



HAL
open science

Consistent estimation of a population barycenter in the Wasserstein space

Jérémie Bigot, Thierry Klein

► **To cite this version:**

Jérémie Bigot, Thierry Klein. Consistent estimation of a population barycenter in the Wasserstein space. 2012. hal-00763668v1

HAL Id: hal-00763668

<https://hal.science/hal-00763668v1>

Preprint submitted on 11 Dec 2012 (v1), last revised 28 Nov 2017 (v7)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Consistent estimation of a population barycenter in the Wasserstein space

Jérémie Bigot and Thierry Klein

*DMIA - Institut Supérieur de l'Aéronautique et de l'Espace**

IMT - Université Paul Sabatier[†]

December 11, 2012

Abstract

We define a notion of barycenter for random probability measures in the Wasserstein space. We give a characterization of the population barycenter in terms of existence and uniqueness for compactly supported measures. Then, the problem of estimating this barycenter from n independent and identically distributed random probability measures is considered. We study the convergence of the empirical barycenter proposed in Agueh and Carlier [2] to its population counterpart as the number of measures n tends to infinity. To illustrate the benefits of this approach for data analysis and statistics, we finally discuss the usefulness of barycenters in the Wasserstein space for curve and image warping.

Keywords: Wasserstein space; Empirical and population barycenters; Fréchet mean; Convergence of random variables; Optimal transport; Curve and image warping.

AMS classifications: Primary 62G05; secondary 49J40.

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS. We would like to also thank Jérôme Bertrand for fruitful discussions on Wasserstein spaces and the optimal transport problem.

1 Introduction

In this paper, we consider the problem of defining the barycenter of random probability measures on \mathbb{R}^d . The set of Radon probability measures endowed with the 2-Wasserstein distance is not an Euclidean space. Consequently, to define a notion of barycenter for random probability measures, it is natural to use the notion of Fréchet mean [13] that is an extension of the usual Euclidean barycenter to non-linear spaces endowed with non-Euclidean metrics. If \mathbf{Y} denotes a random

*Institut Supérieur de l'Aéronautique et de l'Espace, Département Mathématiques, Informatique, Automatique, 10 Avenue Édouard-Belin, BP 54032-31055, Toulouse CEDEX 4, France. Email: jeremie.bigot@isae.fr

[†]Institut de Mathématiques de Toulouse, Université de Toulouse et CNRS (UMR 5219), France. Email: thierry.klein@math.univ-toulouse.fr

variable with distribution \mathbb{P} taking its value in a metric space $(\mathcal{M}, d_{\mathcal{M}})$, then a Fréchet mean (not necessarily unique) of the distribution \mathbb{P} is a point $m^* \in \mathcal{M}$ that is a global minimum of the functional

$$J(m) = \frac{1}{2} \int_{\mathcal{M}} d_{\mathcal{M}}^2(m, y) d\mathbb{P}(y) \quad \text{i.e.} \quad m^* \in \arg \min_{m \in \mathcal{M}} J(m).$$

In this paper, a Fréchet mean of a distribution \mathbb{P} will be also called a barycenter. An empirical Fréchet mean of an independent and identically distributed (iid) sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ of distribution \mathbb{P} is

$$\bar{\mathbf{Y}}_n \in \arg \min_{m \in \mathcal{M}} \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{\mathcal{M}}^2(m, \mathbf{Y}_j).$$

For random variables belonging to nonlinear metric spaces, a well-known example is the computation of the mean of a set of planar shapes in the Kendall's shape space [23] that leads to the Procrustean means studied in [16]. Many properties of the Fréchet mean in finite dimensional Riemannian manifolds (such as consistency and uniqueness) have been investigated in [1, 4, 5, 6, 21]. For random variables taking their value in metric spaces of nonpositive curvature (NPC), a detailed study of various properties of their barycenter can be found in [27]. Recently, some properties of the Fréchet mean in bounded metric spaces have also been studied in [15]. However, there is not so much work on Fréchet means in infinite dimensional metric spaces that do not satisfy the global NPC property as defined in [27].

In this paper, we consider the case where $\mathbf{Y} = \boldsymbol{\mu}$ is a random probability measure belonging to the 2-Wasserstein space on \mathbb{R}^d with distribution \mathbb{P} . More precisely, we propose to study some properties of the barycenter μ^* of $\boldsymbol{\mu}$ defined as the following Fréchet mean

$$\mu^* = \arg \min_{\nu \in \mathcal{M}_+^1(\mathbb{R}^d)} \int_{\mathcal{M}_+^1(\mathbb{R}^d)} \frac{1}{2} d_{W_2}^2(\nu, \mu) d\mathbb{P}(\mu), \quad (1.1)$$

where $\mathcal{M}_+^1(\mathbb{R}^d)$ is the set of Radon probability measures on \mathbb{R}^d , and $d_{W_2}^2$ denotes the squared 2-Wasserstein distance between two probability measures. Note that \mathbb{P} denotes a probability distribution on the space of probability measures $(\mathcal{M}_+^1(\mathbb{R}^d), \mathcal{B}(\mathcal{M}_+^1(\mathbb{R}^d)))$, where $\mathcal{B}(\mathcal{M}_+^1(\mathbb{R}^d))$ is the Borel σ -algebra generated by the topology induced by the distance d_{W_2} . If it exists and is unique, the measure μ^* will be referred to as the population barycenter of the random measure $\boldsymbol{\mu}$ with distribution \mathbb{P} . The empirical counterpart of μ^* is the barycenter $\bar{\boldsymbol{\mu}}_n$ defined as

$$\bar{\boldsymbol{\mu}}_n = \arg \min_{\nu \in \mathcal{M}_+^1(\mathbb{R}^d)} \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{W_2}^2(\nu, \boldsymbol{\mu}_j), \quad (1.2)$$

where $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n$ are iid random measures sampled from the distribution \mathbb{P} . A detailed characterization of $\bar{\boldsymbol{\mu}}_n$ in terms of existence, uniqueness and regularity, together with its link to the multi-marginal problem in optimal transport has been proposed in [2].

The first contribution of this paper is to discuss some assumptions on \mathbb{P} that guaranty the existence and uniqueness of the population barycenter. These results are based on an adaptation of the arguments developed in [2] for the characterization of the empirical barycenter $\bar{\boldsymbol{\mu}}_n$. In particular, we propose a dual formulation of the optimisation problem (1.1) that allows a precise study of some properties of the population barycenter such as its uniqueness. Therefore, our

approach is very much connected with the theory of optimal mass transport, and with the characterization of the Monge-Kantorovich problem via arguments from convex analysis and duality, see [31] for further details on this topic. A second contribution of this paper is to study the convergence of $\bar{\mu}_n$ to μ^* as the number n of measures tends to infinity. Finally, we show that this notion of barycenter of probability measures has interesting applications in various statistical models for data analysis.

The paper is then organised as follows. In Section 2, we give a characterisation of the population barycenter in the case of compactly supported measures. The convergence of the empirical barycenter is discussed in Section 3. As an application of the methodology developed in this paper, we discuss in Section 4 the usefulness of barycenters in the Wasserstein for curve and image warping problems. Finally, we give a conclusion and some perspectives in Section 5.

Throughout the paper, we use bold symbols $\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\theta}, \dots$ to denote random variables.

2 Characterisation of the population barycenter of compactly supported measures

2.1 Some definitions and notations

The notation $|x|$ is used to denote the usual Euclidean norm of a vector $x \in \mathbb{R}^m$, and the notation $\langle x, y \rangle$ denotes the usual inner product for $x, y \in \mathbb{R}^m$. Let Ω be compact set in \mathbb{R}^d , and let $\delta(\Omega) = \sup_{(x,y) \in \Omega \times \Omega} |x - y|$ be its diameter. Let $X = C(\Omega, \mathbb{R})$ be the space of continuous functions $f : \Omega \rightarrow \mathbb{R}$ equipped with the supremum norm

$$\|f\|_X = \sup_{x \in \Omega} \{|f(x)|\}.$$

We denote by $\mathcal{M}(\Omega)$ the space of bounded Radon measures on Ω and by $\mathcal{M}_+^1(\Omega)$ the set of Radon probability measures. Note that any $\nu \in \mathcal{M}_+^1(\Omega)$ can be considered as a probability measure on \mathbb{R}^d having a compact support included in Ω . In this section, we characterize the population barycenter of a specific class of random probability measures taking their values in $\mathcal{M}_+^1(\Omega)$.

Let $X' = \mathcal{M}(\Omega)$ be the topological dual of X . We recall that the squared 2-Wasserstein distance between two probability measures $\mu, \nu \in \mathcal{M}(\Omega)$ is

$$d_{W_2}^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_{\Omega \times \Omega} |x - y|^2 d\gamma(x, y) \right\},$$

where $\Pi(\mu, \nu)$ is the set of all probability measures on $\Omega \times \Omega$ having μ and ν as marginals, see e.g. [31]. We recall that $\gamma \in \Pi(\mu, \nu)$ is called an optimal transport plan between μ and ν if

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

Let $T : \Omega \rightarrow \Omega$ be a measurable mapping, and let $\mu \in \mathcal{M}(\Omega)$. The push-forward measure $T\#\mu$ of μ through the map T is defined as the measure

$$\int_{\Omega} f(x) d(T\#\mu)(x) = \int_{\Omega} f(T(x)) d\mu(x), \text{ for all } f \in X.$$

We also recall the following well known result in optimal transport (see e.g. [31] or Proposition 3.3 in [2]):

Proposition 2.1. *Let $\mu, \nu \in \mathcal{M}(\Omega)$. Then, $\gamma \in \Pi(\mu, \nu)$ is an optimal transport plan between μ and ν if and only if the support of γ is included in the set $\partial\phi$ that is the graph of the subdifferential of a convex and lower semi-continuous function ϕ solution of the problem*

$$\phi = \arg \min_{\psi \in \mathcal{C}} \left\{ \int_{\Omega} \psi(x) d\mu(x) + \int_{\Omega} \psi^*(x) d\nu(x) \right\},$$

where $\psi^*(x) = \sup_{y \in \Omega} \{\langle x, y \rangle - \psi(y)\}$ is the convex conjugate of ψ , and \mathcal{C} denotes the set of convex functions $\psi : \Omega \rightarrow \mathbb{R}$ that are lower semi-continuous

Moreover, if μ admits a density with respect to the Lebesgue measure on \mathbb{R}^d , then there exists a unique optimal transport plan $\gamma \in \Pi(\mu, \nu)$ that is of the form $\gamma = (id, \nabla\phi)\#\mu$ where $\nabla\phi$ denotes the gradient of ϕ . The uniqueness of the transport plan holds in the sense that if $\nabla\phi\#\mu = \nabla\psi\#\mu$, where $\psi : \Omega \rightarrow \mathbb{R}$ is a convex function, then $\phi = \psi$ μ -almost everywhere.

Proposition 2.1 is the key ingredient in the proof of Theorem 2.1 (stated later on in this section) to show the uniqueness of a population barycenter. Let us finally recall that the Wasserstein space $(\mathcal{M}_+^1(\Omega), d_{W_2})$ is a compact metric space since Ω is a compact subset of \mathbb{R}^d .

2.2 A parametric class of random probability measures

Now, let us define a class of random probability measures belonging to $\mathcal{M}_+^1(\Omega)$. Let Θ be a compact subset of \mathbb{R}^p . Let $\phi : (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p)) \rightarrow (\mathcal{M}_+^1(\Omega), \mathcal{B}(\mathcal{M}_+^1(\Omega)))$ be a measurable mapping, where $\mathcal{B}(\mathbb{R}^p)$ is the Borel σ -algebra of \mathbb{R}^p and $\mathcal{B}(\mathcal{M}_+^1(\Omega))$ is the Borel σ -algebra generated by the topology induced by the distance d_{W_2} . Then, let us define

$$M_{\phi}(\Theta) = \{\mu_{\theta} = \phi(\theta), \theta \in \Theta\}$$

as the set of probability measures $\mu_{\theta} \in \mathcal{M}_+^1(\Omega)$ parametrized by the mapping ϕ and the compact set Θ . Throughout the paper, we will suppose that ϕ satisfies the following assumption:

Assumption 1. *For any $\theta \in \Theta$, the measure $\mu_{\theta} = \phi(\theta) \in \mathcal{M}_+^1(\Omega)$ admits a density with respect to the Lebesgue measure on \mathbb{R}^p .*

Let \mathbb{P}_{Θ} be a probability measure on Θ with density $g : \Theta \rightarrow \mathbb{R}^+$ with respect to the Lebesgue measure $d\theta$ on \mathbb{R}^p . We assume that g satisfies the following regularity conditions:

Assumption 2. *The density g is L -Lipschitz for some constant $L > 0$ i.e.*

$$|g(\theta_1) - g(\theta_2)| \leq L|\theta_1 - \theta_2|, \text{ for any } \theta_1, \theta_2 \in \Theta. \quad (2.1)$$

If $\theta \in \mathbb{R}^p$ is a random vector with density g , then $\mu_{\theta} = \phi(\theta)$ is a random probability measure with distribution \mathbb{P}_g on $(\mathcal{M}_+^1(\Omega), \mathcal{B}(\mathcal{M}_+^1(\Omega)))$ that is the push-forward measure defined by

$$\mathbb{P}_g(B) = \mathbb{P}_{\Theta}(\phi^{-1}(B)), \text{ for any } B \in \mathcal{B}(\mathcal{M}_+^1(\Omega)).$$

As explained in the introduction, we want to characterize the barycenter (i.e. the Fréchet mean) of the distribution \mathbb{P}_g when $\mathcal{M}_+^1(\Omega)$ is endowed with the 2-Wasserstein distance d_{W_2} . For this purpose, let us consider the optimization problem: find

$$\mu^* \in \arg \min_{\nu \in \mathcal{M}_+^1(\Omega)} \int_{\mathcal{M}_+^1(\Omega)} \frac{1}{2} d_{W_2}^2(\nu, \mu) d\mathbb{P}_g(\mu) = \arg \min_{\nu \in \mathcal{M}_+^1(\Omega)} \int_{\Theta} \frac{1}{2} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta. \quad (2.2)$$

The main goals of this section are to prove the existence and the uniqueness of μ^* . Since Ω is compact, it is obvious that $d_{W_2}^2(\nu, \mu_\theta) \leq \delta^2(\Omega)$ and thus

$$\int_{\Theta} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta \leq \delta^2(\Omega) < +\infty \text{ for any } \nu \in \mathcal{M}_+^1(\Omega). \quad (2.3)$$

2.3 Primal and dual formulations

Consider the problem

$$(\mathcal{P}) \quad J_{\mathcal{P}} := \inf_{\nu \in \mathcal{M}_+^1(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta. \quad (2.4)$$

To study the existence and uniqueness of μ^* , let us introduce some definitions. The notation $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ will denote any application

$$\begin{cases} f^\Theta : \Theta & \rightarrow X \\ \theta & \mapsto f_\theta \end{cases}$$

such that for any $x \in \Omega$

$$\int_{\Theta} |f_\theta(x)| d\theta < +\infty.$$

Then, following the terminology in [2], we introduce the dual problem

$$(\mathcal{P}^*) \quad J_{\mathcal{P}^*} := \sup \left\{ \int_{\Theta} \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) d\theta; f^\Theta \in L^1(\Theta, X) \text{ such that } \int_{\Theta} f_\theta(x) d\theta = 0, \forall x \in \Omega \right\},$$

where

$$S_{g(\theta)} f(x) := \inf_{y \in \Omega} \left\{ \frac{g(\theta)}{2} |x - y|^2 - f(y) \right\}, \forall x \in \Omega \text{ and } f \in X.$$

Let us also define

$$H_{g(\theta)}(f) := - \int_{\Omega} S_{g(\theta)} f(x) d\mu_\theta(x),$$

and the Legendre-Fenchel transform of $H_{g(\theta)}$ as for $\nu \in X'$

$$H_{g(\theta)}^*(\nu) := \sup_{f \in X} \left\{ \int_{\Omega} f(x) d\nu(x) - H_{g(\theta)}(f) \right\} = \sup_{f \in X} \left\{ \int_{\Omega} f(x) d\nu(x) + \int_{\Theta} \int_{\Omega} S_{g(\theta)} f(x) d\mu_\theta(x) d\theta \right\}.$$

In what follows, we will show that the problems (\mathcal{P}) and (\mathcal{P}^*) are related in the sense that the minimal value $J_{\mathcal{P}}$ in problem (\mathcal{P}) is equal to the supremum $J_{\mathcal{P}^*}$ in problem (\mathcal{P}^*) , see Proposition 2.2 below. We will also show that both problems have optimizers, see Proposition 2.3 below. This duality will then allow us to characterize the uniqueness of the population barycenter, see Theorem 2.1.

Proposition 2.2. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then,*

$$J_{\mathcal{P}} = J_{\mathcal{P}^*}.$$

Proof. 1. Let us first prove that $J_{\mathcal{P}} \geq J_{\mathcal{P}^*}$.

By definition for any $f^{\Theta} \in L^1(\Theta, X)$ such that $\forall x \in \Omega$, $\int_{\Theta} f_{\theta}(x) d\theta = 0$, and for all $y \in \Omega$ we have

$$S_{g(\theta)} f_{\theta}(x) + f_{\theta}(y) \leq \frac{g(\theta)}{2} |x - y|^2.$$

Let $\nu \in \mathcal{M}_+^1(\Omega)$ and $\gamma_{\theta} \in \Pi(\mu_{\theta}, \nu)$ be an optimal transport plan between μ_{θ} and ν . By integrating the above inequality with respect to γ_{θ} we obtain

$$\int_{\Omega} S_{g(\theta)} f_{\theta}(x) d\mu_{\theta}(x) + \int_{\Omega} f_{\theta}(y) d\nu(y) \leq \int_{\Omega \times \Omega} \frac{g(\theta)}{2} |x - y|^2 d\gamma_{\theta}(x, y) = \frac{g(\theta)}{2} W_2(\mu_{\theta}, \nu).$$

Integrating now with respect to $d\theta$ and using Fubini's Theorem we get

$$\int_{\Theta} \int_{\Omega} S_{g(\theta)} f_{\theta}(x) d\mu_{\theta}(x) d\theta \leq \int_{\Theta} \frac{g(\theta)}{2} W_2(\mu_{\theta}, \nu) d\theta.$$

Therefore we deduce that $J_{\mathcal{P}} \geq J_{\mathcal{P}^*}$.

2. Let us now prove the converse inequalities $J_{\mathcal{P}} \leq J_{\mathcal{P}^*}$.

Thanks to the Kantorovich duality formula (see e.g. [31], or Lemma 2.1 in [2]) we have that $H_{g(\theta)}^*(\nu) = \frac{1}{2} d_{W_2}^2(\mu_{\theta}, \nu) g(\theta)$ for any $\nu \in \mathcal{M}_+^1(\Omega)$. Therefore, it follows that

$$J_{\mathcal{P}} = \inf \left\{ \int_{\Theta} H_{g(\theta)}^*(\nu) d\theta, \nu \in X' \right\} = - \left(\int_{\Theta} H_{g(\theta)}^* d\theta \right)^* (0). \quad (2.5)$$

Define the inf-convolution of $(H_{g(\theta)})_{\theta \in \Theta}$ by

$$H(f) := \inf \left\{ \int_{\Theta} H_{g(\theta)}(f_{\theta}) d\theta; f^{\Theta} \in L^1(\Theta, X), \int_{\Theta} f_{\theta}(x) d\theta = f(x), \forall x \in \Omega \right\}, \quad \forall f \in X.$$

We have in the other hand that

$$J_{\mathcal{P}^*} = -H(0).$$

Using Theorem 1.6 in [22], one has that for any $\nu \in \mathcal{M}_+^1(\Omega)$

$$H^*(\nu) = \int_{\Theta} H_{g(\theta)}^*(\nu) d\theta.$$

Then, thanks to (2.5), it follows that

$$J_{\mathcal{P}} = -H^{**}(0) \geq -H(0) = J_{\mathcal{P}^*}.$$

Let us now prove that $H^{**}(0) = H(0)$. Since H is convex it is sufficient to show that H is continuous at 0 for the supremum norm of the space X (see e.g. [12]). For this purpose, let $f^\Theta \in L^1(\Theta, X)$ and remark that it follows from the definition of $H_{g(\theta)}$ that

$$\begin{aligned} H_{g(\theta)}(f_\theta) &= \int_{\Omega} \sup_{y \in \Omega} \left\{ f_\theta(y) - \frac{g(\theta)}{2} |x - y|^2 \right\} d\mu_\theta(x) \\ &\geq f_\theta(0) - \frac{g(\theta)}{2} \int_{\Omega} |x|^2 d\mu_\theta(x), \end{aligned}$$

which implies that

$$H(f) \geq f(0) - \int_{\Theta} \frac{g(\theta)}{2} \int_{\Omega} |x|^2 d\mu_\theta(x) d\theta > -\infty, \quad \forall f \in X.$$

Let $f \in X$ such that $\|f\|_X \leq 1/4$ and choose $f^\Theta \in L^1(\Theta, X)$ defined by $f_\theta(x) = f(x)g(\theta)$ for all $\theta \in \Theta$ and $x \in \Omega$. It follows that

$$\begin{aligned} H(f) &\leq \int_{\Theta} H_{g(\theta)}(f(\cdot)g(\theta)) d\theta \leq \int_{\Theta} \int_{\Omega} \sup_{y \in \Omega} \left\{ \frac{g(\theta)}{4} - \frac{g(\theta)}{2} |x - y|^2 \right\} d\mu_\theta(x) d\theta \\ &\leq \int_{\Theta} \int_{\Omega} \frac{g(\theta)}{4} d\mu_\theta(x) d\theta = \frac{1}{4}. \end{aligned}$$

Hence, the convex function H never takes the value $-\infty$ and is bounded from above in a neighborhood of 0 in X . Therefore, by standard results in convex analysis (see e.g. [12]), H is continuous at 0, and therefore $H^{**}(0) = H(0)$ which completes the proof. \square

Let us now prove the existence of an optimizer for the primal problem (\mathcal{P}) and its dual (\mathcal{P}^*) as formulated in the following proposition:

Proposition 2.3. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then, both problems (\mathcal{P}) and (\mathcal{P}^*) admit an optimizer.*

Proof. 1. Let ν^n be a minimizing sequence of (\mathcal{P}) . Since Ω is compact, the sequence $\int_{\Omega} |x|^2 d\nu^n(x)$ is uniformly bounded. Hence, by Chebyshev's inequality, the sequence ν_n is tight and by Prokhorov's Theorem there exists a (non relabeled) subsequence that weakly converges to some $\mu^* \in \mathcal{M}_+^1(\Omega)$. Since Ω is compact, it can be checked that $\nu \mapsto d_{W_2}^2(\mu_\theta, \nu)$ is lower semi-continuous (l.s.c.) on $\mathcal{M}_+^1(\Omega)$. Therefore by Fatou's Lemma

$$\int_{\Theta} \frac{1}{2} d_{W_2}^2(\mu_\theta, \mu^*) g(\theta) d\theta = \int_{\Theta} \liminf \frac{1}{2} d_{W_2}^2(\mu_\theta, \nu^n) g(\theta) d\theta \leq \liminf \int_{\Theta} \frac{1}{2} d_{W_2}^2(\mu_\theta, \nu_n) g(\theta) d\theta,$$

and therefore $J(v^*) = \inf_{\nu \in \mathcal{M}_+^1(\Omega)} \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta$, which proves that (\mathcal{P}) admits a minimizer.

2. Let $f^\Theta \in L^1(\Theta, X)$ such that $\int_{\Theta} f_\theta d\theta = 0$ and define $h_\theta(x) = S_{g(\theta)} \circ S_{g(\theta)} f_\theta(x)$ for every $x \in \Omega$ and $\theta \in \Theta$. It is easy to check that $f_\theta(x) \leq h_\theta(x)$ and that $h_\theta(x) \leq \frac{g(\theta)}{2} |x|^2 - S_{g(\theta)} f_\theta(0)$. Hence, these two inequalities imply that $\theta \mapsto h_\theta \in L^1(\Theta, X)$. Now, define $\tilde{f}_\theta = h_\theta - \int_{\Theta} h_u du$ for every

$\theta \in \Theta$. Since $\int_{\Theta} h_u du \geq \int_{\Theta} f_u du = 0$, one has that $\tilde{f}_{\theta} \leq h_{\theta}$ which implies that $S_{g(\theta)} \tilde{f}_{\theta} \geq S_{g(\theta)} h_{\theta}$ since $S_{g(\theta)}$ is order-reversing. Since $S_{g(\theta)} h_{\theta} = S_{g(\theta)}^3 f_{\theta}$ it follows that $S_{g(\theta)} h_{\theta} \geq S_{g(\theta)} f_{\theta}$. Moreover, the inequality $f_{\theta} \leq h_{\theta}$ implies that $S_{g(\theta)} f_{\theta} \geq S_{g(\theta)} h_{\theta}$ which finally shows that $S_{g(\theta)} f_{\theta} = S_{g(\theta)} h_{\theta}$ and therefore

$$\int_{\Theta} \int_{\Omega} S_{g(\theta)} \tilde{f}_{\theta}(x) d\mu_{\theta}(x) d\theta \geq \int_{\Theta} \int_{\Omega} S_{g(\theta)} f_{\theta}(x) d\mu_{\theta}(x) d\theta.$$

Hence, one may assume that the supremum in (\mathcal{P}^*) can be restricted to the $f^{\Theta} \in L^1(\Theta, X)$ satisfying $\tilde{f}_{\theta} = h_{\theta} - \int_{\Theta} h_u du$ with $S_{g(\theta)}^2 h_{\theta} = h_{\theta}$ for every $\theta \in \Theta$. Note that one may also assume that $h_{\theta}(0) = 0$ since the functional $\int_{\Theta} \int_{\Omega} S_{g(\theta)} f_{\theta}(x) d\mu_{\theta}(x) d\theta$ in problem (\mathcal{P}^*) is invariant when one adds to the f_{θ} 's constants c_{θ} that integrate to zero namely $\int_{\Theta} c_{\theta} d\theta = 0$.

Now, let $\tilde{f}^{\Theta, n} \in L^1(\Theta, X)$ be a maximizing sequence for problem (\mathcal{P}^*) that can thus be chosen such that $\tilde{f}_{\theta}^n = h_{\theta}^n - \int_{\Theta} h_u^n du$ with $h_{\theta}^n = S_{g(\theta)} g_{\theta}^n$, $g_{\theta}^n = S_{g(\theta)} h_{\theta}^n$ and $h_{\theta}^n(0) = 0$ for every $\theta \in \Theta$. By using (2.1) that for any $x, z \in \Omega$, and $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned} \left| \frac{g(\theta_1)}{2} |x-y|^2 - \frac{g(\theta_2)}{2} |z-y|^2 \right| &= \left| \frac{g(\theta_1)}{2} (|x-y|^2 - |z-y|^2) + \left(\frac{g(\theta_1)}{2} - \frac{g(\theta_2)}{2} \right) |z-y|^2 \right| \\ &\leq \frac{g(\theta_1)}{2} |x-y| - |z-y| (|x-y| + |z-y|) \\ &\quad + \frac{|z-y|^2}{2} |g(\theta_1) - g(\theta_2)| \\ &\leq \delta(\Omega) g(\theta_1) |x-z| + \frac{L\delta^2(\Omega)}{2} |\theta_1 - \theta_2| \\ &\leq K \max(|x-z|, |\theta_1 - \theta_2|), \end{aligned}$$

with $K = \max\left(\delta(\Omega) \|p\|_{\infty}, \frac{L\delta^2(\Omega)}{2}\right)$.

Therefore, the function $(x, \theta) \mapsto \frac{g(\theta)}{2} |x-y|^2 - g_{\theta}^n(y)$ is K -Lipschitz on the compact set $\Omega \times \Theta$. Thus, the function $(x, \theta) \mapsto h_{\theta}^n(x)$ is also K -Lipschitz, since it is an infimum of K -Lipschitz functions, where K is a constant not depending on $y \in \Omega$. Hence, it follows that for any $x, z \in \Omega$ and any $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned} |\tilde{f}_{\theta_1}^n(x) - \tilde{f}_{\theta_2}^n(z)| &\leq |h_{\theta_1}^n(x) - h_{\theta_2}^n(z)| + \int_{\Theta} |h_u^n(x) - h_u^n(z)| du \\ &\leq K \left(1 + \int_{\Theta} du \right) \max(|x-z|, |\theta_1 - \theta_2|). \end{aligned} \tag{2.6}$$

Hence, the functions $(x, \theta) \mapsto \tilde{f}_{\theta}^n(x)$ are \tilde{K} -Lipschitz (with $\tilde{K} = K(1 + \int_{\Theta} du)$) on $\Omega \times \Theta$. Moreover, using the fact that $h_{\theta}^n(0) = 0$, it follows that for any $x \in \Omega$ and any $\theta \in \Theta$

$$|\tilde{f}_{\theta}^n(x)| = |\tilde{f}_{\theta}^n(x) - \tilde{f}_{\theta}^n(0)| \leq K|x| \leq K\delta(\Omega). \tag{2.7}$$

Now, and let us define, the set

$$\mathcal{A} = \left\{ (x, \theta) \mapsto \tilde{f}_{\theta}^n(x), n \in \mathbb{N} \right\}.$$

One can check that \mathcal{A} is a subset of $C(\Omega \times \Theta, \mathbb{R})$. By inequalities (2.6) and (2.7), the set \mathcal{A} is equicontinuous and uniformly bounded. Thus, one can use the Ascoli-Arzelà theorem to obtain

that there exists a subsequence of functions $(x, \theta) \mapsto \tilde{f}_\theta^{\varphi(n)}(x)$ that converge uniformly on $\Omega \times \Theta$ to some $(x, \theta) \mapsto \tilde{f}_\theta(x) \in C(\Omega \times \Theta, \mathbb{R})$, where $\varphi(n)$ is an increasing sequence of positive integers. It is clear that $\int_\Theta |\tilde{f}_\theta(x)| d\theta < +\infty$. Moreover, by inequality (2.7) and Lebesgue's dominated convergence theorem, one finally obtains that $\int_\Theta \tilde{f}_\theta(x) d\theta = \lim_{n \rightarrow +\infty} \int_\Theta \tilde{f}_\theta^{\varphi(n)}(x) d\theta = 0$ for every $x \in \Omega$. Therefore, one has that $\tilde{f}^\Theta = (\tilde{f}_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ with $\int_\Theta \tilde{f}_\theta(x) d\theta = 0$ for every $x \in \Omega$.

Since $S_{g(\theta)}$ is upper semi-continuous (u.s.c.) on X it follows that

$$\begin{aligned} \limsup_n S_{g(\theta)} \tilde{f}_\theta^{\varphi(n)}(x) &\leq \inf_{y \in \mathbb{R}^d} \left\{ \limsup_n \left(\frac{g(\theta)}{2} |x - y|^2 - \tilde{f}_\theta^{\varphi(n)}(y) \right) \right\} \\ &\leq \inf_{y \in \mathbb{R}^d} \left\{ \frac{g(\theta)}{2} |x - y|^2 - \tilde{f}_\theta(y) \right\} = S_{g(\theta)} \tilde{f}_\theta(x). \end{aligned}$$

Using that $\frac{g(\theta)}{2} |\cdot|^2 - S_{g(\theta)} \tilde{f}_\theta^{\varphi(n)}(\cdot)$ is a non-negative function and given that the function $(x, \theta) \mapsto \frac{g(\theta)}{2} |x|^2$ is integrable on $\Omega \times \Theta$ with respect to the measure $d\mu_\theta(x) d\theta$, Fatou's Lemma implies that

$$\begin{aligned} \limsup_n \int_\Theta \int_{\mathbb{R}^d} S_{g(\theta)} \tilde{f}_\theta^{\varphi(n)}(x) d\mu_\theta(x) d\theta &\leq \int_\Theta \int_{\mathbb{R}^d} \limsup_n S_{g(\theta)} \tilde{f}_\theta^{\varphi(n)}(x) d\mu_\theta(x) d\theta \\ &\leq \int_\Theta \int_{\mathbb{R}^d} S_{g(\theta)} \tilde{f}_\theta(x) d\mu_\theta(x) d\theta, \end{aligned}$$

which shows that

$$J_{\mathcal{P}^*} = \int_\Theta \int_{\mathbb{R}^d} S_{g(\theta)} \tilde{f}_\theta(x) d\mu_\theta(x) d\theta,$$

and thus that \tilde{f}^Θ is a maximizer of problem (\mathcal{P}^*) . □

2.4 Uniqueness of the population barycenter

Let us now use the duality between problems (\mathcal{P}) and (\mathcal{P}^*) to characterize more precisely the population barycenter. The following theorem is the main result of this section.

Theorem 2.1. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then, the population barycenter μ^* defined by (2.2) exists and is unique. Moreover, the following statements are equivalent:*

1. *The measure μ^* is the unique minimizer of problem (\mathcal{P})*
2. *If $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ is a maximizer of problem (\mathcal{P}^*) , then for every $\theta \in \Theta$ such that $g(\theta) > 0$*

$$\mu^* = \nabla \phi_\theta \# \mu_\theta \tag{2.8}$$

where $\phi_\theta : \Omega \rightarrow \mathbb{R}$ is the convex function defined by

$$\phi_\theta(x) = \frac{1}{2} |x|^2 - \frac{1}{g(\theta)} S_{g(\theta)} f_\theta(x), \text{ for all } x \in \Omega.$$

Proof. We proceed in a way that is similar to what has been done in [2] to characterize an empirical barycenter. In the proof, we denote by

$$\Theta_g = \{\theta \in \Theta : g(\theta) > 0\}$$

the support of g .

Let $f^\Theta \in L^1(\Theta, X)$ be a maximizer of problem (\mathcal{P}^*) . By Proposition 2.2 and Proposition 2.3 it follows that there exists a minimizer μ^* of (\mathcal{P}) such that

$$\begin{aligned} \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_\theta) g(\theta) d\theta &= \int_{\Theta} \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) d\theta \\ &= \int_{\Theta} \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) d\theta + \int_{\Theta} \int_{\Omega} f_\theta(x) d\mu^*(x) d\theta, \end{aligned} \quad (2.9)$$

using Fubini's theorem and the fact that $\int_{\Theta} f_\theta(x) d\theta = 0$ for all $x \in \Omega$ to obtain the last equality. Thanks to the Kantorovich duality formula (see e.g. [31], or Lemma 2.1 in [2]) we have that

$$\begin{aligned} \frac{1}{2} d_{W_2}^2(\mu^*, \mu_\theta) g(\theta) &= H_{g(\theta)}^*(\mu^*) \\ &= \sup_{f \in X} \left\{ \int_{\Omega} S_{g(\theta)} f(x) d\mu_\theta(x) + \int_{\Omega} f(x) d\mu^*(x) \right\} \\ &\geq \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(x) d\mu^*(x). \end{aligned} \quad (2.10)$$

Therefore by combining (2.9) and (2.10), we necessarily have that

$$\frac{1}{2} d_{W_2}^2(\mu^*, \mu_\theta) g(\theta) = \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(x) d\mu^*(x), \quad (2.11)$$

for every $\theta \in \Theta_g$.

Now, let $\gamma_\theta \in \Pi(\mu_\theta, \mu^*)$ be an optimal transport plan between μ_θ and μ^* . By definition of γ_θ and by (2.11), one obtains that for every $\theta \in \Theta_g$,

$$\begin{aligned} \frac{g(\theta)}{2} \int_{\Omega \times \Omega} |x - y|^2 d\gamma_\theta(x, y) &= \frac{g(\theta)}{2} d_{W_2}^2(\mu^*, \mu_\theta) \\ &= \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(y) d\mu^*(y) \\ &= \int_{\Omega \times \Omega} (S_{g(\theta)} f_\theta(x) + f_\theta(y)) d\gamma_\theta(x, y). \end{aligned} \quad (2.12)$$

Since $\frac{g(\theta)}{2} |x - y|^2 \geq S_{g(\theta)} f_\theta(x) + f_\theta(y)$ (by definition of $S_{g(\theta)} f_\theta(x)$), equality (2.12) implies that

$$\frac{g(\theta)}{2} |x - y|^2 = S_{g(\theta)} f_\theta(x) + f_\theta(y), \quad \gamma_\theta - a.e., \quad (2.13)$$

where the notation $\gamma_\theta - a.e.$ means that the above equality holds for all (x, y) in a set $A_\theta \subset \Omega \times \Omega$ of measure $\gamma_\theta(A_\theta) = 1$.

It is not difficult to check that $S_{g(\theta)}(S_{g(\theta)}f_\theta) \geq f_\theta$. Therefore, by equality (2.13) one obtains that

$$f_\theta(y) = \frac{g(\theta)}{2}|x-y|^2 - S_{g(\theta)}f_\theta(x) \geq S_{g(\theta)}(S_{g(\theta)}f_\theta)(y), \quad \gamma_\theta - a.e.$$

and thus

$$f_\theta = S_{g(\theta)}(S_{g(\theta)}f_\theta), \quad \mu^* - a.e., \quad (2.14)$$

for every $\theta \in \Theta_g$. Thus, by the constraint that $\int_\Theta f_\theta(x)d\theta = 0$ for all $x \in \Omega$, one has that

$$\int_\Theta S_{g(\theta)}(S_{g(\theta)}f_\theta)(x)d\theta = 0, \quad \mu^* - a.e. \quad (2.15)$$

For every $\theta \in \Theta_g$, introduce the convex function ϕ_θ defined by

$$\phi_\theta(x) = \frac{1}{2}|x|^2 - \frac{1}{g(\theta)}S_{g(\theta)}f_\theta(x), \quad (2.16)$$

and its conjugate ϕ_θ^* defined by

$$\phi_\theta^*(y) = \frac{1}{2}|y|^2 - \frac{1}{g(\theta)}S_{g(\theta)}(S_{g(\theta)}f_\theta(y)).$$

Let us denote by

$$\partial\phi_\theta = \{(x, y) \in \Omega \times \Omega : \phi_\theta(x) + \phi_\theta^*(y) = \langle x, y \rangle\}$$

the graph of its subdifferential. Let (x, y) be in the support of the measure γ_θ . By (2.13) and (2.14) it follows that

$$\begin{aligned} g(\theta)\langle x, y \rangle &= -S_{g(\theta)}f_\theta(x) + \frac{g(\theta)}{2}|x|^2 - f_\theta(y) + \frac{g(\theta)}{2}|y|^2 \\ &= g(\theta)\phi_\theta(x) - S_{g(\theta)}(S_{g(\theta)}f_\theta)(y) + \frac{g(\theta)}{2}|y|^2 = g(\theta)\phi_\theta(x) + g(\theta)\phi_\theta^*(y). \end{aligned} \quad (2.17)$$

By equality (2.17), it follows that if $\theta \in \Theta_g$, then $(x, y) \in \partial\phi_\theta$, which shows that the support of γ_θ is included in $\partial\phi_\theta$. Moreover, one can check that if $\theta \in \Theta_g$, then ϕ_θ is the solution of

$$\phi_\theta = \arg \min_{\phi \in \mathcal{C}} \left\{ \int_\Omega \phi(x)d\mu_\theta(x) + \int_\Omega \phi^*(x)d\mu^*(x) \right\}, \quad (2.18)$$

where \mathcal{C} denotes the set of convex functions $\phi : \Omega \rightarrow \mathbb{R}$ that are lower semi-continuous.

Thanks to Assumption 1, the measure μ_θ admits a density with respect to the Lebesgue measure for every $\theta \in \Theta$. Then, let us recall that we have shown previously that, if $\theta \in \Theta_g$, then the support of the optimal transport plan γ_θ between μ_θ and μ^* is included in $\partial\phi_\theta$. Hence, by Proposition 2.1, it follows that there exists a unique convex function $\phi_\theta : \Omega \rightarrow \mathbb{R}$, solution of the optimisation problem (2.18), such that

$$\mu^* = \nabla\phi_\theta\#\mu_\theta \quad (2.19)$$

for every θ in the support Θ_g of g . Since the convex function ϕ_θ is defined by the equation (2.16), it is clear that ϕ_θ does not depend on μ^* but only on f_θ and $g(\theta)$ for $\theta \in \Theta_g$. Therefore, by equation (2.19), the population barycenter μ^* is necessarily unique, which completes the proof of Theorem 2.1. □

3 Convergence of the empirical barycenter

Let us now prove the convergence of the empirical barycenter for the set of compactly supported measures introduced in Section 2. Let $\theta_1, \dots, \theta_n$ be iid random variables in \mathbb{R}^p with distribution \mathbb{P}_Θ . Then, let us define the functional

$$J_n(\nu) = \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}), \quad \nu \in \mathcal{M}_+^1(\Omega), \quad (3.1)$$

and consider the optimization problem: find an empirical barycenter

$$\bar{\mu}_n \in \arg \min_{\nu \in \mathcal{M}_+^1(\Omega)} J_n(\nu) = \arg \min_{\nu \in \mathcal{M}_+^1(\Omega)} \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}), \quad (3.2)$$

Thanks to the results in [2], the following lemma holds:

Lemma 3.1. *Suppose that Assumption 1 holds. Then, for any $n \geq 1$, there exists a unique minimizer $\bar{\mu}_n$ of $J_n(\cdot)$ over $\mathcal{M}_+^1(\Omega)$.*

Let us now give our main result on the convergence of the empirical barycenter $\bar{\mu}_n$.

Theorem 3.1. *Suppose that Assumption 1 and Assumption 2 hold. Let μ^* be the population barycenter defined by (2.2), and $\bar{\mu}_n$ be the empirical barycenter defined by (3.2). Then,*

$$\lim_{n \rightarrow +\infty} d_{W_2}(\bar{\mu}_n, \mu^*) = 0 \text{ almost surely (a.s.)}$$

Proof. Some part of the proof is inspired by the proof of Theorem 1 in [15]. For $\nu \in \mathcal{M}_+^1(\Omega)$, let us define

$$\Delta_n(\nu) = J_n(\nu) - J(\nu).$$

The proof is divided in two steps. First, we prove the uniform convergence to zero of Δ_n over $\mathcal{M}_+^1(\Omega)$. Then, we show that any converging subsequence of $\bar{\mu}_n$ converges a.s. to μ^* for the 2-Wasserstein distance.

Step 1. For $\nu \in \mathcal{M}_+^1(\Omega)$, let us denote by $f_\nu : \mathcal{M}_+^1(\Omega) \rightarrow \mathbb{R}$ the real-valued function defined by

$$f_\nu(\mu) = \frac{1}{2} d_{W_2}^2(\nu, \mu).$$

Then, let us define the following class of functions

$$\mathcal{F} = \{f_\nu, \nu \in \mathcal{M}_+^1(\Omega)\}.$$

Since Ω is compact with diameter $\delta(\Omega)$, \mathcal{F} is a class of functions uniformly bounded by $\frac{1}{2}\delta^2(\Omega)$ (for the supremum norm). Now, let $\nu, \mu, \mu' \in \mathcal{M}_+^1(\Omega)$. By the triangle reverse inequality

$$\begin{aligned} |f_\nu(\mu) - f_\nu(\mu')| &= \frac{1}{2} |d_{W_2}^2(\nu, \mu) - d_{W_2}^2(\nu, \mu')| \leq \delta(\Omega) |d_{W_2}(\nu, \mu) - d_{W_2}(\nu, \mu')| \\ &\leq \delta(\Omega) d_{W_2}(\mu, \mu'). \end{aligned}$$

The above inequality proves that \mathcal{F} is an equicontinuous family of functions. Now, let $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ be iid random vectors in \mathbb{R}^p with density g , and let us define the random empirical measure on $(\mathcal{M}_+^1(\Omega), \mathcal{B}(\mathcal{M}_+^1(\Omega)))$

$$\mathbb{P}_g^n = \frac{1}{n} \sum_{i=1}^n \delta_{\mu_{\boldsymbol{\theta}_i}} = \frac{1}{n} \sum_{i=1}^n \delta_{\phi(\boldsymbol{\theta}_i)},$$

where δ_μ denotes the Dirac measure at $\nu = \mu$. It is clear that

$$\Delta_n(\nu) = \int_{\mathcal{M}_+^1(\Omega)} f_\nu(\mu) d\mathbb{P}_g^n(\mu) - \int_{\mathcal{M}_+^1(\Omega)} f_\nu(\mu) d\mathbb{P}_g(\mu).$$

Let $f : \mathcal{M}_+^1(\Omega) \rightarrow \mathbb{R}$ be a real-valued function that is continuous (for the topology induced by d_{W_2}) and bounded. Thanks to the mesurability of the mapping ϕ , one has that the real random variable $\int_{\mathcal{M}_+^1(\Omega)} f(\mu) d\mathbb{P}_g^n(\mu)$ converges a.s. to $\int_{\mathcal{M}_+^1(\Omega)} f(\mu) d\mathbb{P}_g(\mu)$ as $n \rightarrow +\infty$, meaning that the random measure \mathbb{P}_g^n a.s. converges to \mathbb{P}_g in the weak sense. Therefore, since \mathcal{F} is a uniformly bounded and equicontinuous family of functions, one can use Theorem 6.2 in [25] to obtain that

$$\sup_{\nu \in \mathcal{M}_+^1(\Omega)} |\Delta_n(\nu)| = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{M}_+^1(\Omega)} f(\mu) d\mathbb{P}_g^n(\mu) - \int_{\mathcal{M}_+^1(\Omega)} f(\mu) d\mathbb{P}_g(\mu) \right| \rightarrow 0 \text{ as } n \rightarrow +\infty, \text{ a.s.} \quad (3.3)$$

which proves the uniform convergence of Δ_n to zero over $\mathcal{M}_+^1(\Omega)$.

Step 2. Suppose that Assumption 1 and Assumption 2 hold. By Lemma 3.1, there exists a unique sequence $(\bar{\boldsymbol{\mu}}_n)_{n \geq 1}$ of empirical barycenters defined by (3.2). Thanks to the compactness of the Wasserstein space $(\mathcal{M}_+^1(\Omega), d_{W_2})$, one can extract a converging sub-sequence of empirical barycenters $(\bar{\boldsymbol{\mu}}_{n_k})_{k \geq 1}$ such that $\lim_{k \rightarrow +\infty} d_{W_2}(\bar{\boldsymbol{\mu}}_{n_k}, \bar{\boldsymbol{\mu}}) = 0$ for some measure $\bar{\boldsymbol{\mu}} \in \mathcal{M}_+^1(\Omega)$.

Let us now prove that $\bar{\boldsymbol{\mu}} = \mu^*$. To this end, let us first note that by the definition of $\bar{\boldsymbol{\mu}}_{n_k}$ and μ^* as the unique minimizer of $J_{n_k}(\cdot)$ and $J(\cdot)$ respectively, it follows that

$$\begin{aligned} |J(\bar{\boldsymbol{\mu}}_{n_k}) - J(\mu^*)| &= J(\bar{\boldsymbol{\mu}}_{n_k}) - J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) + J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) - J_{n_k}(\mu^*) + J_{n_k}(\mu^*) - J(\mu^*) \\ &\leq 2 \sup_{\nu \in \mathcal{M}_+^1(\Omega)} |\Delta_{n_k}(\nu)|, \end{aligned}$$

where we have used the fact that $J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) - J_{n_k}(\mu^*) \leq 0$. Therefore, thanks to the uniform convergence (3.3) of Δ_n to zero over $\mathcal{M}_+^1(\Omega)$, one obtains that

$$\lim_{k \rightarrow +\infty} J(\bar{\boldsymbol{\mu}}_{n_k}) = J(\mu^*). \quad (3.4)$$

Therefore, using that

$$\begin{aligned} |J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) - J(\mu^*)| &\leq |J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) - J(\bar{\boldsymbol{\mu}}_{n_k})| + |J(\bar{\boldsymbol{\mu}}_{n_k}) - J(\mu^*)| \\ &\leq \sup_{\nu \in \mathcal{M}_+^1(\Omega)} |\Delta_{n_k}(\nu)| + |J(\bar{\boldsymbol{\mu}}_{n_k}) - J(\mu^*)|, \end{aligned}$$

one finally obtains by (3.3) and (3.4) that

$$\lim_{k \rightarrow +\infty} J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) = J(\mu^*). \quad (3.5)$$

Since $|J_{n_k}(\bar{\boldsymbol{\mu}}) - J(\bar{\boldsymbol{\mu}})| \leq \sup_{\nu \in \mathcal{M}_+^1(\Omega)} |\Delta_{n_k}(\nu)|$, it follows by equation (3.3) that

$$\lim_{k \rightarrow +\infty} J_{n_k}(\bar{\boldsymbol{\mu}}) = J(\bar{\boldsymbol{\mu}}) \text{ a.s.} \quad (3.6)$$

Moreover, for any $\epsilon > 0$, there exists $k_\epsilon \in \mathbb{N}$ such that $d_{W_2}(\bar{\boldsymbol{\mu}}_{n_k}, \bar{\boldsymbol{\mu}}) \leq \epsilon$ for all $k \geq k_\epsilon$. Therefore, using the triangle inequality, it follows that for all $k \geq k_\epsilon$

$$\begin{aligned} (J_{n_k}(\bar{\boldsymbol{\mu}}))^{1/2} &= \left(\frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\boldsymbol{\mu}}, \mu_{\boldsymbol{\theta}_j}) \right)^{1/2} \\ &\leq \left(\frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\boldsymbol{\mu}}_{n_k}, \mu_{\boldsymbol{\theta}_j}) \right)^{1/2} + \left(\frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}_{n_k}) \right)^{1/2} \\ &\leq \left(\frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\boldsymbol{\mu}}_{n_k}, \mu_{\boldsymbol{\theta}_j}) \right)^{1/2} + \frac{\epsilon}{\sqrt{2}}, \end{aligned}$$

and thus by equations (3.5) and (3.6), we obtain that

$$J(\bar{\boldsymbol{\mu}}) \leq \lim_{k \rightarrow +\infty} J_{n_k}(\bar{\boldsymbol{\mu}}_{n_k}) = J(\mu^*) \text{ a.s.} \quad (3.7)$$

which finally proves that $\bar{\boldsymbol{\mu}} = \mu^*$ a.s. since μ^* is the unique minimizer of $J(\nu)$ over $\nu \in \mathcal{M}_+^1(\Omega)$.

Hence, any converging subsequence of empirical barycenters converges a.s. to μ^* for the 2-Wasserstein distance. Since $(\mathcal{M}_+^1(\Omega), d_{W_2})$ is compact, this finally shows that $(\bar{\boldsymbol{\mu}}_n)_{n \geq 1}$ is a converging sequence such that $\lim_{n \rightarrow +\infty} d_{W_2}(\bar{\boldsymbol{\mu}}_n, \mu^*) = 0$ a.s., which completes the proof of Theorem 3.1. \square

4 Application to statistical models for curve and image warping

In this section, we present some statistical models for which the notion of population and empirical barycenters in the 2-Wasserstein space are useful.

In many applications observations are in the form of a set of n gray-level curves or images $\mathbf{X}_1, \dots, \mathbf{X}_n$ (e.g. in geophysics, biomedical imaging or in signal processing for neurosciences), which can be considered as iid random variables belonging to the set $L^2(\Omega)$ of square-integrable and real-valued functions on a compact domain of Ω of \mathbb{R}^d . In many situations the observed curves or images share the same structure. This may lead to the assumption that these observations are random elements which vary around the same but unknown mean pattern (also called reference template). Estimating such a mean pattern and characterizing the modes of individual variations around this template is of fundamental interest.

Due to additive noise and geometric variability in the data, this mean pattern is typically unknown, and it has to be estimated. In this setting, a widely used approach is Grenander's pattern theory [17, 18, 29, 30] that models geometric variability by the action of a Lie group on an infinite dimensional space of curves (or images). Following the ideas of Grenander's pattern

theory, a simple assumption is to consider that the data $\mathbf{X}_1, \dots, \mathbf{X}_n$ are obtained through the deformation of the same reference template $h \in L^2(\Omega)$ via the so-called deformable model

$$\mathbf{X}_i = h \circ \varphi_i^{-1}, \quad i = 1, \dots, n, \quad (4.1)$$

where $\varphi_1, \dots, \varphi_n$ are iid random variables belonging to the set of smooth diffeomorphisms of Ω . In signal and image processing, there has been recently a growing interest on the statistical analysis of deformable models (4.1) using either rigid or non-rigid random diffeomorphisms φ_i , see e.g. [3, 7, 8, 9, 10, 14, 32] and references therein. In a data set of curves or images, one generally observes not only a source of variability in geometry, but also a source of photometric variability (the intensity of a pixel changes from one image to another) that cannot be only captured by a deformation of the domain Ω via a diffeomorphism as in model (4.1).

It is always possible to transform the data $\mathbf{X}_1, \dots, \mathbf{X}_n$ into a set of n iid random probability densities by computing the random variables

$$\mathbf{Y}_i(x) = \frac{\widetilde{\mathbf{X}}_i(x)}{\int_{\Omega} \widetilde{\mathbf{X}}_i(u)} du, \quad x \in \Omega, \quad \text{where } \widetilde{\mathbf{X}}_i(x) = \mathbf{X}_i(x) - \min_{u \in \Omega} \{\mathbf{X}_i(u)\}, \quad i = 1, \dots, n.$$

Let $q \in L^2(\Omega)$ be a probability density function, and consider the deformable model of densities

$$\mathbf{Y}_i(x) = |\det(D\varphi_i^{-1})(x)| q(\varphi_i^{-1}(x)) \quad x \in \Omega, \quad i = 1, \dots, n, \quad (4.2)$$

where $\det(D\varphi_i^{-1})(x)$ denotes the determinant of the Jacobian matrix of the random diffeomorphism φ_i^{-1} at point x . If we denote by $\mu_1, \dots, \mu_n \in \mathcal{M}_+^1(\Omega)$ the random probability measures with densities $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, and by μ the measure with density q , then (4.2) can also be written as the following deformable model of measures

$$\mu_i = \varphi_i \# \mu, \quad i = 1, \dots, n. \quad (4.3)$$

In model (4.3), computing the empirical barycenter in the Wasserstein space of the random measures μ_1, \dots, μ_n may lead to consistent and meaningful estimators of the reference measure μ and thus of the mean pattern q . In the rest of this section, we discuss some examples of model (4.3). In particular, we show how the results of Section 2 can be used to characterise the population barycenter of random measures satisfying the deformable model (4.3).

4.1 A parametric class of diffeomorphisms

Let μ be a measure on \mathbb{R}^d having a density q (with respect to the Lebesgue measure dx on \mathbb{R}^d) whose support is contained in compact set $\Omega_q \subset \mathbb{R}^d$. We propose to characterise the population barycenter of a random measure μ satisfying the deformable model

$$\mu = \varphi \# \mu, \quad (4.4)$$

for a specific class of random diffeomorphisms $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Let $\mathbb{S}_d^+(\mathbb{R})$ be the set of non-negative definite $d \times d$ symmetric matrices with real entries. Let

$$\phi : (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p)) \rightarrow \left(\mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d, \mathcal{B} \left(\mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d \right) \right)$$

be a measurable mapping, where $\mathcal{B}(\mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d)$ is the Borel σ -algebra of $\mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d$. For $\theta \in \mathbb{R}^p$, we will use the notations

$$\phi(\theta) = (A_\theta, b_\theta), \text{ with } A_\theta \in \mathbb{S}_d^+(\mathbb{R}), b_\theta \in \mathbb{R}^d,$$

and

$$\varphi_\theta(x) = A_\theta x + b_\theta, x \in \mathbb{R}^d.$$

For any $\theta \in \mathbb{R}^d$, one has that $\varphi_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a smooth and bijective affine application with

$$\varphi_\theta^{-1}(x) = A_\theta^{-1}(x - b_\theta), x \in \mathbb{R}^d.$$

Let $\Theta \subset \mathbb{R}^p$ be a compact set. One can then define a parametric class of diffeomorphisms of \mathbb{R}^d as follows

$$D_\phi(\Theta) = \{\varphi_\theta, \theta \in \Theta\}. \quad (4.5)$$

Finally, let $\boldsymbol{\theta} \in \mathbb{R}^p$ be a random vector with density g (with respect to the Lebesgue measure $d\theta$ on \mathbb{R}^p) having a support included in the compact set Θ . We propose to study the population barycenter in the 2-Wasserstein space of the random measure $\mu_{\boldsymbol{\theta}}$ satisfying the deformable model

$$\mu_{\boldsymbol{\theta}} = \varphi_{\boldsymbol{\theta}} \# \mu. \quad (4.6)$$

For any $\theta \in \Theta$ (not necessarily a random vector), we define $\mu_\theta = \varphi_\theta \# \mu$. Since φ_θ is a smooth diffeomorphism and μ is a measure with density q whose support is included in the compact set Ω_q , it follows that μ_θ admits a density q_θ on \mathbb{R}^d given by

$$q_\theta(x) = \begin{cases} \det(A_\theta^{-1}) q(A_\theta^{-1}(x - b_\theta)) & \text{if } x \in \mathcal{R}(\varphi_\theta), \\ 0 & \text{if } x \notin \mathcal{R}(\varphi_\theta). \end{cases} \quad (4.7)$$

where $\mathcal{R}(\varphi_\theta) = \{\varphi_\theta(y), y \in \Omega_q\} = \{A_\theta y + b_\theta, y \in \Omega_q\}$. Before stating our main result on the population barycenter of the random measure $\mu_{\boldsymbol{\theta}}$ (4.6), let us make the following regularity assumption on the mapping ϕ .

Assumption 3. *The mapping $\phi : \Theta \rightarrow \mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d$ is continuous.*

Under Assumption 3, it follows that there exists a compact set $\Omega \subset \mathbb{R}^d$ such that $\mathcal{R}(\varphi_\theta) \subset \Omega$ for all $\theta \in \Theta$. Thus, under this assumption, the random measure $\mu_{\boldsymbol{\theta}}$ takes its values in $\mathcal{M}_+^1(\Omega)$.

4.2 Characterization of the population barycenter for parametric diffeomorphisms

Let us now give a characterization of the population barycenter of a random measure following the deformable model (4.6) with random diffeomorphism $\varphi_{\boldsymbol{\theta}}$ taking their value in the parametric class D_ϕ defined by (4.5). Before stating the main result of this section, we define, for any $\theta \in \Theta$, the following quantities

$$\bar{A}_\theta = A_\theta \bar{A}^{-1} \quad \text{and} \quad \bar{b}_\theta = b_\theta - A_\theta \bar{A}^{-1} \bar{b}, \quad (4.8)$$

where $\bar{A} = \mathbb{E}(A_{\boldsymbol{\theta}})$ and $\bar{b} = \mathbb{E}(b_{\boldsymbol{\theta}})$.

Theorem 4.1. *Let $\theta \in \mathbb{R}^p$ be a random vector with a density $g : \Theta \rightarrow \mathbb{R}$ that is continuously differentiable and such that $g(\theta) > 0$ for all $\theta \in \Theta$. Let μ_θ be the random measure defined by the deformable model (4.6). Suppose that Assumption 3 holds.*

Then, the population barycenter μ^ defined by (2.2) exists and is unique. Moreover, let us define the density*

$$q^*(x) = \det(\bar{A}^{-1})q(\bar{A}^{-1}(x - \bar{b})), \quad x \in \Omega, \quad (4.9)$$

where $\bar{A} = \mathbb{E}(A_\theta)$, $\bar{b} = \mathbb{E}(b_\theta)$, and $\bar{A}_\theta, \bar{b}_\theta$ are the random variables defined by (4.8). Then, the following statements hold:

1. *The primal problem (\mathcal{P}) satisfies*

$$J_{\mathcal{P}} = \inf_{\nu \in \mathcal{M}_+^1(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_\theta) g(\theta) d\theta = \frac{1}{2} \int_{\Omega} \mathbb{E}(|\bar{A}_\theta u + \bar{b}_\theta - u|^2) q^*(u) du, \quad (4.10)$$

and the dual problem (\mathcal{P}^*) admits a maximizer at $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ where, for $\theta \in \Theta$,

$$f_\theta(x) = -\frac{g(\theta)}{2} \langle (\bar{A}_\theta - I)x, x \rangle - g(\theta) \langle \bar{b}_\theta, x \rangle, \quad x \in \Omega, \quad (4.11)$$

where I is the $d \times d$ identity matrix.

2. *The population barycenter is the measure $\mu^* \in \mathcal{M}_+^1(\Omega)$ with density q^* (with respect to the Lebesgue measure on \mathbb{R}^d) given by equation 4.9.*

Proof. Under the assumptions of Theorem 4.1, it is clear that Assumption 1 and Assumption 2 are satisfied. Therefore, by Theorem (2.1), there exists a unique population barycenter μ^* of the random measure μ_θ defined by (4.6). To prove the results stated in Theorem 4.1, we will use the characterization (2.8) of the barycenter μ^* . For this purpose, we need to find a maximizer $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ of the dual problem (\mathcal{P}^*).

Let $\bar{A} = \mathbb{E}(A_\theta)$ and $\bar{b} = \mathbb{E}(b_\theta)$. By defining the density

$$\bar{q}(x) = \det(\bar{A}^{-1})q(\bar{A}^{-1}(x - \bar{b})), \quad x \in \mathbb{R}^d,$$

one can re-parametrize the density q_θ , given by (4.7) for any $\theta \in \Theta$, as follows

$$q_\theta(x) = \det(\bar{A}_\theta^{-1}) \bar{q}(\bar{A}_\theta^{-1}(x - \bar{b}_\theta)), \quad x \in \mathbb{R}^d, \quad (4.12)$$

where

$$\bar{A}_\theta = A_\theta \bar{A}^{-1} \quad \text{and} \quad \bar{b}_\theta = b_\theta - A_\theta \bar{A}^{-1} \bar{b}.$$

In the proof, we will denote by $\Omega_{\bar{q}}$ the support of the density \bar{q} . Note that the random variables \bar{A}_θ and \bar{b}_θ are such that

$$\mathbb{E}(\bar{A}_\theta) = \int_{\Theta} \bar{A}_\theta g(\theta) d\theta = I \quad \text{and} \quad \mathbb{E}(\bar{b}_\theta) = \int_{\Theta} \bar{b}_\theta g(\theta) d\theta = 0,$$

where I denotes the $d \times d$ identity matrix.

Proof of statement 1. of Theorem 4.1.

a) Let us first compute an upper bound of $J_{\mathcal{P}^*}$. Let $f^\Theta \in L^1(\Theta, X)$ be such that $\int_\Theta f_\theta(x) d\theta = 0$ for all $x \in \Omega$. By definition of $S_{g(\theta)} f_\theta(x)$ one has that

$$S_{g(\theta)} f_\theta(x) \leq \frac{g(\theta)}{2} |x - y|^2 - f_\theta(y) \quad (4.13)$$

for any $y \in \Omega$. By using equation (4.12) and inequality (4.13) with $y = \bar{A}_\theta^{-1}(x - \bar{b}_\theta)$ one obtains that

$$\begin{aligned} \int_\Theta \int_\Omega S_{g(\theta)} f_\theta(x) q_\theta(x) dx d\theta &\leq \int_\Theta \int_\Omega \left(\frac{g(\theta)}{2} |x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 - f_\theta(\bar{A}_\theta^{-1}(x - \bar{b}_\theta)) \right) q_\theta(x) dx d\theta \\ &\leq \int_\Theta \int_{\Omega_{\bar{q}}} \left(\frac{g(\theta)}{2} |\bar{A}_\theta u + \bar{b}_\theta - u|^2 - f_\theta(u) \right) \bar{q}(u) du d\theta \\ &\leq \int_\Theta \int_{\Omega_{\bar{q}}} \left(\frac{g(\theta)}{2} |\bar{A}_\theta u + \bar{b}_\theta - u|^2 \right) \bar{q}(u) du d\theta \end{aligned}$$

Note that to obtain the second inequality above, we have used the change of variable $u = \bar{A}_\theta^{-1}(x - \bar{b}_\theta)$, while the third inequality has been obtained using with the fact that $\int_\Theta f_\theta(u) d\theta = 0$ for any $u \in \Omega_{\bar{q}}$ combined with Fubini's theorem. Thanks to the compactness of Θ and $\Omega_{\bar{q}}$, and using Assumption 3, it follows that

$$\int_{\Omega_{\bar{q}}} \mathbb{E} (|\bar{A}_\theta u + \bar{b}_\theta - u|^2) \bar{q}(u) du < +\infty.$$

Therefore, we have shown that

$$J_{\mathcal{P}^*} \leq \frac{1}{2} \int_{\Omega_{\bar{q}}} \mathbb{E} (|\bar{A}_\theta u + \bar{b}_\theta - u|^2) \bar{q}(u) du. \quad (4.14)$$

b) Let us recall that we have assumed that $g(\theta) > 0$ for any $\theta \in \Theta$. Now, for $\theta \in \Theta$ we define the function

$$f_\theta(x) = -\frac{g(\theta)}{2} \langle (\bar{A}_\theta - I)x, x \rangle - g(\theta) \langle \bar{b}_\theta, x \rangle.$$

First, one can note that $f^\Theta = (f_\theta)_{\theta \in \Theta}$ belongs to $L^1(\Theta, X)$. Since $\int_\Theta \bar{A}_\theta g(\theta) d\theta = I$ and $\int_\Theta \bar{b}_\theta g(\theta) d\theta = 0$, one has also that $\int_\Theta f_\theta(x) d\theta = 0$. Let us now consider the function $F = \mathbb{R}^d \rightarrow \mathbb{R}$ defined as

$$F(y) = \frac{g(\theta)}{2} |x - y|^2 + \frac{g(\theta)}{2} \langle (\bar{A}_\theta - I)y, y \rangle + g(\theta) \langle \bar{b}_\theta, y \rangle, \quad y \in \mathbb{R}^d.$$

Searching for some $y \in \mathbb{R}^d$, where the gradient of F vanishes, leads to the equation

$$0 = -g(\theta)(x - y) + g(\theta) ((\bar{A}_\theta - I)y + \bar{b}_\theta) = -g(\theta)x + g(\theta) (\bar{A}_\theta y + \bar{b}_\theta).$$

Hence, the function $y \mapsto F(y)$ has a minimum at $y = \bar{A}_\theta^{-1}(x - \bar{b}_\theta)$. Therefore,

$$\begin{aligned}
S_{g(\theta)}f_\theta(x) &= \frac{g(\theta)}{2}|x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 + \frac{g(\theta)}{2}\langle (\bar{A}_\theta - I)(\bar{A}_\theta^{-1}(x - \bar{b}_\theta)), \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&\quad + g(\theta)\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&= \frac{g(\theta)}{2}|x|^2 - g(\theta)\langle x, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle + \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \\
&\quad + \frac{g(\theta)}{2}\langle (x - \bar{b}_\theta), \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle - \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \\
&\quad + g(\theta)\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&= \frac{g(\theta)}{2}|x|^2 - \frac{g(\theta)}{2}\langle x, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle + \frac{g(\theta)}{2}\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \tag{4.15}
\end{aligned}$$

$$\begin{aligned}
&= \frac{g(\theta)}{2}|x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 + \frac{g(\theta)}{2}\langle x + \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&\quad - \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \tag{4.16}
\end{aligned}$$

Let us introduce the notation $J^*(f^\ominus) = \int_\Theta \int_\Omega S_{g(\theta)}f_\theta(x)d\mu_\theta(x)d\theta$. By equation (4.16) and using the re-parametrization (4.12) of q_θ combined with the change of variable $u = \bar{A}_\theta^{-1}(x - \bar{b}_\theta)$, it follows that

$$\begin{aligned}
J^*(f^\ominus) &= \int_\Theta \int_\Omega \frac{g(\theta)}{2}|x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 q_\theta(x)dx d\theta + \int_\Theta \int_\Omega \frac{g(\theta)}{2}\langle x + \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle q_\theta(x)dx d\theta \\
&\quad - \int_\Theta \int_\Omega \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 q_\theta(x)dx d\theta \\
&= \int_\Theta \int_{\Omega_{\bar{q}}} \frac{g(\theta)}{2}|\bar{A}_\theta u + \bar{b}_\theta - u|^2 \bar{q}(u)du d\theta + \int_\Theta \int_{\Omega_{\bar{q}}} \frac{g(\theta)}{2}\langle \bar{A}_\theta u + 2\bar{b}_\theta, u \rangle \bar{q}(u)du d\theta \\
&\quad - \int_\Theta \int_{\Omega_{\bar{q}}} \frac{g(\theta)}{2}|u|^2 \bar{q}(u)du d\theta \\
&= \frac{1}{2} \int_{\Omega_{\bar{q}}} \mathbb{E}(|\bar{A}_\theta u + \bar{b}_\theta - u|^2) \bar{q}(u)du,
\end{aligned}$$

where we have used Fubini's theorem combined with the fact that $\int_\Theta \bar{A}_\theta g(\theta)d\theta = I$ and $\int_\Theta \bar{b}_\theta g(\theta)d\theta = 0$ to obtain the last equality.

Hence, thanks to the upper bound (4.14), we finally have that

$$J^*(f^\ominus) = J_{\mathcal{P}^*} = \frac{1}{2} \int_{\Omega_{\bar{q}}} \mathbb{E}(|\bar{A}_\theta u + \bar{b}_\theta - u|^2) \bar{q}(u)du,$$

which proves that f^\ominus is a maximizer of the dual problem (\mathcal{P}^*), and this completes the proof of statement 1. of Theorem 4.1.

Proof of statement 2. of Theorem 4.1.

Since we have found a solution $f^\Theta = (f_\theta)_{\theta \in \Theta}$ of the dual problem (\mathcal{P}^*) , it follows from Theorem 2.1 that the population barycenter is given by $\mu^* = \nabla \phi_\theta \# \mu_\theta$ where

$$\phi_\theta(x) = \frac{1}{2}|x|^2 - \frac{1}{g(\theta)} S_{g(\theta)} f_\theta(x), \text{ for all } x \in \Omega,$$

for every $\theta \in \Theta$. By equation (4.15), one has that

$$\phi_\theta(x) = \frac{1}{2} \langle x, \bar{A}_\theta^{-1} (x - \bar{b}_\theta) \rangle - \frac{1}{2} \langle \bar{b}_\theta, \bar{A}_\theta^{-1} (x - \bar{b}_\theta) \rangle = \frac{1}{2} \langle x - \bar{b}_\theta, \bar{A}_\theta^{-1} (x - \bar{b}_\theta) \rangle,$$

which implies that

$$\nabla \phi_\theta = \bar{A}_\theta^{-1} (x - \bar{b}_\theta).$$

Since μ_θ is the measure with density $q_\theta(x) = \det(\bar{A}_\theta^{-1}) \bar{q}(\bar{A}_\theta^{-1} (x - \bar{b}_\theta))$, one finally has that that μ^* is a measure having a density q^* given by

$$q^*(x) = \det(\bar{A}_\theta) q_\theta(\bar{A}_\theta x + \bar{b}_\theta) = \bar{q}(x),$$

which completes the proof of statement 2. of Theorem 4.1.

□

Hence, by Theorem 4.1, the population barycenter μ^* in the deformable model (4.6) is related to the template measure μ by the equation

$$\mu^* = \bar{\varphi} \# \mu \text{ with } \bar{\varphi}(x) = \mathbb{E}(A_\theta) x + \mathbb{E}(b_\theta), \quad x \in \mathbb{R}^d.$$

Note that the mapping $\bar{\varphi}$ is the expectation of the random diffeomorphism $\varphi_\theta(x) = A_\theta x + b_\theta$, $x \in \mathbb{R}^d$. Therefore, computing the population barycenter in the Wasserstein space of a measure from the deformable model (4.6) with such random diffeomorphism amounts to transport the template measure μ by the averaged amount of deformation measured by $\bar{\varphi}$. In the case where $\bar{\varphi} = I$ is the $d \times d$ identity matrix (which correspond to the assumption that $\mathbb{E}(A_\theta) = I$ and $\mathbb{E}(b_\theta) = 0$), the population barycenter μ^* is equal to the template measure μ .

4.3 The case of randomly shifted densities

To illustrate Theorem 4.1, let us consider the simplest deformable model of randomly shifted curves or images with

$$\varphi_i^{-1}(x) = x - \theta_i, \quad x \in \mathbb{R}^d,$$

in equation (4.1) for some random shift $\theta_i \in \mathbb{R}^d$. This model has recently received a lot of attention in the literature, see e.g. [7, 9, 10, 14, 32], since it represents a benchmark for the statistical analysis of deformable models. In the one-dimensional case ($d = 1$), the model of shifted curves has applications in various fields such as as neurosciences [28] or biology [26].

Let $q : \mathbb{R}^d \rightarrow \mathbb{R}^+$ be a probability density function with compact support included in $[-A, A]^d$ for some constant $A > 0$. For θ a random vector in \mathbb{R}^d , we define the random density

$$q_\theta(x) = q(x - \theta), \quad x \in \mathbb{R}^d, \tag{4.17}$$

and the associated random measure

$$d\mu_{\boldsymbol{\theta}}(x) = q_{\boldsymbol{\theta}}(x)dx.$$

Note that equation (4.17) corresponds to the deformable model (4.6) with $\varphi_{\boldsymbol{\theta}}(x) = x + \boldsymbol{\theta}$, $x \in \mathbb{R}^d$.

Now let us suppose that $\boldsymbol{\theta}$ has a continuously differentiable density g with compact support $\Theta = [-\epsilon, \epsilon]^d$ for some $\epsilon > 0$. If $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ is an iid sample of random shifts with density g , then the empirical Euclidean barycenter (standard notion of averaging) of the random densities $q_{\boldsymbol{\theta}_1}, \dots, q_{\boldsymbol{\theta}_n}$ is the probability density given by

$$\bar{q}_n(x) = \frac{1}{n} \sum_{j=1}^n q_{\boldsymbol{\theta}_j}(x). \quad (4.18)$$

By the law of large number, one has that

$$\lim_{n \rightarrow +\infty} \bar{q}_n(x) = \int_{\mathbb{R}^d} q(x - \theta)g(\theta)d\theta \text{ a.s. for any } x \in \mathbb{R}^d.$$

Therefore, the Euclidean barycenter \bar{q}_n converges to the convolution of the reference template q by the density g of the random shift $\boldsymbol{\theta}$. Hence, under mild assumptions, \bar{q}_n is not a consistent estimator of the mean pattern q .

Let us now see the benefits of using the notion of empirical barycenter in the 2-Wasserstein space to consistently estimate q . It is clear that the set of shifted measures $(\mu_{\theta})_{\theta \in \Theta}$ with densities $q_{\theta}(x) = q(x - \theta)$ is included in $\mathcal{M}_+^1(\Omega)$ with $\Omega = [-(A + \epsilon), (A + \epsilon)]^d$. Hence, Assumption 1 and Assumption 2 are satisfied. It is also clear that the mapping $\phi : \Theta \rightarrow \mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d$ defined by

$$\phi(\theta) = (I, \theta), \theta \in \Theta, \text{ where } I \text{ is the } d \times d \text{ identity matrix,}$$

is continuous, and thus Assumption 3 holds. Therefore, by Theorem 4.1, one immediately has the following result:

Corollary 4.1. *Suppose that $\boldsymbol{\theta}$ is random vector in \mathbb{R}^d having a continuously differentiable density g (with respect to the Lebesgue measure $d\theta$ on \mathbb{R}^d). Assume that g has a compact support $\Theta = [-\epsilon, \epsilon]^d$ for some $\epsilon > 0$. Let $\mu_{\boldsymbol{\theta}}$ be the random measure with density $q_{\boldsymbol{\theta}}(x) = q(x - \boldsymbol{\theta})$ (with respect to the Lebesgue measure dx) where $q : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is probability density function with compact support included in $[-A, A]^d$.*

Then, the population barycenter μ^ in the 2-Wasserstein space exists and is unique. It is the measure with density $q(x - \mathbb{E}(\boldsymbol{\theta}))$, namely*

$$d\mu^*(x) = q(x - \mathbb{E}(\boldsymbol{\theta}))dx.$$

The primal problem (\mathcal{P}) satisfies

$$J_{\mathcal{P}} = \inf_{\nu \in \mathcal{M}_+^1(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_{\theta})g(\theta)d\theta = \frac{1}{2} \mathbb{E}(|\boldsymbol{\theta} - \mathbb{E}(\boldsymbol{\theta})|^2), \quad (4.19)$$

Moreover, $f^{\Theta} = (f_{\theta})_{\theta \in \Theta} \in L^1(\Theta, X)$, with

$$f_{\theta}(x) = -g(\theta)(\theta - \mathbb{E}(\boldsymbol{\theta}))x, \quad (4.20)$$

is a maximizer of the dual problem (\mathcal{P}^*) .

Hence, if it is assumed that the random shifts have zero expectation i.e. $\mathbb{E}(\boldsymbol{\theta}) = 0$, then the density of the population barycenter μ^* is the reference template q . In this setting, thanks to Theorem 3.1, the empirical barycenter $\bar{\mu}_n$ in the 2-Wasserstein space of the randomly shifted densities $q_{\boldsymbol{\theta}_1}, \dots, q_{\boldsymbol{\theta}_n}$ is a consistent estimator of q . This illustrates the advantages of using the notion of barycenter in the Wasserstein space rather than the Euclidean barycenter \bar{q}_n , defined in (4.18), which may yield to non-consistent estimators of a mean pattern.

4.4 Related results in the literature

In the literature, there exists various applications of the notion of an empirical barycenter in the Wasserstein space for signal and image processing. For example, it has been successfully used for texture analysis in image processing [24]. The theory of optimal transport for image warping has also been shown to be useful in various applications, see e.g. [19, 20] and references therein.

Some properties of the empirical barycenter $\bar{\mu}_n$ (in the 2-Wasserstein space) of random measures satisfying a deformable model similar to (4.3) have been studied in [11]. For a specific set of measurable maps $(\varphi_i)_{i=1, \dots, n}$ that is an admissible family of deformations (in the sense of Definition 4.2 in [11]), it has been shown in [11] that the empirical barycenter $\bar{\mu}_n$ of the measures $\mu_i = \varphi_i \# \mu$, $i = 1, \dots, n$ converges almost surely to the template μ as $n \rightarrow +\infty$ for the d_{W_2} distance under the additional assumptions that the expectation of the φ_i 's is equal to the identity and in the case where the measure μ is compactly supported (see Theorem 4.4 in [11]). Therefore, the results in [11] are consistent with those obtained in Section 3 of this paper on the convergence of the empirical barycenter of random measures that are compactly supported. Nevertheless, the results that we have obtained in this paper are more general than those in [11] since our study on the consistency of the empirical barycenter is not restricted to the deformable model (4.3) with measurable maps $(\varphi_i)_{i=1, \dots, n}$ being in the so-called class of admissible deformations that is introduced in [11]. Moreover, the problem of proving the existence and the unicity of a population barycenter of a random measure, as defined by (1.1), is not considered in [11]. In this paper, we have also shown the benefits of considering the dual formulation (\mathcal{P}^*) of the (primal) problem (2.2) to characterize the population barycenter in the 2-Wasserstein space for a large class of deformable models of measures.

5 Beyond the compactly supported case

To conclude the paper, we briefly discuss the case of a random measure μ whose support is not included in a compact set Ω of \mathbb{R}^d .

Firstly, let us consider the one-dimensional case i.e. $d = 1$. Let $\mathcal{M}_+^2(\mathbb{R})$ be the set of Radon probability measures ν on \mathbb{R} having a finite second order moment (i.e. $\int_{\mathbb{R}} |x|^2 d\nu(x) < +\infty$). For $\nu \in \mathcal{M}_+^2(\mathbb{R})$, we denote by F_ν its cumulative distribution function, and by F_ν^{-1} its generalized inverse (quantile function). In the one-dimensional case ($d = 1$), it is well known that

$$d_{W_2}^2(\nu, \mu) = \int_0^1 |F_\nu^{-1}(y) - F_\mu^{-1}(y)|^2 dy$$

for any ν and μ belonging to $\mathcal{M}_+^2(\mathbb{R})$. Hence, if μ denotes a random variable with distribution

\mathbb{P} taking its values in $\mathcal{M}_+^2(\mathbb{R})$, one has that for any $\nu \in \mathcal{M}_+^2(\mathbb{R})$

$$\begin{aligned}
\mathbb{E}(d_{W_2}^2(\nu, \boldsymbol{\mu})) &= \int_{\mathcal{M}_+^2(\mathbb{R})} d_{W_2}^2(\nu, \mu) d\mathbb{P}(\mu) = \int_{\mathcal{M}_+^2(\mathbb{R})} \left(\int_0^1 |F_\nu^{-1}(y) - F_\mu^{-1}(y)|^2 dy \right) d\mathbb{P}(\mu) \\
&= \mathbb{E} \left(\int_0^1 |F_\nu^{-1}(y) - F_{\boldsymbol{\mu}}^{-1}(y)|^2 dy \right) \\
&= \int_0^1 \mathbb{E} |F_\nu^{-1}(y) - F_{\boldsymbol{\mu}}^{-1}(y)|^2 dy \\
&\geq \mathbb{E} \left(\int_0^1 \left| \mathbb{E}(F_{\boldsymbol{\mu}}^{-1}(y)) - F_{\boldsymbol{\mu}}^{-1}(y) \right|^2 dy \right). \tag{5.1}
\end{aligned}$$

by applying Fubini's theorem, and by using the inequality $\mathbb{E}|x - \mathbf{X}|^2 \geq \mathbb{E}|\mathbb{E}(\mathbf{X}) - \mathbf{X}|^2$ that holds for any random variables $\mathbf{X} \in \mathbb{R}$, and for any $x \in \mathbb{R}$. Hence, if one can define a measure $\mu^* \in \mathcal{M}_+^2(\mathbb{R})$ such that $F_{\mu^*}^{-1}(y) = \mathbb{E}(F_{\boldsymbol{\mu}}^{-1}(y))$ for all $y \in [0, 1]$, then it is clear from inequality (5.1) that

$$\int_{\mathcal{M}_+^2(\mathbb{R})} d_{W_2}^2(\nu, \mu) d\mathbb{P}(\mu) \geq \int_{\mathcal{M}_+^2(\mathbb{R})} d_{W_2}^2(\mu^*, \mu) d\mathbb{P}(\mu), \text{ for any } \nu \in \mathcal{M}_+^2(\mathbb{R}),$$

implying that μ^* is a population barycenter of the random measure $\boldsymbol{\mu}$ with distribution \mathbb{P} . Under additional assumptions on $\boldsymbol{\mu}$ (e.g. if it admits a density with respect to the Lebesgue measure on \mathbb{R}) then one can show that such a μ^* exists and is unique. Therefore, in the one-dimensional case, extending some of our results to random measures that are non-compactly supported is certainly not too difficult.

The multi-dimensional case (i.e. $d \geq 2$) is more involved. Indeed, the arguments that we used to prove the existence of an optimizer of the dual problem (\mathcal{P}^*) as well as those used to show the convergence of the empirical barycenter to its population counterpart strongly depend on the compactness assumption for the support of the random measure $\boldsymbol{\mu}$. Adapting these arguments to non-compactly supported measures to study the dual problem (\mathcal{P}^*) and to show the consistency of the empirical barycenter is an interesting topic for future investigations.

References

- [1] Bijan Afsari. Riemannian l^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.
- [2] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.
- [3] S. Allasonnière, Y. Amit, and A. Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(1):3–29, 2007.

- [4] Marc Arnaudon, Clément Dombry, Anthony Phan, and Le Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Process. Appl.*, 122(4):1437–1455, 2012.
- [5] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (i). *Annals of statistics*, 31(1):1–29, 2003.
- [6] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (ii). *Annals of statistics*, 33:1225–1259, 2005.
- [7] J. Bigot and B. Charlier. On the consistency of Fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, 5:1054–1089, 2011.
- [8] J. Bigot, S. Gadat, and J.M. Loubes. Statistical M-estimation and consistency in large deformable models for image warping. *Journal of Mathematical Imaging and Vision*, 34(3):270–290, 2009.
- [9] J. Bigot, J.M. Loubes, and M. Vimond. Semiparametric estimation of shifts on compact Lie groups for image registration. *Probability Theory and Related Fields*, 148:1–49, 2010.
- [10] Jérémie Bigot and Sébastien Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.
- [11] Emmanuel Boissard, Thibault Le Gouic, and Jean-Michel Loubes. Distribution’s template estimate with Wasserstein metrics. *Preprint*, 2011.
- [12] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999. Translated from the French.
- [13] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H.Poincaré, Sect. B, Prob. et Stat.*, 10:235–310, 1948.
- [14] Fabrice Gamboa, Jean-Michel Loubes, and Elie Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.
- [15] Cedric E. Ginestet. Strong and weak laws of large numbers for Frechet sample means in bounded metric spaces. *Preprint*, 2012.
- [16] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B*, 53(2):285–339, 1991.
- [17] U. Grenander. *General pattern theory - A mathematical study of regular structures*. Clarendon Press, Oxford, 1993.
- [18] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.

- [19] Steven Haker and Allen Tannenbaum. On the Monge-Kantorovich problem and image warping. In *Mathematical methods in computer vision*, volume 133 of *IMA Vol. Math. Appl.*, pages 65–85. Springer, New York, 2003.
- [20] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal on Computer Vision*, 60:225–240, 2004.
- [21] Stephan F. Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.*, 39(2):1098–1124, 2011.
- [22] A. D. Ioffe and V. M. Tihomirov. Duality of convex functions, and extremal problems. *Uspehi Mat. Nauk*, 23(6 (144)):51–116, 1968.
- [23] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math Soc.*, 16:81–121, 1984.
- [24] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wassertein barycenter and its applications to texture mixing. In *LNCS, Proc. SSVM'11*, volume 6667, pages 435–446. Springer, 2011.
- [25] R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.
- [26] B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):243–259, 2001.
- [27] Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, volume 338 of *Contemp. Math.*, pages 357–390. Amer. Math. Soc., Providence, RI, 2003.
- [28] Thomas Trigano, Uri Isserles, and Ya'acov Ritov. Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.*, 59(5):1970–1984, 2011.
- [29] A. Trounev and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59 (electronic), 2005.
- [30] A. Trounev and L. Younes. Shape spaces. In *Handbook of Mathematical Methods in Imaging*. Springer, 2011.
- [31] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.
- [32] Myriam Vimond. Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3):1885–1912, 2010.