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.archives-ouvertes.fr/hal-00763668v3
Submitted on 17 Mar 2014 (v3), last revised 28 Nov 2017 (v7)
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 \( Y \) denotes a random variable with distribution \( P \) taking its value in a metric space \((M, d_M)\), then a Fréchet mean (not necessarily unique) of the distribution \( P \) is a point \( m^* \in M \) that is a global minimum (if any) of the functional
\[
J(m) = \frac{1}{2} \int_M d^2_M(m, y) dP(y) \quad \text{i.e.} \quad m^* \in \arg \min_{m \in M} J(m).
\]
In this paper, a Fréchet mean of a random variable \( Y \) with distribution \( P \) will be also called a barycenter. An empirical Fréchet mean of an independent and identically distributed (iid) sample \( Y_1, \ldots, Y_n \) of distribution \( P \) is
\[
\bar{Y}_n \in \arg \min_{m \in M} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} d^2_M(m, Y_j).
\]
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 \( Y = \mu \) is a random probability measure belonging to the 2-Wasserstein space on \( \mathbb{R}^d \) with distribution \( P \). More precisely, we propose to study some properties of the barycenter \( \mu^* \) of \( \mu \) defined as the following Fréchet mean
\[
\mu^* = \arg \min_{\nu \in \mathcal{M}^2_+(\mathbb{R}^d)} \int_{\mathcal{M}^2_+(\mathbb{R}^d)} \frac{1}{2} d^2_W(\nu, \mu) dP(\mu), \tag{1.1}
\]
where \( \mathcal{M}^2_+(\Omega) \) is the set of Radon probability measures with finite second order moment, and \( d^2_W \) denotes the squared 2-Wasserstein distance between two probability measures. Note that \( P \) 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 \( \sigma \)-algebra generated by the topology induced by the distance \( d_W \). If it exists and is unique, the measure \( \mu^* \) will be referred to as the population barycenter of the random measure \( \mu \) with distribution \( P \). 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 \( \mu^* \) is the barycenter \( \bar{\mu}_n \) defined as
\[
\bar{\mu}_n = \arg \min_{\nu \in \mathcal{M}^2_+(\mathbb{R}^d)} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} d^2_W(\nu, \mu_j), \tag{1.2}
\]
where $\mu_1, \ldots, \mu_n$ are iid random measures sampled from the distribution $P$. 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 $P$ 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 $\mu_0$ denotes some reference measure that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, then the barycenter $\mu^*$ of a random measure $\mu \in \mathcal{M}_1^+(\mathbb{R})$ is given by the push-forward of $\mu_0$ through the mapping $\mathbb{E}(T)$, namely

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

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

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

We recall that the squared 2-Wasserstein 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 $\mu$ and $\nu$ as marginals, see e.g. [36]. We recall that $\hat{\gamma} \in \Pi(\mu, \nu)$ is called an optimal transport plan between $\mu$ and $\nu$ if

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

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

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

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 $\mu$ and $\nu$ if and only if the support of $\gamma$ is included in the set $\partial \phi$ that is the graph of the subdifferential of a convex and lower semi-continuous function $\phi$ 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 $\psi$, and $\mathcal{C}$ denotes the set of convex functions $\psi : \Omega \to \mathbb{R}$ that are lower semi-continuous.

If $\mu$ admits a density with respect to the Lebesgue measure on $\mathbb{R}^d$, then there exists a unique optimal transport plan $\gamma = (id, \nabla\phi) \# \mu$ where $\nabla\phi$ (the gradient of $\phi$) is called the optimal mapping between $\mu$ and $\nu$. The uniqueness of the transport plan holds in the sense that if $\nabla\phi \# \mu = \nabla\psi \# \mu$, where $\psi : \Omega \to \mathbb{R}$ is a convex function, then $\nabla\phi = \nabla\psi$ $\mu$-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 $\Theta$ be a subset of $\mathbb{R}^p$. Let $\varphi : (\mathbb{R}^p, B(\mathbb{R}^p)) \to (\mathcal{M}_2^+(\Omega), B(\mathcal{M}_2^+(\Omega)))$ be a measurable mapping, where $B(\mathbb{R}^p)$ is the Borel $\sigma$-algebra of $\mathbb{R}^p$ and $B(\mathcal{M}_2^+(\Omega))$ is the Borel $\sigma$-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 $\varphi$ and the set $\Theta$. Throughout the paper, we will suppose that $\varphi$ 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 $P_\Theta$ be a probability measure on $\Theta$ with density $g : \Theta \to \mathbb{R}^+$ with respect to the Lebesgue measure $d\theta$ on $\mathbb{R}^p$. It follows that if $\theta \in \mathbb{R}^p$ is a random vector with density $g$, then $\mu_\theta = \varphi(\theta)$ is a random probability measure with distribution $P_g$ on $(\mathcal{M}_2^+(\Omega), B(\mathcal{M}_2^+(\Omega)))$ that is the push-forward measure defined by

$$P_g(B) = P_\Theta(\varphi^{-1}(B)), \text{ for any } B \in B(\mathcal{M}_2^+(\Omega)).$$

A similar class of random probability measures has been considered in [27] for the purpose of studying the existence and uniqueness of a population bar ycenter 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 $\Theta$ is one-dimensional.

As explained in the introduction, we want to characterize the barycenter (i.e. the Fréchet mean) of the distribution $P_g$ 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

$$\mu^* \in \operatorname{arg \min}_{\nu \in \mathcal{M}_2^+(\Omega)} J(\nu),$$

where

$$J(\nu) = \int_{\mathcal{M}_2^+(\Omega)} \frac{1}{2} d_{W_2}^2(\nu, \mu) dP_g(\mu) = \int_{\Theta} \frac{1}{2} d_{W_2}^2(\nu, \mu_\theta) g(\theta) d\theta, \nu \in \mathcal{M}_2^+(\Omega).$$

(2.2)

Using standard arguments, the existence and the uniqueness of $\mu^*$ 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 $J : \mathcal{M}_2^+(\Omega) \to \mathbb{R}$ is strictly convex in the sense that

$$J(\lambda \mu + (1-\lambda)\nu) < \lambda J(\mu) + (1-\lambda)J(\nu), \text{ for any } \lambda \in [0,1[ \text{ and } \mu, \nu \in \mathcal{M}_2^+(\Omega) \text{ with } \mu \neq \nu. \quad (2.3)$$

**Proof.** Inequality (2.3) follows immediately from Assumption 1 and the use of Lemma 3.2.1 in [27].
Hence, thanks to the strict convexity of $J$, it follows that, if a barycenter $\mu^*$ exists, then it is necessarily unique. The existence of $\mu^*$ 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 $\mu^*$ depending on the law of $\mu_\theta$, see Section 5.

3 Barycenter for measures supported on the real line

In this section, we prove the existence of the population barycenter $\mu^*$ of random measures supported on the real line i.e. we consider the case $d=1$ where $\Omega$ is a subset of $\mathbb{R}$. In this setting, the proof of the existence of $\mu^*$ follows from the well known fact that if $\mu$ and $\nu$ are measures belonging to $\mathcal{M}_2^+(\Omega)$ then

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

where $F_{\nu}^{-1}$ (resp. $F_{\mu}^{-1}$) is the quantile function of $\nu$ (resp. $\mu$). 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 $\mu_0$ onto $\mu_\theta$.

**Theorem 3.1.** Let $\Omega$ be a subset of $\mathbb{R}$ and $\mu_0$ 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 $\mu_\theta$ 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 $\mu_\theta$ and $F_{\mu_0}$ is the cumulative distribution function of $\mu_0$.

Then, $T_{\theta}$ is the optimal mapping between $\mu_0$ and $\mu_\theta$ that is

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

Moreover, the barycenter of $\mu_\theta$ exists, it is unique, and it satisfies the equation

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

Finally, the quantile function of $\mu^*$ is $F_{\mu^*}^{-1} = \mathbb{E}\left(F_{\mu_\theta}^{-1}\right)$, and thus $\mu^*$ does not depend on the choice of $\mu_0$.

**Proof.** Let $\nu \in \mathcal{M}_2^+(\Omega)$ then

$$J(\nu) = \int \frac{1}{2} d^2_{W_2}(\nu, \mu_\theta)g(\theta)d\theta = \frac{1}{2} \int \int_0^1 \|F_{\nu}^{-1}(y) - F_{\mu_\theta}^{-1}(y)\|^2 dy g(\theta)d\theta.$$
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
\[
\int_{\Theta} d_{W_2}(\nu, \mu_\theta) g(\theta) d\theta = \int_0^1 \int_{\Theta} |F_{\nu_1}^{-1}(y) - F_{\nu_0}^{-1}(y)|^2 g(\theta) dy d\theta = \int_0^1 \mathbb{E} |F_{\nu_1}^{-1}(y) - F_{\nu_0}^{-1}(y)|^2 dy \\
\geq \int_0^1 \mathbb{E} \mathbb{E} \left( F_{\nu_1}^{-1}(y) \right) - F_{\nu_0}^{-1}(y) \right|^2 dy \\
= \int_0^1 \int_{\Theta} \mathbb{E} \left( F_{\nu_1}^{-1}(y) \right) - F_{\nu_0}^{-1}(y) \right|^2 g(\theta) dy d\theta = \int_{\Theta} d_{W_2}(\mu^*, \mu_\theta) g(\theta) d\theta,
\]
where \( \mu^* \) is the measure in \( \mathcal{M}_+^2(\Omega) \) with quantile function given by \( F_{\mu_0}^{-1} = \mathbb{E} \left( F_{\nu_0}^{-1} \right) \). The above inequality shows that \( J(\nu) \geq J(\mu^*) \) for any \( \nu \in \mathcal{M}_+^2(\Omega) \). Therefore, \( \mu^* \) is a barycenter of the random measure \( \mu_\theta \), and the unicity of \( \mu^* \) follows from the strict convexity of the functional \( J \) as defined in Lemma (2.1). Finally, let \( \mu_0 \) be any fixed measure in \( \mathcal{M}_+^2(\Omega) \) that is absolutely continuous with respect to the Lebesgue measure. Hence, one has that \( F_{\mu_0}^{-1} \circ F_{\mu_0}^{-1}(t) = t \) for any \( t \in [0, 1] \). Therefore, equation (3.1) follows from the equalities
\[
F_{\mu_0}^{-1} = \mathbb{E} \left( F_{\nu_0}^{-1} \right) \circ F_{\nu_0}^{-1} = \mathbb{E} (T_\theta) \circ F_{\mu_0}^{-1},
\]
which completes the proof since it is clear that \( T_\theta = F_{\nu_0}^{-1} \circ F_{\mu_0} \) is the optimal mapping between \( \mu_0 \) and \( \mu_\theta \).

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 \( \theta = (a, b) \in [0, +\infty) \times \mathbb{R} \) be a two dimensional random vector with density \( g \). We denote by \( \mu_\theta \) the random probability measure admitting the density
\[
f_\theta(x) = \frac{1}{a} \bar{f} \left( \frac{x - b}{a} \right), \quad x \in \mathbb{R}.
\]
The cdf and quantile function of \( \mu_\theta \) are thus
\[
F_{\nu_\theta}(x) = \bar{F} \left( \frac{x - b}{a} \right), \quad x \in \mathbb{R}, \quad \text{and} \quad F_{\nu_\theta}^{-1}(y) = a\bar{F}^{-1}(y) + b, \quad y \in [0, 1].
\]
By Theorem 3.1, it follows that the barycenter of \( \mu_\theta \) is the probability measure \( \mu^* \) whose quantile function is given by
\[
F_{\mu^*}^{-1}(y) = \mathbb{E}(a)\bar{F}^{-1}(y) + \mathbb{E}(b), \quad y \in [0, 1].
\]
Therefore, \( \mu^* \) admits the density
\[
f^*(x) = \frac{1}{\mathbb{E}(a)} \bar{f} \left( \frac{x - \mathbb{E}(b)}{\mathbb{E}(a)} \right), \quad x \in \mathbb{R}
\]

7
with respect to the Lebesgue measure on \( \mathbb{R} \). Moreover, if \( \mu_0 \) is any fixed measure in \( \mathcal{M}^2_+(\Omega) \), that is absolutely continuous with respect to the Lebesgue measure, then
\[
\mu^* = T_0 \# \mu_0, \quad \text{where} \quad T_0(x) = \mathbb{E}(a) \tilde{F}^{-1}(F_{\mu_0}(x)) + \mathbb{E}(b), \ x \in \mathbb{R}.
\]

The meaning of Theorem 3.1 is that, in dimension \( d = 1 \), the computation the barycenter of a random probability measure \( \mu_\theta \) (as defined in Section 2.2) can be done by simply taking the expectation (in the usual sense) of the optimal mapping \( T_\theta = F^{-1}_\mu \circ F_{\mu_0} \) between \( \mu_0 \) and \( \mu_\theta \), where \( \mu_0 \) 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: \( d^2_{W_2}(\nu, \mu_\theta) = \int_0^1 |F_{\nu}^{-1}(y) - F_{\mu_\theta}^{-1}(y)|^2 \, dy \). However, this property which explicitly relates the Wasserstein distance \( d_{W_2}(\nu, \mu_\theta) \) to the marginal distributions \( \nu \) and \( \mu_\theta \) 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 \( \Omega \) of the measures \( \nu \in \mathcal{M}^2_+(\Omega) \) is a compact set of \( \mathbb{R}^d \), and \( \Theta \) 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 \( \nu^n \) 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 \( \nu^n \) is a minimizing sequence of (2.1), there exists a constant \( C > 0 \) such that \( \int_\Theta \frac{1}{2} d^2_{W_2}(\mu_\theta, \nu_n) g(\theta) d\theta \leq C \), for all \( n \geq 1 \). Since \( g \) is a probability density function on \( \Theta \) 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 \( \int_{\Theta'} \frac{1}{2} d^2_{W_2}(\mu_\theta, \nu_n) g(\theta) d\theta \leq C \), one obtains that there exists \( \theta^* \in \Theta' \), such that \( g(\theta^*) \neq 0 \) and
\[
\frac{1}{2} d^2_{W_2}(\mu_{\theta^*}, \nu_n) g(\theta^*) \leq C'
\]
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
\[
\frac{1}{2} d^2_{W_2}(\mu_{\theta^*}, \nu_n) g(\theta^*) \geq \int_{\Omega} |x|^2 d\nu_n(x) + \int_{\Omega} C g(\theta^*)(x) d\mu_{\theta^*}(x),
\]

8
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, \( \Omega \) is a compact set, and therefore the function \( x \mapsto |C_g(\theta^*)(x)| \) is bounded on \( \Omega \). 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 \( \nu_n \) 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, \( \frac{1}{2}d_{W_2}(\mu, \mu^*) \leq \liminf_{n \to +\infty} \frac{1}{2}d_{W_2}(\mu, \nu^n), \) and thus, by Fatou’s Lemma
\[
\int_{\Theta} \frac{1}{2}d_{W_2}^2(\mu, \mu^*)g(\theta) d\theta \leq \int_{\Theta} \liminf_{n \to +\infty} \frac{1}{2}d_{W_2}^2(\mu, \nu^n)g(\theta) d\theta \leq \liminf_{n \to +\infty} \int_{\Theta} \frac{1}{2}d_{W_2}^2(\mu, \nu_n)g(\theta) d\theta.
\]
Therefore, \( J(\mu^*) = \inf_{\nu \in \mathcal{M}_2^+(\Omega)} \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu)g(\theta) d\theta, \) 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 \( \mu_\theta \) 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 \( \Omega \) 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 (\( \mathcal{P} \))

Let us recall the optimisation problem (2.1) as
\[
(\mathcal{P}) \quad J_P := \inf_{\nu \in \mathcal{M}_2^+(\Omega)} J(\nu), \text{ where } J(\nu) = \frac{1}{2} \int_{\Theta} d_{W_2}^2(\nu, \mu_\theta)g(\theta) d\theta.
\]
Then, let us introduce some definitions. Let \( \delta(\Omega) = \sup_{(x,y) \in \Omega \times \Omega} |x - y| \) be the diameter of \( \Omega \). Let \( X = C(\Omega, \mathbb{R}) \) be the space of continuous functions \( f : \Omega \to \mathbb{R} \) equipped with the supremum norm
\[
\|f\|_X = \sup_{x \in \Omega} |f(x)| .
\]
We also denote by \( X' = M(\Omega) \) the topological dual of \( X \).

The notation \( f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X) \) will denote any mapping \( \Theta \to X \) 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
\[
(P^*) \quad J_{P^*} := \sup \left\{ \int_{\Theta} \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) d\theta; \ f^\Theta \in L^1(\Theta, X) \text{ such that } \int_{\Theta} f_\theta(x) d\theta = 0, \forall x \in \Omega \right\},
\]
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
\[
H_{g(\theta)}(f) := -\int_{\Theta} S_{g(\theta)} f(x) d\mu_\theta(x),
\]
and the Legendre-Fenchel transform of \( H_{g(\theta)} \) for \( \nu \in X' \) as
\[
H^*_{g(\theta)}(\nu) := \sup_{f \in X} \left\{ \int_{\Omega} f(x) d\nu(x) - H_{g(\theta)}(f) \right\} = \sup_{f \in X} \left\{ \int_{\Omega} f(x) d\nu(x) + \int_{\Theta} S_{g(\theta)} f(x) d\mu_\theta(x) \right\}.
\]

In what follows, we will show in Proposition 5.1 below that the problems \((P)\) and \((P^*)\) are dual to each other in the sense that the minimal value \( J_P \) in problem \((P)\) is equal to the supremum \( J_{P^*} \) in problem \((P^*)\). Then, we show in Proposition 5.2 below that the dual problem \((P^*)\) 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,
\[
J_P = J_{P^*}.
\]

**Proof.** 1. Let us first prove that \( J_P \geq J_{P^*} \).

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 \( \mu_\theta \) and \( \nu \). By integrating the above inequality with respect to \( \gamma_\theta \) we obtain

\[
\int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(y) d\nu(y) \leq \int_{\Omega \times \Omega} \frac{g(\theta)}{2} |x - y|^2 d\gamma_\theta(x,y) = \frac{g(\theta)}{2} \mathcal{W}_2^2(\mu_\theta, \nu).
\]

Integrating now with respect to \( \theta \) and using Fubini’s Theorem we get

\[
\int_{\Theta} \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) d\theta \leq \int_{\Theta} \frac{g(\theta)}{2} \mathcal{W}_2^2(\mu_\theta, \nu) d\theta.
\]

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 \( H_{g(\theta)}^*(\nu) = \frac{1}{2} \mathcal{W}_2^2(\mu_\theta, \nu) g(\theta) \) for any \( \nu \in \mathcal{M}_2^+(\Omega) \). Therefore, it follows that

\[
J_\mathcal{P} = \inf \left\{ \int_{\Theta} H_{g(\theta)}^*(\nu) d\theta \mid \nu \in X \right\} = - \left( \int_{\Theta} H_{g(\theta)}^* d\theta \right)^*(0). \tag{5.2}
\]

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

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

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

\[
H_{g(\theta)}(f_\theta) = \int_{\Omega} \sup_{\gamma_\theta \in \Pi} \left\{ f_\theta(y) - \frac{g(\theta)}{2} |x - y|^2 \right\} d\mu_\theta(x)
\geq f_\theta(0) - \frac{g(\theta)}{2} \int_{\Omega} |x|^2 d\mu_\theta(x),
\]

which implies that

\[
H(f) \geq f(0) - \frac{g(\theta)}{2} \int_{\Omega} |x|^2 d\mu_\theta(x) d\theta > -\infty, \quad \forall f \in X.
\]
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
\[
H(f) \leq \int_{\Theta} \int_{\Omega} H_{g(\theta)}(f(\cdot)g(\theta))d\theta d\mu_\theta \leq \int_{\Theta} \int_{\Omega} \sup_{y \in \Omega} \left\{ \frac{g(\theta)}{4} - \frac{g(\theta)}{2} |x - y|^2 \right\} d\mu_\theta d\theta \leq \int_{\Theta} \int_{\Omega} \frac{g(\theta)}{4} d\mu_\theta d\theta = \frac{1}{4}.
\]
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.

Let us now prove the existence of an optimizer for the dual problem ($P^*$) as formulated in the following proposition:

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

**Proof.** Let $f^{\Theta} \in L^1(\Theta, X)$ such that $\int_{\Theta} f_\theta d\theta = 0$ and define $h_\theta(x) = S_{g(\theta)} \circ S_{y(\theta)} f_\theta(x)$ 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 $\theta \mapsto h_\theta \in L^1(\Theta, X)$. Define $f_\tilde{\theta} = h_\theta - \int_{\Theta} h_n d\mu$ for every $\theta \in \Theta$. Since $\int_{\Theta} h_n d\mu \geq \int_{\Theta} f_\theta d\mu = 0$, 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 $S_{g(\theta)} h_\theta = S_{g(\theta)} f_\theta$ 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 ($P^*$) can be restricted to the $	ilde{f}^{\Theta} = (\tilde{f}_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ satisfying $f_\theta = h_\theta - \int_{\Theta} h_n d\mu$ 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 ($P^*$) is invariant when one adds to the $f_\theta$’s constants $c_\theta$ that integrate to zero namely $\int_{\Theta} c_\theta d\theta = 0$.

Now, let $\tilde{f}^{\Theta,n} = L^1(\Theta, X)$ be a maximizing sequence for problem ($P^*$) that can thus be chosen such that $\tilde{f}^{\Theta,n}_\theta = h^{\Theta,n}_\theta - \int_{\Theta} h_n d\mu$ with $h^{\Theta,n}_\theta = S_{g(\theta)} h^{\Theta,n}_\theta$, $h^{\Theta,n}_\theta = S_{g(\theta)} h^{\Theta,n}_\theta$ and $h^{\Theta,n}(0) = 0$ for every $\theta \in \Theta$.

Let us denote by $L^1(\Theta)$ the set of functions $f : \Theta \to \mathbb{R}$ 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 $\Omega$ to $L^1(\Theta)$ defined by
\[
\mathcal{A} = \left\{ x \mapsto (\tilde{f}^{\Theta,n}_\theta(x))_{\theta \in \Theta}, n \in \mathbb{N} \right\}. \tag{5.3}
\]
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-Arzela theorem:
Theorem 5.1. Let $(Z,d_Z)$ be a compact metric space and $(Y,d_Y)$ be a complete metric space. Let 
$(f_n)_{n \geq 1} \subset C(Z,Y)$ where $C(Z,Y)$ 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 $f,f' : Z \to Y$. 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 $f_0 \in C(Z,Y)$ and a constant $C > 0$ such that

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

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

Therefore, the function $x \mapsto \frac{\delta(\theta)|x-y|^2 - \hat{h}^\alpha(y)}{2}$ is $K$-Lipschitz on the compact set $\Omega$ with

$K = \delta(\Omega) g(\theta)$. Thus, the function $x \mapsto \hat{h}^\alpha(x) = S_{g(\theta)} \hat{h}^\alpha(x)$ 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 $x,z \in \Omega$

$$\int_\Theta |\hat{h}^\alpha(x) - \hat{h}^\alpha(z)|d\theta \leq \delta(\Omega) |x-z|,$$

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

$$d_1 \left( \left( \tilde{f}^\alpha_\theta(x) \right)_{\theta \in \Theta}, \left( \tilde{f}^\alpha_\theta(z) \right)_{\theta \in \Theta} \right) = \int_\Theta |\tilde{f}^\alpha_\theta(x) - \tilde{f}^\alpha_\theta(z)|d\theta \leq \int_\Theta |\hat{h}^\alpha(x) - \hat{h}^\alpha(z)|d\theta + \left( \int_\Theta d\theta \right) \left( \int_\Theta |\hat{h}^\alpha_n(x) - \hat{h}^\alpha_n(z)|d\theta \right) \leq \delta(\Omega) \left( 1 + \int_\Theta d\theta \right) |x-z|. \hfill (5.5)$$

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

$$\int_\Theta |\tilde{f}^\alpha_\theta(x) - \tilde{f}^\alpha_\theta(0)|d\theta \leq K|x| \leq K \delta(\Omega). \hfill (5.6)$$
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 \( x \in \Omega \)
\[
\int_\Theta |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^1(x)|d\theta \leq \int_\Theta |\tilde{f}_\theta^n(x) - \tilde{f}_\theta^0(0)|d\theta + \int_\Theta |\tilde{f}_\theta^1(x) - \tilde{f}_\theta^0(0)|d\theta 
\leq 2\tilde{K}\delta(\Omega),
\]
which proves that the family \( A \) 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 \( x \mapsto (\tilde{f}_\theta(x))_{\theta \in \Theta} \in C(\Omega, L^1(\Theta)) \), 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 \( x \in \Omega \), \( \lim_{n \to +\infty} \int_\Theta |\tilde{f}_\theta(x) - \tilde{f}_\theta^{\varphi(n)}(x)|d\theta = 0 \) 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
\[
\limsup_n S_{g(\theta)}\tilde{f}_\theta^{\varphi(n)}(x) \leq \inf_{y \in \mathbb{R}^d} \left\{ \limsup_n \left( \frac{g(\theta)}{2} |x - y|^2 - \tilde{f}_\theta^{\varphi(n)}(y) \right) \right\} 
\leq \inf_{y \in \mathbb{R}^d} \left\{ \frac{g(\theta)}{2} |x - y|^2 - \tilde{f}_\theta(y) \right\} = S_{g(\theta)}\tilde{f}_\theta(x).
\]

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 \( d\mu(\theta) d\theta \), Fatou’s Lemma implies that
\[
\limsup_n \int_{\mathbb{R}^d} \int_\Theta S_{g(\theta)}\tilde{f}_\theta^{\varphi(n)}(x)d\mu(\theta)(x)d\theta \leq \int_{\mathbb{R}^d} \int_\Theta \limsup_n S_{g(\theta)}\tilde{f}_\theta^{\varphi(n)}(x)d\mu(\theta)(x)d\theta 
\leq \int_{\mathbb{R}^d} \int_\Theta S_{g(\theta)}\tilde{f}_\theta(x)d\mu(\theta)(x)d\theta,
\]
which shows that
\[
J_{P^*} = \int_{\mathbb{R}^d} \int_\Theta S_{g(\theta)}\tilde{f}_\theta(x)d\mu(\theta)(x)d\theta,
\]
and thus that \( \tilde{f}^\Theta \) is a maximizer of problem \((P^*)\).

\[\square\]

5.2 Characterization of the population barycenter by duality

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

**Theorem 5.2.** Suppose that Assumption 1 and Assumption 2 are satisfied. Then, the measure \( \mu^* \in M_2^+(\Omega) \) is the unique minimizer of problem \((P)\) if and only if
\[
\mu^* = \nabla \phi_\theta \# \mu_\theta
\]
for every \( \theta \in \Theta \) such that \( g(\theta) > 0 \), where \( \phi_\theta : \Omega \to \mathbb{R} \) 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 \((P^*)\).

**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 \((P^*)\). By Proposition 4.1, Proposition 5.1 and Proposition 5.2 it follows that there exists a unique minimizer \( \mu^* \) of \((P)\) such that

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

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

\[
\frac{1}{2} d_{W_2}(\mu^*, \mu_\theta) g(\theta) = H_{g(\theta)}(\mu^*)
\]

\[
= \sup_{f \in \mathcal{X}} \left\{ \int_{\Omega} S_{g(\theta)} f(x) d\mu_\theta(x) + \int_{\Omega} f(x) d\mu^*(x) \right\}
\]

\[
\geq \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(x) d\mu^*(x).
\]

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

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

for every \( \theta \in \Theta_g \).

Now, let \( \gamma_\theta \in \Pi(\mu_\theta, \mu^*) \) be an optimal transport plan between \( \mu_\theta \) and \( \mu^* \). By definition of \( \gamma_\theta \) and by (5.11), one obtains that for every \( \theta \in \Theta_g \),

\[
\frac{g(\theta)}{2} \int_{\Omega \times \Omega} |x - y|^2 d\gamma_\theta(x, y) = \frac{g(\theta)}{2} d_{W_2}(\mu^*, \mu_\theta)
\]

\[
= \int_{\Omega} S_{g(\theta)} f_\theta(x) d\mu_\theta(x) + \int_{\Omega} f_\theta(y) d\mu^*(y)
\]

\[
= \int_{\Omega \times \Omega} (S_{g(\theta)} f_\theta(x) + f_\theta(y)) d\gamma_\theta(x, y).
\]

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

\[
\frac{g(\theta)}{2} |x - y|^2 = S_{g(\theta)} f_\theta(x) + f_\theta(y), \; \gamma_\theta - \text{a.e.},
\]

(5.13)
where the notation $\gamma_\theta \setminus \text{a.e.}$ 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
\[
 f_\theta(y) = \frac{g(\theta)}{2} |x - y|^2 - S_{g(\theta)} f_\theta(x) \geq S_{g(\theta)} (S_{g(\theta)} f_\theta)(y), \quad \gamma_\theta - \text{a.e.}
\]
and thus
\[
 f_\theta = S_{g(\theta)} (S_{g(\theta)} f_\theta), \quad \mu^* - \text{a.e.},
\]
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
\[
 \int_\Theta S_{g(\theta)} (S_{g(\theta)} f_\theta)(x) d\theta = 0, \quad \mu^* - \text{a.e.}
\]
For every $\theta \in \Theta_g$, introduce the convex function $\phi_\theta$ defined by
\[
 \phi_\theta(x) = \frac{1}{2} |x|^2 - \frac{1}{g(\theta)} S_{g(\theta)} f_\theta(x),
\]
and its conjugate $\phi_\theta^*$ 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 $\gamma_\theta$. By (5.13) and (5.14) it follows that
\[
 g(\theta) \langle x, y \rangle = -S_{g(\theta)} f_\theta(x) + \frac{g(\theta)}{2} |x|^2 - f_\theta(y) + \frac{g(\theta)}{2} |y|^2
 = g(\theta) \phi_\theta(x) - S_{g(\theta)} (S_{g(\theta)} f_\theta)(y) + \frac{g(\theta)}{2} |y|^2 = g(\theta) \phi_\theta(x) + g(\theta) \phi_\theta^*(y).
\]
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 $\gamma_\theta$ is included in $\partial \phi_\theta$. Moreover, one can check that if $\theta \in \Theta_g$, then $\phi_\theta$ is the solution of
\[
 \phi_\theta = \arg \min \left\{ \int_\Omega \phi(x) d\mu_\theta(x) + \int_\Omega \phi^*(x) d\mu^*(x) \right\},
\]
where $C$ denotes the set of convex functions $\phi : \Omega \to \mathbb{R}$ that are lower semi-continuous.

Thanks to Assumption 1, the measure $\mu_\theta$ 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 $\gamma_\theta$ between $\mu_\theta$ and $\mu^*$ is included in $\partial \phi_\theta$. Hence, by Proposition 2.1, it follows that there exists a unique convex function $\phi_\theta : \Omega \to \mathbb{R}$, solution of the optimisation problem (5.18), such that
\[
 \mu^* = \nabla \phi_\theta \# \mu_\theta
\]
for every \( \theta \) in the support \( \Theta_g \) of \( g \). Since the convex function \( \phi_\theta \) is defined by the equation (5.16), it is clear that \( \phi_\theta \) does not depend on \( \mu^* \) but only on \( f_\theta \) and \( g(\theta) \) for \( \theta \in \Theta_g \). Therefore, by equation (5.19), the population barycenter \( \mu^* \) 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 \( \theta_1, \ldots, \theta_n \) be iid random variables in \( \mathbb{R}^p \) with distribution \( \mathbb{P}_\Theta \). Then, let us define the functional
\[
J_n(\nu) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}), \quad \nu \in \mathcal{M}_2^+(\Omega),
\]
and consider the optimization problem: find an empirical barycenter
\[
\bar{\mu}_n \in \arg\min_{\nu \in \mathcal{M}_2^+(\Omega)} J_n(\nu), \text{ where } J_n(\nu) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} d_{W_2}^2(\nu, \mu_{\theta_j}).
\]

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 \( \mu^* \) be the population barycenter defined by (2.1), and \( \bar{\mu}_n \) be the empirical barycenter defined by (6.2). Then,
\[
\lim_{n \to +\infty} d_{W_2}(\bar{\mu}_n, \mu^*) = 0 \text{ almost surely (a.s.)}
\]

**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 \( \Delta_n \) over \( \mathcal{M}_2^+(\Omega) \). Then, we show that any converging subsequence of \( \bar{\mu}_n \) converges a.s. to \( \mu^* \) for the 2-Wasserstein distance.

Step 1. For \( \nu \in \mathcal{M}_2^+(\Omega) \), let us denote by \( f_\nu : \mathcal{M}_2^+(\Omega) \to \mathbb{R} \) 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 $\Omega$ 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
\[
|f_\nu(\mu) - f_\nu(\mu')| = \frac{1}{2}|d_{W_2}(\nu, \mu) - d_{W_2}(\nu, \mu')| \leq \delta(\Omega)|d_{W_2}(\nu, \mu) - d_{W_2}(\nu, \mu')|
\]
\[
\leq \delta(\Omega)d_{W_2}(\mu, \mu').
\]
The above inequality proves that $\mathcal{F}$ is an equicontinuous family of functions. Now, let $\theta_1, \ldots, \theta_n$ be iid random vectors in $\mathbb{R}^p$ with density $g$, and let us define the random empirical measure on $(\mathcal{M}_+^2(\Omega), \mathcal{B}(\mathcal{M}_+^2(\Omega)))$
\[
\mathbb{P}_g^n = \frac{1}{n} \sum_{i=1}^n \delta_{\theta_i},
\]
where $\delta_\mu$ 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 $f : \mathcal{M}_+^2(\Omega) \to \mathbb{R}$ be a real-valued function that is continuous (for the topology induced by $d_{W_2}$) and bounded. Thanks to the measurability of the mapping $\phi$, 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 $n \to +\infty$, 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
\[
\sup_{\nu \in \mathcal{M}_+^2(\Omega)} |\Delta_n(\nu)| = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g^n(\mu) - \int_{\mathcal{M}_+^2(\Omega)} f(\mu) d\mathbb{P}_g(\mu) \right| \to 0 \text{ as } n \to +\infty, \text{ a.s. (6.3)}
\]
which proves the uniform convergence of $\Delta_n$ to zero over $\mathcal{M}_+^2(\Omega)$.

Step 2. Suppose that Assumption 1 holds. By Lemma 6.1, there exists a unique sequence $(\bar{\mu}_n)_{n \geq 1}$ 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 $(\bar{\mu}_{n_k})_{k \geq 1}$ such that $\lim_{k \to +\infty} d_{W_2}(\bar{\mu}_{n_k}, \bar{\mu}) = 0$ for some measure $\bar{\mu} \in \mathcal{M}_+^2(\Omega)$.

Let us now prove that $\bar{\mu} = \mu^*$. To this end, let us first note that by the definition of $\bar{\mu}_{n_k}$ and $\mu^*$ as the unique minimizer of $J_{n_k}(\cdot)$ and $J(\cdot)$ respectively, it follows that
\[
|J(\bar{\mu}_{n_k}) - J(\mu^*)| = J(\bar{\mu}_{n_k}) - J_{n_k}(\bar{\mu}_{n_k}) + J_{n_k}(\bar{\mu}_{n_k}) - J_{n_k}(\mu^*) + J_{n_k}(\mu^*) - J(\mu^*) \leq 2 \sup_{\nu \in \mathcal{M}_+^2(\Omega)} |\Delta_{n_k}(\nu)|,
\]
where we have used the fact that $J_{n_k}(\bar{\mu}_{n_k}) - J_{n_k}(\mu^*) \leq 0$. Therefore, thanks to the uniform convergence (6.3) of $\Delta_n$ to zero over $\mathcal{M}_+^2(\Omega)$, one obtains that
\[
\lim_{k \to +\infty} J(\bar{\mu}_{n_k}) = J(\mu^*). \tag{6.4}
\]
Therefore, using that
\[
|J_{n_k}(\bar{\mu}_{n_k}) - J(\mu^*)| \leq |J_{n_k}(\bar{\mu}_{n_k}) - J(\bar{\mu}_{n_k})| + |J(\bar{\mu}_{n_k}) - J(\mu^*)| \\
\leq \sup_{\nu \in \mathcal{M}_2^2(\Omega)} |\Delta_{n_k}(\nu)| + |J(\bar{\mu}_{n_k}) - J(\mu^*)|
\]
one finally obtains by (6.3) and (6.4) that
\[
\lim_{k \to +\infty} J_{n_k}(\bar{\mu}_{n_k}) = J(\mu^*). 
\]
(6.5)

Since \(|J_{n_k}(\bar{\mu}) - J(\bar{\mu})| \leq \sup_{\nu \in \mathcal{M}_2^2(\Omega)} |\Delta_{n_k}(\nu)|\), it follows by equation (6.3) that
\[
\lim_{k \to +\infty} J_{n_k}(\bar{\mu}) = J(\bar{\mu}) \text{ a.s.} 
\]
(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 \(k \geq k_\epsilon\)
\[
(J_{n_k}(\bar{\mu}))^{1/2} = \left( \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\mu}, \mu_{\theta_j}) \right)^{1/2} \\
\leq \left( \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\mu}_{n_k}, \mu_{\theta_j}) \right)^{1/2} + \left( \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\mu}, \bar{\mu}_{n_k}) \right)^{1/2} \\
\leq \left( \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} d_{W_2}^2(\bar{\mu}_{n_k}, \mu_{\theta_j}) \right)^{1/2} + \frac{\epsilon}{\sqrt{2}},
\]
and thus by equations (6.5) and (6.6), we obtain that
\[
J(\bar{\mu}) \leq \lim_{k \to +\infty} J_{n_k}(\bar{\mu}_{n_k}) = J(\mu^*) \text{ a.s.} 
\]
(6.7)

which finally proves that \(\bar{\mu} = \mu^*\) a.s. since \(\mu^*\) is the unique minimizer of \(J(\nu)\) over \(\nu \in \mathcal{M}_2^2(\Omega)\).

Hence, any converging subsequence of empirical barycenters converges a.s. to \(\mu^*\) for the 2-Wasserstein distance. Since \((\mathcal{M}_2^2(\Omega), d_{W_2})\) is compact, this finally shows that \((\bar{\mu}_n)_{n \geq 1}\) is a converging sequence such that \(\lim_{n \to +\infty} d_{W_2}(\bar{\mu}_n, \mu^*) = 0\) 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 \(\mu_{\theta} \in \mathcal{M}_2^2(\Omega)\) denote some random measure (as defined in Section 2.2) with \(\Omega \subset \mathbb{R}^d\), and let \(\mu_0\) be a fixed measure in \(\mathcal{M}_2^2(\Omega)\) admitting a density with respect to the Lebesgue measure on \(\mathbb{R}^d\). Then, by Proposition 2.1, there exists a unique optimal mapping \(\varphi_{\theta} : \Omega \to \Omega\) such that
In this section, we show that, for some specific probability models described below, the barycenter of $\mu_\theta$ is given by $\mu^* = E (\varphi_\theta) \# \mu_0$ (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 $\varphi_\theta$ to transport $\mu_0$ on $\mu_\theta$. 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 $X_1, \ldots, X_n$ (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 $\Omega$ 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 \cite{Grenander1975, Grenander1977, Grenander1981, Grenander1997} 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 $X_1, \ldots, X_n$ are obtained through the deformation of the same reference template $h \in L^2(\Omega)$ via the so-called deformable model

$$X_i = h \circ \varphi_i^{-1}, \quad i = 1, \ldots, n,$$

(7.1)

where $\varphi_1, \ldots, \varphi_n$ are iid random variables belonging to the set of smooth diffeomorphisms of $\Omega$. 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 $\varphi_i$, see e.g. \cite{Bouix2008, Baker2001, Trouve2003, Trouve2004, Trouve2005, Trouve2006, Trouve2007} 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 $\Omega$ via a diffeomorphism as in model (7.1).

It is always possible to transform the data $X_1, \ldots, X_n$ into a set of $n$ iid random probability densities by computing the random variables

$$Y_i(x) = \frac{\bar{X}_i(x)}{\int_{\Omega} \bar{X}_i(u)} du, \quad x \in \Omega,$$

where $\bar{X}_i(x) = X_i(x) - \min_{u \in \Omega} \{X_i(u)\}, \quad i = 1, \ldots, n.$

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

$$Y_i(x) = \left| \det D(\varphi_i^{-1})(x) \right| q_0(\varphi_i^{-1}(x)), \quad x \in \Omega, \quad i = 1, \ldots, n,$$

(7.2)
where \( \det \left(D\varphi_i^{-1}\right)(x) \) denotes the determinant of the Jacobian matrix of the random diffeomorphism \( \varphi_i^{-1} \) at point \( x \). If we denote by \( \mu_1, \ldots, \mu_n \in \mathcal{M}_2^+(\Omega) \) the random probability measures with densities \( Y_1, \ldots, Y_n \), and by \( \mu_0 \) the measure with density \( q_0 \), then (7.2) can also be written as the following deformable model of measures

\[
\mu_i = \varphi_i^#\mu_0, \quad i = 1, \ldots, n. 
\] (7.3)

In model (7.3), computing the empirical barycenter in the Wasserstein space of the random measures \( \mu_1, \ldots, \mu_n \) may lead to consistent and meaningful estimators of the reference measure \( \mu_0 \) 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 \( \mu_0 \) 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 \( \mu \) satisfying the deformable model

\[
\mu = \varphi^#\mu_0, 
\] (7.4)

for a specific class of random diffeomorphisms \( \varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d \). Let \( S^+_d(\mathbb{R}) \) be the set of non-negative definite \( d \times d \) symmetric matrices with real entries. Let \( \varphi : (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p)) \rightarrow \left(S^+_d(\mathbb{R}) \times \mathbb{R}^d, \mathcal{B} \left(S^+_d(\mathbb{R}) \times \mathbb{R}^d\right)\right) \) be a measurable mapping, where \( \mathcal{B}(S^+_d(\mathbb{R}) \times \mathbb{R}^d) \) is the Borel \( \sigma \)-algebra of \( S^+_d(\mathbb{R}) \times \mathbb{R}^d \). For \( \theta \in \mathbb{R}^p \), we will use the notations

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

and

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

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

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

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

\[
D_{\varphi}(\Theta) = \{\varphi_\theta, \ \theta \in \Theta\}. 
\] (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 \( \Theta \). We propose to study the population barycenter in the 2-Wasserstein space of the random measure \( \mu_\theta \) satisfying the deformable model

\[
\mu_\theta = \varphi_\theta^#\mu_0. 
\] (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 $\mu_\theta = \phi_\theta \# \mu_0$. Since $\phi_\theta$ is a smooth diffeomorphism and $\mu_0$ is a measure with density $q$ whose support is included in the compact set $\Omega_{q_0}$, it follows that $\mu_\theta$ admits a density $q_\theta$ on $\mathbb{R}^d$ given by

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

(7.7)

where $\mathcal{R}(\phi_\theta) = \{\phi_\theta(y), y \in \Omega_{q_\theta}\} = \{A_\theta y + b_\theta, y \in \Omega_{q_0}\}$. Before stating our main result on the population barycenter of the random measure $\mu_\theta$ (7.6), let us make the following regularity assumption on the mapping $\phi$.

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

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

### 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 $\phi_\theta$ taking their value in the parametric class $D_\phi$ defined by (7.5). Before stating the main result of this section, we define, for any $\theta \in \Theta$, the following quantities

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

(7.8)

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

**Theorem 7.1.** Let $\theta \in \mathbb{R}^p$ be a random vector with a density $g : \Theta \to \mathbb{R}$ that is continuously differentiable and such that $g(\theta) > 0$ for all $\theta \in \Theta$. Let $\mu_\theta$ be the random measure defined by the deformable model (7.6). Suppose that Assumption 3 holds.

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

$$q^*(x) = \det(\bar{A}^{-1}) q_0(\bar{A}^{-1}(x - \tilde{b})),$$

(7.9)

where $\bar{A} = \mathbb{E}(A_\theta), \tilde{b} = \mathbb{E}(b_\theta)$, and $\bar{A}_\theta, \bar{b}_\theta$ are the random variables defined by (7.8). Then, the following statements hold:

1. The primal problem $(P)$ satisfies

$$J_P = \inf_{\nu \in \mathcal{M}_+^1(\Omega)} J(\nu) = \frac{1}{2} \int_{\Theta} \int_{\Omega} dW_2^2(\mu^*, \mu_\theta) g(\theta) d\theta = \frac{1}{2} \int_{\Omega} \mathbb{E} \left( \left| \bar{A}_\theta u + \bar{b}_\theta - u \right|^2 \right) q^*(u) du,$$

(7.10)

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

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

(7.11)
where $I$ is the $d \times d$ identity matrix.

2. The population barycenter is the measure $\mu^* \in \mathcal{M}_2^d(\Omega)$ with density $q^*$ (with respect to the Lebesgue measure on $\mathbb{R}^d$) given by equation (7.9), that is

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

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 $\mu_0$ by the averaged amount of deformation measured by $\varphi$. In the case where $\varphi = I$ is the $d \times d$ identity matrix (which correspond to the assumption that $\mathbb{E}(A_\theta) = I$ and $\mathbb{E}(b_\theta) = 0$), the population barycenter $\mu^*$ is equal to the template measure $\mu_0$. 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 $\mu^*$ of the random measure $\mu_\theta$ defined by (7.6). To prove the results stated in Theorem 7.1, we will use the characterization (5.8) of the barycenter $\mu^*$. For this purpose, we need to find a maximizer $f_\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X)$ of the dual problem $(\mathcal{P}^*)$.

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

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

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

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

where

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

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

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

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

Proof of statement 1. of Theorem 7.1.

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

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

(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

$$
\int_{\Theta} \int_{\Omega} S_{\theta(x)} f_{\theta}(x) q_\theta(x) dx d\theta \leq \int_{\Theta} \int_{\Omega} \left( \frac{g(\theta)}{2} |x - \bar{A}_\theta^{-1} (x - \bar{b}_\theta)|^2 - f_{\theta}(\bar{A}_\theta^{-1} (x - \bar{b}_\theta)) \right) q_\theta(x) dx d\theta
$$

$$
\leq \int_{\Theta} \int_{\Omega} \left( \frac{g(\theta)}{2} \bar{A}_\theta u + \bar{b}_\theta - u|^2 - f_{\theta}(u) \right) q(u) dud\theta
$$

$$
\leq \int_{\Theta} \int_{\Omega} \left( \frac{g(\theta)}{2} \bar{A}_\theta u + \bar{b}_\theta - u|^2 \right) q(u) dud\theta
$$

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 $\Theta$ and $\Omega_{\bar{q}}$, and using Assumption 3, it follows that

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

Therefore, we have shown that

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

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

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

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

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

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

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

\[
S_{g(\theta)} f_\theta(x) = \frac{g(\theta)}{2} |x - A^{-1}_\theta (x - \bar b_\theta)|^2 + \frac{g(\theta)}{2} ((A_\theta - I) (A^{-1}_\theta (x - \bar b_\theta)), A^{-1}_\theta (x - \bar b_\theta)) \\
+ g(\theta)(\bar b_\theta, A^{-1}_\theta (x - \bar b_\theta)) \\
= g(\theta) |x|^2 - g(\theta) \langle x, A^{-1}_\theta (x - \bar b_\theta) \rangle + \frac{g(\theta)}{2} |A^{-1}_\theta (x - \bar b_\theta)|^2 \\
+ \langle (x - \bar b_\theta) , A^{-1}_\theta (x - \bar b_\theta) \rangle - g(\theta)(\bar b_\theta, A^{-1}_\theta (x - \bar b_\theta)) \\
= g(\theta) |x|^2 - \frac{g(\theta)}{2} \langle x, A^{-1}_\theta (x - \bar b_\theta) \rangle + \frac{g(\theta)}{2} (\bar b_\theta, A^{-1}_\theta (x - \bar b_\theta)) \\
= g(\theta) |x|^2 - \frac{g(\theta)}{2} (A^{-1}_\theta (x - \bar b_\theta))^2 + \frac{g(\theta)}{2} \langle x + \bar b_\theta, A^{-1}_\theta (x - \bar b_\theta) \rangle \\
- \frac{g(\theta)}{2} |A^{-1}_\theta (x - \bar b_\theta)|^2
\]  
(7.15)

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

\[
J^* f^\Theta = \int \int \int \frac{g(\theta)}{2} |x - A^{-1}_\theta (x - \bar b_\theta)|^2 q_\theta(x) dxd\theta + \int \int \frac{g(\theta)}{2} (x + \bar b_\theta, A^{-1}_\theta (x - \bar b_\theta)) q_\theta(x) dxd\theta \\
- \int \int \int \frac{g(\theta)}{2} |A^{-1}_\theta (x - \bar b_\theta)|^2 q_\theta(x) dxd\theta \\
= \int \int \int \frac{g(\theta)}{2} (A_\theta u + \bar b_\theta - u)^2 \tilde q(u) dudu \theta + \int \int \int \frac{g(\theta)}{2} (A_\theta u + 2\bar b_\theta, u) \tilde q(u) dudu \theta \\
- \int \int \int \frac{g(\theta)}{2} |u|^2 \tilde q(u) dudu \theta \\
= \frac{1}{2} \int \int \mathbb{E} (|\tilde A_\theta u + \bar b_\theta - u|^2) \tilde q(u) du,
\]

where we have used Fubini’s theorem combined with the fact that \( \int A_\theta g(\theta) d\theta = I \) and \( \int b_\theta g(\theta) d\theta = 0 \) to obtain the last equality.

Hence, thanks to the upper bound (7.14), we finally have that

\[
J^* (f^\Theta) = J_{P^*} = \frac{1}{2} \int \int \mathbb{E} (|\tilde A_\theta u + \bar b_\theta - u|^2) \tilde q(u) du,
\]

which proves that \( f^\Theta \) is a maximizer of the dual problem \( (P^*) \), 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

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

for every $\theta \in \Theta$. By equation (7.15), one has that

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

which implies that

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

Since $\mu_\theta$ is the measure with density $q_\theta(x) = \det (\bar{A}_\theta) \bar{q} (\bar{A}_\theta^{-1} (x - \bar{b}_\theta))$, one finally has that that $\mu^*$ 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.

\[\square\]

### 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 $q_0 : \mathbb{R}^d \to \mathbb{R}^+$ be a probability density function with compact support included in $[-A, A]^d$ for some constant $A > 0$. For $\theta$ a random vector in $\mathbb{R}^d$, we define the random density

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

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

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

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

(7.18)
By the law of large number, one has that
\[
\lim_{n \to +\infty} \bar{q}_n(x) = \int_{\mathbb{R}^d} q_0(x - \theta) g(\theta) d\theta \quad \text{a.s. for any } x \in \mathbb{R}^d.
\]

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 \( \theta \). 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 bar ycenter 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 \( M^+_{2d}(\mathbb{R}) \) where \( \Theta = [-A + \epsilon, A + \epsilon]^d \). Hence, Assumption 1 is satisfied. It is also clear that the mapping \( \phi : \Theta \to \mathbb{S}_d^+(\mathbb{R}) \times \mathbb{R}^d \) defined by
\[
\phi(\theta) = (I, \theta), \quad \theta \in \Theta,
\]
where \( I \) is the \( d \times d \) 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 \( \theta \) is random vector in \( \mathbb{R}^d \) having a continuously differentiable density \( g \) (with respect to the Lebesgue measure \( d\theta \) on \( \mathbb{R}^d \)). Assume that \( g \) has a compact support \( \Theta = [-\epsilon, \epsilon]^d \) for some \( \epsilon > 0 \). Let \( \mu_\theta \) be the random measure with density \( q_\theta(x) = q_0(x - \theta) \) (with respect to the Lebesgue measure \( dx \)) where \( q_0 : \mathbb{R}^d \to \mathbb{R}^+ \) is a probability density function with compact support included in \( [-A, A]^d \).

Then, the population barycenter \( \mu^* \) in the 2-Wasserstein space exists and is unique. It is the measure with density \( q_0(x - \mathbb{E}(\theta)) \), namely
\[
d\mu^*(x) = q_0(x - \mathbb{E}(\theta)) dx.
\]

The primal problem \((P)\) satisfies
\[
J_P = \inf_{\nu \in M^+_{2d}(\mathbb{R})} J(\nu) = \frac{1}{2} \int_{\Theta} d^2_{W_2}(\mu^*,\mu_\theta) g(\theta) d\theta = \frac{1}{2} \mathbb{E} \left( |\theta - \mathbb{E}(\theta)|^2 \right).
\]

(7.19)

Moreover, the family \( f^\Theta = (f_\theta)_{\theta \in \Theta} \in L^1(\Theta, X) \), defined by
\[
f_\theta(x) = -g(\theta)(\theta - \mathbb{E}(\theta)) x,
\]

(7.20)
is a maximizer of the dual problem \((P^*)\).

Hence, if it is assumed that the random shifts have zero expectation i.e. \( \mathbb{E}(\theta) = 0 \), then the density of the population barycenter \( \mu^* \) 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 \( q_{\theta_1}, \ldots, q_{\theta_n} \) 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 \( \mu_{\theta_i} = \varphi_{\theta_i} \# \mu_0, i = 1, \ldots, n \), 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 \( \mu_{\theta_1}, \ldots, \mu_{\theta_n} \) exists and is unique. Moreover, it is given by

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

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

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

which implies that

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

\[
\leq \left( 2 \int_\Omega |x|^2 d\mu_0(x) \right) \left( \left\| \frac{1}{n} \sum_{i=1}^{n} A_{\theta_i} - \bar{A} \right\|^2 + 2 \left\| \frac{1}{n} \sum_{i=1}^{n} b_{\theta_i} - \bar{b} \right\|^2 \right), \tag{7.22}
\]

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 \( \mu^* \), 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 \( X_1, \ldots, X_n \) be a sequence of independent random matrices in \( \mathbb{S}^d_+(\mathbb{R}) \). Suppose that \( \mathbb{E}X_i = 0 \) for all \( i = 1, \ldots, n \), and that there exist two positive constants \( B_1 \) and \( \sigma_1^2 \) such that

\[
\|X_i\| \leq B_1, \quad a.s. \quad i = 1, \ldots, n \quad \text{and} \quad \left\| \sum_{i=1}^{n} \mathbb{E}X_i^2 \right\| \leq \sigma_1^2.
\]

Then, for all \( t \geq 0 \),

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

**Proof.** We refer to [33]. \qed
Theorem 7.3 (Vector Bernstein inequality). Let $Y_1, \ldots, Y_n$ be a sequence of independent random vectors in $\mathbb{R}^d$. Suppose that $EY_i = 0$ for all $i = 1, \ldots, n$, and that there exist two positive constants $B_2$ and $\sigma_2^2$ such that

$$|Y_i| \leq B_2, \text{ a.s. } \quad \text{and} \quad \sum_{i=1}^n E|Y_i|^2 \leq \sigma_2^2.$$

Then, for all $t \geq 0$,

$$P\left(\left|\sum_{i=1}^n Y_i\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2/2}{\sigma_2^2 + B_2t/3}\right).$$

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

By Assumption 3, the mapping $\phi : \Theta \to S^+_d(\mathbb{R}) \times \mathbb{R}^d$ is continuous. Since $\Theta$ 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 $\sigma_1^2 = nE\|A_{\theta}\|^2$, $\sigma_2^2 = nE|b_{\theta}|^2$ 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

$$P\left(d_{W_2}^2(\tilde{\mu}_n, \mu^*) \geq t\right) \leq 2d \exp\left(-\frac{nt}{8\epsilon_0^2E\|A_{\theta}\|^2 + \frac{1}{2}B_1\epsilon_0\sqrt{t}}\right) + 2 \exp\left(-\frac{nt}{8E|b_{\theta}|^2 + \frac{1}{3}B_2\sqrt{t}}\right)$$

(7.23)

for any $t \geq 0$. Hence, the rate of concentration of $\tilde{\mu}_n$ to $\mu^*$ depends on the amount of variance of the reference measure $\mu_0$, and the amount of variability of the random mapping $\phi_{\theta}$. Finally, the concentration inequality (7.23) for $d_{W_2}^2(\tilde{\mu}_n, \mu^*)$ can be used to prove that $\tilde{\mu}_n$ converges in probability to $\mu^*$ 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 $\mu \in \mathcal{M}^2_+(\mathbb{R}^d)$ with distribution $\mathbb{P}$ whose support is not included in a compact set $\Omega$ of $\mathbb{R}^d$. In the one-dimensional case i.e. $d = 1$, let us denote by $F_\mu$ its cumulative distribution function, and by $F_\mu^{-1}$ its generalized inverse (quantile function). Then, one can define the measure $\mu^* \in \mathcal{M}^2_+(\mathbb{R})$ such that its quantile function is $F_\mu^{-1}(y) = \mathbb{E} \left( F_\mu^{-1}(y) \right)$ 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 $\mu^*$ is the unique population barycenter of the random measure $\mu$ 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 $\mu$. 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


31


