Statistical data depth for covariance and spectral density matrices

Joris Chau*
Université catholique de Louvain

Hernando Ombao
University of California at Irvine

Rainer von Sachs
Université catholique de Louvain

February 11, 2017

1. Introduction

In multivariate time series analysis, we study the covariance and correlation structures in a time series by means of its autocovariance matrices in the time domain, or by means of its spectral density matrix in the frequency domain. Non-degenerate covariance matrices or spectral density matrices are elements of the space of Hermitian positive-definite (PD) matrices. The space of Hermitian PD matrices, although very well-structured, is a non-Euclidean space, and to perform data analysis we need to generalize statistical procedures (e.g. regression, clustering, outlier detection, etc.) taking into account the inherent non-Euclidean geometry of this space. Recent works developing statistical tools for Hermitian PD matrix-valued data include [4], [5], [13], [2], [15], and [14] among others. In this work, we generalize the notion of a statistical data depth for data observations in the space of Hermitian PD matrices, with in mind the application to covariance or spectral density matrices. Data depth is an important tool in statistical data analysis providing a center to outward ordering of multivariate data observations. This allows one to characterize central points and central regions of the data, detect outlying observations, but also provides a framework for nonparametric rank-based hypothesis testing.

2. Preliminaries

2.1. Geometric tools

In order to construct a notion of data depth for observations in the space of Hermitian PD matrices, we exploit the geometric properties of this space as a Riemannian manifold, for a more detailed description we refer to [13] or [1, Chapter 6]. Let us write $\mathcal{M} := \mathbb{P}_{d\times d}$ for the space of $(d \times d)$-dimensional Hermitian PD matrices, which is a well-studied differentiable Riemannian manifold. The tangent space $\mathcal{T}_p(\mathcal{M})$ at a point, i.e. a matrix, $p \in \mathcal{M}$ can be identified by the real vector space $\mathcal{H} := \mathbb{H}_{d\times d}$ of $(d \times d)$-dimensional Hermitian matrices, and the inner product on $\mathcal{H}$ leads to the natural (also invariant, Fisher information, or Fisher-Rao) Riemannian metric on the manifold $\mathcal{M}$ given by the family of inner products:

$$\langle h_1, h_2\rangle_p = \mathrm{Tr}((p^{-1/2} * h_1)(p^{-1/2} * h_2))$$

with $h_1, h_2 \in \mathcal{H}$, using the notation $y * x := y^* h y$ for matrix congruence transformation, where * denotes the conjugate transpose of a matrix. Here and throughout the document, $p^{1/2}$ denotes the Hermitian square root matrix of $p \in \mathcal{M}$. The natural Riemannian distance on $\mathcal{M}$ derived from the Riemannian metric is given by:

$$\delta(p_1, p_2) = \|\mathrm{Log}(p_1^{-1/2} * p_2)\|_F \quad (2.1)$$

where $\mathrm{Log}(\cdot)$ denotes the matrix logarithm. By [1, Prop. 6.2.2], $(\mathcal{M}, \delta)$ is a complete metric space, which by the Hopf-Rinow Theorem implies that every geodesic curve can be extended indefinitely. The mapping $x \mapsto a * x$ for any $a \in \mathrm{GL}(d, \mathbb{C})$ is an isometry, i.e. it is distance-preserving with respect to the Riemannian distance function:

$$\delta(p_1, p_2) = \delta(a * p_1, a * p_2), \quad a \in \mathrm{GL}(d, \mathbb{C})$$

The exponential maps $\mathrm{Exp}_p : \mathcal{T}_p(\mathcal{M}) \simeq \mathcal{H} \to \mathcal{M}$ are diffeomorphic maps from the tangent space attached at a point $p \in \mathcal{M}$ given by,

$$\mathrm{Exp}_p(h) = p^{1/2} * \mathrm{Exp}\left(p^{-1/2} * h\right)$$

where $\mathrm{Exp}(\cdot)$ is the matrix exponential. In particular, $\gamma(t) = \mathrm{Exp}_p(th)$ is the geodesic emanating from

*Corresponding author, Ph.D. student (boursier FRIA) at the Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Belgium, j.chau@uclouvain.be

p, such that $\gamma(0) = p$, with unit tangent vector h. Since $\mathcal{M}$ is a geodesically complete manifold and the minimizing geodesics are always unique, it follows by [3, Chapter 13] that for each $p \in \mathcal{M}$ the image of the exponential map Exp_p is the entire manifold $\mathcal{M}$, (i.e. there is no cut-locus). This implies that the exponential maps are *global* diffeomorphisms. In the other direction, the logarithmic maps are global diffeomorphic maps from the manifold to the tangent space attached at a point $p \in \mathcal{M}$, and are defined as the unique inverse exponential maps $\mathrm{Log}_p(\tilde{p}) : \mathcal{M} \to \mathcal{T}_p(\mathcal{M})$ given by:

$$\mathrm{Log}_p(\tilde{p}) \quad = \quad p^{1/2} * \mathrm{Log}\left(p^{-1/2} * \tilde{p}\right)$$

The Riemannian distance function can now also be expressed in terms of the logarithmic map as:

$$\delta(p_1, p_2) \;=\; \|\mathrm{Log}_{p_1}(p_2)\|_{p_1} \;=\; \|\mathrm{Log}_{p_2}(p_1)\|_{p_2}$$

where $\|h\|_p := \langle h, h\rangle_p$ is the norm of $h \in \mathcal{T}_p(\mathcal{M})$ induced by the Riemannian metric.

Furthermore, since the Riemannian manifold $\mathcal{M}$ is geodesically complete and there exist unique geodesic curves connecting any two points $p_1, p_2 \in \mathcal{M}$, geodesically convex sets are well-defined, and we say that a subset $\mathcal{K} \subseteq \mathcal{M}$ is convex or geodesically convex if for each pair of points $p_1, p_2 \in \mathcal{K}$, the geodesic segment $[p_1, p_2]$ lies entirely in $\mathcal{K}$. If $\mathcal{S}$ is any subset of $\mathcal{M}$, then the convex hull of $\mathcal{S}$, denoted by $\mathrm{conv}(\mathcal{S})$, is the smallest convex set containing $\mathcal{S}$. For more details on the construction of (approximate) convex hulls on the manifold $\mathcal{M}$, we refer to [4].

2.2. Geometric mean and median

Let a random variable on the manifold $X : \Omega \to \mathcal{M}$ be a measurable function from some probability space $(\Omega, \mathcal{A}, \nu)$ to the measurable space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, where $\mathcal{B}(\mathcal{M})$ is the Borel algebra, i.e. the smallest σ-algebra containing all open sets in the metric space $(\mathcal{M}, \delta)$. Below, we work directly with the induced probability on $\mathcal{M}$, i.e. $\nu(B) = \nu(\{\omega \in \Omega : X(\omega) \in B\})$. To characterize the center of a random variable X with probability distribution ν, we first consider the geometric (also Karcher or Fréchet) mean. The geometric mean plays a key role in the manifold zonoid depth defined below, and is given by the point that minimizes the variance with respect to the Riemannian distance,

$$\begin{aligned}\mathbb{E}_\nu[X] \;&:=\; \arg\min_{y \in \mathrm{supp}(\nu)} \boldsymbol{E}_\nu[\delta(y, X)^2] \\ &=\; \arg\min_{y \in \mathrm{supp}(\nu)} \int_{\mathcal{M}} \delta(y, x)^2 \; d\nu(x)\end{aligned}$$

where we assume that the distribution has finite second moments with respect to the Riemannian distance, i.e. $\boldsymbol{E}_\nu[\delta(y, X)^2] < \infty$ for every $y \in \mathcal{M}$. In Euclidean space, the geometric expectation reduces to the ordinary Euclidean expectation, which is the point that minimizes the variance with respect to the Euclidean distance. By [12, Corollary 2], since the manifold $\mathcal{M}$ is a space of nonpositive curvature with no cut-locus, the geometric mean $\mu := \mathbb{E}_\nu[X]$ exists and is unique for any distribution with compact support. Recall that the cut-locus at a point $p \in \mathcal{M}$ is the complement of the image of the exponential map Exp_p, which is the empty set for each $p \in \mathcal{M}$ as the image of Exp_p is the entire manifold $\mathcal{M}$. Furthermore, for a distribution with compact support, the geometric mean is uniquely characterized by the point $\mu \in \mathcal{M}$ that satisfies,

$$\boldsymbol{E}_\nu[\mathrm{Log}_\mu(X)] \;=\; \boldsymbol{0}_{d\times d} \tag{2.2}$$

where $\boldsymbol{0}_{d\times d}$ is the zero matrix and $\boldsymbol{E}_\nu[\cdot]$ denotes the Euclidean expectation of a Hermitian matrix. The sample geometric mean of a set of manifold-valued observations minimizes a sum of squared Riemannian distances, and is computed efficiently by e.g. the gradient descent algorithm in [12].

As a second measure of centrality, we consider the geometric median of a random variable X with probability distribution ν, which plays an important role in the geodesic distance depth defined below. In contrast to the geometric mean, the geometric median is based on minimizing an expectation of non-squared distances, and is given by:

$$\begin{aligned}\mathrm{GM}_\nu(X) \;&:=\; \arg\min_{y \in \mathrm{supp}(\nu)} \boldsymbol{E}_\nu[\delta(y, X)] \\ &=\; \arg\min_{y \in \mathrm{supp}(\nu)} \int_{\mathcal{M}} \delta(y, x) \; d\nu(x) \qquad (2.3)\end{aligned}$$

where we assume that the distribution has finite first moments with respect to the Riemannian distance, i.e. $\boldsymbol{E}_\nu[\delta(y, X)] < \infty$ for every $y \in \mathcal{M}$. It can be shown that on $\mathcal{M}$, a space with nonpositive curvature and no cut-locus, the geometric median exists and is unique for any distribution with compact support. The proof runs along the same lines as that of [5, Theorem 1], combined with an application of Leibniz's integral rule. Furthermore, similar to the characterization of the geometric mean in eq.(2.2), for a distribution with compact support, the geometric median is uniquely characterized by the point $m \in \mathcal{M}$ that satisfies,

$$\boldsymbol{E}_\nu\left[\frac{\mathrm{Log}_m(X)}{\delta(m, X)}\right] \;=\; \boldsymbol{0}_{d\times d} \tag{2.4}$$

If the distribution ν of a random variable X has compact support and is *centrally symmetric* around $\mu \in \mathcal{M}$ in the sense that $\mathrm{Log}_\mu(X) \stackrel{d}{=} -\mathrm{Log}_\mu(X)$, then the unique geometric mean and median coincide and are equal to μ. For the geometric mean, this follows by observing that $\boldsymbol{E}_\nu[\mathrm{Log}_\mu(X)] = \mathbf{0}_{d\times d}$, which implies that μ is the unique geometric mean of the random variable X. For the geometric median, if X is centrally symmetric around μ, then X is also *angularly symmetric* around μ in the sense that $\mathrm{Log}_\mu(X)/\|\mathrm{Log}_\mu(X)\|_\mu \stackrel{d}{=} -\mathrm{Log}_\mu(X)/\|\mathrm{Log}_\mu(X)\|_\mu$. Since $\|\mathrm{Log}_\mu(X)\|_\mu = \delta(\mu, X)$ by definition of the Riemannian metric and distance function, it follows that $\boldsymbol{E}_\nu[\mathrm{Log}_\mu(X)/\delta(\mu, X)] = \mathbf{0}_{d\times d}$, implying that μ is also the unique geometric median of the random variable X with probability measure ν.

3. Data depth on the manifold

In order to construct a proper data depth function for data observations taking values in the space of $(d \times d)$-dimensional Hermitian positive definite matrices, we first present the properties the depth function should ideally satisfy. This is in the same spirit as [17], with regard to depth functions acting on multivariate observations in Euclidean space. In the following, we focus our attention on depth functions that are nonnegative and bounded, and distributions with compact support on the manifold.

P1. *(Congruence invariance)* For a data depth function in Euclidean space, it is desirable that it is affine invariant in the following sense: consider a random vector $X \in \mathbb{R}^d$ with distribution ν, such that $aX + b$ has distribution $\nu_{a,b}$, with $a \in \mathrm{GL}(d, \mathbb{R})$ and $b \in \mathbb{R}^d$, then we require that,

$$D(\nu, y) \quad = \quad D(\nu_{a,b}, ay + b), \quad \forall y \in \mathbb{R}^d$$

where $D(\nu, y)$ is the Euclidean depth of a vector $y \in \mathbb{R}^d$ with respect to the distribution ν. In the context of covariance or spectral density matrices, we are concerned with second-order behavior of random variables. For a random vector X with location μ and covariance matrix Σ, the covariance matrix of the affine transformation $aX + b$ is given by $a * \Sigma = a\Sigma a^T$. This means that for a depth function acting on a collection of spectral matrices or covariance matrices, the natural equivalent of the affine invariance property above is invariance of the depth function with regard to congruence transformations of the form $x \mapsto a * x$, where $a \in \mathrm{GL}(d, \mathbb{C})$. That is, for a proper depth function $D(\cdot, \cdot)$ on the manifold $\mathcal{M}$ we require that,

$$D(\nu, y) \quad = \quad D(\nu_a, a * y), \quad \forall y \in \mathcal{M}$$

where ν_a is the distribution of the transformed random variable $a * X$, where X has probability distribution ν on the manifold. Note that by Sylvester's law of inertia for Hermitian matrices, it holds that $a * x \in \mathcal{M}$ for each $x \in \mathcal{M}$ and $a \in \mathrm{GL}(d, \mathbb{C})$.

P2. *(Maximality at center)* The depth function should attain its maximum value, i.e. deepest point or point with maximum depth, at a well-defined unique *center* of the distribution, such as the geometric mean or median, which are characterized as points of central and angular symmetry respectively. Let $\mu \in \mathcal{M}$ be a unique central point of the distribution ν, then

$$D(\nu, \mu) \quad = \quad \sup_{y \in \mathcal{M}} D(\nu, y)$$

P3. *(Monoticity relative to center)* As the point $y \in \mathcal{M}$ moves away from the deepest point μ along a geodesic curve emanating from μ, the depth of the point y with respect to the distribution ν should be monotone decreasing. Let $\mathrm{Exp}_\mu(th)$ be the geodesic emanating from μ such that $\mathrm{Exp}_\mu(0) = \mu$ with unit tangent vector h, then we require that for each $0, \leq t_1 \leq t_2$,

$$D(\nu, \mathrm{Exp}_\mu(t_1 h)) \quad \geq \quad D(\nu, \mathrm{Exp}_\mu(t_2 h))$$

P4. *(Vanishing at infinity)* The depth of a point $y \in \mathcal{M}$ should approach zero as the point y converges to a singular matrix, i.e. a matrix with zero or infinite eigenvalues,

$$\lim_{M\to\infty} \sup_{\|\mathrm{Log}(y)\|_F \geq M} D(\nu, y) \quad = \quad 0$$

If a function $D : \mathcal{F} \times \mathcal{M} \to \mathbb{R}$ satisfies the above four properties, we say that it is a proper data depth function on the manifold $\mathcal{M}$. Here, $\mathcal{F}$ denotes the class of all distributions ν with compact support on the manifold.

4. Manifold zonoid depth

As convex hulls are well-defined on the manifold $\mathcal{M}$, there exist natural manifold generalizations of the notions of simplicial or convex hull peeling depth for multivariate observations in Euclidean space as detailed in e.g. [7]. However, the simplicial depth requires the computation of possibly many (approximate) convex hulls, which quickly becomes computationally expensive, especially in higher dimensions.

In this section, instead we focus on a different notion of data depth: the zonoid depth as described in e.g. [10], for which we propose a straightforward generalization to the manifold $\mathcal{M}$. The manifold zonoid depth can be computed by the same tools used to calculate the classical zonoid depth in Euclidean space and its computation remains efficient, also for higher-dimensional covariance or spectral density matrices.

First, we recall the definition of a zonoid α-trimmed region in Euclidean space according to [10]. Let ν be a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$ with finite first moment, then the zonoid α-trimmed region for $0 < \alpha \leq 1$ is defined as,

$$D_\alpha(\nu) \quad := \quad \left\{ \int_{\mathbb{R}^d} x g(x)\, d\nu(x) \ : \ g : \mathbb{R}^d \to \left[0, \frac{1}{\alpha}\right] \right. \\ \left. \text{measurable, s.t.} \int_{\mathbb{R}^d} g(x)\, d\nu(x) = 1 \right\}$$

If $\alpha = 0$, we set $D_0(\nu) = \mathbb{R}^d$. By [10, Chapter 3], $D_\alpha(\nu)$ is a convex set and monotone decreasing in α, thereby creating a nested sequence of convex sets for a decreasing sequence $\alpha_1 \geq \ldots \geq \alpha_n$. If $\alpha = 1$, $D_\alpha(\nu)$ consists of a single point $\boldsymbol{E}_\nu[X]$, the Euclidean mean of the distribution ν. For a discrete set of observations $x_1, \ldots, x_n \in \mathbb{R}^d$ with the empirical probability measure ν_n, the sample zonoid α-trimmed region becomes,

$$D_\alpha(\nu_n) \quad = \quad \left\{ \sum_{i=1}^n \lambda_i x_i \ : \ \sum_{i=1}^n \lambda_i = 1, \lambda_i \in \left[0, \frac{1}{n\alpha}\right] \right\}$$

This follows by writing $g(x_i)\nu(\{x_i\}) = \lambda_i$ in the population zonoid α-trimmed region above. If $0 < \alpha \leq \frac{1}{n}$, $D_\alpha(\nu_n)$ is the convex hull of the data points, and if $\alpha = 1$, $D_\alpha(\nu_n)$ consists of a single point $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$, the empirical Euclidean mean of the observations. The zonoid depth in Euclidean space of a point $y \in \mathbb{R}^d$ with respect to a distribution ν is characterized by the most central α-trimmed region still containing y as:

$$\text{ZD}_{\mathbb{R}^d}(\nu, y) \quad := \quad \sup\{\alpha : y \in D_\alpha(\nu)\}$$

The zonoid data depth is extended to the manifold as follows.

Definition 4.1. *(Manifold zonoid depth)* Let ν be a probability measure on the manifold $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ with $\boldsymbol{E}_\nu[\delta(y, X)^2] < \infty$ for each $y \in \mathcal{M}$. Let ζ_y be the probability measure on $(\mathbb{R}^{d^2}, \mathcal{B}(\mathbb{R}^{d^2}))$ of the random variable $\text{Log}_y(X) \in \mathcal{H} \cong \mathbb{R}^{d^2}$, where X has probability measure ν. The manifold zonoid depth of a point $y \in \mathcal{M}$ with respect to the distribution ν is defined as:

$$\text{ZD}_{\mathcal{M}}(\nu, y) \quad := \quad \sup\left\{\alpha : \vec{d} \in D_\alpha(\zeta_y)\right\}$$

where $\vec{0}$ is a d^2-dimensional vector of zeros, and $D_\alpha(\zeta_y)$ the zonoid α-trimmed region of the distribution ζ_y defined on $(\mathbb{R}^{d^2}, \mathcal{B}(\mathbb{R}^{d^2}))$.

Theorem 4.1. *The manifold zonoid depth is a proper data depth function in the sense of Section 3, satisfying properties **P1**–**P4** for distributions with compact support on the manifold and finite second moments (with respect to the Riemannian distance), in which case the unique point of maximum depth is characterized by the geometric mean of the distribution.*

Proof. **P1.** This follows from the fact that for each $0 \leq \alpha \leq 1$, $\{\mathbf{0}_{d\times d} \in D_\alpha(\zeta_y)\} \Leftrightarrow \{\mathbf{0}_{d\times d} \in \boldsymbol{D}_\alpha(\zeta_{a,y})\}$, where $\zeta_{a,y}$ is the distribution of $\text{Log}_{a*y}(a * X)$. This is shown by using that $\text{Log}_{a*y}(a * X) \overset{d}{=} a * \text{Log}_y(X)$ for every $a \in \text{GL}(d, \mathbb{C})$.
P2. The zonoid trimmed region $D_1(\zeta_y)$ contains the single point $\boldsymbol{E}_\nu[\text{Log}_y(X)]$. The deepest point $y \in \mathcal{M}$ is therefore characterized by the point that satisfies $\boldsymbol{E}_\nu[\text{Log}_y(X)] = \mathbf{0}_{d\times d}$. By eq.(2.2), for a distribution with compact support, this point is the uniquely existing geometric expectation of the distribution ν.
P3. For $\alpha \in [0, 1]$, we can rewrite $\text{ZD}_{\mathcal{M}}(\nu, y) = \sup\{\alpha : y \in D_\alpha^{\mathcal{M}}(\nu)\}$, where,

$$D_\alpha^{\mathcal{M}}(\nu) \quad = \quad \left\{ y : y = \text{Exp}_y\left(\int_{\mathcal{M}} \text{Log}_y(x) g(x)\, d\nu(x) \right), \right. \\ \left. g : \mathcal{M} \to [0, 1/\alpha], \int_{\mathcal{M}} g(x)\, d\nu(x) = 1 \right\}$$

with g measurable. $D_\alpha^{\mathcal{M}}(\nu)$ is geodesically convex and contains the geometric mean $\mu := \mathbb{E}_\nu[X]$ for each $\alpha \in [0, 1]$, therefore $D_\alpha^{\mathcal{M}}(\nu)$ is geodesically starshaped about μ. Also, $D_{\alpha_1}^{\mathcal{M}}(\nu) \subseteq D_{\alpha_2}^{\mathcal{M}}(\nu)$ for each $1 \geq \alpha_1 \geq \alpha_2 \geq 0$. Combining the above arguments, it follows that a geodesic curve $\text{Exp}_\mu(th)$, with $t \geq 0$ increasing, has monotone decreasing depth as it moves further away from the center μ.
P4. With the same notation as above, for $\alpha \in (0, 1]$ the sets $D_\alpha^{\mathcal{M}}(\nu)$ are closed and bounded, and therefore also compact by the Hopf-Rinow theorem. Boundedness follows from the observation that,

$$\left\| \boldsymbol{E}_\nu[\text{Log}_y(X) g_\alpha(X)] \right\|_y \quad \leq \quad \frac{1}{\alpha} \boldsymbol{E}_\nu\left[\|\text{Log}_y(x)\|_y \right] \\ = \quad \frac{1}{\alpha} \boldsymbol{E}_\nu[\delta(y, X)] \ < \ \infty$$

using that ν has finite first moments with respect to the Riemannian distance function. In particular, if $(y_n)_{n\in\mathbb{N}} \in \mathcal{M}$ is an unbounded sequence, such that $\|\mathrm{Log}(y_n)\|_F \to \infty$ as $n \to \infty$, then $\mathrm{ZD}_{\mathcal{M}}(\nu, y_n) \to 0$, otherwise the boundedness (or compactness) of $D_\alpha^{\mathcal{M}}(\nu)$ for $\alpha > 0$ would be violated. □

In the context of spectral density matrices, which are curves in the space of Hermitian PD matrices along frequency, we wish to generalize the concept of manifold zonoid depth of a point on the manifold to the depth of an entire curve $y(t) \in \mathcal{M}$ with respect to a collection of marginal measures $\nu(t)$ for $t \in \mathcal{I}$, where $\mathcal{I} \subset \mathbb{R}$. To this purpose, we consider the functional manifold zonoid depth given by,

$$\mathrm{fZD}_{\mathcal{M}}(\nu, y) = \int_{\mathcal{I}} \sup\{\alpha : \mathbf{0}_{d\times d} \in D_\alpha(\zeta_y(t))\}\ dt$$

where $\zeta_y(t)$ is the probability measure of the random variable $\mathrm{Log}_{y(t)}(X(t)) \in \mathcal{H} \cong \mathbb{R}^{d^2}$, where $X(t)$ has probability measure $\nu(t)$. Note that this is essentially analogous to the construction of modified band depth (MBD) for a curve in Euclidean space as an integrated version of the pointwise simplicial depths $y(t)$ integrated over the measurement domain $t \in \mathcal{I}$. For more details on functional data depth in Euclidean space we refer to e.g. [9] or [16].

5. Geodesic distance depth

As a second notion of data depth on the manifold, we consider the *geodesic distance depth*, which is the natural analog on the manifold of the arc distance depth detailed in [8] with regard to data observations on circles and spheres. The geodesic distance depth is straightforward to calculate as it requires only the computation of the Riemannian distance between points on the manifold, which remains computationally efficient also for high-dimensional covariance or spectral matrices.

Definition 5.1. *(Geodesic distance depth)* Let ν be a probability measure on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ with $\boldsymbol{E}_\nu[\delta(y, X)] < \infty$ for all $y \in \mathcal{M}$. The geodesic distance depth of a point $y \in \mathcal{M}$ with respect to the distribution $\nu \in \mathcal{M}$ is defined as:

$$\mathrm{GDD}(\nu, y) = \exp\left(-\int_{\mathcal{M}} \delta(y, x)\ d\nu(x)\right)$$

Theorem 5.1. *The geodesic distance depth is a proper data depth function in the sense of Section 3, satisfying **P1**–**P4** for distributions with compact support on the manifold and finite first moments (with respect to the Riemannian distance), in which case the unique point of maximum depth is characterized by the geometric median of the distribution.*

Proof. **P1.** This follows directly by the fact that the map $x \mapsto a * x$ with $a \in \mathrm{GL}(d, \mathbb{C})$ is distance preserving, i.e. $\delta(a * x, a * y) = \delta(x, y)$ for each $x, y \in \mathcal{M}$.
P2. Since $\int_{\mathcal{M}} \delta(y, x)\ d\nu(x) \geq 0$ and $\exp(-z)$ is strictly decreasing in $z \geq 0$, the point of maximum depth is attained at $y = \arg\min_{z\in\mathcal{M}} \int_{\mathcal{M}} \delta(z, x)\ d\nu(x)$. By eq.(2.3), for a distribution with compact support, this point is the uniquely existing geometric median of the distribution ν.
P3. By the proof of [5, Theorem 1] and an application of Leibniz's integral rule, $y \mapsto \boldsymbol{E}_\nu[\delta(y, X)]$ is a (strictly) convex function, and by **P2** it attains its unique minimum at $m := \mathrm{GM}\nu(X)$. This implies that $\boldsymbol{E}_\nu[\delta(\mathrm{Exp}_m(th), X)]$ is a nondecreasing function for $t \geq 0$, where $\mathrm{Exp}_m(th)$ is a geodesic curve emanating from m with unit tangent vector h. As a consequence $\mathrm{GDD}(\nu, \mathrm{Exp}_m(th)) = \exp\left(-\boldsymbol{E}_\nu[\delta(\mathrm{Exp}_m(th), X)]\right)$ is nonincreasing for $t \geq 0$.
P4. Let $(y_n)_{n\in\mathbb{N}}$ be an unbounded sequence such that $\|\mathrm{Log}(y_n)\|_F \to \infty$ as $n \to \infty$, then also $\delta(x, y_n) = \|\mathrm{Log}(x^{-1/2} * y_n)\|_F \to \infty$ for each $x \in \mathcal{M}$, and as a consequence $\mathrm{GDD}(\nu, y_n) = \exp(-\boldsymbol{E}_\nu[\delta(y_n, X)]) \to 0$. □

In order to generalize the geodesic distance depth of a point y, with respect to a single measure ν, to the depth of an entire curve $y(t) \in \mathcal{M}$, with respect to a collection of marginal measures $\nu(t) = \nu_t$, $t \in \mathcal{I} \subset \mathbb{R}$, we consider the integrated Riemannian distance between curves instead of the Riemannian distance between points. In this way, the functional geodesic distance depth is given by:

$$\mathrm{fGDD}(\nu, y) = \exp\left(-\int_{\mathcal{I}}\int_{\mathcal{M}} \delta(y(t), x)\ d\nu_t(x)\ dt\right)$$

Remark We point out that integrated distance measures between spectral density curves of Hermitian PD matrices can also found in e.g. [6] in the context of clustering of spectral density matrices. Here, the considered integrated disparity measures between spectral density matrices are of the form:

$$D_H(f, g) = \int_0^\pi H(f(\omega)g(\omega)^{-1})\ d\omega$$

The integrated Riemannian distance function is seen to be a specific case of such integrated disparity measures, since we can write,

$$\begin{aligned} H(f(\omega)g(\omega)^{-1}) &= \|\mathrm{Log}(f(\omega)g(\omega)^{-1})\|_F \\ &= \delta(f(\omega), g(\omega)) \end{aligned}$$

by definition of the Riemannian distance function. In [6], the considered disparity measures include symmetric J-divergence and symmetric Chernoff-information divergence. It is possible to replace the Riemannian distance in the data depth function by the disparity measures considered in [6], but properties **P1–P4** will no longer be valid, as they are specifically tailored to the metric space $(\mathcal{M}, \delta)$. Moreover, the symmetric J-divergence and symmetric Chernoff-information divergence are only *quasi*-distance functions in the cone of Hermitian PD matrices embedded in Euclidean space, as the triangle inequality does not hold, whereas the Riemannian distance is a proper distance function on the manifold.

6. Conclusion

In this work we develop the concept of data depth for observations in the space of symmetric or Hermitian PD matrices. Hermitian PD matrices are encountered as covariance matrices or spectral density matrices in multivariate time series analysis, but also play an important role in the fields of medical imaging, computer vision, or radar signal processing as discussed in e.g. [13], [14], [4], or [5]. The sample data depth values are straightforward to compute and remain computationally efficient also for relatively high-dimensional matrices. As such, the data depth functions serve as an exploratory data analysis tool in order to characterize central regions of the data or to detect outlying observations. A strong visualization tool that can be used in this context is the *functional boxplot*, as described in e.g. [16] or [11]. The data depth functions also provide a practical framework to perform nonparametric rank-based hypothesis testing for observations taking values in the space of Hermitian PD matrices, replacing the usual ranks by the ranks based on the computed data depths.

References

[1] R. Bhatia. *Positive Definite Matrices.* Princeton university press, 2009.

[2] J. Chau and R. von Sachs. Positive-definite multivariate spectral estimation: a geometric wavelet approach. *ArXiv preprint 1701.03314*, 2017.

[3] M.P. do Carmo. *Riemannian Geometry.* Birkhäuser, Boston, 1992.

[4] P.T. Fletcher, J. Moeller, J.M. Phillips, and S. Venkatasubramanian. Horoball hulls and extents in positive definite space. In *Workshop on Algorithms and Data Structures*, pages 386–398, 2011.

[5] P.T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1):S143–S152, 2009.

[6] Y. Kakizawa, R.H. Shumway, and M. Taniguchi. Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340, 1998.

[7] R.Y. Liu, J.M. Parelius, K. Singh, et al. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.

[8] R.Y. Liu and K. Singh. Ordering directional data: concepts of data depth on circles and spheres. *The Annals of Statistics*, pages 1468–1484, 1992.

[9] S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734, 2009.

[10] K. Mosler. *Multivariate Dispersion, Central Regions, and Depth: The Lift Zonoid Approach.* Springer Science & Business Media, New York, 2002.

[11] D. Ngo, Y. Sun, M.G. Genton, J. Wu, R. Srinivasan, S.C. Cramer, and H. Ombao. An exploratory data analysis of electroencephalograms using the functional boxplots approach. *Frontiers in Neuroscience*, 9:282, 2015.

[12] X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[13] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[14] S. Said, L. Bombrun, Y. Berthoumieu, and J. Manton. Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *ArXiv preprint 1507.01760*, 2015.

[15] S.T. Smith. Intrinsic Cramér-Rao bounds and subspace estimation accuracy. In *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, pages 489–493. IEEE, 2000.

[16] Y. Sun and M.G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2012.

[17] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000.