BACKWARD NESTED DESCRIPTORS ASYMPTOTICS WITH INFERENCE ON STEM CELL DIFFERENTIATION[1]

BY STEPHAN F. HUCKEMANN AND BENJAMIN ELTZNER

Georg-August-Universität Göttingen

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, Dryden and Marron [*Biometrika* **99** (2012) 551–568] 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.

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, for example, by Anderson (1963), Ruymgaart and Yang (1997), Watson (1983), 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 μ, maximizing the projected variance. These subspaces have representa-

tions as $p^k = \mu + \text{span}\{\gamma_1, \ldots, \gamma_k\}$, the affine translates of spans from an eigen-vector 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 the full dimensional space and remove least descriptive dimensions (backward) to obtain the same nested sequence of subspaces

$$(1) \qquad \{\mu\} = p^0 \subset p^1 \subset \cdots \subset p^m = Q.$$

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 an *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 PCA 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, Hotz and Munk (2010b), Huckemann and Ziezold (2006), 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, Foskey and Marron (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, Dryden and Marron (2012) give a tool for descriptive shape analysis that often strikingly outperforms tangent space PCA, for example, Pizer et al. (2013). Here, the data space is a unit sphere $Q = \mathbb{S}^m$ 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 Bhattacharya and Patrangenaru (2003, 2005), Hendriks and Landsman (1996), Huckemann (2011a), Ziezold (1977), 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 nondata space valued descriptors, geodesics, say, are available [cf. Huckemann (2011b, 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, Hotz and Munk (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, for example, 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, for example, 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 nonmanifold Kendall's shape spaces are deferred to the Supplement, Appendices A and B [Huckemann and Eltzner (2018)]. 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, for example, 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 nontrivial 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
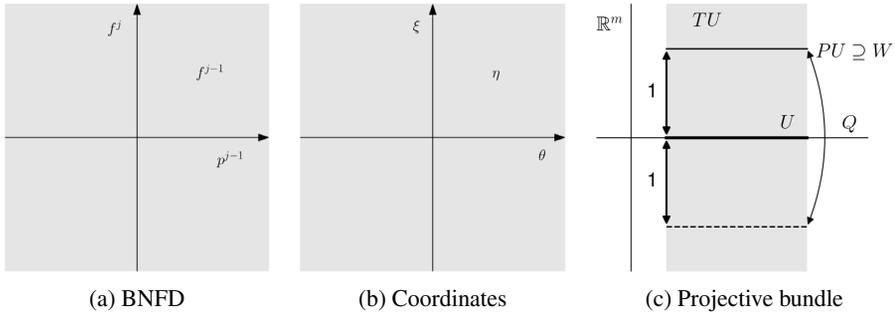
FIG. 1. *The chart ψ factors, if the projection of η to the θ coordinates of the chart [Display (b)] equals the projection of f^{j-1} to p^{j-1} in the descriptor space [Display (a)] followed by a chart for p^{j-1}. For the IMo1GPC example, where the descriptor space carries the structure of the projective bundle PQ, a suitable subset $W \subseteq PU$ over a neighborhood $U \subset Q$ factors, because the tangent bundle TU does so.*

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 σ-algebra of Q. In the following, *smooth* refers to existing continuous second-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 2.1. 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) $P_m = \{Q\}$;

(iii) every $p \in P_j$ ($j = 1, \ldots, m$) is itself a topological space and gives rise to a topological space $\varnothing \neq S_p \subset P_{j-1}$ 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

$$f = (p^{j-k}, \ldots, p^j) \qquad \text{with } p^{l-1} \in S_{p^l}, l = j - k + 1, \ldots, j$$

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 $f = (p^{m-l})_{l=0}^k$ set

$$\pi_f = \pi_{p^{m-k+1}, p^{m-k}} \circ \cdots \circ \pi_{p^m, p^{m-1}} : p^m \to p^{m-k}$$

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

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

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

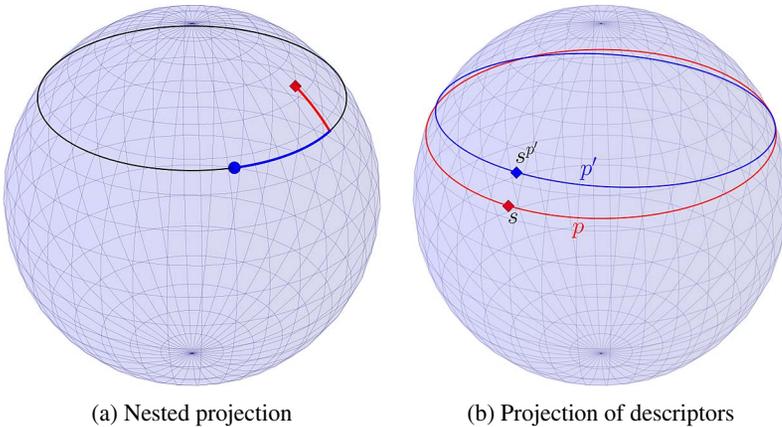(a) Nested projection   (b) Projection of descriptors

FIG. 2. *PNS illustration. Left: Projection of X (red) in $Q = \mathbb{S}^2$ onto small circle p (blue) and further onto s (green). Right: Projection $s^{p'}$ (blue) onto $S_{p'}$ (which is p' in this case) of s (red) on S_p (which is p in this case).*

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

$$\left(E^{f^j} : j = m, \ldots, 0\right), \qquad \left(E_n^{f_n^j} : j = m, \ldots, 0\right)$$

recursively defined via $E^m = \{Q\} = E_n^m$, that is, $p^m = Q = p_n^m$ and

$$E^{f^{j-1}} = \operatorname*{argmin}_{s \in S_{p^j}} \mathbb{E}\left[\rho_{p^j}(\pi_{f^j} \circ X, s)^2\right], \qquad f^j = (p^k)_{k=j}^m,$$

$$E_n^{f_n^{j-1}} = \operatorname*{argmin}_{s \in S_{p_n^j}} \sum_{i=1}^n \rho_{p_n^j}(\pi_{f_n^j} \circ X_i, s)^2, \qquad f_n^j = (p_n^k)_{k=j}^m,$$

where $p^j \in E^{f^j}$ and $p_n^j \in E^{f_n^j}$ 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 $f^j = (p^k)_{k=j}^m$ for all $j = 0, \ldots, m$.

Each of the $E^{f^{j-1}}$ and $E_n^{f_n^{j-1}}$ 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

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

we say that the *chart ψ factors* [cf. Figure 1(a) and (b)], if with the projections

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

we have

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

2.2. *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 $T_q Q$
$(q \in Q)$, induced metric $d : Q \times Q \to [0, \infty)$, *projective tangent bundle* $PQ = \{(q, \{v, -v\}) : q \in Q, v \in T_q Q, \|v\| = 1\}$ and space of classes of geodesics given
by their point sets

$$P_1 = \{[\gamma_{q,v}] : (q, \{v, -v\}) \in PQ\},$$

$$[\gamma_{q,v}] = \{\gamma_{r,w} : \gamma_{q,v}(t) = r, \dot{\gamma}_{q,v}(t) = w \text{ for some } t\},$$

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

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

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

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

Then every class of geodesics determined by

$$\underset{(q,v) \in TQ}{\operatorname{argmin}} \mathbb{E}[\rho(X, [\gamma_{q,v}])^2] \quad \text{or} \quad \underset{(q,v) \in TQ}{\operatorname{argmin}} \sum_{k=1}^n \rho(X_k, [\gamma_{q,v}])^2$$

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 $p_n = [\gamma_{q_n,v_n}]$, with the orthogonal
projection

$$\pi_{Q,p} : Q \to p = [\gamma_{q,v}], \qquad q' \mapsto \underset{\gamma_{q,v}(t)}{\operatorname{argmin}} d(q', \gamma_{q,v}(t))$$

which is well defined outside a set of zero Riemannian volume [e.g., Huckemann,
Hotz and Munk (2010b), Theorem 2.6], we have the *intrinsic population means on*
p and *intrinsic sample means on* p_n determined by

$$\underset{[\gamma_{q,v}]}{\operatorname{argmin}} \mathbb{E}[\rho_p(\pi_{Q,p} \circ X, \gamma_{q,v}(t))^2] \quad \text{or} \quad \underset{[\gamma_{q,v}]}{\operatorname{argmin}} \sum_{k=1}^n \rho_{p_n}(\pi_{Q,p_n} \circ X_k, \gamma_{q,v}(t))^2,$$

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

$$(2) \qquad\qquad T_{1,1} \to PQ, \qquad (s, [\gamma_{q,v}]) \mapsto (s, \{w, -w\}),$$

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

Recall that the tangent bundle $TQ = \{(q, v) : q \in Q, v \in T_q Q\}$ admits *local trivializations*, that is, 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 $v' \in T_{q'} Q$, $q' \in U$ and the second coordinate τ_2 is an isometric vector space isomorphism. In consequence, for a given $(q, \{v, -v\}) \in PQ$, $\|v\| = 1$, 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

$$W = \{(q', \{v', -v'\}) : (q', v') \in TU, \|v'\| = 1, \{\tau_2(q', v'), -\tau_2(q', v')\} \in H\}$$
$$\subset PU,$$

the mapping

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

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 ϕ_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 IMo1GPC 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 $\mathbb{E}[\rho_{p^j}(\pi_f \circ X, s)^2] < \infty$ for all BNFDs f ending at p^j, $s \in S_{p^j}$.

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

$$S_{p'}^s \text{ has exactly one element}$$

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

For $s \in S_p$ and $p, p' \in P_j$ sufficiently close, let $s^{p'} \in S_{p'}^s$ 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 $\varepsilon > 0$, there is $\delta > 0$ such that

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

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 $\varepsilon > 0$, there are $\delta_1, \delta_2 > 0$ such that

$$\left| \rho_p(\pi_f(q), s) - \rho_{p'}(\pi_{f'}(q), s') \right| < \varepsilon \qquad \forall q \in Q$$

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 Supplementary Material [Huckemann and Eltzner (2018)].

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 $Q = \Sigma_m^k$, $0 < m < k$, 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, Hotz and Munk (2010b), Theorem A.5, as there, in Claim II, a neighborhood of a geodesic p is constructed, restricted to which the orthogonal projection π_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

$$\psi : W \to \mathbb{R}^{\dim(U)}, \qquad f^{j-1} = (p^{j-1}, \ldots, p^{m-1}) \mapsto \eta,$$

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

$$\eta \mapsto f^{l-1} \mapsto \rho_{p^l}(\pi_{f^l} \circ X, p^{l-1})^2 := \tau^l(\eta, X),$$

here the first map $\eta \mapsto f^{l-1}$ is ψ^{-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 η', that is, for $\delta \to 0$ we have

$$\mathbb{E}\Big[\sup_{\|\eta - \eta'\| < \delta} \|\text{grad}_\eta \tau^l(\eta, X) - \text{grad}_\eta \tau^l(\eta', X)\|\Big] \to 0,$$

$$\mathbb{E}\Big[\sup_{\|\eta - \eta'\| < \delta} \|\text{Hess}_\eta \tau^l(\eta, X) - \text{Hess}_\eta \tau^l(\eta', X)\|\Big] \to 0.$$

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 Appendix A with projections (8) from the supplement (see also Proposition A.2) and distances (12) from the supplement smooth away from singularity sets. For IMo1GPCs on Kendall's shape spaces, this has been provided in Appendix B; cf. also Huckemann (2011b).

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 (2014). 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* $X_1, \ldots,$ $X_n \overset{\text{i.i.d.}}{\sim} X$ *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* $(E_n^{f_n^k}, \ldots, E_n^{f_n^m})$ *due to a measurable selection* $p_n^j \in E_n^{f_n^j}$ *giving rise to BNFDs* $f_n^j = (p_n^l)_{l=j}^m$, $j = k, \ldots, m$. *If Assumptions* 3.1–3.6 *are valid for all* $j = k, \ldots, m - 1$, *and every* $\bigcup_{n=1}^\infty E_n^{f_n^j}$ *is a.s.* d_j-*Heine Borel* $(j = k, \ldots, m)$, *then* $(E_n^{f_n^k}, \ldots, E_n^{f_n^m})$ *converges a.s. to* $(p^k, \ldots, p^m)$ *in the sense that* $\exists \Omega' \subset \Omega$ *measurable with* $\mathbb{P}(\Omega') = 1$ *such that for all* $j = k, \ldots, m$, $\varepsilon > 0$ *and* $\omega \in \Omega'$, $\exists N = N(\varepsilon, \omega)$ *with*

$$(3) \qquad \bigcup_{r=n}^\infty E_r^{f_r^j} \subset \{p \in P_j : d_j(p^j, p) \leq \varepsilon\} \qquad \forall n \geq N, \omega \in \Omega'.$$

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 (2011b).

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

$$F_f(s) = \mathbb{E}[\rho_p(\pi_f \circ X, s)^2], \qquad s \in S_p,$$

$$F_{n, f'}(s) = \frac{1}{n} \sum_{i=1}^{n} \rho_{p'}(\pi_{f'} \circ X_i, s)^2, \qquad s \in S_{p'},$$

$$\ell_f = \inf_{s \in S_p} F_f(s), \qquad \ell_{n, f'} = \inf_{s \in S_{p'}} F_{n, f'}(s).$$

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

$$\{s^*\} = \operatorname*{argmin}_{s \in S_p} F_f(s), \qquad E_n^{f_n} = \operatorname*{argmin}_{s \in S_{p_n}} F_{n, f_n}(s).$$

To complete the proof, we first show in the supplement

$$(4) \qquad \bigcap_{n=1}^{\infty} \overline{\bigcup_{r=n}^{\infty} E_r^{f_r}} \subset \{s^*\} \qquad \text{a.s.}$$

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). $\qquad \square$

REMARK 4.2. Careful inspection of the proof yields that we have only used that the "distances" d_j vanish on the diagonal $d_j(p, p) = 0$ for all $p \in P_j$; they need not be definite, that is, 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 Supplementary Material [Huckemann and Eltzner (2018)].

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*:

$$\frac{1}{n} \sum_{k=1}^{n} \operatorname{grad}_{\eta} \tau^l(\eta_n, X_k) \xrightarrow{\mathbb{P}} \mathbb{E}[\operatorname{grad}_{\eta} \tau^l(\eta', X)],$$

$$\frac{1}{n} \sum_{k=1}^{n} \operatorname{Hess}_{\eta} \tau^l(\eta_n, X_k) \xrightarrow{\mathbb{P}} \mathbb{E}[\operatorname{Hess}_{\eta} \tau^l(\eta', X)].$$

Here, η_n denotes the image of f_n^{j-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^{j-1} of BN sample means minimizes

$$\frac{1}{n} \sum_{k=1}^{n} \rho_{p_n^j}\big(\pi_{f_n^j} \circ X_k, p_n^{j-1}\big)^2$$

under the constraints that it minimizes each of

$$\frac{1}{n} \sum_{k=1}^{n} \rho_Q(X_k, p_n^{m-1})^2, \qquad \ldots, \qquad \frac{1}{n} \sum_{k=1}^{n} \rho_{p_n^{j+1}}\big(\pi_{f_n^{j+1}} \circ X_k, p_n^{j}\big)^2.$$

Similarly, the BN population mean f'^{j-1} minimizes

$$\mathbb{E}\big[\rho_{p^j}\big(\pi_{f^j} \circ X, p^{j-1}\big)^2\big]$$

under the constraints that it minimizes each of

$$\mathbb{E}\big[\rho_Q(X, p^{m-1})^2\big], \qquad \ldots, \qquad \mathbb{E}\big[\rho_{p^{j+1}}\big(\pi_{f^{j+1}} \circ X, p^{j}\big)^2\big].$$

In consequence, due to differentiability guaranteed by Assumption 3.10, with the notation of τ^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:

(5)
$$\operatorname{grad}_\eta G_n(\eta_n) = 0$$

$$\text{with } G_n(\eta) := \frac{1}{n} \sum_{k=1}^{n} \tau^j(\eta, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \frac{1}{n} \sum_{k=1}^{n} \tau^l(\eta, X_k),$$

(6)
$$\operatorname{grad}_\eta G(\eta') = 0$$

$$\text{with } G(\eta) := \mathbb{E}\big[\tau^j(\eta, X)\big] + \sum_{l=j+1}^{m} \lambda^l \mathbb{E}\big[\tau^l(\eta, X)\big].$$

The following corollary rests on, among others, 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*

$$\lambda_n^l \xrightarrow{\mathbb{P}} \lambda^l \qquad \text{for } l = j+1, \ldots, m.$$

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 Supplementary Material [Huckemann and Eltzner (2018)].

THEOREM 4.5. *Let $j \in \{1, \ldots, m-1\}$ and consider random data $X_1, \ldots,$ $X_n \overset{\text{i.i.d.}}{\sim} X$ on a data space Q admitting BNFDs 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 $(E_n^{f_n^{j-1}}, \ldots,$ $E_n^{f_n^{m-1}})$ due to a measurable selection $p_n^l \in E_n^{f_n^l}$, $f_n^{j-1} = (p_n^{j-1}, \ldots, p_n^{m-1})$, $l = j-1, \ldots, m-1$.*

(i) *Assuming that Assumption 3.10 holds as well as (3) for all $j \in \{j - 1, \ldots, m-1\}$, we have that*

$$\sqrt{n} H_\psi \big(\psi \big(f_n^{j-1} \big) - \psi \big(f'^{j-1} \big) \big) \to \mathcal{N}(0, B_\psi)$$

with a chart ψ as specified in Assumption 3.10 as well as

$$H_\psi = \mathbb{E}\left[\text{Hess}_\eta \, \tau^j(\eta', X) + \sum_{l=j+1}^m \lambda^l \, \text{Hess}_\eta \, \tau^l(\eta', X) \right] \quad \text{and}$$

$$B_\psi = \text{Cov}\left[\text{grad}_\eta \, \tau^j(\eta', X) + \sum_{l=j+1}^m \lambda^l \, \text{grad}_\eta \, \tau^l(\eta', X) \right],$$

with the notation from Assumption 3.10 where $\lambda_{j+1}, \ldots, \lambda_m \in \mathbb{R}$ are suitable such that

$$\text{grad}_\eta \, \mathbb{E}[\tau^j(\eta, X)] + \sum_{l=j+1}^m \lambda^l \, \text{grad}_\eta \, \mathbb{E}[\tau^l(\eta, X)]$$

vanishes at $\eta = \eta'$.

(ii) *If, additionally, H_ψ is positive definite, then f_n^{j-1} satisfies a Gaussian $\sqrt{n}$-CLT*

$$\sqrt{n}\big(\psi\big(f_n^{j-1}\big) - \psi\big(f'^{j-1}\big) \big) \to \mathcal{N}(0, \Sigma_\psi), \qquad \Sigma_\psi = H_\psi^{-1} B_\psi H_\psi^{-1}.$$

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

$$\sqrt{n}\big(\phi\big(p_n^{j-1}\big) - \phi\big(p'^{j-1}\big) \big) \to \mathcal{N}(0, \Sigma_\phi), \qquad \Sigma_\phi = (\Sigma_{\psi_{ik}})_{i,k=1}^{\dim(P_{j-1})}$$

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 $b' = 1, \ldots, B'$),

$$T^{*2}(A) = \left(\left(\phi(\hat{p}_{nb}^{X*}) - \phi(\hat{p}_n^X)\right) - \left(\phi(\hat{p}_{mb'}^{Y*}) - \phi(\hat{p}_m^Y)\right)\right)^T A$$
$$\times \left(\left(\phi(\hat{p}_{nb}^{X*}) - \phi(\hat{p}_n^X)\right) - \left(\phi(\hat{p}_{mb'}^{Y*}) - \phi(\hat{p}_m^Y)\right)\right),$$

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 $\sqrt{n}((\phi(\hat{p}_n^X) - \phi(\hat{p}^X)))$, and $\sqrt{m}((\phi(\hat{p}_m^Y) - \phi(\hat{p}^Y)))$, guaranteed by Theorem 4.5, extend to the same asymptotic normality for $\sqrt{n}((\phi(\hat{p}_{nb}^{X*}) - \phi(\hat{p}_n^X)))$, and $\sqrt{m}((\phi(\hat{p}_{mb'}^{Y*}) - \phi(\hat{p}_m^Y)))$, respectively. We have then under H_0 that $c_{1-\alpha}^*$ gives an asymptotic coverage of $1 - \alpha$ for $T^2(A)$, that is, $\mathbb{P}\{T^{*2}(A) \leq c_{1-\alpha}^*\} \to 1 - \alpha$ 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 Σ_ψ or Σ_ϕ, respectively, from Theorem 4.5 is invertible.

(a) Data set I   (b) Data set II   (c) Data set III
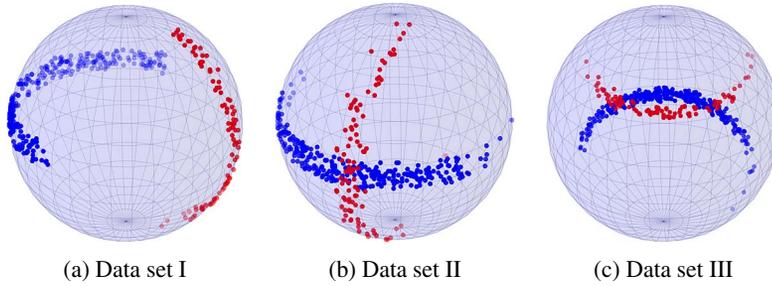
FIG. 3. *Simulated datasets* I (*left*) *and* II (*middle*) *on* S^3 *concentrate on a common proper small* S^2, *their projections to estimated small two-spheres is depicted. The simulated dataset* III (*right*) *is on* S^2.

5. Applications.

5.1. *Simulations.* To illustrate our CLT for principle nested spheres (PNS) and principle 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.

We apply PNS and PNGS to the simulated data and perform the two-sample test for identical respective nested submanifolds (means, small and great circles) and for identical small and great two-spheres. 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 Appendix E of the Supplementary Material [Huckemann and Eltzner (2018)], we report further simulations, in particular that for small and moderate spread bootstrap sample variances following 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). On Kendall's planar shape spaces, smoothness of the distance ρ between a shape and a geodesic has been established if this distance is less than $\pi/4$; cf. Huckemann (2011b), Theorem 2.2. The greater picture, verifying these assumptions more generally, in particular for higher order descriptors remains challenging research.
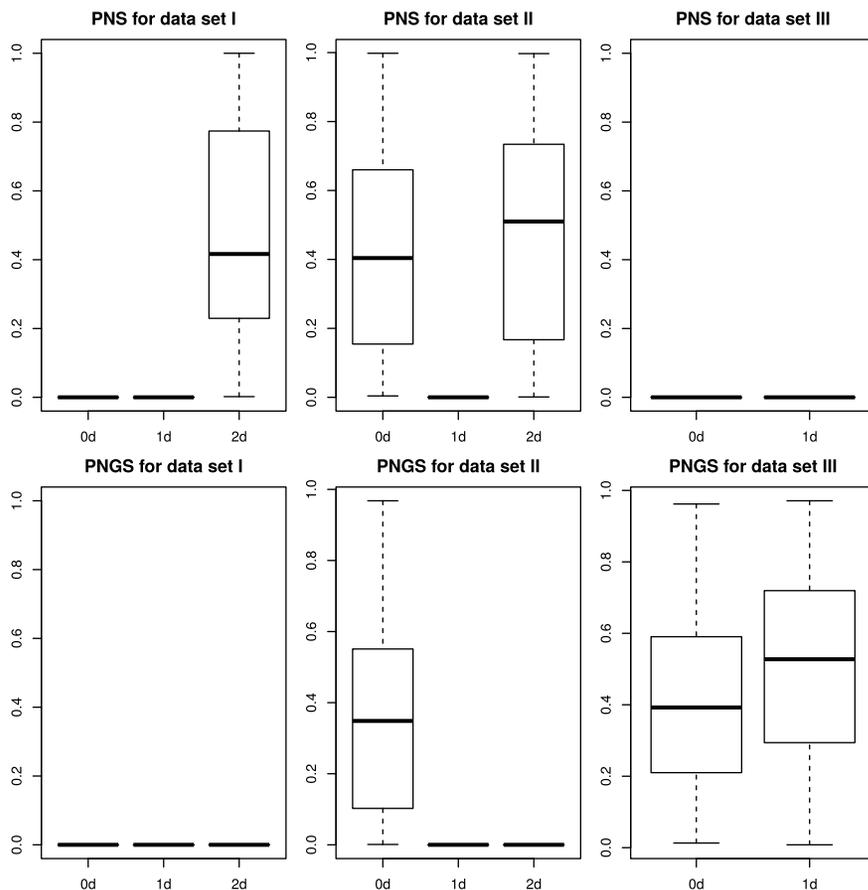
FIG. 4. *Boxplots displaying the distribution of* 100 *p-values for PNS and PNGS from the two-sample test on identical nested mean* (0d), *on identical nested small and great circle, respectively,* (1d) *and on identical small and great two-sphere* (2d). *Using* $B = 1000$ *bootstrap samples each.*

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, for example, Discher, Janmey and Wang (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 fila-*

TABLE 1
Sample sizes of hMSC skeleton images over varying Young's moduli and cultivation time

Time	1 kPa	10 kPa and 30 kPa
4 h	159	321
8 h	163	317
12 h	176	344
16 h	135	274
20 h	138	253
24 h	166	304

ment 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, Hummel and Zucker (1976). The source code of our implementation is available as Supplementary Material [Huckemann and Eltzner (2018)]. A detailed description is deferred to a future publication. The algorithm results in a set of contiguous areas with slowly varying local orientation, and corresponding to each of these areas, a set of filaments which closely follow the local orientation. Also, these data will be published and made available after completion of current research; cf. Wollnik and Rehfeldt (2016).
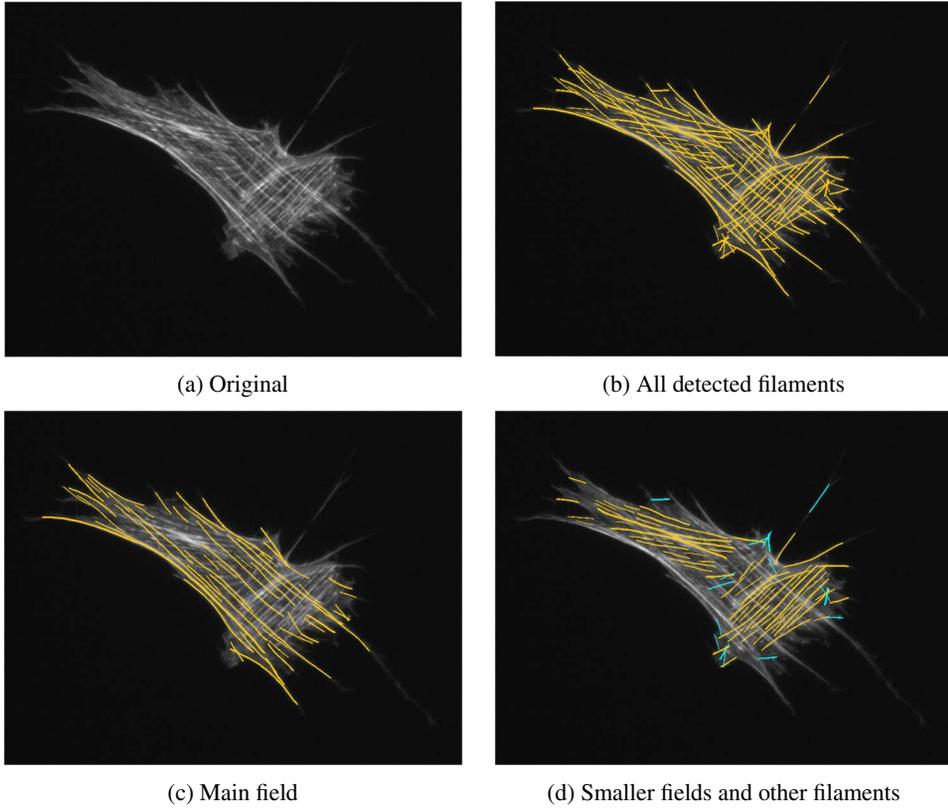
(a) Original

(b) All detected filaments

(c) Main field

(d) Smaller fields and other filaments

FIG. 5. (a): *Fluorescence image of an immuno-stained human mesenchymal stem cell after culti-vation for* 16 *hours at Young's modulus* 10 *kPa.* (b): *Automatically extracted filament structure using the Filament Sensor from* Eltzner et al. (2015). (c): *Filaments of the largest orientation field.* (d): *Filaments of smaller orientation fields* (yellow) *and filaments not belonging to any orientation field* (cyan).

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 $x = (x_1, x_2, x_3) := (\sqrt{m_1/M}, \sqrt{m_2/M}, \sqrt{1 - (m_1 + m_2)/M})^T \in Q = \mathbb{S}^2$ where the square roots ensure that x does not describe relative areas but rather relative diameters of fields. This representation is confined to the $\mathbb{S}^2$ 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
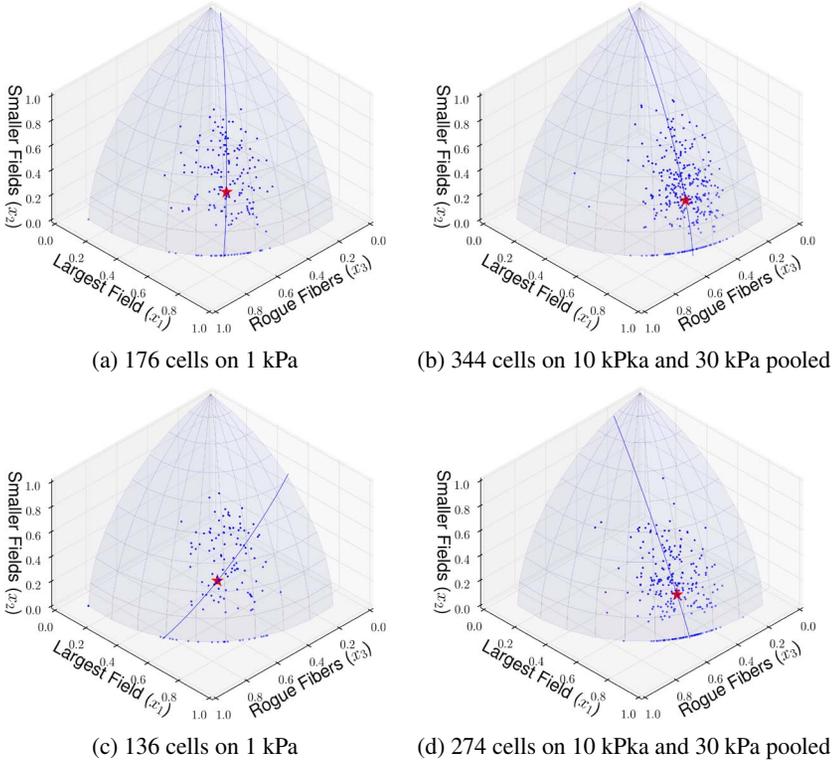
(a) 176 cells on 1 kPa

(b) 344 cells on 10 kPka and 30 kPa pooled

(c) 136 cells on 1 kPa

(d) 274 cells on 10 kPka and 30 kPa pooled

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 (read star).*

than small circle fits and we use the nested two-sample tests for PNGS with the following null hypothesis:

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

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.

TABLE 2
Displaying p-values of two-sample tests for PNGS of filament orientation field distribution data.
The test uses $B = 1000$ bootstrap samples, therefore, the penultimate p-value is 10^{-3}

| Time | Nested great circle mean | | Jointly great circle and nested mean | |
Gel	1 kPa	10 kPa and 30 kPa	1 kPa	10 kPa and 30 kPa
4 h vs. 8 h	0.120	$<10^{-3}$	0.308	$<10^{-3}$
8 h vs. 12 h	$<10^{-3}$	$<10^{-3}$	0.024	$<10^{-3}$
12 h vs. 16 h	0.126	$<10^{-3}$	0.008	$<10^{-3}$
16 h vs. 20 h	0.468	0.626	0.494	0.462
20 h vs. 24 h	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.014

Visualization in Figure 7 reveals further details. As seen from the loci of the
nested means, hMSCs on the soft gel [Figure 7(a)] 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 be-
ginning 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 hM-
SCs on harder gels [cf. Figure 7(b)] 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 criti-
cal 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'
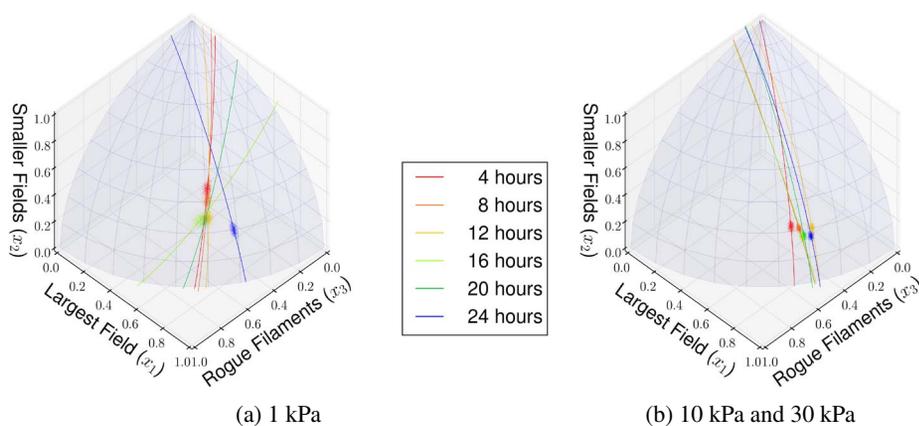level, though. In contrast, for the hMSCs on the soft gel, the critical slot seems to

(a) 1 kPa (b) 10 kPa and 30 kPa

FIG. 7. *Spherical representations of bootstrapped BNFDs (nested means on mean great circles)*
for the four data set at 6 *time points.*

TABLE 3
Displaying p-values of two-sample 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}

Gels	1 kPa vs. 10 kPa and 30 kPa	
Time	Nested great circle mean	Jointly great circle and nested mean
4 h	$<10^{-3}$	$<10^{-3}$
8 h	$<10^{-3}$	$<10^{-3}$
12 h	$<10^{-3}$	$<10^{-3}$
16 h	$<10^{-3}$	$<10^{-3}$
20 h	$<10^{-3}$	$<10^{-3}$
24 h	0.010	0.061

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.

On suggestion by the associate editor, for comparison, we include an application of a multinomial test to the same data, in the supplement. As literature on that topic seems very sparse, we follow standard folklore.

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.

SUPPLEMENTARY MATERIAL

Supplement to "Backward nested descriptors asymptotics with inference on stem cell differentiation" (DOI: 10.1214/17-AOS1609SUPP; .pdf). The supplement contains five appendices. Enumeration of equations, tables and images continues consecutively into the Appendices. Appendix A shows that the relevant assumptions hold true for principal nested spheres (PNS) analysis. Appendix B shows the same for principal components for Kendall's shape spaces. Appendix C contains full proofs of the theorems. Appendix D consists of a numerical study of the convergence rate of PNS estimators and finally, Appendix E discusses a possible alternative treatment of the data application and its shortcomings.

REFERENCES

AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620

ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. LePage and L. Billard, eds.). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 13–47. Wiley, New York. MR1197777

BHATTACHARYA, R. and LIN, L. (2016). Omnibus CLT for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* To appear.

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634

BIANCO, P., CAO, X., FRENETTE, P. S., MAO, J. J., ROBEY, P. G., SIMMONS, P. J. and WANG, C.-Y. (2013). The meaning, the sense and the significance: Translating the science of mesenchymal stem cells into medicine. *Nat. Med.* **19** 35–42.

CHENG, G. (2015). Moment consistency of the exchangeably weighted bootstrap for semiparametric M-estimation. *Scand. J. Stat.* **42** 665–684. MR3391685

DISCHER, D. E., JANMEY, P. and WANG, Y. (2005). Tissue cells feel and respond to the stiffness of their substrate. *Science* **310** 1139–1143.

ELTZNER, B., WOLLNIK, C., GOTTSCHLICH, C., HUCKEMANN, S. and REHFELDT, F. (2015). The filament sensor for near real-time detection of cytoskeletal fiber structures. *PLoS ONE* **10** e0126346.

ENGLER, A. J., SEN, S., SWEENEY, H. L. and DISCHER, D. E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell* **126** 677–689.

FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. C. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.

GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179

GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725

HENDRIKS, H. and LANDSMAN, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statist. Probab. Lett.* **26** 169–178. MR1381468

HOTZ, T. and HUCKEMANN, S. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.* **67** 177–193. MR3297863

HUCKEMANN, S. (2011a). Inference on 3D Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Stat.* **38** 424–446. MR2833839

HUCKEMANN, S. F. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349

HUCKEMANN, S. F. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259. MR2981621

HUCKEMANN, S. (2014). (Semi-)intrinsic statistical analysis on non-Euclidean spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics* Springer, Berlin.

HUCKEMANN, S. F. and ELTZNER, B. (2018). Supplement to "Backward nested descriptors asymptotics with inference on stem cell differentiation." DOI:10.1214/17-AOS1609SUPP.

HUCKEMANN, S. and HOTZ, T. (2014). On means and their asymptotics: Circles and shape spaces. *J. Math. Imaging Vision* **50** 98–106. MR3233136

HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010a). Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 593–603.

HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010b). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statist. Sinica* **20** 1–100.

HUCKEMANN, S. and ZIEZOLD, H. (2006). Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Adv. in Appl. Probab.* **38** 299–319. MR2264946

HUCKEMANN, S., KIM, K.-R., MUNK, A., REHFELDT, F., SOMMERFELD, M., WEICKERT, J. and WOLLNIK, C. (2016). The circular SiZer, inferred persistence of shape parameters and application to early stem cell differentiation. *Bernoulli* **22** 2113–2142. MR3498025

JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568. MR2966769

JUNG, S., FOSKEY, M. and MARRON, J. S. (2011). Principal arc analysis on direct product manifolds. *Ann. Appl. Stat.* **5** 578–603. MR2810410

KENT, J. T. and TYLER, D. E. (1996). Constrained *M*-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370. MR1401854

LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. MR2954043

PENNEC, X. (2015). Barycentric subspaces and affine spans in manifolds. In *Geometric Science of Information*. *Lecture Notes in Computer Science* **9389** 12–21. Springer, Cham. MR3442180

PENNEC, X. (2016). Barycentric subspace analysis on manifolds. *Ann. Statist.* To appear. Available at arXiv:1607.02833.

PITTENGER, M. F., MACKAY, A. M., BECK, S. C., JAISWAL, R. K., DOUGLAS, R., MOSCA, J. D., MOORMAN, M. A., SIMONETTI, D. W., CRAIG, S. and MARSHAK, D. R. (1999). Multilineage potential of adult human mesenchymal stem cells. *Science* **284** 143–147.

PIZER, S. M., JUNG, S., GOSWAMI, D., VICORY, J., ZHAO, X., CHAUDHURI, R., DAMON, J. N., HUCKEMANN, S. and MARRON, J. S. (2013). Nested sphere statistics of skeletal models. In *Innovations for Shape Analysis*. *Math. Vis.* 93–115. Springer, Heidelberg. MR3075829

ROSENFELD, A., HUMMEL, R. A. and ZUCKER, S. W. (1976). Scene labeling by relaxation operations. *IEEE Trans. Syst. Man Cybern. Syst.* **SMC-6** 420–433. MR0418553

RUYMGAART, F. H. and YANG, S. (1997). Some applications of Watson's perturbation approach to random matrices. *J. Multivariate Anal.* **60** 48–60. MR1441458

SHAPIRO, A. (2000). On the asymptotics of constrained local *M*-estimators. *Ann. Statist.* **28** 948–960. MR1792795

SOMMER, S. (2013). Horizontal dimensionality reduction and iterated frame bundle development. In *Geometric Science of Information* 76–83. Springer, Berlin.

WATSON, G. S. (1983). *Statistics on Spheres*. *University of Arkansas Lecture Notes in the Mathematical Sciences* **6**. Wiley, New York. MR0709262

WOLLNIK, C. and REHFELDT, F. (2016). Quantitative live-cell analysis of BM-hMSCs on elastic substrates during early differentiation. Manuscript.

ZEMEL, A., REHFELDT, F., BROWN, A. E. X., DISCHER, D. E. and SAFRAN, S. A. (2010). Optimal matrix rigidity for stress-fibre polarization in stem cells. *Nature Physics* **6** 468–473.

ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory*, *Statistical Decision Functions*, *Random Processes and of the Eighth European Meeting of Statisticians* (*Tech. Univ. Prague*, *Prague*, 1974), *Vol. A* 591–602. Reidel, Dordrecht. MR0501230

FELIX-BERNSTEIN-INSTITUT FÜR MATHEMATISCHE STATISTIK
IN DEN BIOWISSENSCHAFTEN
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
GOLDSCHMIDTSTRASSE 7
D-37077 GÖTTINGEN
GERMANY
E-MAIL: stephan.huckemann@mathematik.uni-goettingen.de
benjamin.eltzner@mathematik.uni-goettingen.de