

# On Means and Their Asymptotics: Circles and Shape Spaces

Stephan Huckemann\*and Thomas Hotz†

## Abstract

We survey some effects that singular strata may have in the positive curvature context of circles and shape spaces when conducting (semi-)intrinsic statistical analyses. Here, the analysis of data on a stratified space is based on statistical descriptors defined in a possibly different stratified space. E.g. in geodesic principal component analysis for shape spaces, shape data are described by generalized geodesics which naturally form a shape space of their own, different from the original one. In a general context, if the descriptors are obtained as generalized Fréchet means, under rather general circumstances, a strong law of large numbers is valid. If furthermore the descriptors are sufficiently well behaved, a classical central limit theorem can be adopted. One of the crucial conditions is that hitting of singular strata as well as of cut loci, if present, must be controlled. We review the statistical role of the cut locus of intrinsic means for circles as well as that of singular strata for shape spaces (occurring where the group action is degenerate) and conclude with an identification of potential research directions.

*Key words and phrases:* stratified spaces, Fréchet mean, statistical analysis, geodesic principal components, strong law, central limit theorem, manifold stability, cut locus, convergence rate, asymptotic normality

*AMS 2000 Subject Classification:* Primary 62H11  
Secondary 60F05

## 1 Introduction

A large branch of statistics is concerned with data spaces that have a linear structure i.e. with the crucial property that the sum of data elements is again in the data space. This allows for the concept of a unique data mean being the data's linear superimposition and principal components being eigenvectors of covariance matrices, say. In this context, a beautiful interplay between linear algebra and the Gaussian distribution allowed to derive powerful asymptotic theorems.

In the last century an increased demand arose for the statistical analysis of data with an inherent non-linear structure. Such structures naturally occur if the data are subject to non-linear constraining conditions and/or if data objects carry geometrical information.

A basic example is that of wind directions living on the circle  $S^1 := \{(\cos \theta, \sin \theta) : 0 \leq \theta \leq 2\pi\}$ . Obviously, two directions cannot simply be added. Hence the elementary statistical concept of a mean is not readily at hand, let alone that of principal components.

More elaborate is the example of the statistical analysis of shape based on landmarks. Here  $m$ -dimensional objects are labeled by placing  $k \in \mathbb{N}$  landmark columns  $x_1, \dots, x_k \in \mathbb{R}^m$  on them, i.e. each object is described by its *landmark matrix*

$$X = (x_1, \dots, x_k) \in M(m, k), \text{ the space of } m \times k \text{ matrices.}$$

Two such objects are considered identical if their landmark matrices  $X$  and  $Y$  are related by a similarity transform

$$Y = \lambda g(X + a\mathbf{1}^T)$$

---

\*supported by DFG HU 1575/2-1 and the Niedersachsen Vorab of the Volkswagen Foundation

†supported by DFG CRC 803

conveyed by a displacement vector  $a \in \mathbb{R}^m$  ( $\mathbf{1}^T$  is a  $k$ -dimensional row vector of ones), a rotation  $g \in SO(m)$  and a scaling factor  $\lambda > 0$ . The equivalence class of  $X$  is denoted by  $[X]$  and the challenge of statistical shape analysis consists in devising a toolkit similar to that of linear statistics for a statistical analysis of objects in a shape space  $\{[X] : X \in M(m, k)\}$ .

Factoring out translations only, retains the linear structure of the data space, e.g. by reducing the number of landmarks by one via Helmertizing (cf. Dryden and Mardia (1998)). Factoring out scaling only, can be achieved by taking away the singular matrix  $X = 0$  and subsequently mapping to a unit sphere  $X \rightarrow X/\|X\|$  (here  $\|\cdot\|$  denotes the usual Euclidean norm). This comes at the cost of losing the linear structure. Factoring out rotations, however, is more complicated. In fact, beginning with *Procrustes analysis*, a term coined by Gower (1975), it has taken the statistical community a while to fully realize the statistical implications of the formal series of mappings envisioned by Kendall (1977) from the *configuration space*  $M^*(m, k) := M(m, k) \setminus \{(a, \dots, a) \in M(m, k) : a \in \mathbb{R}^m\}$  over the *pre-shape sphere*  $S_m^k$  (the hyper unit sphere in  $M(m, k-1)$ ) to *Kendall's shape space*  $\Sigma_m^k$ :

$$M^*(m, k) \rightarrow M(m, k-1) \setminus \{0\} \rightarrow S_m^k \rightarrow S_m^k/SO(m) =: \Sigma_m^k. \quad (1)$$

In case of  $m \geq 3$ , it turns out that the shape space is a non-manifold union of manifold strata of varying dimensions (cf. Kendall et al. (1999)). In particular it is not a symmetric space. At this point we note that a statistically meaningful map  $M^*(m, k) \rightarrow M(m, k-1)$  is given by an isometry from the space orthogonal to  $\{(x_1, \dots, x_k) \in M(m, k) : x_1 + \dots + x_k = 0\}$  in  $M(m, k)$ , e.g. by Helmertizing, cf. Kendall (1984).

The abstract concept of a general *shape space* is that of a canonical topological quotient  $Q = M/G$  of a complete finite-dimensional Riemannian manifold  $M$  on which a Lie group  $G$  acts isometrically thereby conveying shape equivalence. If the action is proper (which essentially means that  $Q$  is Hausdorff), the quotient carries a natural stratified topology

$$Q = Q^* \cup Q_0 \quad (2)$$

where  $Q^*$  is a manifold that is open and dense in  $Q$ , called the *manifold part*; the disjoint *singular part*  $Q_0$  – if non-void – is again stratified, i.e. decomposing into an open and dense manifold part and a disjoint stratified singular part, and so on, cf. Bredon (1972). The curvature of  $Q$  is *positive with respect to  $M$*  in the sense that in consequence of the curvature formula for a submersion by O'Neill (1966), the sectional curvatures of  $Q^*$  are equal or higher than those of  $M$  above. Additionally to the singularity posed by  $Q_0$ , potential singularities due to non-void cut loci in positive sectional curvature contexts enter the play. The situation of “non-positive curvature contexts” is briefly discussed in the discussion in Section 6. In this discussion we identify open research issues which arise from the work surveyed here.

In the (statistical) analysis of geometrical objects many other shape spaces are of high interest. Building on landmark configurations, considering the action of the group of Euclidean motions only, leads to *size and shape spaces* (e.g. Dryden and Mardia (1998)), considering larger groups leads to *affine and projective shape spaces*. For the latter two – as with similarity shape, ensuring a Hausdorff shape space – the group action can only be considered on a proper subset of the configuration space, cf. Mardia and Patrangenaru (2001, 2005). A statistically ingenuous setup as in (1) is to date only known for quadrilateral one-dimensional projective shapes, cf. Kent and Mardia (2012).

The theory of non-landmark based shape spaces is even more challenging as the underlying manifolds are usually infinite-dimensional and statistical subtleties are often only rudimentary explored to date (here are a few and in no way exhaustive references to a fast growing field: Zahn and Roskies (1972), Klassen et al. (2004), Cremers (2006), Michor and Mumford (2006), Mio et al. (2007), Younes (2010), Srivastava et al. (2011)).

The purpose of this paper is to give a survey over some relevant results in this context concerning means and their asymptotics. We begin with various concepts of means on shape spaces (Section 2), generalize these to descriptors taking values in spaces possibly different from the

data space (Section 4), give examples how such generalized Fréchet means can be used in (semi)-intrinsic statistical analysis (Section 3), review some asymptotic theory for such means (Section 4) and finally (Section 5) scrutinize three of the central hypothesis (uniqueness, a locally Euclidean neighborhood and a.s. twice continuous differentiability of the squared distance function) guaranteeing the validity of the  $\sqrt{n}$ -Central-Limit-Theorem 4.9. We conclude with an outlook to future research issues.

For the rest of this paper suppose that  $X, X_1, X_2, \dots$  are i.i.d. random elements on  $Q$ , i.e. mappings from an abstract probability space  $(\Omega, \mathcal{A}, \mathcal{P})$  to  $Q$  that are measurable w.r.t. the Borel  $\sigma$ -algebra of  $Q$ .

## 2 A Multitude of Shape Means

During the process of realizing the importance of the geometrical structure of non-linear data spaces in view of statistical analysis, several extensions of the concept of the linear mean have been introduced. Among others there are *centers of gravity* by Kobayashi and Nomizu (1969), *Fréchet means* by Ziezold (1977), *Procrustes means* by Gower (1975), *residual means* by Jupp (1988), *Karcher means* (cf. Karcher (1977)) by Kendall (1990), *perturbation means* by Goodall (1991) and *mean locations* by Hendriks and Landsman (1996). In fact, it turns out that most of these extensions fall into the classes of *intrinsic* and *extrinsic* means (coined by Bhattacharya and Patrangenaru (2003)) as well as residual means. The latter two can be generalized in a quotient context to *Procrustes* and *Ziezold means* (cf. Ziezold (1994)). These concepts will be further generalized by Definition 4.1.

For now, with every metric  $\rho$  on the data space come the sets of population and sample *Fréchet  $\rho$ -means*, respectively:

$$E^{(\rho)} = \operatorname{argmin}_{\mu \in Q} \mathbb{E}[\rho(X, \mu)^2], \quad E_n^{(\rho)} = \operatorname{argmin}_{\mu \in Q} \sum_{j=1}^n \rho(X_j, \mu)^2. \quad (3)$$

In particular with every Riemannian structure on the manifold part  $Q^*$  inducing a unique metric on  $Q$ , called an *intrinsic metric*, comes the concept of *intrinsic means* on  $Q$ . With every embedding of  $Q$  into a suitable Euclidean space  $\mathbb{R}^D$  ( $D \in \mathbb{N}$ ) the restriction of the Euclidean metric to the embedded space, called an *extrinsic metric*, comes the concepts of *extrinsic means*. Given such an embedding, for  $x \in \mathbb{R}^D$  consider  $\operatorname{argmin}_{q \in Q} \|x - q\|$  with the Euclidean norm  $\|\cdot\|$  of  $\mathbb{R}^D$ .

If  $Q$  is a closed subset of  $\mathbb{R}^D$ , then in consequence of Federer's Theorem (e.g. Stachó (1979)) there is a set of Lebesgue measure zero outside of which the *orthogonal projection*  $\Phi : x \mapsto \operatorname{argmin}_{q \in Q} \|x - q\|$  maps to a well defined unique point on  $Q$ . If  $Q$  is a smooth manifold its differential  $d\Phi_x$  is well defined for all  $x \in Q$ , in fact, it is the orthogonal projection to the tangent space of  $Q$  embedded in the tangent space of  $\mathbb{R}^D$ . Then, with the generalized *residual distance*

$$\rho(x, y) := \|d\Phi_x(x - y)\|, \quad x, y \in Q,$$

*residual means* are obtained when generalizing (3) to non-metrics.

Finally, in case of a shape space  $Q = M/G$  with a proper Lie group action on a smooth closed manifold  $M$  embedded in a Euclidean space  $\mathbb{R}^D$  with norm  $\|\cdot\|$  and a.e. well defined orthogonal projection  $\Phi$  onto  $M$ , if the action of  $G$  is also isometric w.r.t. the extrinsic metric of  $M$ , one may take a "quotient" of extrinsic or residual generalized distances on  $M$  to obtain the generalized *Ziezold distance*  $\rho_Z$  and *Procrustes distance*  $\rho_P$ , respectively, for  $x, y \in Q$  and any  $\tilde{x} \in [\tilde{x}] = x$ ,  $\tilde{y} \in [\tilde{y}] = y$ :

$$\begin{aligned} \rho_Z(x, y) &:= \inf_{g \in G} \|\tilde{x} - g\tilde{y}\|, \\ \rho_P(x, y) &:= \inf_{g, h \in G} \|d\Phi_{h\tilde{x}}(h\tilde{x} - g\tilde{y})\|, \end{aligned}$$

yielding with (3) the concepts of *Ziezold means* and *Procrustes means* on  $Q$ , respectively. We note that due to the Hausdorff structure of  $Q$ , the infima above are actually minima.

Additionally note that Ziezold and Procrustes means can also be defined with respect to a discrete group action.

### 3 (Semi-)Intrinsic Statistical Analysis

In the literature, the non-linear means of the preceding section have often been used to apply methods of linear statistics to data in a neighborhood of such a non-linear mean mapped into a Euclidean space. Typically, principal component analysis (PCA) is employed such as *generalized Procrustes analysis*, e.g. Dryden and Mardia (1998), or *principal geodesic analysis* by Fletcher et al. (2004). Obviously, this approach destroys the curvature information in the data. In order to preserve this information one would define non-Euclidean analogs of principal components, e.g. geodesics and perform the statistical analysis in the space of geodesics considered as point sets. In fact, for Kendall's shape space  $\Sigma_m^k$ ,  $m, k \in \mathbb{N}$ ,  $m < k$ , the space of its geodesics is again a shape space  $P$  which has thus a stratified structure  $P = P^* \cup P_0$  as in (2), cf. (Huckemann et al., 2010, Theorem 5.3).

#### 3.1 Geodesic Principal Component Analysis

In general, geodesics on a shape space  $Q = M/G$  are defined as canonical projections of *horizontal geodesics* on  $M$ ; here, a geodesic curve  $I \rightarrow M, t \mapsto \alpha(t)$  on a suitable interval  $I \subset \mathbb{R}$  is horizontal if its velocity  $\dot{\alpha}(t) \in H_{\alpha(t)}M$  for all  $t \in I$  with the orthogonal tangent space decomposition  $T_{\alpha(t)}M = T_{\alpha(t)}[\alpha(t)] \oplus H_{\alpha(t)}M$ . The quotient structure of the space of geodesics is obtained by identifying all maximal geodesic curves giving the same point set. For many spaces, such as Euclidean spaces, spheres and Kendall's shape spaces, this leads to Hausdorff quotients. On tori, however, most geodesics are dense, giving a statistically not meaningful non-Hausdorff quotient. In the following we consider only quotients  $Q$  with spaces of geodesics  $P$  that are Hausdorff in the canonical topology.

Linking a datum  $q \in Q$  to a geodesic  $\gamma \in P$  which is also a subset of  $Q$  via the geodesic distance  $d$  in  $Q$ ,

$$\rho(q, \gamma) := \inf_{q' \in \gamma} d(q, q'),$$

a Fréchet  $\rho$ -mean in  $P$  in the spirit of (3) – this will be made precise in Definition 4.1 – is called a *first geodesic principal component* (GPC). Given a unique first GPC  $\gamma_1$ , and noting that by horizontally lifting to  $M$ , orthogonality can be extended to  $P$ , define

$$P_1 := \{\gamma \in P : \gamma_1 \cap \gamma \neq \emptyset \text{ in } Q, \gamma_1 \perp \gamma \text{ at all intersection points}\}.$$

Then every Fréchet  $\rho$ -mean on  $P_1$  is called a *second geodesic principal component*. Similarly, higher order GPCs are defined, cf. Huckemann et al. (2010).

A similar approach can be taken for the shape spaces of closed planar curves introduced by Zahn and Roskies (1972) and given a quotient structure by Klassen et al. (2004). Among others, however, difficulties may arise because singularities have to be removed from the corresponding top space rendering it no longer complete.

#### 3.2 The General Approach and More Examples

In general, what we will call *(semi-)intrinsic* data analysis considers data on a stratified space  $Q$ , links these data via a *linking distance*  $\rho$  to descriptors on another stratified space  $P$  and conducts asymptotic analysis of the descriptor on  $P$ , thereby retaining in particular all curvature information. For geodesics, such an approach has been taken by Huckemann (2011b). Other descriptors allowing for (semi)-intrinsic data analysis can be found in Jung et al. (2012, 2011) where data descriptors of spherical data are constant intrinsic curvature small spheres or by Schulz et al. (2012) where descriptors are unions of concentric small circles. Each of these descriptor spaces allows for statistically meaningful distances and linking functions to the data spaces, respectively.

Indeed, this kind of analysis is “intrinsic” because only intrinsic data descriptors are used. It is semi-intrinsic, however, if for the descriptors’ asymptotics extrinsic or Ziezold means are employed as these are often computationally simpler and in view of the asymptotics of the related extrinsic means in Corollary 5.2 they typically are asymptotically normal as opposed to intrinsic means with their possibly non-Gaussian asymptotics in Theorem 5.12.

## 4 Generalized Fréchet means and their Asymptotics

In the abstract setting we have two topological spaces:

$Q$  is the *data space* and  $P$  is the *descriptor space*

where  $P$  is equipped with a continuous mapping  $d : P \times P \rightarrow [0, \infty)$  vanishing on the diagonal  $\{(p, p) : p \in P\}$ . Further, we assume that data and descriptors are linked via a continuous function

$$\rho : Q \times P \rightarrow [0, \infty)$$

which takes the role of a distance between a datum and a descriptor.

Recall that  $X, X_1, X_2, \dots$  are i.i.d. random elements on  $Q$ , i.e. mappings from an abstract probability space  $(\Omega, \mathcal{A}, \mathcal{P})$  to  $Q$  that are measurable w.r.t. the Borel  $\sigma$ -algebra of  $Q$ .

**Definition 4.1.** Define the set of population Fréchet  $\rho$ -means of  $X$  on  $P$  by

$$E^{(\rho)}(X) = \operatorname{argmin}_{\mu \in P} \mathbb{E}(\rho(X, \mu)^2).$$

For  $\omega \in \Omega$  denote by

$$E_n^{(\rho)}(\omega) = \operatorname{argmin}_{\mu \in P} \sum_{j=1}^n \rho(X_j(\omega), \mu)^2$$

the set of sample Fréchet  $\rho$ -means on  $P$ . In case of  $Q = P$  and  $\rho$  being a quasi-metric (different points are allowed to have zero distance), we speak of quasi-metrical means, and if  $\rho$  is additionally a metric we speak of metrical means.

By continuity of  $\rho$ , the mean sets are random closed sets. For our purpose here, we rely on the definition of *random closed sets* as introduced and studied by Choquet (1954), Kendall (1974) and Matheron (1975). Metrical means have been first introduced by Fréchet (1948) and generalized to quasi-metrical means by Ziezold (1977).

For an overview over the well-developed asymptotic theory of random closed sets in Banach spaces cf. Molchanov (2005). As it seems, a more general asymptotic theory for random closed sets in non-linear spaces, where there is no longer a (linear) concept of sums of sets, which concerns here, has received less attention in the literature.

**Definition 4.2.** Let  $E_n^{(\rho)}(\omega)$ ,  $n \in \mathbb{N}$ , be a sequence of random closed sets and  $E^{(\rho)}$  a fixed, deterministic closed set in  $P$ . Then we say that

(Z-SC)  $E_n^{(\rho)}(\omega)$  is a Ziezold strongly consistent estimator of  $E^{(\rho)}$  if a.s. for  $\omega \in \Omega$

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \subset E^{(\rho)},$$

(BP-SC)  $E_n^{(\rho)}(\omega)$  is a Bhattacharya-Patrangeanu strongly consistent estimator of  $E^{(\rho)}$  if  $E^{(\rho)} \neq \emptyset$  and if for every  $\epsilon > 0$  and for a.e.  $\omega \in \Omega$  there is a number  $n = n(\epsilon, \omega) > 0$  such that

$$\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega) \subset \{p \in P : d(E^{(\rho)}, p) \leq \epsilon\}.$$

In linear spaces, usually convergence w.r.t. the Hausdorff distance (which is a metric on the space of compact sets) is considered, cf. Molchanov (2005). We note that already in linear spaces for random sets in general the various notions may be incommensurable. If curvature is involved, there may for instance be convergence for the set of intrinsic means in the above sense but no longer convergence w.r.t. the Hausdorff distance. As an example, recall the uniform distribution on a sphere, the set of its intrinsic means being that entire sphere. Every sample mean, however, is atomic with probability one (for the detailed construction cf. Bhattacharya and Patrangenaru (2003, Remark 2.6.)).

Ziezold (1977) introduced (Z-SC) and proved it for quasi-metrical means on separable (i.e. containing a dense countable subset) spaces. Bhattacharya and Patrangenaru (2003) introduced (BP-SC) and proved it for metrical means on spaces that enjoy the stronger Heine-Borel property. For our purpose this translates to the  $d$ -Heine-Borel property, i.e. that every  $d$ -bounded ( $A$  is  $d$ -bounded if there is a point  $p \in A$  such that  $d(p, p_n)$  is bounded for every sequence  $p_n \in A$ ) closed set is compact. Both properties, (Z-SC) and (BP-SC) have been called ‘strong consistency’ by their respective authors. While the Heine-Borel property is essential for property (PB-SC) to hold, the condition that  $\rho$  be metric or a quasi-metric, respectively, can be relaxed. It suffices that  $\rho$  satisfies the following two properties of *continuity* in the second component *uniform* over the first component – a consequence of the triangle inequality if  $P = Q$  and  $\rho$  is a quasi-metric – and of a version of *coercivity* in the second component – again valid if  $\rho$  is a quasi-metric. For many non-metric distances possibly not satisfying the triangle inequality, e.g. for the energy based semi-metric curve distance in Schmidt et al. (2007), it is open whether the following assumptions are valid.

**Asumption 4.3.** For every  $p \in P$  and  $\epsilon > 0$  there is a  $\delta = \delta(\epsilon, p) > 0$  such that  $|\rho(x, p') - \rho(x, p)| < \epsilon$  for all  $x \in Q, p' \in P$  with  $d(p, p') < \delta$ .

**Asumption 4.4.** There are  $p_0 \in P$  and  $C > 0$  such that  $\mathbb{P}\{\rho(X, p_0) < C\} > 0$  and that for every sequence  $p_n \in P$  with  $d(p_0, p_n) \rightarrow \infty$  there is a sequence  $M_n \rightarrow \infty$  with  $\rho(x, p_n) > M_n$  for all  $x \in Q$  with  $\rho(x, p_0) < C$ . Moreover, if  $p_n \in P$  with  $d(p^*, p_n) \rightarrow \infty$  for some  $p^* \in M$ , then  $d(p_0, p_n) \rightarrow \infty$ .

Then we have the following generalization of Ziezold (1977, Theorem) and Bhattacharya and Patrangenaru (2003, Theorem 2.3), cf. Huckemann (2011b, Theorems A.3 and A.4).

**Theorem 4.5.** Suppose that the data space  $Q$  is separable,  $\mathbb{E}(\rho(X, p)^2) < \infty$  for all  $p \in P$  and that Assumption 4.3 is valid. Then property (Z-SC) (Ziezold strong consistency) holds for the set of Fréchet  $\rho$ -means on  $P$ .

If additionally  $E^{(\rho)} \neq \emptyset$ ,  $\overline{\cup_{n=1}^{\infty} E_n^{(\rho)}}(\omega)$  enjoys the  $d$ -Heine-Borel property for a.e.  $\omega \in \Omega$  and the coercivity Assumption 4.4 is valid then property (BP-SC) (Bhattacharya-Patrangenaru strong consistency) holds for the set of Fréchet  $\rho$ -means on  $P$ .

In order to formulate a  $\sqrt{n}$ -Gaussian central-limit theorem, we require additional properties.

**Asumption 4.6.** The population Fréchet- $\rho$  mean is unique up to a discrete group action, i.e. there are a discrete group  $H$  acting on  $P$  and  $\mu \in P$  such that  $\{h\mu : h \in H\} = E^{(\rho)}$  and there is an open neighborhood  $U$  of  $\mu$  in  $P$  that is a  $D$ -dimensional smooth manifold,  $D \in \mathbb{N}$ .

**Definition 4.7.** Under Assumption 4.6 we say that a sequence  $\mu_n(\omega)$  ( $n \in \mathbb{N}$ ) of  $P$ -valued estimates of  $\mu$  satisfies a Central-Limit-Theorem (CLT), if in any local chart  $(\phi, U)$  near  $\mu = \phi^{-1}(0)$  there are a suitable deterministic  $D \times D$  matrix  $A_\phi$  and a random Gaussian  $D \times D$  matrix  $\mathcal{G}_\phi$  with zero mean and semi-definite symmetric covariance matrix  $\Sigma_\phi$  such that

$$\sqrt{n}A_\phi(\phi(\mu_n) - \phi(\mu)) \rightarrow \mathcal{G}_\phi$$

in distribution as  $n \rightarrow \infty$ .

In most applications  $A_\phi$  seems non-singular, then in consequence of the “ $\delta$ -method”, for any other chart  $(\phi', U)$  near  $\mu = \phi'^{-1}(0)$  we have simply

$$A_{\phi'}^{-1} \Sigma_{\phi'} (A_{\phi'}^{-1})^T = J(\phi' \circ \phi^{-1})_0 A_\phi^{-1} \Sigma_\phi (A_\phi^{-1})^T J(\phi' \circ \phi^{-1})_0^T$$

where  $J(\cdot)_0$  denotes the Jacobi-matrix of first derivatives at the origin. In fact, the argument for a central-limit theorem rests on the “ $\delta$ -method” used for analyzing  $M$ -estimators, cf. van der Vaart (2000). To this end we require additional smoothness assumptions.

**Asumption 4.8.** *Under Assumption 4.6 assume further that for every local chart  $(\phi, U)$  near  $\mu = \phi^{-1}(0)$  the mapping  $x \mapsto \rho(X, \phi^{-1}(x))^2$  is twice differentiable in  $U$  a.s. and that*

$$\left. \begin{array}{ll} \mathbb{E}(\text{grad}_2 \rho(X, \mu)^2) & \text{exists,} \\ \text{Cov}(\text{grad}_2 \rho(X, \mu)^2) & \text{exists,} \\ \mathbb{E}(\text{Hess}_2 \rho(X, \nu)^2) & \text{exists for } \nu \text{ near } \mu \text{ and is continuous at } \nu = \mu, \end{array} \right\} \quad (4)$$

where  $\text{grad}_2 \rho(q, \nu)^2$  and  $\text{Hess}_2 \rho(q, \nu)^2$  denote the gradient and the Hessian of the above mapping.

Obviously the validity of (4) is independent of the particular chart chosen. The following Theorem is a generalization of (Bhattacharya and Patrangenaru, 2005, p.1230) who provided a proof for the case of  $P = Q$  a manifold and  $\rho$  either an intrinsic or an extrinsic distance. For the general case see (Huckemann, 2011a, Theorem 6).

**Theorem 4.9.** *Under Assumptions 4.6 and 4.8 suppose that  $E_n^{(\rho)}$  is a uniform strongly consistent estimator of a Fréchet population  $\rho$ -mean set  $E^{(\rho)}$ . Then for every measurable choice  $\mu_n(\omega) \in E_n^{(\rho)}$  there is a random sequence  $h_n(\omega) \in H$  such that  $h_n(\omega)\mu_n(\omega)$  satisfies a CLT. In a suitable chart  $(\phi, U)$  the corresponding matrices from Definition 4.7 are given by*

$$A_\phi = \mathbb{E}(\text{Hess}_2 \rho(X, \mu)^2), \quad \Sigma_\phi = \text{Cov}(\text{grad}_2 \rho(X, \mu)^2).$$

## 5 A Closeup on the Hypotheses of the Central Limit Theorem

Let us now review under which settings the Gaussian  $\sqrt{n}$ -Central-Limit-Theorem 4.9 is valid. We note that Assumptions 4.4 and 4.3 of Section 4 are valid if the spaces  $Q$  and  $P$  as well as their linking distance  $\rho$  are sufficiently well behaved, e.g. this is the case if  $P$  and  $Q$  are compact and  $d$  a metric. In the general case, if additionally to Assumptions 4.4 and 4.3

- (i) there is a unique (up to a discrete group action) Fréchet  $\rho$ -mean  $\mu \in P$ ,
- (ii)  $P$  is locally Euclidean in neighborhood  $U$  around  $\mu$  and
- (iii)  $x \mapsto \rho(X, \phi^{-1}(x))^2$  is twice continuously differentiable in  $U$  a.s. for a local chart  $(U, \phi)$ ,

then every measurable selection  $\mu_n \in E_n^{(\rho)}$  gives a uniformly consistent estimator of  $\mu$  for which Theorem 4.9 is valid.

### 5.1 Uniqueness of Fréchet $\rho$ -Means

For extrinsic means in case  $P = Q$  being a complete smooth manifold embedded in a Euclidean space  $\mathbb{R}^D$  and  $\rho$  being the restriction of the Euclidean distance to  $Q$ , we have the two exhaustive results due to Bhattacharya and Patrangenaru (2003) and Hendriks and Landsman (1996). Before we state these, recall from Section 2 that in this case the orthogonal projection  $\Phi : \mathbb{R}^D \rightarrow Q$  is well defined and smooth outside the closure of the points with multivalent orthogonal projection, which forms a set of Lebesgue measure zero in  $\mathbb{R}^D$ , cf. Huckemann et al. (2010, Theorem A.5), see also Bhattacharya and Patrangenaru (2003, Theorem 5.3).

This closed set comprises the *focal points* (with multivalent orthogonal projection, so-called by Bhattacharya and Patrangenaru (2003) and corresponding to the medial axis of Blum and Nagel (1978)) and the minimizing foci (where the endpoint map is singular – those are focal points in the sense of Milnor (1969), also called ambiguity set by Hastie and Stuetzle (1989) – and where the endpoint map gives an orthogonal projection; more detail and more generality for shape spaces can be found in Huckemann et al. (2010, Appendix A)). For an ellipse, the open line segment between the usual foci – which are the minimizing foci – give the focal points.

Moreover, denote by  $\mathbb{E}[X]$  the usual expected value in  $\mathbb{R}^D$  of  $X \in Q$  embedded in  $\mathbb{R}^D$ , if existent.

**Theorem 5.1.** *The extrinsic mean on a complete smooth manifold is unique if  $\mathbb{E}[X]$  exists and is not a cluster point of focal points.*

In consequence of the asymptotic distribution for extrinsic means can be easily computed.

**Corollary 5.2.** *If  $\mathbb{E}[X]$  exists and is not a cluster point of focal points, then*

$$\sqrt{n}(\mu_n - \mu) \rightarrow \mathcal{N}(0, d\Phi_{\mathbb{E}[X]} \text{Cov}[X] d\Phi_{\mathbb{E}[X]}^T)$$

*in distribution for the unique extrinsic population mean  $\mu = \Phi(\mathbb{E}[X])$  and any measurable selection  $\mu_n$  of extrinsic sample means.*

For residual means less is known. On embedded hyperspheres  $S \subset \mathbb{R}^D$  it follows at once from the definition that  $\mu \in S$  is a residual mean if and only if  $\mu$  is an eigenvector of  $\mathbb{E}[XX^T]$ . In particular, residual means are only unique up to the group action of sign changes. This yields at once the following.

**Theorem 5.3.** *On an embedded hypersphere, the residual mean set contains just one pair of antipodal points, if and only if the largest eigenvalue of  $\mathbb{E}[XX^T]$  has multiplicity one.*

Note that the uniqueness results on spheres for extrinsic and residual means extend at once to Ziezold and Procrustes means on real projective spaces.

For intrinsic means even less is known. The full picture is available only on circles, cf. (Hotz and Huckemann, 2011, Proposition 3.4).

**Theorem 5.4.** *Consider the distribution of  $X$  on the circle  $S^1$ , decomposed into the part  $\lambda$  which is absolutely continuous w.r.t. Lebesgue measure, with density  $f$ , and the part  $\eta$  singular to Lebesgue measure. Moreover, assume that  $S_1, \dots, S_k$  be the maximal arcs distinct up to their endpoints on which  $f \leq \frac{1}{2\pi}$  such that their interiors  $\text{int } S_j$  are disjoint from  $\text{supp } \eta$ , and that  $\{x \in S^1 : f(x) = \frac{1}{2\pi}\}$  is a Lebesgue null-set. Then  $X$  has at most  $k$  intrinsic means and every  $\text{int } S_j := \{\tilde{x} : x \in \text{int } S_j\}$  contains at most one candidate. Here,  $x \mapsto \tilde{x}$  denotes the antipodal map on  $S^1$ .*

In consequence one can show that candidates for intrinsic means of samples of size  $n$  lie on the vertices of a regular polygon, thus reducing the computational effort to  $O(n \log n)$ , cf. Hotz and Huckemann (2011, Remark 3.8) and Charlier (2011, Section 5.3).

The general case for  $Q$  being a complete Riemannian manifold has been subsequently attacked by Karcher (1977), Kendall (1990), Le (2001) and Groisser (2005) with the best result to date due to Afsari (2011):

**Theorem 5.5.** *Let  $r$  be the injectivity radius of  $Q$  and suppose that its sectional curvatures are bounded from above by  $c > 0$ . Then the intrinsic mean of  $X$  is unique if the support of  $X$  is contained within a geodesic ball of radius strictly less than  $\frac{1}{2} \min \left\{ r, \frac{\pi}{\sqrt{c}} \right\}$ .*

The assumption on boundedness of sectional curvatures is in general not given for the manifold part  $Q^*$  of shape spaces. Indeed, from dimension 3 on, for Kendall's shape spaces near every point of the singularity set  $Q_0$ , some sectional curvatures tend to infinity, cf. (Kendall et al., 1999, pp. 149–156) and (Huckemann et al., 2010, Theorem 5.2).



## 5.2 Locally Euclidean Neighborhoods of Fréchet $\rho$ -Means

Recall the stratification (2) of a general shape space  $M/G = Q = Q^* \cup Q_0$ . We consider in the following the case of continuous functions  $\rho : Q \times Q \rightarrow [0, \infty)$  yielding Fréchet  $\rho$ -means on  $Q = P$ .

**Definition 5.6.** *We say that a random shape  $X$  on  $Q$  is regular with respect to  $\rho$  if its Fréchet  $\rho$ -mean  $E_\rho(X)$  exists, if  $\mathbb{P}\{X \in Q^*\} > 0$  and if the distribution of  $X$  has at most countably many point masses on  $Q_0$ .*

**Definition 5.7.** *We call  $\rho$  manifold stable if  $E_\rho(X) \subset Q^*$  for all  $X$  regular with respect to  $\rho$ ; otherwise we say that  $Q^0$  is hit.*

The following results are taken from Huckemann (2012).

**Theorem 5.8.** *On a shape space the intrinsic geodesic distance is manifold stable. Hence for every random shape regular w.r.t. the geodesic distance, its intrinsic means lie on the manifold part.*

**Theorem 5.9.** *On Kendall's shape spaces the Ziezold distance is manifold stable. Hence for every regular random Kendall shape, its Ziezold means lie on the manifold part.*

**Theorem 5.10.** *On Kendall's shape spaces  $\Sigma_m^k$  for  $m \geq 3$  the singular stratum is hit by the Procrustes distance. Hence there are regular Kendall shapes such that their Procrustes means intersect with the singular part.*

Theorem 5.8 applied to the finite dimensional subspaces exhausting the quotient shape space of closed planar curves with arbitrary initial point introduced by Zahn and Roskies (1972); Klassen et al. (2004) gives that the shape of the circle, since it is a singularity, can never be an intrinsic shape mean of non-circular curves.

## 5.3 A.S. Local Twice Differentiability

We consider here the condition that  $x \mapsto \rho(X, \phi^{-1}(x))^2$  is twice continuously differentiable in  $U$  a.s. for a local chart  $(U, \phi)$ , in the case of  $\rho$  being a geodesic intrinsic distance. If  $Q = P$  is a manifold, since the squared geodesic distance is not differentiable at a cut locus, this condition translates to the fact that

$$X \notin \mathcal{C}(\nu) \text{ a.s. for all } \nu \in U. \quad (5)$$

Here,  $\mathcal{C}(\nu)$  denotes the *cut locus* of  $\nu$  which comprises all points  $q \in Q$  such that the extension of a length minimizing geodesic joining  $\nu$  with  $q$  is no longer length minimizing beyond  $q$ . In fact, this condition is much more restrictive than it may seem. It excludes most parametric models: e.g. on spheres, all of the distributions, Bingham, Fisher, Fisher-Bingham, etc. have a nowhere vanishing continuous density, cf. Mardia and Jupp (2000).

On the circle  $S^1$  which we represent by  $[-\pi, \pi)$ , in Hotz and Huckemann (2011) we have relaxed condition (5) and studied the following implications on the Gaussian  $\sqrt{n}$ -Central-Limit-Theorem 4.9. The first two assertions of the following Theorem 5.11 can also be found in Charlier (2011). Independently, McKilliam et al. (2012) have established the second assertion of Theorem 5.11 and the first of Theorem 5.12.

**Theorem 5.11.** *Let  $X$  be a random variable on the circle  $S^1$  with intrinsic mean  $\mu = 0$ . Then*

$$\mathbb{P}\{X = -\pi\} = 0.$$

*If  $X$  restricted to some neighborhood of  $-\pi$  features a continuous density  $f$ , then*

$$f(-\pi) \leq \frac{1}{2\pi}.$$

*Moreover,  $\mu = 0$  is contained in a whole continuum of intrinsic means if there is  $\epsilon > 0$  such that  $f(x - \pi) = \frac{1}{2\pi} = f(\pi - x)$  for all  $0 \leq x \leq \epsilon$ .*

**Theorem 5.12.** *Let  $X$  be a random variable on the circle  $S^1$  with unique intrinsic mean  $\mu = 0$  featuring a continuous density  $f$  near  $-\pi$ . Assume that  $\mathbb{E}(X^2) = \sigma^2$  where  $X$  is viewed as taking values in  $[-\pi, \pi)$  and that  $\mu_n$  is an intrinsic sample mean. If  $f(-\pi) < \frac{1}{2\pi}$  then*

$$\sqrt{n} \mu_n \xrightarrow{D} \mathcal{N} \left( 0, \frac{\sigma^2}{(1 - 2\pi f(-\pi))^2} \right).$$

*If  $f(-\pi) = \frac{1}{2\pi}$  suppose that there are  $\delta > 0$  and  $k \in \mathbb{N}$  such that  $f$  is  $k - 1$  times continuously differentiable in  $(-\pi, \delta - \pi)$  and  $(\pi - \delta, \pi)$  with  $f^{(j)}(\pi-) = f^{(j)}(-\pi+) = 0$  for all  $1 \leq j \leq k$ , and that there are  $k$ -th order continuous directional derivatives with  $0 \neq f^{(k)}(\pi-) = (-1)^k f^{(k)}(-\pi+) < \infty$ . Then*

$$\sqrt{n} \operatorname{sign}(\mu_n) |\mu_n|^{k+1} \xrightarrow{D} \mathcal{N} \left( 0, \frac{\sigma^2 ((k+1)!)^2}{(2\pi f^{(k)}(-\pi+))^2} \right).$$

In consequence, on the circle, mild violations of condition (5) increase the variance in the Gaussian  $\sqrt{n}$ -Central-Limit-Theorem 4.9. For stronger violations, the assertion of Theorem 4.9 becomes false and the asymptotic rate can be arbitrarily slow with the limiting case of a whole continuum of intrinsic means.

## 6 Outlook: Geometrical Statistics and Statistical Geometry

In this paper, we have briefly sketched a path from earlier origins to some recent results concerning the statistical analysis of non-linear data. It became clear that with the desire to retain so called second order data information, curvature has to be dealt with in a sophisticated way. In this (positive curvature) context three major issues materialized that to date are still very far from their full resolution:

- (i) uniqueness of means and their
- (ii) manifold stability, as well as the
- (iii) impact of mass near the means' cut loci.

Although in a non-positive curvature context, issues (i) and (iii) are non-existent (for the uniqueness of intrinsic means under non-positive curvature see Karcher (1977) and Sturm (2003)), issue (ii) comes in a slightly different – but none the less challenging disguise: In the general context of stratified spaces – in particular for non-positive curvature spaces that are not of the form  $M/G$  with a curvature positive w.r.t.  $M$  – manifold stability translates into its opposite phenomenon of *stickiness* to the singular part. Indeed, it has been shown that the geodesic intrinsic distance on the “open book” not only hits the singular “spine” but actually “sticks” to it, cf. Hotz et al. (2012); Barden et al. (2013). In fact, it seems that manifold stability vs. stickiness appears to be related to a positive curvature vs. a non-positive curvature context. Even more, this overview and Hotz et al. (2012); Barden et al. (2013) seem to suggest that limiting laws on positive curvature spaces tend to be of equal or slower rate than  $1/\sqrt{n}$  while on non-positive curvature spaces the rates tend to be equal or higher than  $1/\sqrt{n}$ .

In this general context we believe that we see a new field of research emerging. Here are three research traits we would like to identify:

- (I) to tackle the above issues (i), (ii) and (iii) towards a satisfactory treatment;
- (II) to make these results available in the context of (semi-)intrinsic statistical analysis built on descriptors not valued in the the data space, e.g. geodesics, or constant intrinsic curvature subspaces, in particular totally geodesic subspaces (subspaces of zero intrinsic curvature), etc.; and

- (III) to fully understand how the obtainable limit theorems, in particular their associated rates, relate to topological and geometric invariants of the underlying spaces – this might be called “Statistical Geometry”.

## References

- Afsari, B. (2011). Riemannian  $L^p$  center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* 139, 655–773.
- Barden, D., H. Le, and M. Owen (2013). Central limit theorems for fréchet means in the space of phylogenetic trees. *Electron. J. Probab* 18(25), 1–25.
- Bhattacharya, R. N. and V. Patrangenaru (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics* 31(1), 1–29.
- Bhattacharya, R. N. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics* 33(3), 1225–1259.
- Blum, H. and R. N. Nagel (1978). Shape description using weighted symmetric axis features. *Pattern Recognition* 10(3), 167–180.
- Bredon, G. E. (1972). *Introduction to Compact Transformation Groups*, Volume 46 of *Pure and Applied Mathematics*. New York: Academic Press.
- Charlier, B. (2011). Necessary and sufficient condition for the existence of a fréchet mean on the circle. *arXiv preprint arXiv:1109.1986*.
- Choquet, G. (1954). Theory of capacities. *Annales de l’Institut de Fourier* 5, 131–295.
- Cremers, D. (2006). Dynamical statistical shape priors for level set based tracking. 28(8), 1262–1273.
- Dryden, I. L. and K. V. Mardia (1998). *Statistical Shape Analysis*. Chichester: Wiley.
- Fletcher, P. T., C. Lu, S. M. Pizer, and S. C. Joshi (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging* 23(8), 995–1005.
- Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’Institut de Henri Poincaré* 10(4), 215–310.
- Goodall, C. R. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *Journal of the Royal Statistical Society, Series B* 53, 285–339.
- Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* 40, 33–51.
- Groisser, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics: Internatl. J. Probab. Stochastic Processes* 77(1), 51–60.
- Hastie, T. and W. Stuetzle (1989). Principal curves. *Journal of the American Statistical Association* 84(406), 502–516.
- Hendriks, H. and Z. Landsman (1996). Asymptotic behaviour of sample mean location for manifolds. *Statistics & Probability Letters* 26, 169–178.
- Hotz, T. and S. Huckemann (2011). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *arXiv.org*, 1108.2141.
- Hotz, T., S. Huckemann, H. Le, J. S. Marron, J. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, and S. Skwerer (2012). Sticky central limit theorems on open books. *Annals of Applied Probability*. accepted.

- Huckemann, S. (2011a). Inference on 3D Procrustes means: Tree boles growth, rank-deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics* 38(3), 424–446.
- Huckemann, S. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics* 39(2), 1098–1124.
- Huckemann, S. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Mathematical Statistics* 64(6), 1227–1259.
- Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica* 20(1), 1–100.
- Jung, S., I. L. Dryden, and J. S. Marron (2012). Analysis of principal nested spheres. *Biometrika* 99(3), 551–568.
- Jung, S., M. Foskey, and J. S. Marron (2011). Principal arc analysis on direct product manifolds. *The Annals of Applied Statistics* 5, 578–603.
- Jupp, P. E. (1988). Residuals for directional data. *Journal of Applied Statistics* 15(2), 137–147.
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics* XXX, 509–541.
- Kendall, D. (1974). Foundations of a theory of random sets. In *Stochastic Geometry, Tribute Memory Rollo Davidson*, New York, pp. 322–376. Wiley.
- Kendall, D. G. (1977). The diffusion of shape. *Adv. Appl. Prob.* 9, 428–430.
- Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.* 16(2), 81–121.
- Kendall, D. G., D. Barden, T. K. Carne, and H. Le (1999). *Shape and Shape Theory*. Chichester: Wiley.
- Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proceedings of the London Mathematical Society* 61, 371–406.
- Kent, J. T. and K. V. Mardia (2012). A geometric approach to projective shape and the cross ratio. *Biometrika* 99(4), 833–849.
- Klassen, E., A. Srivastava, W. Mio, and S. Joshi (2004, March). Analysis on planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(3), 372–383.
- Kobayashi, S. and K. Nomizu (1969). *Foundations of Differential Geometry*, Volume II. Chichester: Wiley.
- Le, H. (2001). Locating Fréchet means with an application to shape spaces. *Advances of Applied Probability (SGSA)* 33(2), 324–338.
- Mardia, K. and V. Patrangenaru (2001). On affine and projective shape data analysis. *Functional and Spatial Data Analysis, Proceedings of the 20th LASR Workshop (Eds: K.V. Mardia and R.G. Aykroyd)*, 39–45.
- Mardia, K. and V. Patrangenaru (2005). Directions and projective shapes. *The Annals of Statistics* 33, 1666–1699.
- Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*. New York: Wiley.

- Matheron, G. (1975). *Random sets and integral geometry*. Wiley Series in Probability and Mathematical Statistics. New York.
- McKilliam, R. G., B. G. Quinn, and I. V. L. Clarkson (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing* 60(5), 2115–2124.
- Michor, P. W. and D. Mumford (2006). Riemannian geometries on spaces of plane curves. *J. of the European Math. Soc.* 8, 1–48.
- Milnor, J. W. (1969). *Morse Theory*. Princeton: Princeton University Press. 3rd printing with corrections.
- Mio, W., A. Srivastava, and S. Joshi (2007). On shape of plane elastic curves. *International Journal of Computer Vision* 73(3), 307–324.
- Molchanov, I. (2005). *Theory of random sets*. Probability and Its Applications, London: Springer. xvi.
- O’Neill, B. (1966). The fundamental equations of a submersion. *Michigan Math. J.* 13(4), 459–469.
- Schmidt, F. R., E. Töppe, D. Cremers, and Y. Boykov (2007). Intrinsic mean for semi-metrical shape retrieval via graph cuts. In F. A. Hamprecht, C. Schnörr, and B. Jähne (Eds.), *DAGM-Symposium*, Volume 4713 of *Lecture Notes in Computer Science*, pp. 446–455. Springer.
- Schulz, J., S. Jung, S. Huckemann, J. Marron, and S. Pizer (2012). Analysis of rotational deformations from directional data. preprint.
- Srivastava, A., E. Klassen, S. Joshi, and I. Jermyn (2011). Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(7), 1415–1428.
- Stachó, L. L. (1979). On curvature measures. *Acta Sci. Math* 41, 191–207.
- Sturm, K. (2003). Probability measures on metric spaces of nonpositive curvature. *Contemporary mathematics* 338, 357–390.
- van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ. Press.
- Younes, L. (2010). *Shapes and diffeomorphisms*, Volume 171. Springer Verlag.
- Zahn, C. and R. Roskies (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers* C-21, 269–281.
- Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes A*, 591–602.
- Ziezold, H. (1994). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biometrical Journal* (36), 491–510.