



Kurtosis removal for data pre-processing

Nicola Loperfido¹

Received: 3 December 2020 / Revised: 11 October 2021 / Accepted: 21 February 2022 /

Published online: 19 March 2022

© The Author(s) 2022, corrected publication 2022

Abstract

Mesokurtic projections are linear projections with null fourth cumulants. They might be useful data pre-processing tools when nonnormality, as measured by the fourth cumulants, is either an opportunity or a challenge. Nonnull fourth cumulants are opportunities when projections with extreme kurtosis are used to identify interesting nonnormal features, as for example clusters and outliers. Unfortunately, this approach suffers from the curse of dimensionality, which may be addressed by projecting the data onto the subspace orthogonal to mesokurtic projections. Nonnull fourth cumulants are challenges when using statistical methods whose sampling properties heavily depend on the fourth cumulant themselves. Mesokurtic projections ease the problem by allowing to use the inferential properties of the same methods under normality. The paper shows necessary and sufficient conditions for the existence of mesokurtic projections and compares them with other gaussianization methods. Theoretical and empirical results suggest that mesokurtic transformations are particularly useful when sampling from finite normal mixtures. The practical use of mesokurtic projections is illustrated with the AIS and the RANDU datasets.

Keywords Closed skew-normal · Finite mixture · Gaussianization · High-dimensional · Kurtosis · Projection pursuit

Mathematics Subject Classification 15A18 · 62H10

1 Introduction

A fourth cumulant of a random vector with finite fourth moments is the fourth derivative, evaluated at the origin, of the cumulant generating function of the random vector itself, that is the logarithm of its characteristic function. All fourth cumulants of a

✉ Nicola Loperfido
nicola.loperfido@uniurb.it

¹ Dipartimento di Economia, Società e Politica (DESP), Università degli Studi di Urbino “Carlo Bo”, Via Saffi 42, 61029 Urbino, PU, Italy

multivariate normal distribution equal zero, and we refer to distributions with the same property as to mesokurtic distributions. Also, we refer to transformations leading to null fourth cumulants as to mesokurtic transformations. In particular, when the mesokurtic transformation is a linear function of the original variables, we refer to it as to a mesokurtic projection. With a little abuse of language, and for ease of description, we write that mesokurtic transformations remove kurtosis. Mesokurtic transformations include as special cases transformations to normality, also known as gaussianizing transformations. They have been extensively studied in scientific fields other than Statistics, as for example Physics (Yu et al. 2016). Mesokurtic transformations are of interest whenever the performance of the chosen statistical method either increases or decreases with the absolute values of fourth-order cumulants.

Researchers from social and life sciences are often concerned with the lack of normality of their data, since many multivariate statistical techniques are not robust when the sampled distribution is wrongly assumed to be normal. Sometimes the performance of the statistical technique depends on the fourth cumulants of the sampled distribution, as exemplified by the following cases.

Inference on covariance matrices is notoriously very sensitive to fourth cumulants. Mardia (1974) showed that the performance of a likelihood test based on the erroneous assumption of normality crucially depends on a measure of multivariate kurtosis which is a simple function of fourth cumulants. Schott (2002) used fourth cumulants to derive a robust procedure for testing the equality of the population covariance and a given matrix. Yanagihara et al. (2005) extended Mardia's results to several likelihood tests on covariance matrices based on the normality assumption, and showed that their performances depend on their fourth cumulants.

Variogram estimation is of paramount importance in spatial statistics but spatial data are often nonnormal, also due to preferential sampling (Loperfido and Guttorp 2008). Genton et al. (2001) modelled nonnormality with the multivariate skew-normal distribution and derived analytical formulae for covariances of variogram estimators, and showed that they depend on the fourth cumulants of the same distribution. Kim (2005) extended their results to mixtures of skew-normal distributions. Rezvandehy and Deutsch (2018) addressed preferential sampling with fourth cumulants.

Relevance of fourth cumulants is not limited to covariance testing and variogram estimation. Yanagihara (2007) stressed the importance of fourth cumulants in the multivariate linear model. Arevalillo and Navarro (2012) investigated the effect of fourth cumulants on the Fisher discriminant function (i.e. the projection which best separates the means of two multivariate distributions), and found it to be nonnegligible. The large-sample approximation of the multivariate sample mean by the skew-normal distribution improves when the fourth cumulants of the two distributions are close to each other (Christiansen and Loperfido 2014).

Multivariate normality is usually pursued by means of the componentwise, non-linear, univariate transformations proposed by either Box and Cox (1964) or Tukey (1977). Tsay et al. (2017) thoroughly review these and other transformations to normality, while proposing a new one which is particularly apt at dealing with platykurtic distributions. Unfortunately, nonlinearly transformed variables might not be easily interpretable nor jointly normal, as pointed out by Lin and Lin (2010), among others. Moreover, the same methods are inappropriate when the joint distribution is not

normal, but univariate marginals are, as it happens for the distributions described in Arnold et al. (2001). Loperfido (2014) and Loperfido (2019) used a simple argument to show how nonlinear transformations to normality are inappropriate when using Hotelling's one-sample test.

Mesokurtic projections do not suffer from these limitations. We illustrate this point with the bivariate distribution $2\phi(x)\phi(y)\Phi(\lambda xy)$ introduced by Arnold et al. (2001), where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability and the cumulative density functions of a standard normal distribution, while λ is a real value. Adcock (2021) thoroughly investigated its properties and proposed some generalizations. The default approach to the gaussianization of the random vector (X, Y) would be raising both its components to an appropriate power. However, there do not exist two real values α and β such that the joint distribution of (X^α, Y^β) is bivariate normal. On the other hand, there are two normal, and therefore mesokurtic, projections of (X, Y) , that is X and Y .

The performance of kurtosis-based projection pursuit tends to increase with the absolute values of fourth cumulants. Projection pursuit is a multivariate statistical technique aimed at finding interesting data projections, with interestingness quantified by the projection index: the data projection with the highest value of the projection index is regarded as the most interesting. The normal distribution is commonly regarded as the least interesting (Huber 1985), so that the projection index often measures a nonnormality feature such as skewness, kurtosis or multimodality. In particular, kurtosis-based projection pursuit uses the absolute value of the fourth standardized cumulant as a projection index, consistently with the projection pursuit criteria stated in Huber (1985). Kurtosis-based projection pursuit therefore looks for interesting projections by either maximizing or minimizing kurtosis, that is the fourth standardized moment. Kurtosis-based projection pursuit appears in cluster analysis (Peña and Prieto 2000, 2001b), outlier detection (Galeano et al. 2006; Peña and Prieto 2001a), normality testing (Malkovich and Afifi 1973), independent component analysis (Girolami and Fyfe 1996), invariant coordinate selection (Alashwali and Kent 2016), chemometrics (Hou and Wentzell 2014), finance (Loperfido 2020b).

Projection pursuit is commonly regarded as more problematic when applied to datasets with more variables than units (see, for example, Hui and Lindsay 2010; Lindsay and Yao 2012). Bickel et al. (2018) thoroughly investigated the asymptotic properties of projection pursuit for several ratios of the number of variables to the number of units, finding serious shortcomings of the method when the number of variables is much greater than the number of units. Pires and Branco (2019) showed that two-dimensional projections of datasets with more variables than units could approximate any given set of bivariate data with the same number of units. Lee and Cook (2010) illustrated the limitation of projection pursuit for classification with a real dataset containing 3571 variables recorded from 72 cases. However, the performance of kurtosis-based projection pursuit might rapidly deteriorate as the number of variables increases while the number of units remains fixed, even if the former remains much smaller than the latter (Loperfido 2020a).

The problems posed by high-dimensional data might be eased by means of sparse projection pursuit, that is projection pursuit performed on a small subset of variables, either original or projected (Bickel et al. 2018). Jones and Sibson (1987) proposed a linear transformation of the data into two mutually disjoint subsets of variables, one

deemed uninteresting and the other deemed interesting. The former should be discarded while the latter should be analysed by means of projection pursuit. Following this approach, Blough (1989), Loperfido (2017), Franceschini and Loperfido (2019) removed skewness from the data by means of linear transformations, within the framework of skewness-based projection pursuit. Hui and Lindsay (2010) and Lindsay and Yao (2012) linearly transformed the data into two mutually orthogonal subsets of interesting and uninteresting projections. Ray (2010) used a simulated dataset with many more variables than units to illustrate the merits of this approach. In kurtosis-based projection pursuit, where the least interesting distributions are those with null fourth order cumulants, data are linearly transformed into two mutually orthogonal sets of variables, of which only one is mesokurtic. Then the mesokurtic set of variables is removed and kurtosis optimization is carried out on the remaining variables.

This paper uses several kurtosis matrices to obtain mesokurtic projections. It states both sufficient and necessary conditions for the existence of mesokurtic projections. The approach is algebraic in nature, since it relies on matrix concepts, including null linear spaces, spectra and generalized eigenvalues. As a major advantage, sampling properties of mesokurtic projections might be easily derived from the theory of random matrices and their spectra. The remainder of the paper is organized as follows. Section 2 uses kurtosis matrices to investigate the existence and the properties of mesokurtic projections. Section 3 contains some simulation studies related to model-based clustering. Section 4 applies the proposed method to a subset of the AIS dataset containing 23 units and 11 variables, which is a nonnegligible variables-to-units ratio. Section 5 uses the RANDU dataset to show that mesokurtic projections might be a viable alternative to nonlinear transformations to normality. Section 6 contains some concluding remarks, mentions further applications of mesokurtic projections, discusses their limitations and suggests some extensions of the proposed method to datasets with more variables than units. The proofs of the theorems are in the Appendix.

2 Theory

This section describes a method for obtaining projections with null or negligible fourth cumulants, that is mesokurtic and nearly mesokurtic projections. Firstly, it uses bivariate random vectors to illustrate situations where mesokurtic projections do not exist. Secondly, provides necessary and sufficient conditions for the existence of mesokurtic projections of random vectors in any dimension. Thirdly, the section briefly discusses the sampling properties of mesokurtic projections.

The $ijkl$ -th cumulant of a d -dimensional random vector $x = (X_1, \dots, X_d)^T$ satisfying $E(X_i^4) < +\infty$ for $i = 1, \dots, d$ is

$$\kappa_{ijkl} = \frac{\partial^4 \log E[\exp(\iota^T x)]}{\partial \iota_i \partial \iota_j \partial \iota_h \partial \iota_k} \bigg|_{\iota=0_d}, \quad (1)$$

where $\iota = \sqrt{-1}$, $\iota \in \mathbb{R}^d$ and $E[\exp(\iota^T x)]$ is the cumulant generating function. In the univariate case, the only fourth cumulant is

$$\kappa_4(X) = E[(X - \mu)^4] - 3\sigma^4, \quad (2)$$

where μ and σ^2 are the mean and the variance of a random variable X satisfying $E(X^4) < \infty$.

The bivariate case provides a good insight into nonexistence of mesokurtic projections. Let U and W be two standardized, independent random variables whose fourth cumulants $\kappa_4(U)$ and $\kappa_4(W)$ are both positive. Elementary properties of cumulants imply that the fourth cumulant of the projection $hU + kW$ is positive, too:

$$\kappa_4(hU + kW) = h^4\kappa_4(U) + k^4\kappa_4(W) > 0. \quad (3)$$

As a direct consequence, there does not exist a mesokurtic projection of (U, W) .

Nonexistence of mesokurtic projections of bivariate and standardized random vectors becomes more difficult to ascertain when the independence assumption is removed. Let (X, Y) be a bivariate and standardized random vector:

$$E(X) = E(Y) = E(XY) = 0, E(X^2) = E(Y^2) = 1.$$

Also, let $\beta_i = E(X^i Y^{4-i})$, for $i = 0, 1, 2, 3, 4$. The fourth cumulant of the linear combination $aX + Y$ is a fourth-order polynomial in a :

$$E[(aX + Y)^4] - 3(a^2 + 1)^2 = (\beta_4 - 3)a^4 + 4\beta_3a^3 + 6(\beta_2 - 1)a^2 + 4\beta_1a + \beta_0 - 3.$$

A mesokurtic linear function of X and Y exists if and only if the polynomial has real roots.

The problem of ascertain the existence of mesokurtic projections becomes even more complicated when considering standardized random vectors with more than two components. We address the problem by recalling that fourth cumulants are simple functions of the fourth and second central moments. The $ijhk$ -th moment of x is the expectation $\mu_{ijhk} = E(X_i X_j X_h X_k)$, for $i, j, h, k = 1, \dots, d$. The $ijhk$ -th central moment of x is the $ijhk$ -th moment of $x - \mu$, where $\mu = (\mu_1, \dots, \mu_d)^T$ is the mean of x :

$$\bar{\mu}_{ijhk} = E[(X_i - \mu_i)(X_j - \mu_j)(X_h - \mu_h)(X_k - \mu_k)]. \quad (4)$$

The $ijhk$ -cumulant can then be represented as

$$\kappa_{ijhk} = \bar{\mu}_{ijhk} - \sigma_{ij}\sigma_{hk} - \sigma_{ih}\sigma_{jk} - \sigma_{ik}\sigma_{jh}, \quad (5)$$

where σ_{ab} is the covariance between the a -th and the b -th components of x , with $a, b=1, \dots, d$. For example, the fourth cumulants of $(X_1, X_2)^T$, expressed as functions of its second and fourth central moments, are

$$\begin{aligned} \kappa_{1111} &= \bar{\mu}_{1111} - 3\sigma_{11}^2, \kappa_{2222} = \bar{\mu}_{2222} - 3\sigma_{22}^2, \kappa_{1112} = \bar{\mu}_{1112} - 3\sigma_{11}\sigma_{12}, \\ \kappa_{1122} &= \bar{\mu}_{1122} - \sigma_{11}\sigma_{22} - 2\sigma_{12}^2, \kappa_{1222} = \bar{\mu}_{1222} - 3\sigma_{12}\sigma_{22}. \end{aligned}$$

We first consider the fourth moment (matrix) of a d -dimensional random vector $x = (X_1, \dots, X_d)^T$ satisfying $E(X_i^4) < +\infty$, for $i = 1, \dots, d$, that is the $d^2 \times d^2$ matrix $M_{4,x} = E(x \otimes x^T \otimes x \otimes x^T)$, where “ \otimes ” denotes the Kronecker product. It conveniently arranges all the fourth-order moments $\mu_{ijhk} = E(X_i X_j X_h X_k)$ of x (where $i, j, h, k = 1, \dots, d$) and admits the block matrix representation $M_{4,x} = \{B_{pq}\}$, where $B_{pq} = E(X_p X_q x x^T)$, for $p, q = 1, \dots, d$. For example, the fourth moment of $(X_1, X_2)^T$ is

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} \mu_{1111} & \mu_{1112} & \mu_{1112} & \mu_{1122} \\ \mu_{1112} & \mu_{1122} & \mu_{1122} & \mu_{1222} \\ \mu_{1112} & \mu_{1122} & \mu_{1122} & \mu_{1222} \\ \mu_{1122} & \mu_{1222} & \mu_{1222} & \mu_{2222} \end{pmatrix}. \quad (6)$$

If the variance Σ of x is positive definite, its fourth standardized moment $M_{4,z}$ is the fourth moment of $z = \Sigma^{-1/2}(x - \mu)$, where $\Sigma^{-1/2}$ is the symmetric, positive definite square root of the concentration matrix Σ^{-1} . Most measures of multivariate kurtosis are either scalar-type or matrix-type functions of $M_{4,z}$ (Loperfido 2017, 2020a). In the univariate case $M_{4,z}$ coincides with Pearson’s measure of kurtosis:

$$\beta_2(X) = \frac{E[(X - \mu)^4]}{\sigma^4}. \quad (7)$$

The eigenstructure of $M_{4,z}$ has several interesting features: eigenvectors of $M_{4,z}$ associated with its positive eigenvalues are vectorized, symmetric matrices and the eigenvector of $M_{4,z}$ associated with its largest eigenvalue (that is the dominant eigenvector of $M_{4,z}$) is a semidefinite matrix (Loperfido 2017). The following theorem gives a necessary condition for the existence of mesokurtic projections based on the eigenstructure of the fourth standardized moment matrix. It is reminiscent of the distinction between platykurtic and leptokurtic random variables, whose fourth standardized moments are smaller and greater than three, respectively.

Proposition 1 *Let $M_{4,z}$ be the fourth moment matrix of the d -dimensional, standardized random vector z and let v be an eigenvector of $M_{4,z}$ which might be represented as a symmetric, definite matrix. Then the dominant eigenvalue of $M_{4,z}$ is simple and v is the associated eigenvector. Also, mesokurtic projections of z do not exist if all positive eigenvalues of $M_{4,z}$ are either greater or smaller than three.*

Proposition 1 suggests the following procedure for computing projections with negligible fourth cumulants. First, compute the nonnull eigenvalues of the fourth standardized moment matrix. If they are all greater (smaller) than three the projections with fourth cumulants closer to zero are those which minimize (maximize) kurtosis. In these circumstances, kurtosis alleviation coincides with kurtosis optimization, that is kurtosis-based projection pursuit. The computational issues of the latter might be addressed by the method proposed in Franceschini and Loperfido (2018), which are implemented in the R package Kurt (Franceschini and Loperfido 2020).

On the other hand, if some nonnull eigenvalues of the fourth standardized moment are smaller than three while other eigenvalues are greater than three, it is worth looking

for mesokurtic projections. An intuitively appealing approach has been proposed by Peña and Prieto (2001a) and Peña and Prieto (2001b) within the framework of outlier detection. First, find the projection with maximal kurtosis. Second, project the data onto its orthogonal subspace. Then repeat these steps until all remaining projections are mesokurtic or nearly so. This approach, which may be referred to as to iterative projection pursuit approach, proved to be useful when nonnormality is due to the presence of outliers, but may not detect mesokurtic projections when data are generated from different models, as it happens in independent component analysis.

Independent component analysis (ICA) is a multivariate statistical technique aimed at recovering independent, unobserved signals by appropriate data projections. The basic ICA model is $x = b + As$, where x is a d -dimensional random vector, b is a d -dimensional real vector, A is a $d \times d$ invertible real matrix and the components of s are mutually independent, standardized random variables of which at most one is normal (Miettinen et al. 2015).

Iterative projection pursuit does not detect mesokurtic projections when the fourth cumulants of the signals are nonnull and take both signs. We illustrate the point with the bivariate ICA model

$$x = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, s = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix},$$

where the first, second, fourth cumulants of S_1 and S_1 are

$$\kappa_1(S_1) = \kappa_1(S_2) = 0, \kappa_2(S_1) = \kappa_2(S_2) = 0, \kappa_4(S_1) = 1, \kappa_4(S_2) = -1.$$

The projection of x with maximal kurtosis coincides, up to location and scale changes, with the first signal, which is leptokurtic:

$$S_1 = 2X_2 - 3X_1 = 2(2S_1 + 3S_2) - 3(S_1 + 2S_2).$$

The projection of x orthogonal to $2X_2 - 3X_1$ coincides, up to location and scale changes, with the second signal, which is platykurtic:

$$S_2 = 2X_1 - X_2 = 2(S_1 + 2S_2) - (2S_1 + 3S_2).$$

Neither projection is mesokurtic, but the model implies the existence of the mesokurtic projection $X_1 + X_2$.

A sufficient condition for the existence of mesokurtic projections might be established using the cokurtosis matrix (see, for example, Jondeau and Rockinger 2006), that is the $d \times d^3$ matrix

$$cok(x) = E \left[(x - \mu) \otimes (x - \mu)^T \otimes (x - \mu)^T \otimes (x - \mu)^T \right]. \quad (8)$$

For example, the cokurtosis of $(X_1, X_2)^T$ is

$$\text{cok}(X_1, X_2) = \begin{pmatrix} \bar{\mu}_{1111} & \bar{\mu}_{1112} & \bar{\mu}_{1112} & \bar{\mu}_{1122} & \bar{\mu}_{1112} & \bar{\mu}_{1122} & \bar{\mu}_{1122} & \bar{\mu}_{1222} \\ \bar{\mu}_{1112} & \bar{\mu}_{1122} & \bar{\mu}_{1122} & \bar{\mu}_{1222} & \bar{\mu}_{1122} & \bar{\mu}_{1222} & \bar{\mu}_{1222} & \bar{\mu}_{2222} \end{pmatrix}. \quad (9)$$

The standardized cokurtosis of x is just the cokurtosis of $z = \Sigma^{-1/2}(x - \mu)$. The fourth standardized matrix $M_{4,z} = \{M_{pq} = E(Z_p Z_q z z^T)\}$ and the standardized cokurtosis $\text{cok}(z)$ might be regarded as block matrices where the same blocks are arranged in different ways. The latter aligns side by side the blocks of the fourth standardized moment in such a way that the blocks with the smallest first indices appear first, followed by those having the smallest second indices: $\text{cok}(x) = (M_{11}, M_{12}, \dots, M_{dd})$ where the block M_{ij} appears before the block M_{aj} if $i < a$, and the block M_{ij} appears before the block M_{ib} if $j < b$.

Both $M_{4,z}$ and $\text{cok}(z)$ may contain up to $d(d+1)(d+2)(d+3)/24$ distinct elements. Since this number grows very quickly with the vector's dimension, it is convenient to summarize them with smaller matrices. Cardoso (1989) proposed $K_z = E(z^T z z z^T)$ as a kurtosis matrix. Mõri et al. (1993) independently proposed and discussed its variant $K_z - (d+2)I_d$. Statistical applications of both kurtosis matrices include independent component analysis (Cardoso 1989), generalized principal components (Caussinus and Ruiz-Gazen 2009), invariant coordinate selection (Alashwali and Kent 2016) and cluster analysis (Peña et al. 2010). Loperfido (2017) highlighted the connection between K_z and tensor contraction. The kurtosis matrix K_z is a function of the fourth moment matrix of a standardized random vector and is commonly regarded as a good compromise between detail and synthesis, when studying the kurtosis structure of the random vector itself. The kurtosis matrix K_z does not account for expectations of products of four different components of z : elements in K_z do not depend on $E(Z_i Z_j Z_h Z_k)$, when i, j, h, k differ from each other. For this reason we refer to K_z as the partial kurtosis matrix. In the bivariate case it is

$$\begin{bmatrix} E(Z_1^4) + E(Z_1^2 Z_2^2) & E(Z_1^3 Z_2) + E(Z_1 Z_2^3) \\ E(Z_1^3 Z_2) + E(Z_1 Z_2^3) & E(Z_2^4) + E(Z_1^2 Z_2^2) \end{bmatrix}. \quad (10)$$

The following theorem establishes a sufficient condition for the existence of mesokurtic projections. It describes a matrix depending on both standardized cokurtosis and partial kurtosis which recovers mesokurtic projections, when it is not of full rank.

Proposition 2 *Let $\text{cok}(z)$ and K_z be the cokurtosis and the partial kurtosis of the standardized, d -dimensional random vector z . Also, let the columns of the $d \times h$ matrix B^T , with $h > 0$, span the null space of the matrix*

$$\text{cok}(z) \text{cok}^T(z) - 6K_z + 3(d+2)I_d, \quad (11)$$

where I_d is the d -dimensional identity matrix. Then the fourth cumulant of the random vector Bz is a null matrix.

In practice, kurtosis matrices need to be estimated. To this end, we use a $n \times d$ data matrix X and the corresponding sample covariance matrix S , which is assumed to be of full rank. This assumption implies that there are more units than variables, the opposite case falling outside the scope of the present paper. Possible extensions of the proposed method to datasets with more variables than units are discussed in the Conclusions section.

First we obtain the standardized data matrix $Z = H_n X S^{-1/2}$, where H_n is the $n \times n$ centring matrix and $S^{-1/2}$ is the symmetric square root of S^{-1} . The fourth sample standardized moment, the sample standardized cokurtosis and the sample partial kurtosis are

$$\begin{aligned} M_{4,Z}^{(n)} &= \frac{1}{n} \sum_{i=1}^d z_i \otimes z_i^T \otimes z_i \otimes z_i^T, \\ \text{cok}^{(n)}(Z) &= \frac{1}{n} \sum_{i=1}^d z_i \otimes z_i^T \otimes z_i^T \otimes z_i^T \\ \text{and } K_{4,Z}^{(n)} &= \frac{1}{n} \sum_{i=1}^d z_i^T z_i z_i z_i^T, \end{aligned}$$

where z_i^T is the i -th row of Z . Under random sampling, the sequences

$$\left\{ M_{4,Z}^{(n)} \right\}, \left\{ \text{cok}^{(n)}(Z) \right\} \text{ and } \left\{ K_{4,Z}^{(n)} \right\} \quad (12)$$

almost surely converge to the matrices $M_{4,z}$, $\text{cok}(z)$ and K_z , as long as $E(X_i^4) < \infty$, for $i = 1, \dots, d$. Since eigenvalues are continuous functions of their matrices (Ortega 1987, page 41) eigenvalues of matrices in the three sequences converge almost surely to eigenvalues of the corresponding matrices. Convergence of eigenvectors and convergence in law require some additional assumptions, which are thoroughly described in Rublik (2001). Propositions 1, 2 connect mesokurtic projections to the eigenvalues of symmetric, positive semidefinite matrices, thus easing the task of making inference on the former by using the well-established theory regarding the latter.

Consistently with Proposition 2, mesokurtic projections might be obtained as follows, if some nonnull eigenvalues of the fourth sample standardized moment being smaller than three while other eigenvalues are greater than three. First, compute the eigenvectors corresponding to the smallest eigenvalues of the matrix

$$Q_n = \text{cok}^{(n)}(Z) \left[\text{cok}^{(n)}(Z) \right]^T - 6K_{4,Z}^{(n)} + 3(d+2)I_d. \quad (13)$$

For the sake of clarity we assume that all eigenvalues of Q_n are distinct, as it often occurs in practice, according to our personal experience. The matrix ZB_n estimates the mesokurtic projections that can be obtained from the sampled distributions, where the columns of the matrix B_n are the eigenvectors associated with the k smallest eigenvalues of Q_n , with $k < d$. Alternatively, we could look for random projections

onto the subspace generated by the same eigenvectors. Random projections proved to be very useful for finding interesting projections (Peña and Prieto 2007), but they might also be useful for detecting uninteresting ones, since most data projections are approximatively normal, under mild assumptions (Diaconis and Freedman 1984). The number of projections might be chosen on the ground of subject-matter considerations or after testing for nullity the smallest eigenvalues of Q_n .

The above mentioned approach aimed at finding mesokurtic projections relies on little restrictive assumptions, being Propositions 1, 2 nonparametric in nature. However, a practitioner would greatly appreciate more precise indications when looking for mesokurtic projections. We address the problem with another kurtosis matrix and within the framework of closed skew-normal distributions. Fourth cumulants are often arranged in the exkurtosis, that is the $d \times d^3$ block matrix

$$exk(x) = (K_{11}, K_{12}, \dots, K_{dd}), \text{ where } K_{hk} = \{\kappa_{ijhk}\} \in \mathbb{R}^d \times \mathbb{R}^d$$

and the blocks with smaller first indices appear first, followed by those having smaller second indices: the block K_{ij} appears before the block K_{qj} if $i < q$, and the block K_{ij} appears before the block K_{ip} if $j < p$. The name “exkurtosis” is a shorthand for “excess kurtosis matrix”, since the best known measures of multivariate excess kurtosis are functions of the exkurtosis. Loperfido (2020c) investigates some properties of the exkurtosis matrix. In particular, we have

$$\begin{aligned} exk(x+b) &= exk(x), \\ exk(x+y) &= exk(x) + exk(y) \text{ and} \\ exk(Ax) &= Aexk(x) \left(A^T \otimes A^T \otimes A^T \right), \end{aligned}$$

where b is a d -dimensional real vector, y is a d -dimensional random vector with finite fourth-order moments and independent of x , A is a $k \times d$ real matrix.

The closed skew-normal distribution introduced by Gonzalez-Farias et al. (2003) provides another useful tool for modelling the sample bias. Its name reminds that it is closed with respect to conditioning, affine transformations and convolutions. The random vector x has a closed skew-normal distribution of parameters $\xi, \Omega, \Psi, \eta, \Delta$, and we write $x \sim CSN(\xi, \Omega, \Psi, \eta, \Delta)$, if its density function is

$$\frac{\phi(x; \xi, \Omega) \Phi[\Psi(x - \xi); \eta, \Delta]}{\Phi(0_h; \eta, \Delta + \Psi \Omega \Psi^T)}, \text{ where} \\ x \in \mathbb{R}^d, \xi \in \mathbb{R}^d, \eta \in \mathbb{R}^h, \Psi \in \mathbb{R}^h \times \mathbb{R}^d, \Omega \in \mathbb{R}^d \times \mathbb{R}^d, \Delta \in \mathbb{R}^h \times \mathbb{R}^h,$$

and $\Phi(z; \mu, \Sigma)$ is the cdf of $N_p(\mu, \Sigma)$ evaluated at $z \in \mathbb{R}^p$, while Ω and Δ are symmetric, positive definite matrices. The following proposition gives a sufficient condition for the existence of a linear projection of a closed skew-normal random vector which is normal (and hence mesokurtic).

Proposition 3 Let $x \sim CSN(\xi, \Omega, \Psi, \eta, \Delta)$ be a d -dimensional random vector whose distribution is closed skew-normal with parameters

$$\xi \in \mathbb{R}^d, \eta \in \mathbb{R}^h, \Psi \in \mathbb{R}^h \times \mathbb{R}^d, \Omega \in \mathbb{R}^d \times \mathbb{R}^d, \Delta \in \mathbb{R}^h \times \mathbb{R}^h, \text{ with } h < d.$$

Then there exists a linear projection Bx of x which is normal, where B is a $(d - h) \times d$ full-rank matrix whose rows belong to the null space of the transposed exkurtosis of x , i.e. $B \text{exk}(x)$ is a $(d - h) \times d^3$ null matrix.

Proposition 3 and known inferential results on the closed skew-normal distribution pave the way toward a criterion aimed at finding normal (and therefore mesokurtic) projections. First, estimate the parameters of the closed skew-normal distribution either by maximum likelihood or by alternative methods (Flecher et al. 2009; He et al. 2019). The latter methods should be preferred when the likelihood function is deemed too difficult to compute, since it involves the evaluation of multivariate integrals.

Second, estimate the exkurtosis matrix $\text{exk}(x)$ by its sample counterpart \hat{E} and compute the normalized eigenvectors associated with the $d - \hat{h}$ smallest eigenvalues of $\hat{E}\hat{E}^T$, where the estimate \hat{h} of h is just the size of the symmetric matrix $\hat{\Delta}$, i.e. the estimate of Δ . Finally, estimate the matrix B with the matrix \hat{B} whose i -th row is the normalized eigenvector associated with the above mentioned matrix. Convergence of \hat{B} to a matrix whose columns belong to the null space of $\text{exk}(x)^T$ follows from the mild assumptions stated in Tyler (1981) for eigenvectors of symmetric random matrices.

3 Simulations

The scope of the simulation studies in this section is twofold. Firstly, simulations are used to assess the relevance of the problems posed by projection pursuit when the number of variables approaches the number of units. Secondly, simulations are used to assess the number of mesokurtic projections detected by the method described in the previous section, when sampling from finite normal mixtures.

We addressed the first question by simulating 1000 samples of size $n = 25$ and $n = 50$ from

$$x \sim \pi_1 N_d(0_d, I_d) + (1 - \pi_1) N_d(\mu 1_d, I_d), \quad (14)$$

where 0_d , 1_d and I_d are the d -dimensional null vector, the d -dimensional vector of ones and the $d \times d$ identity matrix, $\pi_1 = 0.1, 0.2, 0.3$, $\mu = 1, 2, 3, 4$, $d = 3, 6, 12, 24$ if $n = 25$ and $d = 6, 12, 24, 48$ if $n = 50$. The ratio of the number of variables to the number of units increases from 12 to 96%, with greater values of μ denoting better separations between the mixture's components and greater values of π_1 denoting smaller skewness and kurtosis. The projection maximizing kurtosis is also the one which best separates the two normal components and is proportional to $x^T 1_d$ (see, for example, Loperfido 2020b). For each simulated sample, we computed the squared correlation between the projection maximizing sample kurtosis and the projection onto the direction of 1_d : higher values of the squared correlation indicate a better

Table 1 Results of the first simulation study

μ	d	$n = 25$			d	$n = 50$		
1		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$
	3	36	34	30	6	23	19	15
	6	21	19	16	12	15	11	9
	12	12	10	9	24	6	4	4
	24	7	6	6	48	2	2	2
2		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$
	3	56	39	26	6	52	22	14
	6	40	23	16	12	36	14	9
	12	22	13	8	24	10	5	4
	24	14	9	6	48	2	2	2
3		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$
	3	72	45	24	6	69	25	13
	6	58	29	17	12	49	16	9
	12	33	15	10	24	15	5	5
	24	18	10	6	48	2	2	2
4		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$		$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$
	3	78	49	24	6	73	27	14
	6	59	31	18	12	53	17	9
	12	39	17	9	24	23	6	5
	24	21	10	6	48	2	2	2

The table reports the integer part of the average squared correlation between the projection maximizing sample kurtosis and the data projection on the direction of 1_d

performance of kurtosis-based projection pursuit in detecting the cluster structure, that is in separating the mixture components.

The simulation's results in Table 1 clearly indicate that kurtosis-based projection pursuit is highly sensitive to the ratio of the number of variables to the number of units: higher ratios lead to worse performances. Kurtosis-based projection pursuit becomes virtually useless when the number of variables is just slightly smaller than the number of units. Another simulations study (not reported here) suggests that under this circumstance the projection maximizing sample kurtosis and the projection of the data onto the direction of 1_d are uncorrelated. Quite surprisingly, the problem gets more serious when the number of units increases. On the other hand, the performance of kurtosis-based projection pursuit improves when the components are better separated and nonnormality increases.

A natural question to ask is whether mesokurtic projections exist for widely used families of multivariate distributions, such as normal mixtures. Let the distribution of the d -dimensional random vector x be the mixture, with weights π_1, \dots, π_k of the normal distributions $N_d(\mu_1, \Sigma), \dots, N_d(\mu_k, \Sigma)$, with $k < d$. In the general case, the fourth cumulants of x are nonnull. Consider now a $k \times d$ matrix B satisfying $B\mu_i = 0_k$ for $i = 1, \dots, k$. It follows that Ax is normally distributed and hence mesokurtic. Does the method described in the previous section detect all these mesokurtic projections?

The question becomes more difficult to answer when the covariances of the normal components are not assumed to coincide. As an example, consider the mixture of two normal distributions with different means and proportional covariance matrices: $\pi_1 N_d(\mu_1, \Sigma) + \pi_2 N_d(\mu_2, \alpha \Sigma)$, with $\pi_1 + \pi_2 = 1$, $\mu_1 \neq \mu_2$, $\pi_1, \pi_2 > 0$, $\alpha \neq 1$. There is no projection whose distribution is normal. Does this mean that there are not mesokurtic projections?

We use simulations to address these problems. Multivariate samples of size 100 were simulated from mixtures of two multivariate normal distributions. Other simulation studies, not reported here, were based on samples of sizes 200 and 300 and led to conclusions similar to the below-mentioned ones.

We first simulated 1000 samples from the normal, homoscedastic mixture $\pi_1 N_d(0_d, I_d) + (1 - \pi_1) N_d(5 \cdot 1_d, I_d)$, where π_1 is either 0.1 or 0.5, for $d = 8, 12, 16, 20$. The sampled distribution is skewed and leptokurtic (symmetric and platykurtic) when $\pi_1 = 0.1$ ($\pi_1 = 0.5$), and admits $d - 1$ jointly normal projections onto the linear subspace orthogonal to 1_d . All samples exhibit either moderate or large units-to-variables ratios, thus motivating mesokurtic projections, either for variable selection or for parametric inference.

Then we simulated 1000 samples from the normal, heteroscedastic mixture

$$0.05 N_d(0_d, \alpha I_d) + 0.95 N_d(51_d, I_d), \quad (15)$$

where α is either 0.5 or 2, for $d = 8, 12, 16, 20$. It models the presence of multivariate outliers, by assuming that the sampled distribution is a two-component mixture, with one mixture weight much smaller than the other. The components with the largest and smallest mixture weights model the bulk of the data and the outliers, respectively (Loperfido 2020b). The outliers generated from the model are said to be concentrated or dispersed depending on whether the parameter α equals 0.5 or 2.

We measure kurtosis by Koziol's statistic (Koziol 1989), that is the squared Euclidean norm of the fourth standardized moment:

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left[(x_i - m)^T S^{-1} (x_j - m) \right]^4, \quad (16)$$

where m is the sample mean, S is the sample variance (which is assumed to be non-singular) and x_h^T is the h -th row of the $n \times d$ data matrix X , for $h = 1, \dots, d$. We favored Koziol's kurtosis over the more popular Mardia's kurtosis because the former, unlike the latter, depends on all fourth-order standardized moments. Other properties of Koziol's kurtosis, including its advantages over Mardia's kurtosis, are discussed in Loperfido (2020a).

The Koziol's kurtosis of either a random vector or a sampled dataset increases with its dimension: the Koziol's kurtosis of a d -dimensional normal random vector is $3d(d + 2)$. In order to compare the nonnormality of datasets with different numbers of variables we use the relative difference between their Koziol's kurtoses and the Koziol's kurtoses of normal random vectors with the same dimensions. For example, if the dataset at hand contains three variables and its Koziol's kurtosis is 54, its relative

difference to the Koziol's kurtosis of a three-dimensional normal random vector is $(54 - 45)/45 = 0.2$.

We applied the proposed method to each sample to obtain $d - 1$, $d/2$, \sqrt{d} projections. Ideally, a method purported to obtain mesokurtic projections should detect all the $d - 1$ jointly normal projections. When this does not happen, it is desirable to have some guidance about the achievable number of jointly normal linear projections, as for example $d/2$ or \sqrt{d} . For both the original and the projected data we computed the absolute relative difference between the observed Koziol measure of kurtosis and its expectation under the normality assumption. We would like this ratio to be much lower for the projected data than for the original ones. For each sample we also computed the p -value of Koziol kurtosis, using its asymptotically normal approximation. We would like it to be, on average and for the projected data, well above the commonly used threshold value of 0.05.

Tables 2 and 3 contain the results of the two simulation studies. Columns **A** and **B** report the integer part of the average absolute relative difference, multiplied by 100, between the Koziol measure of kurtosis and the value which it is expected to attain in the normal case, for the original and the projected variables, respectively. Column **C** reports the integer part of the average p values, multiplied by 100, corresponding to the Koziol measure of multivariate kurtosis computed for the projected variables. The subtables denoted with $d - 1$, $d/2$ and \sqrt{d} report the simulation's results corresponding to a number of projections equal to the number of variables minus one, half the number of variables and the integer part of the square root of the number of variables, respectively.

The subtables of Table 2 with headers "Leptokurtic" and "Platykurtic" report the results for samples simulated from mixtures where the weight of the standard normal component are 0.1 and 0.5, respectively. The results of the first simulation, related to homoscedastic normal mixtures, may be summarized as follows. The ratio statistic is always much smaller for the projected data than for the original ones. It increases (decreases) with the number of variables when $d - 1$ (\sqrt{d}) mesokurtic projections are sought. When $d/2$ mesokurtic projections are sought, the ratio statistic first decreases and then increases. Normality of the projections, as measured by Koziol's kurtosis, is better achieved when their number is small compared to the number of variables, and the latter is small, either.

The subtables of Table 3 with headers "Concentrated" and "Dispersed" report the results for samples simulated from mixtures where the outliers generated from the model are more concentrated and more dispersed, respectively. The proposed method appears to perform better in the second simulation study. The differences between the ratio statistics for the original and the projected variables are always greater in the second simulation study. The performances of the proposed method are worst and best when $d - 1$ and \sqrt{d} mesokurtic projections are sought, and when the number of original variables is small. The performances of mesokurtic transformations appear to be similar in the presence of concentrated and dispersed outliers. Similar remarks hold when considering the p values of Koziol's kurtosis tests for the projected variables.

We conclude that the proposed method succeeds in either removing or alleviating kurtosis. It is particularly successful when the number of projections is moderate with

Table 2 Results of the second simulation study

Projections	Variables	Concentrated			Dispersed		
		A	B	C	A	B	C
$d - 1$	8	183	24	2	266	27	3
	12	140	52	0	235	56	0
	16	150	87	0	243	96	0
	20	175	132	0	276	144	0
$d/2$	8	178	7	26	268	7	27
	12	146	5	29	231	8	21
	16	147	13	4	236	18	2
	20	182	30	0	275	34	0
\sqrt{d}	8	173	13	26	272	13	26
	12	146	10	25	240	9	27
	16	146	5	31	242	5	31
	20	178	9	24	273	8	25

Columns **A** and **B** report the integer part of the average absolute relative difference, multiplied by 100, between the Koziol measure of kurtosis and the value which it is expected to attain in the normal case, for the original and the projected variables, respectively. Column **C** reports the integer part of the average p values, multiplied by 100, corresponding to the Koziol measure of multivariate kurtosis computed for the projected variables

Table 3 Results of the third simulation study

Projections	Variables	Concentrated			Dispersed		
		A	B	C	A	B	C
$d - 1$	8	183	24	2	266	27	3
	12	140	52	0	235	56	0
	16	150	87	0	243	96	0
	20	175	132	0	276	144	0
$d/2$	8	178	7	26	268	7	27
	12	146	5	29	231	8	21
	16	147	13	4	236	18	2
	20	182	30	0	275	34	0
\sqrt{d}	8	173	13	26	272	13	26
	12	146	10	25	240	9	27
	16	146	5	31	242	5	31
	20	178	9	24	273	8	25

Columns **A** and **B** report the integer part of the average absolute relative difference, multiplied by 100, between the Koziol measure of kurtosis and the value which it is expected to attain in the normal case, for the original and the projected variables, respectively. Column **C** reports the integer part of the average p values, multiplied by 100, corresponding to the Koziol measure of multivariate kurtosis computed for the projected variables

respect to the number of variables, and the latter is moderate, too. As a major drawback, it may not detect all mesokurtic projections, especially in high dimensions.

4 AIS data

In this section we illustrate the use of mesokurtic projections in projection pursuit, as described in Sect. 2, with the data collected from the Australian Institute of Sport by Telford and Cunningham (1991), also known as AIS dataset. It contains eleven blood and body measurements from 202 athletes of both genders competing in different sport events: red blood cell count, white blood cell count, hematocrit, hemoglobin concentration, plasma ferritins, body mass index, sum of skin folds, body fat, lean body mass, height and weight. Here, we focus on the 23 female athletes playing netball, so that the number of units is about twice the number of variables.

Nine variables in the dataset are mildly platykurtic, while the remaining two variables are mildly leptokurtic: their kurtoses range from 2.079 to 4.681, with just two of them being greater than 3. We assess the joint normality of the eleven variables by means of Koziol's kurtosis tests for multivariate normality, whose p value is virtually zero. The maximum number of jointly mesokurtic projection detected by the method described in Sect. 2 is three (the p value of Koziol's kurtosis test for multivariate normality is greater than 0.17), which is much smaller than the number of original variables. Further dimension reduction might be achieved by looking for two sets of mutually orthogonal projections, one nearly mesokurtic and the other containing the information on the interesting structure.

Mild nonnormality might be due to the presence of outliers, often detectable by means of visual inspection. The boxplots in Fig. 1 suggest the presence of several

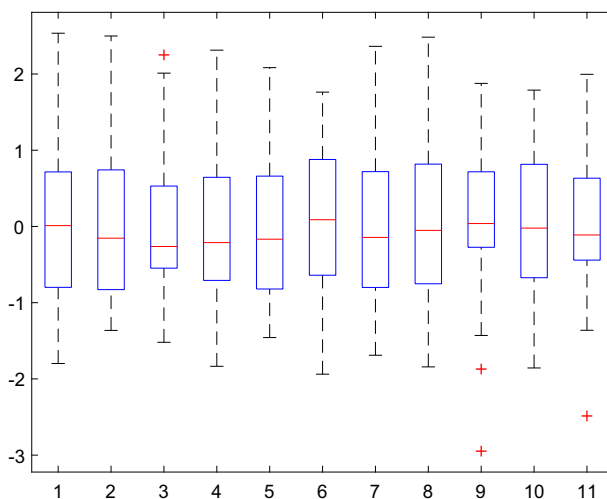


Fig. 1 Boxplots of the standardized variables in the AIS dataset recorded from female netball players

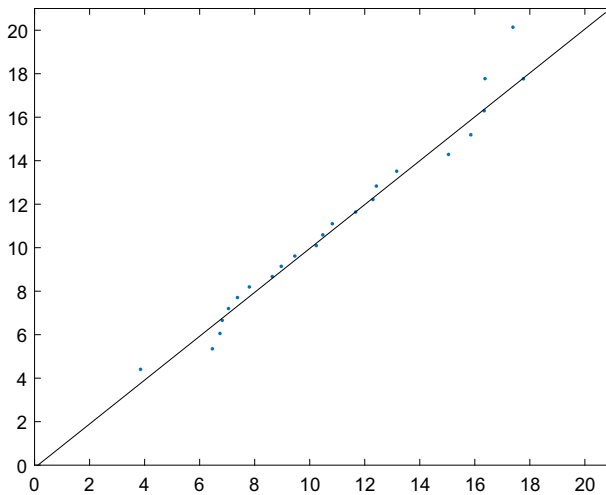


Fig. 2 Healy's plot of the eleven variables in the AIS dataset recorded from female netball players

potential outliers, represented by the asterisks above and below the whiskers. The third, ninth and eleventh variables correspond to the third, ninth and eleventh boxplots starting from the left of Fig. 1. They hint that the potential outliers might be the first, the second and the fifteenth units. On the other hand, the Healy's plot in Fig. 1 clearly suggests that the twenty-first unit is the most likely to be an outlier, since it has the highest Mahalanobis distance from the mean and it is represented by the point farthest away from the bisector line. We conclude that visual inspection of the original variables gives contradictory results about the presence of outliers (Fig. 2).

We addressed outlier detection by means of kurtosis maximization, as suggested by several authors (Peña and Prieto 2001a; Hou and Wentzell 2014; Loperfido 2020b). The projection with maximal kurtosis is represented by the last boxplot from the left in Fig. 3, which clearly shows the presence of a single outlier, that is the twenty-first observation. Then we performed kurtosis maximization on the projections orthogonal to those with small absolute fourth cumulants, as computed with the method described in Sect. 2. The first, second, third, fourth and fifth boxplots from the left of Fig. 3 represent the projections with maximal kurtoses orthogonal to 10, 9, 8, 7 and 6 nearly mesokurtic projections. All boxplots clearly suggest that the twenty-first unit is the most likely to be an outlier and all but the first two boxplots on the left suggest that it is the only potential outlier. All projections maximizing kurtosis are strongly correlated with each other: the correlations between them are never smaller than 0.93. This statistical analysis is exploratory in nature and more sophisticated, inferential methods are needed to assess the presence of outliers. This section is just aimed at showing how mesokurtic projections might be helpful for dimension reduction while retaining the relevant data structure to be uncovered by kurtosis-based projection pursuit.

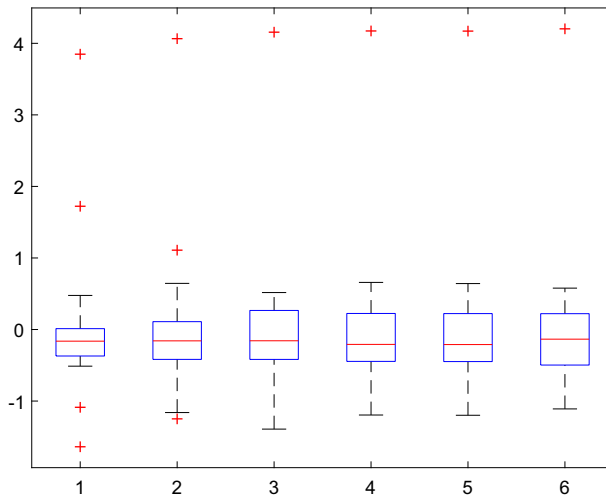


Fig. 3 Boxplots of the standardized projections with maximal kurtoses of the variables in the AIS dataset recorded from female netball players

5 Randu data

In simulation studies, a normal random sample is often obtained by applying the normal quantile transformation to data generated from a uniform distribution. The method might lead to unsatisfactory results, even if visual and formal procedures support the hypothesis of the randomly generated data being uniform. We illustrate the problem with the RANDU data and show how mesokurtic projections might be helpful in alleviating it. The same dataset provides a good case of mesokurtic projections virtually identical to those with extreme kurtosis.

The RANDU dataset contains 400 cases and 3 variables. Each datum was generated by the RANDU multiplicative congruential scheme $x_{n+1} = (2^{16} + 3)x_n \pmod{2^{31}}$. RANDU data satisfy the constraint $x_n - 6x_{n+1} + x_{n+2} \equiv 0 \pmod{2^{31}}$, so that all triplets (x_n, x_{n+1}, x_{n+2}) lie on 15 parallel lines through the unit cube. This structure might be detected by means of appropriate rotations. Despite that, each variable in the dataset appear to be generated from a uniform distribution in the interval $[0, 1]$. For example, their means, standard deviations, skewnesses and kurtoses (reported in Table 4) are very close to their expected values 0.5, 0.2889, 0 and 1.8, respectively. These features made the RANDU dataset a perfect case for projection pursuit (see, for example, Huber 1985).

Table 4 Descriptive statistics of the RANDU data

	First	Second	Third
Mean	0.5264	0.4861	0.4810
Deviation	0.2847	0.2934	0.2787
Skewness	-0.1028	0.0045	0.0867
Kurtosis	1.8665	1.7626	1.8747

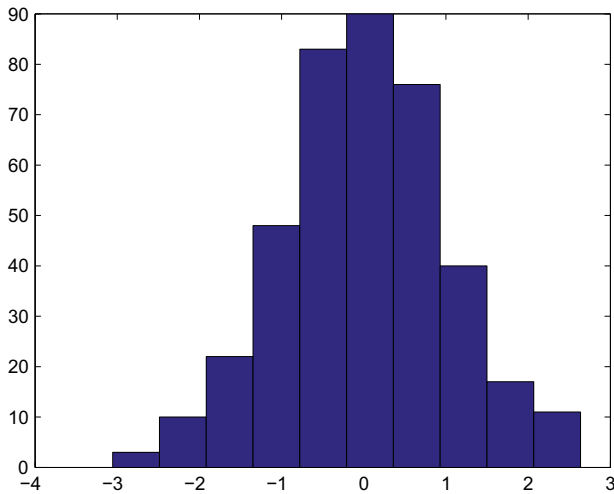


Fig. 4 Histogram of the first projection

Since each variable appears to be uniformly distributed, it is natural to transform them to normality by applying the normal quantile transformation. Unfortunately, this method fails to achieve normality, as it is apparent from their univariate kurtoses: 3.626, 3.449, 3.795. The p values of the kurtosis-based tests for normality of the three transformed variables are smaller than 0.05. The Box-Cox transformation gives even worse results. It leads to variables which are significantly platykurtic: their kurtoses are 2.056, 2.026, 2.117. The p values of both the skewness and the kurtosis tests for normality of the three transformed variables are smaller than 0.05.

Finally, we applied the proposed method to obtain two data projections. The first one is virtually normal (its skewness and kurtosis are -0.0359 and 2.9815), while the second one is only slightly so (its skewness and kurtosis are -0.1784 and 2.5485). The p values of the kurtosis tests for testing the normality of the first and second projections are 0.94 and 0.0654, respectively. The two projections are not jointly normal: the p -value of the Koziol test for multivariate normality is 0.0067. The histograms (Figs. 4 and 5) and the scatterplot of the two projections (Fig. 6) are consistent with these findings.

Interestingly, the first projection coincides, up to location and scale changes, with the projection achieving maximal kurtosis. Hence the RANDU dataset provides a good example of kurtosis-based projection pursuit as a tool for recovering normality, rather than detecting nonnormal features.

The empirical findings in this section might be summarized as follows. We applied the proposed method to a well-known, three-dimensional, platykurtic dataset, obtaining a normal univariate projection and a slightly nonnormal bivariate projection. We therefore succeeded in achieving univariate normality and in alleviating bivariate nonnormality. Commonly used, componentwise, nonlinear methods failed in both recovering either univariate or bivariate normality. We conclude that the RANDU dataset supports the use of mesokurtic projections when a randomly generated normal sample is sought.

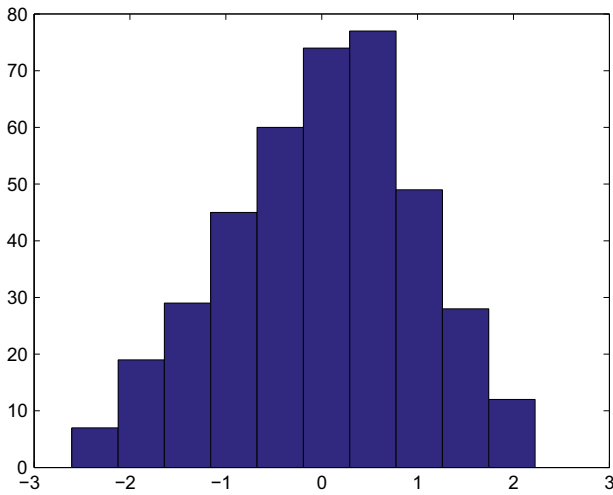


Fig. 5 Histogram of the second projection

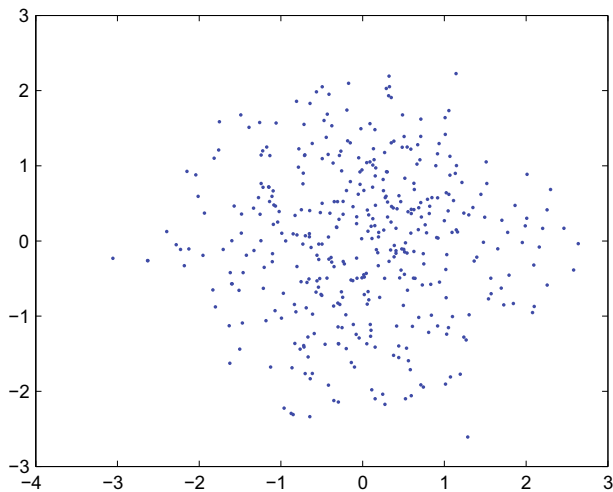


Fig. 6 Scatterplot of the two projections

6 Conclusions

The paper describes a method for obtaining data projection with null or negligible fourth cumulants. It uses several kurtosis matrices to establish either sufficient or necessary conditions for the existence of mesokurtic projections. It also relates mesokurtic projections to spectra of kurtosis matrices, thus easing inferential tasks. Mesokurtic projections address the problems arising in the application of kurtosis-based projection pursuit to datasets where the number of variables is only slightly smaller than the

number of units. The proposed method has been implemented in the R package Kurt (Franceschini and Loperfido 2020).

Data preprocessing is a default practice in projection pursuit (see, for example, Jones and Sibson (1987)). Laa and Cook (2020) proposed to remove skewness from the data when preprocessing them, which can be done by means of linear projections (Loperfido 2014; Franceschini and Loperfido 2019). We recommend the use of mesokurtic projections after linear projections to symmetry when preprocessing the data. Centering, sphering, symmetrizing and mesokurtic projections conveniently address the first, second, third and fourth sample cumulants.

Mesokurtic projections might also be used in independent component analysis, a multivariate statistical method aimed at recovering independent random variables from the observed data which are their one-to-one affine transformations. The model is identifiable if at most one independent variable is normal (Miettinen et al. 2015). The nonnormal independent variables are commonly assumed to be leptokurtic, so that fourth cumulants may be used to recover them (Girolami and Fyfe 1996). The same assumption implies that the model is identifiable if there is at most one linear combination of the observed variables which is mesokurtic. Hence model identification is related to the rank of the matrix which characterizes mesokurtic projections.

Mesokurtic projections might also preprocess the data before using a statistical method whose performance heavily depends on the sampled distribution having null or at least negligible cumulants. Examples include inference on covariance matrices (Yanagihara et al. 2005), linear discriminant analysis (Arealillo and Navarro 2012), multivariate linear regression (Yanagihara 2007) and variogram estimation (Rezvan-dehy and Deutsch 2018). Mesokurtic projections might provide a viable alternative to other statistical methods which are commonly used to address these problems: flexible modelling, nonparametric inference and nonlinear transformations.

Despite its theoretical properties and practical usefulness, the proposed method suffers from several drawbacks. In the first place, mesokurtic projections do not exist for some distributions, as for example scale mixtures of multivariate normal distributions. In the second place, the proposed method may not detect all mesokurtic projections, especially in high dimensional cases. In the third place, mesokurtic projections are not appropriate when the sampled distribution is nonnormal, but all its fourth cumulants equal zero. In the fourth place, mesokurtic projections imply some information loss, since the original data are projected onto a lower dimensional space.

A natural question to ask is whether the proposed method might be extended to datasets with more variables than units. Both Propositions 1, 2 rely on the standardized random variable $z = \Sigma^{-1/2} (x - \mu)$, which is defined only if the number of units is greater than the number of variables. In the opposite case, a common solution is the replacement of the covariance matrix Σ with the positive definite convex linear combination $\Gamma = \lambda \Sigma + (1 - \lambda) \Omega$ of Σ itself with a full rank matrix Ω of the same size, where $0 < \lambda < 1$ (see, for example, Hastie et al. 2008). This approach already appeared in the literature on projection pursuit (Lee and Cook 2010; Hui and Lindsay 2010). We are currently investigating the generalizations of Propositions 1, 2 to situations where $z = \Sigma^{-1/2} (x - \mu)$ is replaced by $y = \Gamma^{-1/2} (x - \mu)$, where $\Gamma^{-1/2}$ is the positive definite symmetric square root of the inverse Γ^{-1} of Γ .

Acknowledgements The author would like to thank the Coordinating Editor and two anonymous Reviewers for their insightful comments, which greatly helped to improve the quality of this paper.

Funding Open access funding provided by Università degli Studi di Urbino Carlo Bo within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

7 Appendix

We recall some fundamental properties of the Kronecker product which we shall use repeatedly in the following proofs (see, for example, Rao and Rao 1998, pages 194–201): (P1) the Kronecker product is associative: $(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C$; (P2) if matrices A, B, C and D are of appropriate size, then $(A \otimes B)(C \otimes D) = AC \otimes BD$; (P3) the transpose of a Kronecker product of two matrices is the Kronecker product of the transposed matrices: $(A \otimes B)^T = A^T \otimes B^T$; (P4) if a and b are two vectors, then ab^T , $a \otimes b^T$ and $b^T \otimes a$ denote the same matrix; (P5) $\text{tr}(A^T B) = \text{vec}^T(B) \text{vec}(A)$ for any two $m \times n$ matrices A and B ; (P6) $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$, when $A \in \mathbb{R}^p \times \mathbb{R}^q$, $B \in \mathbb{R}^q \times \mathbb{R}^r$, $C \in \mathbb{R}^r \times \mathbb{R}^s$. Finally, we recall some properties of the commutation matrix $C_{p,q} \in \mathbb{R}^{pq} \times \mathbb{R}^{pq}$ (Kollo and von Rosen 2005, page 79): (P7) $\text{vec}(M^T) = C_{p,q} \text{vec}(M)$ for any $p \times q$ matrix M ; (P8) $C_{p,q}^T = C_{q,p}$; (P9) $C_{d,d} = \sum_{i,j} (e_i e_j^T) \otimes (e_j e_i^T)$, where e_i is the i -th column of I_d (Kollo and von Rosen 2005, page 82); (P10) $C_{d,d} = C_{d,d}^{-1}$.

Proof of Proposition 1 Let a and b be two eigenvectors of $M_{4,z}$ corresponding to different, positive eigenvalues. Both of them might be represented as vectorized, symmetric matrices (see, for example, Loperfido 2017): $a = \text{vec}(A)$, $b = \text{vec}(B)$, $A = A^T$, $B = B^T$. Since a and b correspond to different eigenvalues of the symmetric matrix $M_{4,z}$, they are mutually orthogonal: $b^T a = \text{vec}^T(B) \text{vec}(A) = 0 = \text{tr}(AB)$, the last equality following from P5 and symmetry of the two matrices. Suppose now that both A and B were positive definite, symmetric matrices. Then all eigenvalues of AB would be positive (Ortega 1987, page 232). This would in turn imply that $\text{tr}(AB)$ were positive, thus leading to a contradiction. We conclude that A and B cannot be both positive definite. Equivalently, we showed that $M_{4,z}$ has at most one eigenvector which might be represented as a vectorized, definite symmetric matrix. We denote this eigenvector by $v = \text{vec}(V)$, where V is a vectorized, symmetric and positive definite $d \times d$ matrix.

Assume now that v is not an eigenvector associated to the dominant eigenvalue λ of $M_{4,z}$, and denote with u such eigenvector. Using an argument similar to the previous one, we can show that $u^T v = 0$, which would be possible if and only if u were

a vectorized, symmetric and indefinite matrix. However, u is a vectorized, positive semidefinite symmetric matrix (Loperfido 2017), thus leading to a contradiction. We have therefore proved that v must be an eigenvector associated with the dominant eigenvalue of $M_{4,z}$.

We now prove by contradiction that λ is simple. Assume that λ is not a simple eigenvalue of $M_{4,z}$, so that there is another eigenvector $m = \text{vec}(M)$ associated to it, where M is a vectorized and symmetric $d \times d$ matrix of unit norm. Without loss of generality we can assume that v and m are mutually orthogonal: $v^T m = \text{tr}(VM) = 0$. The last equality, together with positive definiteness of V , imply that M is indefinite. The inequality

$$E \left[\left(z^T M z \right)^2 \right] \geq E \left[\left(z^T S z \right)^2 \right], \quad (17)$$

where S is a symmetric, $d \times d$ matrix of unit norm, follows from P6 and from m being an eigenvector associated to the dominant eigenvalue of $M_{4,z}$. Since V is symmetric, it might be decomposed into $\Omega \Delta \Omega^T$, where the columns $\omega_1, \dots, \omega_d$ of $\Omega \in \mathbb{R}^d \times \mathbb{R}^d$ are the normalized eigenvectors of M and Δ is a diagonal matrix whose i -th diagonal entry is the i -th eigenvalue of M : $\Delta = \text{diag}(\delta_1, \dots, \delta_d)$. It follows that

$$E \left[\left(z^T M z \right)^2 \right] = E \left[\left(\lambda_1 Y_1^2 + \dots + \lambda_d Y_d^2 \right)^2 \right], \quad (18)$$

where $y = (Y_1, \dots, Y_d)^T = \Omega^T z$. By assumption, the covariance matrix of z is nonsingular, so that $E(Y_i^2) > 0$ for $i = 1, \dots, d$. Suppose now that M is indefinite, so that the smallest k eigenvalues of M are negative, for some integer k between 1 and $d - 1$, implying

$$E \left[\left(\delta_1 Y_1^2 + \dots + \delta_d Y_d^2 \right)^2 \right] < E \left[\left(\delta_1 Y_1^2 + \dots + \delta_{k-1} Y_{k-1}^2 - \delta_k Y_k^2 \dots + \delta_d Y_d^2 \right)^2 \right]. \quad (19)$$

Equivalently, the inequality

$$E \left[\left(z^T M z \right)^2 \right] < E \left[\left(z^T Q z \right)^2 \right] \quad (20)$$

would hold true, for $Q = \Omega \Delta_+^T \Omega$ and $\Delta_+ = \text{diag}(\delta_1, \dots, \delta_{k-1}, -\delta_k, \dots, -\delta_d)$. Also, Q would be positive semidefinite with the same norm of M : $\|M\|^2 = \delta_1^2 + \dots + \delta_d^2 = \|Q\|^2$. As a direct consequence, $m = \text{vec}(M)$ would not be an eigenvector associated with the dominant eigenvalue of $M_{4,z}$, thus leading to a contradiction. We conclude that the dominant eigenvalue λ of $M_{4,z}$ is simple.

We now prove that no mesokurtic projection of z exists when all positive eigenvalues of $M_{4,z}$ are greater than three. Let $w = \text{vec}(W)$ be the eigenvector associated with the smallest positive eigenvalue $\eta > 3$ of $M_{4,x}$, where W is a vectorized, symmetric $d \times d$ matrix of unit norm. Also, let c be a d -dimensional real vector of unit length, so that $c^T z$ is a standardized random variable: $E(c^T z) = 0$ and $E(c^T z)^2 = 1$. The

inequality

$$E \left[\left(z^T W z \right)^2 \right] \leq E \left[\left(z^T S z \right)^2 \right], \quad (21)$$

where S is a symmetric, $d \times d$ matrix of unit norm, follows from $P6$ and from m being an eigenvector associated to the smallest positive eigenvalue of $M_{4,z}$. The kurtosis of $c^T z$ is

$$E \left[\left(z^T c \right)^4 \right] = E \left[\left(z^T c c^T z \right)^2 \right] \quad (22)$$

and the $d \times d$ matrix $c c^T$ is real, symmetric and of unit norm. Hence the kurtosis of $c^T z$ is always greater than η , which in turn is greater than three by assumption. The proof that no mesokurtic projection of z exists when all positive eigenvalues of $M_{4,z}$ are smaller than three is very similar to the previous one and is therefore omitted. \square

Proof of Proposition 2 Repeated use of $P4$ leads to the identities

$$z \otimes z^T \otimes z^T \otimes z^T = z z^T \otimes z^T \otimes z^T = z z^T \otimes \text{vec}^T \left(z z^T \right) = z^T \otimes z^T \otimes z z^T. \quad (23)$$

Now apply $P2$, $P6$ and $P4$ to obtain

$$\begin{aligned} \left(z z^T \otimes z^T \otimes z^T \right) [I_d \otimes \text{vec} (I_d)] &= \left(z z^T I_d \right) \otimes \left(z^T \otimes z^T \right) \text{vec} (I_d) \\ &= z z^T \otimes \left(z^T z \right) = z^T z z z^T. \end{aligned} \quad (24)$$

Similarly, apply $P2$ and $P8$ and $P7$ to obtain

$$\begin{aligned} \left[z z^T \otimes \text{vec}^T \left(z z^T \right) \right] (I_d \otimes C_{d,d}) &= z z^T I_d \otimes \text{vec}^T \left(z z^T \right) C_{d,d} \\ &= z z^T \otimes \left[C_{d,d} \text{vec} \left(z z^T \right) \right]^T = z z^T \otimes \text{vec}^T \left(z z^T \right). \end{aligned} \quad (25)$$

The definitions of $\text{cok} (z) = E \left(z \otimes z^T \otimes z^T \otimes z^T \right)$ and $K_z = E \left(z^T z z z^T \right)$, together with linear properties of the expected value, leads to

$$K_z = \text{cok} (z) [I_d \otimes \text{vec} (I_d)] = \left[I_d \otimes \text{vec}^T (I_d) \right] \text{cok}^T (z) = \text{cok} (z) [\text{vec} (I_d) \otimes I_d]. \quad (26)$$

We now prove the identity $[\text{vec}^T (I_d) \otimes I_d] (I_d \otimes C_{d,d}) [\text{vec} (I_d) \otimes I_d] = I_d$. First apply properties $P9$, $P2$ and $P6$ to obtain

$$\begin{aligned} [\text{vec}^T (I_d) \otimes I_d] (I_d \otimes C_{d,d}) &= \sum_{i,j=1}^d \text{vec}^T (I_d) \left[I_d \otimes \left(e_i e_j^T \right) \right] \otimes I_d \left(e_j e_i^T \right) \\ &= \sum_{i,j=1}^d \left[\left[I_d \otimes \left(e_j e_i^T \right) \right] \text{vec} (I_d) \right]^T \otimes \left(e_j e_i^T \right) = \sum_{i,j=1}^d \text{vec}^T \left(e_j e_i^T \right) \otimes \left(e_j e_i^T \right). \end{aligned}$$

The following identities are direct consequences of $P2$ and $P5$, as well as ordinary properties of the matrix's trace:

$$\begin{aligned}
 & \left[\text{vec}^T(I_d) \otimes I_d \right] (I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \\
 &= \sum_{i,j=1}^d \text{vec}^T(e_j e_i^T) \otimes (e_j e_i^T) [\text{vec}(I_d) \otimes I_d] \\
 &= \sum_{i,j=1}^d \left[\text{vec}^T(e_j e_i^T) \text{vec}(I_d) \right] \otimes (e_j e_i^T) I_d = \sum_{i,j=1}^d \text{tr}(I_d e_j e_i^T) \otimes (e_j e_i^T) \\
 &= \sum_{i,j=1}^d e_i^T e_j \otimes (e_j e_i^T).
 \end{aligned}$$

By definition, e_i is the d -dimensional vector whose i -th component is one while the others are zero. Hence $e_i^T e_j$ is either one or zero depending on whether $i = j$ or $i \neq j$. We can then write

$$\left[\text{vec}^T(I_d) \otimes I_d \right] (I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] = \sum_{i=1}^d e_i e_i^T = I_d. \quad (27)$$

The last identity follows from $e_i e_i^T$ being the matrix whose only nonnull element is the i -th diagonal element, which equals one. Similarly, we have $[\text{vec}^T(I_d) \otimes I_d] (I_d \otimes C_{d,d}) [I_d \otimes \text{vec}(I_d)] = I_d$. This identity, together with $P2$ and $P10$, implies

$$\begin{aligned}
 & \left[I_d \otimes \text{vec}^T(I_d) \right] [\text{vec}(I_d) \otimes I_d] \\
 &= \left[I_d \otimes \text{vec}^T(I_d) \right] (I_d \otimes C_{d,d}) (I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \\
 &= \left[I_d \otimes \text{vec}^T(I_d) \right] (I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \\
 &= \left\{ \left[\text{vec}^T(I_d) \otimes I_d \right] (I_d \otimes C_{d,d}) [I_d \otimes \text{vec}(I_d)] \right\}^T = I_d.
 \end{aligned}$$

Apply now $P2$ and $P5$: $[I_d \otimes \text{vec}^T(I_d)] [I_d \otimes \text{vec}(I_d)] = dI_d$. Apply properties $P2$ and $P10$ to obtain

$$\begin{aligned}
 & \left[\text{vec}^T(I_d) \otimes I_d \right] (I_d^3 + I_d \otimes C_{d,d}) (I_d^3 + I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \\
 &= 2 \left[\text{vec}^T(I_d) \otimes I_d \right] (I_d^3 + I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \\
 &= 2 \left[\text{vec}^T(I_d) \otimes I_d \right] I_d^3 [\text{vec}(I_d) \otimes I_d] \\
 &\quad + 2 \left[\text{vec}^T(I_d) \otimes I_d \right] (I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] = 2dI_d + 2I_d.
 \end{aligned}$$

All fourth-order cumulants $\kappa_{ijhk} = \partial^4 \log E [\exp (u^T x)] / \partial t_i \partial t_j \partial t_h \partial t_k$, where $u = \sqrt{-1}$ and $t^T = (t_1, \dots, t_d)$, might be conveniently arranged into the $d \times d^3$ matrix $F = \partial^4 \log E [\exp (u^T x)] / \partial t \partial^3 t^T$, which admits the representation

$$F = \text{cok}(z) - I_d \otimes \text{vec}^T(I_d) - [\text{vec}^T(I_d) \otimes I_d] (I_{d^3} + I_d \otimes C_{d,d}), \quad (28)$$

(Kollo and von Rosen 2005, page 187). The above identities, together with P3, helps in simplifying the product of F and its transpose F^T :

$$\begin{aligned} FF^T &= \text{cok}(z) \left\{ \text{cok}^T(z) - I_d \otimes \text{vec}(I_d) - (I_{d^3} + I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \right\} \\ &\quad - [I_d \otimes \text{vec}^T(I_d)] \left\{ \text{cok}^T(z) - I_d \otimes \text{vec}(I_d) - (I_{d^3} + I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \right\} \\ &\quad - [\text{vec}^T(I_d) \otimes I_d] (I_{d^3} + I_d \otimes C_{d,d}) \left\{ \text{cok}^T(z) - I_d \otimes \text{vec}(I_d) - (I_{d^3} + I_d \otimes C_{d,d}) [\text{vec}(I_d) \otimes I_d] \right\} \\ &= [\text{cok}(z) \text{cok}^T(z) - 3K] - (K - 2I_d - dI_d) - (2K - 2dI_d - 4I_d) \\ &= \text{cok}(z) \text{cok}^T(z) - 6K + 3(d+2)I_d. \end{aligned}$$

Without loss of generality we shall assume that the columns of B^T are mutually orthogonal, normalized vectors, so that $BB^T = I_h$. Let

$$G = \frac{\partial^4 \log E [\exp (u^T y)]}{\partial t \partial^3 t^T}, \quad (29)$$

where $y = Bz$. By ordinary properties of cumulants, the assumption $BB^T = I_h$ and P2, we have

$$GG^T = BF (B^T \otimes B^T \otimes B^T) (B \otimes B \otimes B) F^T B^T = BFF^T B^T. \quad (30)$$

By definition, the columns of B^T span the null space of $\text{cok}(z) \text{cok}^T(z) - 6K + 3(d+2)I_d$, so that GG^T is a null matrix, which implies that all fourth-order cumulants of Bz equal zero, and this completes the proof. \square

Proof of Proposition 3 Let u and v be a d -dimensional and a h -dimensional random vectors whose joint distribution is

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N_{d+h} \left[\begin{pmatrix} \xi \\ -\eta \end{pmatrix}, \begin{pmatrix} \Omega & \Omega \Psi^T \\ \Psi \Omega & \Delta + \Psi \Omega \Psi^T \end{pmatrix} \right].$$

Consider now the decomposition $u = u - Cv + Cv$, where

$$C = \Omega \Psi^T (\Delta + \Psi \Omega \Psi^T)^{-1} \in \mathbb{R}^d \times \mathbb{R}^h.$$

Basic properties of normal random vectors imply that $u - Cv$ and v are independent, normal random vectors. Gonzalez-Farias et al. (2003) showed that x and $u|v > 0$ are identically distributed, so that we can write

$$x \sim u - Cv + Cv_+, \text{ where } v_+ = v|v > 0.$$

The exkurtosis of the sum of independent random vectors is the sum of their exkurtoses:

$$exk(x) = exk(u - Cv) + exk(Cv_+).$$

The identity $exk(x) = exk(Cv_+)$ follows from $u - Cv$ being a normally distributed random vector. Apply now multilinear properties of the fourth-order cumulants (Loperfido 2020c):

$$exk(x) = C exk(v_+) (C^T \otimes C^T \otimes C^T).$$

By definition, C is a $d \times h$ matrix and by assumption $d > h$, so that there exists a full rank $(d - h) \times d$ matrix B such that BC is a null matrix. As a first implication, the rows of B belongs to the null space of the transposed exkurtosis:

$$exk^T(x) B^T = (C \otimes C \otimes C) exk^T(v_+) C^T B^T$$

is a $d^3 \times (d - h)$ null matrix. As a second implication, Bx is a normally distributed random vector:

$$Bx \sim Bu - BCv + BCv_+ = Bu.$$

The proof is then complete. \square

References

- Adcock C (2021) Copulaesque versions of the skew-normal and skew-student distributions. *Symmetry* 13:815
- Alashwali F, Kent JT (2016) The use of a common location measure in the invariant coordinate selection and projection pursuit. *J Multivar Anal* 152:145–161
- Arevalillo JM, Navarro H (2012) A study of the effect of kurtosis on discriminant analysis under elliptical populations. *J Multivar Anal* 107:53–63
- Arnold B, Castillo E, Sarabia J (2001) Conditionally specified distributions: an introduction. *Stat Sci* 16:249–274
- Bickel P.J, Kur G, Nadler B (2018) Projection pursuit in high dimensions. *Proc Natl Acad Sci USA* 115(37):9151–9156
- Blough DK (1989) Multivariate symmetry via projection pursuit. *Ann Inst Stat Math* 41:461–475
- Box GEP, Cox DR (1964) An analysis of transformations. *J R Stat Soc B* 26:211–252
- Cardoso JF (1989) Source separation using higher order moments. In: *Proc. ICASSP'89*, pp 2109–2112
- Causinus H, Ruiz-Gazen A (2009) Exploratory projection pursuit. In: Govaert G (ed) *Data analysis*. Wiley, Amsterdam, pp 76–92
- Christiansen M, Loperfido N (2014) Improved approximation of the sum of random vectors by the skew-normal distribution. *J Appl Probab* 51:466–482
- Diaconis P, Freedman D (1984) Asymptotics of graphical projection pursuit. *Ann Stat* 12:793–815
- Flecher C, Naveau P, Allard D (2009) Estimating the closed skew-normal distribution parameters using weighted moments. *Stat Probab Lett* 79:1977–1984

- Franceschini C, Loperfido N (2018) An algorithm for finding projections with extreme kurtosis. In: Perna C, Pratesi M, Ruiz-Gazen A (eds), *Studies in theoretical and applied statistics: SIS2016-48th meeting of the Italian statistical society*, Salerno 8–10 June 2016. Springer
- Franceschini C, Loperfido N (2019) MaxSkew and MultiSkew, Two R packages for detecting, Measuring and removing multivariate skewness. *Symmetry* 11(8):970
- Franceschini C, Loperfido N (2020) Kurt: performs kurtosis-based statistical analysis. R package version 1.0, <https://CRAN.R-project.org/package=Kurt>
- Galeano P, Peña D, Tsay RS (2006) Outlier detection in multivariate time series by projection pursuit. *J. Am. Stat. Assoc.* 101:654–669
- Genton MG, He L, Liu X (2001) Moments of skew-normal random vectors and their quadratic forms. *Stat. Prob. Lett.* 51:319–325
- Girolami M, Fyfe C (1996) Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms. In: *advances in neural information processing systems workshop* p. 9
- Gonzalez-Farias G, Dominguez-Molina JA, Gupta AK (2003) Additive properties of skew-normal random vectors. *J Stat Plan Inference* 126:521–534
- Hastie T, Tibshirani R, Friedman J (2008) *The elements of statistical learning: data mining, inference, and prediction*, 2nd edn. Springer, New York, NY
- He L, Chen J, Qi Y (2019) Event-based state estimation: optimal algorithm with generalized closed skew normal distribution. *IEEE Trans Autom Control* 64:321–328
- Hou S, Wentzell PD (2014) Re-centered kurtosis as a projection pursuit index for multivariate data analysis. *J Chemom* 28:370–384
- Huber PJ (1985) Projection pursuit. *Ann Stat* 13:435–475
- Hui G, Lindsay BG (2010) Projection pursuit via white noise matrices (with discussion). *Sankhya B* 72:123–153
- Jondeau E, Rockinger M (2006) Optimal portfolio allocation under higher moments. *Eur Financ Manag* 12:29–55
- Jones MC, Sibson R (1987) What is projection pursuit? (with discussion). *J R Stat Soc Ser A* 150:1–38
- Kim H-M (2005) Moments of variogram estimator for a generalized skew-t distribution. *J Korean Stat Soc* 34:109–123
- Kollo T, von Rosen D (2005) *Advanced multivariate statistics with matrices*. Springer, Dordrecht
- Kozioł JA (1989) A note on measures of multivariate kurtosis. *Biom J* 31:619–624
- Laa U, Cook D (2020) Using tours to visually investigate properties of new projection pursuit indexes with application to problems in physics. *Comput Stat*. <https://doi.org/10.1007/s00180-020-00954-8>
- Lee EK, Cook D (2010) A projection pursuit index for large p small n data. *Stat Comput* 20:381–392
- Lin TC, Lin TI (2010) Supervised learning of multivariate skew normal mixture models with missing information. *Comput Stat* 25:183–201
- Lindsay BG, Yao W (2012) Fisher information matrix: a tool for dimension reduction, projection pursuit, independent component analysis, and more. *Can J Stat* 40:712–730
- Loperfido N (2014) Linear transformations to symmetry. *J Multivar Anal* 129:186–192
- Loperfido N (2017) A new kurtosis matrix, with statistical applications. *Linear Algebra Appl* 512:1–17
- Loperfido N (2019) Finite mixtures, projection pursuit and tensor rank: a triangulation. *Adv Data Anal Classif* 31:145–173
- Loperfido N (2020a) Some remarks on kozioł's kurtosis. *J Multivar Anal* 175:104565
- Loperfido N (2020b) Kurtosis-based projection pursuit for outlier detection in financial time series. *Eur J Finance* 26:142–164
- Loperfido N (2020) Representing kozioł's kurtoses. *Mathematical and statistical methods for actuarial sciences and finance MAF 2020*. Springer, New York, p 5
- Loperfido N, Guttorp P (2008) Network bias in air quality monitoring design. *Environmetrics* 19:661–671
- Malkovich JF, Afifi AA (1973) On tests for multivariate normality. *J. Am. Stat. Assoc.* 68:176–179
- Mardia KV (1974) Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B* 36:115–128
- Miettinen J, Taskinen S, Nordhausen K, Oja H (2015) Fourth moments and independent component analysis. *Stat Sci* 3:372–390
- Möri TF, Rohatgi VK, Székely GJ (1993) On multivariate skewness and kurtosis. *Theory Probab Appl* 38:547–551
- Ortega JM (1987) *Matrix theory: a second course*. Plenum Publishing Corporation, New York, NY

- Peña D, Prieto FJ (2000) The kurtosis coefficient and the linear discriminant function. *Stat Probab Lett* 49:257–261
- Peña D, Prieto FJ (2001) Multivariate Outlier Detection and Robust Covariance Estimation (with discussion). *Technometrics* 43:286–310
- Peña D, Prieto FJ (2001) Multivariate outlier detection and robust covariance estimation (with discussion). *J Am Stat Assoc* 96:1433–1445
- Peña D, Prieto FJ (2007) Combining random and specific directions for outlier detection and robust estimation of high-dimensional multivariate data. *J Comput Graph Stat* 16:228–254
- Peña D, Prieto FJ, Viladomat J (2010) Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. *J Multivar Anal* 101:1995–2007
- Pires AM, Branco JM (2019) High dimensionality: The latest challenge to data analysis. [arXiv:1902.04679](https://arxiv.org/abs/1902.04679) [stat.ME]
- Rao CR, Rao MB (1998) *Matrix Algebra and its applications to statistics and econometrics*. World Scientific Co. Pte. Ltd., Singapore
- Ray S (2010) Discussion of Projection pursuit via white noise matrices. by G. Hui and B. Lindsay. *Sankhya B* 72:147–151
- Rezvandehy M, Deutsch CV (2018) Declustering experimental variograms by global estimation with fourth order moments. *Stoch Environ Res Risk Assess* 32:261–277
- Rublik F (2001) Tests of some hypotheses on characteristic roots of covariance matrices not requiring normality assumptions. *Kybernetika* 37:61–78
- Schott JR (2002) Inferences using a structured fourth-order moment matrix. *Sankhyā B* 64:11–25
- Telford RD, Cunningham RB (1991) Sex, sport and body-size dependency of hematology in highly trained athletes. *Med Sci Sports Exerc* 23:788–794
- Tsay AC, Liou M, Simak M, Cheng PE (2017) On hyperbolic transformations to normality. *Comput Stat Data Anal* 115:250–266
- Tukey JW (1977) Modern techniques in data analysis. Nsp-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, Massachusetts
- Tyler DE (1981) Asymptotic inference for eigenvectors. *Ann Stat* 9:725–736
- Yanagihara H (2007) A family of estimators for multivariate kurtosis in a nonnormal linear regression model. *J Multivar Anal* 98:1–29
- Yanagihara H, Tonda T, Matsumoto C (2005) The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption. *J Multivar Anal* 96:237–264
- Yu Y, Zhang P, Jing Y (2016) Fast generation of weak lensing maps by the inverse-Gaussianization method. *Phys Rev D* 94:083520