ARTICLE IN PRESS

Journal of Multivariate Analysis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotic and bootstrap tests for subspace dimension

Klaus Nordhausen [a], Hannu Oja [b,*], David E. Tyler [c]

[a] Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland
[b] Department of Mathematics and Statistics, University of Turku, Turku, Finland
[c] Department of Statistics, The State University of New Jersey, Piscataway, United States of America

ARTICLE INFO

Article history:
Received 12 August 2021
Received in revised form 31 August 2021
Accepted 31 August 2021
Available online xxxx

AMS 2020 subject classifications:
primary 62H12
secondary 62F12

Keywords:
Independent component analysis
Order determination
Principal component analysis
Sliced inverse regression

ABSTRACT

Many linear dimension reduction methods proposed in the literature can be formulated using an appropriate pair of scatter matrices. The eigen-decomposition of one scatter matrix with respect to another is then often used to determine the dimension of the signal subspace and to separate signal and noise parts of the data. Three popular dimension reduction methods, namely principal component analysis (PCA), fourth order blind identification (FOBI) and sliced inverse regression (SIR) are considered in detail and the first two moments of subsets of the eigenvalues are used to test for the dimension of the signal space. The limiting null distributions of the test statistics are discussed and novel bootstrap strategies are suggested for the small sample cases. In all three cases, consistent test-based estimates of the signal subspace dimension are introduced as well. The asymptotic and bootstrap tests are illustrated in real data examples.

1. Introduction

Dimension reduction (DR) plays an increasingly important role in high dimensional data analysis. In linear dimension reduction for a random vector $\boldsymbol{x} \in \mathbb{R}^p$, the idea is to try to find a transformation matrix $\boldsymbol{W} \in \mathbb{R}^{q \times p}$, $q \ll p$, such that the interesting features of the distribution of $\boldsymbol{x}$ are captured by $\boldsymbol{W}\boldsymbol{x}$ only, that is,

(i) $\boldsymbol{x}|\boldsymbol{W}\boldsymbol{x}$ is viewed as noise (unsupervised DR), or
(ii) $y \perp\!\!\!\perp \boldsymbol{x} \mid \boldsymbol{W}\boldsymbol{x}$ for the response of interest y (supervised DR).

In this paper we consider three classical but diverse linear dimension reduction methods: principal component analysis, independent component analysis and sliced inverse regression. As an introduction to our approach, we first highlight the similarities between these three approaches and show that the different methods can be presented in a joint framework.

Write $F_{\boldsymbol{x}}$ and $\boldsymbol{S} = \boldsymbol{S}(F_{\boldsymbol{x}})$ for the cumulative distribution function and covariance matrix of $\boldsymbol{x}$. To simplify the notation, we assume in the following that $\mathbb{E}(\boldsymbol{x}) = \boldsymbol{0}$.

(i) In the principal component analysis (PCA), one finds the $p \times p$ transformation matrix $\boldsymbol{W}$ such that

$$\boldsymbol{W}\boldsymbol{W}^\top = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^\top = \boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements $d_1 \geq \cdots \geq d_p \geq 0$. If $d_1 \geq \cdots \geq d_q > d_{q+1} = \cdots = d_p \geq 0$ and $\boldsymbol{W}$ is partitioned accordingly as $\boldsymbol{W} = (\boldsymbol{W}_1^\top, \boldsymbol{W}_2^\top)^\top$, then $\boldsymbol{W}_1\boldsymbol{x}$ is often seen as the q-variate signal part and $\boldsymbol{W}_2\boldsymbol{x}$ as the $(p-q)$-variate noise part. Hence, $\boldsymbol{W}_2\boldsymbol{x}$ is considered noise if and only if the eigenvalues of $\boldsymbol{W}_2\boldsymbol{S}\boldsymbol{W}_2^\top$ are all equal.

* Corresponding author.
E-mail address: hannu.oja@utu.fi (H. Oja).

https://doi.org/10.1016/j.jmva.2021.104830

Please cite this article as: K. Nordhausen, H. Oja and D.E. Tyler, Asymptotic and bootstrap tests for subspace dimension, Journal of Multivariate Analysis (2021) 104830, https://doi.org/10.1016/j.jmva.2021.104830.

(ii) In the independent component analysis (ICA) with q non-Gaussian and $p-q$ Gaussian components, the fourth order blind identification (FOBI) method finds a transformation matrix $\boldsymbol{W} \in \mathbb{R}^{p\times p}$ such that

$$\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^\top = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{W}\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{S}^{-1}\boldsymbol{x}\boldsymbol{x}^\top\right]\boldsymbol{W}^\top = \boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix with the diagonal elements ordered so that $(d_1-(p+2))^2 \geq \cdots \geq (d_q-(p+2))^2 > (d_{q+1}-(p+2))^2 = \cdots = (d_p-(p+2))^2 = 0$. Then $\boldsymbol{W}$ can again be partitioned as $\boldsymbol{W} = (\boldsymbol{W}_1^\top, \boldsymbol{W}_2^\top)^\top$ so that, under weak assumptions, $\boldsymbol{W}_1\boldsymbol{x}$ is the q-variate non-Gaussian signal and $\boldsymbol{W}_2\boldsymbol{x}$ the $(p-q)$-variate Gaussian noise. If we further write $\boldsymbol{S}_1 := \boldsymbol{S}$ and $\boldsymbol{S}_2 := \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{S}_1^{-1}\boldsymbol{x}\boldsymbol{x}^\top\right]$ then, $\boldsymbol{W}_2\boldsymbol{x}$ is considered noise if the eigenvalues of $\boldsymbol{W}_2\boldsymbol{S}_2\boldsymbol{W}_2^\top$ are all equal to $p+2$.

(iii) In the sliced inverse regression (SIR) with a p-variate random vector $\boldsymbol{x}$ and the response (dependent) variable y, one finds a matrix $\boldsymbol{W} \in \mathbb{R}^{p\times p}$ which satisfies

$$\boldsymbol{W}\boldsymbol{S}_1\boldsymbol{W}^\top = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{W}\boldsymbol{S}_2\boldsymbol{W}^\top = \boldsymbol{D}$$

where $\boldsymbol{S}_1 := \boldsymbol{S}$ and $\boldsymbol{S}_2 := \mathbb{E}\left[\mathbb{E}(\boldsymbol{x}|y)\mathbb{E}(\boldsymbol{x}|y)^\top\right]$ and $\boldsymbol{D}$ is a diagonal matrix with the diagonal elements $d_1 \geq \cdots \geq d_p \geq 0$. Under appropriate assumptions on the distribution of $(\boldsymbol{x}, y)$, we have $d_1 \geq \cdots \geq d_q > d_{q+1} = \cdots = d_p = 0$ with the corresponding partitioning $\boldsymbol{W} = (\boldsymbol{W}_1^\top, \boldsymbol{W}_2^\top)^\top$. It is then thought that $(\boldsymbol{W}_1\boldsymbol{x}, y)$ carries all the information about the dependence between $\boldsymbol{x}$ and y, and $\boldsymbol{W}_2\boldsymbol{x}$ just presents noise. Thus, $\boldsymbol{W}_2\boldsymbol{x}$ is thought to be noise if the eigenvalues of $\boldsymbol{W}_2\boldsymbol{S}_2\boldsymbol{W}_2^\top$ are all equal to zero.

To test and estimate the dimension of the signal space (also called order determination) and to separate signal and noise, we thus utilize, for empirical versions of appropriate choices of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, the eigen-decomposition of $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$, or that of the symmetric matrix $\boldsymbol{R} := \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$. For the PCA case, we take $\boldsymbol{S}_1 = \boldsymbol{I}_p$ and $\boldsymbol{S}_2 = \hat{\boldsymbol{S}}$, the sample covariance matrix, or some other scatter matrix, as defined later in Section 2. The tests are based on the first two moments of selected subsets of the eigenvalues of $\boldsymbol{R}$ and the corresponding estimates are obtained applying different sequential testing strategies. The sequential testing procedures for the order determination problem in SIR have been suggested earlier by Li [23] and Bura and Cook [4]. Zhu et al. [46,47] used the eigenvalues with BIC-type penalties to find consistent estimates for the dimension of the signal subspace of a regression model. In other general approaches, Ye and Weiss [45] considered eigenvectors rather than eigenvalues and proposed an estimation procedure that was based on the bootstrap variation of the subspace estimates for different dimensions. In a general approach, Luo and Li [26] combined the eigenvalues and bootstrap variation of eigenvectors for consistent estimation of the dimension. The last two approaches are based on the notion that the variation of eigenvectors is large for the eigenvalues that are close together and their variability tends to be small for far apart eigenvalues.

In PCA the eigenvalues of $\hat{\boldsymbol{S}}$ are generally used to make inference on the dimension of the signal space, see e.g. Jolliffe [19] and Schott [36] and references therein. Early papers on the use of bootstrap estimation and testing (via confidence intervals) in principal component analysis are Beran and Srivastava [1], Daudin et al. [11], Eaton and Tyler [13] and Jackson [18]. For the use of permutation tests in restricting the number of principal components, see Dray [12] and Vieira [42].

In the independent component analysis (ICA) the fourth-order blind identification (FOBI) by Cardoso [6] uses the regular covariance matrix and the scatter matrix based on fourth moments and the eigenvalues provide measures of marginal kurtosis. These two matrices can be replaced by any two matrices possessing the so called independence property, see Oja et al. [34], Tyler et al. [41] and Nordhausen and Tyler [33]. Very recently, Nordhausen et al. [30] used the empirical eigenvalues of $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$ to test and estimate the dimensions of Gaussian and non-Gaussian subspaces.

PCA and FOBI are examples of unsupervised dimension reduction procedures as they do not use information on any response variable y. Other examples of unsupervised dimension reduction methods are invariant coordinate selection (ICS), see Tyler et al. [41], and generalized principal components analysis (GPCA), see Caussinus and Ruiz-Gazen [7]. Sliced inverse regression (SIR) uses the regular covariance matrix of $\boldsymbol{x}$ and the covariance matrix of the conditional expectation $E(\boldsymbol{x}|y)$. Other examples on supervised dimension reduction methods are the canonical correlation analysis (CCA), sliced average variance estimate (SAVE) and principal Hessian directions (PHD), for example, and they all can be formulated using two scatter matrices. For these methods and estimation of the dimension of the signal subspace, also with regular bootstrap sampling, see Li [23], Cook and Weisberg [9], Li [24], Bura and Cook [4], Cook [8], Zhu et al. [46,47], Bura and Yang [5] and Luo and Li [26] and the references therein. For nice reviews on supervised dimension reduction, see Ma and Zhu [27], Li [25].

The plan of this paper is as follows. In Section 2 we introduce the tools for our analysis, that is, the notion of a scatter matrix with some preliminary theory. In all three cases in Section 3 (PCA), 4 (FOBI) and 5 (SIR), respectively, we first specify a natural semiparametric model: $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b}$ where $\boldsymbol{A}$ and $\boldsymbol{b}$ are the parameters and the distribution of the standardized $\boldsymbol{z}$ is only partially specified. The null hypothesis says that $\boldsymbol{z}$ can be partitioned as $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top)^\top$ and the first part $\boldsymbol{z}_1$ carries the interesting variation. In the paper, the empirical version of the eigenvalues of $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$, that is, the eigenvalues of $\boldsymbol{R} = \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$, are utilized in this partitioning and used to build tests and estimates for the dimension of $\boldsymbol{z}_1$. We discuss the asymptotic tests with corresponding estimates and provide different strategies for bootstrap testing. Different approaches are illustrated with real data examples. All the proofs are postponed to the final section.

We adapt the following notation. $\mathbb{R}^{p\times p}_{sym}$ and $\mathbb{R}^{p\times p}_{sym,+}$ are the sets of symmetric and positive definite symmetric $p \times p$ matrices, respectively. The first and second moments and the variance of the eigenvalues of $\boldsymbol{R} \in \mathbb{R}^{p\times p}_{sym}$ are denoted by

$$m_1(\boldsymbol{R}) := \mathrm{tr}(\boldsymbol{R})/p, \quad m_2(\boldsymbol{R}) := m_1(\boldsymbol{R}^2), \quad s^2(\boldsymbol{R}) := m_2(\boldsymbol{R}) - m_1^2(\boldsymbol{R}),$$

respectively. If $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top \in \mathbb{R}^{p\times p}_{sym,+}$ is a eigen-decomposition of $\boldsymbol{R}$ then $\boldsymbol{R}^{1/2} := \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^\top \in \mathbb{R}^{p\times q}$ (symmetric version of the square root matrix). Given k matrices $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k$, we write

$$\mathrm{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_k \end{pmatrix}.$$

The vectorization of a matrix $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{p\times q}$, denoted by $\mathrm{vec}(\boldsymbol{A})$, is a qp-vector obtained by stacking the columns of $\boldsymbol{A}$ on top of each other, that is, $\mathrm{vec}(\boldsymbol{A}) = (a_{11}, \ldots, a_{p1}, a_{12}, \ldots, a_{p2}, \ldots, a_{1q}, \ldots, a_{pq})^\top$. We further write $\mathcal{O}^{p\times k}$, $k \le p$, for the set of column orthonormal $p \times k$ matrices, i.e., $\boldsymbol{U} \in \mathcal{O}^{p\times k}$ implies $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_k$. Hence, given $\boldsymbol{U} \in \mathcal{O}^{p\times k}$, $\boldsymbol{P}_{\boldsymbol{U}} := \boldsymbol{U}\boldsymbol{U}^\top$ is the orthogonal projection onto the range of $\boldsymbol{U}$, and $\boldsymbol{Q}_{\boldsymbol{U}} = \boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{U}}$ is the orthogonal projection onto its orthogonal complement, i.e., onto the null space of $\boldsymbol{U}^\top$. Let $\boldsymbol{e}_i \in \mathbb{R}^p$ denote the ith Euclidean basis element, i.e., a vector with a one in the ith position and zeros elsewhere. For two random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we write $\boldsymbol{x} \sim \boldsymbol{y}$ if $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same distribution. The random vector $\boldsymbol{z} \in \mathbb{R}^p$ has a spherical distribution if $\boldsymbol{U}\boldsymbol{z} \sim \boldsymbol{z}$ for all $\boldsymbol{U} \in \mathcal{O}^{p\times p}$. The distribution of $\boldsymbol{z}$ is subspherical with dimension k, $k < p$, if $\boldsymbol{U}^\top\boldsymbol{z}$ is spherical for some $\boldsymbol{U} \in \mathcal{O}^{p\times k}$. The distribution of $\boldsymbol{x}$ is elliptical if $\boldsymbol{x} \sim \boldsymbol{A}\boldsymbol{z}+\boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{b} \in \mathbb{R}^p$ and $\boldsymbol{z} \in \mathbb{R}$ has a spherical distribution.

2. Scatter matrices

In this chapter, we state what we mean by a scatter matrix and a supervised scatter matrix and provide some preliminary results. Let $\boldsymbol{F}_{\boldsymbol{x}}$ be the cumulative distribution function (cdf) of a p-variate random vector $\boldsymbol{x}$ and $\boldsymbol{F}_{\boldsymbol{x},y}$ the cdf of the joint distribution of p-variate $\boldsymbol{x}$ and univariate y.

Definition 1. (i) The functional $\boldsymbol{S}(F_{\boldsymbol{x}}) \in \mathbb{R}^{p\times p}_{sym,+}$ is a scatter matrix (functional) if it is affine equivariant in the sense that $\boldsymbol{S}(F_{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}}) = \boldsymbol{A}\boldsymbol{S}(F_{\boldsymbol{x}})\boldsymbol{A}^\top$ for all non-singular $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ and all $\boldsymbol{b} \in \mathbb{R}^p$.

(ii) The functional $\boldsymbol{S}(F_{\boldsymbol{x},y}) \in \mathbb{R}^{p\times p}_{sym}$ is a supervised scatter matrix (functional) if it is affine equivariant in the sense that $\boldsymbol{S}(F_{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b},y}) = \boldsymbol{A}\boldsymbol{S}(F_{\boldsymbol{x},y})\boldsymbol{A}^\top$ for all non-singular $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ and all $\boldsymbol{b} \in \mathbb{R}^p$.

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n\times p}$ be a random sample from a distribution $F_{\boldsymbol{x}}$. The estimate $\widehat{\boldsymbol{S}}$ of the population value $\boldsymbol{S}(F_{\boldsymbol{x}})$ is obtained as the value of the functional at the empirical distribution F_n of $\boldsymbol{X}$. We also write $\boldsymbol{S}(\boldsymbol{X})$ for this estimate. Let $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{A}^\top + \boldsymbol{1}_n\boldsymbol{b}^\top$ where $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^\top$ is a random sample from a spherical distribution $F_{\boldsymbol{z}}$ with $\boldsymbol{S}(F_{\boldsymbol{z}}) = \boldsymbol{I}_p$. (Note that, for any scatter matrix $\boldsymbol{S}$, $\boldsymbol{S}(F_{\boldsymbol{z}}) \propto \boldsymbol{I}_p$ and can the rescaled to satisfy the last condition.) Then $\boldsymbol{X}$ is a random sample from an elliptical distribution with $\boldsymbol{S}(F_{\boldsymbol{x}}) = \boldsymbol{A}\boldsymbol{A}^\top$.

Under general assumptions, the limiting distribution of $\sqrt{n}\,\mathrm{vec}(\boldsymbol{S}(\boldsymbol{Z}) - \boldsymbol{I}_p)$ is

$$N_{p^2}\left(\boldsymbol{0}, \sigma_1(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{p,p}) + \sigma_2\mathrm{vec}(\boldsymbol{I}_p)\mathrm{vec}(\boldsymbol{I}_p)^\top\right)$$

where $\boldsymbol{K}_{p,p} = \sum_{i=1}^p \sum_{j=1}^p (\boldsymbol{e}_i\boldsymbol{e}_j^\top) \otimes (\boldsymbol{e}_j\boldsymbol{e}_i^\top)$ is the commutation matrix, see Theorem 1 in Tyler [38]. The limiting distribution is known if the following two constants, same for any $i \neq j$,

$$\sigma_1 = \mathrm{AsVar}(\boldsymbol{S}(\boldsymbol{Z})_{ij}), \quad \sigma_2 = \mathrm{AsCov}(\boldsymbol{S}(\boldsymbol{Z})_{ii}, \boldsymbol{S}(\boldsymbol{Z})_{jj})$$

are known and then $\mathrm{AsVar}(\boldsymbol{S}(\boldsymbol{Z})_{ii}) = 2\sigma_1 + \sigma_2$. Also, under general conditions, the influence function of the scatter functional $\boldsymbol{S}(F)$ at a spherical $F_{\boldsymbol{z}}$ is given by

$$\mathrm{IF}(\boldsymbol{x}; \boldsymbol{S}, F_{\boldsymbol{z}}) = \alpha(r)\boldsymbol{u}\boldsymbol{u}^T - \beta(r)\boldsymbol{I}_p,$$

where $r = \|\boldsymbol{x}\|$ and $\boldsymbol{u} = \|\boldsymbol{x}\|^{-1}\boldsymbol{x}$, see Hampel et al. [16]. If $\boldsymbol{S}(F)$ is the covariance matrix and $\boldsymbol{S}(F_{\boldsymbol{z}}) = \boldsymbol{I}_p$, then $\alpha(r) = r^2$ and $\beta(r) = 1$ and if $\boldsymbol{z} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ then $\sigma_1 = 1$ and $\sigma_2 = 0$. For Tyler's shape estimate (proposed in Tyler [40] and scaled so that its trace equals p) which we use as a robust alternative for the covariance matrix in our example in Section 3, one gets $\alpha(r) = (p + 2)$ and $\beta(r) = (p + 2)/p$.

In the following we often need to estimate σ_1. It then follows, as noted in Croux and Haesbroeck [10], that $\sigma_1 = E(\alpha^2(r))/(p(p + 2))$. Due to affine equivariance of the scatter matrix, the limiting distribution of $\sqrt{n}\,\mathrm{vec}(\boldsymbol{S}(\boldsymbol{X}) - \boldsymbol{A}\boldsymbol{A}^\top) = (\boldsymbol{A} \otimes \boldsymbol{A})\sqrt{n}\,\mathrm{vec}(\boldsymbol{S}(\boldsymbol{Z}) - \boldsymbol{I}_p)$ and, using $\widehat{\boldsymbol{S}}$ with a companion location estimate $\hat{\boldsymbol{\mu}}$, σ_1 can often be consistently estimated by

$$\hat{\sigma}_1 = \frac{1}{p(p+2)}\frac{1}{n}\sum_{i=1}^n \alpha^2(\hat{r}_i), \quad \hat{r}_i = \left((\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})\right)^{1/2}.$$

3. Testing for subspace dimension in PCA

3.1. The model, null hypothesis and test statistic

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ be a random sample from a p-variate elliptical distribution $F_{\boldsymbol{x}}$, that is, from the distribution of a random p-vector $\boldsymbol{x}$ generated by

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b},$$

where $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is non-singular, $\boldsymbol{b} \in \mathbb{R}^p$ and $\boldsymbol{z}$ has a spherical distribution around the origin, that is, $\boldsymbol{U}\boldsymbol{z} \sim \boldsymbol{z}$ for all $\boldsymbol{U} \in \mathcal{O}^{p\times p}$. The distribution of $\boldsymbol{z}$ is then fully determined by the distribution of its radius $r := \|\boldsymbol{z}\|$. We assume that $\boldsymbol{S}(F_{\boldsymbol{z}}) = \boldsymbol{I}_p$ for the scatter matrix functional used in the analysis. For a general overview of spherical and elliptical distributions, see Kelker [21] or Bilodeau and Brenner [2].

As the matrix of eigenvectors and the corresponding eigenvalues of $\boldsymbol{S}(F_x)$ are equivariant and invariant, respectively, under orthogonal transformations of $\boldsymbol{x}$, it is not a restriction to assume in our derivations that $\boldsymbol{A}$ is diagonal with positive and descending entries and $\boldsymbol{b} = \boldsymbol{0}$ so that $\boldsymbol{S}(F_{\boldsymbol{x}})$ is a diagonal matrix $\boldsymbol{D} = \boldsymbol{A}^2$ with diagonal entries $d_1 \geq \cdots \geq d_p > 0$. Let $\widehat{\boldsymbol{S}}$ be the value of the scatter functional at the empirical distribution of $\boldsymbol{X}$. For the asymptotic results, we assume that $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{S}} - \boldsymbol{D})$ has a limiting multivariate normal distribution with zero mean vector and the covariance structure as described in Section 2. We wish to test the null hypothesis

$$H_{0k}: \quad d_1 \geq \cdots \geq d_k > d_{k+1} = \cdots = d_p = d \quad \text{for some unknown } d,$$

stating that the dimension of the signal space is k. Under H_{0k}, the distribution of $\boldsymbol{x}$ is subspherical, that is, the distribution of the subvector of the last $p-k$ principal components is spherical. In principal component analysis, the scree plot is often used to figure out how many components to include in the final model. The null hypothesis H_{0k} then implies that there is the elbow on the scree plot at the kth eigenvalue. Also, sphericity and subsphericity (in a weaker sense) are important in the analysis of the repeated measures data, for example.

To test the null hypothesis, we use the variance of the $p-k$ smallest eigenvalues, that is,

$$T_k := s^2(\widehat{\boldsymbol{U}}_k^\top \widehat{\boldsymbol{S}} \widehat{\boldsymbol{U}}_k), \quad \widehat{\boldsymbol{U}}_k = \arg \min_{\boldsymbol{U}\in\mathcal{O}^{p\times(p-k)}} m_1(\boldsymbol{U}^\top \widehat{\boldsymbol{S}} \boldsymbol{U}),$$

as a test statistic. It follows from the Poincaré separation theorem that a solution $\widehat{\boldsymbol{U}}_k \in \mathcal{O}^{p\times(p-k)}$ is the matrix of the eigenvectors associated with the $p-k$ smallest eigenvalues of $\widehat{\boldsymbol{S}}$ and other solutions are obtained by post-multiplying it by an orthogonal $(p-k)\times(p-k)$ matrix. The projection matrices $\widehat{\boldsymbol{P}}_k := \widehat{\boldsymbol{U}}_k\widehat{\boldsymbol{U}}_k^\top$ and $\widehat{\boldsymbol{Q}}_k := \boldsymbol{I}_p - \widehat{\boldsymbol{P}}_k$ are unique and satisfy $\widehat{\boldsymbol{P}}_k\widehat{\boldsymbol{S}}\widehat{\boldsymbol{Q}}_k = \boldsymbol{0}$ and provide the noise-signal decomposition $\boldsymbol{x} = \widehat{\boldsymbol{P}}_k\boldsymbol{x} + \widehat{\boldsymbol{Q}}_k\boldsymbol{x}$ with uncorrelated $\widehat{\boldsymbol{P}}_k\boldsymbol{x}$ and $\widehat{\boldsymbol{Q}}_k\boldsymbol{x}$.

Other possible measures for the variation of the smallest eigenvalues are $s(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_k)/m_1(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_k)$, i.e., the coefficient of variation, or the log ratio of the arithmetic mean $m_1(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_k)$ to the geometrical mean $\det(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_k)^{1/(p-k)}$. If $\widehat{\boldsymbol{S}}$ is the covariance matrix, then the latter measure corresponds to the likelihood ratio criterion for H_{0k} in the multivariate normal case.

If one wishes to test a related null hypothesis that $\boldsymbol{S}(F_{\boldsymbol{x}})$ has $k+1$ distinct eigenvalues with multiplicities $1, \ldots, 1, p-k$, then a natural test statistic is

$$V_k := \min_{\boldsymbol{U}\in\mathcal{O}^{p\times(p-k)}: \boldsymbol{P}_{\boldsymbol{U}}\widehat{\boldsymbol{S}}\boldsymbol{Q}_{\boldsymbol{U}}=\boldsymbol{0}} s^2\left(\boldsymbol{U}^\top\widehat{\boldsymbol{S}}\boldsymbol{U}\right).$$

A solution $\widehat{\boldsymbol{U}}_k$ for which the minimum value is attained consists of the eigenvectors of $\widehat{\boldsymbol{S}}$ associated with the eigenvalues closest together (in the variance sense). This is seen as follows. Let $\boldsymbol{U} \in \mathcal{O}^{p\times(p-k)}$ and $\boldsymbol{P}_{\boldsymbol{U}}\widehat{\boldsymbol{S}}\boldsymbol{Q}_{\boldsymbol{U}} = \boldsymbol{0}$. Then $\boldsymbol{P}_{\boldsymbol{U}}\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}\boldsymbol{P}_{\boldsymbol{U}}$. As the symmetric matrices commute if and only if they have the same eigenvectors, $\boldsymbol{U}$ is a matrix of $p-k$ eigenvectors of $\widehat{\boldsymbol{S}}$, say $\boldsymbol{U}_0 \in \mathcal{O}^{p\times(p-k)}$, post-multiplied by an orthogonal $(p-k)\times(p-k)$ matrix. Consequently, $\boldsymbol{U}^\top\widehat{\boldsymbol{S}}\boldsymbol{U}$ and $\boldsymbol{U}_0^\top\widehat{\boldsymbol{S}}\boldsymbol{U}_0$ have the same eigenvalues and $s^2(\boldsymbol{U}^\top\widehat{\boldsymbol{S}}\boldsymbol{U}) = s^2(\boldsymbol{U}_0^\top\widehat{\boldsymbol{S}}\boldsymbol{U}_0)$. Thus the problem of minimizing $s^2(\boldsymbol{U}^\top\widehat{\boldsymbol{S}}\boldsymbol{U})$ under the constraint $\boldsymbol{P}_{\boldsymbol{U}}\widehat{\boldsymbol{S}}\boldsymbol{Q}_{\boldsymbol{U}} = 0$ reduces to that of minimizing $s^2(\boldsymbol{U}_0^\top\widehat{\boldsymbol{S}}\boldsymbol{U}_0)$ over the $p-k$ subsets of eigenvectors of $\widehat{\boldsymbol{S}}$.

3.2. Asymptotic tests for dimension

Assume now that $\boldsymbol{x}$ is elliptical with diagonal scatter matrix $\boldsymbol{D} = \boldsymbol{A}^2$. Let q denote the true value of the dimension of the signal space, that is, H_{0q} is true, and consider the limiting distribution of $T_q = s^2(\widehat{\boldsymbol{U}}_q^\top\widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_q)$. With a correct value q we have the partitions

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{0} \\ \boldsymbol{0} & d\boldsymbol{I}_{p-q} \end{pmatrix}, \quad \widehat{\boldsymbol{S}} = \begin{pmatrix} \widehat{\boldsymbol{S}}_{11} & \widehat{\boldsymbol{S}}_{12} \\ \widehat{\boldsymbol{S}}_{21} & \widehat{\boldsymbol{S}}_{22} \end{pmatrix},$$

respectively, and the diagonal elements in $\boldsymbol{D}_1$ are strictly larger than d. Under our assumptions, $\sqrt{n}(\widehat{\boldsymbol{S}} - \boldsymbol{D}) = O_P(1)$ and we have the following.

Lemma 1. *Under the stated assumptions and H_{0q}, $nT_q = ns^2(\widehat{\boldsymbol{S}}_{22}) + O_P(n^{-1/2})$.*

Under our assumptions stated in Section 2, $\sqrt{n}\ \mathrm{vec}(\boldsymbol{S}(\boldsymbol{Z}) - \boldsymbol{I}_p)$ where $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{D}^{-1/2}$ converges in distribution to a p^2-variate normal distribution with zero mean vector and the covariance matrix $\sigma_1(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{p,p}) + \sigma_2 \mathrm{vec}(\boldsymbol{I}_p)\mathrm{vec}(\boldsymbol{I}_p)^\top$. Then we have the following.

Theorem 1. *Under the previously stated assumptions and under H_{0q},*

$$\frac{n(p-q)T_q}{2d^2\sigma_1} \xrightarrow{d} \chi^2_{\frac{1}{2}(p-q-1)(p-q+2)}.$$

If the multiplicities of the eigenvalues of $\boldsymbol{D}_1$ are smaller than $p-q$ then $P(V_q = T_q) \to 1$ and the limiting distributions of nV_q and nT_q are the same.

For the test construction in practice we thus need to estimate two population constants σ_1 and d, both of which are invariant under orthogonal transformations to $\boldsymbol{x}$. The limiting distribution in Theorem 1 stays the same even if σ_1 and d are replaced by their consistent estimates, say $\hat{\sigma}_1$ and $\hat{d}$. Construction of a consistent estimate for σ_1 has already been discussed in Section 2. The unknown d can be consistently estimated by the average of the $p-q$ smallest eigenvalues, that is, by $\hat{d} = m_1(\widehat{\boldsymbol{U}}_q^\top \widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_q)$. Note also that the test statistic in Theorem 1 with these replacements depends on the smallest eigenvalues through their coefficient of variation, a test statistic suggested by Schott [36]. As noted previously, a possible test statistic for H_{0q} is also the log of the ratio of the arithmetic and geometric means of the smallest $p-q$ eigenvalues of $\widehat{\boldsymbol{S}}$, say L_q. Then under the null hypotheses as well as under certain contiguous alternatives, $n(T_q - 2d^2L_q) \xrightarrow{p} 0$ and then, under H_{0q}, $n(p-q)L_q/\hat{\sigma}_1 \xrightarrow{d} \chi^2_{(p-q-1)(p-q+2)/2}$. See Theorem 5.1 and 5.2 and their proofs in Tyler [39].

We now utilize the test statistics T_k, $k \in \{0, 1, \ldots, p-1\}$, for the estimation problem and collect some useful limiting properties in the following theorem.

Theorem 2. *Under the previously stated assumptions and under H_{0q},*

(i) *for $k < q$, $T_k \xrightarrow{P} c_k$ for some $c_1, \ldots, c_{q-1} > 0$,*
(ii) *for $k = q$, $n(p-q)T_q/(2d^2\sigma_1) \xrightarrow{d} \chi^2_{\frac{1}{2}(p-q-1)(p-q+2)}$,*
(iii) *for $k > q$, $nT_k \le (\frac{p-q}{p-k})^2 nT_q = O_P(1)$.*

A consistent estimate $\hat{q}$ of the unknown dimension $q \le p-1$ can then be based on the test statistics T_k, $k \in \{0, 1, \ldots, p-1\}$, as follows.

Corollary 1. *For all $k \in \{0, 1, \ldots, p-1\}$, let $(c_{k,n})$ be a sequence of positive real numbers such that $c_{k,n} \to 0$ and $nc_{k,n} \to \infty$ as $n \to \infty$. Then, under the assumptions of Theorem 2,*

$$\mathbb{P}(T_k \ge c_{k,n}) \ \to \ \begin{cases} 1, & \text{if } k < q, \\ 0, & \text{if } k \ge q, \end{cases}$$

and $\hat{q} = \min\{k \ : \ T_k < c_{k,n}\} \xrightarrow{P} q$.

Note that, by definition, $T_{p-1} = 0$ and the maximum value of q is $p-1$, which corresponds to the smallest eigenvalue being distinct. The estimate $\hat{q}$ is easily found by using the so called bottom-up testing strategy: Start with tests for H_{00}, H_{01} and so on, and stop when you get the first acceptance. An alternative consistent estimate with a top-down testing strategy is $\hat{q} = \max\{k \ : \ T_{k-1} \ge c_{k-1,n}\}$ using successive tests for $H_{0,p-2}, H_{0,p-3}, \ldots$, and stopping after the first rejection. For large p, faster strategies such as the divide and conquer algorithm are naturally available in the estimation.

Let F_k be the limiting distribution of nT_k under H_{0k}. The sequences of critical values $(c_{k,n})$ for testing H_{0k} can be determined by the corresponding sequences of asymptotical test sizes $(\alpha_{k,n})$ satisfying $\alpha_{k,n} = 1 - F_k(nc_{k,n})$ A simple and practical choice of the sequences of the test sizes is for example $\alpha_{k,n} = (n_0/n)\alpha_k$, $k \le p-2$ and $n \ge n_0$. Then $nc_{k,n} \to \infty$ as $\alpha_{k,n} = 1 - F_k(nc_{k,n}) \to 0$, and $c_{k,n} \to 0$ as $nc_{k,n}\alpha_{k,n} = nc_{k,n}(1 - F_k(nc_{k,n})) \to 0$.

To end the discussion on asymptotics, suppose we relax now the ellipticity assumption and consider a model for which $\mathrm{diag}(\boldsymbol{I}_q, \boldsymbol{U})\boldsymbol{z} \sim \boldsymbol{z}$ for all $\boldsymbol{U} \in \mathcal{O}^{(p-q)\times(p-q)}$. Since $\boldsymbol{D} = \boldsymbol{A}^2 = \mathrm{diag}(\boldsymbol{D}_1, d\boldsymbol{I}_{p-q})$, $\boldsymbol{x}$ is subspherical but not necessarily elliptical. It is then easy to show that, for the covariance matrix and finite fourth moments, Lemma 1 and Theorem 1 still hold true with $\sigma_1 = 1$. For other scatter matrices, however, the asymptotic behavior in this wider model is not known.

Lemma 1 shows the remarkable fact that under the null hypothesis H_{0q} the limiting distributions of $nT_q = ns^2(\widehat{\boldsymbol{U}}_q^\top \widehat{\boldsymbol{S}}\widehat{\boldsymbol{U}}_q)$ and that of $ns^2(\boldsymbol{U}_q^\top \widehat{\boldsymbol{S}}\boldsymbol{U}_q)$ with known noise subspace are the same. If, in the small sample case, the p-values are obtained from the limiting distribution of the test statistic, the variation coming from the estimation of the subspace is thus ignored in the null asymptotic approximation. In the following we therefore propose that the small sample null distribution of a test statistic be estimated by resampling the data from a distribution obeying the null hypothesis and being as close as possible to the empirical distribution.

3.3. Bootstrap tests for dimension

Again, let q denote the true dimension of the signal space and we wish to test the null hypothesis

$$H_{0k}: \quad d_1 \geq \cdots \geq d_k > d_{k+1} = \cdots = d_p = d \ \text{ for some } d.$$

It is important to stress that, in the practical testing situation, we do not know whether H_{0k} is true ($k = q$) or whether it is false ($k \neq q$) but we still wish to compute the p-values for true H_{0k}. See Hall and Wilson [15] for some guidelines in bootstrap hypothesis testing. For testing, we start with a scatter matrix estimate $\widehat{\boldsymbol{S}}$ and a companion location estimate $\widehat{\boldsymbol{\mu}}$ and compute $\widehat{\boldsymbol{U}}_k$ and $T_k = s^2(\widehat{\boldsymbol{U}}_k^\top \widehat{\boldsymbol{S}} \widehat{\boldsymbol{U}}_k)$, the variance of $p-k$ smallest eigenvalues of $\widehat{\boldsymbol{S}}$. We further write $\widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{U}}_k^\top$ and $\widehat{\boldsymbol{Q}}_k = \boldsymbol{I}_p - \widehat{\boldsymbol{P}}_k$ for the estimated projection matrices to the noise and signal subspace under true H_{0k}, respectively.

The basic idea in the bootstrap testing strategy is that the bootstrap samples $\boldsymbol{X}^*$ for H_{0k} should be generated from a distribution $F_{n,k}$

(i) for which the null hypothesis H_{0k} is true (even if $k \neq q$) and
(ii) which is as close as possible to the empirical distribution F_n of $\boldsymbol{X}$.

We suggest the following two procedures. In the first procedure, the bootstrap samples come from a subspherical and elliptical distribution (with the distribution of the radius estimated from the data) and, in the second procedure, they come a subspherical distribution (not assuming full ellipticity). It is important that the dimension of the subspherical part is $p-k$ even when $k \neq q$. If one wishes to assume multivariate normality then the first procedure can be further modified accordingly.

Bootstrap strategy PCA-I (elliptical subspherical distribution):

1. Starting with $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, compute $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{S}}$ with the estimated matrix of eigenvectors in $\widehat{\boldsymbol{U}}$ and corresponding estimated eigenvalues in $\widehat{\boldsymbol{D}}$.
2. Take a bootstrap sample $\widetilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n)^\top$ of size n from $(\boldsymbol{X} - \boldsymbol{1}_n \widehat{\boldsymbol{\mu}}^\top)\widehat{\boldsymbol{U}}\widehat{\boldsymbol{D}}^{-1/2}$.
3. For ellipticity to be true, transform
$$\boldsymbol{z}_i^* = \boldsymbol{O}_i \tilde{\boldsymbol{z}}_i, \quad i \in \{1, \ldots, n\},$$
and $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_n \in \mathcal{O}^{p\times p}$ are i.i.d. from the Haar distribution.
4. For subsphericity to be true as well, the bootstrap sample is
$$\boldsymbol{X}^* = \boldsymbol{Z}^* \widehat{\boldsymbol{D}}_k^{1/2} \widehat{\boldsymbol{U}}^\top + \boldsymbol{1}_n \widehat{\boldsymbol{\mu}}^\top,$$
where $\widehat{\boldsymbol{D}}_k = \mathrm{diag}(\hat{d}_1, \ldots, \hat{d}_k, \sum_{i=k+1}^p \hat{d}_i/(p-k), \ldots, \sum_{i=k+1}^p \hat{d}_i/(p-k))$.

Bootstrap strategy PCA-II (subspherical distribution):

1. Starting with $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, compute $\widehat{\boldsymbol{S}}$, $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{U}}_k$, $\widehat{\boldsymbol{P}}_k$ and $\widehat{\boldsymbol{Q}}_k$.
2. Take a bootstrap sample $\widetilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n)^\top$ of size n from $\boldsymbol{X}$.
3. For subsphericity to be true, transform
$$\boldsymbol{x}_i^* = \left[\widehat{\boldsymbol{Q}}_k + \widehat{\boldsymbol{U}}_k \boldsymbol{O}_i \widehat{\boldsymbol{U}}_k^\top\right](\tilde{\boldsymbol{x}}_i - \widehat{\boldsymbol{\mu}}) + \widehat{\boldsymbol{\mu}}, \quad i \in \{1, \ldots, n\},$$
and $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_n \in \mathcal{O}^{(p-k)\times(p-k)}$ are i.i.d. from the Haar distribution.
4. The bootstrap sample is $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)$.

For both strategies and for $k \in \{0, \ldots, p-1\}$, the hypothesis H_{0k} is true for the corresponding bootstrap null distribution, say $F_{n,k}$. For the PCA-I strategy,

$$F_{n,k}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\boldsymbol{O}_{i,p}} \left[\mathbb{I}\left(\widehat{\boldsymbol{U}}_k \widehat{\boldsymbol{D}}_k^{1/2} \boldsymbol{O}_{i,p} \widehat{\boldsymbol{D}}^{-1/2} \widehat{\boldsymbol{U}}_k^\top (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}) + \widehat{\boldsymbol{\mu}} \leq \boldsymbol{x} \right) \right]$$

with random matrices $\boldsymbol{O}_{1,p}, \ldots, \boldsymbol{O}_{n,p} \in \mathcal{O}^{p\times p}$ from the Haar distribution. Similarly, for the PCA-II strategy,

$$F_{n,k}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\boldsymbol{O}_{i,p-k}} \left[\mathbb{I}\left((\widehat{\boldsymbol{Q}}_k + \widehat{\boldsymbol{U}}_k \boldsymbol{O}_{i,p-k} \widehat{\boldsymbol{U}}_k^\top)(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}) + \widehat{\boldsymbol{\mu}} \leq \boldsymbol{x} \right) \right],$$

where $O_{1,p-k}, \ldots, O_{n,p-k} \in \mathcal{O}^{(p-k)\times(p-k)}$ are from the Haar distribution.

Consider next the distribution of $nT_k(\boldsymbol{X}^*)$ for the PCA-I strategy. Let then $\boldsymbol{X}_N^* \in \mathbb{R}^{N\times p}$ be a random sample of size N from $F_{n,k}$. Note that $F_{n,k}$ is an elliptical distribution with true H_{0k} and with data dependent parameters, namely, symmetry center $\boldsymbol{\mu} := \widehat{\boldsymbol{\mu}}$, covariance matrix $\boldsymbol{S} := \widehat{\boldsymbol{U}}\widehat{\boldsymbol{D}}_k \widehat{\boldsymbol{U}}^\top$ and

$$d := \hat{d} = \frac{1}{p-k} \sum_{i=k+1}^p \hat{d}_i, \qquad \sigma_1 := \hat{\sigma}_1 = \frac{1}{p(p+2)} \frac{1}{n} \sum_{i=1}^n \alpha^2(\hat{r}_i),$$

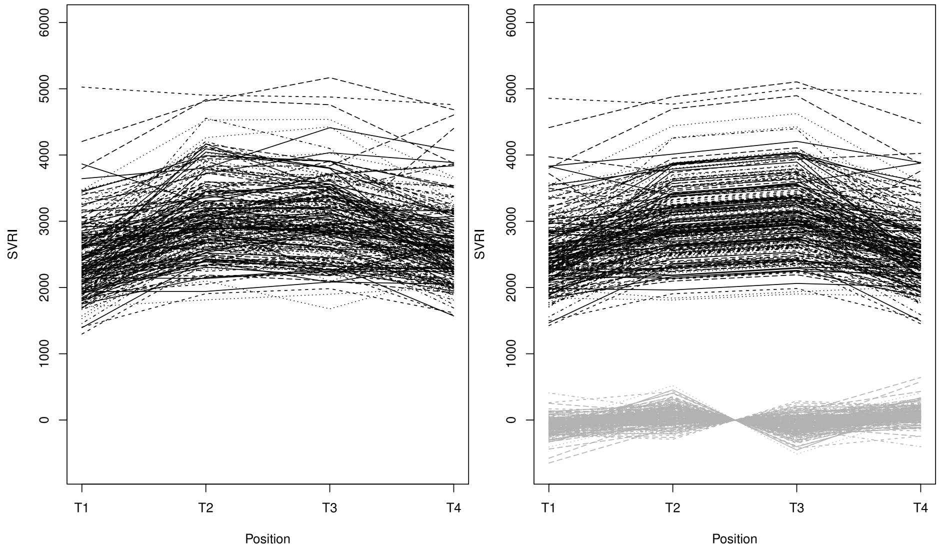

Fig. 1. Left figure: The original data set consisting of the SVRI values measured on 223 subjects at 4 time points. Right figure: The estimated signal part (upper curves) and noise part (lower part) of the same data set.

where $\hat{r}_i = ((\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{S}}^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}))^{1/2}$, $i \in \{1, \dots, n\}$. Theorem 1 then implies that, given $\boldsymbol{X}$, $N(p-k)T_k(\boldsymbol{X}_N^*)/(2\hat{d}^2\hat{\sigma}_1) \xrightarrow{d} \chi^2_{\frac{1}{2}(p-k-1)(p-k+2)}$ (a.s.) which provides, for large n, the same asymptotic chi-squared approximation for the distribution of the unconditional $n(p-k)T_k(\boldsymbol{X}^*)/(2\hat{d}^2\hat{\sigma}_1)$ as well. Theorem 1 gave the same approximation for $n(p-k)T_k(\boldsymbol{X})/(2\hat{d}^2\hat{\sigma}_1)$. For the PCA-I strategy applied to the covariance matrix, similar arguments can be used to get the same approximations for the distributions of $n(p-k)T_k(\boldsymbol{X}^*)/(2\hat{d}^2)$ and $n(p-k)T_k(\boldsymbol{X})/(2\hat{d}^2)$.

In practice, the exact p-values are not computed but estimated as follows. Let $T = T(\boldsymbol{X})$ be a test statistic for H_{0k} such as T_k, that is, the variance of the $p-k$ smallest eigenvalues of $\widehat{\boldsymbol{S}}$. If $\boldsymbol{X}_1^*, \dots, \boldsymbol{X}_M^*$ are independent bootstrap samples of size n as described above and $T_i^* = T(\boldsymbol{X}_i^*)$, $i \in \{1, \dots, M\}$, then the bootstrap p-value is given by

$$\hat{p} = \frac{\#(T_i^* \geq T) + 1}{M+1}.$$

Note that, conditioned on $\boldsymbol{X}$, $\hat{p}$ is a random variable whose variance around the true p-value can be estimated by $\frac{1}{M}\hat{p}(1-\hat{p})$.

The asymptotic and bootstrap tests discussed here have been extended to a noisy latent model framework, for example, in Virta and Nordhausen [43].

3.4. An example

The standard repeated measures ANOVA needs the assumption of spherical multivariate normality. Sphericity has then been defined both in terms of the variances of difference scores and in terms of the variances and covariances of orthogonal contrasts to be used in the analysis, see e.g., Lane [22]. Preliminary testing for sphericity or subsphericity is then of interest in this context. Subsphericity indicates that there are no latent subgroups or clusters in that part of the data, and the subspherical part may then be seen simply as noise. To illustrate the methodology we use some data from the LASERI study (Cardiovascular risk in young Finns study) which is available in the R package ICSNP [32]. To collect these data, 223 subjects took part in a tilt-table test. For the first ten minutes the subjects were lying on a motorized table in a supine position, then the table was tilted to a head-up position for five minutes, and thereafter returned to the supine position for the last five minutes. Various hemodynamic variables were measured during the experiment. The variable considered here consists of the four measurements of the systemic vascular resistance index (SVRI) on all subjects. The four time points were (i) the tenth supine minute before the tilt, the (ii) second and (iii) fifth minute during the tilt and (iv) the fifth minute in supine position after the tilting. The 223 SVRI values at the 4 time points are shown in Fig. 1 (left figure).

To illustrate the three testing strategies from above we use as scatter matrix the sample covariance matrix and Tyler's shape matrix where the location is estimates as specified in Hettmansperger and Randles [17]. The obtained eigenvalues of the sample covariance matrix and Tyler's shape matrix are then 982935.95, 176465.68, 36213.91, 25865.65 and 8.94,

Table 1
The p-values for testing $q = 0$ (H_{00}), $q = 1$ (H_{01}) and $q = 2$ (H_{02}) based on the covariance matrix and Tyler's shape matrix for the SVRI data. The p-values are calculated using three different testing strategies.

	Cov			Tyler's shape matrix		
	Asymp	PCA-I	PCA-II	Asymp	PCA-I	PCA-II
H_{00}	0.000	0.002	0.002	0.000	0.002	0.002
H_{01}	0.000	0.002	0.002	0.000	0.002	0.002
H_{02}	0.104	0.130	0.142	0.064	0.072	0.064

1.78, 0.30, 0.21, respectively, and the corresponding eigenvectors are the columns of

$$\begin{pmatrix} -0.48 & 0.46 & -0.42 & 0.62 \\ -0.51 & -0.53 & -0.56 & -0.38 \\ -0.52 & -0.44 & 0.64 & 0.36 \\ -0.50 & 0.56 & 0.31 & -0.59 \end{pmatrix}, \quad \begin{pmatrix} -0.47 & 0.52 & -0.13 & 0.70 \\ -0.51 & -0.48 & -0.70 & -0.11 \\ -0.53 & -0.47 & 0.69 & 0.12 \\ -0.48 & 0.52 & 0.10 & -0.70 \end{pmatrix}.$$

Both scatter matrices seem to suggest that $q = 2$ and that the principal components are (close) to the average and the contrast comparing the supine and tilted positions and the two contrasts within positions. The suggestion $q = 2$ is supported by the p-values for H_{00}, H_{01} and H_{02} using the two scatter matrices and three testing strategies, see Table 1. The estimated signal and noise parts of the data using Tyler's scatter matrix are given in Fig. 1 (right figure).

4. Testing for subspace dimension in FOBI

4.1. The model, null hypothesis and test statistic

In the independent component (IC) model it is assumed that $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ is a random sample from a distribution of the form

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b}$$

where $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is non-singular, $\boldsymbol{b} \in \mathbb{R}^p$, and $\boldsymbol{z}$ is a random p-vector with independent components standardized so that $\mathbb{E}(\boldsymbol{z}) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{z}) = \boldsymbol{I}_p$. We further assume that $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top)^\top$ where the components of $\boldsymbol{z}_1 \in \mathbb{R}^q$ (signal) are non-Gaussian and the components of $\boldsymbol{z}_2 \in \mathbb{R}^{p-q}$ (noise) are Gaussian. The general idea then is to make inference on the unknown q, $0 \le q \le p$, and to estimate the non-Gaussian signal and Gaussian noise subspaces. In this chapter we discuss some recent tests and estimates for q introduced in Nordhausen et al. [30] that are based on the joint use of the covariance matrix and the matrix of fourth moments. Throughout this chapter we therefore need to assume that the fourth moments of $\boldsymbol{z}$ exist.

In the independent component analysis (ICA) it is usually assumed that q is $p - 1$ or p. If $1 \le q \le p$ is allowed as in our case, the approach is sometimes called non-Gaussian independent component analysis (NGICA). In the non-Gaussian component/subspace analysis (NGCA), $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent, $\boldsymbol{z}_1$ is non-Gaussian and $\boldsymbol{z}_2$ is Gaussian, that is, there is no $\boldsymbol{a}_1 \in \mathbb{R}^q$ such that $\boldsymbol{a}_1^\top \boldsymbol{z}_1$ has a normal distribution while $\boldsymbol{a}_2^\top \boldsymbol{z}_2$ has a normal distribution for all $\boldsymbol{a}_2 \in \mathbb{R}^{p-q}$. The components of $\boldsymbol{z}_1$ are thus allowed to be dependent in the NGCA model. See Blanchard et al. [3], Theis et al. [37] and Nordhausen et al. [30].

In fourth order blind identification (FOBI) an unmixing matrix $\boldsymbol{W} \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{p\times p}$ are found such that

$$\boldsymbol{W}\boldsymbol{S}_1\boldsymbol{W}^\top = \boldsymbol{I}_p, \quad \boldsymbol{W}\boldsymbol{S}_2\boldsymbol{W}^\top = \boldsymbol{D},$$

where $\boldsymbol{S}_1 = \mathbb{E}\left[(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\top\right]$ and $\boldsymbol{S}_2 = \mathbb{E}\left[r^2(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\top\right]$ with $r^2 = (\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\top \boldsymbol{S}_1^{-1}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))$ is the scatter matrix based on fourth moments. The matrix $\boldsymbol{W}$ is called an unmixing matrix as $\boldsymbol{W}\boldsymbol{x}$ has independent components under the assumption that $E(z_1^4), \ldots, E(z_q^4)$ are distinct from one another and from 3 (normal case). Write $\boldsymbol{U}^\top = \boldsymbol{W}\boldsymbol{S}_1^{1/2}$. As $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_p$, $\boldsymbol{U}$ is orthogonal and $\boldsymbol{W} = \boldsymbol{U}^\top\boldsymbol{S}_1^{-1/2}$. If

$$\boldsymbol{R} := \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2},$$

then $\boldsymbol{W}\boldsymbol{S}^{1/2}\boldsymbol{R}\boldsymbol{S}^{1/2}\boldsymbol{W}^\top = \boldsymbol{U}^\top\boldsymbol{R}\boldsymbol{U} = \boldsymbol{D}$ and $\boldsymbol{U}$ is therefore obtained from the eigen-decomposition $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$. The eigenvalue d_i in $\boldsymbol{D}$ is then $p + 2$ if and only if $E(z_i^4) = 3$, $i \in \{1, \ldots, p\}$, and, under mild assumptions, the eigenvalues can be used to separate the Gaussian and non-Gaussian components. As $\boldsymbol{W}(F_{\boldsymbol{A}\boldsymbol{x}})\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{W}(F_{\boldsymbol{x}})\boldsymbol{x}$ are the same up to sign changes, location shifts and perturbations of the coordinates and the ordered eigenvalues of $\boldsymbol{D}(F_{\boldsymbol{A}\boldsymbol{x}})$ and of $\boldsymbol{D}(F_{\boldsymbol{x}})$ are the same, we can in our derivations assume without any loss of generality that $\boldsymbol{A} = \boldsymbol{I}_p$, $\boldsymbol{b} = \boldsymbol{0}$ and $\boldsymbol{S}_1 = \boldsymbol{I}_p$, $\boldsymbol{S}_2 = \boldsymbol{R} = \boldsymbol{D} = \mathrm{diag}(\boldsymbol{D}_1, (p+2)\boldsymbol{I}_{p-q})$. For our approach, we also need the assumption that the diagonal elements in $\boldsymbol{D}_1$ are distinct from $p + 2$.

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ be a random sample from the stated independent component model with q non-Gaussian and $p-q$ Gaussian independent components with an unknown dimension q. Write $\widehat{\boldsymbol{S}}_1$, $\widehat{\boldsymbol{S}}_2$ and $\widehat{\boldsymbol{R}}$ for the values of functionals $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and $\boldsymbol{R}$, respectively, at the empirical distribution of $\boldsymbol{X}$. If $\sqrt{n}(\widehat{\boldsymbol{S}}_1 - \boldsymbol{I}_p) = O_P(1)$ and $\sqrt{n}(\widehat{\boldsymbol{S}}_2 - \boldsymbol{D}) = O_P(1)$ then, by Slutsky's theorem,

$$\sqrt{n}(\widehat{\boldsymbol{R}} - \boldsymbol{D}) = \sqrt{n}(\widehat{\boldsymbol{S}}_2 - \boldsymbol{D}) - \frac{1}{2}\left[\sqrt{n}(\widehat{\boldsymbol{S}}_1 - \boldsymbol{I}_p)\boldsymbol{D} + \boldsymbol{D}\sqrt{n}(\widehat{\boldsymbol{S}}_1 - \boldsymbol{I}_p)\right] + o_P(1)$$

and the limiting multivariate normality of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{R}} - \boldsymbol{D})$ follows from the joint limiting multivariate normality of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{S}}_1 - \boldsymbol{I}_p, \widehat{\boldsymbol{S}}_2 - \boldsymbol{D})$ which holds if the eight moments of $\boldsymbol{z}$ exist. We wish to test the null hypothesis

$$H_{0k}: \quad \text{exactly } p-k \text{ eigenvalues in } \boldsymbol{D} \text{ are } p+2$$

stating that the dimension of the signal space is k. To test the null hypothesis H_{0k}, we use the test statistic

$$T_k := \min_{\boldsymbol{U} \in \mathcal{O}^{p\times(p-k)}} m_2\left(\boldsymbol{U}^\top(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)\boldsymbol{U}\right) = \min_{\boldsymbol{U} \in \mathcal{O}^{p\times(p-k)}} m_1\left(\boldsymbol{U}^\top(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)^2\boldsymbol{U}\right).$$

Recall that Kankainen et al. [20] used $T_0 = m_2\left(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p\right)$ to test for full multivariate normality of $\boldsymbol{x}$. If

$$\widehat{\boldsymbol{U}}_k = \arg\min_{\boldsymbol{U} \in \mathcal{O}^{p\times(p-k)}} m_1\left(\boldsymbol{U}^\top(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)^2\boldsymbol{U}\right),$$

then, again according to the Poincaré separation theorem, a solution of $\widehat{\boldsymbol{U}}_k$ is the matrix of the eigenvectors associated with the $p-k$ eigenvalues of $\boldsymbol{R}$ that are closest to $p+2$. We can then also write

$$T_k = m_2\left(\widehat{\boldsymbol{U}}_k^\top(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)\widehat{\boldsymbol{U}}_k\right) = s^2\left(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_k\right) + \left[m_1\left(\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_k\right) - (p+2)\right]^2$$

and $\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}_1^{-1/2}\boldsymbol{x}$ is, under H_{0k}, an estimate for the Gaussian noise vector.

4.2. Asymptotic tests for dimension

Consider the independent component model and, without loss of generality, presume $\boldsymbol{A} = \boldsymbol{I}_p$ and $\boldsymbol{b} = \boldsymbol{0}$. Let q denote the dimension of the non-Gaussian signal space, and denote the corresponding partition by

$$\widehat{\boldsymbol{R}} = \begin{pmatrix} \widehat{\boldsymbol{R}}_{11} & \widehat{\boldsymbol{R}}_{12} \\ \widehat{\boldsymbol{R}}_{21} & \widehat{\boldsymbol{R}}_{22} \end{pmatrix}.$$

We then have the following result.

Lemma 2. *Under the previously stated assumptions and under H_{0q},*

$$nT_q = n \cdot m_2\left(\widehat{\boldsymbol{R}}_{22} - (p+2)\boldsymbol{I}_{p-q}\right) + O_P(n^{-1/2}) = n \cdot s^2\left(\widehat{\boldsymbol{R}}_{22}\right) + n\left[m_1(\widehat{\boldsymbol{R}}_{22}) - (p+2)\right]^2 + O_P(n^{-1/2}).$$

Note that the first term in the sum on the second row provides a test statistic for the equality of $p-q$ eigenvalues closest to $p+2$ and the second term measures the deviation of their average from $p+2$ (Gaussian case). Under our assumptions and under H_{0q}, these two random variables are asymptotically independent and we have the following.

Theorem 3. *Under the previously stated assumptions and under H_{0q},*

$$n(p-q)T_q \xrightarrow{d} 2\sigma_1\chi^2_{\frac{1}{2}(p-q-1)(p-q+2)} + (2\sigma_1 + \sigma_2(p-q))\,\chi^2_1$$

with independent chi squared variables $\chi^2_{\frac{1}{2}(p-q-1)(p-q+2)}$ and χ^2_1, and $\sigma_1 = Var\left(\|\boldsymbol{z}\|^2\right) + 8$ and $\sigma_2 = 4$.

Recall that $T_q = T_{q,1} + T_{q,2}$ where $T_{q,1} = s^2(\widehat{\boldsymbol{U}}_q^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_q)$ and $T_{q,2} = [m_1\left(\widehat{\boldsymbol{U}}_q^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_q\right) - (p+2)]^2$ provide two asymptotically independent test statistics for H_{0q} as seen from the proof of the theorem. Under the assumptions in Theorem 3, $n(p-q)T_{q,1} \xrightarrow{d} 2\sigma_1\chi^2_{\frac{1}{2}(p-q-1)(p-q+2)}$ and $n(p-q)T_{q,2} \xrightarrow{d} (2\sigma_1 + \sigma_2(p-q))\,\chi^2_1$. For deriving the values of σ_1 and σ_2, see the appendix in Nordhausen et al. [30]. They show that the result is true even in the wider NGCA model. As seen in the proof, $\sigma_1 = \mathrm{AsVar}((\widehat{\boldsymbol{R}}_{22})_{12})$ and $\sigma_2 = \mathrm{AsCov}((\widehat{\boldsymbol{R}}_{22})_{11}, (\widehat{\boldsymbol{R}}_{22})_{22})$. In the independent component model, we simply have $\sigma_1 = \sum_{k=1}^p E(z_k^4) - p + 8$ with a consistent estimate $\hat{\sigma}_{1a} = \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^p(\hat{z}_i)_k^4 - p + 8$ where $\hat{\boldsymbol{z}}_i = \widehat{\boldsymbol{W}}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$, $i \in \{1, \ldots, n\}$. In the wider NCGA model, the parameter σ_1 can be consistently estimated by $\hat{\sigma}_{1b} = \frac{1}{n}\sum_{i=1}^n\|\hat{\boldsymbol{z}}_i\|^4 - p^2 + 8$. Both estimates, $\hat{\sigma}_{1a}$ and $\hat{\sigma}_{1b}$, are consistent in the case of the independent component model even for unknown q.

To estimate q, we consider the joint limiting behavior of test statistics $n(p-k)T_k$ for H_{0k}, $k \in \{0, \ldots, p-1\}$, but under true H_{0q}. For $k \in \{0, \ldots, p-1\}$, write

$$T_k^* = m_2\left((\boldsymbol{0}, \boldsymbol{I}_{p-k})(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)(\boldsymbol{0}, \boldsymbol{I}_{p-k})^\top\right).$$

Then $T_k \le T_k^*$, $k \in \{0, \dots, p-1\}$, and we have the following [30].

Theorem 4. *Under the previously stated assumptions and under H_{0q},*

(i) *for $k < q$, $T_k \xrightarrow{P} c_k$ for some $c_1, \dots, c_{q-1} > 0$,*
(ii) *for $k = q$, $n(p-k)T_k \xrightarrow{d} C_k$,*
(iii) *for $k > q$, $n(p-k)T_k \le n(p-k)T_k^* \xrightarrow{d} C_k$,*

where

$$C_k \sim 2\sigma_1 \chi^2_{(p-k-1)(p-k+2)/2} + (2\sigma_1 + \sigma_2(p-k))\,\chi^2_1$$

with independent chi squared variables $\chi^2_{(p-k-1)(p-k+2)/2}$ and χ^2_1 and σ_1, and σ_2 as in Theorem 4.

As in PCA, a consistent estimate $\hat{q}$ of the unknown dimension q can be based on sequential testing using the test statistics T_k and corresponding critical values $c_{k,n}$, $k \in \{0, \dots, p-1\}$, as suggested in the following. Other (top-down or divide and conquer) strategies again provide alternative consistent estimates.

Corollary 2. *For all $k \in \{0, \dots, p-1\}$, let $(c_{k,n})$ be a sequence of positive real numbers such that $c_{k,n} \to 0$ and $nc_{k,n} \to \infty$ as $n \to \infty$. Then*

$$\mathbb{P}(T_k \ge c_{k,n}) \;\to\; \begin{cases} 1, & \text{if } k < q, \\ 0, & \text{if } k \ge q, \end{cases}$$

and

$$\hat{q} = \min\{k \;:\; T_k < c_{k,n}\} \xrightarrow{P} q.$$

4.3. Bootstrap tests for dimension

Let q denote the true dimension and consider the test statistic $T_k = m_2\left(\widehat{\boldsymbol{U}}_k^\top(\widehat{\boldsymbol{R}} - (p+2)\boldsymbol{I}_p)\widehat{\boldsymbol{U}}_k\right)$ for H_{0k}, $k \in \{0, \dots, p-1\}$. In the following we also need

$$\widehat{\boldsymbol{P}}_k = \widehat{\boldsymbol{S}}_1^{1/2}\widehat{\boldsymbol{U}}_k\widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}_1^{-1/2}, \quad \widehat{\boldsymbol{Q}}_k = \boldsymbol{I}_p - \widehat{\boldsymbol{P}}_k,$$

which are the estimated projection matrices (with respect to Mahalanobis inner product) to the noise and signal subspaces, respectively.

To obtain the p-value for T_k, the bootstrap samples are generated, as in PCA, from a distribution for which the null hypothesis H_{0k} is true under the stated model (even if $k \ne q$) and which is as similar as possible to the empirical distribution of $\boldsymbol{X}$. We suggest again two procedures. The first one is for testing the hypothesis H_{0k} in the IC model and the second one in the wider NGCA model, see Nordhausen et al. [30]. The bootstrap p-values are obtained as in PCA with M bootstrap samples.

Bootstrap strategy FOBI-I (IC model):

1. Start with centered $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ and compute $\bar{\boldsymbol{x}}$ and $\widehat{\boldsymbol{W}} = (\widehat{\boldsymbol{W}}_1^\top, \widehat{\boldsymbol{W}}_2^\top)^\top$ where $\widehat{\boldsymbol{W}}_2 = \widehat{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{S}}_1^{-1/2}$.
2. Write $\widehat{\boldsymbol{Z}} = (\boldsymbol{X} - \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top)\widehat{\boldsymbol{W}}^\top$ and further $\widehat{\boldsymbol{Z}} = (\widehat{\boldsymbol{Z}}_1, \widehat{\boldsymbol{Z}}_2)$ where $\widehat{\boldsymbol{Z}}_2 \in \mathbb{R}^{n\times(p-k)}$.
3. Let $\boldsymbol{Z}_1^* \in \mathbb{R}^{n\times k}$ for a matrix of independent componentwise bootstrap samples of size n from $\widehat{\boldsymbol{Z}}_1$.
4. Let $\boldsymbol{Z}_2^* \in \mathbb{R}^{n\times(p-k)}$ be a random sample of size n from $N_{p-k}(\boldsymbol{0}, \boldsymbol{I}_{p-k})$.
5. Write $\boldsymbol{Z}^* = (\boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*)$.
6. Write $\boldsymbol{X}^* = \boldsymbol{Z}^*(\widehat{\boldsymbol{W}}^\top)^{-1} + \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top$.

Bootstrap strategy FOBI-II (NGCA model):

1. Start with $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, compute $\bar{\boldsymbol{x}}, \widehat{\boldsymbol{S}}_1, \widehat{\boldsymbol{S}}_2, \widehat{\boldsymbol{R}}, \widehat{\boldsymbol{U}}_k, \widehat{\boldsymbol{P}}_k$ and $\widehat{\boldsymbol{Q}}_k$.
2. Take a bootstrap sample $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_n)^\top$ of size n from $\boldsymbol{X}$.
3. For the noise space to be Gaussian, transform
$$\boldsymbol{x}_i^* = [\widehat{\boldsymbol{Q}}_k(\tilde{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}) + \widehat{\boldsymbol{S}}_1^{1/2}\widehat{\boldsymbol{U}}_k\boldsymbol{o}_i] + \bar{\boldsymbol{x}}, \quad i \in \{1, \dots, n\},$$
where $\boldsymbol{o}_1, \dots, \boldsymbol{o}_n$ are i.i.d. from $N_{p-k}(\boldsymbol{0}, \boldsymbol{I}_{p-k})$.
4. $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \dots, \boldsymbol{x}_n^*)^\top$.

In the case of the FOBI-I strategy, the bootstrap null distribution $F_{k,n}(\boldsymbol{x})$ is the average

$$\frac{1}{n^k}\sum_{i_1,\dots,i_k=1}^{n} \mathbb{E}_{\boldsymbol{o}_{i_1\cdots i_k}}\left[\mathbb{I}\left(\widehat{\boldsymbol{W}}^{-1}\begin{pmatrix} (\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_k})^\top(\boldsymbol{X} - \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top)\widehat{\boldsymbol{W}}_1^\top \\ \boldsymbol{o}_{i_1\cdots i_k}\end{pmatrix} + \bar{\boldsymbol{x}} \le x\right)\right],$$

Table 2
The asymptotic and bootstrapping based p-values for H_{01}-H_{04} for the image data when using FOBI. Either an IC model or a NGCA model was assumed. The null hypothesis $q = 1$ (H_{01}) is rejected by all four tests and the true value $q = 2$ (H_{02}) is the smallest one to be accepted.

	ICA		NGCA	
	Asymp	Boot	Asymp	Boot
H_{01}	0.000	0.002	0.000	0.002
H_{02}	0.211	0.082	0.206	0.116
H_{03}	0.878	0.940	0.873	0.880
H_{04}	0.810	0.778	0.806	0.729

where the $\boldsymbol{o}_{i_1\cdots i_k}^\top$s are from $N_{p-k}(\mathbf{0}, \boldsymbol{I}_{p-k})$ and the $\boldsymbol{e}_i^\top$s (with the ith element one and other elements zero) are in $\mathbb{R}^n$, and in the FOBI-II strategy, the bootstrap samples for H_{0k} are generated from the distribution $F_{k,n}(\boldsymbol{x})$ that is the average

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{o_i}\left[\mathbb{I}\left([\widehat{\boldsymbol{Q}}_k(\boldsymbol{x}_i-\bar{\boldsymbol{x}})+\widehat{\boldsymbol{S}}_1^{1/2}\widehat{\boldsymbol{U}}_k\boldsymbol{o}_i]+\bar{\boldsymbol{x}}\le \boldsymbol{x}\right)\right],$$

where $\boldsymbol{o}_1,\ldots,\boldsymbol{o}_n \sim N_{p-k}(\mathbf{0}, \boldsymbol{I}_{p-k})$.

As in the PCA bootstrap asymptotics, let $\boldsymbol{X}_N^*$ be a random sample of size N from $F_{n,k}$. As these observations come from the ICA and NGCA models, respectively, with true H_{0k} and known (data based) parameters $\sigma_1 = \hat{\sigma}_{1a}$ or $\sigma_1 = \hat{\sigma}_{1b}$ and $\sigma_2 = 4$, the limiting (conditional and unconditional) distribution of $NT_k(\boldsymbol{X}_N^*)$ is as given in Theorem 3. For large n, the limiting distribution then provides the approximation for $nT_k(\boldsymbol{X}^*)$ as well.

The bootstrapping testing strategy was explored for any pair of two scatter matrices in [35] in is quite similar than the approach described above.

4.4. An example

ICA is often illustrated using mixed images. Following this tradition, we mix 6 gray scale images: Two of the images are the pictures of a cat and a forest road, available in the R package ICS [29], and the remaining four images are just Gaussian noise. The images have 130×130 pixels and the six original images can be presented as a matrix $\boldsymbol{Z} \in \mathbb{R}^{n\times p}$ with $n = 16900$ pixels and $p = 6$ columns identifying the 6 images. The observed mixed images are then $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{A}^\top + \mathbf{1}_n\boldsymbol{b}^\top$ and the idea is to recover the two (signal) images. Note that the rows of $\boldsymbol{X}$ are not independent in this example but FOBI uses the marginal distribution of the column elements rather than their joint distribution.

The first three columns of the $\boldsymbol{Z}$ and $\widehat{\boldsymbol{Z}} = \boldsymbol{X}\widehat{\boldsymbol{W}}^\top$ are given on the first and second row of Fig. 2, respectively. Note that the result on the second row would be the same for any choices of $\boldsymbol{A}$ and $\boldsymbol{b}$. The ordered eigenvalues (with respect to the squared deviation from $p + 2 = 8$) of $\widehat{\boldsymbol{R}}$ are $9.00, 8.27, 7.92, 8.04, 7.97$ and 8.00. The p-values for H_{01}-H_{04} both all the tests are given Table 2. Note that the bootstrap tests here use $m = 500$ bootstrap samples. In this examples all four tests nicely agree and the false hypothesis H_{01} is rejected and the true hypothesis H_{02} is the first to be accepted at level $\alpha = 0.05$.

5. Testing for subspace dimension in SIR

5.1. The model, null hypothesis and test statistic

In this section we assume that

$$(\boldsymbol{y}, \boldsymbol{X}) = \left(\begin{pmatrix} y_1 \\ \boldsymbol{x}_1 \end{pmatrix}, \ldots, \begin{pmatrix} y_n \\ \boldsymbol{x}_n \end{pmatrix}\right)^\top \in \mathbb{R}^{n\times(p+1)}$$

is a random sample from a distribution of $(y, \boldsymbol{x}^\top)^\top$ where

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b},$$

$\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is non-singular, $\boldsymbol{b} \in \mathbb{R}^p$ and $\boldsymbol{z} = (\boldsymbol{z}_i^\top, \boldsymbol{z}_2^\top)^\top$ is a random p-vector with $\mathbb{E}(\boldsymbol{z}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{z}) = \boldsymbol{I}_p$ and $(y, \boldsymbol{z}_1^\top)^\top \perp\!\!\!\perp \boldsymbol{z}_2$. If $\boldsymbol{z}_1 \in \mathbb{R}^q$ and $\boldsymbol{z}_2 \in \mathbb{R}^{p-q}$, with q being the smallest value for which this condition holds, then they correspond respectively to the signal and noise parts of $\boldsymbol{z}$. The partition $\boldsymbol{z} = (\boldsymbol{z}_i^\top, \boldsymbol{z}_2^\top)^\top$ is then unique up to transformations $\boldsymbol{z}_1 \to \boldsymbol{O}_1\boldsymbol{z}_1$ and $\boldsymbol{z}_2 \to \boldsymbol{O}_2\boldsymbol{z}_2$ with $\boldsymbol{O}_1 \in \mathcal{O}^{q\times p}$ and $\boldsymbol{O}_2 \in \mathcal{O}^{(p-q)\times(p-q)}$. The aim is again to test and estimate the unknown dimension q and then find the projections to the well defined signal and noise subspaces of $\boldsymbol{x}$.

Remark 1. Note that our assumption $(y, \boldsymbol{z}_1^\top)^\top \perp\!\!\!\perp \boldsymbol{z}_2$ is stronger than the regular assumptions in sliced inverse regression and related methods: In classical SIR and SAVE approaches the dependence conditions are for example (i) $y \perp\!\!\!\perp \boldsymbol{z}_2|\boldsymbol{z}_1$

Fig. 2. The first row shows the original signal images plus one exemplary noise component from $\boldsymbol{Z}$. The second row shows the first three estimated components $\hat{\boldsymbol{Z}}$ when using FOBI. All components not shown look like the noise components (third column).

and $E(\boldsymbol{z}_2|\boldsymbol{z}_1) = \boldsymbol{0}$ *a.s.* (linearity condition) for SIR and (ii) $y \perp\!\!\!\perp \boldsymbol{z}_2|\boldsymbol{z}_1$, $E(\boldsymbol{z}_2|\boldsymbol{z}_1) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{z}_2|\boldsymbol{z}_1) = \boldsymbol{I}_{p-q}$ *a.s.* for SAVE. Alternative or additional assumptions needed for easy and tractable asymptotics have been given in the literature such as the assumption that $\boldsymbol{z}$ is multivariate normal [23] or that the conditional covariance $\text{Cov}(\boldsymbol{z}|y)$ is constant [4]. See Section 5.2 for more discussion. Under our strong assumption, bootstrap samples from a true null distributions are easily generated as shown in Section 5.3.

In the sliced inverse regression (SIR) one finds a transformation matrix $\boldsymbol{W} \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{p\times p}$ such that

$$\boldsymbol{W}\boldsymbol{S}_1\boldsymbol{W}^\top = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{W}\boldsymbol{S}_2\boldsymbol{W}^\top = \boldsymbol{D}$$

with $\boldsymbol{S}_1 := \mathbb{E}\left[(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\top\right]$ and $\boldsymbol{S}_2 := \mathbb{E}\left[\mathbb{E}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})|y)\mathbb{E}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})|y)^\top\right]$. Under our assumptions, the diagonal elements in $\boldsymbol{D}$ are

$$d_1 \geq \cdots \geq d_q \geq d_{q+1} = \cdots = d_p = 0.$$

Again, as in ICA, $\boldsymbol{W} = \boldsymbol{U}^\top\boldsymbol{S}_1^{-1/2}$ with some orthogonal $\boldsymbol{U} \in \mathbb{R}^{p\times p}$ and, if $\boldsymbol{R} := \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ then $\boldsymbol{U}$ is the matrix of eigenvectors of $\boldsymbol{R}$.

In practice, the random variable y is replaced by its discrete approximation as follows. Let $\mathbb{S}_1, \ldots, \mathbb{S}_H$ be H disjoint intervals (slices) such that $\mathbb{R} = \mathbb{S}_1 + \cdots + \mathbb{S}_H$ and let $y^d := \sum_{h=1}^{H} y_h \mathbb{I}(y \in \mathbb{S}_h)$ for some choices $y_h \in \mathbb{S}_h$, $h \in \{1, \ldots, H\}$, independent of $\boldsymbol{z}$. ($\mathbb{I}(y \in \mathbb{S}_h) = 1$ if $y \in \mathbb{S}_h$ and zero otherwise.) The random variable y^d can then be seen as a discrete approximation of a continuous random variable y. Naturally also $(y^d, \boldsymbol{z}_1^\top)^\top \perp\!\!\!\perp \boldsymbol{z}_2$. The sliced inverse regression (SIR) then just refers to the use of the inverse regression $\mathbb{E}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})|y^d)$ and the corresponding supervised scatter matrix

$$\boldsymbol{S}_2 = \mathbb{E}\left[\mathbb{E}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})|y^d)\mathbb{E}(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})|y^d)^\top\right]$$

in the analysis of the data. With this choice of $\boldsymbol{S}_2$, we still have $d_1 \geq \cdots \geq d_q \geq d_{q+1} = \cdots = d_p = 0$. Next write $\boldsymbol{\mu} := \mathbb{E}(\boldsymbol{x})$ and $\boldsymbol{\Sigma} := \text{Cov}(\boldsymbol{x})$, and $\boldsymbol{\mu}_h := \mathbb{E}(\boldsymbol{x}|y \in \mathbb{S}_h)$, $\boldsymbol{\Sigma}_h := \text{Cov}(\boldsymbol{x}|y \in \mathbb{S}_h)$ and $p_h = \mathbb{P}(y \in \mathbb{S}_h)$, $h \in \{1, \ldots, H\}$. Then

$$\boldsymbol{S}_1 = \boldsymbol{\Sigma}, \quad \boldsymbol{S}_2 = \sum_{h=1}^{H} p_h(\boldsymbol{\mu}_h - \boldsymbol{\mu})(\boldsymbol{\mu}_h - \boldsymbol{\mu})^\top.$$

Consider next the corresponding sample statistics. For the estimates of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, write

$$\widehat{\boldsymbol{S}}_1 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top, \quad \widehat{\boldsymbol{S}}_2 = \frac{1}{n}\sum_{h=1}^{H} n_h(\bar{\boldsymbol{x}}_h - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_h - \bar{\boldsymbol{x}})^\top,$$

where $\bar{\boldsymbol{x}}_h = \frac{1}{n_h}\sum_{i=1}^n \boldsymbol{x}_i \mathbb{I}(y_i \in \mathbb{S}_h)$ and $n_h = \sum_{i=1}^n \mathbb{I}(y_i \in \mathbb{S}_h)$, $h \in \{1, \ldots, H\}$. Note that $np \cdot m_1(\widehat{\boldsymbol{S}}_1^{-1}\widehat{\boldsymbol{S}}_2)$ is the well-known Pillai's trace statistic for testing $H_0: \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_H$ under the assumption that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_H$ with the limiting null distribution $\chi^2_{(H-1)p}$.

Furthermore, let $\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{S}}_1^{-1/2}\widehat{\boldsymbol{S}}_2\widehat{\boldsymbol{S}}_1^{-1/2}$. We wish to test the null hypothesis

$$H_{0k}: \quad d_1 \geq \cdots \geq d_k > d_{k+1} = \cdots = d_p = 0$$

stating that the dimension of the signal space is exactly k. To test the null hypothesis, we use a natural test statistic, that is, the average of the $p-k$ smallest eigenvalues of $\widehat{\boldsymbol{R}}$, that is,

$$T_k := m_1(\widehat{\boldsymbol{U}}_k^\top \widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_k),$$

where the columns of $\widehat{\boldsymbol{U}}_k \in \mathcal{O}^{p\times(p-k)}$ are the eigenvectors corresponding the smallest $p-k$ eigenvalues of $\widehat{\boldsymbol{R}}$.

5.2. Asymptotic tests for dimension

As the eigenvalues of $\widehat{\boldsymbol{R}}$ are invariant under affine transformations, we can assume without loss of generality that $(\boldsymbol{y}, \boldsymbol{X})$ is a random sample from a SIR model with $\boldsymbol{A} = \boldsymbol{I}_p$ and $\boldsymbol{b} = \boldsymbol{0}$. This implies $\boldsymbol{S}_1 = \boldsymbol{I}_p$ and $\boldsymbol{\mu} = \boldsymbol{0}$. We assume that the number of slices $H > q+1$, the slices $\mathbb{S}_1, \ldots, \mathbb{S}_H$ do not change with n, and the related $\boldsymbol{S}_2 = \boldsymbol{R} = \boldsymbol{D} = \mathrm{diag}(\boldsymbol{D}_1, \boldsymbol{0})$ with a full-rank $\boldsymbol{D}_1 \in \mathbb{R}^{q\times q}$. The assumption thus states that, with selected H slices and by using SIR, one can find the full q-dimensional signal space.

Let $f_h = n_h/n$, $h \in \{1, \ldots, H\}$, and write

$$\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{S}}_1^{-1/2}\left(\sqrt{f_1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}), \quad \ldots, \quad \sqrt{f_H}(\bar{\boldsymbol{x}}_H - \bar{\boldsymbol{x}})\right).$$

Then $\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^\top$ and, with $\boldsymbol{\pi} = (\sqrt{p}_1, \ldots, \sqrt{p}_H)^\top$,

$$\widehat{\boldsymbol{B}} \to \boldsymbol{B} := (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_H)\,\mathrm{diag}(\boldsymbol{\pi}) = \begin{pmatrix} \boldsymbol{D}_1^{1/2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{Q}$$

for some $\boldsymbol{Q} \in \mathcal{O}^{H\times H}$, where $\boldsymbol{Q} = (\boldsymbol{Q}_1^\top, \boldsymbol{Q}_2^\top)^\top$ and $\boldsymbol{Q}_1 \in \mathcal{O}^{q\times H}$ satisfies $\boldsymbol{Q}_1\boldsymbol{\pi} = \boldsymbol{0}$. With the correct $\boldsymbol{Q}$ and correct dimension q, we have the partitions

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} \widehat{\boldsymbol{B}}_1 \\ \widehat{\boldsymbol{B}}_2 \end{pmatrix}, \quad \widehat{\boldsymbol{B}}\boldsymbol{Q}^\top = \begin{pmatrix} \widehat{\boldsymbol{B}}_1\boldsymbol{Q}_1^\top & \widehat{\boldsymbol{B}}_1\boldsymbol{Q}_2^\top \\ \widehat{\boldsymbol{B}}_2\boldsymbol{Q}_1^\top & \widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top \end{pmatrix}.$$

An asymptotic approximation to the distribution of $T_q = m_1(\widehat{\boldsymbol{U}}_q^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_q)$ can now be stated as follows:

Lemma 3. *Under the previously stated assumptions and under H_{0q},*

$$n \cdot T_q = n \cdot m_1(\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top) + O_P(n^{-1/4}).$$

Note that, in this setting, with $\boldsymbol{U}_q^\top = (\boldsymbol{0}, \boldsymbol{I}_{p-q})$,

$$\boldsymbol{U}_q^\top\widehat{\boldsymbol{R}}\boldsymbol{U}_q = \widehat{\boldsymbol{B}}_2\widehat{\boldsymbol{B}}_2^\top = \widehat{\boldsymbol{B}}_2\boldsymbol{Q}_1^\top\boldsymbol{Q}_1\widehat{\boldsymbol{B}}_2^\top + \widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top.$$

Consequently, unlike in Lemmas 1 and 2 for PCA and ICA asymptotics, the asymptotic approximation given in Lemma 3 is not obtained by simply replacing $\widehat{\boldsymbol{U}}_q$ by $\boldsymbol{U}_q$ within the definition of T_q. The limiting distribution of $n(p-q)T_q$ is then given in the following theorem.

Theorem 5. *Under our assumptions and under H_{0q}, $n(p-q)T_q \xrightarrow{d} \chi^2_{(p-q)(H-q-1)}$.*

The same limiting distribution is given in Theorem 5.1 in Li [23] and in Corollary 1 in Bura and Cook [4] under the conditional independence relation $y \perp\!\!\!\perp \boldsymbol{z}_2|\boldsymbol{z}_1$ and under the linearity condition $E(\boldsymbol{z}_2|\boldsymbol{z}_1) = \boldsymbol{0}$, *a.s.*. In the former, the theorem is stated under an additional assumption that the distribution of $\boldsymbol{z}$ is multivariate normal, but within the proof it is noted that it in fact holds if $\mathrm{Cov}(\boldsymbol{z}_2|y)$ does not depend on y. In the latter, the above theorem is stated under the additional assumption that $\mathrm{Cov}(\boldsymbol{z}|y)$ does not depend on y, but from their proof it can be noted that they only need this to hold for $\mathrm{Cov}(\boldsymbol{z}_2|y)$. In our setting, this condition obviously holds since $\boldsymbol{z}_2 \perp\!\!\!\perp y$. For completeness, a proof to Theorem 5 is given in the Appendix. Note that for $q \geq H-1$, $T_q = 0$.

To estimate q, we consider the limiting behavior of the test statistics $n(p-k)T_k$ for H_{0k}, $k \in \{0, \ldots, H-1\}$, when in fact H_{0q} is true. We write

$$T_k^* = m_1((\boldsymbol{I}_{p-k}, \boldsymbol{0})\widehat{\boldsymbol{U}}_q^\top\widehat{\boldsymbol{R}}\widehat{\boldsymbol{U}}_q(\boldsymbol{I}_{p-k}, \boldsymbol{0})^\top), \quad k \in \{q+1, \ldots, H-1\}$$

and then have the following theorem.

Table 3
The p-values for H_{00}-H_{03} with two testing strategies for the Australian athletes data. The null hypotheses $q = 0$ (H_{00}) and $q = 1$ (H_{01}) are both rejected and both tests suggest an estimate $\hat{q} = 2$.

	H_{00}	H_{01}	H_{02}	H_{03}
SIR-I	0.002	0.002	0.090	0.349
Asymp	0.000	0.001	0.121	0.458

Theorem 6. *Under the previously stated assumptions and under H_{0q},*

(i) *for $k < q$, $T_k \xrightarrow{P} c_k$ for some $c_1 > 0, \ldots, c_{q-1} > 0$,*
(ii) *for $k = q$, $n(p-k)T_k \xrightarrow{d} \chi^2_{(p-q)(H-q-1)}$,*
(iii) *for $k > q$, $\mathbb{P}(T_k \le T_k^*) \to 1$ and $n(p-k)T_k^* \xrightarrow{d} \chi^2_{(p-k)(H-q-1)}$*

As in PCA and ICA, a consistent estimate $\hat{q}$ of the unknown dimension q can found with the bottom-up sequential testing strategy as follows. Again alternative testing strategies may be used to find computationally faster and consistent estimates.

Corollary 3. *For all $k \in \{0, \ldots, H-1\}$, let $(c_{k,n})$ be a sequence of positive real numbers such that $c_{k,n} \to 0$ and $nc_{k,n} \to \infty$ as $n \to \infty$. Then $\hat{q} = \min\{k \,:\, T_k < c_{k,n}\} \xrightarrow{P} q$.*

5.3. A bootstrap test for dimension

We consider the hypotheses H_{0k} saying that the rank of $\boldsymbol{D}$ is k, $k \in \{1, \ldots, H-1\}$. Bootstrap samples are then to be generated from a null distribution for which $(y, \boldsymbol{z}_1^\top)^\top \perp\!\!\!\perp \boldsymbol{z}_2$ and $\boldsymbol{z}_1 \in R^k$ even if the true dimension $p \neq k$. Bootstrap sampling from a null distribution obeying the weaker assumptions such as $y \perp\!\!\!\perp \boldsymbol{z}_2 | \boldsymbol{z}_1$ and $\mathbb{E}(\boldsymbol{z}_2|\boldsymbol{z}_1) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{z}_2|y) = \boldsymbol{I}_{p-k}$ seems much more difficult to carry out and not developed here. Sampling under our strong assumption is described in the following.

Bootstrap strategy SIR: Generate from the SIR model.

1. Starting from $\boldsymbol{X}$, find $\bar{\boldsymbol{x}}$ and $\widehat{\boldsymbol{W}} = (\widehat{\boldsymbol{W}}_1^\top, \widehat{\boldsymbol{W}}_2^\top)^\top$ where $\widehat{\boldsymbol{W}}_1 \in \mathbb{R}^{k\times p}$ and write $\widehat{\boldsymbol{Z}}_i = (\boldsymbol{X} - \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top)\widehat{\boldsymbol{W}}_i^\top$, $i \in \{1, 2\}$.
2. Let $(y^*, \boldsymbol{Z}_1^*)$ be a bootstrap sample of size n from $(y, \widehat{\boldsymbol{Z}}_1)$.
3. Let $\boldsymbol{Z}_2^*$ be a bootstrap sample of size n from $\widehat{\boldsymbol{Z}}_2$. (Bootstrap samples in 2 and 3 are independent)
4. Write $\boldsymbol{Z}^* = (\boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*)$.
5. Write $(\boldsymbol{y}^*, \boldsymbol{X}^*) = \left(\boldsymbol{y}^*, \widehat{\boldsymbol{Z}}^*(\widehat{\boldsymbol{W}}^\top)^{-1} + \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top\right)$.

In other terms, the bootstrap null distribution $F_{k,n}$ at $(y, \boldsymbol{x}^\top)^\top$ is now obtained as the average

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{I}\left(\left(\begin{array}{c} \boldsymbol{y}^\top\boldsymbol{e}_i \\ \widehat{\boldsymbol{W}}^{-1}\begin{pmatrix} \widehat{\boldsymbol{W}}_1(\boldsymbol{X} - \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top)^\top\boldsymbol{e}_i \\ \widehat{\boldsymbol{W}}_2(\boldsymbol{X} - \boldsymbol{1}_n\bar{\boldsymbol{x}}^\top)^\top\boldsymbol{e}_j \end{pmatrix}\end{array}\right) + \begin{pmatrix} 0 \\ \bar{\boldsymbol{x}} \end{pmatrix} \le \begin{pmatrix} y \\ \boldsymbol{x} \end{pmatrix}\right),$$

where the $\boldsymbol{e}^\top$s are in $\mathbb{R}^n$. As for PCA and ICA bootstrap strategies, let $\boldsymbol{X}_N^*$ be a sample of size N from $F_{k,n}$ for which the null hypothesis H_{0k} and our model assumptions naturally hold true. Then $NT_k(\boldsymbol{X}_N^*) \xrightarrow{d} \chi^2_{(p-k)(H-k-1)}$ and therefore, for large n, also the distribution of $nT_k(\boldsymbol{X}_n^*)$ can be approximated by the same distribution. The estimated bootstrap p-value is obtained as in the previous cases.

5.4. An example

For the illustration we revisit the Australian Athletes data available in the R package `dr` [44]. The response variable y is the lean body mass the predictors in $\boldsymbol{x}$ are given by the logarithms of height (Ht), weight (Wt), red cell count (RCC), white cell count (WCC), Hematocrit (Hc), Hemoglobin (Hg), plasma ferritin concentration (Ferr) and sum of skin folds (SSF). The same data was analyzed e.g., by Cook [8], who developed tests of the hypothesis of no effect for a selected subset of predictors. The data for all 202 athletes is shown in Fig. 3 and the SIR eigenvalues are, rounding to two decimal places, 0.95, 0.21, 0.11, 0.07, 0.04, 0.02, 0.01 and 0.00.

The observed p values for successive testing of hypotheses H_{00} to H_{04} are reported in Table 3. The number of bootstrap samples was $M = 500$ and the bootstrap test as well as the asymptotic test suggest that the signal space has dimension two. Note that the p-values of the asymptotic tests differ slightly from those in Cook [8], perhaps due to different number of slices and different numbers of observations in slices.

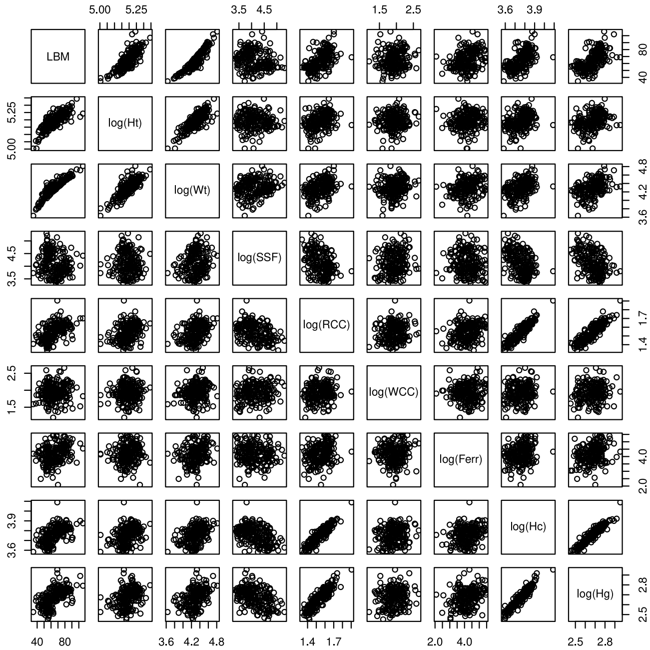

Fig. 3. Pairwise scatter plots for 9 variables in the Australian athletes data. The first variable LBM is the response variable to be explained by the 8 remaining variables.

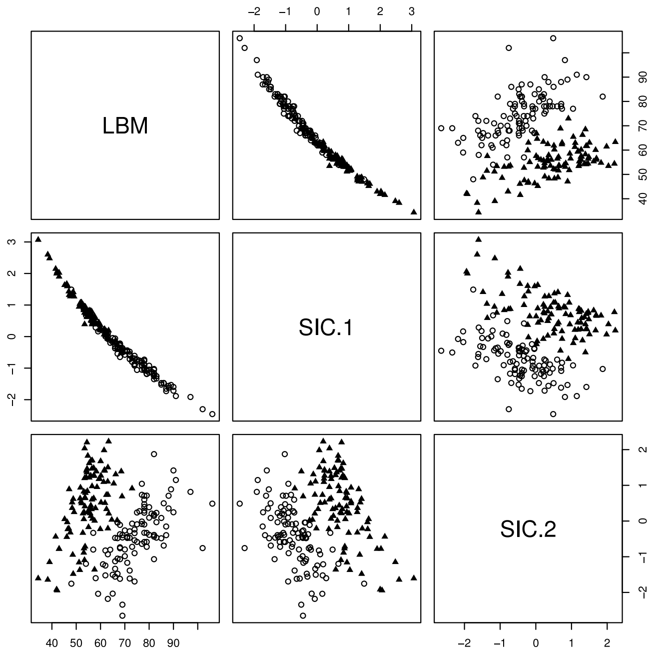

Fig. 4. The pairwise scatter plots for the response LBM and the two first SIR components SIC.1 and SIC.2. In the plots different symbols are used for men and women. The gender was not used in the analysis.

The two signal components are plotted against the response in Fig. 4 where the plotting symbols differ for female and male athletes. The figure nicely shows that both components contain information about the response. The gender of the athletes was not used in the analysis. However, the first two SIR components seem also to separate the female and male athletes.

Table 4
The similarities and differences between PCA, ICA and SIR and their use of scatter matrices in the data analysis.

	PCA	FOBI	SIR
Supervised	No	No	Yes
Model	Elliptical model	Independent component model	Regression model
Data	Numeric vector $\boldsymbol{x}$	Numeric vector $\boldsymbol{x}$	Response y and numeric vector $\boldsymbol{x}$
Scatter matrix $\boldsymbol{S}_1$	Identity matrix $\boldsymbol{I}_p$	Covariance matrix	Covariance matrix
Scatter matrix $\boldsymbol{S}_2$	Any scatter matrix $\boldsymbol{S}$	Scatter matrix based on fourth moments	Supervised SIR scatter matrix
Signal	Non-spherical principal components	Non-Gaussian independent components	Components sufficient to explain y
Noise	Spherical principal components	Gaussian independent components	Components conditionally independent of y
Hypothesis	Multiplicity of smallest eigenvalue is $p-k$	Multiplicity of eigenvalue $p+2$ is $p-k$	Multiplicity of zero eigenvalue is $p-k$
Test statistics	Variance of the $p-k$ smallest eigenvalues	Smallest sum of squared distances between $p-k$ eigenvalues and $p+2$	Mean of the $p-k$ smallest eigenvalues
Limiting distributions	chi-square	Weighted sum of independent chi-square variables	chi-square
Bootstrapping	Two different strategies	Two different strategies	One strategy

6. Final remarks

In this paper, we considered three dimension reduction methods based on the use of a pair of sample matrices, principal component analysis, fourth order blind identification and sliced inverse regression, and showed how first two moments of the eigenvalues of one matrix with respect to another can be used to test for signal (and noise) dimension. The concluding joint framework for the three methods is summarized in Table 4. In all three cases, the asymptotic null distributions of the test statistics were given and bootstrap strategies were provided for the testing problems. The asymptotic and bootstrap tests were compared in real data examples. These three methods serve here as examples and it is obvious that our approach can be extended to other pairs of scatter matrices tailored for the multivariate semiparametric goodness-of-fit problems at hand, see e.g., Nordhausen et al. [28].

The R code for all computations in the paper is available upon request from Klaus Nordhausen and almost all methods are implemented in the R package `ICtest` [31]. Simulation results are given in an extend version of this paper on Arxiv. However larger simulation studies as well as theoretical studies in various contexts are still necessary in the future to compare the estimates here to other consistent estimates suggested in the literature [26,45,47] and to compare different sequential testing strategies (bottom-up, top-down, divide and conquer).

Technical details

Proofs for Section 3

Proof of Lemma 1. Let $\widehat{\boldsymbol{d}} = (\hat{d}_{q+1}, \ldots, \hat{d}_p)$ denote the $r = p-q$ smallest ordered eigenvalues of $\widehat{\boldsymbol{S}}$ and let $\widehat{\boldsymbol{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_r)$ denote the ordered eigenvalues of $\widehat{\boldsymbol{S}}_{22}$. Lemma 3.1 in Eaton and Tyler (1991) then states that $\widehat{\boldsymbol{d}} - \widehat{\boldsymbol{\delta}} = O_P\left(n^{-1}\right)$ and, applying Theorem 3.2 in Eaton and Tyler (1991), $\widehat{\boldsymbol{\delta}} - d\boldsymbol{1}_r = O_P\left(n^{-1/2}\right)$ then implies that $\widehat{\boldsymbol{d}} - d\boldsymbol{1}_r = O_P\left(n^{-1/2}\right)$. Setting $\boldsymbol{P}_r = \boldsymbol{I}_r - r^{-1}\boldsymbol{1}_r\boldsymbol{1}_r^\top$, we then have $r \cdot T_q = \widehat{\boldsymbol{d}}^\top \boldsymbol{P}_r \widehat{\boldsymbol{d}} = (\widehat{\boldsymbol{d}} - d\boldsymbol{1}_r)^\top \boldsymbol{P}_r (\widehat{\boldsymbol{d}} - d\boldsymbol{1}_r)$ and $r \cdot s^2(\widehat{\boldsymbol{S}}_{22}) = \widehat{\boldsymbol{\delta}}^\top \boldsymbol{P}_r \widehat{\boldsymbol{\delta}} = (\widehat{\boldsymbol{\delta}} - d\boldsymbol{1}_r)^\top \boldsymbol{P}_r (\widehat{\boldsymbol{\delta}} - d\boldsymbol{1}_r)$. Hence,

$$r\left(T_q - s^2(\widehat{\boldsymbol{S}}_{22})\right) = 2(\widehat{\boldsymbol{\delta}} - d\boldsymbol{1}_r)^\top \boldsymbol{P}_r (\widehat{\boldsymbol{d}} - \widehat{\boldsymbol{\delta}}) + (\widehat{\boldsymbol{d}} - \widehat{\boldsymbol{\delta}})^\top \boldsymbol{P}_r (\widehat{\boldsymbol{d}} - \widehat{\boldsymbol{\delta}}),$$

which is $O_P\left(n^{-3/2}\right) + O_P\left(n^{-2}\right) = O_P\left(n^{-3/2}\right)$.

Proof of Theorem 1. By Lemma 1 it is sufficient to consider the limiting distribution of $n \cdot s^2(\widehat{\boldsymbol{S}}_{22})$. Let again $r = p-q$ and $\boldsymbol{Z}_{22} = \sqrt{n}(\widehat{\boldsymbol{S}}_{22} - d\boldsymbol{I}_r)/d$. Then

$$nr \cdot s^2(\widehat{\boldsymbol{S}}_{22})/d^2 = n \cdot \mathrm{vec}(\widehat{\boldsymbol{S}}_{22})^\top \boldsymbol{\Gamma} \mathrm{vec}(\widehat{\boldsymbol{S}}_{22})/d^2 = \mathrm{vec}(\boldsymbol{Z}_{22})^\top \boldsymbol{\Gamma} \mathrm{vec}(\boldsymbol{Z}_{22}),$$

where $\boldsymbol{\Gamma} = \boldsymbol{I}_{r^2} - r^{-1}\mathrm{vec}(\boldsymbol{I}_r)\mathrm{vec}(\boldsymbol{I}_r)^\top$ is idempotent. The second identity follows since $\boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{I}_r) = \boldsymbol{0}$. Under H_{0q}, $\boldsymbol{Z}_{22} \xrightarrow{d} \mathrm{Z}$ with $\mathrm{vec}(\mathrm{Z}) \sim N_{r^2}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sigma_1(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r,r}) + \sigma_2\mathrm{vec}(\boldsymbol{I}_r)\mathrm{vec}(\boldsymbol{I}_r)^\top$. This implies

$$nr \cdot s^2(\widehat{\boldsymbol{S}}_{22})/d^2 \xrightarrow{d} 2\sigma_1 \mathrm{z}^\top \mathrm{z}, \text{ with } \mathrm{z} = \boldsymbol{\Gamma}\mathrm{vec}(\mathrm{Z})/\sqrt{2\sigma_1} \sim N_{r^2}(\boldsymbol{0}, \boldsymbol{\Sigma}_o),$$

where

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Gamma}\frac{1}{2}(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r,r})\boldsymbol{\Gamma} = \frac{1}{2}\left(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r,r} - \frac{2}{r}\mathrm{vec}(\boldsymbol{I}_r)\mathrm{vec}(\boldsymbol{I}_r)^\top\right).$$

Now $\boldsymbol{\Sigma}_0$ is symmetric and idempotent with $\mathrm{rank}(\boldsymbol{\Sigma}_0) = (r^2 + r - 2)/2 = (r + 2)(r - 1)/2$, and so $\mathrm{z}^\top\mathrm{z} \sim \chi^2_{(r+2)(r-1)/2}$ and the first part of the theorem follows. The second part follows as V_q is the minimum of the variance over all $(p-q)$-subsets of the ordered eigenvalues of $\widehat{\boldsymbol{S}}$. The variance of the $p - q$ smallest eigenvalues, that is, T_q converges in probability to 0, and the variance for any other $\binom{p}{q} - 1$ choices of subsets converges in probability to a positive constant.

Proof of Theorem 2. (i) T_k converges in probability to the variance of $p - k$ smallest eigenvalues which is positive for $k < q$. (ii) is given in the previous theorem. (iii) follows as, for $k \in \{q, \ldots, p - 1\}$,

$$T_k = \frac{1}{2(p-k)^2}\sum_{i=k+1}^{p}\sum_{j=k+1}^{p}(\hat{d}_i - \hat{d}_j)^2 \le \left(\frac{p-q}{p-k}\right)^2 \frac{1}{2(p-q)^2}\sum_{i=q+1}^{p}\sum_{j=q+1}^{p}(\hat{d}_i - \hat{d}_j)^2 = \left(\frac{p-q}{p-k}\right)^2 T_q.$$

Proofs for Section 4

Proof of Lemma 2. This proof is similar to the proof of Lemma 1. Again set $r = p - q$. Rather than using the ordering of the roots given in Section 4, let $\lambda_1, \ldots, \lambda_p$ denote the ordered eigenvalues of $\boldsymbol{R}$, and so for some $0 \le m \le q$, $\lambda_m > p + 2$, $\lambda_{m+1} = \cdots = \lambda_{m+r} = p + 2$ and $\lambda_{m+r+1} < p + 2$. Also, let $\widehat{\boldsymbol{\lambda}} = (\hat{\lambda}_{m+1}, \ldots, \hat{\lambda}_{m+r})^\top$ denote the $(m + 1)th$ to $(m + r)th$ ordered eigenvalues of $\widehat{\boldsymbol{R}}$ and let $\widehat{\boldsymbol{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_r)^\top$ denote the ordered eigenvalues of $\widehat{\boldsymbol{R}}_{22}$. Again using [13], applying its Lemma 3.1 twice gives $\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\delta}} = O_P\left(n^{-1}\right)$ and applying its Theorem 3.2 gives $\widehat{\boldsymbol{\lambda}} - (p + 2)\boldsymbol{1}_p = O_P\left(n^{-1/2}\right)$. Now, $r \cdot T_q = (\widehat{\boldsymbol{\lambda}} - (p + 2)\boldsymbol{1}_r)^\top(\widehat{\boldsymbol{\lambda}} - (p + 2)\boldsymbol{1}_r)$ and $r \cdot s^2(\widehat{\boldsymbol{S}}_{22}) = (\widehat{\boldsymbol{\delta}} - (p + 2)\boldsymbol{1}_r)^\top(\widehat{\boldsymbol{\delta}} - (p + 2)\boldsymbol{1}_r)$. Hence,

$$r\left(T_q - m_2(\widehat{\boldsymbol{R}}_{22})\right) = 2(\widehat{\boldsymbol{\delta}} - (p + 2)\boldsymbol{1}_r)^\top(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\delta}}) + (\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\delta}})^\top(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\delta}}),$$

which is $O_P\left(n^{-3/2}\right) + O_P\left(n^{-2}\right) = O_P\left(n^{-3/2}\right)$.

Proof of Theorem 3. By Lemma 2 it is sufficient to consider the joint limiting distribution of $n(s^2(\widehat{\boldsymbol{R}}_{22}), m_1^2(\widehat{\boldsymbol{R}}_{22}))$. Set again $r = p - q$. The arguments for obtaining the limiting distribution of $n \cdot s^2(\widehat{\boldsymbol{R}}_{22})$ are analogous to those used in the proof of Theorem 1, and we use the same notation but now with $\boldsymbol{Z}_{22} = \sqrt{n}(\widehat{\boldsymbol{R}}_{22} - (p + 2)\boldsymbol{I}_r)/(p + 2) \to \mathrm{Z}$ with the property that $\boldsymbol{U}^\top\mathrm{Z}\boldsymbol{U} \sim \mathrm{Z}$ for all $\boldsymbol{U} \in \mathcal{O}^{r\times r}$. Then again $\mathrm{vec}(\mathrm{Z}) \sim N_{r^2}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sigma_1(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r,r}) + \sigma_2\mathrm{vec}(\boldsymbol{I}_r)\mathrm{vec}(\boldsymbol{I}_r)^\top$ with two population constants σ_1 and σ_2. Using arguments analogous to those in the proof of Theorem 1, we again obtain under the null hypothesis that $nr \cdot s^2(\widehat{\boldsymbol{R}}_{22})/(p + 2)^2 \to \chi^2_{(r+2)(r-1)/2}$. Next, $r\sqrt{n} \cdot m_1(\widehat{\boldsymbol{R}}_{22}) = \mathrm{vec}(\boldsymbol{I}_r)^\top\mathrm{vec}(\boldsymbol{Z}_{22}) \xrightarrow{d} \mathrm{vec}(\boldsymbol{I}_r)^\top\mathrm{vec}(\mathrm{Z}) \sim N(0, \sigma^2)$, with $\sigma^2 = \mathrm{vec}(\boldsymbol{I}_r)^\top\boldsymbol{\Sigma}\mathrm{vec}(\boldsymbol{I}_r) = 2r\sigma_1 + r^2\sigma_2$. Thus $r^2 n \cdot m_1^2(\widehat{\boldsymbol{R}}_{22}) \xrightarrow{d} \sigma^2\chi^2_1$. Finally, recall that, as in the proof of Theorem 1, $n \cdot s^2(\widehat{\boldsymbol{R}}_{22}) = \mathrm{vec}(\boldsymbol{Z}_{22})^\top\boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{Z}_{22})$ where $\boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{I}_r) = \boldsymbol{0}$. This establishes the independence of the limiting distributions of the component variables in $(n \cdot s^2(\widehat{\boldsymbol{R}}_{22}), n \cdot m_1^2(\widehat{\boldsymbol{R}}_{22}))$, and consequently Theorem 3 follows with some constants σ_1 and σ_2. The values of σ_1 and σ_2 are derived in the Appendix in [30].

Proof of Theorem 4. (i) T_k converges in probability to the sum of $p - k$ smallest eigenvalues of $(\boldsymbol{D} - (p + 2)\boldsymbol{I}_p)^2$ which is positive for $k < q$. (ii) is given in the previous theorem. (iii) follows as

$$T_k = \min_{\boldsymbol{U}\in\mathcal{O}^{p\times(p-k)}} m_1\left(\boldsymbol{U}^\top(\widehat{\boldsymbol{R}} - (p + 2)\boldsymbol{I}_p)^2\boldsymbol{U}\right) \le m_1\left((\boldsymbol{0}, \boldsymbol{I}_{p-k})(\widehat{\boldsymbol{R}} - (p + 2)\boldsymbol{I}_p)^2(\boldsymbol{0}, \boldsymbol{I}_{p-k})^\top\right)$$

and the result follows as, for $k \in \{q, \ldots, p - 1\}$, $(\boldsymbol{0}, \boldsymbol{I}_{p-k})\widehat{\boldsymbol{R}}(\boldsymbol{0}, \boldsymbol{I}_{p-k})^\top$ is a $(p - k) \times (p - k)$-submatrix of $\widehat{\boldsymbol{R}}_{22}$ with the known limiting distribution.

Proofs for Section 5

Proof of Lemma 3. For $H \ge p$, let $\widehat{\boldsymbol{\gamma}} = (\hat{\gamma}_{q+1}, \ldots, \hat{\gamma}_p)^\top$ denote the $p - q$ smallest ordered singular values of $\widehat{\boldsymbol{B}}\boldsymbol{Q}^\top$. When $q + 1 < H < p$, we use the same notation while noting $\hat{\gamma}_{H+1} = \cdots = \hat{\gamma}_p = 0$. Likewise, let $\widehat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_{p-q})^\top$ denote the ordered singular values of $\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top$. Since $\sqrt{n}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}^\top = O_P(1)$, it follows respectively from Theorems 4.1 and 4.2 in [14] that $\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\eta}} = O_P\left(n^{-3/4}\right)$ and $\widehat{\boldsymbol{\gamma}} = O_P\left(n^{-1/2}\right)$. Next, observe that $(p - q)T_q = \widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\gamma}}$ and $(p - q)m_1(\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top) = \widehat{\boldsymbol{\eta}}^\top\widehat{\boldsymbol{\eta}}$. Hence,

$$(p - q)\{T_q - m_1(\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top)\} = 2\widehat{\boldsymbol{\eta}}^\top(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\eta}}) + (\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\eta}})^\top(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\eta}}),$$

which is $O_P\left(n^{-5/4}\right) + O_P\left(n^{-3/2}\right) = O_P\left(n^{-5/4}\right)$.

Proof of Theorem 5. By Lemma 3, the limiting distributions of $n{\cdot}T_q$ and $n{\cdot}m_1(\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top)$ are the same. Let $\boldsymbol{x}_h^* \in \mathbb{R}^{p-q}$ refer to the last $p-q$ components of $\mathbb{I}(y_i \in \mathbb{S}_h)\boldsymbol{x} \in \mathbb{R}^p$, $h \in \{1, \ldots, H\}$. Hence, under H_{0q}, $\boldsymbol{x}^* = \boldsymbol{z}_2$ is independent of the response y. Since $f_h \cdot \bar{\boldsymbol{x}}_h^* = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{(h),i}^*$, where $\boldsymbol{x}_{(h),i}^* = \boldsymbol{x}_i^*\mathbb{I}(y_i \in \mathbb{S}_h)$, with $\mathbb{E}(\boldsymbol{x}_{(h)}^*) = p_h\mathbb{E}(\boldsymbol{x}^*) = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{x}_{(h)}^*) = p_h\mathrm{Cov}(\boldsymbol{x}^*) = p_h\boldsymbol{I}_{p-q}$, $\mathrm{Cov}(\boldsymbol{x}_{(h)}^*, \boldsymbol{x}_{(m)}^*) = \boldsymbol{0}$ for $h \neq m$, and $f_h \xrightarrow{P} p_h$, it follows from the central limit theorem and from Slutsky's theorem that $\sqrt{n}\left(\sqrt{f_1}\,\bar{\boldsymbol{x}}_1^*, \;\ldots\;, \sqrt{f_H}\,\bar{\boldsymbol{x}}_H^*\right) \xrightarrow{d} \mathrm{Z}$, where the elements of the $(p-q)\times H$ random matrix Z are i.i.d. $N(0, 1)$.

Since $\widehat{\boldsymbol{S}}_1 \xrightarrow{p} \boldsymbol{I}_p$ and $\bar{\boldsymbol{x}}^* = \sum_{h=1}^H f_h\bar{\boldsymbol{x}}_h^*$, we obtain $\sqrt{n}\cdot\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top \xrightarrow{d} \mathrm{Z}(\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\boldsymbol{Q}_2^\top$ with $\boldsymbol{\pi}^\top = (\sqrt{p_1}, \ldots, \sqrt{p_H})$. Hence $n\cdot\widetilde{\boldsymbol{B}}_2\widetilde{\boldsymbol{B}}_2^\top \xrightarrow{d} \mathrm{Z}\boldsymbol{P}\mathrm{Z}^\top$, where $\boldsymbol{P} = (\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\boldsymbol{Q}_2^\top\boldsymbol{Q}_2(\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top)$. It is shown below that $\boldsymbol{P}$ is idempotent with rank $H-q-1$, which implies $\mathrm{Z}\boldsymbol{P}\mathrm{Z}^\top \sim \mathit{Wishart}_{p-q}(H-q-1, \boldsymbol{I}_{p-q})$, and consequently, $n\cdot tr(\widehat{\boldsymbol{B}}_2\boldsymbol{Q}_2^\top\boldsymbol{Q}_2\widehat{\boldsymbol{B}}_2^\top) \xrightarrow{d} \mathrm{tr}(\mathrm{Z}\boldsymbol{P}\mathrm{Z}^\top) \sim \chi^2_{(p-q)(H-q-1)}$.

To complete the proof, note that since $\boldsymbol{\mu} = \boldsymbol{0}$, it follows that $\boldsymbol{B}\boldsymbol{\pi} = \boldsymbol{0}$ and hence $\boldsymbol{Q}_1\boldsymbol{\pi} = \boldsymbol{0}$. Also, since $\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top$ is idempotent with rank $H-1$, we have

$$\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top = (\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\boldsymbol{Q}^\top\boldsymbol{Q}(\boldsymbol{I}_H - \boldsymbol{\pi}\boldsymbol{\pi}^\top) = \boldsymbol{Q}_1^\top\boldsymbol{Q}_1 + \boldsymbol{P},$$

which implies $\boldsymbol{P}$ is idempotent with rank $H-q-1$.

CRediT authorship contribution statement

Klaus Nordhausen: Development and design of methodology, Programming, Software development. **Hannu Oja:** Writing – original draft, Development and design of methodology. **David E. Tyler:** Development and design of methodology, First versions of the proofs.

Acknowledgments

David E. Tyler's research was partially supported by the National Science Foundation, United States Grant No. DMS-1407751. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation. The authors wish to thank the editors and reviewers for their valuable comments.

References

[1] R. Beran, M.S. Srivastava, Bootstrap tests and confidence regions for functions of a covariance matrix, Ann. Statist. 15 (1) (1987) 470–471.
[2] M. Bilodeau, D. Brenner, Theory of Multivariate Statistics, Springer-Verlag, New York, 1999.
[3] G. Blanchard, M. Sugiyama, M. Kawanabe, V. Spokoiny, K.-R. Müller, Non-Gaussian component analysis: a semi-parametric framework for linear dimension reduction, in: Y. Weiss, B. Schölkopf, J. Platt (Eds.), Advances in Neural Information Processing Systems, Vol. 18, MIT Press, Cambridge, MA, 2006, pp. 131–138.
[4] E. Bura, R.D. Cook, Extending sliced inverse regression: the weighted chi-squared test, J. Amer. Statist. Assoc. 96 (455) (2001) 996–1003.
[5] E. Bura, J. Yang, Dimension estimation in sufficient dimension reduction: a unifying approach, J. Multivariate Anal. 102 (1) (2011) 130–142.
[6] J.-F.c. Cardoso, Source separation using higher order moments, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1989, pp. 2109–2112.
[7] H. Caussinus, A. Ruiz-Gazen, Projection pursuit and generalized principal component analysis, in: S. Morgenthaler, E. Ronchetti, W. Stahel (Eds.), New Directions in Statistical Data Analysis and Robustness, Birkhäuser, Basel, 1933, pp. 328–332.
[8] R.D. Cook, Testing predictor contributions in sufficient dimension reduction, Ann. Statist. 32 (3) (2004) 1062–1092.
[9] R. Cook, S. Weisberg, Sliced inverse regression for dimension reduction: comment, J. Amer. Statist. Assoc. 86 (414) (1991) 328–332.
[10] C. Croux, G. Haesbroeck, Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, Biometrika 87 (3) (2000) 603–618.
[11] J.-J. Daudin, C. Duby, P. Trécourt, Stability of principal component analysis studied by the bootstrap method, Stat. A J. Theor. Appl. Stat. 19 (2) (1988) 241–258.
[12] S. Dray, On the number of principal components: a test of dimensionality based on measurements of similarity between matrices, Comput. Statist. Data Anal. 52 (4) (2008) 2228–2237.
[13] M.L. Eaton, D.E. Tyler, On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix, Ann. Statist. 19 (1) (1991) 260–271.
[14] M.L. Eaton, D. Tyler, The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis, J. Multivariate Anal. 50 (2) (1994) 238–264.
[15] P. Hall, S.R. Wilson, Two guidelines for bootstrap hypothesis testing, Biometrics 47 (2) (1991) 757–762.
[16] F. Hampel, E. Ronchetti, P. Rousseeuw, W. Stahel, Robust Statistics: the Approach Based on Influence Functions, Wiley, New York, 1986.
[17] T.P. Hettmansperger, R.H. Randles, A practical affine equivariant multivariate median, Biometrika 89 (4) (2002) 851–860.
[18] D.A. Jackson, Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches, Ecology 74 (8) (1993) 2204–2214.
[19] I. Jolliffe, Principal Component Analysis, second ed., in: Springer Series in Statistics, Springer-Verlag, New York, 2002, p. xxx+487.
[20] A. Kankainen, S. Taskinen, H. Oja, Tests of multinormality based on location vectors and scatter matrices, Stat. Methods Appl. J. Ital. Stat. Soc. 16 (3) (2007) 357–379.
[21] D. Kelker, Distribution theory of spherical distributions and a location-scale parameter generalization, SankhyĀ 32 (1970) 419–438.
[22] D.M. Lane, The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated, Quant. Methods Psychol. 12 (2) (2016) 114–122.
[23] K.-C. Li, Sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (414) (1991) 316–342, With discussion and a rejoinder by the author.

[24] K.-C. Li, On principal hessian directions for data visualization and dimension reduction: another application of Stein's lemma, J. Amer. Statist. Assoc. 87 (420) (1992) 1025–1039.
[25] B. Li, Sufficient Dimension Reduction Methods and Applications with R, CRC Press, Boca Raton, 2018.
[26] W. Luo, B. Li, Combining eigenvalues and variation of eigenvectors for order determination, Biometrika 103 (4) (2016) 875–887.
[27] Y. Ma, L. Zhu, A review on dimension reduction, Int. Stat. Rev. Rev. Int. Stat. 81 (1) (2013) 134–150.
[28] K. Nordhausen, H. Oja, E. Ollila, Multivariate models and the first four moments, in: Nonparametric Statistics and Mixture Models, World Sci. Publ., Hackensack, NJ, 2011, pp. 267–287.
[29] K. Nordhausen, H. Oja, D.E. Tyler, Tools for exploring multivariate data: The package ICS, J. Stat. Softw. 28 (6) (2008) 1–31.
[30] K. Nordhausen, H. Oja, D.E. Tyler, J. Virta, Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace, IEEE Signal Process. Lett. 24 (6) (2017) 887–891.
[31] K. Nordhausen, H. Oja, D.E. Tyler, J. Virta, ICtest: Estimating and testing the number of interesting components in linear dimension reduction, 2021, R package version 0.3-4.
[32] K. Nordhausen, S. Sirkia, H. Oja, D.E. Tyler, ICSNP: Tools for multivariate nonparametrics, 2018, R package version 1.1-1.
[33] K. Nordhausen, D.E. Tyler, A cautionary note on robust covariance plug-in methods, Biometrika 102 (3) (2015) 573–588.
[34] H. Oja, S. Sirkiä, J. Eriksson, Scatter matrices and independent component analysis, Aust. J. Stat. 35 (2&3) (2016) 175–189.
[35] U. Radojičić, K. Nordhausen, Non-Gaussian component analysis: Testing the dimension of the signal subspace, in: M. Maciak, M. Pešta, M. Schindler (Eds.), Analytical Methods in Statistics, Springer International Publishing, Cham, 2020, pp. 101–123.
[36] J.R. Schott, A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix, J. Multivariate Anal. 97 (4) (2006) 827–843.
[37] F.J. Theis, M. Kawanabe, K.-R. Müller, Uniqueness of non-Gaussianity-based dimension reduction, IEEE Trans. Signal Process. 59 (9) (2011) 4478–4482.
[38] D.E. Tyler, Radial estimates and the test for sphericity, Biometrika 69 (2) (1982) 429–436.
[39] D.E. Tyler, The asymptotic distribution of principal component roots under local alternatives to multiple roots, Ann. Statist. 11 (4) (1983) 1232–1242.
[40] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, Ann. Statist. 15 (1) (1987) 234–251.
[41] D.E. Tyler, F. Critchley, L. Dümbgen, H. Oja, Invariant co-ordinate selection, J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (3) (2009) 549–592.
[42] V.M.N.C.S. Vieira, Permutation tests to estimate significances on principal components analysis, Comput. Ecol. Softw. 2 (2012) 103–123.
[43] J. Virta, K. Nordhausen, Estimating the number of signals using principal component analysis, Stat 8 (1) (2019) e231.
[44] S. Weisberg, Dimension reduction regression in R, J. Stat. Softw. Art. 7 (1) (2002) 1–22.
[45] Z. Ye, R.E. Weiss, Using the bootstrap to select one of a new class of dimension reduction methods, J. Amer. Statist. Assoc. 98 (464) (2003) 968–979.
[46] L. Zhu, B. Miao, H. Peng, On sliced inverse regression with high-dimensional covariates, J. Amer. Statist. Assoc. 101 (474) (2006) 630–643.
[47] L. Zhu, T. Wang, L. Zhu, L. Ferré, Sufficient dimension reduction through discretization-expectation estimation, Biometrika 97 (2) (2010) 295–304.