



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. 2013. hal-00763668v3

HAL Id: hal-00763668

<https://hal.science/hal-00763668v3>

Preprint submitted on 17 Mar 2014 (v3), 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[†]

March 17, 2014

Abstract

We define a notion of barycenter for random probability measures in the Wasserstein space. We study the population barycenter in terms of existence and uniqueness. Using a duality argument, we give a precise characterization of the population barycenter for compactly supported measures, and we make a connection between averaging in the Wasserstein space and taking the expectation of optimal transport maps. Then, the problem of estimating this barycenter from n independent and identically distributed random probability measures is considered. To this end, 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 show the usefulness of averaging in the Wasserstein space for curve and image warping. In this setting, we also study the rate of convergence of the empirical barycenter to its population counterpart for some semi-parametric models of random densities.

Keywords: Wasserstein space; Empirical and population barycenters; Fréchet mean; Convergence of random variables; Optimal transport; Duality; 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

*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

Euclidean space. Consequently, to define a notion of barycenter for random probability measures, it is natural to use the notion of Fréchet mean [15] that is an extension of the usual Euclidean barycenter to non-linear spaces endowed with non-Euclidean metrics. If \mathbf{Y} denotes a random 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 (if any) 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 random variable \mathbf{Y} with 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 [25] that leads to the Procrustean means studied in [18]. Many properties of the Fréchet mean in finite dimensional Riemannian manifolds (such as consistency and uniqueness) have been investigated in [1, 4, 5, 6, 23]. 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 [31]. Recently, some properties of the Fréchet mean in bounded metric spaces have also been studied in [17]. 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 [31].

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

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

where $\mathcal{M}_+^2(\Omega)$ is the set of Radon probability measures with finite second order moment, 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 $(\mathcal{M}_+^2(\mathbb{R}^d), \mathcal{B}(\mathcal{M}_+^2(\mathbb{R}^d)))$, where $\mathcal{B}(\mathcal{M}_+^2(\mathbb{R}^d))$ 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} . A similar notion (to the one in this paper) of a population barycenter and its connection to optimal transportation with infinitely many marginals have been recently studied in [27]. Throughout the paper, we shall thus explain the differences and the similarities between the approach that we follow and the one in [27].

The empirical counterpart of μ^* is the barycenter $\bar{\boldsymbol{\mu}}_n$ defined as

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

where μ_1, \dots, μ_n are iid random measures sampled from the distribution \mathbb{P} . A detailed characterization of $\bar{\mu}_n$ 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 warrant the existence and uniqueness of the population barycenter. In the one-dimensional case, we obtain an explicit characterization that can be (informally) stated as follows: in the case $d = 1$, if μ_0 denotes some reference measure that is absolutely continuous with respect to the Lebesgue measure on \mathbb{R} , then the barycenter μ^* of a random measure $\mu \in \mathcal{M}_+^2(\mathbb{R})$ is given by the push-forward of μ_0 through the mapping $\mathbb{E}(\mathbf{T})$, namely

$$\mu^* = \mathbb{E}(\mathbf{T})\#\mu_0, \tag{1.3}$$

where \mathbf{T} is the optimal mapping to transport μ_0 onto μ (i.e. $\mu = \mathbf{T}\#\mu_0$), and \mathbb{E} denotes the usual expectation of random variables. Moreover, μ^* does not depend on the choice of μ_0 .

One of the purposes of this paper is to extend equation (1.3) to higher dimensions $d \geq 2$ for some specific probability models on μ . To this end, we propose a dual formulation of the optimisation problem (1.1) that allows a precise study of some properties of the population barycenter. These results are based on an adaptation of the arguments developed in [2] for the characterization of the empirical barycenter $\bar{\mu}_n$. 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 [36] for further details on this topic.

Another 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 for which analogs of equation (1.3) may hold.

The paper is then organised as follows. In Section 2, we introduce the general framework of the paper, and we describe a specific probability model of random measures. In Section 3, we characterize the population Barycenter in the one-dimensional case, i.e. for random measures supported on \mathbb{R} . In Section 4, we prove the existence of the population barycenter in dimension $d \geq 2$ within our framework and for random measures with a compact support. In Section 5, in the case of compactly supported measures and for $d \geq 1$, we introduce a dual formulation of the optimisation problem (1.1), and we give a characterisation of the population barycenter. The convergence of the empirical barycenter is discussed in Section 6. As an application of the methodology developed in this paper, we discuss in Section 7 the usefulness of barycenters in the Wasserstein space for curve and image warping problems. In this setting, we discuss the extension of equation (1.3) to dimension $d \geq 2$ for some semi-parametric models of random densities, and we study the rate of convergence of the empirical barycenter to its population counterpart. Finally, we give a conclusion and some perspectives in Section 8.

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

2 General framework

2.1 Some definitions and notation

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 a convex subset of \mathbb{R}^d . We denote by $\mathcal{M}(\Omega)$ the space of bounded Radon measures on Ω and by $\mathcal{M}_+^2(\Omega)$ the set of Radon probability measures with finite second order moment.

We recall that the squared 2-Wassertein distance between two probability measures $\mu, \nu \in \mathcal{M}_+^2(\Omega)$ 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. [36]. We recall that $\tilde{\gamma} \in \Pi(\mu, \nu)$ is called an optimal transport plan between μ and ν if

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

Let $T : \Omega \rightarrow \Omega$ be a measurable mapping, and let $\mu \in \mathcal{M}_+^2(\Omega)$. The push-forward measure $T\#\mu$ of μ through the map T is the measure defined by duality as

$$\int_{\Omega} f(x) d(T\#\mu)(x) = \int_{\Omega} f(T(x)) d\mu(x), \quad \text{for all continuous and bounded functions } f : \Omega \rightarrow \mathbb{R}.$$

We also recall the following well known result in optimal transport (see e.g. [36] or Proposition 3.3 in [2]):

Proposition 2.1. *Let $\mu, \nu \in \mathcal{M}_+^2(\Omega)$. 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

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 $\nabla\phi$ (the gradient of ϕ) is called the optimal mapping between μ and ν . 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 $\nabla\phi = \nabla\psi$ μ -almost everywhere. Moreover, one has that

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

2.2 A parametric class of random probability measures

We will now define the model of random measures that will be studied throughout the paper. Let Θ be a subset of \mathbb{R}^p . Let $\varphi : (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p)) \rightarrow (\mathcal{M}_+^2(\Omega), \mathcal{B}(\mathcal{M}_+^2(\Omega)))$ be a measurable mapping, where $\mathcal{B}(\mathbb{R}^p)$ is the Borel σ -algebra of \mathbb{R}^p and $\mathcal{B}(\mathcal{M}_+^2(\Omega))$ is the Borel σ -algebra generated by the topology induced by the distance d_{W_2} . Then, let us define

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

as the set of probability measures $\mu_\theta \in \mathcal{M}_+^2(\Omega)$ parametrized by the mapping φ and the set Θ . Throughout the paper, we will suppose that φ satisfies the following assumption:

Assumption 1. *For any $\theta \in \Theta$, the measure $\mu_\theta = \varphi(\theta) \in \mathcal{M}_+^2(\Omega)$ admits a density with respect to the Lebesgue measure on \mathbb{R}^d .*

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 . It follows that if $\boldsymbol{\theta} \in \mathbb{R}^p$ is a random vector with density g , then $\mu_{\boldsymbol{\theta}} = \varphi(\boldsymbol{\theta})$ is a random probability measure with distribution \mathbb{P}_g on $(\mathcal{M}_+^2(\Omega), \mathcal{B}(\mathcal{M}_+^2(\Omega)))$ that is the push-forward measure defined by

$$\mathbb{P}_g(B) = \mathbb{P}_\Theta(\varphi^{-1}(B)), \text{ for any } B \in \mathcal{B}(\mathcal{M}_+^2(\Omega)).$$

A similar class of random probability measures has been considered in [27] for the purpose of studying the existence and uniqueness of a population barycenter in the 2-Wasserstein space. However, the setting in [27] is somewhat more restrictive than the one considered in this paper, since it is assumed in [27] that the parameter set Θ is one-dimensional.

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

$$\mu^* \in \arg \min_{\nu \in \mathcal{M}_+^2(\Omega)} J(\nu), \tag{2.1}$$

where

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

Using standard arguments, the existence and the uniqueness of μ^* are not difficult to show. To this end, a key property of the functional J defined in (2.2) is the following:

Lemma 2.1. *Suppose that Assumption 1 holds. Then, the functional $J : \mathcal{M}_+^2(\Omega) \rightarrow \mathbb{R}$ is strictly convex in the sense that*

$$J(\lambda\mu + (1-\lambda)\nu) < \lambda J(\mu) + (1-\lambda)J(\nu), \text{ for any } \lambda \in]0, 1[\text{ and } \mu, \nu \in \mathcal{M}_+^2(\Omega) \text{ with } \mu \neq \nu. \tag{2.3}$$

Proof. Inequality (2.3) follows immediately from Assumption 1 and the use of Lemma 3.2.1 in [27]. \square

Hence, thanks to the strict convexity of J , it follows that, if a barycenter μ^* exists, then it is necessarily unique. The existence of μ^* is then proved in Section 3 and Section 4.

One of the main goals of the paper is then to introduce a dual formulation of the optimization problem (2.2) that will allow us to give an explicit characterization of μ^* depending on the law of μ_θ , see Section 5.

3 Barycenter for measures supported on the real line

In this section, we prove the existence of the population barycenter μ^* of random measures supported on the real line i.e. we consider the case $d = 1$ where Ω is a subset of \mathbb{R} . In this setting, the proof of the existence of μ^* follows from the well known fact that if μ and ν are measures belonging to $\mathcal{M}_+^2(\Omega)$ then

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

where F_ν^{-1} (resp. F_μ^{-1}) is the quantile function of ν (resp. μ). This explicit expression for the Wasserstein distance allows a simple characterization of the barycenter of random measures. In particular, we prove equation (1.3) which shows that (for $d = 1$) computing a barycenter in the Wasserstein space amounts to take the expectation (in the usual sense) of the optimal mapping to transport a non-random reference measure μ_0 onto μ_θ .

Theorem 3.1. *Let Ω be a subset of \mathbb{R} and μ_0 be any fixed measure in $\mathcal{M}_+^2(\Omega)$ that is absolutely continuous with respect to the Lebesgue measure. Suppose that Assumption 1 holds. Let μ_θ be a random measure as defined in Section 2.2. Let $T_\theta : \Omega \rightarrow \Omega$ be the random mapping defined by $T_\theta(x) = F_{\mu_\theta}^{-1}(F_{\mu_0}(x))$, $x \in \Omega$, where $F_{\mu_\theta}^{-1}$ is the quantile function of μ_θ and F_{μ_0} is the cumulative distribution function of μ_0 .*

Then, T_θ is the optimal mapping between μ_0 and μ_θ that is

$$\mu_\theta = T_\theta \# \mu_0.$$

Moreover, the barycenter of μ_θ exists, it is unique, and it satisfies the equation

$$\mu^* = \mathbb{E}(T_\theta) \# \mu_0. \tag{3.1}$$

Finally, the quantile function of μ^ is $F_{\mu^*}^{-1} = \mathbb{E}(F_{\mu_\theta}^{-1})$, and thus μ^* does not depend on the choice of μ_0 .*

Proof. Let $\nu \in \mathcal{M}_+^2(\Omega)$ then

$$J(\nu) = \int_{\Theta} \frac{1}{2} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta = \frac{1}{2} \int_{\Theta} \int_0^1 |F_\nu^{-1}(y) - F_{\mu_\theta}^{-1}(y)|^2 dy g(\theta) d\theta.$$

Applying Fubini's Theorem and the fact that $\mathbb{E}|X - a|^2 \geq \mathbb{E}|X - \mathbb{E}(X)|^2$ for any squared integrable real random variable X and real number a , we obtain that

$$\begin{aligned} \int_{\Theta} d_{W_2}^2(\nu, \mu_{\theta})g(\theta)d\theta &= \int_0^1 \int_{\Theta} |F_{\nu}^{-1}(y) - F_{\mu_{\theta}}^{-1}(y)|^2 g(\theta)d\theta dy = \int_0^1 \mathbb{E} \left| F_{\nu}^{-1}(y) - F_{\mu_{\theta}}^{-1}(y) \right|^2 dy \\ &\geq \int_0^1 \mathbb{E} \left| \mathbb{E} \left(F_{\mu_{\theta}}^{-1}(y) \right) - F_{\mu_{\theta}}^{-1}(y) \right|^2 dy \\ &= \int_0^1 \int_{\Theta} \left| \mathbb{E} \left(F_{\mu_{\theta}}^{-1}(y) \right) - F_{\mu_{\theta}}^{-1}(y) \right|^2 g(\theta)d\theta dy = \int_{\Theta} d_{W_2}^2(\mu^*, \mu_{\theta})g(\theta)d\theta, \end{aligned}$$

where μ^* is the measure in $\mathcal{M}_+^2(\Omega)$ with quantile function given by $F_{\mu^*}^{-1} = \mathbb{E} \left(F_{\mu_{\theta}}^{-1} \right)$. The above inequality shows that $J(\nu) \geq J(\mu^*)$ for any $\nu \in \mathcal{M}_+^2(\Omega)$. Therefore, μ^* is a barycenter of the random measure μ_{θ} , and the unicity of μ^* follows from the strict convexity of the functional J as defined in Lemma (2.1). Finally, let μ_0 be any fixed measure in $\mathcal{M}_+^2(\Omega)$ that is absolutely continuous with respect to the Lebesgue measure. Hence, one has that $F_{\mu_0} \circ F_{\mu_0}^{-1}(t) = t$ for any $t \in [0, 1]$. Therefore, equation (3.1) follows from the equalities

$$F_{\mu^*}^{-1} = \mathbb{E} \left(F_{\mu_{\theta}}^{-1} \right) = \mathbb{E} \left(F_{\mu_{\theta}}^{-1} \right) \circ F_{\mu_0} \circ F_{\mu_0}^{-1} = \mathbb{E} (T_{\theta}) \circ F_{\mu_0}^{-1},$$

which completes the proof since it is clear that $T_{\theta} = F_{\mu_{\theta}}^{-1} \circ F_{\mu_0}$ is the optimal mapping between μ_0 and μ_{θ} . \square

To illustrate Theorem 3.1, we consider a simple construction of random probability measures in the case $\Omega = \mathbb{R}$. Let $\bar{\mu} \in \mathcal{M}_+^2(\mathbb{R})$ admitting the density \bar{f} with respect to the Lebesgue measure on \mathbb{R} , and cumulative distribution function (cdf) \bar{F} . Let $\theta = (\mathbf{a}, \mathbf{b}) \in]0, +\infty[\times \mathbb{R}$ be a two dimensional random vector with density g . We denote by μ_{θ} the random probability measure admitting the density

$$f_{\theta}(x) = \frac{1}{\mathbf{a}} \bar{f} \left(\frac{x - \mathbf{b}}{\mathbf{a}} \right), \quad x \in \mathbb{R}.$$

The cdf and quantile function of μ_{θ} are thus

$$F_{\mu_{\theta}}(x) = \bar{F} \left(\frac{x - \mathbf{b}}{\mathbf{a}} \right), \quad x \in \mathbb{R}, \quad \text{and} \quad F_{\mu_{\theta}}^{-1}(y) = \mathbf{a} \bar{F}^{-1}(y) + \mathbf{b}, \quad y \in [0, 1].$$

By Theorem 3.1, it follows that the barycenter of μ_{θ} is the probability measure μ^* whose quantile function is given by

$$F_{\mu^*}^{-1}(y) = \mathbb{E}(\mathbf{a}) \bar{F}^{-1}(y) + \mathbb{E}(\mathbf{b}), \quad y \in [0, 1].$$

Therefore, μ^* admits the density

$$f^*(x) = \frac{1}{\mathbb{E}(\mathbf{a})} \bar{f} \left(\frac{x - \mathbb{E}(\mathbf{b})}{\mathbb{E}(\mathbf{a})} \right), \quad x \in \mathbb{R}$$

with respect to the Lebesgue measure on \mathbb{R} . Moreover, if μ_0 is any fixed measure in $\mathcal{M}_+^2(\Omega)$, that is absolutely continuous with respect to the Lebesgue measure, then

$$\mu^* = T_0 \# \mu_0, \quad \text{where} \quad T_0(x) = \mathbb{E}(\mathbf{a})\bar{F}^{-1}(F_{\mu_0}(x)) + \mathbb{E}(\mathbf{b}), \quad x \in \mathbb{R}.$$

The meaning of Theorem 3.1 is that, in dimension $d = 1$, the computation the barycenter of a random probability measure μ_θ (as defined in Section 2.2) can be done by simply taking the expectation (in the usual sense) of the optimal mapping $T_\theta = F_{\mu_\theta}^{-1} \circ F_{\mu_0}$ between μ_0 and μ_θ , where μ_0 is any fixed measure in $\mathcal{M}_+^2(\Omega)$. However, extending Theorem 3.1 in dimension $d \geq 2$ is not straightforward. Indeed, a key ingredient in the proof of Theorem 3.1 is the use of the well-known characterization of the Wasserstein distance in dimension $d = 1$ via the quantile functions: $d_{W_2}^2(\nu, \mu_\theta) = \int_0^1 |F_\nu^{-1}(y) - F_{\mu_\theta}^{-1}(y)|^2 dy$. However, this property which explicitly relates the Wasserstein distance $d_{W_2}(\nu, \mu_\theta)$ to the marginal distributions ν and μ_θ is not valid in dimension $d \geq 2$. Nevertheless, one of the purposes of this paper is to show that analogs of Theorem 3.1 can still be obtained in dimension $d \geq 2$.

4 Existence of the population barycenter in dimension $d \geq 2$

In this section, we study the existence of a minimizer for the optimization problem (2.1) in the case $d \geq 2$. Nevertheless, we restrict this study to the following case:

Assumption 2. *The support Ω of the measures $\nu \in \mathcal{M}_+^2(\Omega)$ is a compact set of \mathbb{R}^d , and Θ is a compact subset of \mathbb{R}^p*

A similar assumption is made in [27] for proving the existence of a population barycenter of random measures whose support is supposed to be contained in a bounded domain of \mathbb{R}^d .

Proposition 4.1. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then, the optimization problem (2.1) admits a unique minimizer.*

Proof. Let ν^n be a minimizing sequence of the optimization problem (2.1). Let us first show that the sequence $\int_\Omega |x|^2 d\nu^n(x)$ is uniformly bounded. Since ν^n is a minimizing sequence of (2.1), there exists a constant $C > 0$ such that $\int_\Theta \frac{1}{2} d_{W_2}^2(\mu_\theta, \nu_n) g(\theta) d\theta \leq C$, for all $n \geq 1$. Since g is a probability density function on Θ with respect to the Lebesgue measure, it follows that there exists $\Theta' \subset \Theta$ with $0 < \int_{\Theta'} d\theta < +\infty$ and $g(\theta) \neq 0$ for all $\theta \in \Theta'$. Since $\int_{\Theta'} \frac{1}{2} d_{W_2}^2(\mu_\theta, \nu_n) g(\theta) d\theta \leq C$, one obtains that there exists $\theta^* \in \Theta'$, such that $g(\theta^*) \neq 0$ and

$$\frac{1}{2} d_{W_2}^2(\mu_{\theta^*}, \nu_n) g(\theta^*) \leq C' \tag{4.1}$$

where $C' = \frac{C}{\int_{\Theta'} d\theta}$. Then, thanks to the Kantorovich duality formula (see e.g. [36], or Lemma 2.1 in [2]), it follows that

$$\frac{1}{2} d_{W_2}^2(\mu_{\theta^*}, \nu_n) g(\theta^*) \geq \int_\Omega |x|^2 d\nu_n(x) + \int_\Omega C_{g(\theta^*)}(x) d\mu_{\theta^*}(x),$$

where $C_{g(\theta^*)}(x) = \inf_{y \in \Omega} \left\{ \frac{g(\theta^*)}{2} |x - y|^2 - |y|^2 \right\}$. Note that by Assumption 2, Ω is a compact set, and therefore the function $x \mapsto |C_{g(\theta^*)}(x)|$ is bounded on Ω . Hence, combining the above inequality with (4.1), one finally obtains that

$$\int_{\Omega} |x|^2 d\nu_n(x) \leq C' - \int_{\Omega} C_{g(\theta^*)}(x) d\mu_{\theta^*}(x), \text{ for all } n \geq 1,$$

which shows that $\int_{\Omega} |x|^2 d\nu^n(x)$ is a uniformly bounded sequence.

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 $\mu^* \in \mathcal{M}_+^2(\Omega)$. Therefore, $\frac{1}{2}d_{W_2}^2(\mu_{\theta}, \mu^*) \leq \liminf_{n \rightarrow +\infty} \frac{1}{2}d_{W_2}^2(\mu_{\theta}, \nu^n)$, and thus, by Fatou's Lemma

$$\int_{\Theta} \frac{1}{2}d_{W_2}^2(\mu_{\theta}, \mu^*)g(\theta)d\theta \leq \int_{\Theta} \liminf_{n \rightarrow +\infty} \frac{1}{2}d_{W_2}^2(\mu_{\theta}, \nu^n)g(\theta)d\theta \leq \liminf_{n \rightarrow +\infty} \int_{\Theta} \frac{1}{2}d_{W_2}^2(\mu_{\theta}, \nu_n)g(\theta)d\theta.$$

Therefore, $J(\mu^*) = \inf_{\nu \in \mathcal{M}_+^2(\Omega)} \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu_{\theta})g(\theta)d\theta$, which proves that the optimization problem (2.1) admits a minimizer.

By the strict convexity of the functional J defined in (2.2), it follows that, under Assumption 1, the barycenter of μ_{θ} is necessarily unique. \square

5 Characterisation of the population barycenter for compactly supported measures

In this section, we give a more precise characterisation of the population barycenter beyond the proof of its existence (see Proposition 4.1). For this purpose, we shall introduce a dual formulation of problem (2.1) that is inspired by the one proposed in [2] to study the properties of empirical barycenters.

In the rest of this section, it is supposed that Assumption 2 is satisfied i.e. this study will be restricted to the case where Ω is a compact set of \mathbb{R}^d . We recall that this assumption implies that the Wasserstein space $(\mathcal{M}_+^2(\Omega), d_{W_2})$ is compact. Finally, it should be noted that this characterisation of a barycenter by a duality argument will allow us to extend the results of Theorem 3.1 to dimensions $d \geq 2$ (see Section 7 below). We may also remark that a dual formulation of problem (2.1) has not been considered in [27] for the characterisation of a population barycenter. The results in [27] are rather focussed on the connection between barycenters in the Wasserstein space and optimal transportation with infinitely many marginals.

5.1 A dual formulation of problem (P)

Let us recall the optimisation problem (2.1) as

$$(P) \quad J_{\mathcal{P}} := \inf_{\nu \in \mathcal{M}_+^2(\Omega)} J(\nu), \text{ where } J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu_{\theta})g(\theta)d\theta. \quad (5.1)$$

Then, let us introduce some definitions. Let $\delta(\Omega) = \sup_{(x,y) \in \Omega \times \Omega} |x - y|$ be the diameter of Ω . 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 also denote by $X' = \mathcal{M}(\Omega)$ the topological dual of X .

The notation $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ will denote any mapping

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

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 optimization 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 $H_{g(\theta)}$ for $\nu \in X'$ as

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

In what follows, we will show in Proposition 5.1 below that the problems (\mathcal{P}) and (\mathcal{P}^*) are dual to each other 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}^*) . Then, we show in Proposition 5.2 below that the dual problem (\mathcal{P}^*) admits an optimizer. This duality will then allow us to give a nice characterization the population barycenter via the use of a solution of the dual problem, see Theorem 5.2.

Proposition 5.1. *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 $\nu \in \mathcal{M}_+^2(\Omega)$ 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

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

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

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

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. [36], 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 $\nu \in \mathcal{M}_+^2(\Omega)$. Therefore, it follows that

$$J_{\mathcal{P}} = \inf \left\{ \int_{\Theta} H_{g(\theta)}^*(\nu) d\theta, \nu \in X' \right\} = - \left(\int_{\Theta} H_{g(\theta)}^* d\theta \right)^* (0). \quad (5.2)$$

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 [24], one has that for any $\nu \in \mathcal{M}_+^2(\Omega)$

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

Then, thanks to (5.2), 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. [14]). 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. [14]), 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 dual problem (\mathcal{P}^*) as formulated in the following proposition:

Proposition 5.2. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then, the dual problem (\mathcal{P}^*) admits an optimizer.*

Proof. 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 $\tilde{f}^{\Theta} = (\tilde{f}_{\theta})_{\theta \in \Theta} \in L^1(\Theta, X)$ 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 $h_{\theta}^n = S_{g(\theta)} \tilde{h}_{\theta}^n$, $\tilde{h}_{\theta}^n = S_{g(\theta)} h_{\theta}^n$ and $h_{\theta}^n(0) = 0$ for every $\theta \in \Theta$.

Let us denote by $L^1(\Theta)$ the set of functions $f : \Theta \rightarrow \mathbb{R}$ such that $\int_{\Theta} |f(\theta)| d\theta < +\infty$. The space $L^1(\Theta)$, endowed with the metric $d_1(f, f') = \int_{\Theta} |f(\theta) - f'(\theta)| d\theta$ for $f, f' \in L^1(\Theta)$, is complete. Now, let us consider the family \mathcal{A} of functions from Ω to $L^1(\Theta)$ defined by

$$\mathcal{A} = \left\{ x \mapsto \left(\tilde{f}_{\theta}^n(x) \right)_{\theta \in \Theta}, n \in \mathbb{N} \right\}. \quad (5.3)$$

To prove that one may extract a converging subsequence from the elements in \mathcal{A} , we will use the following result (see Theorem 8.33 and Corollary 8.34 in [13]) which is an extension of the usual Ascoli-Arzelà theorem:

Theorem 5.1. *Let (Z, d_Z) be a compact metric space and (Y, d_Y) be a complete metric space. Let $(f_n)_{n \geq 1} \subset \mathcal{C}(Z, Y)$ where $\mathcal{C}(Z, Y)$ is the set of continuous functions from Z to Y for the topology induced by the uniform distance $d_{sup}(f, f') := \sup_{z \in Z} \{d_Y(f(z), f'(z))\}$ for $f, f' : Z \rightarrow Y$. If the sequence $(f_n)_{n \geq 1}$ is equicontinuous and bounded, then it admits a subsequence which is convergent for the uniform distance.*

In Theorem 5.1 above, the boundedness of the sequence $(f_n)_{n \geq 1}$ means that there exists $f_0 \in \mathcal{C}(Z, Y)$ and a constant $C > 0$ such that

$$(f_n)_{n \geq 1} \subset \{f \in \mathcal{C}(Z, Y) : d_{sup}(f_0, f) \leq C\}. \quad (5.4)$$

We refer to Definition 6.4 and Theorem 8.32 in [13] for further details on this notion. In what follows, we will apply Theorem 5.1 with $Z = \Omega$, $d_Z = |\cdot|$, $Y = L^1(\Theta)$ and $d_Y = d_1$.

One has that for any $x, z \in \Omega$, and $\theta \in \Theta$

$$\begin{aligned} \left| \frac{g(\theta)}{2} |x - y|^2 - \frac{g(\theta)}{2} |z - y|^2 \right| &= \frac{g(\theta)}{2} \left| |x - y|^2 - |z - y|^2 \right| \\ &\leq \frac{g(\theta)}{2} \left| |x - y| - |z - y| \right| (|x - y| + |z - y|) \\ &\leq \delta(\Omega) g(\theta) |x - z|. \end{aligned}$$

Therefore, the function $x \mapsto \frac{g(\theta)}{2} |x - y|^2 - \tilde{h}_\theta^n(y)$ is K -Lipschitz on the compact set Ω with $K = \delta(\Omega) g(\theta)$. Thus, the function $x \mapsto h_\theta^n(x) = S_{g(\theta)} \tilde{h}_\theta^n(x)$ is also K -Lipschitz, since it is an infimum of K -Lipschitz functions. Using that $\int_\Theta g(\theta) d\theta = 1$, this implies that for any $x, z \in \Omega$

$$\int_\Theta |h_\theta^n(x) - h_\theta^n(z)| d\theta \leq \delta(\Omega) |x - z|,$$

which proves that the function $x \mapsto (h_\theta^n(x))_{\theta \in \Theta}$ is $\delta(\Omega)$ -Lipschitz, as a mapping from Ω to $(L^1(\Theta), d_1)$. Hence, it follows that for any $x, z \in \Omega$

$$\begin{aligned} d_1 \left(\left(\tilde{f}_\theta^n(x) \right)_{\theta \in \Theta}, \left(\tilde{f}_\theta^n(z) \right)_{\theta \in \Theta} \right) &= \int_\Theta |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^n(z)| d\theta \leq \int_\Theta |h_\theta^n(x) - h_\theta^n(z)| d\theta \\ &\quad + \left(\int_\Theta d\theta \right) \left(\int_\Theta |h_u^n(x) - h_u^n(z)| du \right) \\ &\leq \delta(\Omega) \left(1 + \int_\Theta d\theta \right) |x - z|. \end{aligned} \quad (5.5)$$

Hence, the functions $x \mapsto \left(\tilde{f}_\theta^n(x) \right)_{\theta \in \Theta}$ are \tilde{K} -Lipschitz (with $\tilde{K} = \delta(\Omega) (1 + \int_\Theta d\theta)$) from the compact set Ω to the complete space $(L^1(\Theta), d_1)$. By inequality (5.5), the set $\mathcal{A} \subset \mathcal{C}(\Omega, L^1(\Theta))$ is thus equicontinuous. Moreover, from the Lipschitz continuity of this mapping, it follows that

$$\int_\Theta |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^n(0)| d\theta \leq \tilde{K} |x| \leq \tilde{K} \delta(\Omega). \quad (5.6)$$

Then, using the fact that $h_\theta^n(0) = 0$, it follows that $\tilde{f}_\theta^n(0) = 0$ and thus by inequality (5.6) we obtain that for any $x \in \Omega$

$$\begin{aligned} \int_{\Theta} |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^1(x)| d\theta &\leq \int_{\Theta} |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^n(0)| d\theta + \int_{\Theta} |\tilde{f}_\theta^1(x) - \tilde{f}_\theta^1(0)| d\theta \\ &\leq 2\tilde{K}\delta(\Omega), \end{aligned} \quad (5.7)$$

which proves that the family \mathcal{A} is bounded in the sense of equation (5.4).

Thus, one can use the Ascoli-Arzelà's Theorem 5.1 to obtain that there exists a subsequence of functions $x \mapsto \left(\tilde{f}_\theta^{\varphi(n)}(x)\right)_{\theta \in \Theta}$ that converges uniformly to some $x \mapsto (\tilde{f}_\theta(x))_{\theta \in \Theta} \in \mathcal{C}(\Omega, L^1(\Theta))$, where $\varphi(n)$ is an increasing sequence of positive integers. It is clear that $\int_{\Theta} |\tilde{f}_\theta(x)| d\theta < +\infty$. Moreover, since, for every $x \in \Omega$, $\lim_{n \rightarrow +\infty} \int_{\Theta} |\tilde{f}_\theta(x) - \tilde{f}_\theta^{\varphi(n)}(x)| d\theta = 0$ and $\int_{\Theta} \tilde{f}_\theta^{\varphi(n)}(x) d\theta = 0$, it follows that $\int_{\Theta} \tilde{f}_\theta(x) d\theta = 0$. 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}^*) . □

5.2 Characterization of the population barycenter by duality

Let us now use the duality between problems (\mathcal{P}) and (\mathcal{P}^*) to characterize more precisely the population barycenter.

Theorem 5.2. *Suppose that Assumption 1 and Assumption 2 are satisfied. Then, the measure $\mu^* \in \mathcal{M}_+^2(\Omega)$ is the unique minimizer of problem (\mathcal{P}) if and only if*

$$\mu^* = \nabla \phi_\theta \# \mu_\theta \quad (5.8)$$

for every $\theta \in \Theta$ such that $g(\theta) > 0$, where $\phi_\theta : \Omega \rightarrow \mathbb{R}$ is the convex function defined by

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

and where $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ is a maximizer of problem (\mathcal{P}^*) .

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 which is necessarily such that $\int_{\Theta_g} d\theta \neq 0$.

Let $f^\Theta \in L^1(\Theta, X)$ be a maximizer of problem (\mathcal{P}^*) . By Proposition 4.1, Proposition 5.1 and Proposition 5.2 it follows that there exists a unique minimizer μ^* of (\mathcal{P}) such that

$$\begin{aligned} \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_\theta)g(\theta)d\theta &= \int_{\Theta} \int_{\Omega} S_{g(\theta)}f_\theta(x)d\mu_\theta(x)d\theta \\ &= \int_{\Theta} \int_{\Omega} S_{g(\theta)}f_\theta(x)d\mu_\theta(x)d\theta + \int_{\Theta} \int_{\Omega} f_\theta(x)d\mu^*(x)d\theta, \end{aligned} \quad (5.9)$$

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. [36], or Lemma 2.1 in [2]) we have that

$$\begin{aligned} \frac{1}{2}d_{W_2}^2(\mu^*, \mu_\theta)g(\theta) &= H_{g(\theta)}^*(\mu^*) \\ &= \sup_{f \in X} \left\{ \int_{\Omega} S_{g(\theta)}f(x)d\mu_\theta(x) + \int_{\Omega} f(x)d\mu^*(x) \right\} \\ &\geq \int_{\Omega} S_{g(\theta)}f_\theta(x)d\mu_\theta(x) + \int_{\Omega} f_\theta(x)d\mu^*(x). \end{aligned} \quad (5.10)$$

Therefore by combining (5.9) and (5.10), we necessarily have that

$$\frac{1}{2}d_{W_2}^2(\mu^*, \mu_\theta)g(\theta) = \int_{\Omega} S_{g(\theta)}f_\theta(x)d\mu_\theta(x) + \int_{\Omega} f_\theta(x)d\mu^*(x), \quad (5.11)$$

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 (5.11), one obtains that for every $\theta \in \Theta_g$.

$$\begin{aligned} \frac{g(\theta)}{2} \int_{\Omega \times \Omega} |x - y|^2 d\gamma_\theta(x, y) &= \frac{g(\theta)}{2} d_{W_2}^2(\mu^*, \mu_\theta) \\ &= \int_{\Omega} S_{g(\theta)}f_\theta(x)d\mu_\theta(x) + \int_{\Omega} f_\theta(y)d\mu^*(y) \\ &= \int_{\Omega \times \Omega} (S_{g(\theta)}f_\theta(x) + f_\theta(y)) d\gamma_\theta(x, y). \end{aligned} \quad (5.12)$$

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 (5.12) implies that

$$\frac{g(\theta)}{2}|x - y|^2 = S_{g(\theta)}f_\theta(x) + f_\theta(y), \quad \gamma_\theta - a.e., \quad (5.13)$$

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 (5.13) 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

$$f_\theta = S_{g(\theta)}(S_{g(\theta)}f_\theta), \quad \mu^* - a.e., \quad (5.14)$$

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

$$\int_\Theta S_{g(\theta)}(S_{g(\theta)}f_\theta)(x)d\theta = 0, \quad \mu^* - a.e. \quad (5.15)$$

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

$$\phi_\theta(x) = \frac{1}{2}|x|^2 - \frac{1}{g(\theta)}S_{g(\theta)}f_\theta(x), \quad (5.16)$$

and its conjugate ϕ_θ^* that satisfies the following equality

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

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 (5.13) and (5.14) it follows that

$$\begin{aligned} g(\theta)\langle x, y \rangle &= -S_{g(\theta)}f_\theta(x) + \frac{g(\theta)}{2}|x|^2 - f_\theta(y) + \frac{g(\theta)}{2}|y|^2 \\ &= g(\theta)\phi_\theta(x) - S_{g(\theta)}(S_{g(\theta)}f_\theta)(y) + \frac{g(\theta)}{2}|y|^2 = g(\theta)\phi_\theta(x) + g(\theta)\phi_\theta^*(y). \end{aligned} \quad (5.17)$$

By equality (5.17), 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

$$\phi_\theta = \arg \min_{\phi \in \mathcal{C}} \left\{ \int_\Omega \phi(x)d\mu_\theta(x) + \int_\Omega \phi^*(x)d\mu^*(x) \right\}, \quad (5.18)$$

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 (5.18), such that

$$\mu^* = \nabla\phi_\theta\#\mu_\theta \quad (5.19)$$

for every θ in the support Θ_g of g . Since the convex function ϕ_θ is defined by the equation (5.16), 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 (5.19), the population barycenter μ^* is necessarily unique, which completes the proof of Theorem 5.2. □

6 Convergence of the empirical barycenter

Let us now prove the convergence of the empirical barycenter for the set of measures introduced in Section 5 under Assumption 2 (i.e. that they are compactly supported). 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

$$J_n(\nu) = \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}), \quad \nu \in \mathcal{M}_+^2(\Omega), \quad (6.1)$$

and consider the optimization problem: find an empirical barycenter

$$\bar{\mu}_n \in \arg \min_{\nu \in \mathcal{M}_+^2(\Omega)} J_n(\nu), \quad \text{where } J_n(\nu) = \frac{1}{n} \sum_{j=1}^n \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}), \quad (6.2)$$

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

Lemma 6.1. *Suppose that Assumption 1 holds. Then, for any $n \geq 1$, there exists a unique minimizer $\bar{\mu}_n$ of $J_n(\cdot)$ over $\mathcal{M}_+^2(\Omega)$.*

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

Theorem 6.1. *Suppose that Assumption 1 and Assumption 2 hold. Let μ^* be the population barycenter defined by (2.1), and $\bar{\mu}_n$ be the empirical barycenter defined by (6.2). 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 [17]. For $\nu \in \mathcal{M}_+^2(\Omega)$, 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 $\mathcal{M}_+^2(\Omega)$. Then, we show that any converging subsequence of $\bar{\mu}_n$ converges a.s. to μ^* for the 2-Wasserstein distance.

Step 1. For $\nu \in \mathcal{M}_+^2(\Omega)$, let us denote by $f_\nu : \mathcal{M}_+^2(\Omega) \rightarrow \mathbb{R}$ 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

$$\mathcal{F} = \{f_\nu, \nu \in \mathcal{M}_+^2(\Omega)\}.$$

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 $\nu, \mu, \mu' \in \mathcal{M}_+^2(\Omega)$. 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 $(\mathcal{M}_+^2(\Omega), \mathcal{B}(\mathcal{M}_+^2(\Omega)))$

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

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

$$\Delta_n(\nu) = \int_{\mathcal{M}_+^2(\Omega)} f_\nu(\mu) d\mathbb{P}_g^n(\mu) - \int_{\mathcal{M}_+^2(\Omega)} f_\nu(\mu) d\mathbb{P}_g(\mu).$$

Let $f : \mathcal{M}_+^2(\Omega) \rightarrow \mathbb{R}$ 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 $\int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g^n(\mu)$ converges a.s. to $\int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g(\mu)$ 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 [29] to obtain that

$$\sup_{\nu \in \mathcal{M}_+^2(\Omega)} |\Delta_n(\nu)| = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g^n(\mu) - \int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g(\mu) \right| \rightarrow 0 \text{ as } n \rightarrow +\infty, \text{ a.s.} \quad (6.3)$$

which proves the uniform convergence of Δ_n to zero over $\mathcal{M}_+^2(\Omega)$.

Step 2. Suppose that Assumption 1 holds. By Lemma 6.1, there exists a unique sequence $(\bar{\boldsymbol{\mu}}_n)_{n \geq 1}$ of empirical barycenters defined by (6.2). Thanks to the compactness of the Wasserstein space $(\mathcal{M}_+^2(\Omega), d_{W_2})$, 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 $\bar{\boldsymbol{\mu}} \in \mathcal{M}_+^2(\Omega)$.

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

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

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 (6.3) of Δ_n to zero over $\mathcal{M}_+^2(\Omega)$, one obtains that

$$\lim_{k \rightarrow +\infty} J(\bar{\boldsymbol{\mu}}_{n_k}) = J(\mu^*). \quad (6.4)$$

Therefore, using that

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

one finally obtains by (6.3) and (6.4) that

$$\lim_{k \rightarrow +\infty} J_{n_k}(\bar{\mu}_{n_k}) = J(\mu^*). \quad (6.5)$$

Since $|J_{n_k}(\bar{\mu}) - J(\bar{\mu})| \leq \sup_{\nu \in \mathcal{M}_+^2(\Omega)} |\Delta_{n_k}(\nu)|$, it follows by equation (6.3) that

$$\lim_{k \rightarrow +\infty} J_{n_k}(\bar{\mu}) = J(\bar{\mu}) \text{ a.s.} \quad (6.6)$$

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

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

and thus by equations (6.5) and (6.6), we obtain that

$$J(\bar{\mu}) \leq \lim_{k \rightarrow +\infty} J_{n_k}(\bar{\mu}_{n_k}) = J(\mu^*) \text{ a.s.} \quad (6.7)$$

which finally proves that $\bar{\mu} = \mu^*$ a.s. since μ^* is the unique minimizer of $J(\nu)$ over $\nu \in \mathcal{M}_+^2(\Omega)$.

Hence, any converging subsequence of empirical barycenters converges a.s. to μ^* for the 2-Wasserstein distance. Since $(\mathcal{M}_+^2(\Omega), d_{W_2})$ is compact, this finally shows that $(\bar{\mu}_n)_{n \geq 1}$ is a converging sequence such that $\lim_{n \rightarrow +\infty} d_{W_2}(\bar{\mu}_n, \mu^*) = 0$ a.s. which completes the proof of Theorem 6.1. \square

7 Characterization of the empirical and population barycenters for some semi-parametric models of random densities

In this section, we propose to extend the results of Theorem 3.1 to dimension $d \geq 2$. More precisely, let $\mu_\theta \in \mathcal{M}_+^2(\Omega)$ denote some random measure (as defined in Section 2.2) with $\Omega \subset \mathbb{R}^d$, and let μ_0 be a fixed measure in $\mathcal{M}_+^2(\Omega)$ admitting a density with respect to the Lebesgue measure on \mathbb{R}^d . Then, by Proposition 2.1, there exists a unique optimal mapping $\varphi_\theta : \Omega \rightarrow \Omega$ such that

$\mu_{\boldsymbol{\theta}} = \varphi_{\boldsymbol{\theta}} \# \mu_0$. In this section, we show that, for some specific probability models described below, the barycenter of $\mu_{\boldsymbol{\theta}}$ is given by $\mu^* = \mathbb{E}(\varphi_{\boldsymbol{\theta}}) \# \mu_0$ (see Theorem 7.1 below), which means that computing a barycenter in the Wasserstein space amounts to take the expectation (in the usual sense) of the optimal mapping $\varphi_{\boldsymbol{\theta}}$ to transport μ_0 on $\mu_{\boldsymbol{\theta}}$. As shown below, the use of the dual problem (\mathcal{P}^*) is the key step to prove such a result. Moreover, we also study the convergence rate of the empirical barycenter to its population counterpart.

7.1 A connection with statistical models for curve and image warping

To define probability models where averaging in the Wasserstein space amounts to take the expectation of an optimal transport mapping, we first introduce some statistical models for which the notion of population and empirical barycenters in the 2-Wasserstein space is relevant.

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 [19, 20, 34, 35] 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

$$\mathbf{X}_i = h \circ \varphi_i^{-1}, \quad i = 1, \dots, n, \quad (7.1)$$

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 (7.1) using either rigid or non-rigid random diffeomorphisms φ_i , see e.g. [3, 7, 9, 10, 8, 16, 37] 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 (7.1).

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 $q_0 \in L^2(\Omega)$ be a probability density function, and consider the deformable model of densities

$$\mathbf{Y}_i(x) = |\det(D\varphi_i^{-1})(x)| q_0(\varphi_i^{-1}(x)) \quad x \in \Omega, \quad i = 1, \dots, n, \quad (7.2)$$

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 $\mu_1, \dots, \mu_n \in \mathcal{M}_+^2(\Omega)$ the random probability measures with densities $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, and by μ_0 the measure with density q_0 , then (7.2) can also be written as the following deformable model of measures

$$\mu_i = \varphi_i \# \mu_0, \quad i = 1, \dots, n. \quad (7.3)$$

In model (7.3), 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_0 and thus of the mean pattern q_0 . In the rest of this section, we discuss some examples of model (7.3). In particular, we show how the results of Section 5 can be used to characterise the population barycenter of random measures satisfying the deformable model (7.3).

7.2 A parametric class of diffeomorphisms

Let μ_0 be a measure on \mathbb{R}^d having a density q_0 (with respect to the Lebesgue measure dx on \mathbb{R}^d) whose support is contained in compact set $\Omega_{q_0} \subset \mathbb{R}^d$. We propose to characterise the population barycenter of a random measure μ satisfying the deformable model

$$\mu = \varphi \# \mu_0, \quad (7.4)$$

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

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

and

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

Note that the matrix A_θ is nonsingular. For any $\theta \in \mathbb{R}^p$, one has that $\varphi_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a smooth and bijective affine mapping with

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

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

$$D_\phi(\Theta) = \{\varphi_\theta, \theta \in \Theta\}. \quad (7.5)$$

Finally, let $\theta \in \mathbb{R}^p$ 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 μ_θ satisfying the deformable model

$$\mu_\theta = \varphi_\theta \# \mu_0. \quad (7.6)$$

The above equation may also be interpreted as a semi-parametric model of random densities. For any $\theta \in \Theta$ (not necessarily a random vector), we define $\mu_\theta = \varphi_\theta \# \mu_0$. Since φ_θ is a smooth diffeomorphism and μ_0 is a measure with density q whose support is included in the compact set Ω_{q_0} , it follows that μ_θ admits a density q_θ on \mathbb{R}^d given by

$$q_\theta(x) = \begin{cases} \det(A_\theta^{-1}) q_0(A_\theta^{-1}(x - b_\theta)) & \text{if } x \in \mathcal{R}(\varphi_\theta), \\ 0 & \text{if } x \notin \mathcal{R}(\varphi_\theta). \end{cases} \quad (7.7)$$

where $\mathcal{R}(\varphi_\theta) = \{\varphi_\theta(y), y \in \Omega_{q_0}\} = \{A_\theta y + b_\theta, y \in \Omega_{q_0}\}$. Before stating our main result on the population barycenter of the random measure μ_θ (7.6), 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 μ_θ takes its values in $\mathcal{M}_+^2(\Omega)$.

7.3 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 (7.6) with random diffeomorphism φ_θ taking their value in the parametric class D_ϕ defined by (7.5). Before stating the main result of this section, we define, for any $\theta \in \Theta$, the following quantities

$$\bar{A}_\theta = A_\theta \bar{A}^{-1} \quad \text{and} \quad \bar{b}_\theta = b_\theta - A_\theta \bar{A}^{-1} \bar{b}, \quad (7.8)$$

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

Theorem 7.1. *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 (7.6). Suppose that Assumption 3 holds.*

Then, the population barycenter μ^ defined by (2.1) exists and is unique. Moreover, let us define the density*

$$q^*(x) = \det(\bar{A}^{-1}) q_0(\bar{A}^{-1}(x - \bar{b})), \quad x \in \Omega, \quad (7.9)$$

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 (7.8). Then, the following statements hold:

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

$$J_{\mathcal{P}} = \inf_{\nu \in \mathcal{M}_+^2(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_\theta) g(\theta) d\theta = \frac{1}{2} \int_{\Omega} \mathbb{E}(|\bar{A}_\theta u + \bar{b}_\theta - u|^2) q^*(u) du, \quad (7.10)$$

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

$$f_\theta(x) = -\frac{g(\theta)}{2} \langle (\bar{A}_\theta - I)x, x \rangle - g(\theta) \langle \bar{b}_\theta, x \rangle, \quad x \in \Omega, \quad (7.11)$$

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

2. The population barycenter is the measure $\mu^* \in \mathcal{M}_+^2(\Omega)$ with density q^* (with respect to the Lebesgue measure on \mathbb{R}^d) given by equation (7.9), that is

$$\mu^* = \bar{\varphi} \# \mu_0 \quad \text{where} \quad \bar{\varphi}(x) = \mathbb{E}(\varphi_{\boldsymbol{\theta}}(x)) = \mathbb{E}(A_{\boldsymbol{\theta}} x + \mathbb{E}(b_{\boldsymbol{\theta}})), \quad x \in \mathbb{R}^d.$$

Theorem 7.1 shows that computing the population barycenter in the Wasserstein space of a measure from the deformable model (7.6) amounts to transport the reference measure μ_0 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 $\mathbb{E}(A_{\boldsymbol{\theta}}) = I$ and $\mathbb{E}(b_{\boldsymbol{\theta}}) = 0$), the population barycenter μ^* is equal to the template measure μ_0 . Hence, this result represents an extension to the dimension $d \geq 2$ of equation (3.1) in Theorem 3.1

Proof. Under the assumptions of Theorem 7.1, it is clear that Assumption 1 is satisfied. Therefore, by Theorem (5.2), there exists a unique population barycenter μ^* of the random measure $\mu_{\boldsymbol{\theta}}$ defined by (7.6). To prove the results stated in Theorem 7.1, we will use the characterization (5.8) of the barycenter μ^* . For this purpose, we need to find a maximizer $f^{\ominus} = (f_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta} \in L^1(\Theta, X)$ of the dual problem (\mathcal{P}^*).

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

$$\bar{q}(x) = \det(\bar{A}^{-1}) q_0(\bar{A}^{-1}(x - \bar{b})), \quad x \in \mathbb{R}^d,$$

one can re-parametrize the density $q_{\boldsymbol{\theta}}$, given by (7.7) for any $\boldsymbol{\theta} \in \Theta$, as follows

$$q_{\boldsymbol{\theta}}(x) = \det(\bar{A}_{\boldsymbol{\theta}}^{-1}) \bar{q}(\bar{A}_{\boldsymbol{\theta}}^{-1}(x - \bar{b}_{\boldsymbol{\theta}})), \quad x \in \mathbb{R}^d, \quad (7.12)$$

where

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

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

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

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

Proof of statement 1. of Theorem 7.1.

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

$$S_{g(\boldsymbol{\theta})} f_{\boldsymbol{\theta}}(x) \leq \frac{g(\boldsymbol{\theta})}{2} |x - y|^2 - f_{\boldsymbol{\theta}}(y) \quad (7.13)$$

for any $y \in \Omega$. By using equation (7.12) and inequality (7.13) 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

$$J_{\mathcal{P}^*} \leq \frac{1}{2} \int_{\Omega_{\bar{q}}} \mathbb{E} (|\bar{A}_\theta u + \bar{b}_\theta - u|^2) \bar{q}(u) du. \quad (7.14)$$

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

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

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

$$\begin{aligned}
S_{g(\theta)}f_\theta(x) &= \frac{g(\theta)}{2}|x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 + \frac{g(\theta)}{2}\langle (\bar{A}_\theta - I)(\bar{A}_\theta^{-1}(x - \bar{b}_\theta)), \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&\quad + g(\theta)\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&= \frac{g(\theta)}{2}|x|^2 - g(\theta)\langle x, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle + \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \\
&\quad + \frac{g(\theta)}{2}\langle (x - \bar{b}_\theta), \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle - \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \\
&\quad + g(\theta)\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&= \frac{g(\theta)}{2}|x|^2 - \frac{g(\theta)}{2}\langle x, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle + \frac{g(\theta)}{2}\langle \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \tag{7.15}
\end{aligned}$$

$$\begin{aligned}
&= \frac{g(\theta)}{2}|x - \bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 + \frac{g(\theta)}{2}\langle x + \bar{b}_\theta, \bar{A}_\theta^{-1}(x - \bar{b}_\theta) \rangle \\
&\quad - \frac{g(\theta)}{2}|\bar{A}_\theta^{-1}(x - \bar{b}_\theta)|^2 \tag{7.16}
\end{aligned}$$

Let us introduce the notation $J^*(f^\Theta) = \int_\Theta \int_\Omega S_{g(\theta)}f_\theta(x)d\mu_\theta(x)d\theta$. By equation (7.16) and using the re-parametrization (7.12) of q_θ combined with the change of variable $u = \bar{A}_\theta^{-1}(x - \bar{b}_\theta)$, it follows that

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

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 (7.14), we finally have that

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

which proves that f^Θ is a maximizer of the dual problem (\mathcal{P}^*), and this completes the proof of statement 1. of Theorem 7.1.

Proof of statement 2. of Theorem 7.1.

Since we have found a solution $f^\Theta = (f_\theta)_{\theta \in \Theta}$ of the dual problem (\mathcal{P}^*) , it follows from Theorem 5.2 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 (7.15), 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 7.1.

□

7.4 The case of randomly shifted densities

To illustrate Theorem 7.1, 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 (7.1) 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, 10, 8, 16, 37], 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 [32] or biology [30].

Let $q_0 : \mathbb{R}^d \rightarrow \mathbb{R}^+$ 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

$$q_\theta(x) = q_0(x - \theta), \quad x \in \mathbb{R}^d, \tag{7.17}$$

and the associated random measure $d\mu_\theta(x) = q_\theta(x)dx$. Note that equation (7.17) corresponds to the deformable model (7.6) with $\varphi_\theta(x) = x + \theta$, $x \in \mathbb{R}^d$.

Now let us suppose that θ has a continuously differentiable density g with compact support $\Theta = [-\epsilon, \epsilon]^d$ for some $\epsilon > 0$. If $\theta_1, \dots, \theta_n$ is an iid sample of random shifts with density g , then the empirical Euclidean barycenter (standard notion of averaging) of the random densities $q_{\theta_1}, \dots, q_{\theta_n}$ is the probability density given by

$$\bar{q}_n(x) = \frac{1}{n} \sum_{j=1}^n q_{\theta_j}(x). \tag{7.18}$$

By the law of large number, one has that

$$\lim_{n \rightarrow +\infty} \bar{q}_n(x) = \int_{\mathbb{R}^d} q_0(x - \theta)g(\theta)d\theta \text{ a.s. for any } x \in \mathbb{R}^d.$$

Therefore, the Euclidean barycenter \bar{q}_n converges to the convolution of the reference template q_0 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_0 .

Let us now see the benefits of using the notion of empirical barycenter in the 2-Wasserstein space to consistently estimate q_0 . It is clear that the set of shifted measures $(\mu_\theta)_{\theta \in \Theta}$ with densities $q_\theta(x) = q_0(x - \theta)$ is included in $\mathcal{M}_+^2(\Omega)$ with $\Omega = [-(A + \epsilon), (A + \epsilon)]^d$. Hence, Assumption 1 is 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 7.1, one immediately has the following result:

Corollary 7.1. *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 μ_θ be the random measure with density $q_\theta(x) = q_0(x - \theta)$ (with respect to the Lebesgue measure dx) where $q_0 : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a 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 $q_0(x - \mathbb{E}(\boldsymbol{\theta}))$, namely*

$$d\mu^*(x) = q_0(x - \mathbb{E}(\boldsymbol{\theta}))dx.$$

The primal problem (\mathcal{P}) satisfies

$$J_{\mathcal{P}} = \inf_{\nu \in \mathcal{M}_+^2(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\mu^*, \mu_\theta)g(\theta)d\theta = \frac{1}{2} \mathbb{E} (|\boldsymbol{\theta} - \mathbb{E}(\boldsymbol{\theta})|^2). \quad (7.19)$$

Moreover, the family $f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$, defined by

$$f_\theta(x) = -g(\theta)(\theta - \mathbb{E}(\boldsymbol{\theta}))x, \quad (7.20)$$

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_0 . In this setting, thanks to Theorem 6.1, 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_0 . Through this example, we can see the advantages of using the notion of barycenter in the Wasserstein space rather than the Euclidean barycenter \bar{q}_n , defined in (7.18). Indeed, replacing usual averaging by the notion of barycenter in the Wasserstein space yields to consistent estimators of a mean pattern.

7.5 Convergence rate of the empirical barycenter

In this sub-section, we show that it is possible to derive the rate of convergence of the empirical barycenter in some semi-parametric models of random measures. To this end, let us consider a sequence of iid random measures $\mu_{\theta_i} = \varphi_{\theta_i} \# \mu_0, i = 1, \dots, n$, from the deformable model (7.6), and suppose that Assumption 3 holds. Arguing as in the proof of Theorem 7.1, it is clear that the empirical barycenter $\bar{\mu}_n$ of $\mu_{\theta_1}, \dots, \mu_{\theta_n}$ exists and is unique. Moreover, it is given by

$$\bar{\mu}_n = \bar{\varphi}_n \# \mu_0 \quad \text{with} \quad \bar{\varphi}_n(x) = \frac{1}{n} \sum_{i=1}^n \varphi_{\theta_i}(x) = \left(\frac{1}{n} \sum_{i=1}^n A_{\theta_i} \right) x + \frac{1}{n} \sum_{i=1}^n b_{\theta_i}, \quad x \in \mathbb{R}^d. \quad (7.21)$$

The mapping $\bar{\varphi}$, defined above, is the expectation of the random diffeomorphism $\varphi_{\theta}(x) = A_{\theta}x + b_{\theta}$, $x \in \mathbb{R}^d$. Now, let us define the transport plan $\gamma_n = (\bar{\varphi}, \bar{\varphi}_n) \# \mu_0$ where $\bar{\varphi}(x) = \bar{A}x + \bar{b}$, $x \in \mathbb{R}^d$. Since $\mu^* = \bar{\varphi} \# \mu_0$ and $\bar{\mu}_n = \bar{\varphi}_n \# \mu_0$, the transport plan γ_n is thus a probability measure on $\Omega \times \Omega$ having μ^* and $\bar{\mu}_n$ as marginals. Hence, by definition of the squared 2-Wasserstein distance

$$d_{W_2}^2(\bar{\mu}_n, \mu^*) \leq \int_{\Omega} |\bar{\varphi}_n(x) - \bar{\varphi}(x)|^2 d\mu_0(x),$$

which implies that

$$\begin{aligned} d_{W_2}^2(\bar{\mu}_n, \mu^*) &\leq \int_{\Omega} \left| \left(\frac{1}{n} \sum_{i=1}^n A_{\theta_i} - \bar{A} \right) x + \left(\frac{1}{n} \sum_{i=1}^n b_{\theta_i} - \bar{b} \right) \right|^2 d\mu_0(x), \\ &\leq \left(2 \int_{\Omega} |x|^2 d\mu_0(x) \right) \left\| \frac{1}{n} \sum_{i=1}^n A_{\theta_i} - \bar{A} \right\|^2 + 2 \left| \frac{1}{n} \sum_{i=1}^n b_{\theta_i} - \bar{b} \right|^2, \end{aligned} \quad (7.22)$$

where $\|A\|$ denotes the standard operator norm of a matrix $A \in \mathbb{S}_d^+(\mathbb{R})$. Hence, to derive a rate of convergence of $\bar{\mu}_n$ to μ^* , one can use the concentration rate of $\frac{1}{n} \sum_{i=1}^n A_{\theta_i}$ and $\frac{1}{n} \sum_{i=1}^n b_{\theta_i}$ around their expectation \bar{A} and \bar{b} . To this end, we use the following concentration inequalities:

Theorem 7.2 (Matrix Bernstein inequality). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a sequence of independent random matrices in $\mathbb{S}_d^+(\mathbb{R})$. Suppose that $\mathbb{E}\mathbf{X}_i = 0$ for all $i = 1, \dots, n$, and that there exist two positive constants B_1 and σ_1^2 such that*

$$\|\mathbf{X}_i\| \leq B_1, \quad \text{a.s. } i = 1, \dots, n \quad \text{and} \quad \left\| \sum_{i=1}^n \mathbb{E}\mathbf{X}_i^2 \right\| \leq \sigma_1^2.$$

Then, for all $t \geq 0$,

$$\mathbb{P} \left(\left\| \sum_{i=1}^n \mathbf{X}_i \right\| \geq t \right) \leq 2d \exp \left(-\frac{t^2/2}{\sigma_1^2 + B_1 t/3} \right).$$

Proof. We refer to [33]. □

Theorem 7.3 (Vector Bernstein inequality). *Let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be a sequence of independent random vectors in \mathbb{R}^d . Suppose that $\mathbb{E}\mathbf{Y}_i = 0$ for all $i = 1, \dots, n$, and that there exist two positive constants B_2 and σ_2^2 such that*

$$|\mathbf{Y}_i| \leq B_2, \text{ a.s. } i = 1, \dots, n \quad \text{and} \quad \sum_{i=1}^n \mathbb{E}|\mathbf{Y}_i|^2 \leq \sigma_2^2.$$

Then, for all $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \mathbf{Y}_i\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2/2}{\sigma_2^2 + B_2 t/3}\right).$$

Proof. We refer to Chapter 6 in [26]. □

By Assumption 3, the mapping $\phi : \Theta \rightarrow \mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d$ is continuous. Since Θ is compact, this implies that there exist two positive constants B_1 and B_2 such that for $\phi(\theta) = (A_\theta, b_\theta)$, one has $\|A_\theta\| \leq B_1$ and $|b_\theta| \leq B_2$ for any $\theta \in \Theta$. Then, let us define $\sigma_1^2 = n\mathbb{E}\|A_\theta^2\|$, $\sigma_2^2 = n\mathbb{E}|b_\theta|^2$ and $\epsilon_0^2 = \int_\Omega |x|^2 d\mu_0(x)$. By combining inequality (7.22) with Theorem 7.3 and Theorem 7.2, one finally obtains that

$$\mathbb{P}(d_{W_2}^2(\bar{\mu}_n, \mu^*) \geq t) \leq 2d \exp\left(-\frac{nt}{8\epsilon_0^2\mathbb{E}\|A_\theta\|^2 + \frac{4}{3}B_1\epsilon_0\sqrt{t}}\right) + 2 \exp\left(-\frac{nt}{8\mathbb{E}|b_\theta|^2 + \frac{4}{3}B_2\sqrt{t}}\right) \quad (7.23)$$

for any $t \geq 0$. Hence, the rate of concentration of $\bar{\mu}_n$ to μ^* depends on the amount of variance of the reference measure μ_0 , and the amount of variability of the random mapping φ_θ . Finally, the concentration inequality (7.23) for $d_{W_2}^2(\bar{\mu}_n, \mu^*)$ can be used to prove that $\bar{\mu}_n$ converges in probability to μ^* at the rate n^{-1} for the squared 2-Wasserstein distance.

7.6 Related results in the literature on signal and image processing

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 [12, 28]. The theory of optimal transport for image warping has also been shown to be useful in various applications, see e.g. [21, 22] and references therein. Some properties of the empirical barycenter in the 2-Wasserstein space of random measures satisfying a deformable model similar to (7.3) have also been studied in [11].

Nevertheless, the results in this paper are novel in various aspects. First, we have also shown the benefits of considering the dual formulation (\mathcal{P}^*) of the (primal) problem (2.1) to characterize the population barycenter in the 2-Wasserstein space for a large class of deformable models of measures. To the best of our knowledge, the characterization of a population barycenter in deformable models throughout such duality arguments is novel. Moreover, we have studied on the consistency of the empirical barycenter for compactly supported measures, and we have derived its rate of converge in some deformable models.

8 Beyond the compactly supported case

To conclude the paper, we briefly discuss the case of a random measure $\boldsymbol{\mu} \in \mathcal{M}_+^2(\mathbb{R}^d)$ with distribution \mathbb{P} whose support is not included in a compact set Ω of \mathbb{R}^d . In the one-dimensional case i.e. $d = 1$, let us denote by $F_{\boldsymbol{\mu}}$ its cumulative distribution function, and by $F_{\boldsymbol{\mu}}^{-1}$ its generalized inverse (quantile function). Then, one can define the measure $\mu^* \in \mathcal{M}_+^2(\mathbb{R})$ such that its quantile function is $F_{\mu^*}^{-1}(y) = \mathbb{E} \left(F_{\boldsymbol{\mu}}^{-1}(y) \right)$ for all $y \in [0, 1]$. By applying arguments similar to those used in the proof of Theorem 3.1, one can easily show that μ^* is the unique population barycenter of the random measure $\boldsymbol{\mu}$ with distribution \mathbb{P} .

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] M. Agueh and G. 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] M. Arnaudon, C. Dombry, A. Phan, and L. 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 and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.
- [9] 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.

- [10] 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.
- [11] E. Boissard, T. Le Gouic, and J.-M. Loubes. Distribution’s template estimate with Wasserstein metrics. *Preprint*, 2011.
- [12] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and radon wasserstein barycenters of measures. Technical report, Preprint Hal-00881872, 2013.
- [13] M. Crainic. Inleiding topologie. *Lecture notes, available at <http://www.staff.science.uu.nl/~crain101/topologie11/>*, 2011.
- [14] I. Ekeland and R. 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.
- [15] 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.
- [16] F. Gamboa, J.-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.
- [17] C. E. Ginestet. Strong and weak laws of large numbers for Frechet sample means in bounded metric spaces. *Preprint*, 2012.
- [18] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B*, 53(2):285–339, 1991.
- [19] U. Grenander. *General pattern theory - A mathematical study of regular structures*. Clarendon Press, Oxford, 1993.
- [20] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.
- [21] S. Haker and A. 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.
- [22] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal on Computer Vision*, 60:225–240, 2004.
- [23] S. F. Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.*, 39(2):1098–1124, 2011.
- [24] A. D. Ioffe and V. M. Tihomirov. Duality of convex functions, and extremal problems. *Uspehi Mat. Nauk*, 23(6 (144)):51–116, 1968.
- [25] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math Soc.*, 16:81–121, 1984.

- [26] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.
- [27] B. Pass. Optimal transportation with infinitely many marginals. *J. Funct. Anal.*, 264(4):947–963, 2013.
- [28] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its applications to texture mixing. In *LNCS, Proc. SSVM’11*, volume 6667, pages 435–446. Springer, 2011.
- [29] R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.
- [30] B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):243–259, 2001.
- [31] K.-T. 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.
- [32] T. Trigano, U. Isserles, and Y. Ritov. Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.*, 59(5):1970–1984, 2011.
- [33] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.
- [34] A. Trounev and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59 (electronic), 2005.
- [35] A. Trounev and L. Younes. Shape spaces. In *Handbook of Mathematical Methods in Imaging*. Springer, 2011.
- [36] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.
- [37] M. Vimond. Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3):1885–1912, 2010.