Backward Nested Descriptors Asymptotics with Inference on Stem Cell Differentiation

Stephan F. Huckemann*

Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences
Georg-August-Universität Göttingen
Goldschmidtstr. 7
D-37077 Göttingen, Germany
e-mail: huckeman@math.uni-goettingen.de

and

Benjamin Eltzner*

Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences
Georg-August-Universität Göttingen
Goldschmidtstr. 7
D-37077 Göttingen, Germany
e-mail: beltzne@uni-goettingen.de

Abstract: For sequences of random backward nested subspaces as occur, say, in dimension reduction for manifold or stratified space valued data, asymptotic results are derived. In fact, we formulate our results more generally for backward nested families of descriptors (BNFD). Under rather general conditions, asymptotic strong consistency holds. Under additional, still rather general hypotheses, among them existence of a.s. local twice differentiable charts, asymptotic joint normality of a BNFD can be shown. If charts factor suitably, this leads to individual asymptotic normality for the last element, a principal nested mean or a principal nested geodesic, say. It turns out that these results pertain to principal nested spheres (PNS) and principal nested great subsphere (PNGS) analysis by Jung et al. (2012) as well as to the intrinsic mean on a first geodesic principal component (IMo1GPC) for manifolds and Kendall’s shape spaces. A nested bootstrap two-sample test is derived and illustrated with simulations. In a study on real data, PNGS is applied to track early human mesenchymal stem cell differentiation over a coarse time grid and, among others, to locate a change point with direct consequences for the design of further studies.

AMS 2000 subject classifications: Primary 62G20, 62G25; secondary 62H11, 58C06, 60D05.

Keywords and phrases: Fréchet means, dimension reduction on manifolds, principal nested spheres, asymptotic consistency and normality, geodesic principal component analysis, Kendall’s shape spaces, flags of subspaces.

1. Introduction

In this paper, the novel statistical problem of deriving asymptotic results for nested random sequences of statistical descriptors for data in a non-Euclidean space is considered. It can be viewed as a generalization of classical PCA’s asymptotics, e.g. by Anderson (1963); Watson (1983); Ruymgaart and Yang (1997), where, as a consequence of Pythagoras’ theorem, nestedness of approximating subspaces is trivially given and thus requires no special attention.

Classical PCA from a geometric perspective. Given data on $Q = \mathbb{R}^m$, for every $0 \leq k \leq m$ a unique affine subspace $p^k$ of dimension $k$ is determined by equivalently minimizing residual sums of squares or, among those containing the classical mean $\mu$, maximizing the projected variance. These subspaces have representations as $p^k = \mu + \text{span}\{\gamma_1, \ldots, \gamma_k\}$, the affine translates of spans from an eigenvector decomposition $\gamma_1, \ldots, \gamma_m$ of the data’s covariance matrix with descending eigenvalues. In consequence, one may either start from the zero dimensional mean and subsequently add most descriptive dimensions (forward) or start from

*Supported by DFG SFB 755, DFG HU 1575/4 and the Niedersachsen Vorab of the Volkswagen Foundation
the full dimensional space and remove least descriptive dimensions (backward) to obtain the same nested sequence of subspaces
\[ \{\mu\} = p_0 \subset p_1 \subset \ldots \subset p_m = Q. \]  

(1)

**Non-Euclidean PCA.** For dimension reduction of non-Euclidean data, *Procrustes analysis* by Gower (1975) and later *principal geodesic analysis* by Fletcher et al. (2004) are approaches to mimic PCA on shape spaces and Riemannian manifolds, respectively. Both build on the concept of a *Fréchet mean*, a minimizer of expected squared distance, around which classical PCA is conducted for the data mapped to a suitable tangent space. This can be viewed as a *forward nested approach*.

Notably, these *tangent space PCA* methods are in no way canonical, as the statistical outcomes depend on specific choices of tangent space coordinates, and given curvature, no tangent space coordinates can correctly reflect mutual data distances. More severely, for specific data situations as in the following example, such tangent space methods may be meaningless.

**Example 1.1.** Consider data on a two-sphere that is uniformly distributed on its equator. Then there are two intrinsic Fréchet means, each at one of the two poles and tangent space PCA using standard coordinates at either mean yields 2D isotropy, see Huckemann (2012).

To tackle also scenarios where the mean is far from the data, *geodesic principal component analysis* (GPCA) has been introduced by Huckemann and Ziezold (2006); Huckemann et al. (2010), where in particular the first geodesic principal component is a geodesic approximating the data best in intrinsic $L^2$-distance, along with the *intrinsic mean on the first geodesic principal component* (IMo1GPC). For spheres, this *backward nested* approach has been generalized by Jung et al. (2011) to *principal arc analysis* with the *backward nested mean* confined to the best approximating small circle. This method and its generalization *backward nested sphere analysis* (PNS) by Jung et al. (2012) give a tool for descriptive shape analysis that often strikingly outperforms tangent space PCA, e.g. Pizer et al. (2013). Here, the data space is a unit sphere $Q = S^m$ of dimension $m \in \mathbb{N}$, say, and in (1) each of the $p_k$ is a $k$-dimensional (small) subsphere for PNS and for *principal nested great spheres* (PNGS) it is a $k$-dimensional great subsphere. Another generalization of GPCA is given by *iterated frame bundle development* by Sommer (2013) and the recently proposed *barycentric subspaces* by Pennec (2015, 2016).

**Asymptotics for non-Euclidean PCA.** Asymptotics for Fréchet means have been subsequently provided, among others, by Ziezold (1977); Hendriks and Landsman (1996); Bhattacharya and Patrangenaru (2003, 2005); Huckemann (2011a), allowing for inferential methods such as two-sample tests. Asymptotics for the above mentioned tangent space PCA methods, however, reflecting forward nestedness due to random base points (i.e. corresponding means) of tangent spaces with random PCs therein, remain open to date. Similarly, asymptotics for non-data space valued descriptors, geodesics, say, are available (cf. Huckemann (2011, 2014)), but only for ones that are directly defined as minimizers, not indirectly as a nested sequence of minimizers.

**Challenges for and results of this paper.** It is the objective of this paper to close the backward part of this gap by providing asymptotic results for rather general random *backward nested families of descriptors* (BNFDs) on rather general spaces. The challenge here is that random objects that are constrained by other random objects are to be investigated, requiring an elaborate setup. Into this setup, we translate strong consistency arguments of Ziezold (1977) and Bhattacharya and Patrangenaru (2003), and introducing a *constrained* M-estimation technique, we show joint asymptotic normality of an entire BNFD. In the special case of nested subspaces, BNFDs may terminate at any dimension and $p_j$, $j \geq 0$.

As we minimize a functional under the constraining conditions that other functionals are minimized as well, our approach can be called *constrained M-estimation*. In the literature, this term *constrained M-estimation* has been independently introduced by Kent and Tyler (1996) who robustify M-estimators by introducing constraining conditions and by Geyer (1994); Shapiro (2000), who consider M-estimators that are confined to closed subsets of a Euclidean space with specifically regular boundaries. It seems that our M-estimation problem, which is constrained to satisfying other M-estimation problems has not been dealt with before. We solve it using a random Lagrange multiplier approach.
Furthermore, in order to obtain asymptotic normality of each single sequence element, in particular for the last, we require the rather technical concept of factoring charts. Our very general setup will be illustrated, still with some effort, by example of PNS, PNGS and the intrinsic mean on a first geodesic principal component (IMo1GPC).

In order to exploit nested asymptotic normality for a nested two-sample test, we utilize bootstrapping techniques. While for Fréchet means, as they are descriptors assuming values in the data space, one can explicitly model the dependence of the random base points of the tangent spaces as in Huckemann et al. (2010a), so that suitable statistics can be accordingly directly approximated, this modeling and approximation can be avoided using the bootstrap as in Bhattacharya and Patrangenaru (2005). For our application at hand, as data space and descriptor space are different, we cannot approximate the distribution of random descriptors and we fall back on the bootstrap.

Suggestions for live imaging of stem cell differentiation. After illustrations of our nested two-sample test by simulations for PNS and PNGS, we apply it to a cutting edge application in adult human stem cell differentiation research. “Rooted in a line of experimentation originating in the 1960s” (from Bianco et al. (2013)), the promise that stem cells taken from a patient’s bone marrow may be used to rebuild specific, previously lost, patient’s tissue is currently undergoing an abundance of clinical trials. Although the underlying mechanisms are, to date, not fully understood, it is common knowledge that early stem cell differentiation is triggered by biomechanical cues, e.g. Zemel et al. (2010), which result in specific ordering of the cellular actin-myosin filament skeleton. In collaboration with the Third Institute of Physics at the University of Göttingen we map fluorescence images of cell structures to two-spheres, where each point stands for a specific ordering. With our 2D PNGS two-sample test we can track the direction of increased ordering over the first 24 hours. We find, however, a consistent reversal of ordering between hours 16 to 20 which hint toward the effect of cell division. This effect suggests that the commonly used time point of 24 hours for fixated hMSCs imaging, e.g. as in Zemel et al. (2010), may not be ideal for cell differentiation detection. In fact, our method can be used to direct more elaborate and refined imaging techniques, such as time resolved in-vivo cell imaging, using Eltzner et al. (2015), say, to investigate specifically discriminatory time intervals in detail.

2. Backward Nested Families of Descriptors

In this section we first introduce the general framework including the fundamental assumption of factoring charts which is essential to prove asymptotic normality of single nested descriptors in Section 4. Then we give the first example: the intrinsic mean on a first geodesic principal component (IMo1GPC) for Riemannian manifolds. The next two examples of principal nested spheres (PNS) as well as principal nested great spheres (PNGS), and finally the example for the IMo1GPCs also on non-manifold Kendall’s shape spaces are deferred to the Supplement, Sections 6.1 and 6.2. The first example is rather straightforward, the last is slightly more involved and the second and third are much more involved. The differential geometry used here can be found in any standard textbook, e.g. Lee (2013).

First, let us quickly sketch the ideas in case of IMo1GPCs on a Riemannian manifold \( Q \). There, we have the space \( P_1 \) of point sets of geodesics on \( Q \) which is the first non-trivial descriptor space "below" the space \( P_2 = \{Q\} \). In order to show strong asymptotic consistency in Theorem 4.1, on \( P_1 \times P_1 \) we require the concept of a loss function \( d_1 \) that has some properties of a distance between two (point sets of) geodesics. In order to model nestedness, given a geodesic \( p \in P_1 \) we require the set \( S_p \) of lower dimensional descriptors in \( P_0 = Q \) which lie on \( p \). These are the candidate nested means on \( p \), and in this case, \( S_p = p \). Further, we need the data projection \( \pi_{Q,p} : Q \to p \) and we measure the distance \( \rho(\pi_{Q,p}(q), s) \) of the projected data to a candidate nested mean \( s \in S_p \). Then every \( (Q,p,s) \) with \( p \in P_1, s \in S_p \) will be a backward nested family of descriptors (BNFD) and the set \( (p,s) \) with \( p \in P_1, s \in S_p \) carries a natural manifold structure. It is the objective of factoring charts to represent this manifold locally as a direct product of arbitrary variable offsets \( s \in P_0 = Q \) times a suitable space parametrizing directions of geodesics, parametrized independently from the offset \( s \in Q \), cf. Figure 1. We will see that is precisely the geometry of the projective bundle. Once
we establish asymptotic normality of the backward nested descriptor \((p, s)\), asymptotic normality follows at once also for \(s\), because, under factoring charts, \(s\) is given by some coordinates of a Gaussian vector as reasoned in the proof of Theorem 4.5.

### 2.1. General Framework

With a silently underlying probability space \((\Omega, \mathfrak{A}, \mathbb{P})\), random elements on a topological space \(Q\) are mappings \(X : \Omega \to Q\) that are measurable with respect to the Borel \(\sigma\)-algebra of \(Q\). In the following, smooth refers to existing continuous 2nd order derivatives.

For a topological space \(Q\) we say that a continuous function \(d : Q \times Q \to [0, \infty)\) is a loss function if \(d(q, q') = 0\) if and only if \(q = q'\). We say that a set \(A \subset Q\) is \(d\)-bounded if \(\sup_{a, a' \in A} d(a, a') < \infty\). Moreover, we say that \(B \subset Q\) is \(d\)-Heine Borel if all closed \(d\)-bounded subsets of \(B\) are compact.

**Definition 2.2.** A separable topological space \(Q\), called the data space, admits backward nested families of descriptors (BNFDs) if

- (i) there is a collection \(P_j\) \((j = 0, \ldots, m)\) of topological separable spaces with loss functions \(d_j : P_j \times P_j \to [0, \infty)\);
- (ii) \(P_n = \{Q\}\);
- (iii) every \(p \in P_j\) \((j = 1, \ldots, m)\) is itself a topological space and gives rise to a topological space \(\emptyset \neq S_p \subset P_{j-1}\) which comes with a continuous map
  
  \[ \rho_p : p \times S_p \to [0, \infty); \]

- (iv) for every pair \(p \in P_j\) \((j = 1, \ldots, m)\) and \(s \in S_p\) there is a measurable map called projection
  
  \[ \pi_{p,s} : p \to s. \]

For \(j \in \{1, \ldots, m\}\) and \(k \in \{1, \ldots, j\}\) call a family

\[ f = \{p^{j-k}, \ldots, p^j\}, \text{ with } p^{l-1} \in S_{p^l}, l = j - k + 1, \ldots, j \]

a backward nested family of descriptors (BNFD) from \(P_j\) to \(P_{j-k}\). The space of all BNFDs from \(P_j\) to \(P_{j-k}\) is given by

\[ T_{j,k} = \{ f = \{p^{j-l}\}_{l=0}^k : p^{l-1} \in S_{p^l}, l = j - k + 1, \ldots, j \} \subseteq \prod_{l=0}^k P_{j-l}. \]

For \(k \in \{1, \ldots, m\}\), given a BNFD \(f = \{p^{m-l}\}_{l=0}^k\)

\[ \pi_f = \pi_{p^m-k+1, p^m-k} \circ \ldots \circ \pi_{p^m, p^{m-1}} : p^m \to p^{m-k} \]

which projects along each descriptor. For another BNFD \(f' = \{p'^{j-l}\}_{l=0}^k\) \(\in T_{j,k}\) set

\[ d(f, f') = \sum_{l=0}^k d_j(p^{j-l}, p'^{j-l})^2. \]

In case of PNS, the nested projection \(\pi_f\) is illustrated in Figure 2 (a).

**Definition 2.2.** Random elements \(X_1, \ldots, X_n \overset{i.i.d.}{\sim} X\) on a data space \(Q\) admitting BNFDs give rise to backward nested population and sample means (abbreviated as BN means)

\[ \{E^{f_j} : j = 0, \ldots, m\}, \quad \{E^{f_{0}^n} : j = 0, \ldots, m\} \]
Recursively defined via $E^m = \{Q\} = E^m_n$, i.e. $p^m = Q = p^m_n$ and

$$E^{f_{j-1}} = \arg\min_{s \in S_{f_{j-1}}} \mathbb{E}[\rho_p(\pi_{f_j} \circ X, s)^2],$$

$$E^{f_{n-1}}_n = \arg\min_{s \in S_{f_{n-1}}} \sum_{i=1}^n \rho_{p_k}(\pi_{f_j} \circ X_i, s)^2,$$

where $p^j \in E^{f_j}$ and $p^n_j \in E^{f_{n-1}}$ is a measurable choice for $j = 1, \ldots, m$.

We say that a BNFD $f = \{p^k\}_{k=0}^m$ gives unique BN population means if $E^{f_j} = \{p^j\}$ with $f^j = \{p^k\}_{k=j}$ for all $j = 0, \ldots, m$.

Each of the $E^{f_{j-1}}$ and $E^{f_{n-1}}_n$ is also called a generalized Fréchet mean.

Note that by definition there is only one $p^m = Q \in P_m$. For this reason, for notational simplicity, we ignore it from now on and begin all BNFDs with $p^{m-1}$ and consider thus the corresponding $T_{m-1,k}$.

**Definition 2.3** (Factoring Charts). Let $j \in \{0, \ldots, m-1\}$, $k \in \{1, \ldots, j\}$. If $T_{j,k}$ and $P^{j-k}$ carry smooth manifold structures near $f' = (p^{j-k}, \ldots, p^j) \in T_{j,k}$ and $p^{j-k} \in P^{j-k}$, respectively, with open $W \subset T_{j,k}$, $U \subset P^{j-k}$ such that $f' \in W$, $p^{j-k} \in U$, and with local charts

$$\psi : W \to \mathbb{R}^{\dim(W)}, \ f = (p^{j-k}, \ldots, p^j) \mapsto \eta = (\theta, \xi), \ \phi : U \to \mathbb{R}^{\dim(U)}, \ p^{j-k} \mapsto \theta$$

we say that the chart $\psi$ factors, cf. Figure 1 (a) and (b), if with the projections

$$\pi^{P^{j-k}} : T_{j,k} \to P^{j-k}, \ f \mapsto p^{j-k}, \ \pi^{\dim(U)} : \mathbb{R}^{\dim(W)} \to \mathbb{R}^{\dim(U)}, \ (\theta, \xi) \mapsto \theta$$

we have

$$\phi \circ \pi^{P^{j-k}}|_W = \pi^{\dim(U)}|_{\psi(W)} \circ \psi.$$

![Fig 1: Factoring charts.](image)

2.2. Example: The Intrinsic Mean on a First Principal Component Geodesic for Manifolds

Suppose that $X_1, \ldots, X_n \sim X$ are random variables assuming values on a Riemannian manifold $Q$ with Riemannian norm $\|\cdot\|$ for the tangent spaces $T_qQ$ ($q \in Q$), induced metric $d : Q \times Q \to [0, \infty)$, projective
tangent bundle $PQ = \{(q, \{v, -v\}) : q \in Q, v \in T_qQ, \|v\| = 1\}$ and space of classes of geodesics given by their point sets

$$P_1 = \{[\gamma_{q,v}] : (q, \{v, -v\}) \in PQ\}, \quad [\gamma_{q,v}] = \{\gamma_{r,w} : \gamma_{q,v}(t) = r, \; \dot{\gamma}_{q,v}(t) = \gamma_{r,w}(t) = w\; \text{for some } t\}$$

where $t \mapsto \gamma_{q,v}(t)$ denotes the unique maximal geodesic through $q = \gamma_{q,v}(0)$ with unit speed velocity $v = \dot{\gamma}_{q,v}(0)$, $\|v\| = 1$. Then consider

$$P_2 = \{Q\}, \quad S_Q = P_1, \quad P_0 = Q.$$ 

There is a well defined distance between a point $s \in Q$ and a class of geodesics determined by

$$\rho_Q : Q \times P_1 \to [0, \infty), \quad (s, [\gamma_{q,v}]) \mapsto \inf_t d(s, \gamma_{q,v}(t)).$$

Then every class of geodesics determined by

$$\arg\min_{(q,v) \in TKQ} \rho(X, [\gamma_{q,v}])^2 \quad \text{or} \quad \arg\min_{(q,v) \in TKQ} \sum_{k=1}^n \rho(X_k, [\gamma_{q,v}])^2$$

is called a first population principal component geodesic or a first sample principal component geodesic, respectively, cf. Huckemann and Ziezold (2006).

Moreover, given a first population principal component geodesic $p = [\gamma_{q,v}]$ and a first sample principal component geodesic $\gamma_{n} = [\gamma_{q,v,n}]$, with the orthogonal projection

$$\pi_{Q,p} : Q \to p = [\gamma_{q,v}], \quad q' \mapsto \arg\min_{\gamma_{q,v}(t)} d(q', \gamma_{q,v}(t))$$

which is well defined outside a set of zero Riemannian volume, e.g. Huckemann et al. (2010, Theorem 2.6), we have the intrinsic population means on $p$ and intrinsic sample means on $\gamma_{n}$ determined by

$$\arg\min_{[\gamma_{q,v}]} \mathbb{E} \left[ \rho(X, [\gamma_{q,v}])^2 \right] \quad \text{or} \quad \arg\min_{[\gamma_{q,v}]} \sum_{k=1}^n \rho(X_k, [\gamma_{q,v}])^2,$$

respectively, where $\rho_p(q, q') = d(q, q')$ for $q, q' \in p$. In particular, we have the space of backward nested descriptors

$$T_{1,1} = \{(s, p) : p = [\gamma_{q,v}] \in P_1, \; s \in p\}$$

which carries the natural manifold structure of the projective tangent bundle $PQ$ conveyed by the identity

$$T_{1,1} \to PQ, \quad (s, [\gamma_{q,v}]) \mapsto (s, \{w, -w\})$$

where $w = \dot{\gamma}_{q,v}(t)$, $\|w\| = 1$, if $s = \gamma_{q,v}(t)$.

Recall that the tangent bundle $TQ = \{(q, v) : q \in Q, v \in T_qQ\}$ admits local trivializations, i.e. every $q \in Q$ has a local neighborhood $U \subset Q$ with a smooth one-to-one mapping

$$\tau = (\tau_1, \tau_2) : TU \to U \times \mathbb{R}^{\dim(Q)}$$

where the first coordinate satisfies $\tau_1(q', v') = q'$ for all $v' \in T_qQ$, $q' \in U$ and the second coordinate $\tau_2$ is a vector space isomorphism. In consequence, for a given $(q, v) \in PQ$, with local charts $\phi : U \to \mathbb{R}^{\dim(Q)}$ of $Q$ around $q$, and $\chi : H \to \mathbb{R}^{\dim(Q) - 1}$ of the real projective space $P\mathbb{R}^{\dim(Q) - 1}$ of dimension $\dim(Q) - 1$ around $\{\tau_2(q, v), -\tau_2(q, v)\} \in H \subset \mathbb{R}^{\dim(Q) - 1}$, $H$ open, and the open set

$$W = \left\{ (q', \{v', -v'\}) : (q', v') \in PU, \{\tau_2(q', v'), -\tau_2(q', v')\} \in H \right\} \subset PQ,$$

the mapping

$$\psi : W \to \mathbb{R}^{\dim(Q)} \times \mathbb{R}^{\dim(Q) - 1}, \quad (q', \{v', -v'\}) \mapsto \left( \phi(q'), \chi\{\tau_2(q', v'), -\tau_2(q', v')\} \right)$$

yields a local chart that factors as in Definition 2.3. This scenario is sketched in Figure 1 (c).
3. Assumptions for the Main Results

In this section we are back in the general scenario described in Section 2.1. We develop a set of assumptions, among others establishing additional compatibility for the families \(d_j\) and \(\phi_p\), necessary for the general results on asymptotic consistency and asymptotic normality in Section 4.1. We then show that they are fulfilled in case of PNS/PNGS and the IMoIGPC of Kendall’s shape spaces.

3.1. Assumptions for Strong Consistency

For the following assumptions suppose that \(j \in \{1, \ldots, m-1\}\).

**Assumption 3.1.** For a random element \(X\) in \(Q\), assume that \(\mathbb{E}[\rho_p(\pi_f \circ X, s)^2] < \infty\) for all BNFDs \(f\) ending at \(p', s \in S_{p'}\).

In order to measure a difference between \(s \in S_p\) and \(s' \in S_{p'}\) for \(p, p' \in P_j\) define the orthogonal projection of \(s \in S_p\) onto \(S_{p'}\) as

\[
S_{p'}^s = \operatorname{argmin}_{s' \in S_{p'}} d_j^{-1}(s, s').
\]

In case of PNS this is illustrated in Figure 2 (a).

**Assumption 3.2.** For every \(s \in S_p\) there is \(\delta > 0\) such that

\(\operatorname{S}_{p'}^s\) has exactly one element

whenever \(p, p' \in P_j\) with \(d_j(p, p') < \delta\).

---

(a) Nested projection

(b) Projection of descriptors

Fig 2: PNS illustration. Left: Projection of \(X\) (red diamond) in \(Q = \mathbb{S}^2\) onto small circle \(p\) (blue) and further onto \(s\) (blue disk). Right: Projection \(s''\) (blue) onto \(S_{p'}\) (which is \(p'\) in this case) of \(s\) (red) on \(S_p\) (which is \(p\) in this case).
For $s \in S_p$ and $p, p' \in P_j$ sufficiently close let $s^{p'} \in S_p^{p'}$ be the unique element. Note that we have in general for $(s^{p'})^p$, which denotes the projection back to $S_p$ of $s \in S_p$, after it was first projected to $S_p'$,

$$(s^{p'})^p \neq s.$$ 

In the following assumption, however, we will require that they will uniformly not differ too much if $p$ is close to $p'$. Also, we require that $s^{p'}$ and $s$ be close.

**Assumption 3.3.** For $\epsilon > 0$ there is $\delta > 0$ such that

$$d_{j-1}(s^{p'}, s) < \epsilon \text{ and } d_{j-1}((s^{p'})^p, s) < \epsilon \quad \forall s \in S_p$$

whenever $p, p' \in P_j$ with $d_j(p, p') < \delta$.

We will also require the following assumption, which, in conjunction with Assumption 3.3, is a consequence of the triangle inequality, if $d_{j-1}$ is a metric.

**Assumption 3.4.** Suppose that $d_j(p_n, p) \to 0$ and $d_{j-1}(s_n, s) \to 0$ with $p, p_n \in P_j$ and $s \in S_p, s_n \in S_{p_n}$. Then also

$$d_{j-1}(s_n, s^{p_n}) \to 0$$

Moreover, we require uniformity (which, among others guarantees the Ziezold strong law (4)) and coercivity (which, among others prohibits that sample means “diverge to infinity”, thus additionally guaranteeing the Bhattacharya-Patrangenaru strong law, which in our setup is Theorem 4.1) in the following senses.

**Assumption 3.5.** For all $\epsilon > 0$ there are $\delta_1, \delta_2 > 0$ such that

$$|\rho_p(\pi_f(q), s) - \rho_{p'}(\pi_{f'}(q), s')| < \epsilon \quad \forall q \in Q$$

for all BNFDs $f, f' \in T_{m-1,m-j-1}$ ending in $p, p' \in P_j$, respectively, with $d(f, f') < \delta_1$ and $s \in S_p, s' \in S_{p'}$ with $d_{j-1}(s, s') < \delta_2$.

**Assumption 3.6.** If $p_n, p \in P_j$ and $s_n \in S_{p_n}, s \in S_p$ with $d_{j-1}(s_n, s) \to \infty$, then for every $C > 0$ we have that

$$\rho_{p_n}(\pi_{f_n}(q), s_n) \to \infty$$

for every $q \in Q$ with $\rho_p(\pi_{f}(q), s) < C$ and BNFDs $f, f_n \in T_{m-1,m-j-1}$ ending at $p, p_n$ respectively.

**Remark 3.7.** Due to continuity, Assumptions 3.1 and 3.5 hold if $Q$ is compact and Assumption 3.6 if each $P_j$ is compact.

The proof of the following Theorem can be found in the supplement.

**Theorem 3.8.** Assumptions 3.1 – 3.6 hold for PNS and PNGS for all $j = 1, \ldots, m - 1$. Moreover, each $P_j$ is $d_j$-Heine Borel for $j = 0, \ldots, m$.

**Theorem 3.9.** For iMo1GPCs on Kendall’s shape spaces $Q = \Sigma^k_m$, $0 < m < k$, Assumptions 3.1 – 3.6 hold for $j = 1$. Moreover, $P_j$ is $d^j$ Heine-Borel for $j = 0, 1$.

**Proof.** Assumption 3.2 follows at once from the compactness of $Q$, hence the geodesics $p \subset Q$ are also compact, and the proof of *Huckemann et al. (2010, Theorem A.5)*, as there, in Claim II, a neighborhood of a geodesic $p$ is constructed, restricted to which the orthogonal projection $\pi_p$ is well defined and continuous in $p$. Compactness and continuity also imply Assumptions 3.3 and 3.4. Assumptions 3.1, 3.5 and 3.6 follow from Remark 3.7.
3.2. An Additional Assumption for Asymptotic Normality.

Again, let \( j \in \{1, \ldots, m-1\} \).

Assumption 3.10. Assume that \( T_{m-1, m-j} \) carries a smooth manifold structure near the unique BN population mean \( f^{j-1} = (p^{j-1}, \ldots, p^{m-1}) \) such that there is an open set \( W \subset T_{m-1, m-j} \), \( f^{j-1} \in W \) and a local chart

\[
\psi : W \rightarrow \mathbb{R}^{\dim(U)} \quad f^{j-1} = (p^{j-1}, \ldots, p^{m-1}) \mapsto \eta,
\]

and we set \( \eta' = \psi(f^{j-1}) \). Further, assume that for every \( l = j, \ldots, m \) the mapping

\[
\eta \mapsto f^{l-1} \mapsto \rho_l (\pi_l \circ X, p^{l-1})^2 := \tau_l(\eta, X),
\]

here the first map \( \eta \mapsto f^{l-1} \) is \( \psi^{-1} \) followed by a projection to \((p^{l-1}, \ldots, p^{m-1})\), has first and second derivatives, such that for all \( l = j, \ldots, m \),

\[
\text{Cov}[\text{grad}_{\eta'} \tau_l(\eta', X)], \quad \text{and} \quad \mathbb{E}[\text{Hess}_{\eta} \tau_l(\eta', X)]
\]

exist and are in expectation continuous near \( \eta' \), i.e. for \( \delta \rightarrow 0 \) we have

\[
\mathbb{E} \left[ \sup_{\|\eta - \eta'\| < \delta} \| \text{grad}_{\eta'} \tau_l(\eta, X) - \text{grad}_{\eta'} \tau_l(\eta', X) \| \right] \rightarrow 0,
\]

\[
\mathbb{E} \left[ \sup_{\|\eta - \eta'\| < \delta} \| \text{Hess}_{\eta} \tau_l(\eta, X) - \text{Hess}_{\eta} \tau_l(\eta', X) \| \right] \rightarrow 0.
\]

Finally, assume that the vectors \( \mathbb{E}[\text{grad}_{\eta} \tau^{j+1}(\eta', X)], \ldots, \mathbb{E}[\text{grad}_{\eta} \tau^m(\eta', X)] \) are linearly independent.

Remark 3.11. For PNS and PNGS a global, manifold structure has been derived in Section 6.1 with projections (8) (see also Proposition 6.2) and distances (12) smooth away from singularity sets. For IMo1GPCs on Kendall’s shape spaces, this has been provided in Section 6.2, cf. also Huckemann (2011, Theorem 2.2), where for planar space shapes, smoothness of the distance \( \rho \) between a shape and a geodesic has been established if this distance is less than \( \pi/4 \).

In general, however, it is unclear under which circumstances (if the second derivatives are continuous in both arguments where \( X \) is supported in a compact set, then convergence to zero holds not only in expectation but also a.s.) the three assumptions above, uniqueness, existence of first and second moments of second and first derivatives and their continuity in expectation are valid. Even for the much simpler case of intrinsic means on manifolds this is only very partially known, cf. the discussion in Huckemann and Hotz (2013). It seems that only for the most simple non-Euclidean case of intrinsic means on circles the full picture is available (Hotz and Huckemann (2015)). Recently, rather generic conditions for densities have been derived by Bhattacharya and Lin (2016), ensuring \( \sqrt{n} \)-Gaussian asymptotic normality.

The condition on linear independence is rather natural for realistic scenarios where each constraining condition adds a new constraint, not covered by the previous, as introduced after Corollary 4.3. For example, if charts factor, then with decreasing \( l \), every constraining condition results in conditions on new coordinates.

4. The Main Results

4.1. Asymptotic Theorems

Theorem 4.1. Let \( k \in \{0, \ldots, m-1\} \) and consider random data \( X_1, \ldots, X_n \overset{iid}{\sim} X \) on a data space \( Q \) admitting BN descriptor families from \( P_m \) to \( P_k \), unique BN population means \( \{p^k, \ldots, p^m\} \) and BN sample
means \( \{ E_{nk}^j, \ldots, E_{nm}^j \} \) due to a measurable selection \( p_{n}^j \in E_{nk}^j \) giving rise to BNFDs \( f_n^j = \{ p_{n}^j \}_{j=k}^m \), \( j = k, \ldots, m \). If Assumptions 2.1 – 3.6 are valid for all \( j = k, \ldots, m-1 \), and every \( \cup_{n=1}^{\infty} E_{nk}^j \) is a.s. \( d_j \)-Heine Borel (\( j = k, \ldots, m \)) then \( E_{nk}^j, \ldots, E_{nm}^j \) converges a.s. to \( \{ p_k^j, \ldots, p_m^j \} \) in the sense that \( \exists \Omega' \subset \Omega \) measurable with \( \Pr(\Omega') = 1 \) such that for all \( j = k, \ldots, m, \epsilon > 0 \) and \( \omega \in \Omega' \), \( \exists N = \Omega(\epsilon, \omega) \) with

\[
\bigcup_{r=n}^{\infty} E_{nk}^r \subset \{ p \in P_j : d_j(p^j, p) \leq \epsilon \} \quad \forall n \geq N, \ \omega \in \Omega'.
\]

**Proof.** We proceed by backward induction on \( j \). The case \( j = m \) is trivial and the case \( j = m - 1 \) has been covered by Theorems A.3 and A.4 from Huckemann (2011).

Now suppose that (3) have been established for \( j + 1 \in \{ k+1, \ldots, m \} \). Set \( P = P_{j+1}, p = p_{j+1}^j, f = \{ p_{j+1}^j, \ldots, p_m^m \}, p_n = p_{n+1}^{j+1}, f_n = \{ p_{n+1}^{j+1}, \ldots, p_m^m \} \) and for an arbitrary BNFD \( f' \) ending at \( p' \in P_{j+1} \)

\[
F_f(s) = \mathbb{E}[\rho_p(\pi_f \circ X, s)^2], \quad s \in S_p \quad F_n.f(s) = \frac{1}{n} \sum_{i=1}^{n} \rho_n(\pi_{f_n} \circ X_i, s)^2, \quad s \in S_{p_n} \quad \ell_f = \inf_{s \in S_p} F_f(s), \quad \ell_n.f = \inf_{s \in S_{p_n}} F_n.f(s)
\]

Then, \( F_f(s) < \infty \) for all \( s \in S_p \), by hypothesis, and with \( s^* = p^j \),

\[
\{ s^* \} = \arg\min_{s \in S_p} F_f(s), \quad E_{n,f_n}^j = \arg\min_{s \in S_{p_n}} F_n.f_n(s).
\]

To complete the proof we first show in the supplement

\[
\bigcap_{n=1}^{\infty} \bigcup_{r=n}^{\infty} E_{rk}^r \subset \{ s^* \} \text{ a.s. .}
\]

This is Ziezold’s version of strong consistency (cf. Ziezold (1977)). Further, we show that this implies the Bhattacharya-Patrangenaru version (cf. Bhattacharya and Patrangenaru (2003)) of strong consistency which takes here the form (3).

**Remark 4.2.** Careful inspection of the proof yields that we have only used that the “distances” \( d_j(p, p') = 0 \) for all \( p \in P_j \); they need not be definite, i.e. it is not necessary that \( d_j(p, p') = 0 \Rightarrow p = p' \).

Moreover, note that the \( d_j \)-Heine Borel property holds trivially in case of unique sample descriptors.

The following Corollary rests on a standard argument involving Chebyshev’s inequality, the proof is detailed in the supplement

**Corollary 4.3.** Suppose that (3) holds together with Assumption 3.10. Then we have for \( l = j, \ldots, m \) the following convergence in probability

\[
\frac{1}{n} \sum_{k=1}^{n} \text{grad}_n r^{l}(\eta_n, X_k) \xrightarrow{p} \mathbb{E} \left[ \text{grad}_n r^{l}(\eta', X) \right], \quad \frac{1}{n} \sum_{k=1}^{n} \text{Hess}_n r^{l}(\eta_n, X_k) \xrightarrow{p} \mathbb{E} \left[ \text{Hess}_n r^{l}(\eta', X) \right].
\]

Here, \( \eta_n \) denotes the image of \( f_n^{-1} \) in the chart from Assumption 3.10.

We now introduce notation we use for the central limit theorem. Let \( j \in \{ 1, \ldots, m \} \). By construction, every measurable selection \( f_n^{-1} \) of BN sample means minimizes

\[
\frac{1}{n} \sum_{k=1}^{n} \rho_{p_n}(\pi_{f_n} \circ X_k, p_n^{-1})^2
\]
under the constraints that it minimizes each of
\[ \frac{1}{n} \sum_{k=1}^{n} \rho_Q(X_k, p_n^{m-1})^2, \quad \ldots, \quad \frac{1}{n} \sum_{k=1}^{n} \rho_{p_{n}^{j+1}}(\pi_{f_{j+1}} \circ X_k, p_{n}^{j})^2. \]

Similarly, the BN population mean \( f^{j-1} \) minimizes
\[ \mathbb{E} [\rho_{\mathbb{P}}(\pi_{f_{j}} \circ X, p^{j-1})^2] \]
under the constraints that it minimizes each of
\[ \mathbb{E} [\rho_{\mathbb{P}}(\pi_{f_{j}} \circ X, p^{m-1})^2], \quad \ldots, \quad \mathbb{E} [\rho_{\mathbb{P}}^{j+1}(\pi_{f_{j+1}} \circ X, p^{j})^2]. \]

In consequence, due to differentiability guaranteed by Assumption 3.10, with the notation of \( \tau^j \) there, suitable random Lagrange multipliers \( \lambda_n^{j+1}, \ldots, \lambda_n^{m} \in \mathbb{R} \) and deterministic Lagrange multipliers \( \lambda^{j+1}, \ldots, \lambda^{m} \in \mathbb{R} \) exist such that for \( \eta_n = \psi^{-1}(f_{n}^{j-1}) \) and \( \eta' = \psi^{-1}(f^{j-1}) \) the following hold
\[
\begin{align*}
\text{grad}_{\eta} G_n(\eta_n) &= 0 \quad \text{with} \quad G_n(\eta) := \frac{1}{n} \sum_{k=1}^{n} \tau^j(\eta, X_k) + \sum_{l=j+1}^{m} \lambda_n^{l} \frac{1}{n} \sum_{k=1}^{n} \tau^l(\eta, X_k) \quad \text{(5)} \\
\text{grad}_{\eta} G(\eta') &= 0 \quad \text{with} \quad G(\eta) := \mathbb{E}[\tau^j(\eta, X)] + \sum_{l=j+1}^{m} \lambda^l \mathbb{E}[\tau^l(\eta, X)] \quad \text{(6)}
\end{align*}
\]

The following Corollary rests, among others, on the linear independence of the nested constraining conditions from Assumption 3.10. Its proof is detailed in the supplement.

**Corollary 4.4.** Suppose that (3) holds together with Assumption 3.10. Then the random Lagrange multipliers \( \lambda_n^{j+1}, \ldots, \lambda_n^{m} \) in (5) and \( \lambda^{j+1}, \ldots, \lambda^{m} \) in (6) satisfy
\[ \lambda_n^l \xrightarrow{p} \lambda^l \quad \text{for} \quad l = j + 1, \ldots, m. \]

Having dealt in the two preceding corollaries with the complications caused by nestedness, the following central limit theorem can now be reduced to a standard Taylor expansion argument as in Bhattacharya and Patrangenaru (2005); Huckemann (2011a). The proof is detailed in the supplement.

**Theorem 4.5.** Let \( j \in \{1, \ldots, m - 1\} \) and consider random data \( X_1, \ldots, X_n \overset{iid}{\sim} X \) on a data space \( Q \) admitting BNFD\( s \) from \( P_{m-1} \) to \( P_{j-1} \), a unique BN population mean \( f^{j-1} = \{p^{j-1}, \ldots, p^{m-1}\} \) and BN sample means \( \{E_{n}^{l-1}, \ldots, E_{n}^{m-1}\} \) due to a measurable selection \( p_{n}^{l} \in E_{n}^{l}, \quad f_{n}^{j-1} = \{p_{n}^{j-1}, \ldots, p_{n}^{m-1}\}, \quad l = j - 1, \ldots, m - 1. \)

(i) Assuming that Assumption 3.10 hold as well as (3) for all \( j \in \{j - 1, \ldots, m - 1\} \), we have that
\[ \sqrt{n} H_\psi(\psi(f_{n}^{j-1}) - \psi(f^{j-1})) \rightarrow N(0, B_\psi) \]
with a chart \( \psi \) as specified in Assumption 3.10 as well as
\[ H_\psi = \mathbb{E} \left[ \text{Hess}_{\eta} \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \text{Hess}_{\eta} \tau^l(\eta', X) \right] \]
and
\[ B_\psi = \text{Cov} \left[ \text{grad}_{\eta} \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \text{grad}_{\eta} \tau^l(\eta', X) \right], \]
with the notation from Assumption 3.10 where \( \lambda_{j+1}, \ldots, \lambda_m \in \mathbb{R} \) are suitable such that

\[
\nabla \eta \mathbb{E}[\tau^i(\eta, X)] + \sum_{l=j+1}^m \lambda_l \nabla \eta \mathbb{E}[\tau^i(\eta, X)]
\]

vanishes at \( \eta = \eta' \).

(ii) If additionally \( H_\psi > 0 \), then \( f^{j-1}_n \) satisfies a Gaussian \( \sqrt{n} \)-CLT

\[
\sqrt{n}(\psi(f^{j-1}_n) - \psi(f^{j-1})) \xrightarrow{\text{d}} N(0, \Sigma_\psi), \quad \Sigma_\psi = H^{-1}_\psi B_\psi H^{-1}_\psi.
\]

(iii) If additionally the chart \( \psi \) factors as in Definition 2.3, then also \( p^{j-1}_n \) satisfies a Gaussian \( \sqrt{n} \)-CLT

\[
\sqrt{n}(\phi(p^{j-1}_n) - \phi(p^{j-1})) \xrightarrow{\text{d}} N(0, \Sigma_\phi), \quad \Sigma_\phi = (\Sigma_{\psi_{ik}})_{i,k=1}^{\dim(P_j)}
\]

with the notation of Definition 2.3.

### 4.2. A Nested Two-Sample Bootstrap Test

Suppose that we have two independent i.i.d. samples \( X_1, \ldots, X_n \sim X \in Q \), \( Y_1, \ldots, Y_m \sim Y \in Q \) in a data space \( Q \) admitting BNFDs and we want to test

\[
H_0 : X \sim Y \quad \text{versus} \quad H_1 : X \not\sim Y
\]

using descriptors in \( p \in P \). Here, \( p \in P \) stands either for a single \( p_j \in P_j \) for which we have established factoring charts, or for a suitable sequence \( f \in T_{j,k} \). We assume that the first sample gives rise to \( \hat{p}_n^X \in P \), the second to \( \hat{p}_m^Y \in P \), and that these are unique. Under the corresponding assumptions of Theorem 4.5, define a statistic

\[
T^2(A) = \left( \phi(\hat{p}_n^X) - \phi(\hat{p}_m^Y) \right)^T A \left( \phi(\hat{p}_n^X) - \phi(\hat{p}_m^Y) \right).
\]

Under \( H_0 \) and the assumptions in Theorem 4.5, including invertibility of the Hessians, up to a suitable factor, this is Hotelling \( T^2 \) distributed if \( A^{-1} \) exists and is the corresponding empirical covariance matrix. Therefore, for \( A^{-1} \) we use the empirical covariance matrix from bootstrap samples. Notably, under additional regularity assumptions, this estimator is asymptotically consistent, cf. (Cheng, 2015, Corollary 1).

With this fixed \( A \), we simulate a slightly changed statistic under \( H_0 \), by again bootstrapping from \( X_1, \ldots, X_n \) (for \( b = 1, \ldots, B \)) and \( Y_1, \ldots, Y_m \) (for \( b' = 1, \ldots, B' \)),

\[
T^{*2}(A) = \left( \left( \phi(\hat{p}_{n,b}^X) - \phi(\hat{p}_{m,b}^Y) \right) - \left( \phi(\hat{p}_{m,b}^X) - \phi(\hat{p}_{m,b}^Y) \right) \right)^T A \left( \left( \phi(\hat{p}_{n,b}^X) - \phi(\hat{p}_{m,b}^Y) \right) - \left( \phi(\hat{p}_{m,b}^X) - \phi(\hat{p}_{m,b}^Y) \right) \right),
\]

to obtain higher power. From these, for a given level \( \alpha \in (0, 1) \) we compute the empirical quantile \( c_{1-\alpha} \) such that

\[
\mathbb{P}\{T^{*2}(A) \leq c_{1-\alpha} | X_1, \ldots, X_n, Y_1, \ldots, Y_m \} = 1 - \alpha.
\]

Then, in consequence of (Arcones and Giné, 1992, Theorems 3.2 and 3.5), asymptotic normality of \( \sqrt{n}(\phi(\hat{p}_n^X) - \phi(\hat{p}_n^X)) \), and \( \sqrt{m}(\phi(\hat{p}_n^Y) - \phi(\hat{p}_m^Y)) \), guaranteed by Theorem 4.5, extends to the same asymptotic normality for \( \sqrt{n}(\phi(\hat{p}_{n,b}^X) - \phi(\hat{p}_{m,b}^X)) \), and \( \sqrt{m}(\phi(\hat{p}_{n,b}^Y) - \phi(\hat{p}_{m,b}^Y)) \), respectively. We have then under \( H_0 \) that \( c_{1-\alpha} \) gives an asymptotic coverage of \( 1 - \alpha \) for \( T^2(A) \), i. e. \( \mathbb{P}\{T^{*2}(A) \leq c_{1-\alpha} \} \to 1 - \alpha \) as \( B = n, B' = m \to \infty \) if \( n/m \to c \) with a fixed \( c \in (0, \infty) \).

We note that also the argument from Bhattacharya and Patrangenaru (2005, Corollary 2.3 and Remark 2.6) extends at once to our setup, as we assume that the corresponding population covariance matrix \( \Sigma_\psi \) or \( \Sigma_\phi \), respectively, from Theorem 4.5 is invertible.
5. Applications

5.1. Simulations

To illustrate our CLT for principal nested spheres (PNS) and principal nested great spheres (PNGS), we simulate three data sets, each from two paired random variables $X$ and $Y$, displayed in Figure 3.

I) Data on an $S^3$ concentrate on the same proper small $S^2$ and there on segments of orthogonal great circles such that their nested means are antipodal.

II) Data on an $S^3$ concentrate on the same proper small $S^2$ and there on segments of orthogonal great circles such that their nested means coincide.

III) Data on an $S^2$ concentrate on segments of different small circles, have different nested means under PNS, but, under PNGS, coinciding principal geodesics and nested means.

![Data set I](image1)
![Data set II](image2)
![Data set III](image3)

Fig 3: Simulated datasets I (left) and II (middle) on $S^3$ concentrate on a common proper small $S^2$, their projections to estimated small two-spheres is depicted. The simulated dataset III (right) is on $S^2$.

We apply PNS and PNGS to the simulated data and perform the nested two-sample bootstrap test for identical respective nested submanifolds (means, small and great circles) and for identical small and great two-spheres, from Section 4.2. The distributions of resulting p-values are displayed in Figure 4. These values are in perfect agreement with the intuition guiding the design of the data.

In the supplement, Section 6.6, we report further simulations, in particular that for small and moderate spread, bootstrap sample variances follow indeed an asymptotic $1/n$ law. For larger spread, it is unclear whether the assumptions of Theorem 4.5 are met. For example for the intrinsic mean on general manifolds, uniqueness can only be guaranteed if, among others, the data are supported within a geodesic half ball, cf. Afsari (2011), and only for the circle the full picture is known, cf. Hotz and Huckemann (2015). Recall that on Kendall’s planar shape spaces, smoothness of the distance $\rho$ between a shape and a geodesic has been established if this distance is less than $\pi/4$, cf. (Huckemann, 2011, Theorem 2.2). The greater picture, verifying these assumptions more generally, in particular for higher order descriptors remains challenging research.

5.2. Early Human Mesenchymal Stem Cell Differentiation

Understanding differentiation of adult human stem cells with the perspective of clinical use (see e.g. Pittenger et al. (1999) who emphasize their potential for cartilage and bone reconstruction) is an ongoing fundamental challenge in current medical research, still with many open questions (e.g. Bianco et al. (2013)). To investigate mechanically guided differentiation, human mesenchymal stem cells (hMSCs, pluripotent adult stem
cells taken from the bone marrow) are placed on gels of varying elasticity, quantified by the Young’s modulus, to mimic different environments in the human body, e.g. Discher et al. (2005). It is well known that within the first day the surrounding elasticity measured in Kilopascal (kPa) induces differentiation through biomechanical cues, cf. Engler et al. (2006); Zemel et al. (2010), where the changes manifest in orientation and ordering of the actin-myosin filament skeleton. In particular, in order to direct future, more focused research, it is of high interest to more precisely identify time intervals in which such changes of ordering occur and to separate changes due to differentiation from changes due to other causes.

**Experimental setup.** We compare hMSC skeletons that have been cultured at the Third Institute of Physics of the University of Göttingen on gels with Young’s moduli of 1 kPa mimicking neural tissue, 10 kPa mimicking muscle tissue, and 30 kPa mimicking bone tissue. The cells have been fixed after multiples of 4 hours on the respective gel and have then been immuno-stained for NMM IIa, the motor proteins making up small filaments that are responsible for cytoskeletal tension and imaged (as described in Zemel et al. (2010)). In particular every cell is observed only once guaranteeing independence. Table 1 shows their sample sizes and...
the data will be published and made available after completion of current research, cf. Wollnik and Rehfeldt (2016). Because earlier research (Huckemann et al. (2016)) suggests that during the first 24 hours, 10 kPa and 30 kPa hMSCs develop rather similarly and quite differently from 1 kPa hMSCs, for this investigation, we pool the former.

The actin-myosin filament structure has been automatically retrieved from the fluorescence images using the Filament Sensor from Eltzner et al. (2015). Since neighboring filaments share the same orientation, the 3D structure of the cellular skeleton can be retrieved by separating the filament structure into different orientation fields, cf. Figure 5.

Orientation fields for filament structures are determined via a relaxation labeling procedure, see Rosenfeld et al. (1976). The source code of our implementation is available as supplementary material. A detailed description is deferred to a future publication. The algorithm results in a set of contiguous areas with slowly
Table 1

Sample sizes of hMSC skeleton images over varying Young’s moduli and cultivation time.

<table>
<thead>
<tr>
<th>Time</th>
<th>1 kPa</th>
<th>10 kPa and 30 kPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>4h</td>
<td>159</td>
<td>321</td>
</tr>
<tr>
<td>8h</td>
<td>163</td>
<td>317</td>
</tr>
<tr>
<td>12h</td>
<td>176</td>
<td>344</td>
</tr>
<tr>
<td>16h</td>
<td>135</td>
<td>274</td>
</tr>
<tr>
<td>20h</td>
<td>138</td>
<td>253</td>
</tr>
<tr>
<td>24h</td>
<td>166</td>
<td>304</td>
</tr>
</tbody>
</table>

Data analysis. For each single hMSC image, let $M$ be the number of pixels of all detected filaments, $m_1$ the number of all filament pixels of filaments of the largest orientation field and $m_2$ the number of all filament pixels of filaments of all smaller orientation fields. $M - m_1 - m_2$ is then the number of pixels in all “rogue” filaments which are not associated to any field, because they are too inconsistent with neighboring filaments. Define $x = (x_1, x_2, x_3) := (\sqrt{m_1/M}, \sqrt{m_2/M}, \sqrt{1 - (m_1 + m_2)/M})^T \in Q = S^2$ where the square roots ensure that $x$ does not describe relative areas but rather relative diameters of fields. This representation is confined to the $S^2$ part in the first octant and every sample shows a distinct accumulation of points in the $x_2 = 0$ plane, corresponding to cells with only one orientation field. Figure 6 shows typical data representations. As common with biological data, especially from primary cells, their variance is rather high. In consequence, great circle fits are more robust under bootstrapping than small circle fits and we use the nested two-sample bootstrap tests for PNGS with the following null hypothesis.

$H_0$: hMSC orientation and ordering measured by random loci on $S^2$ as above does not change between successive time points.

Table 2

Displaying $p$-values of the nested two-sample bootstrap tests for PNGS of filament orientation field distribution data. The test uses $B = 1000$ bootstrap samples, therefore the penultimate $p$-value is $10^{-3}$.

<table>
<thead>
<tr>
<th>Time</th>
<th>nested great circle mean</th>
<th>jointly great circle and nested mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 kPa</td>
<td>10 kPa and 30 kPa</td>
</tr>
<tr>
<td>Gel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4h vs. 8h</td>
<td>0.120</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>8h vs. 12h</td>
<td>$&lt; 10^{-3}$</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>12h vs. 16h</td>
<td>0.126</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>16h vs. 20h</td>
<td>0.468</td>
<td>0.626</td>
</tr>
<tr>
<td>20h vs. 24h</td>
<td>$&lt; 10^{-3}$</td>
<td>$&lt; 10^{-3}$</td>
</tr>
</tbody>
</table>

Results. As visible in Table 2, while for hMSCs on harder gels (10 kPa and 30 kPa), nested means and the joint descriptor of nested mean and great circle change over each 4 hour interval until 16 hours – for both the null hypothesis is rejected at the highest level possible – similar changes are less clearly visible for hMSCs on the soft gel (1 kPa) between the intervals between 8 and 16 hours and not at all visible for the first time interval. Strikingly, for hMSCs on all gels, no changes seem to occur between 16 and 20 hours. In contrast, in the final interval between 20 and 24 hours, nested means and great circles clearly change for hMSCs on the soft gel – rejecting the null hypothesis at the highest level possible. This effect is also there for the nested mean of hMSCs on the harder gels, but not as clearly visible for the joint descriptor including the circle.

Visualization in Figure 7 reveals further details. As seen from the loci of the nested means, hMSCs on
Fig 6: Representing samples of cells by their relative orientation field shares, grown for 12 hours (top row) and 16 hours (bottom row) on gels of different elasticity, with best approximating great circle (blue line) and nested mean (red star).

Table 3
Displaying p-values of nested two-sample bootstrap tests for equality of PNGS of filament orientation field distributions of 1 kPa cells with those of 10 and 30 kPa cells, for all time points. We use B = 1000 bootstrap samples, thus the penultimate p-value is $10^{-3}$.

<table>
<thead>
<tr>
<th>Gels</th>
<th>1 kPa vs. 10 kPa and 30 kPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>nested great circle mean</td>
</tr>
<tr>
<td>4 h</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>8 h</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>12 h</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>16 h</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>20 h</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>24 h</td>
<td>0.010</td>
</tr>
</tbody>
</table>

the soft gel (Figure 7a) tend to lose minor orientation field filaments with a nearly constant ratio of large orientation field filaments and rogue filaments until the critical slot, the time interval between 16 and 20
hours. Their great circles, indicating the direction of largest spread, change at the beginning of the critical slot, suggesting that the major variation there occurs in the amount of rogue filaments. While, until the critical slot, the temporal motion of nested means for 1 kPa is mainly vertical, the corresponding motion for the hMSCs on harder gels (cf. Figure 7b) is horizontal, indicating that the number of rogue filaments decreases in favor of the main orientation field. Curiously for the nested means, there is also a sharp drop in height at the beginning of the critical slot as well as a backward horizontal motion. After the critical slot, hMSCs seem to continue the direction of their previous journey, at a lower smaller fields’ level, though. In contrast, for the hMSCs on the soft gel, the critical slot seems to represent a true change point since afterward, the nested mean travels not much longer towards reducing the smaller fields, but like hMSCs on harder gels, mainly reduces the number of rogue filaments. Indeed, taking into account the auxiliary mesh lines, it can be seen that descriptors are rather close at time 24 hours, cf. Table 3, where, in contrast they are rather far away from each other for all other time points.

Discussion. We conclude that hMSCs react clearly distinctly and differently on both gels already for short time intervals, where at the critical time slot some kind of reboot happens. A generic candidate for this effect is cell division. As all cells used in the experiments were thawed at the same time (72 hours before seeding) and treated identically, cell division is expected to occur at similar (at least for each environment) time points. Dividing cells completely reorganize their cell skeleton which would explain the change point found. In particular, it seems that due to cell division, the time point 24 hours (as used in Zemel et al. (2010)) may not be ideal if differences in hMSCs differentiation due to different Young’s moduli are to be detected. Our results clearly warrant further analysis using higher time resolution, in particular time resolved in-vivo imaging, that among others, allow to register cell division times.
Acknowledgment

We thank Rabi Bhattacharya and Vic Patrangenaru for their valuable comments on the bootstrap, Peter Jupp for his comments on the multinomial test and our collaborators Florian Rehfeldt and Carina Wollnik for their stem cell data.

References


Huckemann, S., T. Hotz, and A. Munk (2010a). Intrinsic MANOVA for Riemannian manifolds with an
application to Kendall’s space of planar shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(4), 593–603.


Wollnik, C. and F. Rehfeldt (2016). Quantitative live-cell analysis of BM-hMSCs on elastic substrates during early differentiation. manuscript.


6. Supplement to “Backward Nested Descriptors Asymptotics with Inference on Stem Cell Differentiation” by Stephan Huckemann and Benjamin Eltzner

6.1. The Geometric Framework of PNS and PNGS

The (nested) projections detailed below are illustrated in Figure 2.

**Notation.** Consider the standard inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^{m+1}$ with norm $\|x\| = (x, x)^{1/2}$ and the $m$-dimensional unit sphere $Q = S^m = \{x \in \mathbb{R}^{m+1} : \|x\| = 1\}$, with interior $\mathbb{D}^{m+1} = \{x \in \mathbb{R}^{m+1} : \|x\| < 1\}$. For any matrix $v = (v_1, \ldots, v_k) \in \mathbb{R}^{m \times k}$ we have the Frobenius norm $\|v\| = \sqrt{\sum_{j=1}^k \|v_j\|^2}$ and the inner product $\langle v, w \rangle = \text{trace}(vw^T)$ for $v, w \in \mathbb{R}^{n \times k}$. The $k \times k$ dimensional unit matrix is $I_k$ and $O(k) = \{R \in \mathbb{R}^{k \times k} : R^T R = I_k\}$ is the orthogonal group. Moreover, e.g. James (1976, p. 1),

$$O(k, m) = \left\{ v = (v_1, \ldots, v_k) \in \mathbb{R}^{m \times k} : \langle v_i, v_k \rangle = \delta_{ik}, \; 1 \leq i \leq j \leq k \right\} \cong O(m)/O(m - k),$$

denotes the Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^m$. For every such orthonormal $k$-frame we have a non-unique orthonormal complement $\tilde{v} \in O(m - k, m)$ such that $(v, \tilde{v}) \in O(m)$.

In PNS and PNGS, for a top sphere $Q = S^m$ a sequence of nested subspheres is sought for. Only in the scenario of PNS it is required that each subsphere is a great subsphere. In general, every $j$-dimensional subsphere $(j = 1, \ldots, m - 1)$ is the intersection of a $j + 1$ dimensional affine subspace of $\mathbb{R}^{m+1}$ with $S^m$. Recall that every $j + 1$ dimensional affine subspace

$$A = \{ x \in \mathbb{R}^{m+1} : \langle x, v_k \rangle = \alpha_k, \; k = 1, \ldots, m - j \}$$

is determined by a matrix

$$v = (v_1, \ldots, v_{m-j})$$

determined by a matrix $v$ of $m + 1 - (j + 1) = m - j$ orthonormal column vectors that are orthogonal to $A$ and a vector of signed distances from the origin

$$\alpha = (\alpha_1, \ldots, \alpha_{m-j})^T.$$

In particular $\|\alpha\| < 1$ ensures that $A$ intersects with $S^m$ in a $j$ dimensional subsphere. Obviously, $A$ determines $v$ and $\alpha$ up to an action of $O(m - j)$, i.e. with every $R \in O(m - j)$, $vR$ and $R^T \alpha$ determine the same $A$.

In PNGS only great subspheres are allowed as intersections. Hence all affine spaces under consideration above pass through the origin, i.e. $\alpha = 0$ above.

**The parameter space.** In consequence we have that the family of $j$-dimensional subspheres $(j = 1, \ldots, m - 1)$ is given by the smooth manifold

$$P_j = M_j/O(m - j) = \{ [z] : z \in M_j \} \text{ with } [z] = \{ zR : R \in O(m - j) \}$$

where

$$M_j = \begin{cases} O(m - j, m + 1) \times \mathbb{D}^{m-j} = \left\{ \begin{pmatrix} v \\ \alpha^T \end{pmatrix} : v \in O(m - j, m + 1), \alpha \in \mathbb{D}^{m-j} \right\} & \text{for PNS} \\ O(m - j, m + 1) \times \{0\} & \text{for PNGS} \end{cases}$$

with $\{0\} \subset \mathbb{R}^{m-j}$ above, for compatibility, becoming clear in the considerations below. Indeed, the smooth action from the right of the compact Lie group $O(m - j)$ on $M_j$ is free (for $v \in O(m - j, m + 1)$ and $R \in O(m - j)$, $vR = v$ implies that $R = I_{m-j}$), giving rise to a smooth quotient manifold, e.g. Lee (2013, Theorem 7.10), of dimension $(j + 2)(m - j)$ for PNS and of dimension $(j + 1)(m - j)$ for PNGS. Notably, in the latter case, $P_j$ is just a Grassmannian.
In the above setup, we had excluded the cases $j = 0, m$. For $j = m$, for PNS and PNGS it is natural to set, as in Definition 2.1,

$$P_m = \{S^m\}.$$ 

In case of $j = 0$, the setup above would yield pairs of points (the intersections of suitable lines with $S^m$ give topologically zero-dimensional spheres $S^0$). In order to have a single nested mean as a zero dimensional descriptor only, for PNS and PNGS (in order to represent nestedness) we use the convention

$$P_0 = S^m \text{ i.e. } M_0 = O(m, m + 1) \times S^{m-1} \text{ with } S^{m-1} = \{\alpha \in \mathbb{R}^m : \|\alpha\| = 1\}.$$ 

**Distance between subspheres of equal dimension.** For PNS and PNGS, on $M_j$ we have the *extrinsic metric*

$$(z, z') \mapsto \|z - z'\|.$$

Since the extrinsic metric is invariant under the action of $O(m - j)$ it gives rise to the well defined quotient metric, called the *Ziezold metric*, cf. Ziezold (1994)) on $P_j$ given by

$$d_j(p, p') = \min_{R \in O(m-j)} \|z - z'R\|$$

for arbitrary representatives $z, z' \in M_j$ of $p, p' \in P_j$, respectively, as the following Lemma teaches.

**Lemma 6.1.** The mapping $d_j : P_j \to [0, \infty)$ satisfies the triangle inequality and it is definite, i.e. $d_j(p, p') = 0$ implies $p = p'$.

**Proof.** Suppose that $z = (v^T, \alpha)^T, z' = (v'^T, \alpha')^T, z'' \in M_j$ are representatives of $p, p', p'' \in P_j$ with the property, w.l.o.g., that $d_j(p, p') = \|z - z'\|, d_j(p, p'') = \|z - z''\|$. Then the usual triangle inequality yields

$$d_j(p', p') \leq \|z' - z''\| \leq \|z' - z\| + \|z - z'\| = d_j(p', p) + d_j(p, p'').$$

Moreover, we have

$$\|z - z'R\|^2 = 2(m - j) + \|\alpha\|^2 + \|\alpha'\|^2 - 2\text{trace}(v^Tv'R) - 2\text{trace}(\alpha^T R^T \alpha')$$

where $\text{trace}(v^Tv') \leq m - j$ and $\text{trace}(\alpha^T R^T \alpha') \leq \|\alpha\| \|\alpha'\|$ with equality if and only if $v = v'R$ and $\alpha = cR^T \alpha'$ with $c \geq 0$, yielding that the above vanishes if and only if this is the case with $c = 1$. \hfill \Box

**Backwards nesting.** In PNS, the space of subspheres $S_p \subset P_{j-1}$ within a given subsphere $p = [v^T, \alpha]^T \in P_j$ ($j = 2, \ldots, m$) can be given the following structure.

$$S_p = \left\{ \begin{bmatrix} u,v_{m-j+1} \\ \alpha^T, \alpha_{m-j+1} \end{bmatrix} : v_{m-j+1} \in S^m, \alpha_{m-j+1} \in \mathbb{R} \text{ with } v_{m-j+1}^Tv = 0, \alpha_{m-j+1}^2 < 1 - \|\alpha\|^2 \right\} \subset P_{j-1}.$$ 

Indeed, if $(R^Tv^T, R^T \alpha)^T, R \in O(m - j)$ is another representative for $p$, then $v_{m-j+1}^Tv = 0$ if and only if $v_{m-j+1}^TvR = 0$. In case of PNGS, the condition on entries of $\alpha$ above is simply $\alpha_{m-j+1} = 0$ because $\alpha = 0$.

According to our convention of $P_0 = S^m$, in case of $j = 1$ the inequality above needs to be replaced with $\alpha_{m-j+1}^2 = 1 - \|\alpha\|^2$ (for PNS) and with $\alpha_{m-j+1}^2 = 1$ for (PNGS).

**Projections.** For all $j = 0, \ldots, m - 1$ we have the intrinsic orthogonal projection onto $p = [v^T, \alpha]^T \in P_j$

$$\pi_{S^m, p_j} : S^m \to p_j, \quad q \mapsto y = v\alpha + \sqrt{1-\|\alpha\|^2} \frac{(I_{m+1} - v^Tv)q}{\|I_{m+1} - v^Tv\|} = v\alpha + \sqrt{1-\|\alpha\|^2} \frac{v^Tvq}{\|v^Tv\|}$$

(8)
which is independent of the representative \((v^T, \alpha)^T\) chosen and independent of the specific orthogonal complement \(\bar{v} \in O(j+1, m+1)\) of \(v\) chosen; and it is well defined except for a set
\[
\left\{ q \in S^m : q = \sum_{i=1}^{m-j} (q_i v_i) \right\}
\]
of spherical measure zero. Note that we have \(\|\alpha\| = 1\) for \(j = 0\) and hence the constant mapping \(y \mapsto \alpha \).

**Nested projections.** More generally, if \(p^j, p^{j'}\) are from a family of backward nested subspaces \(S^m \supset p^{m-1} \supset \ldots \supset p^0, 0 \leq j' < j \leq m - 1\), we may choose representatives \((v^T, \alpha)^T\) of \(p^j\) and \((v'^T, \alpha')^T\) of \(p^{j'}\) such that
\[
v = (v_1, \ldots, v_{m-j}), \quad v' = (v_1, \ldots, v_{m-j'})
\]
\[
\alpha^T = (\alpha_1, \ldots, \alpha_{m-j}), \quad \alpha'^T = (\alpha_1, \ldots, \alpha_{m-j'})
\]

With (8) and the arbitrary but fixed complement \(\bar{v}\) of \(v\) chosen, embed \(p^j\) in \(\mathbb{R}^{j+1}\) (first a translation, possible in PNS, and then a blow up), depending on the specific choice of \(\bar{v}\), via
\[
g_{p^j, \bar{v}} : p^j \to \mathbb{R}^{j}, \quad y \mapsto z = \frac{\bar{v}^T y}{\|\bar{v}^T y\|} = \frac{\bar{v}^T q}{\|\bar{v}^T q\|} \quad \text{for} \quad y = \pi_{S^m, p^j}(q), \quad g_{p^j, \bar{v}}(z) = v\alpha + \sqrt{1 - \|\alpha\|^2} \bar{v} z. \quad (9)
\]

Now, \(g_{p^j, \bar{v}}\) embeds \(p^j\) as a \(j'\)-dimensional (possibly small) subsphere \(p_{j'}^j = g_{p^j, \bar{v}}(p^j) \subset S^j\) in \(\mathbb{R}^{j+1}\), given by \([w^T, \beta]^T = p_{j'}^j\), for suitable \((w^T, \beta)^T \in O(j-j', j+1) \times \mathbb{D}^{j'-j}\), and there is an orthogonal complement \(\bar{w} \in O(j+1, j+1)\) of \(w\), such that
\[
v' = (v, \bar{v} w), \quad \bar{v}' = v\bar{w}, \quad \alpha'^T = (\alpha^T, \beta^T \sqrt{1 - \|\alpha\|^2}) \quad \text{and} \quad (v', \bar{v}') \in O(m+1). \quad (10)
\]

This gives the definition of the projection, independent of the specific orthogonal complements chosen,
\[
\pi_{p^j, p^{j'}} = g_{p_{j'}^j, \bar{v}}^{-1} \circ \pi_{S^j, p_{j'}^j} \circ g_{p^j, \bar{v}} : p^j \to p^{j'}, \\
y \mapsto v\alpha + \sqrt{1 - \|\alpha\|^2} \bar{v} \left( w\beta + \sqrt{1 - \|\beta\|^2} \frac{\bar{w}w^T v^T y}{\|\bar{v}^T q \|} \right) \\
= v\alpha' + \sqrt{1 - \|\alpha'\|^2} \frac{\bar{v} v'^T y}{\|\bar{v}'^T \|}. \quad (11)
\]

Plugging in (8) into the above equality and taking into account that \(\bar{v}^T y/\|\bar{v}^T y\| = \bar{v}^T q/\|\bar{v}^T q\|\), cf. (9), yields at once the following proposition which asserts that projections along nested subspaces only depend on the final subsphere at which it ends. Recall that \(\alpha = 0\) (= \(\beta\) if \(j' > 0\)) in case of PNS.

**Proposition 6.2.** With the above notation \(\pi_{p^j, p^{j'}} \circ \pi_{S^m, p^j} = \pi_{S^m, p^{j'}}\).

**Distance between projected data and next subsphere.** Here, we compute the intrinsic geodesic distance \(\rho_{p^j}(y, p^{j-1})\) between \(y\) and \(p^{j-1} = [v^T, \alpha]^T \in S^j\) in a subsphere \(p^j = [v^T, \alpha]^T \in P_j\) \((j = 1, \ldots, m)\). Note that only in case of \(p^j\) being a great subsphere, this distance agrees with the spherical distance arccos \((y^T \pi_{p^j, p^{j-1}}(y))\) in the top sphere \(S^m\). If \(p^j\) is a proper subsphere \((0 < \|\alpha\| < 1)\), assuming w.l.o.g. that \(v' = (v, v_{m-j+1})\) and \(\alpha'^T = (\alpha^T, \alpha_{m-j+1})^T\) we have
\[
\rho_{p^j}(y, p^{j-1}) \\
= \sqrt{1 - \|\alpha\|^2} \arccos \left( \frac{\langle g_{p^j, \bar{v}}(y) \rangle^T g_{p^j, \bar{v}} \circ \pi_{p^j, p^{j-1}}(y) \rangle}{1 - \|\alpha\|^2} \left( y^T v_{m-j+1} \alpha_{m-j+1} + \sqrt{1 - \|\alpha'\|^2} y^T (I_{m+1} - v' v'^T) y \right) \right). \quad (12)
\]
Indeed with \( j' = j - 1 \) and the notation from (8) to (11) and orthogonal complements \( \tilde{v} = (v_{m-j+1}, \tilde{v}') \) of \( v \) and \( \tilde{v}' \) of \( v' \), respectively, we have

\[
(g_{\tilde{p}'_j, \tilde{p}'_j}(y))^T g_{\tilde{p}'_j, \tilde{p}'_j} \circ \pi_{\tilde{p}'_j, \tilde{p}'_j}(y) = \frac{y^T \tilde{v}}{||\tilde{v}||} \left( \tilde{v}' \alpha' + \sqrt{1 - ||\alpha'||^2} \tilde{v}' \alpha y \right) ||\tilde{v}' y|| \left( \tilde{v}' \alpha' + \sqrt{1 - ||\alpha'||^2} \tilde{v}' y \right)
\]

since \( ||\tilde{v}' y|| = \sqrt{1 - ||\alpha'||^2} \) due to (8), \( y^T \tilde{v} \tilde{v}' y = y^T v_{m-j+1} \alpha_{m-j+1} + y^T \tilde{v} \tilde{v}' y = y^T \tilde{v}' y \).

**Optimal positioning.** On \( M_j \) we have the extrinsic metric due to its embedding in \( \mathbb{R}^{(m+1)\times(m-j)} \times \mathbb{R}^{m-j} \), cf. (7), and its thus induced Riemannian metric. Obviously \( O(m-j) \) acts on \( M_j \) isometrically w.r.t. the Riemannian metric because the action also preserves geodesics on \( O(m-j, m) \), e.g. Edelman et al. (1998, p. 309). In consequence, for both metrics, we say for given \( z, z' \in M_j \) that \( R \in O(m-j) \) puts \( z \) i.o.p. (in optimal position) to \( z' \), if

\[
||z' - zR|| = \min_{R' \in O(m-j)} ||z' - zR'||.
\]

**Lemma 6.3.** Let \( j \in \{0, \ldots, m-1\} \) and \( z, z' \in M_j \) be sufficiently close. Then, there is a unique \( R_{z,z'} \in O(m-j) \) such that

(i) \( R_{z,z'} \) puts \( z \) uniquely i.o.p. \( zR_{z,z'} \) to \( z' \),

(ii) if \( z^{\top} z \) is symmetric then \( R_{z,z'} = I_{m-j} \).

**Proof.** In order that \( R \in O(m-j) \) minimizes the r.h.s of (7), given by \( ||z' - zR||^2 = ||z'||^2 + ||z||^2 - 2 \text{trace}(AR) \), with \( A = z^{\top} z \), it maximizes \( \text{trace}(AR) \). This is so if \( R = QS^T \) for a singular value decomposition (svd) \( SAQ^T = A \) stemming from a spectral decomposition \( AA^T = SA^2S^T \). Since for \( z = z' = (v^{\top}, \alpha')^T \) we have that \( A = I_{m-j} + \alpha'^{\top} \alpha' \) is of full rank for all \( \alpha' \in \mathbb{R}^{m-j} \), there is locally a full rank svd, which is unique up to \( S \mapsto SL \) and \( Q = A^T \Lambda^{-1} \mapsto A^T SLA^{-1} \) for any \( L \in O(m-j) \) with \( LL^T = \Lambda \). However, \( R_{z,z'} = A = A^T = Q \Lambda S^T \) allows to choose \( Q = S \), i.e. \( R_{z,z'} = I_{m-j} \).

This Lemma has the following immediate consequence.

**Corollary 6.4.** If \( z, z' \in M_j \) are sufficiently close then

\[
z, z' \text{ are i.o.p. } \iff z^{\top} z' - z'^{\top} z = 0.
\]

Thus, in further consequence, for every \( z' \in M_j \) there is \( \epsilon > 0 \) such that

\[
U_{z'} = \{ z \in M_j : z^{\top} z' - z'^{\top} z = 0, ||z - z'|| < \epsilon \} \tag{13}
\]

is a smooth manifold and every \( p' \in P_j \) has a neighborhood \( V \) with a smooth diffeomorphism

\[
V \to U_{z'}^z, \quad p \mapsto z = zR_{z,z'} \tag{14}
\]

where \( z' \in M_j \) is a fixed representative of \( p' \) and \( z \in M_j \) is an arbitrary representative of \( p \), and \( z \) is a representative i.o.p to \( z' \). This follows from the fact that locally a point is i.o.p to \( z' \) if and only if it can be reached by a horizontal geodesic from \( z' \), and from the fact that all geodesics in \( P_j \) through \( p' \) lift to horizontal geodesics in \( M_j \) through \( z' \) and all horizontal geodesics in \( M_j \) project to geodesics in \( P_j \). For a detailed discussion e.g. Huckemann et al. (2010).

**Representing spaces of nested subspaces: Factoring charts.** For \( j \in \{1, \ldots, m-1\} \) recall

\[
T_{j,1} = \{(s,p) : p \in P_j, s \in S_p\},
\]
and that if \( y = (v^T, \alpha)^T \in M_j \) is a representative of \( p \in P_j \), then every \( s \in S_p \) is represented by some

\[
z = \left( v, v_{m-j+1} \right) \in M_{j-1}
\]

with suitable choices \( v_{m-j+1} \in S^m \), \( v^T v_{m-j+1} = 0 \) and \( \alpha_{m-j+1}^2 < 1 - \|\alpha\|^2 \). For PNGS, \( j > 1 \), \( \alpha_{m-j+1} = 0 = \|\alpha\| \) (for PNGS, \( j > 1 \)) and \( \alpha_{m-j+1}^2 = 1 - \|\alpha\|^2 \). Vice versa, from every representative \( z = (w^T, \beta)^T \in M_{j-1} \) of \( s \in S_p \), all of the \( p \in P_j \) with \( s \in S_p \) are determined by choosing \( m - j \) orthonormal vectors from the column span of \( w \) (the columns of \( wB \) below) with suitable \( m - j \) distances (given by the vector \( \beta^T B \) below), i.e. every such \( p \) has a representation

\[
p = [zB]
\]

as \( B \) ranges over \( O(m - j, m - j + 1) \). In case of PNGS and \( j = 1 \), for compatibility we set \( (0, \ldots, 0, 1)B = (0, \ldots, 0) \). Of course, different \( B \) may give same the \( p \). More precisely, here is the central observation.

Given \( s = [z] \in P_{j-1} \), every \( p \in P_j \) with \( s \in S_p \) is uniquely determined as \( p = [zB] \) by choice of \( b \in S^{m-j} \), where \( B \) is an arbitrary orthonormal complement of \( b \) in \( O(m - j + 1) \), i.e. \((B, b) \in O(m - j + 1) \). This means that \( S^{m-j} \) uniquely parametrizes \( \{ p \in P_j : s \in S_p \} \). The reason is the well known fact that the Grassmannian \( O(m - j, m - j + 1)/SO(m - j) = SO(m - j + 1)/SO(m - j) \) is diffeomorphic with \( S^{m-j} \). Again, in case of PNGS and \( j = 1 \), for compatibility we set \((0, \ldots, 0, 1)B = 0 \).

This observation gives rise to factoring charts as introduced in Definition 2.3. Here and below in Theorem 6.6, \( U_{j-1}^\epsilon \) assumes the role of \( U \) from Definition 2.3.

**Lemma 6.5.** Every \((p', s') \in T_{j, 1}\) has a neighborhood \( W \subset T_{j, 1} \) and a smooth diffeomorphism

\[
\phi : W \rightarrow U_{j-1}^\epsilon \times V, \quad (p, s) \mapsto (z, b)
\]

where \( U_{j-1}^\epsilon \) is from (13), \( V \subset S^{m-j} \) is a neighborhood of \((0, \ldots, 0, 1)^T \in R^{m-j+1} \) and

\[
z' = \left( v', v'_{m-j+1} \right) \quad \text{and} \quad \left( v' \right)
\]

are representatives of \( s' \) and \( p' \) respectively. In particular,

\[
s = [z] \quad \text{and} \quad p = [zB]
\]

with arbitrary orthonormal complement \( B \) of \( b \).

**Proof.** Let \( s = [\tilde{z}] \) for some \( \tilde{z} \in M_j \) be sufficiently close to \( s' \). Then, according to Lemma 6.3, \( s \) has a unique representative \( z = \tilde{z}R_{\tilde{z}, z'} \) in optimal position to \( z' \), varying smoothly in \( \tilde{z} \), hence smoothly in \( s \). Moreover, as elaborated above, every \( p \in P_j \) with \( s \in S_p \) depends uniquely upon the choice of \( b \in S^{m-j} \) and any orthonormal complement \( B \), i.e. \( p = [zB] \). Although \(-b\) determines the same \( m - j \) dimensional subspace, a one-to-one mapping is obtained choosing the neighborhood \( V \) of \((0, \ldots, 0, 1) \) suitably small not containing antipodals. \( \square \)

The following is a straightforward generalization to spaces of sequences of several nested subspheres.

**Theorem 6.6** (Factoring Charts). Let \( j \in \{1, \ldots, m - 1\} \) and \( k \in \{1, \ldots, j\} \). Then every \((p^{j-k}, \ldots, p^j) \in T_{j, k} \) has a neighborhood \( W \subset T_{j, k} \) and a smooth diffeomorphism

\[
\phi : W \rightarrow U_{j-k}^\epsilon \times \prod_{r=1}^k V_r, \quad (p^{j-k}, \ldots, p^j) \mapsto (z, b_1, \ldots, b_k)
\]
where \( U_{j-k}^{t} \) is from (13), each \( V_r \subset S^{m-j+r-1} \) is a neighborhood of \((0, \ldots, 0, 1)^T \in \mathbb{R}^{m-j+r} \),

\[
z' = \begin{pmatrix} v'_1, \ldots, v'_{m-j+k} \\ \alpha'_1, \ldots, \alpha'_{m-j+k} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} v''_{m-j} \\ \vdots \\ \alpha''_{m-j} \end{pmatrix}
\]

are representatives of \( p^{j-k} \) and \( p^{j-r} \) respectively \((r = 1, \ldots, k)\). In particular,

\[
p^{j-k} = [z] \quad \text{and} \quad p^{j-r+1} = [z_k \cdots B_r]
\]

with arbitrary orthonormal complements \( B_r \) in \( O(m-j+r) \) of \( b_r \in V_r \) \((r = 1, \ldots, k)\).

A joint representation. Here we assume that \( f' = (p'^{m-1}, \ldots, p'^0) \) and \( f = (p^{m-1}, \ldots, p^0) \) are two sufficiently close families of backward nested spheres with representatives

\[
p^j = \begin{pmatrix} v'_1, \ldots, v'_{m-j} \\ \alpha'_1, \ldots, \alpha'_{m-j} \end{pmatrix}, \quad p^j = \begin{pmatrix} v_1, \ldots, v_{m-j} \\ \alpha_1, \ldots, \alpha_{m-j} \end{pmatrix}, \quad j = 0, \ldots, m
\]

such that

\[
\begin{pmatrix} v'_1, \ldots, v'_{m} \\ \alpha'_1, \ldots, \alpha'_{m} \end{pmatrix}, \quad \begin{pmatrix} v_1, \ldots, v_{m} \\ \alpha_1, \ldots, \alpha_{m} \end{pmatrix}
\]

are in o. \( O(m)\)-p. Notably, there are then uniquely determined \( SO(m+1) \) complements \( v'_m+1 \) and \( v_{m+1} \) respectively. For a sufficiently small neighborhood \( W \subset T_{m-1,m-1} \) of \( f' \) and neighborhoods \( V_r \subset S^r \) of \((0, \ldots, 0, 1)^T \in S^r \) \((r = 1, \ldots, k)\) we have then a smooth mapping

\[
W \rightarrow \prod_{r=1}^{m} V_r, \quad f \mapsto (b_1, \ldots, b_m)
\]

determined by the following algorithm (here is only the PNGS version). For arbitrary \( v \in \mathbb{R}^{\dim(v)}, \Pi_{v^\perp} : x \mapsto (I_{\dim(v)} - vv^T)x \) denotes the orthogonal projection to the complement of \( v \).

- \( b_m \in S^m \) is the unique element such that \( v_m = (v'_1, \ldots, v'_{m+1})b_m \).
- \( b_{m-1} \in S^{m-1} \) is the unique element such that \( v_{m-1} = (v'_1, \ldots, v'_{m})b_{m-1} \). Here, \( (v'_1, \ldots, v'_{m}) \) is obtained from \( (\Pi_{v^\perp}, v'_1, \ldots, \Pi_{v^\perp}v'_{m+1}) \) by removing one column such that \( v_1, \ldots, v_{m-1} \) is in the span of the rest (usually \( \Pi_{v^\perp}v'_{m} \)).
- \( \vdots \)
- \( b_1 \in S^1 \) is the unique element such that \( v_1 = (v'_{m-1}, v'_{m})b_1 \). Here, \( (v'_{m-1}, v'_{m}) \) is obtained from \( (\Pi_{v^\perp}, v'_{m-2}, \Pi_{v^\perp}v'_{m-2}, \Pi_{v^\perp}v'_{m-2}) \) by removing one column such that \( v_1 \) is in the span of the other two.

Local charts. For \( P_j \) and \( T_{j,k} \), we have derived in (13) and Theorem 6.6, respectively, local smooth diffeomorphic representations in a Euclidean space of suitable dimension, \( r \), say, that can be generically written as

\[
U = \{ x \in V : \Phi(x) = 0 \}
\]

with a neighborhood \( V \subset \mathbb{R}^r \) of some \( x' \in \mathbb{R}^r \) and a smooth mapping \( \Phi : V \rightarrow \mathbb{R}^s, r, s \in \mathbb{N}, r > s \) with full rank derivative \( d\Phi_{x'} \) at \( x' \). Here, the corresponding matrices are viewed as vectorized by stacking their columns on top of one another. With \( \Pi_{\mathbb{R}^sd\Phi_{x'}} \), the orthogonal projection to the row space of the \( s \times r \) matrix of the derivative at \( x' \in V \) and orthogonal \( r \)-vectors \( y_1, \ldots, y_r \), spanning the kernel of \( d\Phi_{x'} \) of dimension \( t = r - s \), obtain the local chart

\[
U \rightarrow \mathbb{R}^t, \quad x \mapsto (y_1, \ldots, y_t)(I_r - \Pi_{\mathbb{R}^sd\Phi_{x'}})(x - x').
\]
6.2. Intrinsic Mean on a First Geodesic Principal Component for Kendall's Shape Spaces

Another application is given by the intrinsic mean on a geodesic principal component on a quotient space $Q = M/G$ due to an isometric action of a Lie group $G$ on a Riemannian manifold $M$. We treat here the prominent application of Kendall’s shape spaces

$$Q = \Sigma^k_m = \mathbb{S}^{m(k-1)-1}/SO(m) = \{ [x] : x \in \mathbb{S}^{m(k-1)-1} \}, \quad [x] = \{ gx : g \in SO(m) \}$$

where the space $\mathbb{S}^{m(k-1)-1}$ of unit size $m \times (k-1)$ matrices – corresponding to normed and centered configuration of $k$ landmarks in $m$-dimensional Euclidean space – is taken modulo the group of rotations in $m$-dimensional space. This space models all $m$-dimensional configurations of $k$ landmarks, not all coinciding, modulo similarity transformations, cf. Dryden and Mardia (1998). For $m \geq 3$ this space is no longer a manifold but decomposes into strata of manifolds of different dimensions, cf. Kendall et al. (1999). Geodesics on the unit sphere $\mathbb{S}^{m(k-1)-1}$, i.e. great circles, orthogonal to the the orbits $[x]$, called horizontal great circles, project to geodesics in $Q = \Sigma^k_m$ such that the space $P_1$ of geodesics of $Q$ can be given the quotient structure of a stratified space

$$P_1 = O^H(2, m \times (k - 1))/ (SO(m) \times O(2))$$

with the horizontal Stiefel manifold

$$O^H(2, m \times (k - 1)) = \{ (x, v) \in \mathbb{R}^{m \times (k - 1)} \times \mathbb{R}^{m \times (k - 1)} : \text{trace}(x^T v) = 0 = \text{trace}(x^T x) - 1 = \text{trace}(v^T v) - 1, \; xv^T = vx^T \}$$

and the orbits

$$[x, v] = \{ (gx, gv)h : g \in SO(m), h \in O(2) \}$$

for $(x, v) \in O^H(2, m \times (k - 1))$, cf. Huckemann et al. (2010, Theorem 5.2). The action of $SO(m)$ is not free for $m \geq 3$, giving rise to a non-trivial stratified structure. As before in Section 2.2, we set

$$P_2 = \{ Q \}, \; P_0 = Q.$$ 

Having geodesics, orthogonal projections $\pi_{Q,p} : Q \to p, p \in P_1$ can be defined, which are unique outside a set of intrinsic measure zero, cf. Huckemann et al. (2010, Theorem 2.6). The geodesic distance $d$ on $Q$ gives rise to the geodesic distance $\rho_Q = \rho, \rho(q, p) = \min_{q' \in p} d(q, q')$ between a datum $q$ and a geodesic $p$. Similarly the induced intrinsic distance on $p$ gives rise to $\rho_p : p \times S_p \to [0, \infty)$ where $S_p = \{ s \in p : s \in Q \}$ can be identified with $p$. For $d_1$ and $d_0$, for simplicity, not the canonical intrinsic distances on $P_1$ and $P_0 = Q$ but quotient distances due to the embedding of $O^H(2, m \times (k - 1)) \to \mathbb{R}^{m \times (k - 1)} \times \mathbb{R}^{m \times (k - 1)}$ and $\mathbb{S}^{m(k-1)-1} \to \mathbb{R}^{m \times (k-1)}$, respectively, can be used, called the Ziezold distances, cf. Huckemann (2011).

The generalized Fréchet mean corresponding to $\rho$ is the first geodesic principal component (GPC) $p$, the generalized Fréchet mean corresponding to $\rho_p$ is the the intrinsic mean on the first GPC, cf. Huckemann et al. (2010); Huckemann (2011). The latter is again a nested mean. With the horizontal projective bundle over the unit sphere

$$P_1^H \mathbb{S}^{m(k-1)-1} = \bigcup_{x \in \mathbb{S}^{m(k-1)-1}} \{ x \} \times P_x^H \mathbb{S}^{m(k-1)-1}$$

$$P_x^H \mathbb{S}^{m(k-1)-1} = \{ [v, -v] : v \in T_x \mathbb{S}^{m(k-1)-1}, xv^T = vx^T, \text{trace}(v^T v) = 1 \}$$

we have the space

$$T_{1,1} = \{ (s, p) : p \in P_1, s \in S_p \} \cong \{ ([x, v], [x]) : x \in \mathbb{S}^{m(k-1)-1}, \{ v, -v \} \in P_x^H \mathbb{S}^{m(k-1)-1} \}. $$
The Principal Orbit Theorem (e.g. Bredon (1972, p. 199) states in particular, that \( P_0 = Q \) and \( P_1 \) have open and dense subsets \( Q^* \subset Q, P_1^* \subset P_1 \) that are manifolds, in our case smooth manifolds. This gives rise to the manifold

\[ T_{1,1}^* = \{(s, p) \in T_{1,1} : p \in P_1^*, s \in Q^*\} \]

with local smooth coordinates near \((p', s') \in W \subset T_{1,1}^*\)

\[ W \to P^*H \mathbb{S}^{m(k-1)-1}, \quad (s, p) \mapsto (x, \{v - v\}) \]

where \( x \in s \) and \((x, v)\) is a representative of \( p \) i.o.p. to \((x', v')\), under the action of \( SO(m) \times O(2)\), with trace\((v^Tv') > 0\). Here \((x', v')\) is an arbitrary but fixed representative of \( p' \) such that \( x' \) is a representative of \( s' \). Along the lines of Lemma 6.3 one can show that optimal positioning is unique if \( x'x^T + v'v^T \) has rank \( \geq m - 1 \), which may be assumed for most realistic data scenarios.

Arguing as in Section 2.2, with every local trivialization of the horizontal bundle

\[ HS^{m(k-1)-1} = \{(x, v) \in \mathbb{S}^{m(k-1)-1} \times \mathbb{R}^{m(k-1)} : trace(x^Tv) = 0 \text{ and } xv^T = vx^T\} \]

comes a factoring chart.

### 6.3. Proofs

#### 6.3.1. Proof of Theorem 3.8

Recall from Proposition 6.2 that \( \pi_f \) only depends on the final descriptor at which \( f \) ends. For PNS and PNGS we use the notation introduced in Section 2 and show first Assumption 3.2. Let \((v^T, \alpha)^T \) and \((v'^T, \alpha')^T \) be representatives of \( p, p' \in P_j \) in optimal position and \( s = [w^T, \beta]^T \in S_p \) with \( w = (v, v_{m-j+1}) \in O(m - j + 1, m + 1) \) and \( \beta^T = (\alpha^T, \alpha_{m-j+1}) \). Moreover consider a candidate element

\[ s^* = \begin{bmatrix} v', x \\ \alpha'^T, y \end{bmatrix} \in S_{p'}, \quad x \in \mathbb{S}^{m+1}, x^Tv' = 0, \quad \begin{cases} y \in (-1, 1), y^2 + \|o'\|^2 < 1, & \text{for PNS, } j > 1 \\ y = 0, & \text{for PNGS, } j > 1 \\ y \in (-1, 1), y^2 + \|o'\|^2 = 1, & \text{for PNS/PNGS, } j = 1 \end{cases} \]

and its squared distance to \( s \),

\[
d^2_f(s^*, s) = \min_{R \in O(m-j)} \left\| \begin{bmatrix} w \\ \beta^T \end{bmatrix} R - \begin{bmatrix} v', x \\ \alpha'^T, y \end{bmatrix} \right\|^2 = 2(m - j + 1) + \|\beta\|^2 + \|o'\|^2 + \|y\|^2 - 2 \max_{R \in O(m-j)} \text{trace}(HR)
\]

with

\[
H = \begin{pmatrix} v^Tv' + o'o'^T & v^T + \alpha o'y \\ v_{m-j+1} v' + \alpha_{m-j+1} o'^T & v_{m-j+1}^Tv + \alpha_{m-j+1} y \end{pmatrix}.
\]

With the continuous function \( F(o', o') = \min_{x, y} d^2_f(s^*, s) \), we have \( F(v, o) = 0 \), which, by Lemma 6.1, is uniquely assumed for \( x = v_{m-j+1} \) and \( y = \alpha_{m-j+1} \). Due to continuity, for \((o', o')\) sufficiently close to \((v, o)\) all minimizers \((x, y)\) of \( F(o', o') \) are in a neighborhood of \((v_{m-j+1}, \alpha_{m-j+1})\) and there, \( H \) is full rank and hence, arguing as in the proof of (i) of Lemma 6.3, the extremal \( R \in O(m - j) \) is uniquely determined. Let us write \( R = (B, b) \) to obtain

\[
d^2_f(s^*, s) = \min_{B \in O(m-j), m-j + 1} \left\| \begin{bmatrix} w \\ \beta^T \end{bmatrix} B - \begin{bmatrix} v' \\ \alpha'^T \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} w \\ \beta^T \end{bmatrix} b - \begin{bmatrix} x \\ y \end{bmatrix} \right\|^2.
\]
As $b$ is unique, minimizing the above for $x$ is equivalent to minimizing $\|wb - x\|^2$ under the constraining conditions $\|x\|^2 = 1$ and $v'^T x = 0$ which yields the necessary equation

$$x - wb + v' \lambda + \mu x = 0,$$

with suitable Lagrange multipliers $\lambda \in \mathbb{R}^{n-j}, \mu \in \mathbb{R}$.

Right multiplication with $v'^T$ yields at once $\lambda = v'^T wb$ such that $x$ is uniquely determined by

$$x = \frac{(I_{m+1} - v'^{v'^T})wb}{\|(I_{m+1} - v'^{v'^T})wb\|}.$$

Indeed, $x$ is well defined because it is in a neighborhood of $v_{m-j+1}$ and hence $b$ in a neighborhood of $(0, \ldots, 0)^T$. More simply, without Lagrange minimization, we obtain $y = \beta Tb$.

For $p, p'$ sufficiently close, this yields a unique $s^* = s^{p'}$ minimizing $d_j(s^*, s)$ over $s^* \in S_{p'}$, yielding at once Assumptions 3.2, 3.3 and 3.4 (because $d_j$ is a metric).

Due to Remark 3.7, Assumptions 3.1, 3.5 and 3.6 hold. Because each $P_j$ is a finite dimensional manifold and $d_j$ is a topologically compatible metric, $P_j$ is $d_j$-Heine Borel.

6.3.2. Completing the Proof of Theorem 4.1

We continue to use the notation introduced in the sketch of the proof right after Theorem 4.1, in particular, recall that $p$ is the last element in $f$. First we show a crucial Lemma.

**Lemma 6.7.** Fix $s \in S_p$. Then there is a measurable set $\Omega_s \subset \Omega$ with $\mathbb{P}(\Omega_s) = 1$ such that

$$F_{n, f_n}(s^{p_n}) \to F_f(s) \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \rho_{p_n}(\pi_{f_n} \circ X_i, s^{p_n}) \to \mathbb{E}[\rho_p(\pi_f \circ X, s)].$$

**Proof.**

$$F_{n, f_n}(s^{p_n}) = \frac{1}{n} \sum_{i=1}^{n} \rho_{p_n}(\pi_{f_n} \circ X_i, s^{p_n})^2 = \frac{1}{n} \sum_{i=1}^{n} \rho_p(\pi_f \circ X_i, s)^2 + \frac{2}{n} \sum_{i=1}^{n} \rho_p(\pi_f \circ X_i, s)h_i(f, f_n, s, s^{p_n}) + \frac{1}{n} \sum_{i=1}^{n} h_i(f, f_n, s, s^{p_n})^2$$

with

$$h_i(f, f_n, s, s^{p_n}) = \rho_{p_n}(\pi_{f_n} \circ X_i, s^{p_n}) - \rho_p(\pi_f \circ X_i, s)$$

which, in conjunction with Assumption 3.3 and the induction hypothesis $d(f, f_n) \to 0$ a.s. can be made arbitrary small a.s. due to Assumption 3.5. In consequence of the usual strong law the first assertion follows. The second follows with the same argument.

**Showing (4).**

Having established Lemma 6.7, in principle we can now follow the steps laid out by Ziezold (1977). They are, however, more intrigued in our endeavor. By hypothesis we have $d(f_n, f) \to 0$ a.s.

By separability of $P_j$ it follows at once from Lemma 6.7 that there is a measurable set $\Omega' \subset \Omega$ with $\mathbb{P}(\Omega') = 1$ and a dense subset $\{s_k : k \in \mathbb{N}\} \subset S_p$ such that

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_{p_n}(\pi_{f_n} \circ X_i, s^{p_n}) \to \mathbb{E}[\rho_p(\pi_f \circ X, s)] \right\} \quad \forall s \in \{s_k : k \in \mathbb{N}\} \text{ and } \omega \in \Omega'.$$  (16)
In order to obtain (16) for all \( s \in S_p \), consider \( s_n, s'_n \in S_{p_n} \) and the following estimates.

\[
|F_{n,f}(s'_n) - F_{n,f}(s_n)| \\
\leq \frac{1}{n} \sum_{i=1}^{n} (\rho_p(\pi_{f_n} \circ X_i, s'_n) + \rho_p(\pi_{f_n} \circ X_i, s_n)) [\rho_p(\pi_{f_n} \circ X_i, s'_n) - \rho_p(\pi_{f_n} \circ X_i, s_n)] \\
= \left\{ \frac{1}{n} \sum_{i=1}^{n} (2\rho_p(\pi_{f_n} \circ X_i, s'_n) + g_p(\pi_{f_n} \circ X_i, s_n, s'_n)) [g_p(\pi_{f_n} \circ X_i, s_n, s'_n)] \right\}
\]

with

\[
g_p(\pi_{f_n} \circ X_i, s^{p_n}) = \rho_p(\pi_{f_n} \circ X_i, s^{p_n}) - \rho_p(\pi_{f_n} \circ X_i, s^{p_n}).
\]

Now, w.l.o.g., consider \( s_k \to s \) (which implies that \( d_j(s_k, s) \to 0 \)) for which (16) is valid. Using twice the first line in (17) for \( s_n = s^{p_n} \) and \( s'_n = s'^{p_n} \) we obtain

\[
F_{n,f}(s^{p_n}) - \frac{1}{n} \sum_{i=1}^{n} (2\rho_p(\pi_{f_n} \circ X_i, s^{p_n}) + g_p(\pi_{f_n} \circ X_i, s^{p_n}, s^{p_n})) [g_p(\pi_{f_n} \circ X_i, s^{p_n}, s^{p_n})] \\
\leq F_{n,f}(s^{p_n}) \\
F_{n,f}(s^{p_n}) + \frac{1}{n} \sum_{i=1}^{n} (2\rho_p(\pi_{f_n} \circ X_i, s^{p_n}) + g_p(\pi_{f_n} \circ X_i, s^{p_n}, s^{p_n})) [g_p(\pi_{f_n} \circ X_i, s^{p_n}, s^{p_n})]
\]

Due to Assumption 3.5 and the strong law (16) (and the argument applied in the proof of Lemma 6.7), for every \( \epsilon > 0 \) there is \( K = K(\epsilon) \in \mathbb{N} \) such that for all \( k \geq K \) we have

\[
F_f(s_k) - 2(\mathbb{E}[\rho_p(\pi_f \circ X, s_k)]) + \epsilon \leq \liminf_{n \to \infty} F_{n,f}(s^{p_n}) \\
\leq \limsup_{n \to \infty} F_{n,f}(s^{p_n}) \\
\leq F_f(s_k) + 2(\mathbb{E}[\rho_p(\pi_f \circ X, s_k)]) + \epsilon
\]

for all \( \omega \in \Omega' \). Taking into account the continuity of \( F_f \), letting \( \epsilon \to 0 \) yields

\[
F_{n,f}(s^{p_n}) \to F_f(s) \quad \forall s \in S_p \text{ and } \omega \in \Omega' \quad \text{for } n \to \infty.
\]

Similarly we see that

\[
\frac{1}{n} \sum_{i=1}^{n} \rho_p(\pi_{f_n} \circ X_i, s^{p_n}) \to \mathbb{E}[\rho_p(\pi_f \circ X_i, s)] \quad \forall s \in S_p \text{ and } \omega \in \Omega' \quad \text{for } n \to \infty.
\]

Next, we consider a sequence \( S_{p_n} \ni s_n \to s \in S_p \). Note that in consequence of Assumption 3.4 we have that

\[
d_j(s_n, s^{p_n}) \to 0.
\]

Using the bottom line of (17) yields that

\[
|F_{n,f}(s_n) - F_{n,f}(s^{p_n})| \\
\leq \frac{1}{n} \sum_{i=1}^{n} (2\rho_p(\pi_{f_n} \circ X_i, s^{p_n}) + g_p(\pi_{f_n} \circ X_i, s_n, s^{p_n})) [g_p(\pi_{f_n} \circ X_i, s_n, s^{p_n})] \rightarrow 0 \quad \forall \omega \in \Omega'
\]
with the same $\Omega'$ for all $s \in S_p$ due to (19). Hence, in consequence of this and (18), for all $S_p \ni s_n \to s \in S_p$ we have that
\[
|F_{n,f_n}(s_n) - F_f(s)| \leq |F_{n,f_n}(s_n) - F_{n,f_n}(s^{p_n})| + |F_{n,f_n}(s^{p_n}) - F_f(s)| \to 0. \tag{20}
\]
for all $\omega \in \Omega'$.

Finally let us show
\[
\text{if } \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k^{f_k} \neq \emptyset \text{ then } \ell_{n,f_n} \to \ell_f \text{ a.s.} \tag{21}
\]
Note that Assertion (4) is trivial in case of $\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k^{f_k} = \emptyset$. Otherwise, for ease of notation let $B_n := \cup_{k=n}^{\infty} E_k^{f_k}$, $B_n \supset A := \cap_{n=1}^{\infty} B_n$, $b \in B$. Then $b \in B_n$ for all $n \in \mathbb{N}$. Hence, there is a sequence $b_n \in B_n$, $b_n \to b$. Moreover, there is a sequence $k_n$ such that $b_n = s_{k_n} \in E_{k_n}^{f_{k_n}}$ for a suitable $k_n \geq n$. Then $\ell_{n_k,f_{k_n}} = F_{n_k,f_{k_n}}(s_{k_n}) \to F_f(b) \geq \ell$ by (20) a.s.. On the other hand, by Lemma 6.7 for arbitrary fixed $s \in S_p$, there is a sequence $\epsilon_n \to 0$ such that $F_f(s) \geq F_{n,f_n}(s^{p_n}) - \epsilon_n \geq \ell_{n,f_n} - \epsilon_n$. First letting $n \to \infty$ and then considering the infimum over $s \in S_p$ yields
\[
\ell_f \geq \limsup_{n \to \infty} \ell_{n,f_n}.
\]
In consequence
\[
\limsup_{n \to \infty} \ell_{n,f_n} \geq F_f(b) \geq \ell_f \geq \limsup_{n \to \infty} \ell_{n,f_n} \text{ a.s.} \tag{22}
\]
In particular we have shown that $\ell_f = F_f(b)$ which means that $b = s^*$ thus completing the proof of (4)

**Proof of (3).** Using the notation of the previous proof of (4), let $s_n \in S_p$ and consider $r_n = d_j(s^*, s_n)$. If the assertion (3) was false, there would be a measurable set $A \subset \Omega$ with $\mathbb{P}(A) > 0$ such that for every $\omega \in A$ there is $r_0(\omega) > 0$ and $r_n(\omega) \geq r_0(\omega) > 0$.

First, we claim that $\mathbb{P}(B) = 0$ with $B = \{\omega \in A : s_n(\omega) \text{ has a cluster point}\}$.

For if $\omega \in B$ with $s_n(\omega) \to \tilde{s}(\omega) \in S_p$, then by continuity $d_j(s^*(\omega), \tilde{s}(\omega)) \geq r_0(\omega) > 0$ which in conjunction with (4),
\[
\tilde{s} \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k^{f_k} \subset \{s^*\} \text{ a.s.}
\]
implies that $\mathbb{P}(B) = 0$.

In consequence of the Heine Borel property we have thus $r_n \to \infty$ for all $\omega \in A \setminus B$. Since $\mathbb{E}[\rho_p(\pi_f \circ X, s)^2] < \infty$ for all $s \in S_p$, there is a $C > 0$ such that
\[
\mathbb{P}\{\omega \in A : \rho_p(\pi_f \circ X, s) < C\} > 0.
\]
Hence, in consequence of Assumption 3.6 we have thus a subset $A' \subset A \setminus B$ with $\mathbb{P}(A') > 0$ such that for all $\omega \in A'$, due to the usual strong law,
\[
\ell \geq \ell_{n,f_n} = F_{n,f_n}(s_n) \geq \frac{1}{n} \sum_{i=1}^{n} 1_{\{\omega \in A : r_0(\pi_f \circ X_i(s), s) < C\}} \rho_p(n, \pi_f \circ X_i(s), s_n)^2 \to \infty \text{ a.s.}
\]
This is a contradiction to (22). This yields (3) completing the proof of Theorem 4.1.
6.3.3. Proof of Corollary 4.3

Let \( \epsilon > 0 \). Then by Assumption 3.10, Chebyshev’s inequality and (3), there is a sequence \( \delta_n \to 0 \) such that 
\[
\| \eta_n - \eta' \| < \delta_n \text{ a.s. and}
\]
\[
P \left\{ \left\| \frac{1}{n} \sum_{k=1}^{n} \text{grad}_\eta \tau^l(\eta_n, X_k) - E \left[ \text{grad}_\eta \tau^l(\eta', X) \right] \right\| \geq \epsilon \right\}
\]
\[
\leq P \left\{ \sup_{\|\eta - \eta'\| < \delta_n} \left\| \frac{1}{n} \sum_{k=1}^{n} \text{grad}_\eta \tau^l(\eta, X_k) - E \left[ \text{grad}_\eta \tau^l(\eta', X) \right] \right\| \geq \epsilon \right\}
\]
\[
\leq \frac{1}{\epsilon} E \left[ \sup_{\|\eta - \eta'\| < \delta_n} \left\| \frac{1}{n} \sum_{k=1}^{n} \text{grad}_\eta \tau^l(\eta, X_k) - E \left[ \text{grad}_\eta \tau^l(\eta', X) \right] \right\| \right] \to 0
\]
as \( n \to \infty \), yielding the first assertion. The second assertion follows similarly.

6.4. Proof of Corollary 4.4

By hypothesis, the vector
\[
a_n := \frac{1}{n} \sum_{k=1}^{n} \text{grad}_\eta \tau^j(\eta_n, X_k)
\]
is a linear combination of the vectors
\[
b^l_n := \frac{1}{n} \sum_{k=1}^{n} \text{grad}_\eta \tau^l(\eta_n, X_k), \ l = j + 1, \ldots, m
\]
conveyed by \( \lambda^j_n + 1, \ldots, \lambda^m_n \). Similarly, \( a := E[\text{grad}_\eta \tau^j(\eta', X)] \) is a linear combination of the vectors \( b^l := E[\text{grad}_\eta \tau^l(\eta', X)] \), \( l = j + 1, \ldots, m \) conveyed by \( \lambda^{j+1}, \ldots, \lambda^m \). Set \( b = (b^{j+1}, \ldots, b^m) \), \( b_n = (b^{j+1}_n, \ldots, b^m_n) \) and \( \lambda = (\lambda^{j+1}, \ldots, \lambda^m)^T \), \( \lambda_n = (\lambda^{j+1}_n, \ldots, \lambda^m_n)^T \).

By Assumption 3.10 we have \( \text{rank}(b) = m - j \). We set \( \Omega_n = \{ \text{rank}(b_n) = m - j \} \). Since the determinant is continuous, by Corollary 4.3,
\[
P(\Omega_n) \to 1 \text{ as } n \to \infty.
\]
Now consider the function \( g(b, a) = b^+ a \) where \( b^+ \) denotes the Moore-Penrose pseudoinverse of \( b \). Then \( \lambda = b^+ a \) and \( \lambda_n = b^+_n a_n \). In consequence, for arbitrary \( \epsilon > 0 \) and \( n \to \infty \) we have that
\[
P(\{ \| \lambda_n - \lambda \| > \epsilon \}) \leq P \left\{ \| g(b_n, a_n) - g(b, a) \| > \epsilon, \omega \in \Omega_n \right\} + P(\Omega \setminus \Omega_n) \to 0
\]
because, for any \( \delta > 0 \) the first term is smaller than
\[
P \left\{ \| g(b_n, a_n) - g(b, a) \| > \epsilon, \omega \in \Omega_n, \| (b_n, a_n) - (b, a) \| \leq \delta \right\} + P \left\{ \| (b_n, a_n) - (b, a) \| > \delta \right\}.
\]
Here, due to continuity established by Stewart (1969), the first term vanishes for \( \delta \) sufficiently small, and, due to Corollary 4.3, for any fixed \( \delta > 0 \), the second term tends to zero. This yields the assertion.
6.5. Proof of Theorem 4.5

By Taylor expansion we have for $G_n$ defined in (5),

$$0 = \sqrt{n} \text{grad}_n G_n(\eta_n)$$
$$= \sqrt{n} \text{grad}_n G_n(\eta') + \text{Hess}_n G_n(\eta') \sqrt{n} (\eta_n - \eta')$$
$$+ \left( \text{Hess}_n G_n(\eta_n) - \text{Hess}_n G_n(\eta') \right) \sqrt{n} (\eta_n - \eta')$$

(23)

with some random $\tilde{\eta}_n$ between $\eta_n$ and $\eta'$. In consequence of Corollary 4.4 we have with the usual CLT for the first term in (23) that

$$\sqrt{n} \text{grad}_n G_n(\eta') = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left( \text{grad}_n \tau^j(\eta', X_k) + \sum_{l=j+1}^{m} (\lambda_l + o_p(1)) \text{grad}_n \tau^l(\eta', X_k) \right) \to \mathcal{G}$$

with a zero-mean Gaussian vector $\mathcal{G}$ of covariance

$$\text{Cov} \left[ \text{grad}_n \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda_l \text{grad}_n \tau^l(\eta', X) \right].$$

Similarly, we have for the first factor in the second term in (23),

$$\text{Hess}_n G_n(\eta') = \frac{1}{n} \sum_{k=1}^{n} \left( \text{Hess}_n \tau^j(\eta', X_k) + \sum_{l=j+1}^{m} \lambda_l \text{Hess}_n \tau^l(\eta', X_k) \right)$$
$$\xrightarrow{p} \mathbb{E} \left[ \text{Hess}_n \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda_l \text{Hess}_n \tau^l(\eta', X) \right].$$

Finally, for the first factor in the last the term in (23), invoking also Corollary 4.3, we obtain that

$$\|\text{Hess}_n G_n(\tilde{\eta}_n) - \text{Hess}_n G_n(\eta')\| \xrightarrow{p} 0.$$ 

This yields Assertion (i). If $H_\psi$ is invertible, as asserted in (ii), joint normality follows at once for $\sqrt{n}(\eta_n - \eta')$.

(iii): In case of factoring charts we can rewrite

$$\rho_{ps}(\pi_{f^j} \circ X, p^{j-1})^2 = \rho_{ps}(\pi_{f^j} \circ \psi_2^{-1}(\xi)) \left( \pi_{\psi_2^{-1}(\xi)} \circ X, \psi_1^{-1}(\theta) \right)^2 =: \tau^j(\theta, \xi, X)$$

where $\psi(f^{j-1}) = \psi(j^1, p^{j-1}) = (\psi_1(f^{j-1}), \psi_2(f^j)) = \eta$ with $\psi_1(f^{j-1}) = \theta = \phi(p^{j-1})$ and $\psi_2(f^{j}) = \xi$, and 

$\pi_{P_j} : T_{m-1, m-j+1} \to P_j$ is defined by $(p^{m-1}, \ldots, p^j) \mapsto p^j$. With $(\theta_n, \xi_n) = \eta_n$ in Assertion (ii) we obtain thus

$$\sqrt{n} H_\psi \left( \left( \frac{\theta_n}{\xi_n} \right) - \eta' \right) \to \mathcal{N}(0, B_\psi).$$

Since under projection to the first coordinates $\theta = \phi(p^{j-1})$, asymptotic normality is preserved, Assertion (iii) follows at once. \qed
(a) Great circular with $(\sigma, \tau) = (0.25\pi, 0.05\pi)$

(b) Great circular with $(\sigma, \tau) = (\pi, 0.25\pi)$

(c) Small circular with $(\sigma, \tau) = (0.25\pi, 0.05\pi)$

(d) Small circular with $(\sigma, \tau) = (\pi, 0.25\pi)$

Fig 8: Simulating $n = 200$ data points in $S^2$ following a great circle (left displays) and a proper small circle (right displays) with noise $\sigma$ along the circle and noise $\tau$ orthogonal to it. We show a moderate and a large noise level.

6.6. Simulating the Asymptotic Rate of Bootstrapped Descriptor Variance

In order to study the asymptotic rate of variances of BNFDs, which, under the assumptions of Theorem 4.5 scale with inverse sample size, we simulate more data in $S^2$. It is concentrated on a great circle and on a small circle, respectively, where the distribution along the circle is Gaussian with variance $\sigma^2$ and, orthogonal to it, an independent folded Gaussian with variance $\tau^2$. We study the asymptotic rate of the variance of the bootstrapped nested mean as well as that of the joint distribution of the circular descriptor and the nested mean over varying sample sizes $n$ and varying data variances $\sigma^2, \tau^2$. Figure 8 illustrates the data of four of such simulations and in Figure 9 we depict the dependence of the bootstrapped variances of $\|\phi(p_{X_n}^* - \phi(p_{X_n})\|$ over varying sample size $n$ and noise levels $\sigma, \tau$. Indeed, for small and moderate noise levels, bootstrapped variances decay asymptotically with $1/n$. For large data spread, however, there seems no asymptotic decay for some descriptors. This hints to a context, where no longer all assumptions for the Consistency Theorem 4.1 and the Central Limit Theorem 4.5 are met.

6.7. For Comparison: An Asymptotic Multinomial Two-Sample Test

As literature on this topic seems very sparse, we follow standard folklore. Consider two groups of cells $X$ and $Y$ which give rise to accumulated sample pixel counts $m_{1X}, m_{2X}, m_{3X}$ and $m_{1Y}, m_{2Y}, m_{3Y}$ in the corresponding orientation fields, where, as introduced in Section 5, $m_{1}$ stands for the dominating field, $m_{2}$ for the sum over all other fields and $m_{3}$ for all other filaments. Assuming that for every cell in each group the pixel counts $m_{j}$ in each orientation field follow true population pixel ratios $p_{j} (j = 1, 2, 3)$ with $p_{1} + p_{2} + p_{3} = 1$ and that $m_{X}$ and $m_{Y}$, respectively, are the accumulated numbers of pixels in groups $X$ and $Y$, respectively, we have for the classical Pearson’s $\chi^2$-statistic that

$$\sum_{j=1}^{3} \left( \frac{\left(m_{jX} - p_{X} m_{X}\right)^2}{p_{jX} m_{X}} + \frac{\left(m_{jY} - p_{Y} m_{Y}\right)^2}{p_{jY} m_{Y}} \right) \to \chi^2_4 \text{ for } m_{X}, m_{Y} \to \infty, \text{ with } \lim \frac{m_{X}}{m_{Y}} \in (0, \infty).$$

Under equality of filament populations, i.e. $H_0 : (p_{1X}, p_{2X}, p_{3X}) = (p_{1Y}, p_{2Y}, p_{3Y}) = (p_1, p_2, p_3)$ we estimate the unknown $p_j$ via the averaged MLE

$$\hat{p}_j = \frac{m_{jX}^2}{m_{X}^2 + m_{Y}^2}$$
Fig 9: Log-log plots for bootstrapped variances of simulated data (cf. Figure 8) over sample sizes $n$ for various underlying data spread, reporting nested means (dashed) and jointly great and small circles, respectively, with nested means (solid) for PNGS (red) and PNS (blue).

$(j = 1, 2, 3)$, lose two degrees of freedom, to obtain, under $H_0$,

$$
\begin{align*}
\sum_{j=1}^{3} \left( \frac{(m_j^X - \hat{p}_j m^X)^2}{\hat{p}_j m^X} + \frac{(m_j^Y - \hat{p}_j m^Y)^2}{\hat{p}_j m^Y} \right) & \\
= \sum_{j=1}^{3} \frac{m_j^X m_j^Y}{m_j^X + m_j^Y} & \left( \frac{m_j^X}{m_j^X} - \frac{m_j^Y}{m_j^Y} \right)^2 \xrightarrow{D} \chi_2^2 \quad \text{for } m^X, m^Y \to \infty, \lim_{m^X, m^Y} \in (0, \infty). \tag{24}
\end{align*}
$$

When applying the corresponding test to the data, we always reject the null hypothesis. If we normalize total pixel counts in every cell to one, we never reject. In fact, normalization of cell pixel counts to $\beta \in \mathbb{N}$ serves as a tuning parameter and setting this unknown parameter to $\beta = 20$, we obtain in the left display of Table 4 a result similar to our result in Section 5.2, Table 2. Obviously this parameter leads to a less conservative test, as the equality of both cell samples for 24 hours, which is suggested by Table 3, is not detected in the right display of Table 4.
References


---

Table 4
Displaying \( p \)-values of the multinomial two-sample tests based on (24) modified with tuning parameter \( \beta = 20 \) for the cell filament orientation field distribution data from Section 5.2.

<table>
<thead>
<tr>
<th>Time</th>
<th>1 kPa vs. 10 kPa and 30 kPa</th>
<th>4h vs. 8h</th>
<th>8h vs. 12h</th>
<th>12h vs. 16h</th>
<th>16h vs. 20h</th>
<th>20h vs. 24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gel</td>
<td></td>
<td>&lt;10^{-3}</td>
<td>&lt;10^{-3}</td>
<td>0.064</td>
<td>0.259</td>
<td>&lt;10^{-3}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>1 kPa vs. 10.1 kPa and 30.1 kPa</th>
<th>4h vs. 8h</th>
<th>8h vs. 12h</th>
<th>12h vs. 16h</th>
<th>16h vs. 20h</th>
<th>20h vs. 24h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&lt;10^{-3}</td>
<td>&lt;10^{-3}</td>
<td>&lt;10^{-3}</td>
<td>&lt;10^{-3}</td>
<td>&lt;10^{-3}</td>
</tr>
</tbody>
</table>