}_2(H)$ and let $t \in [0,1]$. The curve extension lemma yields a measure $\boldsymbol{\nu}_t$ such that $(\pi^{0,1})_{\#}\boldsymbol{\nu}_t = (\pi^{0,1})_{\#}\boldsymbol{\gamma}$ and $(\pi_t^{0\to 1}, \pi^2)_{\#}\boldsymbol{\nu}_t \in \Gamma_{\mathrm{o}}((\pi_t^{0\to 1})_{\#}\boldsymbol{\gamma}, \nu)$. We have

$$
\begin{aligned}
W_2^2(\gamma_t, \nu) &= \int_{H\times H\times H} |(1-t)x_0 + tx_1 - x_2|^2 \mathrm{d}\boldsymbol{\nu}_t(x_0, x_1, x_2) \\
&= (1-t) \int_{H\times H\times H} |x_0 - x_2|^2 \mathrm{d}\boldsymbol{\nu}_t \\
&\quad + t \int_{H\times H\times H} |x_1 - x_2|^2 \mathrm{d}\boldsymbol{\nu}_t - t(1-t) \int_{H\times H\times H} |x_0 - x_1|^2 \mathrm{d}\boldsymbol{\nu}_t \\
&\geq (1-t)W_2^2(\gamma_0, \nu) + tW_2^2(\gamma_1, \nu) - t(1-t) \int_{H\times H\times H} |x_0 - x_1|^2 \mathrm{d}\boldsymbol{\gamma}.
\end{aligned}
$$

Using the two inequalities and the fact that $\sigma(\mathcal{P}_2(H)) = \alpha$, we obtain

$$
\begin{aligned}
\mathscr{E}(\gamma_t) &= (1+\alpha)W_2^2(\gamma_t, \nu_0) - \int_{\mathcal{P}_2(H)} W_2^2(\gamma_t, \nu)\mathrm{d}\sigma(\nu) \\
&\leq (1-t)\mathscr{E}(\gamma_0) + t\mathscr{E}(\gamma_1) - t(1-t) \int_{H\times H\times H} |x_0 - x_1|^2 \mathrm{d}\boldsymbol{\gamma} \\
&\leq (1-t)\mathscr{E}(\gamma_0) + t\mathscr{E}(\gamma_1) - t(1-t)W_2^2(\gamma_0, \gamma_1),
\end{aligned}
$$

which proves that $\mathscr{E}$ is $1-$convex along generalized geodesics with basepoint ν_0. $\square$

Remark 4 Note that having only one positive weight is crucial: if we were to try to generalize the argument to $n \geq 2$ positive weights, we would have to construct n different generalized geodesics γ_t with basepoints equal to the different measures associated with the positive weights.

Now we repeat the informal argument presented above to show that $\mathscr{E}$ admits a unique minimizer.

Proof of Theorem 1-(ii) Suppose that there are two minimum points η_1 and η_2 for $\mathscr{E}$. Let γ_s be a generalized geodesic with basepoint ν_0. By Theorem 7, for $s = 1/2$,

$$\frac{1}{4}W_2^2(\eta_1, \eta_2) \leq \frac{1}{2}(\mathscr{E}(\eta_1) + \mathscr{E}(\eta_2)) - \mathscr{E}(\gamma_{1/2}) \leq 0,$$

so that $W_2(\eta_1, \eta_2) = 0$. We conclude that $\eta_1 = \eta_2$. □

Corollary 2 Under the hypothesis of 1-(ii), if μ is the minimizer of $\mathscr{E}$ and (μ_n) is a minimizing sequence, then $\mu_n \to \mu$ in $\mathcal{P}_2(H)$ as $n \to \infty$.

Proof Let γ_s^n be a generalized geodesic between μ and μ_n, with basepoint ν_0. For all $\varepsilon > 0$ there exists $\bar{n} \in \mathbb{N}$ such that $\mathscr{E}(\mu_n) \leq \mathscr{E}(\mu) + \varepsilon$ for all $n \geq \bar{n}$. Therefore, by Theorem 7, for $s = 1/2$,

$$\limsup_{n\to\infty} \frac{1}{4}W_2^2(\mu, \mu_n) \leq \limsup_{n\to\infty} \left(\frac{1}{2}(\mathscr{E}(\mu) + \mathscr{E}(\mu_n)) - \mathscr{E}(\gamma_{1/2}^n)\right) \leq \frac{\varepsilon}{2}.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\lim_{n\to\infty} W_2^2(\mu, \mu_n) = 0$. □

5.2 Counterexample to Uniqueness in the Case of Two Positive Weights

We look at a particular example in $\mathcal{P}_2(\mathbb{R}^2)$: let

$$\nu_0 := \delta_{(0,0)}, \quad \nu_1 := \frac{1}{2}(\delta_{(-1,-1)} + \delta_{(1,1)}), \quad \nu_2 := \frac{1}{2}(\delta_{(1,-1)} + \delta_{(-1,1)}),$$

and consider the functional

$$\mathscr{E}(\mu) := -W_2^2(\mu, \nu_0) + W_2^2(\mu, \nu_1) + W_2^2(\mu, \nu_2).$$

Owing to Theorem 1, $\mathscr{E}$ admits a minimizer in $\mathcal{P}_2(\mathbb{R}^2)$; we are going to prove that uniqueness does not hold.

The argument will employ the symmetry of the problem; for convenience, we define the matrix

$$R := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the subsets

$$V := \left\{(x, y) \in \mathbb{R}^2 : |y| > |x|\right\} \quad \text{and} \quad U := \left\{(x, y) \in \mathbb{R}^2 : |x| > |y|\right\}.$$

Denoting $\mathrm{x} = (x, y) \in \mathbb{R}^2$, we consider the functions $S, T : \mathbb{R}^2 \to \mathbb{R}^2$,

$$S(\mathrm{x}) := -(R\mathrm{x})\mathbb{1}_V(\mathrm{x}) + \mathrm{x}\mathbb{1}_{V^c}(\mathrm{x}), \quad \text{and} \quad T(\mathrm{x}) := (R\mathrm{x})\mathbb{1}_U(\mathrm{x}) + \mathrm{x}\mathbb{1}_{U^c}(\mathrm{x}),$$

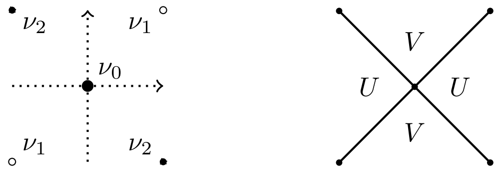

Fig. 3 The fixed measured ν_0, ν_1, and ν_2 (left); the subdivision of $\mathbb{R}^2$ using U and V (right)

where, for $A \subset \mathbb{R}^2$, $\mathbb{1}_A$ is the characteristic function of the set A and $A^c = \mathbb{R}^2 \setminus A$. Let μ be a minimizer of $\mathscr{E}$; we assume that μ is unique and argue by contradiction. Owing to the symmetry of ν_0, ν_1, and ν_2, (see Fig. 3), it can be easily checked that

$$\mathscr{E}(S_\sharp\mu) = \mathscr{E}(\mu) = \mathscr{E}(T_\sharp\mu).$$

By the uniqueness of μ, we infer that $S_\sharp\mu = \mu$, and therefore supp(μ) =supp$(S_\sharp\mu) \subset V^c$. Similarly $T_\sharp\mu = \mu$, and therefore supp(μ)=supp$(T_\sharp\mu) \subset U^c$. In conclusion,

$$\operatorname{supp}(\mu) \subset V^c \cap U^c = \left\{(x, y) \in \mathbb{R}^2 : |x| = |y|\right\}.$$

For all points p in $\{|x| = |y|\}$ let us consider the function

$$f(p) = \min\left\{|p - (-1, 1)|^2, |p - (1, -1)|^2\right\} + \min\left\{|p - (-1, -1)|^2, |p - (1, 1)|^2\right\} - |p|^2.$$

Then

$$\int_{\mathbb{R}^2} f(p) d\mu \leq \mathscr{E}(\mu).$$

Assume that p has the form $p = (x, x)$, then

$$|p - (-1, 1)|^2 = |p|^2 + 2 = |p - (1, -1)|^2.$$

The same computation, with respect to the support of ν_1, holds if p lies on $y = -x$. Therefore, for all measures μ concentrated on $\{|x| = |y|\}$,

$$\mathscr{E}(\mu) \geq \int_{\mathbb{R}^2} \left(|p|^2 + 2 - |p|^2\right) \mathrm{d}\mu(p) = 2.$$

On the other hand, if we take $\eta = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)}$, we have $\mathscr{E}(\eta) = 1$, which contradicts the assumption of μ being the minimizer of $\mathscr{E}$. We conclude that $\mathscr{E}$ does not admit a unique minimizer.

6 Proof of Theorem 2

Let $(\lambda_k)_{k\in\mathbb{N}}$ be a sequence of signed measures on $\mathcal{P}_2(H)$, such that

$$\lambda_k(\mathcal{P}_2(H)) = 1, \qquad M_2^2(|\lambda_k|) = \int_{\mathcal{P}_2(H)} m_2^2(\nu)\,\mathrm{d}|\lambda_k|(\nu) < +\infty.$$

As in the previous sections, the functional $\mathcal{E}_k : \mathcal{P}_2(H) \to \mathbb{R}$ can be decomposed as

$$\begin{aligned}\mathcal{E}_k(\mu) &= \int_{\mathcal{P}_2(H)} W_2^2(\nu,\mu)\mathrm{d}\lambda_k(\nu) \\ &= m_2^2(\mu) + \int_{\mathcal{P}_2(H)} \left(\mathcal{R}(\nu,\mu) + m_2^2(\nu)\right)\mathrm{d}\lambda_k(\nu).\end{aligned}$$

Γ-convergence of functions is defined in [7] for topological spaces. Under additional hypotheses, which are satisfied, e.g., by metric spaces, the Γ-limit has a useful sequential characterization, which we are going to employ. In our setting, the crucial property, that allows us to use the sequential characterization for the Γ-limit in the weak topology, is that sets of measures with bounded second moment are compact and metrizable with respect to the topology of $\mathcal{P}_2^w(H)$ (see Proposition 5 and [6, Proposition 3.4(b)]). Proposition 8.10 in [7], which is originally stated for a Banach space X with separable dual space, relies exactly on the existence of a metric that induces a topology equivalent to the weak topology on norm bounded sets. We can directly adapt Dal Maso's statement to our setting, where X is replaced by $\mathcal{P}_2(H)$ and norm bounded sets in X are replaced by sets bounded in the W_2 metric (i.e., with bounded second moment), and obtain the following characterization.

Proposition 8 Let $\Psi : \mathcal{P}_2(H) \to \mathbb{R} \cup \{+\infty\}$ be a function such that

$$\forall M > 0\ \exists K > 0 : m_2^2(\mu) > K \Rightarrow \Psi(\mu) > M. \tag{27}$$

Let (F_h) be a sequence of functionals on $(\mathcal{P}_2(H), W_2)$, such that $F_h(\mu) \geq \Psi(\mu)$ for all $h \in \mathbb{N}$, for all $\mu \in \mathcal{P}_2(H)$. Then,

$$F : \mathcal{P}_2(H) \to \mathbb{R} \text{ is the } \Gamma\text{-limit of } (F_h) \text{ in the topology of } \mathcal{P}_2^w(H)$$

if and only if

$$\begin{aligned}&(i)\ \liminf_{k\to\infty} F_h(\mu_h) \geq F(\mu) \qquad \forall \mu_h \to \mu \quad \text{in } \mathcal{P}_2^w(H);\\ &(ii)\ \forall \mu \in \mathcal{P}_2(H)\ \exists \mu_h \to \mu \text{ in } \mathcal{P}_2^w(H) : \limsup_{h\to\infty} F_h(\mu_h) \leq F(\mu).\end{aligned}$$

Under the hypotheses of Theorem 2, the sequence of functionals $(\mathcal{E}_k)$ is bounded from below by a function satisfying (27) and we can thus use the sequential characterization to prove the Γ-convergence result.

Lemma 6 Let $(\mathcal{E}_k)$ satisfy (4) and (5). Then, there exists a constant $M_\infty > 0$ such that

$$\mathcal{E}_k(\mu) \geq \frac{1}{2} m_2^2(\mu) - (1 + 2M_\infty) M_\infty$$

for all $\mu \in \mathcal{P}_2(H)$, for all $k \in \mathbb{N}$.

Proof By hypotheses (4) and (5), the variation and the quadratic moment of λ_k are uniformly bounded, i.e., there exists $M_\infty > 0$ such that

$$M(|\lambda_k|) + M_2^2(|\lambda_k|) < M_\infty \quad \forall k \in \mathbb{N}.$$

By Lemma 3, for all $\mu \in \mathcal{P}_2(H)$

$$\mathcal{E}_k(\mu) \geq \frac{1}{2} m_2^2(\mu) - (1 + 2\, M(|\lambda_k|)) M_2^2(|\lambda_k|) \geq \frac{1}{2} m_2^2(\mu) - (1 + 2M_\infty) M_\infty.$$

□

Remark 5 We are going to show that $\mathcal{E}$ is the Γ-limit of $\mathcal{E}_k$ with respect to the (weak) topology of $\mathcal{P}_2^w(H)$ and with respect to the (strong) topology of $(\mathcal{P}_2(H), W_2)$. This notion of convergence is known as Mosco-convergence (see [8]) and can be shown by proving the *liminf* inequality for the weak convergence and the *limsup* inequality for the strong one, as the other two inequalities are a direct implication of the former two.

From now on we denote $X := (\mathcal{P}_2(H), W_2)$ (i.e., the metric space $\mathcal{P}_2(H)$, endowed with the 2-Wasserstein distance). Notice that (4) and (5) imply (see, e.g., [14, Lemma 6.1.5]) that $\nu \mapsto m_2^2(\nu)$ is uniformly integrable with respect to (λ_k^+) and (λ_k^-). Recalling that, by Lemma 4, for all $\mu \in \mathcal{P}_2(H)$ the function $\nu \mapsto \mathcal{R}(\nu, \mu)$ is continuous in $\mathcal{P}_2^w(H)$, and thus in X, and that $|\mathcal{R}(\nu, \mu)| \leq m_2^2(\mu) + m_2^2(\nu)$ (this follows from the elementary inequality $2x \cdot y \leq |x|^2 + |y|^2$), we immediately obtain that for all $\mu \in \mathcal{P}_2(H)$,

$$\lim_{k \to \infty} \int_X W_2^2(\mu, \nu) \mathrm{d}\lambda_k^\pm(\nu) = \int_X W_2^2(\mu, \nu) \mathrm{d}\lambda^\pm(\nu). \tag{28}$$

From (28), the Γ-limsup inequality follows:

Lemma 7 (Γ-limsup) Let (λ_k) and λ be as in the hypotheses of Theorem 2. Then, for all μ in $\mathcal{P}_2(H)$ there exists a sequence (μ_k), converging to μ in X, such that

$$\limsup_{k \to \infty} \mathcal{E}_k(\mu_k) \leq \mathcal{E}(\mu).$$

Proof Let $\mu_k = \mu$ for all $k \in \mathbb{N}$. Then, owing to (28),

$$\begin{aligned}\limsup_{k\to\infty}\mathcal{E}_k(\mu_k) &= \limsup_{k\to\infty}\int_X W_2^2(\nu,\mu)\mathrm{d}\lambda_k(\nu)\\ &= \limsup_{k\to\infty}\left(\int_X W_2^2(\nu,\mu)\mathrm{d}\lambda_k^+(\nu) - \int_X W_2^2(\nu,\mu)\mathrm{d}\lambda_k^-(\nu)\right)\\ &= \int_X W_2^2(\nu,\mu)\mathrm{d}\lambda^+(\nu) - \int_X W_2^2(\nu,\mu)\mathrm{d}\lambda^-(\nu)\\ &= \int_X W_2^2(\nu,\mu)\mathrm{d}\lambda(\nu) = \mathcal{E}(\mu).\end{aligned}$$

□

In order to prove the Γ-liminf inequality, we need Skorohod's representation theorem (see, e.g., [2, Theorem 8.5.4]), which we report here using our notation.

Theorem 9 *(Skorohod) Let X be a complete, separable, metric space. Then, to every Borel probability measure μ on X, one can associate a Borel mapping $\xi_\mu : [0,1] \to X$ such that $\mu = (\xi_\mu)_\sharp\mathcal{L}$, where $\mathcal{L}$ is the Lebesgue measure, and*

$$\xi_{\mu_n}(t) \to \xi_\mu(t) \qquad for almost all t \in [0,1]$$

whenever $\mu_n \to \mu$ narrowly in $\mathcal{P}(X)$.

We are going to apply Skorohod's representation to address the Γ-liminf of the term involving $\mathcal{R}(\nu,\mu)$.

Lemma 8 Let $(\eta_k)_{k\in\mathbb{N}}$ and η be Borel probability measures on $X = (\mathcal{P}_2(H), W_2)$, such that $\eta_k \to \eta$ narrowly and

$$\lim_{k\to\infty}\int_X m_2^2(\nu)\,\mathrm{d}\eta_k(\nu) = \int_X m_2^2(\nu)\,\mathrm{d}\eta(\nu) < \infty. \tag{29}$$

Then, for all $(\mu_k)_{k\in\mathbb{N}}$ and μ in $\mathcal{P}_2(H)$ such that $\mu_k \to \mu$ with respect to the topology of $\mathcal{P}_2^w(H)$,

$$\lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\eta_k(\nu) = \int_X \mathcal{R}(\nu,\mu)\mathrm{d}\eta(\nu).$$

Proof According to Skorohod's representation theorem, there exist Borel mappings $\xi_k, \xi : [0,1] \to X$ such that

$$(\xi_k)_\sharp\mathcal{L} = \eta_k, \quad \xi_\sharp\mathcal{L} = \eta,$$

and $\xi_k(t) \to \xi(t)$ for almost all $t \in [0,1]$, i.e.,

$$\lim_{k\to\infty} W_2(\xi_k(t),\xi(t)) = 0, \qquad \text{for almost all } t \in [0,1]. \tag{30}$$

Therefore

$$\int_X m_2^2(\nu)\mathrm{d}\eta_k(\nu) = \int_0^1 m_2^2(\xi_k(t))\mathrm{d}t \tag{31}$$

and

$$\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\eta(\nu) = \int_0^1 \mathcal{R}(\xi_k(t),\mu_k)\mathrm{d}t.$$

Define

$$f_k : [0,1] \to [0,+\infty), \qquad f_k(t) := m_2^2(\xi_k(t)) = \int_H |x|^2 \mathrm{d}\xi_k(t)(x) \geq 0.$$

Owing to (30), we first remark that f_k converges for almost all $t \in [0,1]$:

$$\lim_{k\to\infty} f_k(t) = \lim_{k\to\infty} \int_H |x|^2 \mathrm{d}\xi_k(t)(x) = \int_H |x|^2 \mathrm{d}\xi(t)(x) =: f(t).$$

Then, owing to (31) and (29)

$$\lim_{k\to\infty} \int_0^1 f_k(t)\mathrm{d}t = \lim_{k\to\infty} \int_X m_2^2(\nu)\mathrm{d}\eta_k(\nu) = \int_X m_2^2(\nu)\mathrm{d}\eta(\nu) = \int_0^1 f(t)\mathrm{d}t.$$

Since $f_k \to f$ almost everywhere and $\|f_k\|_{L^1} \to \|f\|_{L^1}$, we conclude that

$$f_k \to f, \qquad \text{strongly in } L^1(0,1). \tag{32}$$

Since $\mu_k \to \mu$ (weakly) in $\mathcal{P}_2^w(H)$ and for almost all $t \in [0,1]$ $\xi_k(t) \to \xi(t)$ (strongly) in $\mathcal{P}_2(H)$, by Lemma 4

$$\lim_{k\to\infty} \mathcal{R}(\xi_k(t),\mu_k) = \mathcal{R}(\xi(t),\mu) \qquad \text{a.e. in } [0,1].$$

Denoting γ_k an optimal transport plan in $\Gamma_o(\xi_k(t),\mu_k)$, we obtain

$$\begin{aligned} |\mathcal{R}(\xi_k(t),\mu_k)| &= \left|2\int_{H\times H} x\cdot y\,\mathrm{d}\gamma_k(x,y)\right| \\ &\leq \left|\int_{H\times H} |x|^2 + |y|^2 \mathrm{d}\gamma_k(x,y)\right| \\ &= m_2^2(\xi_k(t)) + m_2^2(\mu_k). \end{aligned}$$

Since $m_2^2(\xi_k(t)) = f_k(t)$ converges strongly in $L^1(0,1)$ and $m_2^2(\mu_k)$ is bounded by hypothesis (since $\mu_k \to \mu$ in $\mathcal{P}_2^w(H)$), we may conclude by dominated convergence that

$$\begin{aligned}\lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\eta_k(\nu) &= \lim_{k\to\infty}\int_0^1 \mathcal{R}(\xi_k(t),\mu_k)\mathrm{d}t\\ &= \int_0^1 \mathcal{R}(\xi(t),\mu)\mathrm{d}t = \int_X \mathcal{R}(\nu,\mu)\mathrm{d}\eta(\nu).\end{aligned}$$

□

Lemma 9 (Γ-liminf) Let (λ_k) and λ be as in the hypotheses of Theorem 2. Then, for all (μ_k) and μ in $\mathcal{P}_2(H)$ such that $\mu_k \to \mu$ with respect to the topology of $\mathcal{P}_2^w(H)$,

$$\liminf_{k\to\infty} \mathcal{E}_k(\mu_k) \geq \mathcal{E}(\mu).$$

Proof We consider the Hahn-Jordan decomposition $\lambda_k = \lambda_k^+ - \lambda_k^-$, $\lambda = \lambda^+ - \lambda^-$, and notice that by (4)

$$\lim_{k\to\infty} \lambda_k^+(X) = \lambda^+(X), \qquad \lim_{k\to\infty} \lambda_k^-(X) = \lambda^-(X).$$

Assume first that $\lambda^-(X) \neq 0$: then the sequence $\lambda_k^-(X)$ is definitively uniformly positive, i.e., there exist $\bar{k} \in \mathbb{N}$ and $\epsilon > 0$ such that $\lambda_k^-(X) \geq \epsilon$ for all $k \geq \bar{k}$.

For $k \geq \bar{k}$, consider the sequences of probability measures

$$\eta_k^+ := \frac{\lambda_k^+}{\lambda_k^+(X)} \qquad \text{and} \qquad \eta_k^- := \frac{\lambda_k^-}{\lambda_k^-(X)}.$$

First, we notice that $\eta_k^\pm \to \lambda^\pm/\lambda^\pm(X)$ narrowly: applying Lemma 8, we obtain

$$\begin{aligned}\lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\lambda_k(\nu) &= \lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\lambda_k^+(\nu) - \lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\lambda_k^-(\nu)\\ &= \lim_{k\to\infty}\lambda_k^+(X)\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\eta_k^+(\nu) - \lambda_k^-(X)\lim_{k\to\infty}\int_X \mathcal{R}(\nu,\mu_k)\mathrm{d}\eta_k^-(\nu)\\ &= \int_X \mathcal{R}(\nu,\mu)\mathrm{d}\lambda^+(\nu) - \int_X \mathcal{R}(\nu,\mu)\mathrm{d}\lambda^-(\nu) = \int_X \mathcal{R}(\nu,\mu)\mathrm{d}\lambda(\nu).\end{aligned}$$

Recalling that

$$
\begin{aligned}
\mathcal{E}_k(\mu_k) &= \int_{\mathcal{P}_2(H)} W_2^2(\nu, \mu_k) \mathrm{d}\lambda_k(\nu) \\
&= m_2^2(\mu_k) + \int_{\mathcal{P}_2(H)} \left(\mathcal{R}(\nu, \mu_k) + m_2^2(\nu)\right) \mathrm{d}\lambda_k(\nu)
\end{aligned}
$$

and that

$$
\liminf_{k\to\infty} m_2^2(\mu_k) \geq m_2^2(\mu), \qquad \lim_{k\to\infty} \int_X m_2^2(\nu)\mathrm{d}\lambda_k(\nu) = \int_X m_2^2(\nu)\mathrm{d}\lambda(\nu),
$$

we obtain the thesis.

Now we turn to the case in which $\lambda^-(X) = 0$: in this case, the positive part may be treated as above, and it is enough to prove that

$$
\lim_{k\to\infty} \int_X W_2^2(\mu_k, \nu)\mathrm{d}\lambda_k^-(\nu) = 0.
$$

Using the squared triangular inequality,

$$
\begin{aligned}
\int_X W_2^2(\mu_k, \nu)\, \mathrm{d}\lambda_k^-(\nu) &\leq 2 \int_X W_2^2(\mu_k, \mu)\, \mathrm{d}\lambda_k^-(\nu) + 2 \int_X W_2^2(\mu, \nu)\, \mathrm{d}\lambda_k^-(\nu) \\
&\leq 2\, \lambda_k^-(X)\, W_2^2(\mu_k, \mu) + 2 \int_X W_2^2(\mu, \nu)\, \mathrm{d}\lambda_k^-(\nu).
\end{aligned}
$$

Note that the sequence $W_2^2(\mu_k, \mu)$ is bounded: in fact, $W_2^2(\mu_k, \mu) \leq 2(m_2^2(\mu_k) + m_2^2(\mu))$, and $m_2^2(\mu_k)$ is bounded, since μ_k converges to μ in $\mathcal{P}_2^w(H)$ (see Corollary 3.6 (b) in [6]). Hence, the first term vanishes, since $\lambda_k^-(X) \to 0$. The second term vanishes owing to (28). □

By definition of Γ-convergence, Lemma 7 and Lemma 9 ensure that $\mathcal{E}$ is the Γ-limit of $\mathcal{E}_k$ (or, more precisely, the Mosco-limit) and complete the proof of part (i) of Theorem 2.

If, in particular, μ_k is a minimizer of $\mathcal{E}_k$, by Lemma 6 and Corollary 1-(ii), the sequence $(\mu_k)_{k\in\mathbb{N}}$ is precompact with respect to the topology of $\mathcal{P}_2^w(H)$, which is the statement in Theorem 2-(ii). Finally, by the fundamental theorem of Γ-convergence ([7, Theorem 7.4]), if μ_k is a minimizer of $\mathcal{E}_k$, then

$$
\min_{\mu\in\mathcal{P}_2(H)} \mathcal{E}(\mu) = \lim_{k\to\infty} \mathcal{E}_k(\mu_k),
$$

and if $(\mu_{k_j})_{j\in\mathbb{N}}$ is a subsequence converging to $\overline{\mu}$ in $\mathcal{P}_2^w(H)$, then ([7, Corollary 7.20])

$$
\mathcal{E}(\overline{\mu}) = \min_{\mu\in\mathcal{P}_2(H)} \mathcal{E}(\mu),
$$

which concludes the proof of Theorem 2-(iii).

Acknowledgements We would like to thank Alessandro Spelta and Stefano Gualandi for many stimulating and insightful discussions on the applicative and computational aspects of Wasserstein barycenters. We are also grateful to the anonymous reviewer for the incisive comments and suggestions, that, from our point of view, considerably improved our work.

Author Contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by Francesco Tornabene and Marco Veneroni. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Pavia within the CRUI-CARE Agreement. M.V. has been supported by the MIUR-PRIN Grant 2020F3NCPX Mathematics for industry 4.0 (Math4I4). G.S. has also been supported by IMATI-CNR, Pavia, by the MIUR-PRIN 202244A7YL project Gradient Flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning, and by the INDAM project E53C23001740001. F.T. declares that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. SIAM J. Math. Anal. **43**(2), 904–924 (2011). https://doi.org/10.1137/100805741
2. Bogachev, V.I.: Measure Theory, p. 1075. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-34514-5
3. Bigot, J., Klein, T.: Consistent estimation of a population barycenter in the Wasserstein space. arXiv preprint (2012). https://arxiv.org/abs/1212.2562
4. Bigot, J., Klein, T.: Characterization of barycenters in the Wasserstein space by averaging optimal transport maps. ESAIM: P &S **22**, 35–57 (2018). https://doi.org/10.1051/ps/2017020
5. Le Gouic, T., Loubes, J.-M.: Existence and consistency of Wasserstein barycenters. Probab. Theory Relat. Fields **168**, 901–917 (2017). https://doi.org/10.1007/s00440-016-0727-z
6. Naldi, E., Savaré, G.: Weak topology and Opial property in Wasserstein spaces, with applications to gradient flows and proximal point algorithms of geodesically convex functionals. Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **32**(4), 725–750 (2021). https://doi.org/10.4171/rlm/955
7. Dal Maso, G.: An Introduction to Γ-convergence Progress in Nonlinear Differential Equations and Their Applications, vol. 8. Birkhäuser, Boston (1993) https://doi.org/10.1007/978-1-4612-0327-8
8. Mosco, U.: Approximation of the solutions of some variational inequalities. Ann. Scuola Normale Sup. Pisa Cl. Sci. **21**(3), 373–394 (1967)
9. Villani, C.: Topics in optimal transportation. Graduate Studies in Mathematics, vol. 58, p. 370. American Mathematical Society, Providence (2003). https://doi.org/10.1090/gsm/058
10. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**(1), 153–179 (1997). https://doi.org/10.1006/aima.1997.1634

11. Petersen, A., Müller, H.-G.: Fréchet regression for random objects with Euclidean predictors. Ann. Statist. **47**(2), 691–719 (2019). https://doi.org/10.1214/17-AOS1624
12. Matthes, D., Plazotta, S.: A variational formulation of the BDF2 method for metric gradient flows. ESAIM Math. Model. Numer. Anal. **53**(1), 145–172 (2019). https://doi.org/10.1051/m2an/2018045
13. Gallouët, T.O., Natale, A., Todeschi, G.: From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows. Math. Comp. **93**(350), 2769–2810 (2024). https://doi.org/10.1090/mcom/3951
14. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich, p. 333. Birkhäuser Verlag, Basel (2005). https://doi.org/10.1007/978-3-7643-8722-8
15. Dahlquist, G.: Convergence and stability in the numerical integration of ordinary differential equations. Math. Scand. **4**, 33–53 (1956). https://doi.org/10.7146/math.scand.a-10454
16. Gallouët, T.O., Natale, A., Todeschi, G.: Metric extrapolation in the Wasserstein space. Calc. Var. Partial Differ. Equ. **64**(5), 147 (2025). https://doi.org/10.1007/s00526-025-02996-6
17. Brenier, Y., Grenier, E.: Sticky particles and scalar conservation laws. SIAM J. Numer. Anal. **35**(6), 2317–2328 (1998). https://doi.org/10.1137/S0036142997317353
18. Natile, L., Savaré, G.: A Wasserstein approach to the one-dimensional sticky particle system. SIAM J. Math. Anal. **41**(4), 1340–1365 (2009). https://doi.org/10.1137/090750809
19. Villani, C.: Optimal transport: old and new. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, p. 973. Springer, Berlin (2009). https://doi.org/10.1007/978-3-540-71050-9
20. Santambrogio, F.: Optimal transport for applied mathematicians: calculus of variations, pdes, and modeling. Progress in Nonlinear Differential Equations and their Applications, vol. 87, p. 353. Birkhäuser/Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20828-2
21. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems. Vol. I : Theory. Probability and its Applications (New York), p. 508. Springer, New York (1998). https://doi.org/10.1007/b98893
22. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext, p. 599. Springer, New York (2011). https://doi.org/10.1007/978-0-387-70914-7