Chapter 18
Essentials of backward nested descriptors inference

Stephan F. Huckemann and Benjamin Eltzner

Abstract Principal component analysis (PCA) is a popular device for dimension reduction and their asymptotics are well known. In particular, principal components through the mean span the data with decreasing residual variance, as the dimension increases, or, equivalently maximize projected variance, as the dimensions decrease, and these spans are nested in a backward and forward fashion – all due to Pythagoras Theorem. For non-Euclidean data with no Pythagorian variance decomposition available, it is not obvious what should take the place of PCA and how asymptotic results generalize. For spaces with high symmetry, for instance for spheres, backward nested sphere analysis has been successfully introduced. For spaces with less symmetry, recently, nested barycentric subspaces have been proposed. In this short contribution we sketch how to arrive at asymptotic results for sequences of random nested subspaces.

18.1 Introduction

From the early days of statistics of non-Euclidean data, *Procrustes analysis* proposed by [3] for shape data as a generalization of PCA, has been a successful device of choice. In essence, data are mapped to a tangent space of a Fréchet mean and PCA is performed in that tangent space. Notably, in that setting, not only the PCs but also the base point of the tangent space is random. Because all tangent spaces are the same in a Euclidean space, this complication is non-existent for asymptotics of classical PCA, derived by [1, 10, 9] and others.

Beyond lacking a general asymptotic theory, one may view this and similar methods (e.g. [2]) also as non-satisfactory, because these tangent space PCs neither

Stephan F. Huckemann (✉) and Benjamin Eltzner
Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidstr. 7, 37077 Göttingen, Germany, e-mail: huckeman@math.uni-goettingen.de and e-mail: beltzne@uni-goettingen.de

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_18

minimize residual variance nor maximize projected variance with respect to an invariant distance. To this end *geodesic PCs* that are geodesics minimizing intrinsic residual variance have been considered by [6, 5]. These are non-nested in the sense that the *intrinsic mean* μ is in general not located on the first PC, cf. [6]. From a dimension reduction viewpoint, however, nestedness appears as a desirable feature, where one seeks a sequence of subspaces $\{p_j\}_{j=0}^m$ of the data space Q, where each subspace approximates the data best, in a certain sense, over a family of admissible subspaces, that is nested.

$$\{\mu\} = p^0 \subset p^1 \subset \ldots \subset p^m = Q\,. \tag{18.1}$$

On a sphere, if all the subspaces are small subspheres, this is realized by *principal nested sphere* (PNS) analysis by [7]. For general spaces, if each subspace is a barycentric center of an nested sequence of points, this is realized by *barycentric subspaces* (BS) by [8].

In the following we formulate a general setup for (18.1) and state asymptotic results from which inferential bootstrap procedures can be derived, as detailed in [4]. In particular we have shown that the geometric assumptions below are satisfied for PNS and for the *intrinsic mean on a first principal component geodesic*, cf. [4, 6, 5]. It is still an open problem, to explore under which conditions these assumptions hold also for barycentric subspaces.

18.2 Setup

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

1. there is a collection P_j ($j = 0,\ldots,m$) of topological separable spaces with loss functions $d_j : P_j \times P_j \to [0,\infty)$;
2. $P_m = \{Q\}$;
3. 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
$$\rho_p : p \times S_p \to [0,\infty)\,;$$
4. 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 $j \in \{1,\dots,m\}$ and $k \in \{1,\dots,j\}$ call a family

$$f = \{p^j,\dots,p^{j-k}\}, \text{ with } p^{l-1} \in S_{p^l}, l = j-k+1,\dots,j$$

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

$$T_{j,k} = \left\{ f = \{p^{j-l}\}_{l=0}^{k} : p^{l-1} \in S_{p^l}, l = j-k+1,\dots,j \right\} \subseteq \prod_{l=0}^{k} P_{j-l}\,.$$

For $k \in \{1,\dots,m\}$, given a BNFD $f = \{p^{m-l}\}_{l=0}^{k}$ set

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

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 18.1 (a).

Definition 18.2. Random elements $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} X$ on a data space Q admitting BNFDs give rise to *backward nested population* and *sample means* (abbreviated as BN means)

$$\{E^{f^j} : j = m,\dots,0\}, \quad \{E_n^{f_n^j} : j = m,\dots,0\}$$

recursively defined via $E^m = \{Q\} = E_n^m$, i.e. $p^m = Q = p_n^m$ and

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

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

where $p^j \in E^{f^j}$ and $p_n^j \in E^{f_n^j}$ is a measurable choice for $j = 1,\dots,m$.

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 $j = 0,\dots,m$.

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 18.3 (Factoring Charts). Let $j \in \{0,\dots,m-1\}, k \in \{1,\dots,j\}$. If $T_{j,k}$ and P^{j-k} carry smooth manifold structures near $f' = (p'^j,\dots,p'^{j-k}) \in T_{j,k}$ 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

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

we say that the *chart ψ factors*, if with the projections

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

we have

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

18.3 Assumptions

For the following assumptions suppose that $j \in \{1, \dots, m-1\}$.

Assumption 18.1 *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

$$S^s_{p'} = \operatorname*{argmin}_{s' \in S_{p'}} d_{j-1}(s, s') .$$

In case of PNS this is illustrated in Figure 18.1 (a).

Assumption 18.2 *For every $s \in S_p$ there is $\delta > 0$ such that*

$$|S^s_{p'}| = 1$$

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 $s^{p'} \in S^s_{p'}$ be the unique element. Note that in general

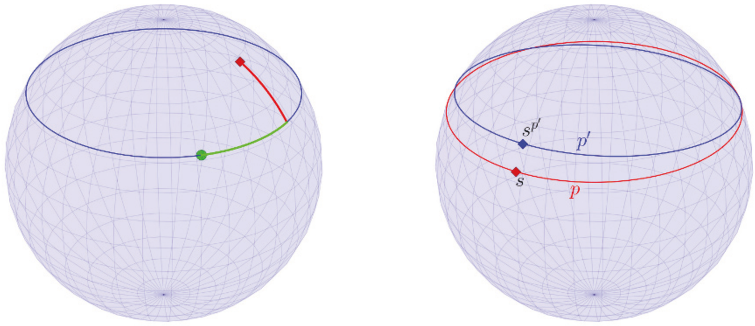
$$(s^{p'})^p \neq s .$$

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 18.3 *For $\varepsilon > 0$ there is $\delta > 0$ such that*

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

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



(a) Nested projection (b) Projection of descriptors

Fig. 18.1: *PNS illustration. Left: Projection of X (filled diamond) in $Q = \mathbb{S}^2$ onto small circle p and further onto s (filled circle). Right: Projection $s^{p'}$ (on the top circle) onto $S_{p'}$ (which is p' in this case) of s (on the lower circle) on S_p (which is p in this case).*

We will also require the following assumption, which, in conjunction with Assumption 18.3, is a consequence of the triangle inequality, if d_{j-1} is a metric.

Assumption 18.4 *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 and coercivity in the following senses.

Assumption 18.5 *For all $\varepsilon > 0$ there are $\delta_1, \delta_2 > 0$ such that*

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

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 18.6 *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 18.4 *Due to continuity, Assumptions 18.1 and 18.5 hold if Q is compact and Assumption 18.6 if each P_j is compact.*

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

Assumption 18.7 *Assume that $T_{m-1,m-j}$ carries a smooth manifold structure near the unique BN population mean ${f'}^{j-1} = ({p'}^{m-1},\ldots,{p'}^{j-1})$ such that there is an open set $W \subset T_{m-1,m-j}$, ${f'}^{j-1} \in W$ and a local chart*

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

Further, assume that for every $l = j,\ldots,m$ the mapping

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

has first and second derivatives, such that for all $l = j,\ldots,m$,

$$\operatorname{Cov}\big[\operatorname{grad}_\eta \tau^l(\eta',X)\big],\ \textit{and}\ \mathbb{E}\big[\operatorname{Hess}_\eta \tau^l(\eta',X)\big]$$

exist and are in expectation continuous near η', i.e. for $\delta \to 0$ we have

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

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

18.4 Asymptotic Theorems

The proofs of the two asymptotic theorems can be found in [4].

Theorem 18.8. *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 $\{p^m,\ldots,p^k\}$ and BN sample means $\{E_n^{f_n^m},\ldots,E_n^{f_n^k}\}$ 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 18.1 – 18.6 are valid for all $j = k,\ldots,m-1$, and every $\overline{\cup_{n=1}^\infty E_n^{f_n^j}}$ is a.s. d_j-Heine Borel ($j = k,\ldots,m$) then $\{E_n^{f_n^m},\ldots,E_n^{f_n^k}\}$ converges a.s. to $\{p^m,\ldots,p^k\}$ 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*

$$\bigcup_{r=n}^\infty E_r^{f_r^j} \subset \{p \in P_j : d_j(p^j,p) \le \varepsilon\} \quad \forall n \ge N,\ \omega \in \Omega'\,. \tag{18.2}$$

Remark 18.5 *In fact, for the proof we require 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, 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.

Theorem 18.9. *Let $j \in \{1,\dots,m-1\}$ and consider random data $X_1,\dots,X_n \overset{iid}{\sim} X$ on a data space Q admitting BNFDs from P_{m-1} to P_{j-1}, a unique BN population mean $f'^{j-1} = \{p'^{m-1},\dots,p'^{j-1}\}$ and BN sample means $\{E_n^{f_n^{m-1}},\dots,E_n^{f_n^{j-1}}\}$ due to a measurable selection $p_n^l \in E_n^{f_n^l}$, $f_n^{j-1} = \{p_n^{m-1},\dots,p_n^{j-1}\}$, $l = j-1,\dots,m-1$.*

(i) Assuming that Assumption 18.7 hold as well as (18.2) for all $j \in \{j-1,\dots,m-1\}$, we have that

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

with a chart ψ as specified in Assumption 18.7 as well as

$$\begin{aligned} H_\psi &= \mathbb{E}\left[\mathrm{Hess}_\eta \tau^j(\eta',X) + \sum_{l=j+1}^m \lambda^l \,\mathrm{Hess}_\eta \tau^l(\eta',X)\right] \text{ and} \\ B_\psi &= \mathrm{Cov}\left[\mathrm{grad}_\eta \tau^j(\eta',X) + \sum_{l=j+1}^m \lambda^l \,\mathrm{grad}_\eta \tau^l(\eta',X)\right], \end{aligned}$$

with the notation from Assumption 18.7 where $\lambda^{j+1},\dots\lambda^m \in \mathbb{R}$ are suitable such that

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

vanishes at $\eta = \eta'$.

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

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

(iii) If additionally the chart ψ factors as in Definition 18.3, then also p_n^{j-1} satisfies a Gaussian $\sqrt{n}$-CLT

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

with the notation of Definition 18.3.

Acknowledgements The authors gratefully acknowledge DFG HU 1575/4, DFG SFB 755 and the Niedersachsen Vorab of the Volkswagen Foundation.

References

[1] Anderson, T.: Asymptotic theory for principal component analysis. Ann. Math. Statist. **34** (1), 122–148 (1963)

[2] Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S. C.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23** (8), 995–1005 (2004)

[3] Gower, J. C.: Generalized Procrustes analysis. Psychometrika **40**, 33–51 (1975)

[4] Huckemann, S.F, Eltzner, B.: Backward Nested Descriptors Asymptotics with Inference on Stem Cell Differentiation, arXiv preprint: 1609.00814 (2016)

[5] Huckemann, S.F, Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). Statist. Sinica **20** (1), 1–100 (2010)

[6] Huckemann, S.F, Ziezold, H.: Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. Advances of Applied Probability (SGSA) **38** (2), 299–319 (2006)

[7] Jung, S., Dryden, I.L., Marron,J.S.: Analysis of principal nested spheres. Biometrika **99** (3), 551–568 (2012)

[8] Pennec, X. Barycentric subspace analysis on manifolds. arXiv preprint:1607.02833 (2016)

[9] Ruymgaart, F. H., Yang, S.: Some applications of Watsons perturbation approach to random matrices. J. Multivariate Anal. **60** (1), 48–60 (1997)

[10] Watson, G: Statistics on Spheres. University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 6. New York, Wiley (1983)