Geodesic PCA in the Wasserstein space

Jérémie Bigot\textsuperscript{1,2}, Raúl Gouet\textsuperscript{3}, Thierry Klein\textsuperscript{2} & Alfredo López\textsuperscript{3}

DMIA - ISAE\textsuperscript{1}
Institut de Mathématiques de Toulouse et CNRS (UMR 5219)\textsuperscript{2}
Université de Toulouse
Depto. de Ingeniería Matemática and CMM (CNRS, UMI 2807) \textsuperscript{3}
Universidad de Chile

May 11, 2014

Abstract

We introduce the method of Geodesic Principal Component Analysis (GPCA) analysis on the space of probability measures on the line, with finite second moments, endowed with the Wasserstein metric. We discuss the advantages of this approach over a standard functional PCA of probability densities in the Hilbert space of square-integrable functions. We establish the consistency of the method by showing that the empirical GPCA converges to its population counterpart as the sample size tends to infinity. We also give illustrative examples on simple statistical models to show the benefits of this approach for data analysis.

Keywords: Wasserstein space, Principal Component Analysis, Fréchet mean, Functional data analysis, Geodesic space, Inference for family of densities.

AMS classifications: Primary 62G08; secondary 62G20.

Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS, and the financial support from the French embassy in Chile and CONICYT PFCHA-2012. J.Bigot would like to thank the Center for Mathematical Modeling (CMM) and the CNRS for financial support and excellent hospitality while visiting Santiago, where part of this work was carried out. R. Gouet acknowledges support from Fondecyt grant 1120408 and project PFB-03-CMM.

1 Introduction

1.1 Main goals of this paper

The main goal of this paper is to define a notion of principal component analysis (PCA) of a family of probability measures $\nu_1, \ldots, \nu_n$, defined on the real line $\mathbb{R}$. In the case where the
measures admit square-integrable densities \( f_1, \ldots, f_n \), with respect to the Lebesgue measure, the standard approach is to use functional PCA (FPCA) (see e.g. [DPR82, RS05, Sil96]) on the Hilbert space \( L^2(\mathbb{R}) \) of square-integrable densities, endowed with its usual inner product. This method has already been applied in [KU01] for analyzing the main modes of variability of a set of densities.

Let us introduce some tools and notations of standard PCA in a separable Hilbert space \( H \), endowed with inner product \( \langle \cdot, \cdot \rangle \) and associated norm \( \| \cdot \| \). The PCA of the data \( x_1, \ldots, x_n \) in \( H \) is carried out by diagonalizing the covariance operator \( K : H \to H \) of \( x_1, \ldots, x_n \), defined by

\[
Kx = \frac{1}{n} \sum_{i=1}^{n} \langle x_i - \bar{x}, x \rangle (x_i - \bar{x}), \quad x \in H,
\]

where, \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) is the Euclidean mean. The eigenvectors of \( K \), associated to the largest eigenvalues, describe the principal modes of data variability around \( \bar{x} \). The first principal mode of linear variation of the data is defined by the \( H \)-valued curve \( g : \mathbb{R} \to H \) given by

\[
g_t = \bar{x} + t \sigma_1 w_1, \quad t \in \mathbb{R}, \tag{1.1}
\]

where \( w_1 \in H \) is the eigenvector associated to the largest eigenvalue \( \sigma_1 \geq 0 \).

On the other hand, it is well known that PCA can be formulated as the problem of finding a sequence of nested affine subspaces minimizing the norms of the projection residuals of the data. In particular, \( w_1 \) is a solution of

\[
\min_{v \in H, \|v\|=1} \frac{1}{n} \sum_{i=1}^{n} d^2(x_i, S_v) = \min_{v \in H, \|v\|=1} \frac{1}{n} \sum_{i=1}^{n} \|x_i - \bar{x}_n - \langle x_i - \bar{x}_n, v \rangle v \|^2, \tag{1.2}
\]

where \( S_v = \{ \bar{x}_n + tv, \ t \in \mathbb{R} \} \) is the affine subspace through \( \bar{x}_n \) with direction \( v \in H \) and \( d(x, S) = \inf_{x' \in S} \|x - x'\| \) denotes the distance from \( x \in H \) to \( S \subset H \).

Let us illustrate the strategy discussed above on a set of Gaussian densities \( f_1, \ldots, f_4 \), shown in Figure 1 in the Hilbert space \( H = L^2(\mathbb{R}) \). We first compute the Euclidean mean \( f_4 \), shown in Figure 1(e), and obtain a bi-modal density which is not a "satisfactory" average of the unimodal densities \( f_1, \ldots, f_4 \). We also compute the first mode of linear variation \( g \) given by (1.1) and observe that it is not a "meaningful" descriptor of the data variability. Indeed, for \( |t| \) sufficiently large, the function \( g_t \) may take negative values and does not integrate to one, as illustrated in Figure 2(e),(f). Moreover, even for small values of \( |t| \), \( g_t \) does not represent the typical shape of the observed densities \( f_1, \ldots, f_4 \), as shown by Figure 2(c),(d). Therefore, functional PCA of densities in \( L^2(\mathbb{R}) \) is not always interpretable as it leads to principal modes of linear variation that may be not coherent with the shape variability that is observed in the data. This drawback of functional PCA comes from the fact that the Euclidean distance \( \| f_1 - f_2 \|_{L^2(\mathbb{R})} \) is not necessarily appropriate to compare two measures \( \nu_1 \) and \( \nu_2 \), admitting \( f_1 \) and \( f_2 \) as densities.

In this paper we suggest to rather consider that the measures \( \nu_1, \ldots, \nu_n \) belong to the Wasserstein space \( W_2 \) of probability measures over \((\mathbb{R}, B(\mathbb{R}))\), with finite second order moment, endowed with the Wasserstein distance \( d_{W_2} \), associated to the quadratic cost; see [Vil03] for information.
on the Wasserstein space. In this setting, it is not possible to define a notion of PCA in the usual sense as the Wasserstein space $W_2$ is not linear. Nevertheless, in this paper we show how to define a proper notion of PCA in $W_2$, relying on the formal Riemannian structure of $W_2$, described in Section 2.3. A first idea in that direction is related to the mean of the data, which is an essential ingredient in any notion of PCA. We propose to use the notion of Fréchet mean (also called barycenter) as introduced in [AC11], whose asymptotic properties have been studied in [BK12]. It is significant to see that the barycenter of $f_1, \ldots, f_4$ in our example above, has a shape which is coherent with the shapes of the densities; see Figure 1(f).

1.2 Analogs of PCA for data belonging to a Riemannian manifold

There is currently a growing interest in the statistical literature on the development of nonlinear analogs of PCA, for the analysis of data belonging to curved Riemannian manifolds; see e.g. [FLPJ04, HHM10, SLHN10] and references therein. These methods, generally referred to as Principal Geodesic Analysis (PGA), extend the notion of classical PCA on Hilbert spaces. In this section we describe some of the main ideas of PGA.

Consider $y_1, \ldots, y_n$ belonging to a complete Riemannian manifold $\mathcal{M}$ admitting a geodesic distance $d_\mathcal{M}$. In order to define a PGA one needs a notion of average on $\mathcal{M}$. It has been suggested [FLPJ04] that the appropriate notion is the Frechet mean, defined as an element $z \in \mathcal{M}$ (not

---

Figure 1: (a,b,c,d) Graphs of Gaussian densities $f_1, \ldots, f_4$, with different means and variances. (e) Euclidean mean of these densities in the Hilbert space $L^2(\mathbb{R})$. (f) Density of the barycenter $\tilde{\nu}_n$ in the Wasserstein space $W_2$ of the probability measures $\nu_1, \ldots, \nu_4$ admitting $f_1, \ldots, f_4$ as densities.
Figure 2: An example of functional PCA of densities. First principal mode of linear variation \( g_t \) in the Hilbert space \( L^2(\mathbb{R}) \) for \(-2 \leq t \leq 2\), see equation \(1.1\), of the densities displayed in Figure 1.

necessarily unique) minimizing the sum of squared distances to the data, namely

\[
z \in \arg \min_{y \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^{n} d^2_{\mathcal{M}}(y, y_i).
\]

We refer to [BP03] for further details and properties of the Fréchet mean on Riemannian manifolds.

Let \( T_z \mathcal{M} \) be the tangent space to \( \mathcal{M} \) at \( z \). If \( v \) denotes a tangent vector in \( T_z \mathcal{M} \), there exists a unique geodesic \( \gamma_v(t) \), having \( v \) as its initial velocity, where \( t \in \mathbb{R} \) is a time parameter. The Riemannian exponential map \( \exp_z : T_z \mathcal{M} \to \mathcal{M} \), defined by \( \exp_z (v) = \gamma_v(1) \) is a diffeomorphism on a neighborhood of zero, and its inverse is the Riemannian log map, denoted by \( \log_z \).

1.2.1 PGA via linearization on the tangent space

In this approach, the data \( y_1, \ldots, y_n \) is first projected on \( T_z \mathcal{M} \) by means of the \( \log_z \) map, so let \( x_i = \log_z (y_i), i = 1, \ldots, n \). Next, we perform a standard PCA of \( x_1, \ldots, x_n \) in the linear space \((T_z \mathcal{M}, \langle \cdot, \cdot \rangle, \| \cdot \|)\), which leads to computing the first principal component \( v^{lin} \), the eigenvector
associated with the largest eigenvalue, of the covariance operator

\[ Kv = \frac{1}{n} \sum_{i=1}^{n} \langle x_i - \bar{x}_n, v \rangle (x_i - \bar{x}_n), \quad v \in T_z M, \]

where \( \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \). Finally, \( v^{lin} \) is projected back onto \( M \) by means of the \( \exp_z \) map, to obtain \( w^{lin} = \exp_z (v^{lin}) \), which represents a first notion of principal direction of geodesic variability. The main drawback of PGA via linearization is the fact that the distances are not preserved by the projection, i.e. \( \|x_i - x_j\| \neq d_M(y_i, y_j) \).

### 1.2.2 PGA along geodesics on \( M \)

Inspired by (1.2) which characterizes standard PCA, we consider the following notion of PGA along geodesics on \( M \). In a first step, one computes

\[ w^{geo} = \arg \min_{v \in T_z M, \|v\|=1} \frac{1}{n} \sum_{i=1}^{n} d_M^2(y_i, G_v), \]

where \( G_v = \{ \exp_z (tv), \; t \in \mathbb{R} \} \) and \( d_M(y, G) = \inf_{y' \in S} d_M(y, y') \) for \( y \in M \) and \( G \subset M \). Then, in a second (and final) step, one projects the element \( w^{geo} \) in \( T_z M \) onto \( M \), by computing \( w^{geo} = \exp_z (w^{geo}) \). This yields another notion of principal direction of geodesic variability of the data and generally one has that \( w^{lin} \neq w^{geo} \), except if \( M \) is a Hilbert space. Therefore, PGA via linearization on the tangent space and PGA along geodesics may lead to different directions of geodesic variability in a curved manifold. A detailed analysis of the differences between these methods can be found in [SLHN10]. In both methods, it is also possible to define subsequent principal directions (second, third, and so on) of geodesic variability in a recursive manner, and we refer to [FLPJ04] for further details.

### 1.3 Main contributions and organization of the paper

Since the Wasserstein space \( W_2 \) is not a Riemannian manifold, one cannot directly use the methods described above to define a notion of PCA. Nevertheless, the space \( W_2 \) has a formal Riemannian structure [AGS04] that we use to define a new notion of Geodesic PCA (GPCA) in \( W_2 \) similar to those described in Section 1.2. Another contribution of this paper is to prove the consistency of empirical GPCA performed on a random data set.

Before precisely defining GPCA, we display in Figure 3 the first principal mode of geodesic variation \( \tilde{g} \) in \( W_2 \), of the data displayed in Figure 1 see equation (1.1). GPCA clearly gives a better description of the variability in the data compared to the results displayed in Figure 2 that correspond to the first principal mode of linear variation \( g \) in \( L^2(\mathbb{R}) \), as defined in (1.1).

The paper is organized as follows. In Section 2 we give a precise definition of the Wasserstein space \( W_2 \) and we describe its formal Riemannian structure. In particular, we recall basic definitions such as tangent space, geodesics, exponential and logarithmic maps in the Wasserstein space framework, having their analogs in the Riemannian manifold setting. The main results are contained in sections 3 and 5. Section 3 is devoted to the construction and existence of

5
Figure 3: An example of GPCA of densities. First principal mode of geodesic variation $\tilde{g}_t$ in the Wasserstein space $W_2$ for $-2 \leq t \leq 2$ of the densities displayed in Figure 1. See equation (4.1).

GPCA in two variants, called global and nested, while Section 5 is dedicated to the consistency of GPCA when the number $n$ of random data points tends to infinity. Finally, we conclude the paper in Section 6 with a discussion of the differences between GPCA and the methods described in Section 1.2. In order to be self-contained, we give in the Appendix some classical results on geodesic spaces, Kuratowski convergence and $\Gamma$-convergence. Throughout the paper, we provide numerical experiments to illustrate the properties of GPCA in simple statistical models.

2 Preliminaries and definitions

2.1 A running example: the homothetic model

We begin this section by describing a simple construction of probabilities on the line, that will be used throughout the paper as illustrative example. Let $\mu_0 \in W_2$ with density $f_0$ and cdf $F_0$. For $(a, b) \in (0, \infty) \times \mathbb{R}$, let us denote by $\nu^{(a,b)}$ the probability admitting the density

$$f^{(a,b)}(x) := \frac{1}{a} f_0 \left( \frac{x - b}{a} \right), \, x \in \mathbb{R}. \quad (2.1)$$
and cdf
\[ F^{(a,b)}(x) := F_0 \left( \frac{x - b}{a} \right), \quad x \in \mathbb{R}. \] (2.2)

Then, if \((a_1,b_1), \ldots, (a_n,b_n)\) belong to \((0,\infty) \times \mathbb{R}\), we consider as a running example the case where the data \(\nu_1, \ldots, \nu_n\) satisfy the model
\[ \nu_i = \nu^{(a_i,b_i)}, \quad i = 1, \ldots, n, \] (2.3)
to be referred as the homothetic model of probabilities in the rest of the paper. The densities displayed in Figure 1 represent an example of realizations of this model with \(f_0\) being the standard normal density and \(n = 4\).

2.2 Optimal transportation theory

Let \(\Omega\) denote either the real line \(\mathbb{R}\) or a compact subset of \(\mathbb{R}\). We denote by \(W_2(\Omega)\) the set of probability measures over \((\Omega, B(\Omega))\) with finite second order moment, where \(B(\Omega)\) is the sigma field of Borel sets; if \(\Omega = \mathbb{R}\), we write \(W_2\) instead of \(W_2(\mathbb{R})\). We will also denote by \(W_{ac}^2(\Omega) \subset W_2(\Omega)\) the set of probability measures that are absolutely continuous with respect to the Lebesgue measure \(dx\) on \(\mathbb{R}\). If \(T : \Omega \to \Omega\) is a measurable map, we also recall that the push-forward measure \(T\#\mu\) of \(\mu \in W_2(\Omega)\) through the map \(T\) is defined as
\[ (T\#\mu)(A) = \mu(T^{-1}(A)) \quad \text{for all } A \in B(\Omega). \]

**Definition 2.1.** Given \(\mu, \nu \in W_2(\Omega)\), the Wasserstein distance between \(\mu\) and \(\nu\) is defined by
\[ d_{W_2}(\mu, \nu) := \left( \inf_{\pi \in \mathcal{P}(\mu, \nu)} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x,y) \right)^{1/2}, \]
where \(\mathcal{P}(\mu, \nu)\) is the set of probability measures on \(\Omega \times \Omega\) with marginals \(\mu\) and \(\nu\).

It can be shown that \(W_2(\Omega)\) endowed with the distance \(d_{W_2}\) is a metric space, the so-called Wasserstein space. For a detailed analysis of its properties, we refer to [Vil03]. Let us now recall the following key theorem from optimal transportation theory, due to Brenier [Bre91].

**Theorem 2.1.** Let \(\mu \in W_{ac}^2(\Omega)\) and \(\nu \in W_2(\Omega)\), then
\[ d_{W_2}^2(\mu, \nu) = \inf_{T \in \text{MP}(\mu, \nu)} \int_{\Omega} |T(x) - x|^2 d\mu(x), \] (2.4)
where \(\text{MP}(\mu, \nu) = \{ T : \Omega \to \Omega \mid T\#\nu = \mu \text{ and } T T^* = 0 \} \). Moreover, there exists \(T^* \in \text{MP}(\mu, \nu)\) such that \(d_{W_2}^2(\mu, \nu) = \int_{\Omega} |T^*(x) - x|^2 d\mu\), characterized as the unique (up to a \(\mu\) negligible set) element in \(\text{MP}(\mu, \nu)\) that can be represented as the gradient of a convex function.

Since we are in dimension one, a mapping \(T : \Omega \to \Omega\) is the gradient of a convex function if and only if \(T\) is increasing. Therefore, if \(F\) and \(G\) are the cumulative distribution functions (cdf) of \(\mu\) and \(\nu\), then the optimal mapping in Theorem 2.1 is given by \(T^* = G^{-} \circ F\) and
\[ d_{W_2}^2(\mu, \nu) = \int_{\Omega} (G^{-} \circ F(x) - x)^2 d\mu(x), \] (2.5)
where \( G^- (y) := \inf \{ x \in \mathbb{R} : G(x) \geq y \} \) is the quantile function of \( \nu \). Note that \( G^- \) is increasing, hence almost everywhere differentiable. By the change of variable \( x = F^- (y) \), we obtain the well-known characterization of the Wasserstein distance via the quantile functions of \( \mu \) and \( \nu \):

\[
d^2_{W_2} (\mu, \nu) = \int_0^1 (G^- (y) - F^- (y))^2 dy.
\]  

(2.6)

2.3 The pseudo-Riemannian structure of the Wasserstein space

The Wasserstein space over \((\Omega, B(\Omega))\) has a formal Riemannian structure considered, for example, in [AGS04]. We first provide some basic definitions in the Wasserstein space framework, having their analogs in the Riemannian manifold setting. Following [AGS04], given \( \mu \in W^{ac}_2 (\Omega) \) we define the tangent space at \( \mu \) as the Hilbert space \( L^2_\mu (\Omega) \) of real-valued, \( \mu \)-square-integrable functions, equipped with the standard inner product and its associated norm denoted by \( \| \cdot \|_\mu \). Furthermore, we define the exponential and the logarithmic maps at \( \mu \) as follows.

**Definition 2.2.** Let \( \mu \in W^{ac}_2 (\Omega) \) be a reference measure. The exponential map \( \exp_\mu : L^2_\mu (\Omega) \to W_2 (\Omega) \) and the logarithmic map \( \log_\mu : W_2 (\Omega) \to L^2_\mu (\Omega) \) are defined respectively as

\[
\exp_\mu (v) = (id + v)#\mu \quad \text{and} \quad \log_\mu (\nu) = G^- \circ F - id,
\]  

(2.7)

where \( F, G \) are the cdf of \( \mu, \nu \) respectively, \( id \) is the identity function on \( \Omega \), and \((id + v)#\mu \) is the push-forward measure of \( \mu \) through the map \( T = id + v \).

Note that the exponential map is well defined in the sense that \( \exp_\mu (v) \in W_2 (\Omega) \) for any \( v \in L^2_\mu (\Omega) \), since

\[
\int_\Omega x^2 d \exp_\mu (v) (x) = \int_\Omega (x + v(x))^2 d\mu (x) \leq 2 \int_\Omega x^2 d\mu (x) + \int_\Omega v^2 (x) d\mu (x) < +\infty.
\]

Similarly, the logarithmic map is well defined in the sense that \( \log_\mu (\nu) \in L^2_\mu (\Omega) \) since \( \| \log_\mu (\nu) \|_\mu^2 = d^2_{W_2} (\mu, \nu) < +\infty \), for all \( \nu \in W_2 (\Omega) \).

**Example 2.1.** In order to illustrate the notions of exponential and logarithmic maps, we consider the homothetic model \([2.3]\). Let \( \mu \in W^{ac}_2 (\Omega) \) be a reference measure with cdf \( F \). Then one has

\[
\log_\mu (\mu^{(a,b)}) (x) = [F^{(a,b)}]^- \circ F (x) - x, \ x \in \mathbb{R}.
\]

Taking \( \mu = \mu_0 \), from \([2.2]\), the expression above simplifies to

\[
\log_{\mu_0} (\nu^{(a,b)}) (x) = (a - 1)x + b, \ x \in \mathbb{R}.
\]

Therefore, if \( v \in L^2_{\mu_0} (\mathbb{R}) \) is the function \( v(x) = (a - 1)x + b, \ x \in \Omega \), then

\[
\exp_{\mu_0} (v) = \nu^{(a,b)}.
\]  

(2.8)
In the setting of Riemannian manifolds, the exponential map at a given point is a local homeomorphism from a neighborhood of the origin at the tangent space onto the manifold. However, this is not the case for the exponential map \( \text{exp}_\mu \) defined above, as it is possible to find two arbitrarily small functions in \( L^2_\mu(\Omega) \) with same exponentials, see e.g. [AGS04]. On the other hand, we will show that \( \text{exp}_\mu \) is an isometry when restricted to a specific set of admissible functions defined below.

**Definition 2.3.** Let \( \mu \in W^{ac}_2(\Omega) \). The function \( v \in L^2_\mu(\Omega) \) is said to be admissible if \( \text{id} + v \) is increasing \( \mu \)-a.s. It is said to be strictly admissible if \( \text{id} + v \) is strictly increasing \( \mu \)-a.s. The set of admissible (resp. strictly) functions is denoted by \( V^*_{\mu} \) (resp. \( V^*_{\mu}^s \)).

**Lemma 2.1.** Let \( (T_k) \) be a sequence in \( L^2_\mu(\Omega) \) such that \( T_k \) is \( \mu \)-a.s increasing, for all \( k \geq 1 \). (a) If \( (T_k) \) converges pointwise \( \mu \)-a.s to \( T \in L^2_\mu(\Omega) \) then \( T \) is \( \mu \)-a.s increasing. (b) If \( (T_k) \) converges in norm to \( T \in L^2_\mu(\Omega) \) then \( T \) is \( \mu \)-a.s increasing.

**Proof.** (a) It is possible to find a measurable set \( A \), such that \( \mu(A) = 1 \), \( T_k \) is increasing in \( A \), for all \( k \geq 1 \) and \( T_k \to T \) pointwise in \( A \). Then it directly follows that \( T \) is increasing in \( A \).

(b) As \( (T_k) \) converges in norm to \( T \) and using a result from measure theory, we know there exists a subsequence \( (T_{k_j}) \) converging \( \mu \)-a.s to \( T \). Then the conclusion follows from (a).

We remark that \( V^*_{\mu} \) is not a linear space. The following result is a key ingredient for the construction of GPCA in the Wasserstein space.

**Proposition 2.1.** Let \( \mu \in W^{ac}_2(\Omega) \), then \( V^*_{\mu} \) is closed and convex in \( L^2_\mu(\Omega) \). Moreover \( V^*_{\mu} \) is convex and dense in \( V^*_{\mu} \).

**Proof.** For convexity take \( \lambda \in [0,1], v_1, v_2 \in V^*_{\mu} \). Then, from Definition 2.3 \( \text{id} + v_1 \) and \( \text{id} + v_2 \) are increasing and so is \( \text{id} + \lambda v_1 + (1-\lambda)v_2 = \lambda(\text{id} + v_1) + (1-\lambda)(\text{id} + v_2) \). The same arguments applies for the convexity of \( V^*_{\mu} \). For closeness, take a sequence \( (v_k)_{k \geq 1} \) in \( V^*_{\mu} \) converging to \( v \in L^2_\mu(\Omega) \). Then \( \text{id} + v_k \to \text{id} + v \) and, by Lemma 2.1, \( \text{id} + v \) is increasing, that is \( v \in V^*_{\mu} \). Finally, we prove that \( V^*_{\mu} \) is dense in \( V^*_{\mu} \) by taking a sequence \( (h_k)_{k \geq 1} \) in \( V^*_{\mu} \) such that \( h_k \to 0 \). Then, for \( v \in V^*_{\mu} \), \( v + h_k \in V^*_{\mu} \) and \( v + h_k \to v \), so the conclusion follows.

The following proposition shows that the exponential map in \( W^2(\Omega) \), restricted to the convex set of admissible functions \( V^*_{\mu} \), is an isometry. This result is a key property that will allow us to define and to compute the GPCA in \( W^2(\Omega) \).

**Proposition 2.2.** Let \( \mu \in W^{ac}_2(\Omega) \), then \( \text{exp}_\mu(V^*_{\mu}) = W^2(\Omega) \) and \( \text{exp}_\mu(V^*_{\mu}^s) = W^{ac}_2(\Omega) \). Also, the exponential map \( \text{exp}_\mu \) restricted to \( V^*_{\mu} \) or \( V^*_{\mu}^s \) is an isometric homeomorphism, with inverse given by \( \log_\mu \).

**Proof.** Let \( v \in W^2(\Omega) \) then, from Theorem 2.1, \( G^- \circ F \) is the unique \( \mu \)-a.s. increasing map such that \( (G^- \circ F)\#\mu = v \), where \( F \) and \( G \) are the cdf of \( \mu \) and \( \nu \) respectively. In other words, \( v := \log_\mu(v) = G^- \circ F - \text{id} \) is the unique element in \( V^*_{\mu} \) such that \( \text{exp}_\mu(v) = \nu \). Now, take \( v \in W^{ac}_2(\Omega) \) and observe that, since \( \nu \) is absolutely continuous with respect to the Lebesgue measure, \( G^- \) is strictly increasing (see for instance [EH13]). As \( F \) is strictly increasing \( \mu \)-a.s., we have that \( G^- \circ F \) is strictly increasing \( \mu \)-a.s. and so, \( v \in V^*_{\mu} \).
Let us now prove the isometry property. From (2.6) and a change of variable we obtain
\[ d_{W_2}^2(v_1, v_2) = \| \log_\mu(v_1) - \log_\mu(v_2) \|_\mu^2, \]
for any \( v_1, v_2 \in W_2^\text{ac}(\Omega) \), and so \( \exp_\mu : V_\mu \to W_2^\text{ac}(\Omega) \) is an isometry. Finally, thanks to Proposition 2.2, \( V_\mu^s \) is dense in \( V_\mu \), which implies that \( \exp_\mu : V_\mu \to W_2(\Omega) \) is an isometry as well. \( \square \)

2.4 Geodesics in the Wasserstein space

A general overview of geodesics in a metric space is given in the Appendix. In this section, we consider the notion of geodesic in the metric space \( W_2(\Omega) \) as given in Definition A.3. A direct consequence of Corollary A.1 and Proposition 2.2 is that geodesics in \( W_2(\Omega) \) are exactly the image under \( \exp_\mu \) of straight lines in \( V_\mu \), where \( \mu \in W_2^\text{ac}(\Omega) \) is a reference measure. In particular, given two measures in \( W_2(\Omega) \), there exists a unique shortest path connecting them.

This property is stated in the following lemma.

**Lemma 2.2.** Let \( \mu \in W_2^\text{ac}(\Omega) \) and \( \gamma : [0, 1] \to W_2(\Omega) \) be a curve, where \( v_0 := \log_\mu(\gamma(0)) \) and \( v_1 := \log_\mu(\gamma(1)) \). Then \( \gamma \) is a geodesic if and only if \( \gamma(t) = \exp_\mu((1-t)v_0 + tv_1), \) for all \( t \in [0, 1] \).

**Example 2.2.** To illustrate Lemma 2.2, let us consider again the homothetic model (2.3) and take \( \mu = \mu_0 \) as reference measure. If we let \( v_0 := \mu_0 \) and \( v_1 = \nu^{(a,b)} \) with \( a > 0, \ b \in \mathbb{R} \), then one has that
\[ v_0(x) := \log_{\nu_0}(v_0) = 0 \text{ and } v_1(x) := \log_{\nu_0}(v_1) = (a-1)x + b, \ x \in \mathbb{R}. \]

From (2.8) and Lemma 2.2, the curve \( \gamma : [0, 1] \to W_2(\Omega), \) defined by
\[ \gamma(t) = \exp_{\mu_0}((1-t)v_0 + tv_1) = \exp_{\mu_0}(t(a-1)x + tb) = \nu^{(a_t,b_t)}, \ t \in [0, 1], \]
is a geodesic such that \( \gamma(0) = \mu_0 = \nu^{(1,0)} \) and \( \gamma(1) = \nu^{(a,b)}, \) where \( a_t = 1 - t + ta \) and \( b_t = tb. \)

Moreover, for each time \( t \in [0, 1] \), the measure \( \gamma(t) \) admits the density
\[ f^{(a_t,b_t)}(x) = \frac{1}{a_t} f_0 \left( \frac{x-b_t}{a_t} \right), \ x \in \mathbb{R}. \tag{2.9} \]

In Figure 4, we display the densities \( f^{(a_t,b_t)}, \) for some values of \( t \in [0, 1] \), in the case where \( \mu_0 \) is the standard Gaussian measure, \( a = 0.5 \) and \( b = 2. \)

Note that, by Lemma 2.2, \( W_2(\Omega) \) endowed with the Wasserstein distance \( d_{W_2} \) is a geodesic space. Moreover, we immediately have the following corollary.

**Corollary 2.1.** Let \( \mu \in W_2^\text{ac}(\Omega) \) be a reference measure. Then \( G \subseteq W_2(\Omega) \) is geodesic (in the sense of Definition A.3) if and only if \( \log_\mu(G) \) is convex.

Finally, to define the GPCA in \( W_2(\Omega) \) we will need the following definition of the dimension of a geodesic subset.

**Definition 2.4.** Let \( G \) be a geodesic subset of \( W_2(\Omega) \). We define \( \dim(G) \), the dimension of \( G \), as the dimension of the smallest affine subspace of \( L^2_\mu(\Omega) \) containing \( \log_\mu(G) \).
Figure 4: Visualization of the densities \( f^{(a_t,b_t)} \) associated to the geodesic curve \( \gamma(t) = \nu^{(a_t,b_t)} \) in \( W_2 \), described in Example 2.2, with \( a = 0.5 \) and \( b = 2 \), in the case where \( \mu = \mu_0 \) is the standard Gaussian measure.

We observe that the previous definition does not depend on the choice of the measure \( \mu \). Indeed take \( \mu' \in W_2(\Omega) \) and \( E \) an affine subspace of \( L^2_\mu(\Omega) \) containing \( \log_\mu(G) \). It is easy to see that \( T_{\mu,\mu'} := \log_{\mu'} \circ \exp_\mu \) is an affine function from \( L^2_\mu(\Omega) \) to \( L^2_{\mu'}(\Omega) \), therefore \( T_{\mu,\mu'}(E) \) is an affine subspace of \( L^2_{\mu'}(\Omega) \) containing \( \log_{\mu'}(G) \) and \( \dim(E) = \dim(T_{\mu,\mu'}(E)) \). Observe also that, if \( \gamma : [0,1] \to W_2(\Omega) \) is a geodesic, then \( \gamma([0,1]) \) is a geodesic space of dimension 1.

3 Population Fréchet mean and principal geodesics

Throughout this section we consider a reference measure \( \mu \in W_2^{ac}(\Omega) \) and a \( W_2(\Omega) \)-valued random element \( \nu \) which we assume to be square-integrable, in the sense of the following definition:

**Definition 3.1.** A random measure \( \nu \in W_2^{ac}(\Omega) \) is said to be square-integrable if

\[
\mathbb{E}(d_{W_2}^2(\nu, \mu')) < +\infty,
\]

for some (thus for all) \( \mu' \in W_2^{ac}(\Omega) \).

From now on all random measure \( \nu \) in \( W_2(\Omega) \) are assumed to be square-integrable, according to the previous definition. Also, we consider \( W_2(\Omega) \) equipped with the Borel sigma field relative to the Wasserstein metric.
3.1 Fréchet Mean

A natural notion of averaging in $W_2(\Omega)$ is the Fréchet mean, studied in [BK12] in a general setting, for random measures with support in $\mathbb{R}^d$, $d \geq 1$. In what follows we define and give some properties of the population Fréchet mean $\mathcal{M}(\nu)$ of a random measure $\nu \in W_2(\Omega)$. Note that results are stated for the one-dimensional case i.e. $d = 1$. The higher dimensional case is much more involved and we refer to [AC11, BK12] for further details.

Let $\nu$ be a random measure in $L^2_\mu(\Omega)$, such that $\mathbb{E}(\|\nu\|_\mu) < +\infty$. Then the expectation $\mathbb{E}(\nu)$ of $\nu$ is defined by $\mathbb{E}(\nu)(x) = \mathbb{E}(\nu(x))$, for all $x \in \mathbb{R}$. Observe that $\|\mathbb{E}(\nu)\|_\mu \leq \mathbb{E}(\|\nu\|_\mu) < \infty$, hence $\mathbb{E}(\nu) \in L^2_\mu(\Omega)$. Also, if $\mathbb{P}(\nu \in V_\mu) = 1$, then clearly $\mathbb{E}(\nu) \in V_\mu$ and if $\mathbb{P}(\nu \in V^*_\mu) = 1$, then $\mathbb{E}(\nu) \in V^*_\mu$.

Proposition 3.1. Let $\nu \in W_2(\Omega)$ be a random measure. Then, one has the following.

(i) There exists a unique element $\mathcal{M}(\nu) \in \arg\min_{\pi \in W_2(\Omega)} \mathbb{E}\left(d^2_{W_2}(\nu, \pi)\right)$ that we define as the population Fréchet mean of $\nu$.

(ii) $\mathcal{M}(\nu) = \exp_\mu(\mathbb{E}(\nu))$, for any $\mu \in W_{2ac}(\Omega)$, where $\nu = \log_\mu(\nu)$.

(iii) $\mathcal{M}(\nu)$ has cdf $\bar{G}$ that satisfies $\bar{G}^- = \mathbb{E}(\bar{G}^-)$, where $\bar{G}$ is the cdf of $\nu$.

(iv) If $\mathbb{P}(\nu \in W_{2ac}(\Omega)) = 1$, then $\mathcal{M}(\nu) \in W_{2ac}(\Omega)$.

Proof. (i),(ii) Let $\mu \in W_{2ac}(\Omega)$, and let $\nu = \log_\mu(\nu)$. From Proposition [2,2] and as $\nu$ is square-integrable, we have that $\mathbb{E}\left(\|\nu - u\|^2_{\mu}\right) = \mathbb{E}\left(d^2_{W_2}(\nu, \exp_\mu(u))\right) < \infty$, for all $u \in L^2_\mu(\Omega)$. Therefore

$$\inf_{\pi \in W_2(\Omega)} \mathbb{E}\left(d^2_{W_2}(\nu, \pi)\right) = \inf_{u \in V_\mu} \mathbb{E}\left(\|\nu - u\|^2_{\mu}\right), \quad (3.1)$$

and $u^*$ is a minimizer of the right hand side (rhs) of (3.1) if and only if $\exp_\mu(u^*)$ is a minimizer of the left hand side (lhs) of (3.1). On the other hand, observe that $\mathbb{E}(\nu)$ belongs to the convex set $V_\mu$ and that it is clearly the unique minimizer of $u \mapsto \mathbb{E}\left(\|\nu - u\|^2_{\mu}\right)$ for $u \in V_\mu$. Hence $\mathbb{E}(\nu)$ is the unique minimizer of the rhs of (3.1) and thus, by Proposition [2,2] this implies that $\mathcal{M}(\nu) = \exp_\mu(\mathbb{E}(\nu))$ is the unique minimizer of the lhs of (3.1), for any $\mu \in W_{2ac}(\Omega)$.

(iii) From [EH13], it can be shown that if $\mu \in W_{2ac}(\Omega)$, with cdf $F$ and $T : \mathbb{R} \to \mathbb{R}$ is an increasing and left-continuous map, then the quantile of $T \# \mu$ is given by $G^- = T \circ F^-$. Therefore, by (ii) and Definition [2.2] we have

$$G^- = (id + \mathbb{E}(\nu)) \circ F^- = \mathbb{E}(id + \log_\mu(\nu)) \circ F^- = \mathbb{E}(G^- \circ F) \circ F^- = \mathbb{E}(G^-),$$

where $\nu = \log_\mu(\nu)$.

(iv) It is a direct consequence of (ii) and Proposition [2,3]. \qed

One can remark that Proposition [3.1](ii) implies that $\exp_\mu(\mathbb{E}(\log_\mu(\nu)))$ does not depend on $\mu$. 

12
3.2 Principal geodesics

Let us first introduce some definitions and results that will be needed to define the notion of principal geodesics of a random measure \( \nu \) in \( W_2(\Omega) \).

**Definition 3.2.** Let \( G \subset W_2(\Omega) \) and \( \nu \in W_2(\Omega) \). We define

\[
d_{W_2}(\nu, G) = \inf_{\pi \in G} d_{W_2}(\nu, \pi),
\]

the distance between \( \nu \) and the set \( G \).

**Definition 3.3.** Let \( \nu \) be a random measure in \( W_2(\Omega) \) and \( G \subset W_2(\Omega) \). We associate to \( G \) the cost given by the expected value of the square residual of projecting \( \nu \) onto \( G \), that is,

\[
\text{cost}(G) = E \left( d_{W_2}^2(\nu, G) \right).
\]

(3.2)

Note that \( \text{cost}(G) \) depends on the random measure \( \nu \) and that it is necessarily finite as \( \nu \) is assumed to be square-integrable. Observe also that \( \text{cost}(G) \) is monotone, in the sense that

\[
\text{cost}(G) \geq \text{cost}(F), \text{ if } G \subset F.
\]

(3.3)

**Definition 3.4.** Let \( CL \) be the metric space of nonempty and closed subsets of \( W_2(\Omega) \), endowed with the Hausdorff distance \( h_{W_2} \). (See Definition A.5). For an integer \( k \geq 1 \) let also

\[
CG_{\mu,k} = \{ G \in CL \mid \mu \in G, G \text{ is a geodesic set and } \dim(G) \leq k \},
\]

where \( \dim(G) \) is defined in Definition 2.4.

In the rest of this Section 3.2 we will concentrate on defining and showing the existence of principal geodesics under the assumption that \( \Omega \) is compact. In this case it is well known that \((W_2(\Omega), d_{W_2})\) is also a compact metric space, see e.g. [Vil03]. Note also that the compactness of \( W_2(\Omega) \) implies that every random measure \( \nu \) in \( W_2(\Omega) \) is square-integrable. Furthermore, compactness of \( W_2(\Omega) \) also implies that \( h_{W_2}(G_n, G) \to 0 \), as \( n \to \infty \) if and only if \( G_n \to G \) in the sense of Kuratowski (see Definition A.3). From Proposition A.1 we obtain the following result.

**Proposition 3.2.** For any \( \nu \in W_2(\Omega) \), the function \( G \in CL \mapsto d_{W_2}(\nu, G) \) is continuous.

**Corollary 3.1.** The function \( G \in CL \mapsto \text{cost}(G) \), defined in (3.2), is continuous.

**Proof.** Let \( \nu \in W_2(\Omega) \) be a random measure and \( (G_n) \) be a sequence in \( CL \) converging to \( G \in CL \), in the sense of Hausdorff’s distance. Take \( \pi \in W_2(\Omega) \) and observe that

\[
d_{W_2}^2(\nu, G_n) \leq 2d_{W_2}^2(\nu, \pi) + 2d_{W_2}^2(\pi, G_n).
\]

Since \( \nu \) is square-integrable, one has that \( E(d_{W_2}^2(\nu, \pi)) < \infty \) and, given that \( W_2(\Omega) \) is compact, \( d_{W_2}^2(\pi, G_n) \leq R \), for some constant \( R > 0 \). Hence, by applying Proposition 3.2 and the dominated convergence theorem, we conclude that \( \text{cost}(G_n) \to \text{cost}(G) \).

\[\square\]
Proposition 3.3. The sets $CL$ and $CG_{\mu,k}$, from Definition 3.4, are compact.

Proof. Since the $\log_\mu$ map is a continuous bijection from the compact set $W_2(\Omega)$ onto $V_\mu$ (see Proposition 2.2), then $V_\mu$ is also compact. Hence, from Proposition A.2, $CL(V_\mu) := \{ C \subset V_\mu \mid C \text{ is nonempty and closed} \}$ and $CC_{k,0}(V_\mu) := \{ C \in CL(V_\mu) \mid C \text{ is convex}, 0 \in C \text{ and } \dim(C) \leq k \}$ are compact metric spaces, for the topology induced by the Hausdorff distance $h_\mu$ of sets in $L^2_{\mu}(\Omega)$ (see Definition A.5). Now, consider the map $i_\mu : (CL(V_\mu), h_\mu) \rightarrow (CL, h_{W_2})$ and observe that $i_\mu(CL(V_\mu)) = CL$, by Proposition 2.2, and $i_\mu(CG_{k,0}(V_\mu)) = CG_{\mu,k}$, by Corollary 2.1. On the other hand, as the exponential map is an isometry, it is easy to see that

$$h_\mu(C_1, C_2) = h_{W_2}(exp_\mu(C_1), exp_\mu(C_2)), \quad C_1, C_2 \in CL(V_\mu).$$

(3.4)

Hence $i_\mu$ is an isometric bijection and so, $CL$ and $CG_{\mu,k}$ are compact.

3.2.1 Global principal geodesics

We now present a first definition of GPCA, namely the notion of global principal geodesic of a random measure $\nu$, with respect to a reference measure $\mu \in W_{ac}^2(\Omega)$.

Definition 3.5. Let $\mu \in W_{ac}^2(\Omega)$ be a reference measure, $\nu \in W_2(\Omega)$ a random measure and $k \geq 1$ an integer. A global $k$-principal $\mu$-geodesics of $\nu$ is an element of

$$G_{\mu,k} := \arg \min \{ \text{cost}(G) \mid G \in CG_{\mu,k} \}. \quad (3.5)$$

The following is the main result about existence of global principal geodesics.

Theorem 3.1. Let $\mu \in W_{ac}^2(\Omega)$ be a reference measure and $k \geq 1$ an integer. Then $G_{\mu,k}$, from Definition 3.5, is nonempty.

Proof. The result is a direct consequence of Corollary 3.1 and Proposition 3.3.

3.2.2 Nested principal geodesics

Here we introduce a second variant of GPCA, that we call nested principal geodesics, defined inductively This approach is inspired by the usual characterization of PCA in terms of a nested sequence of optimal linear subspaces.

Definition 3.6. Let $\mu \in W_{ac}^2(\Omega)$ be a reference measure, $\nu \in W_2(\Omega)$ a random measure and $k \geq 1$ an integer. A nested $k$-principal $\mu$-geodesics of $\nu$ is a $k$-tuple $(G_1, \ldots, G_k)$ such that

$$G_1 \in \arg \min \{ \text{cost}(G) \mid G \in CG_{\mu,1} \}$$

and for $j = 2, \ldots, k$,

$$G_j \in \arg \min \{ \text{cost}(G) \mid G \in CG_{\mu,j}, G \supset G_{j-1} \}.$$

The set of nested $k$-principal $\mu$-geodesics of $\nu$ is denoted by $N_{\mu,k}$.
Theorem 3.2. Let \( \mu \in W_2^a(\Omega) \) be a reference measure and \( k \geq 1 \) an integer. Then \( \mathcal{N}_{\mu,k} \), from Definition 3.6, is nonempty.

Proof. Let \( F, G_n \in \text{CL} \) such that \( F \subset G_n, n \geq 1 \). If \( (G_n) \) converges to \( G \in \text{CL} \) then, thanks to (ii) in Definition A.4, \( F \subset G \). Therefore, the set \( \{G \in \text{CL} : G \supset F\} \) is closed and so \( \{G \in \text{CG}_{\mu,k} : G \supset F\} \) is compact, thanks to Proposition 3.3. Thus, from Corollary 3.1

\[
\text{arg min}\{\text{cost}(G) \mid G \in \text{CG}_{\mu,k}, G \supset F\} \neq \emptyset.
\]

Let us show by induction on \( k \in \mathbb{Z}^+ \) that \( \mathcal{N}_{\mu,k} \) is nonempty. First observe that \( \mathcal{N}_{\mu,1} = \mathcal{G}_{\mu,1} \); then, from Theorem 3.1, we have \( \mathcal{N}_{\mu,1} \neq \emptyset \). Assume \( \mathcal{N}_{\mu,k-1} \neq \emptyset \), let \( (G_1, \ldots, G_{k-1}) \in \mathcal{N}_{\mu,k-1} \) and define \( G_k \) as an element in \( \text{arg min}\{\text{cost}(G) \mid G \in \text{CG}_{\mu,k}, G \supset G_{k-1}\} \), which was shown above to be nonempty. Finally \( (G_1, \ldots, G_{k-1}, G_k) \in \mathcal{N}_{\mu,k} \) by definition.

Remark 3.1. (1) For \( k = 1 \), the notions of global and nested principal geodesics coincide. However, this might be not the case for \( k \geq 2 \).
(2) Our proofs of existence of principal geodesics, in Theorems 3.1 and 3.2, rely on the assumption that \( \Omega \) is compact. However, this assumption is not a necessary condition as seen in Section 4, where we provide an example with \( \Omega = \mathbb{R} \) and compute a principal geodesic for the case of \( k = 1 \). We believe that the compactness assumption can be lifted, but this might significantly complicate our proofs.

3.3 Empirical Fréchet mean and principal geodesics

We define the empirical Fréchet mean of measures \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \) as follows.

**Definition 3.7.** An empirical Fréchet mean of \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \) is defined as an element of

\[
\text{arg min}_{\nu \in W_2(\Omega)} \frac{1}{n} \sum_{i=1}^n d_{W_2}^2(\nu_i, \nu).
\]

**Definition 3.8.** Given \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \), we denote by \( \nu^{(n)} \) the random measure in \( W_2(\Omega) \) such that \( \mathbb{P}(\nu^{(n)} = \nu_i) = 1/n \), for \( i = 1, \ldots, n \).

It is clear that \( \nu^{(n)} \), defined above, is square-integrable, according to Definition 3.1.

**Proposition 3.4.** For any \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \) there exists a unique empirical Fréchet mean, denoted by \( \bar{\nu}_n \). Furthermore,

\[
\bar{G}_n = \frac{1}{n} \sum_{i=1}^n G_i,
\]

where \( \bar{G}_n \) the cdf of \( \bar{\nu}_n \) and \( G_1, \ldots, G_n \) are the cdf of \( \nu_1, \ldots, \nu_n \) respectively.

Proof. First observe that the empirical Fréchet mean can be characterized as the Fréchet mean (Proposition 3.1 (i)) of the random measure \( \nu^{(n)} \) (see Definition 3.8), that is, \( \bar{\nu}_n = \mathcal{M}(\nu^{(n)}) \). Hence, the conclusions follow from Proposition 3.1 (i), (ii).
Observe that \( \{3.6\} \) is the well known formula of quantile averaging in statistics, see e.g. [ZM11, GLM13]. A detailed characterization of the empirical Fréchet mean in the Wasserstein space, can be found in [AC11], for the general case of measures supported on \( \mathbb{R}^d \), for any dimension \( d \geq 1 \).

**Definition 3.9.** Let \( \mu \in W_2^{ac}(\Omega) \) be a reference measure, \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \) and \( k \geq 1 \) an integer. The empirical global and nested \( k \)-principal \( \mu \)-geodesics of \( \nu_1, \ldots, \nu_n \) are defined as in Definitions 3.5 and 3.6 respectively, applied to the random measure \( \nu^{(n)} \), given in Definition 3.8. The sets of empirical global and nested \( k \)-principal \( \mu \)-geodesics are denoted by \( G_{\mu,k,n} \) and \( N_{\mu,k,n} \) respectively.

Note that Theorems 3.1 and 3.2 yield the existence of empirical principal geodesics. Observe also that cost function \( 3.2 \) can be written in this empirical setting as

\[
\text{cost}_n(G) := E\left(d_{W_2}^2(\nu^{(n)}, G)\right) = \frac{1}{n} \sum_{i=1}^{n} d_{W_2}^2(\nu_i, G). \tag{3.7}
\]

In this section we are assuming that the reference measure \( \mu \in W_2^{ac}(\Omega) \) is arbitrary. However, by analogy with PCA in Hilbert spaces, a natural choice for \( \mu \) would be the Fréchet mean \( \mathcal{M}(\nu) \), which belongs to \( W_2^{ac}(\Omega) \) if \( \mathcal{P}(\nu \in W_2^{ac}(\Omega)) = 1 \), thanks to Proposition 3.1(iv).

### 3.4 Formulation of GPCA as an optimization problem in \( L_\mu^2(\Omega) \)

In this section we formulate the empirical GPCA, as an optimization problem in \( L_\mu^2(\Omega) \). Then, in the next section, we use this formulation to compute principal geodesics in the case of the homothetic model. First we introduce some notation to be used throughout this section. For \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L_\mu^2(\Omega), k \geq 1 \), we denote by \( \text{span}(\mathcal{U}) \) the subspace spanned by \( u_1, \ldots, u_k \), by \( \Pi_{\text{span}(\mathcal{U})} \) the projection of \( v \in L_\mu^2(\Omega) \) onto \( \text{span}(\mathcal{U}) \cap V_\mu \) and by \( \Pi_{\text{span}(\mathcal{U}) \cap V_\mu} v \) the projection of \( v \) onto the closed convex set \( \text{span}(\mathcal{U}) \cap V_\mu \).

**Definition 3.10.** Let \( v \) be a random element in \( L_\mu^2(\Omega) \) such that \( E(\|v\|^2) < \infty \) and \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L_\mu^2(\Omega), k \geq 1 \). We associate to \( \mathcal{U} \) the cost given by the expected value of the square residual of projecting \( v \) onto the closed convex set \( \text{span}(\mathcal{U}) \cap V_\mu \), that is,

\[
\text{cost}^*(\mathcal{U}) = E\left(\|v - \Pi_{\text{span}(\mathcal{U}) \cap V_\mu} v\|^2_\mu\right). \tag{3.8}
\]

Note that \( \text{cost}^* \), which depends on the random element \( v \), is necessarily finite as \( \|E(v)\|^2 \leq E(\|v\|^2) < \infty \).

**Definition 3.11.** Let \( \mu \in W_2^{ac}(\Omega) \) be a reference measure and \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L_\mu^2(\Omega), k \geq 1 \). We define the geodesic set generated by \( \mathcal{U} \) as

\[
G_\mathcal{U} := \exp_\mu(\text{span}(\mathcal{U}) \cap V_\mu).
\]

Observe that, by Corollary 3.1, \( G_\mathcal{U} \) is indeed a geodesic set in \( W_2(\Omega) \). In order to simplify the notation, we write \( \text{span}(u) \), \( \text{cost}^*(u) \) or \( G_u \), in the definitions above, if \( \mathcal{U} = \{u\} \).

The following proposition shows that the problem of finding global \( k \)-principal \( \mu \)-geodesics, (see Definition 3.5), can be formulated as an optimization problem in \( (L_\mu^2(\Omega))^k \).
Proposition 3.5. Let \( \mu \in W^{ac}_2(\Omega) \) be a reference measure, \( \nu \in W_2(\Omega) \) a random measure, \( v = \log \mu(\nu) \) and \( k \geq 1 \) an integer. Let \( U^* = \{u^*_1, \ldots, u^*_k\} \) be a minimizer of \( \text{cost}^* \), given by (3.8), over orthonormal sets \( U = \{u_1, \ldots, u_k\} \subset L^2_\mu(\Omega) \), then \( G_{U^*} \in G_{\mu,k} \).

Proof. Recall that, by Corollary 2.1, geodesic sets in \( W_2(\Omega) \) correspond to the image under \( \exp_\mu \) of convex sets in \( V_\mu \). Therefore, as \( \exp_\mu \) is an homeomorphism, we have

\[
G_{U^*} \in CG_{\mu,k}.
\]

On the other hand, from (3.2), (3.8) and as \( \exp_\mu \) is an isometry (see Proposition 2.2), we have

\[
\text{cost}^*(U) = \text{cost}(G_{U}), \quad U = \{u_1, \ldots, u_k\} \subset L^2_\mu.
\]

Again by Corollary 2.1, given \( G \in CG_{\mu,k} \), there exists an orthonormal set \( U = \{u_1, \ldots, u_k\} \subset L^2_\mu(\Omega) \), such that \( G \subset G_U \). Thus, as cost is monotone, in the sense of (3.3), and from (3.10), we have

\[
\text{cost}(G) \geq \text{cost}(G_{U}) = \text{cost}^*(U) \geq \text{cost}^*(U^*) = \text{cost}(G_{U^*})
\]

and the conclusion follows thanks to (3.9).

Similarly to the previous proposition, the next result shows that the problem of finding nested principal geodesics (see Definition 3.9) can be formulated as a sequence of optimization problem in \( L^2_\mu(\Omega) \). The proof is based on the same arguments as the proof of Proposition 3.5, so it is omitted.

Proposition 3.6. Let \( \mu \in W^{ac}_2(\Omega) \) be a reference measure, \( \nu \in W_2(\Omega) \) a random measure, \( v = \log \mu(\nu) \) and \( k \geq 1 \) an integer. Let \( u^*_1, \ldots, u^*_k \in L^2_\mu(\Omega) \) such that

\[
u^*_1 \in \arg\min\{\text{cost}^*(u) \mid u \in L^2_\mu(\Omega), \|u\|_\mu = 1\}
\]

and for \( j = 2, \ldots, k \),

\[
u^*_j \in \arg\min\{\text{cost}^*(u) \mid u \in \text{span}(u^*_1, \ldots, u^*_{j-1})^\perp, \|u\|_\mu = 1\}
\]

where \( \perp \) denotes orthogonal. Then \( (G_{u^*_1}, G_{\{u^*_1, u^*_2\}}, \ldots, G_{\{u^*_1, \ldots, u^*_k\}}) \in N_{\mu,k} \).

Given data \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \), Propositions 3.5 and 3.6 can be applied to the random measure \( \nu^{(n)} \) of Definition 3.8. In this case, the empirical version of (3.8) can be written as

\[
\text{cost}_{n}^*(U) = \frac{1}{n} \sum_{i=1}^{n} \|v_i - \Pi_{\text{span}(U) \cap V_\mu} v_i\|_\mu^2
\]

and an optimal solution \( U^* = \{u^*_1, \ldots, u^*_k\} \) leads to the construction of empirical principal geodesics.
Motivated by the method described in Section 1.2.1, we consider applying the standard PCA to the logarithms of the data. The next proposition provides conditions ensuring that such procedure leads to a solution of GPCA. For the sake of simplicity, we state the result only for global principal geodesics.

**Proposition 3.7.** Let \( \mu \in W_{ac}^2(\Omega) \) be a reference measure, \( \nu \in W_2(\Omega) \) a random measure, \( \nu = \log_{\mu}(\nu) \) and \( k \geq 1 \) an integer. Let \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L^2_{\mu}(\Omega) \) be a set of orthonormal eigenvectors associated to the \( k \) largest eigenvalues of the covariance operator \( K : L^2_{\mu}(\Omega) \to L^2_{\mu}(\Omega) \), given by

\[
K\nu = \mathbb{E}(\nu - \mathbb{E}(\nu), \nu - \mathbb{E}(\nu)), \quad \nu \in L^2_{\mu}(\Omega).
\] (3.12)

If \( \Pi_{\text{span}(\tilde{U})} v \in V_{\mu} \) a.s. then \( G_{\tilde{U}} \in G_{\mu,k} \).

**Proof.** It is well known that \( \tilde{U} \) is minimizer of

\[
\tilde{\text{cost}}(\mathcal{U}) = \mathbb{E}\|v - \Pi_{\text{span}(\tilde{U})} v\|^2_{\mu},
\] (3.13)

over orthonormal sets \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L^2_{\mu}(\Omega) \). On the other hand, from (3.8) and (3.13), it is clear that

\[
\text{cost}^*(\mathcal{U}) \geq \tilde{\text{cost}}(\mathcal{U}), \quad \mathcal{U} = \{u_1, \ldots, u_k\} \subset L^2_{\mu}(\Omega).
\] (3.14)

As we have assumed that \( \Pi_{\text{span}(\tilde{U})} v \in V_{\mu} \) a.s. then, from (3.8) and (3.13), we have

\[
\text{cost}^*(\tilde{\mathcal{U}}) = \tilde{\text{cost}}(\tilde{\mathcal{U}}).
\] (3.15)

From (3.14) and (3.15) we have that \( \tilde{\mathcal{U}} \) is a minimizer of \( \text{cost}^*(\mathcal{U}) \) over orthonormal sets \( \mathcal{U} = \{u_1, \ldots, u_k\} \subset L^2_{\mu}(\Omega) \). Finally, from Proposition 3.7, we obtain the result. \( \square \)

Let \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \) and \( \nu^{(n)} \) from Definition 3.8. If in Proposition 3.7 we replace \( \nu \) by \( \nu^{(n)} \), we obtain the empirical version of this result. In this case, if \( \bar{\mathcal{U}} = \{\bar{u}_1, \ldots, \bar{u}_k\} \subset L^2_{\mu}(\Omega) \) are orthonormal eigenvectors associated to the \( k \) largest eigenvalues of the empirical covariance operator

\[
Kv = \frac{1}{n} \sum_{i=1}^{n} (v_1 - \bar{v}_n)(v_i - \bar{v}_n), \quad v \in L^2_{\mu}(\Omega),
\] (3.16)

with \( v_i = \log_{\mu} \nu_i \) and if \( \Pi_{\text{span}(\bar{\mathcal{U}})} v_i \in V_{\mu}, \quad i = 1, \ldots, n \), then \( G_{\bar{\mathcal{U}}} \in G_{\mu,k,n} \).

From Proposition 3.7, we can informally say that, if the data is sufficiently concentrated around the reference measure \( \mu \), then the GPCA in \( W_2(\Omega) \) can be simply obtained from standard PCA on logarithms.
4 Some numerical examples of GPCA in \( W_2 \)

In Section 4.1 we present an example of concentrated data such that the conditions in Proposition 3.7 are satisfied and so, the problem of finding principal geodesics is reduced to the standard PCA on the data logarithms. On the other hand, in Section 4.2 we present an example of spread out data, where the GPCA cannot be obtained from the standard PCA of the logarithms.

4.1 The case of sufficiently concentrated data

To illustrate the empirical GPCA in \( W_2 \), let us consider the set of \( n = 4 \) probability measures \( \nu_1, \ldots, \nu_4 \) with densities \( f_1, \ldots, f_4 \), displayed in Figure 1. These measures satisfy the homothetic model (2.3), with \( \mu_0 \) being the standard Gaussian measure (with zero mean and unit variance) and the values of \( a_i \)'s and \( b_i \)'s given in Table 1.

| \( i \) | 1  | 2  | 3  | 4  |
|------|----|----|----|----|
| \( a_i \) | 0.4 | 0.8 | 1.2 | 1.6 |
| \( b_i \) | -1.8 | -0.1 | 0.7 | 1.2 |

Table 1: Values of \((a, b)\).

We first calculate the Fréchet mean \( \bar{\nu}_4 \) of \( \nu_1, \ldots, \nu_4 \). To that end, we apply the quantile average formula (3.6), from which we obtain the density \( \bar{g}_4 \) of \( \bar{\nu}_4 \) (displayed in Figure 1(f)), given by

\[
\bar{g}_4(x) = f^{(\bar{a}_4, \bar{b}_4)}(x) = \frac{1}{\bar{a}_4} f_0 \left( \frac{x - \bar{b}_4}{\bar{a}_4} \right), \quad x \in \mathbb{R},
\]

where \( \bar{a}_4 = \frac{1}{4} \sum_{i=1}^4 a_i = 1 \) and \( \bar{b}_4 = \frac{1}{4} \sum_{i=1}^4 b_i = 0 \). Hence, in this example, one has that \( \bar{g}_4 = f_0 \) and \( \bar{\nu}_4 = \mu_0 \).

Observe that \( \nu_1, \ldots, \nu_4 \) are concentrated around their empirical Fréchet mean \( \bar{\nu}_4 \), in the sense that their expectations and variances are not too far from those of \( \bar{\nu}_4 \), see Figure 1. Let us apply Proposition 3.7 to compute explicitly an empirical first principal geodesic, taking \( \mu = \mu_0 \) as reference measure. Let \( \tilde{u} \) be the eigenvector associated to the largest eigenvalue of the empirical covariance operator

\[
Kv = \frac{1}{4} \sum_{i=1}^4 \langle v_i - \bar{v}, v \rangle (v_i - \bar{v}), \quad v \in L^2_{\mu_0}(\Omega)
\]

where

\[
v_i(x) = \log_{\mu_0}(\nu_i)(x) = \left( \frac{a_i}{\bar{a}_4} - 1 \right) x + b_i - \bar{b}_4 \frac{a_i}{\bar{a}_4} = (a_i - 1) x + b_i, \quad x \in \mathbb{R},
\]

for \( i = 1, \ldots, 4 \). Using the fact that the \( v_i \)'s are affine functions belonging to the space generated by the identity function and the constant function 1, which are orthonormal in \( L^2_{\mu_0}(\mathbb{R}) \), the operator \( K \) can be identified with the \( 2 \times 2 \) matrix

\[
M = \frac{1}{4} \sum_{i=1}^4 V_i V_i^t,
\]

19
with \( V_i = (a_i - 1, b_i)' \in \mathbb{R}^2, 1 \leq i \leq 4 \). Therefore, \( \tilde{u} \) is also an affine function i.e. \( \tilde{u}(x) = \tilde{\alpha}x + \tilde{\beta}, x \in \mathbb{R} \), where the vector \( \tilde{U} = (\tilde{\alpha}, \tilde{\beta})' \in \mathbb{R}^2 \) (with \( \tilde{\alpha} = 0.36 \) and \( \tilde{\beta} = 0.93 \)) is the eigenvector associated to the largest eigenvalue of \( M \). In other words, computing the vector \( \tilde{U} \) simply amounts to calculating the first eigenvector associated to the standard PCA of \( V_i \in \mathbb{R}^2, i = 1, \ldots, 4 \), that represent the slope and intercept parameters of the affine functions \( v_i \). In Figure 5 we display the vectors \( V_i \) (blue circles) for \( 1 \leq i \leq 4 \), together with the linear space spanned by \( \tilde{U} \) (red dash-dot line), which corresponds to the first principal direction of variation of this data set. Any affine function \( u(x) = \alpha x + \beta \) in \( L^2_{\mu_0} \) can be represented in \( \mathbb{R}^2 \) by the vector \( U = (\alpha, \beta)' \). If \( \alpha \geq -1 \), then such function belongs to \( V_{\mu_0} \), which corresponds to the case where the point \( (\alpha, \beta) \in \mathbb{R}^2 \) is located to the right of the vertical green dashed line in Figure 5. Hence, it can be seen from the projections of the vectors \( V_i \) onto the linear space spanned by \( \tilde{U} \), that \( \Pi_{\text{span}(\tilde{u})}v_i \in V_{\mu_0} \) for all \( 1 \leq i \leq 4 \), and therefore, from Proposition 3.7, we conclude that the family of probability measures

\[
G_{\tilde{u}} = \{ \tilde{\nu}_t := \exp_{\mu_0}(t\tilde{u}), \text{ for all } t \in \mathbb{R} \text{ such that } 1 + t\tilde{\alpha} \geq 0 \},
\]
is a first empirical principal geodesic. From (2.1) and (2.8), each \( \tilde{\nu}_t \) in \( G_{\tilde{u}} \) admits the density

\[
\tilde{g}_t(x) = \frac{1}{\tilde{a}_t} f_0 \left( \frac{x - \tilde{b}_t}{\tilde{a}_t} \right), \quad x \in \mathbb{R},
\]

with \( \tilde{a}_t = 1 + t\tilde{\alpha} \) and \( \tilde{b}_t = t\tilde{\beta} \). In Figure 5 we display the first principal mode of geodesic variation \( \tilde{g}_t \), for \(-2 \leq t \leq 2\), of the densities displayed in Figure 1. As already mentioned, GPCA in \( W_2 \) gives a better interpretation of the data variability when compared to the results given by the first principal mode of linear variation of these densities (in the Hilbert space \( L^2(\mathbb{R}) \)), displayed in Figure 2.
Figure 5: A two-dimensional representation of the affine functions $u(x) = \alpha x + \beta, x \in \mathbb{R}$ in $L^2_{\mu_0}$. The horizontal axis is the slope parameter $\alpha$, and the vertical axis is the intercept parameter $\beta$. The points to the right of the vertical green dashed line at $\alpha = -1$ correspond to the affine functions belonging to $V_{\mu_0}$. The blue circles correspond to the vectors $V_i = (a_i - 1, b_i)' \in \mathbb{R}^2$ that are associated to the affine functions $v_i(x) = (a_i - 1)x + b_i, x \in \mathbb{R}$, for $1 \leq i \leq 4$, corresponding to the measures admitting as densities the functions displayed in Figure 1. The dash-dot red line is the linear space spanned by the first eigenvector $\tilde{U} \in \mathbb{R}^2$ of a standard PCA of the vectors $V_1, \ldots, V_4$.

4.2 The case of spread out data

In this section, we exhibit a situation such that the GPCA of $\nu_1, \ldots, \nu_n$ in $W_2$ and the standard PCA of the functions $v_i = \log_{\bar{\nu}_n}(\nu_i), i = 1, \ldots, n$ in the Hilbert space $L^2_{\bar{\nu}_n}$ lead to different results. Let $\nu_1, \ldots, \nu_4$ from the homothetic model (2.3) with $\mu_0$ again chosen as the standard Gaussian measure and the values of $a_i$’s and $b_i$’s given in Table 2.

| $i$ | 1 | 2 | 3 | 4 |
|-----|---|---|---|---|
| $a_i$ | 0.2 | 0.2 | 0.2 | 3.4 |
| $b_i$ | -3 | -1 | 1 | 3 |

Table 2: Values of $(a, b)$. 

21
Since $\bar{a}_4 = 1$ and $\bar{b}_4 = 0$, the empirical Fréchet mean of these measures is $\bar{\nu}_4 = \mu_0$, as in the example in Section 4.1. From Figure 6 it can be seen that $\nu_1, \ldots, \nu_4$ are less concentrated around their Fréchet mean when compared with the previous example (see Figure 1).

Figure 6: An example of $n = 4$ Gaussian densities $f_1, \ldots, f_4$ satisfying the homothetic model (2.3) with means and variances given in Table 2. (e) Euclidean mean of these densities in the Hilbert space $L^2(\mathbb{R})$. (f) Density of the barycenter $\bar{\nu}_4$ in the Wasserstein space $W_2$ of the probability measures $\nu_1, \ldots, \nu_4$ with densities $f_1, \ldots, f_4$.

We apply Proposition 3.5 to compute an empirical first principal geodesic. Let

$$v_i(x) = \log_{\mu_0}(\nu_i)(x) = (a_i - 1) x + b_i, \ x \in \mathbb{R},$$

for $i = 1, \ldots, 4$. Note that the affine functions $v_i$’s belong to the space generated by the identity function and the constant function 1, which are orthonormal in $L^2(\mu_0(\Omega))$. Recall also that an affine function $u(x) = \alpha x + \beta, x \in \mathbb{R}$ belongs to $V_{\mu_0}$ if and only if $\alpha \geq -1$. Therefore, if $S = \{(\alpha, \beta) \in \mathbb{R}^2 \mid \alpha \geq -1\}$ then minimizing $\text{cost}_4^* (\text{4.11})$ is equivalent to minimizing

$$U \in \mathbb{R}^2 \mapsto \frac{1}{4} \sum_{i=1}^{4} \|V_i - \Pi_{\text{span}(U) \cap S} V_i\|^2,$$

(4.2)

with $V_i = (a_i - 1, b_i)' \in \mathbb{R}^2, 1 \leq i \leq 4$. We find a unique minimizer $U^* = (\alpha^*, \beta^*)$ of (4.2), and so $u^*(x) = \alpha^* x + \beta^*, x \in \mathbb{R}$, is the unique minimizer of $\text{cost}_4^*$. From Proposition 3.5 the set of probability measures

$$G_{u^*} := \{\nu_t^* := \exp_{\mu_0}(tu^*), \text{ for all } t \in \mathbb{R} \text{ such that } 1 + t\alpha^* \geq 0\},$$

22
is a first empirical principal geodesic, and, from (2.1) and (2.8), each \( \nu^*_t \) in \( G_{\nu^*} \) admits the density

\[
g^*_t(x) = \frac{1}{a^*_t} f_0 \left( \frac{x - b^*_t}{a^*_t} \right), \quad x \in \mathbb{R},
\]

(4.3)

with \( a^*_t = 1 + ta^* \) and \( b^*_t = t\beta^* \).

We also compute \( \tilde{U} = (\tilde{\alpha}, \tilde{\beta})' \in \mathbb{R}^2 \), the eigenvector associated to the largest eigenvalue of the empirical covariance operator of \( V_1, \ldots, V_4 \) and obtain that the affine function \( \tilde{u}(x) = \tilde{\alpha}x + \tilde{\beta}, \quad x \in \mathbb{R} \), is the eigenvector associated to the largest eigenvalue of the empirical covariance operator of \( v_1, \ldots, v_4 \). We see that \( U^* \neq \tilde{U} \), so we have \( \text{cost}_n^*(u^*) < \text{cost}_n^*(\tilde{u}) \) and, from (3.10), we obtain

\[
\text{cost}_n(G_{u^*}) < \text{cost}_n(G_{\tilde{u}}).
\]

The previous inequality proves that computing a GPCA in \( W_2 \) of a set of measures \( \nu_1, \ldots, \nu_n \) cannot always be achieved by simply performing a standard functional PCA in \( L^2_{\bar{\nu}_n}(\mathbb{R}) \) of the functions \( v_i = \log_{\bar{\nu}_n}(\nu_i), i = 1, \ldots, 4 \). Figure 7 shows the vectors \( V_i \)'s and the convex sets \( \text{span}(U^*) \cap S \) and \( \text{span}(\tilde{U}) \cap S \). From this figure, one can see that (4.2) is strictly smaller evaluated at \( U^* \) than evaluated at \( \tilde{U} \).
Figure 7: A two-dimensional representation of the affine functions \( u(x) = \alpha x + \beta, x \in \mathbb{R} \) in \( L^2_{\mu_0}(\Omega) \). The horizontal axis is the slope parameter \( \alpha \), and the vertical axis is the intercept parameter \( \beta \). The points that lie at \( S \) (the region to the right of the vertical green dashed line at \( \alpha = -1 \)) correspond to the affine functions belonging to \( V_{\mu_0} \). The blue circles correspond to the vectors \( V_i = (a_i - 1, b_i)' \in \mathbb{R}^2 \) that are associated to the affine functions \( v_i(x) = (a_i - 1)x + b_i, x \in \mathbb{R} \), for \( 1 \leq i \leq 4 \), which are the logarithms of the measures admitting as densities the functions displayed in Figure 6. The dash-dot red line is the linear space spanned by the first eigenvector \( U' \in \mathbb{R}^2 \) of the standard PCA of \( V_1, \ldots, V_4 \). The black line is the convex set \( \text{span}(U^*) \cap S \) where \( U^* = (\alpha, \beta^*)' \in \mathbb{R}^2 \) is the minimizer of (4.2). The black dot is the projection of \( V_1 = (0.2, -3) \) onto \( \text{span}(U^*) \cap S \), while the red dot is the projection of the vector \( V_1 \) onto \( \text{span}(U') \cap S \).

5 Consistency of the empirical GPCA

As in Section 3, we assume that \( \Omega \) is compact and that \( \nu \) is a square-integrable random element in \( W_2(\Omega) \). We will study the convergence of the empirical Fréchet mean and the empirical global principal geodesics to their population counterparts, when \( \nu_1, \ldots, \nu_n \) are iid copies of \( \nu \).

Proposition 5.1. Let \( \nu \) be a random measure in \( W_2(\Omega) \). Let \( \nu_1, \ldots, \nu_n \) be iid copies of \( \nu \) and denote by \( \bar{\nu}_n \) their empirical Fréchet mean. Then, as \( n \to \infty \),

\[
d_{W_2}(\bar{\nu}_n, \mathcal{M}(\nu)) \to 0 \ a.s.
\]
Proof. Let $\mu \in W^{ac}_2(\Omega)$, then from Proposition \ref{prop:comparison} (ii), $\log_\mu(M(\nu)) = \mathbb{E}(v)$ and $\log_\mu(\nu_n) = \frac{1}{n} \sum_{i=1}^n v_i$, where $v = \log_\mu(\nu)$ and $v_i = \log_\mu(\nu_i), i = 1, \ldots, n$. Observe that $v_1, \ldots, v_n$ are iid copies of $v$ and that $\mathbb{E}(\|v\|^2) = \mathbb{E}(d^2_\nu(\nu, \mu)) < \infty$. Then, By Proposition \ref{prop:comparison} $d^2_\nu(\nu_n, M(\nu)) = \|\frac{1}{n} \sum_{i=1}^n v_i - \mathbb{E}(v)\|^2 \to 0$, a.s., by the strong law of large numbers in a Hilbert space (see e.g. \cite{LT11}).

Observe that the previous lemma is also valid for non compact $\Omega$, provided that $\nu$ is square-integrable. In the following lemma we show that the indicators of $CG_{\mu_n,k}$, $\Gamma$-converge to the indicator of $CG_{\mu,k}$ when $\mu_n$ converges to $\mu$ in $W^{ac}_2(\Omega)$. We refer to Section \ref{sec:Gamma-convergence} in the Appendix for the definitions of $\Gamma$-convergence and of indicator function.

**Lemma 5.1.** Let $(\mu_n)$ be a sequence in $W^{ac}_2(\Omega)$ converging to $\mu \in W^{ac}_2(\Omega)$, then

$$\Gamma\text{-}\lim_{n \to \infty} \chi_{CG_{\mu_n,k}} = \chi_{CG_{\mu,k}}.$$ (5.1)

**Proof.** By Lemma \ref{lem:Gamma-convergence} with $(X, d) = (CL, h_{W_2})$, it is sufficient to show that $CG_{\mu_n,k}$ converges to $CG_{\mu,k}$ in the sense of Kuratowski. That is, we have to show that

(a) for every $G \in CG_{\mu,k}$ there exists a sequence $(G_n)$ converging to $G$ such that $G_n \in CG_{\mu_n,k}$, for every $n \geq 1$, and

(b) if $G$ is an accumulation point of a sequence $(G_n)$, with $G_n \in CG_{\mu_n,k}, n \geq 1$, then $G \in CG_{\mu,k}$.

For (a) let $R_n : L^2_\mu(\Omega) \to L^2_{\mu_n}(\Omega)$ be defined by $R_n(u) = u \circ F^{-1} \circ F_n$, where $F$ and $F_n$ are the cdf of $\mu$ and $\mu_n$ respectively. Let also $T_n : W_2(\Omega) \to W_2(\Omega)$ given by $T_n = \exp_{\mu_n} \circ R_n \circ \log_\mu$, $n \geq 1$. Take $G \in CG_{\mu,k}$ and define $G_n = T_n(G)$, $n \geq 1$. From Corollary \ref{cor:comparison} and as $R_n$ is linear, it is easy to check that $G_n \in CG_{\mu,k}, n \geq 1$. Denote by $d_\mu$ the distance in $L^2_\mu(\Omega)$. As the logarithmic map is an isometry and after some calculation we get that the deviation from $G_n$ to $G$ (see Definition \ref{def:Gamma-convergence}), satisfies

$$d_{W_2}(G, G_n) = d_\mu(\log_\mu(G), \log_\mu(G_n)) \leq \|log_\mu(\mu_n)\|_\mu = d_{W_2}(\mu, \mu_n), n \geq 1.$$

Similarly, $d_{W_2}(G_n, G) \leq d_{W_2}(\mu, \mu_n), n \geq 1$, and we conclude that

$$h_{W_2}(G, G_n) \leq d_{W_2}(\mu, \mu_n) \to 0, \text{ as } n \to \infty.$$

For (b) take $G, G_n, n \geq 1$ as stated above. Since $\mu_n \in G_n, n \geq 1$, and $\mu_n \to \mu$, it follows that $\mu \in G$, by (ii) in Definition \ref{def:Gamma-convergence}.

On the other hand, let $CG_k = \{G \in CL \mid G \text{ is a geodesic set and } \dim(G) \leq k\}$, which is shown to be compact, following the same arguments in Proposition \ref{prop:Gamma-convergence}. Then, as $G_n \in CG_{\mu_n,k}, n \geq 1$, we have $G \in CG_k$ and recalling that $\mu \in G$, we conclude that $G \in CG_{\mu,k}$.

As $CL$ is compact, every sequence $(G_n)$ with $G_n$ belonging to the set of empirical global $k$-principal $\nu$-geodesics $G_{\mu_n,k,n}, n \geq 1$ has a convergent subsequence. The following theorem ensures that the limit set belongs almost surely to the population global $k$-principal $M(\nu)$-geodesics $G_{M(\nu),k}$. This is the main result about consistency of global GPCA.

25
Theorem 5.1. Let $\nu$ be a random measure in $W^\text{ac}_2(\Omega)$ with Fréchet mean $\mu := \mathcal{M}(\nu)$ and let $\nu_1, \ldots, \nu_n$ be iid copies of $\nu$ with empirical Fréchet mean $\mu_n := \bar{\nu}_n$. Let $G_{\mu,k}$ be the global $k$-principal $\mu$-geodesics of $\nu$ and $G_{\mu_n,k,n}$ be the empirical global $k$-principal $\mu_n$-geodesics of $\nu_1, \ldots, \nu_n$.

Then for every sequence $(G_n)$, with $G_n \in G_{\mu_n,k,n}$, one has that

$$\lim_{n \to \infty} \text{cost}_n(G_n) = \min \{ \text{cost}(G) \mid G \in CG_{\mu,k} \} \text{ a.s.}$$

(5.2)

Moreover, the accumulation points of $(G_n)$ belong to $G_{\mu,k}$ a.s. In other words, subsequential limits of $(G_n)$ are global principal population geodesics.

Proof. Observe that $\mu, \mu_n \in W^\text{ac}_2(\Omega)$, $n \geq 1$, by Proposition 3.1(iv). Also, by Proposition 5.1, $\mu_n \to \mu$ a.s. On the other hand, one may remark that the set of global principal geodesics (3.5) can be characterized as

$$G_{\mu,k} = \arg \min \{ \text{cost}(G) + \chi_{CG_{\mu,k}}(G) \mid G \in CL \},$$

(5.3)

where $\chi_{CG_{\mu,k}} : CL \to \mathbb{R} \cup \{+\infty\}$ is the indicator of $CG_{\mu,k}$ according to Definition A.9. Similarly,

$$G_{\mu_n,k,n} = \arg \min \{ \text{cost}_n(G) + \chi_{CG_{\mu_n,k,n}}(G) \mid G \in CL \}.$$

(5.4)

By applying Lemma 5.1 we have that

$$\Gamma- \lim_{n \to \infty} \chi_{CG_{\mu_n,k,n}} = \chi_{CG_{\mu,k}} \text{ a.s.,}$$

(5.5)

where the $\Gamma$-convergence holds in the space CL. From Proposition 3.2 and recalling that $W_2(\Omega)$ is compact, we have that $d^2_{W_2}(\nu, G)$ is separately continuous in $\nu \in W_2(\Omega)$ and $G \in CL$. Hence $d^2_{W_2}(\nu, G)$ is measurable on the product space $W_2(\Omega) \times CL$; see [Joh69] or [Rud81]. Also, from Theorem 2.3 in [AW95], we have the following $\Gamma$-convergence in CL,

$$\Gamma- \lim_{n \to \infty} \text{cost}_n = \text{cost} \text{ a.s.,}$$

(5.6)

On the other hand, as $W_2(\Omega)$ is compact, there exists a constant $R > 0$ such that $d^2_{W_2}(\nu, G) \leq R$, for all $\nu \in W_2(\Omega)$ and $G \in CL$. Also, by Proposition 3.3, CL is a compact set. Therefore, by the uniform strong law of large number (see e.g. Lemma 2.4 in [NM94]), $\text{cost}_n(G) \to \text{cost}(G)$ uniformly in CL a.s. i.e.

$$\lim_{n \to \infty} \sup_{G \in CL} | \text{cost}(G) - \text{cost}_n(G) | = 0 \text{ a.s.}$$

(5.7)

From (5.5) to (5.7) and by Proposition 6.24 in [DM93], we obtain

$$\Gamma- \lim_{n \to \infty} \text{cost}_n + \chi_{CG_{\mu_n,k}} = \text{cost} + \chi_{CG_{\mu,k}} \text{ a.s.}$$

(5.8)

Therefore, from (5.3), (5.4) and (5.8); the compactness of CL and Theorem A.1 the conclusions follow.
The following theorem ensures that the limit set of sequences \((G_n)\) with \(G_n \in \mathcal{N}_{\nu_n,k,n}, n \geq 1\), belongs almost surely to the population nested \(k\)-principal \(\mathcal{M}(\nu)\)-geodesics \(\mathcal{N}_{\mathcal{M}(\nu),k}\). This is the main result about consistency of nested GPCA.

**Theorem 5.2.** Let \(\nu\) be a random measure in \(W^2_{ac}(\Omega)\), with Fréchet mean \(\mu := \mathcal{M}(\nu)\), and let \(\nu_1, \ldots, \nu_n\) be iid copies of \(\nu\), with empirical Fréchet mean \(\mu_n := \nu_n\). Let \(\mathcal{N}_{\mu,k}\) be the nested \(k\)-principal \(\mu\)-geodesics of \(\nu\) and \(\mathcal{N}_{\mu_n,k,n}\) be the empirical nested \(k\)-principal \(\mu_n\)-geodesics of \(\nu_1, \ldots, \nu_n\). If \(G_n = (G_1, n, \ldots, G_{k,n}) \in \mathcal{N}_{\mu_n,k,n}\) and \((G_{n'})\) is a subsequence converging to \(G = (G_1, \ldots, G_k)\), then \(G \in \mathcal{N}_{\mu,k}\) and \(\text{cost}_n(G_{j,n'}) \to \text{cost}(G_j), j = 1, \ldots, k\).

**Proof.** Let us show the result by induction on \(k \in \mathbb{Z}_+\). First observe that \(\mathcal{N}_{\mu_n,1,n} = \mathcal{G}_{\mu_n,1,n}\) and \(\mathcal{N}_{\mu,1} = \mathcal{G}_{\mu,1}\), then the case \(k = 1\) follows from Theorem 5.1. Let us assume that the result is valid for \(k - 1\) and show that it is also true for \(k \geq 2\). Observe that set of population and empirical nested principal geodesic can be expressed as

\[
\mathcal{N}_{\mu,k} = \{(G_1, \ldots, G_k) \mid (G_1, \ldots, G_{k-1}) \in \mathcal{N}_{\mu,k-1}, G_k \in \mathcal{G}_{\mu,k} \text{ and } G_{k-1} \subset G_k\} \tag{5.9}
\]

and

\[
\mathcal{N}_{\mu_n,k,n} = \{(G_1, \ldots, G_k) \mid (G_1, \ldots, G_{k-1}) \in \mathcal{N}_{\mu_n,k-1,n}, G_k \in \mathcal{G}_{\mu_n,k,n} \text{ and } G_{k-1} \subset G_k\}, \tag{5.10}
\]

where \(\mathcal{G}_{\mu,k}\) and \(\mathcal{G}_{\mu_n,k,n}\) are the sets of population and empirical global principal geodesics, respectively. Let \(G_n = (G_1, n, \ldots, G_{k,n}) \in \mathcal{N}_{\mu_n,k,n}\) and \((G_{n'})\) be a subsequence of \((G_n)\) converging to an element \(G = (G_1, \ldots, G_k)\). From \(\text{(5.10)}\) we have \((G_1, n, \ldots, G_{k-1,n}) \in \mathcal{N}_{\mu_n,k-1,n}\) therefore, by hypothesis of induction, \((G_1, \ldots, G_{k-1}) \in \mathcal{N}_{\mu,k-1}\) and \(\text{cost}_n(G_{j,n'}) \to \text{cost}(G_j), j = 1, \ldots, k-1\). Also from \(\text{(5.10)}\), \(G_{k,n} \in \mathcal{G}_{\mu_n,k,n}\), \(n \geq 1\), hence \(G_k \in \mathcal{G}_{\mu,k}\) and \(\text{cost}_n(G_{k,n}) \to \text{cost}(G_k)\), thanks to Theorem 5.1. Finally, from Lemma \(\text{A.3}\) \(G_{k-1} \subset G_k\) and since \((G_1, \ldots, G_{k-1}) \in \mathcal{N}_{\mu,k-1}\) and \(G_k \in \mathcal{G}_{\mu,k}\), the result follows from \(\text{(5.9)}\). \(\square\)

The interpretation of Theorems 5.1 and 5.2 is that the empirical GPCA is strongly consistent in its two variants. From the proofs of these theorems, one can immediately see that \(\mu_n, n \geq 1\), can be arbitrary measures in \(W^2_{ac}(\Omega)\) provided that \(\mu_n \to \mu\). However, it seems natural to use \(\mu_n = \nu_n\) as a reference measures to compute the empirical GPCA in the Wasserstein space \(W_2(\Omega)\). In this case, we obtain convergence to the population GPCA, related to the reference measure \(\mu = \mathcal{M}(\nu)\).

### 6 Comparison between GPCA in \(W_2\) and analogs of PCA on Riemannian manifolds

As already mentioned in the introduction, nonlinear analogs of PCA have been proposed in the literature for the analysis of data belonging to curved Riemannian manifolds [FLPJ04, HHM10]. To perform a PCA-like analysis, two popular approaches are: (1) standard PCA of the data projected onto the tangent space at their Fréchet mean, with back projection onto the manifold, as presented in Section 1.2.1 and (2) PGA along geodesics, as described in Section 1.2.2. These
two approaches lead generally to different directions of geodesic variability in a curved manifold \cite{SLHN10}.

In this paper we consider the analysis of data in the Wasserstein space $W_2(\Omega)$, which is not a Riemannian manifold but has pseudo-Riemannian structure, rich enough to allow the definition of a notion of geodesic PCA. By means of the analogs of the logarithmic and the exponential maps, we also introduce the corresponding version of the standard PCA on the tangent space, with back projection onto $W_2(\Omega)$, thus establishing a parallel to the methodological duality available for data in Riemannian manifolds, described in the introduction. Also, as could be expected, these two approaches yield, in general, different forms of geodesic variability.

There is however a significant distinguishing feature of our methodology, namely the possibility of performing a PCA on the tangent space under convexity restrictions, which is equivalent (after projection) to the geodesic PCA on $W_2(\Omega)$. This restricted PCA on the logarithms of the data is interesting because it is formally simpler than the geodesic PCA in $W_2(\Omega)$ although more complex than standard PCA. In this respect it is also worth noticing that if the data are “sufficiently concentrated”, the standard and the restricted PCA on the tangent space yield the same results.

Finally, it should be mentioned that the terminology geodesic PCA (GPCA) was used previously by Huckemann et al. in \cite{HHM10} to denote a Riemannian manifold generalization of linear PCA. Their approach shares similarities with the PGA method introduced in \cite{FLPJ04}, but optimizes additionally for the placement of the center point (not necessarily equal to the Fréchet mean). Furthermore, it does not use a linear approximation of the manifold and is only suited for Riemannian manifolds, where explicit formulas for geodesics exist. However, it is difficult to compare our approach to the GPCA in \cite{HHM10} since the notion of principal geodesic, that we propose in this paper, is defined with respect to a given reference measure $\mu$ (chosen to be either the population or the empirical Fréchet mean). For a precise comparison it would be necessary to carry out the optimization (3.5) with respect to the reference measure $\mu$, a task which is beyond the scope of this paper.

A Appendix

A.1 Geodesics in metric spaces

We introduce the concept of geodesic in metric spaces; for notations, definitions and results, we follow \cite{Cho11} and references therein.

**Definition A.1.** A curve in a metric space $(X,d)$ is a continuous function $\gamma : I \to X$, where $I \subset \mathbb{R}$ is a closed (not necessarily bounded) interval. Also

(i) $\gamma$ is said to pass through $z \in X$ if $\gamma(t) = z$, for some $t \in I$;

(ii) $\gamma$ joins $x, y \in X$ if there exists $a, b \in I, a \leq b$, such that $\gamma(a) = x$ and $\gamma(b) = y$.

(iii) $\gamma$ is rectifiable if its length $L(\gamma)$ is finite.

For convenience, without loss of generality, we consider $I$ such that $[0,1] \subset I$. 

28
Definition A.2. A metric space \((X,d)\) is said to be geodesic if for every \(x, y \in X\), there exists a rectifiable curve \(\gamma\) joining \(x\) and \(y\), such that \(d(x,y) = L(\gamma)\). Such minimum length curve \(\gamma\) is called a shortest path between \(x\) and \(y\). A curve \(\gamma : I \to X\) is a geodesic if for every \(t \in I\), there exist \(a,b \in I, a < b, a \leq t \leq b\) such that the restriction of \(\gamma\) to \([a,b]\) is a shortest path between \(\gamma(a)\) and \(\gamma(b)\).

The following is a useful characterization of shortest path (See [Cho11] for a proof).

Lemma A.1. For any shortest path, there exists a continuous reparametrization \(\gamma\) on \([0, 1]\) such that
\[
d(\gamma(s), \gamma(t)) = |t - s|d(\gamma(0), \gamma(1))\quad \text{for all } s, t \in [0, 1].
\]

Lemma A.2. Let \(H\) be a Hilbert space and \(x, y \in H\). Then \(\gamma\) is a shortest path joining \(x\) and \(y\) if and only if \(\gamma(t) = (1 - t)x + ty\), for all \(t \in [0, 1]\), up to a continuous reparametrization.

Proof. Denote the inner product and the induced norm in \(H\) by \(\langle \cdot, \cdot \rangle\) and \(\| \cdot \|\) respectively. Let \(\gamma\) be a shortest path between \(x\) and \(y\), and \(t \in [0, 1]\). From Lemma A.1 we have
\[
\|x - \gamma(t)\| = t\|x - y\|
\]
and \(\|\gamma(t) - y\| = (1 - t)\|x - y\|\), then
\[
\|x - \gamma(t)\| + \|\gamma(t) - y\| = \|x - y\|.
\]
Squaring and simplifying the expression above, we get
\[
\|x - \gamma(t)\|\|\gamma(t) - y\| = \langle x - \gamma(t), \gamma(t) - y \rangle.
\]
Hence, by the Cauchy-Schwartz inequality, there exists \(\lambda \geq 0\) such that \(x - \gamma(t) = \lambda(\gamma(t) - y)\). Finally, taking norm we find \(\lambda = \frac{1}{1 - t}\) and the result follows. The other implication is direct.

From the previous lemma, we deduce that in Hilbert spaces, any geodesic is locally a segment, therefore geodesics correspond to straight lines. We state this in the following corollary.

Corollary A.1. Let \(H\) be a Hilbert space and \(\gamma : I \to H\) a curve such that \(\gamma(0) = x \in H\) and \(\gamma(1) = y \in H\). Then, \(\gamma\) is a geodesic if and only if \(\gamma(t) = (1 - t)x + ty\), for all \(t \in I\), up to a continuous reparametrization.

Definition A.3. Let \((X,d)\) be a geodesic space and \(Y \subset X\). We say that \(Y\) is geodesic if the induced metric space \((Y,d)\) is geodesic. In other words, if for any \(x, y \in Y\), there exists a shortest path joining \(x\) and \(y\), totally contained in \(Y\).

Note that, as direct consequence of Lemma A.2, any Hilbert space \(H\) is geodesic and \(C \subset H\) is geodesic if and only if \(C\) is convex.

A.2 \(K\)-convergence

In this section we provide definitions and results that we use for proving the existence of principal geodesics (see Section 3.2). In particular, we define an appropriate concept of convergence for sequences of convex sets in a metric space.
Definition A.4. Let \((X, d)\) be a metric space and \(C, C_n \subset X, n \geq 1\). We say that the sequence 
\((C_n)\) converges to \(C\) in the sense of Kuratowski, denoted by \(K\)-lim_{n \to \infty} C_n = C\), if 
(i) for every \(x \in C\) there exists a sequence \((x_n)\) converging to \(x\) such that \(x_n \in C_n, n \geq 1\);
(ii) if \((x_n)\) is a sequence such that \(x_n \in C_n, n \geq 1\), then any accumulation point of \((x_n)\) belongs to \(C\).

Definition A.5. Let \((X, d)\) be a metric space and \(A, B \subset X\). The deviation from \(x \in X\) to \(B\) is defined by 
\[d(x, B) := \inf_{x' \in B} d(x, x')\] 
the deviation from \(A\) to \(B\) is 
\[d(A, B) := \sup_{x \in A} d(x, B)\] 
and The Hausdorff distance between the sets \(A\) and \(B\) is 
\[h(A, B) := \max\{d(A, B), d(B, A)\}. \tag{A.1}\]

It is well known (see \cite{Pri} and references therein) that convergence with respect to the Hausdorff distance is stronger than convergence in the sense of Kuratowski. Moreover, if \(X\) is compact, then both notions of convergence coincide.

Definition A.6. Let \((X, d)\) be a metric space. We define the metric space 
\[\text{CL}(X) := \{C \subset X \mid C \neq \emptyset, \text{closed, bounded}\}\]
endowed with the Hausdorff distance between sets.

Proposition A.1. Let \((X, d)\) be a compact metric space and \(y \in X\), then the function \(C \mapsto d(y, C)\) is continuous in \(\text{CL}(X)\).

Proof. Let \(y \in X\) and let \((C_n)\) be a sequence in \(\text{CL}(X)\) converging to \(C \in \text{CL}(X)\). Observe that, by the compactness of \(X\), there exists \(x^* \in C\) and \(x_n^* \in C_n, n \geq 1\), such that 
\[d(y, C_n) = d(y, x_n^*), n \geq 1\]. From (i) in Definition A.4, there exists a sequence \((x_n)\) converging to \(x^*\) such that \(x_n \in C_n\) for every \(n \geq 1\). Then
\[\limsup_{n \to \infty} d(y, C_n) \leq \limsup_{n \to \infty} d(y, x_n) = \lim_{n \to \infty} d(y, x_n) = d(y, x^*) = d(y, C). \tag{A.2}\]

Now, let \(\vartheta_n := d(y, C_n)\) and let \((\vartheta_{n_j})\) be a subsequence of \((\vartheta_n)\) such that \(\liminf_{n \to \infty} \vartheta_n = \lim_{j \to \infty} \vartheta_{n_j}\). By compactness, \((x_{n_j}^*)\) converges, up to a subsequence, to an element \(x \in X\). From (ii) in Definition A.4, \(x\) belongs to \(C\), therefore
\[\liminf_{n \to \infty} d(y, C_n) = \lim_{j \to \infty} d(y, C_{n_j}) = \lim_{j \to \infty} d(y, x_{n_j}^*) = d(y, x) \geq d(y, C). \tag{A.3}\]

From (A.2) and (A.3) we obtain the result. \(\square\)

If \(X\) is a subset of a Hilbert space, then, for an integer \(k \geq 1\), we denote by \(\text{CC}_k(X)\) the collection of convex sets \(C\) in \(\text{CL}(X)\), such that \(\dim(C) \leq k\), where \(\dim(C)\) is the dimension of the smaller affine subspace containing \(C\). We denote also by \(\text{CC}_{k,0}(X)\) the collection of sets \(C\) in \(\text{CC}_k(X)\) such that \(0 \in C\).

Proposition A.2. If \(X\) is a compact subset of a Hilbert space, then \(\text{CL}(X)\), \(\text{CC}_k(X)\) and \(\text{CC}_{k,0}(X)\) are compact.
Theorem A.1. Let \( (X,d) \) be a compact metric space and \( (F_n) \) a sequence of functions from \( X \) into \( \mathbb{R} = \mathbb{R} \cup \{\pm \infty, -\infty\} \). We say that \( (F_n) \) \( \Gamma \)-converges to \( F : X \to \mathbb{R} \), denoted \( \Gamma \text{-lim}_{n \to \infty} F_n = F \), if for every \( x \in X \), it holds that

(i) for every sequence \( (x_n) \) converging to \( x \), \( F(x) \leq \liminf_{n \to \infty} F_n(x_n) \), and

(ii) there exists a sequence \( (x_n) \) converging to \( x \), such that \( F(x) = \lim_{n \to \infty} F_n(x_n) \).

Definition A.7. Let \( (X,d) \) be a compact metric space and \( (F_n) \) a sequence of functions from \( X \) into \( \mathbb{R} = \mathbb{R} \cup \{\pm \infty, -\infty\} \). We say that \( (F_n) \) \( \Gamma \)-converges to \( F : X \to \mathbb{R} \), denoted \( \Gamma \text{-lim}_{n \to \infty} F_n = F \), if for every \( x \in X \), it holds that

(i) for every sequence \( (x_n) \) converging to \( x \), \( F(x) \leq \liminf_{n \to \infty} F_n(x_n) \), and

(ii) there exists a sequence \( (x_n) \) converging to \( x \), such that \( F(x) = \lim_{n \to \infty} F_n(x_n) \).

Definition A.8. For \( F : X \to \mathbb{R} \), we denote \( M(F) = \{x \in X : F(x) = \inf_{y \in X} F(y)\} \).

The following result ([DM93], Theorems 7.8 and 7.23) shows that \( \Gamma \)-convergence together with compactness (or more generally equicoercivity) implies convergence of minimum values and minimizers.

Theorem A.1. Let \( (X,d) \) be a compact metric space and \( (F_n) \) a sequence of functions from \( X \) to \( \mathbb{R} \), \( \Gamma \)-converging to \( F : X \to \mathbb{R} \). Then \( M(F) \) is nonempty and

\[
\lim_{n \to \infty} \inf_{x \in X} F_n(x) = \min_{x \in X} F(x)
\]

Moreover, if \( x_n \in M(F_n), n \geq 1 \), then the accumulation points of \( (x_n) \) belong to \( M(F) \).
**Definition A.9.** Let \((X, d)\) be a metric space and \(A \subset X\). The indicator of \(A\) is the function \(\chi_A : X \to \mathbb{R} \cup \{+\infty\}\), with \(\chi_A(x) = 0\) if \(x \in A\), and \(\chi_A(x) = +\infty\), if \(x \notin A\).

The following Proposition ([Att84], Proposition 4.15.) shows the relation between K-convergence (see Definition A.4) and \(\Gamma\)-convergence.

**Lemma A.4.** Let \((X, d)\) be a metric space and \(A, A_n \subset X, n \geq 1\). Then \(\text{K-lim}_{n \to \infty} A_n = A\) if and only if \(\text{\Gamma-lim}_{n \to \infty} \chi_{A_n} = \chi_A\).

**References**

[AC11] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.

[AGS04] L. Ambrosio, N. Gigli, and G. Savaré. Gradient flows with metric and differentiable structures, and applications to the wasserstein space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 15(3-4):327–343, 2004.

[Att84] H. Attouch. *Variational Convergence for Functions and Operators*. Applicable Mathematic series. Pitman, London, 1984.

[AW95] Z. Artstein and R. J. B. Wets. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.*, 2(1-2):1–17, 1995.

[Bee85] G. Beer. On convergence of closed sets in a metric space and distance functions. *Bulletin of the Australian Mathematical Society*, 31:421–432, 6 1985.

[BK12] J. Bigot and T. Klein. Consistent estimation of a population barycenter in the Wasserstein space. *Preprint, ArXiv/1212.2562*, 2012.

[BP03] R. Bhattacharyya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of Statistics*, 31(1):1–29, 2003.

[Bre91] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[Cho11] O. Chodosh. Optimal transport and Ricci curvature: Wasserstein space over the interval. *Preprint, ArXiv/1105.2883*, 2011.

[DM93] G. Dal Maso. *An Introduction to \(\Gamma\)-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, Boston, MA, 1993.

[DPR82] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, 12(1):136–154, 1982.

[EH13] P. Embrechts and M. Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77:423–432, 2013.
[FLPJ04] P. T. Fletcher, C. Lu, Stephen M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

[GLM13] S. Gallón, J.-M. Loubes, and E. Maza. Statistical properties of the quantile normalization method for density curve alignment. *Mathematical biosciences*, 242(2):129–142, 2013.

[HHM10] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric lie group actions. *Statistica Sinica*, 20:1–100, 2010.

[HT13] N. N. Hai and An P. T. A generalization of Blaschke’s convergence theorem in metric spaces. *Journal of Convex Analysis*, to be published, 2013.

[Joh69] B. E. Johnson. Separate continuity and measurability. *Proceedings of the American Mathematical Society*, 20(2):420–422, 1969.

[KU01] A. Kneip and K. J. Utikal. Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.*, 96(454):519–542, 2001. With comments and a rejoinder by the authors.

[LT11] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[NM94] W. K. Newey and D. McFadden. Large Sample Estimation and Hypothesis Testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam, 1994.

[Pri40] B. Price. On the completeness of a certain metric space with an application to Blaschke’s selection theorem. *Bull. Amer. Math. Soc.*, 46:278–280, 1940.

[RS05] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.

[Rud81] W. Rudin. Lebesgue’s first theorem. In *Mathematical analysis and applications, Part B*, volume 7b of *Advances in Math. Suppl. Stud*, pages 741–747. Academic Press, 1981.

[Sil96] B. W. Silverman. Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, 24(1):1–24, 1996.

[SLHN10] S. Sommer, F. Lauze, S. Hauberg, and M. Nielsen. Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, volume 6316 of *Lecture Notes in Computer Science*, pages 43–56. Springer Berlin Heidelberg, 2010.

[Vil03] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.
[ZM11] Z. Zhang and H.-G. Müller. Functional density synchronization. *Comput. Stat. Data Anal.*, 55(7):2234–2249, 2011.