

Article

# Skewness-Based Projection Pursuit as an Eigenvector Problem in Scale Mixtures of Skew-Normal Distributions

Jorge M. Arevalillo \*  and Hilario Navarro 

Department of Statistics and Operational Research, University Nacional Educación a Distancia (UNED), 28040 Madrid, Spain; hnavarro@ccia.uned.es

\* Correspondence: jmartin@ccia.uned.es

**Abstract:** This paper addresses the projection pursuit problem assuming that the distribution of the input vector belongs to the flexible and wide family of multivariate scale mixtures of skew normal distributions. Under this assumption, skewness-based projection pursuit is set out as an eigenvector problem, described in terms of the third order cumulant matrix, as well as an eigenvector problem that involves the simultaneous diagonalization of the scatter matrices of the model. Both approaches lead to dominant eigenvectors proportional to the shape parametric vector, which accounts for the multivariate asymmetry of the model; they also shed light on the parametric interpretability of the invariant coordinate selection method and point out some alternatives for estimating the projection pursuit direction. The theoretical findings are further investigated through a simulation study whose results provide insights about the usefulness of skewness model-based projection pursuit in the statistical practice.

**Keywords:** skewness; scatter matrices; projection pursuit; scale mixtures of skew normal distribution; eigenvector



**Citation:** Arevalillo, J.M.; Navarro, H. Skewness-Based Projection Pursuit as an Eigenvector Problem in Scale Mixtures of Skew-Normal Distributions. *Symmetry* **2021**, *13*, 1056. <https://doi.org/10.3390/sym13061056>

Academic Editor: Nicola Maria Rinaldo Loperfido

Received: 19 May 2021  
Accepted: 3 June 2021  
Published: 11 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The key idea behind the term “projection pursuit” (PP) is concerned with the search of “interesting” low-dimensional representations of multivariate data, an idea dating back to Kruskal’s early work [1], which later on inspired its use as an exploratory data analysis tool for uncovering hidden patterns and structure in data as described in several of the foundational works [2–5]. In order to assess the interestingness of a projection, a gamut of indices have been proposed [3,4,6–12] in combination with proposals for computational developments and the difficulties posed by the need for efficient optimization routines that implement the search of the PP solution [5,11–17]. Although PP was born as an exploratory data analysis technique, the need to enhance progress towards the inferential arena has motivated the use of skewness and kurtosis, based on third and fourth order moments, as projection indices in the context of parametric models either in an explicit way [11,12,18–20] or implicitly [21–23].

This paper revisits the PP problem when the third moment skewness measure is taken as projection index so that PP reduces to finding the direction that yields the maximal skewness projection, an idea already studied in the past [24]. We will assume that the underlying multivariate distribution belongs to the flexible family of scale mixtures of skew normal (SMSN) distributions so the issue is delimited within the skewness model based PP framework. Under this assumption, it is shown that skewness PP can be described in terms of two eigenvector problems: the first problem stems from the third cumulant matrix of the SMSN distribution and the second problem from the simultaneous diagonalization of the covariance and scale scatter matrices of the SMSN model; it can be shown that both approaches have an appealing interpretation in terms of the shape vector that regulates the asymmetry of the model. The theoretical findings ascertain alternative ways to approach

the PP problem with important computational implications when estimating the direction yielding the maximal skewness projection.

The rest of the paper is organized as follows: Section 2 gives an overview about the SMSN family. Section 3 addresses the eigenvector formulation of PP by exploring the theoretical ins and outs of the problem when the input vector follows a SMSN distribution; under this assumption, two eigenvector standpoints—based on third cumulants and on the scatter matrices of the SMSN model—are used to tackle it. Our findings serve to contribute to the invariant coordinate selection (ICS) literature shedding light on the parametric interpretability of this method under the SMSN model; they also give rise to several theoretical grounded proposals for estimating the PP direction from data, which are discussed in Section 3.3. The theoretical results and the proposals herein are further investigated in Section 4 through a simulation study that shows their usefulness when applied to calculate the PP direction; the performance of the estimations is compared by assessing their mean squared errors in well-defined population scenarios. Finally, Section 5 provides some summary remarks for concluding the paper.

## 2. Background

### 2.1. The Skew Normal and the Scale Mixtures of Skew Normal Distributions

The multivariate skew normal (SN) distribution has become a widely used model for regulating asymmetry departures from normality. Its study has originated fruitful research [25–30] with incremental works pursuing its enrichment and extension to wider classes of distributions [31–39]. This paper adopts the formulation from previous work [25,40] to define the density function of a  $p$ -dimensional SN vector with location vector  $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_p)'$  and scale matrix  $\boldsymbol{\Omega}$  as follows:

$$f(\boldsymbol{x}; \boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = 2\phi_p(\boldsymbol{x} - \boldsymbol{\zeta}; \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\boldsymbol{x} - \boldsymbol{\zeta})) \quad : \quad \boldsymbol{x} \in \mathbb{R}^p, \quad (1)$$

where  $\phi_p(\cdot; \boldsymbol{\Omega})$  is the  $p$ -dimensional normal probability density function (pdf) with zero mean vector and covariance matrix  $\boldsymbol{\Omega}$ ,  $\Phi$  is the distribution function of a  $N(0, 1)$  variable,  $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_p)$  is a diagonal matrix with non-negative entries such that  $\overline{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$  is a correlation matrix and  $\boldsymbol{\alpha}$  or  $\boldsymbol{\eta} = \boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$  is the shape vector that regulates the asymmetry of the model. We put  $\boldsymbol{X} \sim SN_p(\boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  to denote that  $\boldsymbol{X}$  follows a SN distribution with pdf (1); note that when  $\boldsymbol{\alpha} = \mathbf{0}$  the normal pdf is recovered.

The SN vector can be represented through the stochastic formulation:  $\boldsymbol{X} = \boldsymbol{\zeta} + \boldsymbol{\omega}\boldsymbol{Z}$ , where  $\boldsymbol{Z}$  is a normalized multivariate skew normal variable with pdf given by

$$f(\boldsymbol{z}; \mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = 2\phi_p(\boldsymbol{z}; \overline{\boldsymbol{\Omega}})\Phi(\boldsymbol{\alpha}'\boldsymbol{z}). \quad (2)$$

The previous stochastic representation highlights the natural extension of the SN to the SMSN family that accounts for asymmetry and tail weight simultaneously [23,41]; due to its analytical tractability, this family has gained increasing interest both from the theory [21,23,42–45] and the applications, with a few works showing its usefulness in real data [20,22,46]. Here, we use the formulation in [23] to define the SMSN as follows:

**Definition 1.** Let  $\boldsymbol{Z}$  be a random vector such that  $\boldsymbol{Z} \sim SN_p(\mathbf{0}, \overline{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$ , with pdf as in (2), and let  $S$  be a non negative scalar variable, independent of  $\boldsymbol{Z}$ . The vector  $\boldsymbol{X} = \boldsymbol{\zeta} + \boldsymbol{\omega}S\boldsymbol{Z}$ , where  $\boldsymbol{\omega}$  is a diagonal matrix with non-negative entries, is said to follow a multivariate SMSN distribution.

In this case, we write  $\boldsymbol{X} \sim SMSN_p(\boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, H)$ , with  $H$  the distribution function of the mixing variable  $S$ , to indicate that  $\boldsymbol{X}$  follows a  $p$ -dimensional SMSN distribution. When  $\boldsymbol{\alpha} = \mathbf{0}$ , the vector  $\boldsymbol{X}$  follows a scale mixture of normal distributions. On the other hand, if  $H$  is degenerate at  $S = 1$  then the input vector  $\boldsymbol{X}$  has a  $SN_p(\boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  distribution.

## 2.2. Examples of SMSN Distributions

There are some well-known multivariate distributions that belong to the SMSN family. Here, we focus on a few representative instances.

### 2.2.1. The Multivariate SN Distribution

We have just mentioned that the SN distribution is obtained when  $S$  is degenerate at  $S = 1$ . Hence, the only source of non-normality is regulated by the shape vector  $\alpha$ .

### 2.2.2. The Multivariate Skew-t Distribution

The multivariate skew-t (ST) distribution arises when the mixing variable in Definition 1 is  $S = V^{-1/2}$  with  $V \sim \chi^2_\nu/\nu$ . Its pdf [46] is given by

$$f(x; \xi, \Omega, \alpha, \nu) = 2 t_p(x; \nu) T_1 \left( \alpha' \omega^{-1} (x - \xi) \left( \frac{\nu + p}{Q_x + \nu} \right)^{1/2}; \nu + p \right) : x \in \mathbb{R}^p, \quad (3)$$

with the quantity  $Q_x$  given by  $Q_x = (x - \xi)' \Omega^{-1} (x - \xi)$ ,  $t_p(\cdot; \nu)$  the pdf of a  $p$ -dimensional ST variate with  $\nu$  degrees of freedom and  $T_1(\cdot; \nu + p)$  the distribution function of a scalar ST variable with  $\nu + p$  degrees of freedom. We put  $X \sim ST_p(\xi, \Omega, \alpha, \nu)$  to indicate that  $X$  follows a ST distribution with pdf as in (3). Note that when  $\nu \rightarrow \infty$  the ST becomes a  $p$ -dimensional SN distribution, i.e.,  $X \sim SN_p(\xi, \Omega, \alpha)$ .

### 2.2.3. The Multivariate Skew-Slash Distribution

Another model that simultaneously tackles asymmetry and tail weight behavior is the multivariate skew-slash (SSL) distribution [47]. It appears when the mixing variable is  $S = U^{-1/q}$ , with  $U \sim U(0, 1)$  and  $q$  a tail weight parameter such that  $q > 0$ . We write  $X \sim SSL_p(\xi, \Omega, \alpha, q)$  to indicate that  $X$  follows a  $p$ -dimensional SSL distribution with location  $\xi$ , scale matrix  $\Omega$ , shape vector  $\alpha$  and tail parameter  $q > 0$ . When  $q \rightarrow \infty$ , the SSL becomes a SN variable, i.e.,  $X \sim SN_p(\xi, \Omega, \alpha)$ .

## 3. Skewness Projection Pursuit

Skewness PP has already been studied for specific models [18,19,48]. Here, we assume that  $X \sim SMSN_p(\xi, \Omega, \alpha, H)$ ; its scaled version is  $U = \Sigma^{-1/2}(X - \mu)$ , where  $\Sigma$  and  $\mu$  are the covariance matrix and mean vector of  $X$ , respectively. PP is aimed at finding the direction  $c$  for which the variable  $c'U$  attains the maximum skewness, which implies solving the following problem:

$$\max_{c \in \mathbb{R}_0^p} \gamma_1(c'U),$$

with  $\gamma_1$  the standard skewness measure:  $\gamma_1(Y) = E \left( \frac{Y - \mu_Y}{\sigma_Y} \right)^3$ , and  $\mathbb{R}_0^p$  the set of all non-null  $p$ -dimensional vectors.

In order to ensure a unique solution, taking into account the scale invariance of  $\gamma_1$ , the problem can be stated by the following equivalent formulations:

$$\max_{c \in \mathbb{S}_p} \gamma_1(c'U) \quad , \quad \max_{d \in \mathbb{S}_p^*} \gamma_1(d'X) \quad (4)$$

where  $\mathbb{S}_p = \{c \in \mathbb{R}^p : c'c = 1\}$  and  $\mathbb{S}_p^* = \{d \in \mathbb{R}^p : d'\Sigma d = 1\}$  with  $d = \Sigma^{-1/2}c$ . The maximal skewness  $\gamma_{1,p}^D = \max_{c \in \mathbb{S}_p} \gamma_1(c'U) = \max_{d \in \mathbb{S}_p^*} \gamma_1(d'X)$ , with the superscript  $D$  standing for directional skewness, is a well-known index of asymmetry [24].

The vectors yielding the maximal skewness projection for  $X$  and its scaled version are denoted by

$$\lambda_X = \arg \max_{d \in \mathbb{S}_p^*} \gamma_1(d'X) \quad , \quad \lambda_U = \arg \max_{c \in \mathbb{S}_p} \gamma_1(c'U) \quad (5)$$

and satisfy that  $\lambda_X \propto \Sigma^{-1/2} \lambda_U$ .

When a simple condition on the moments of the mixing variable is fulfilled, the PP problem can be brought to the parametric arena [22]. Now, we examine the eigenvector formulation of the problem; some instrumental results about moments and cumulants of SMSN vectors are needed in advance [21,49].

**Proposition 1.** *Let  $X$  be a vector such that  $X \sim \text{SMSN}_p(\xi, \Omega, \alpha, H)$ . Provided that the first three moments,  $c_k = E(S^k) : k = 1, 2, 3$ , of the mixing variable exist, the moments of  $X - \xi$  up to the third order are given by*

1.  $M_1 = E[(X - \xi)] = \sqrt{\frac{2}{\pi}} c_1 \gamma.$
  2.  $M_2 = E[(X - \xi)(X - \xi)'] = c_2 \Omega.$
  3.  $M_3 = E[(X - \xi) \otimes (X - \xi)' \otimes (X - \xi)] = \sqrt{\frac{2}{\pi}} c_3 (\gamma \otimes \Omega + \text{vec}(\Omega) \gamma' + \Omega \otimes \gamma - \gamma \gamma' \otimes \gamma)$
- where the symbols  $\text{vec}$  and  $\otimes$  denote the vectorization and Kronecker operators and  $\gamma$  is a shape asymmetry vector given by  $\gamma = \frac{\Omega \eta}{\sqrt{1 + \eta' \Omega \eta}}.$

The next proposition establishes the relationship between cumulants and moments of multivariate SMSN distributions.

**Proposition 2.** *Let  $X$  be a vector such that  $X \sim \text{SMSN}_p(\xi, \Omega, \alpha, H)$ . Provided that the first three moments,  $c_k = E(S^k) : k = 1, 2, 3$ , of the mixing variable exist, the cumulants of  $X - \xi$  up to the third order are given by*

1.  $K_1(X - \xi) = M_1.$
2.  $K_2(X - \xi) = \Sigma = c_2 \Omega - \frac{2}{\pi} c_1^2 \gamma \gamma'.$
3.  $K_3(X - \xi) = M_3 - M_2 \otimes M_1 - M_1 \otimes M_2 - \text{vec}(M_2) M_1' + 2M_1 \otimes M_1' \otimes M_1.$

### 3.1. Projection Pursuit from the Third Cumulant Matrix

Now, skewness PP is revisited using an eigenvector approach from the third cumulant matrix of the scaled input vector  $U = \Sigma^{-1/2}(X - \mu)$ . The approach is motivated by the formulation of directional skewness as the maximum of an homogeneous third-order polynomial obtained from the third cumulant matrix as follows [11]:

$$\gamma_{1,p}^D = \max_{c \in \mathbb{S}_p} \gamma_1(c'U) = \max_{c \in \mathbb{S}_p} (c \otimes c)' K_3(U) c. \tag{6}$$

This formulation inspires the eigenvector approach, based on the third-cumulant tensor, and the use of the higher-order power method (HOPM) to implement the computation [14]. The problem admits a simplification for skew-t vectors, which in turn provides the theoretical grounds and interpretation for the choice of the initial direction in the HOPM algorithm ([11] Proposition 2). Theorem 1 below enriches the interpretation by extending the result to the family of SMSN distributions.

Firstly, we recall the definitions of left and right singular vectors of the third cumulant matrix of a vector  $U$ : We say that the vector  $v$  is a right singular vector of the  $p^2 \times p$  third cumulant matrix  $K_3(U)$  if  $K_3'(U) K_3(U) v$  is proportional to  $v$ . On the other hand, the  $\text{vec}$  operator can be used to define a left singular vector of  $K_3(U)$  as the  $p^2 \times 1$  vectorized matrix,  $w = \text{vec}(W)$  with  $W$  a  $p \times p$  matrix, such that  $K_3(U) K_3'(U) w$  gives a vector proportional to  $w$ .

**Theorem 1.** *Let  $X$  be a vector such that  $X \sim \text{SMSN}_p(\xi, \Omega, \alpha, H)$ , with mixing variable having finite second order moment satisfying the condition  $\frac{4}{\pi} c_1^2 > c_2$ , and let  $U$  be the scaled version of*

the input vector having third cumulant matrix  $\mathbf{K}_3(\mathbf{U})$ . Then, it holds that the right dominant eigenvector of  $\mathbf{K}_3(\mathbf{U})$  is given by  $\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\gamma}$ .

**Proof.** Putting together the results from Propositions 1 and 2, we can write:

$$\begin{aligned} \mathbf{K}_3(\mathbf{X}) &= \sqrt{\frac{2}{\pi}}c_3(\boldsymbol{\gamma} \otimes \boldsymbol{\Omega} + \text{vec}(\boldsymbol{\Omega})\boldsymbol{\gamma}' + \boldsymbol{\Omega} \otimes \boldsymbol{\gamma} - \boldsymbol{\gamma}\boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}) - \sqrt{\frac{2}{\pi}}c_1c_2\boldsymbol{\Omega} \otimes \boldsymbol{\gamma} \\ &\quad - \sqrt{\frac{2}{\pi}}c_1c_2\boldsymbol{\gamma} \otimes \boldsymbol{\Omega} - \sqrt{\frac{2}{\pi}}c_1c_2\text{vec}(\boldsymbol{\Omega})\boldsymbol{\gamma}' + 2\sqrt{\frac{2}{\pi}}\frac{2}{\pi}c_1^3\boldsymbol{\gamma} \otimes \boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}. \end{aligned}$$

Taking into account the expression for the covariance matrix provided in the second item of Proposition 2, we obtain that

$$\boldsymbol{\Omega} \otimes \boldsymbol{\gamma} = \frac{1}{c_2}\boldsymbol{\Sigma} \otimes \boldsymbol{\gamma} + \frac{2}{\pi} \frac{c_1^2}{c_2}\boldsymbol{\gamma} \otimes \boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}. \tag{7}$$

$$\boldsymbol{\gamma} \otimes \boldsymbol{\Omega} = \frac{1}{c_2}\boldsymbol{\gamma} \otimes \boldsymbol{\Sigma} + \frac{2}{\pi} \frac{c_1^2}{c_2}\boldsymbol{\gamma} \otimes \boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}. \tag{8}$$

$$\text{vec}(\boldsymbol{\Omega})\boldsymbol{\gamma}' = \frac{1}{c_2}\text{vec}(\boldsymbol{\Sigma})\boldsymbol{\gamma}' + \frac{2}{\pi} \frac{c_1^2}{c_2}\boldsymbol{\gamma} \otimes \boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}. \tag{9}$$

When these quantities are inserted into the previous expression for  $\mathbf{K}_3(\mathbf{X})$ , we obtain

$$\mathbf{K}_3(\mathbf{X}) = r\boldsymbol{\Sigma} \otimes \boldsymbol{\gamma} + r\boldsymbol{\gamma} \otimes \boldsymbol{\Sigma} + r\text{vec}(\boldsymbol{\Sigma})\boldsymbol{\gamma}' + s\boldsymbol{\gamma} \otimes \boldsymbol{\gamma}' \otimes \boldsymbol{\gamma}, \tag{10}$$

where  $r = \sqrt{\frac{2}{\pi}}\left(\frac{c_3}{c_2} - c_1\right)$  and  $s = \sqrt{\frac{2}{\pi}}\frac{2}{\pi}\frac{1}{c_2}\left(3c_1^2c_3 - c_1^3c_2 - \frac{\pi}{2}c_2c_3\right)$ . Now, taking into account that  $\mathbf{K}_3(\mathbf{U}) = \left(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}\right)\mathbf{K}_3(\mathbf{X})\boldsymbol{\Sigma}^{-1/2}$  [50], an analogous argument as the one used in [11] leads to

$$\mathbf{K}_3(\mathbf{U}) = r[\mathbf{I}_p \otimes \boldsymbol{\lambda} + \boldsymbol{\lambda} \otimes \mathbf{I}_p + \text{vec}(\mathbf{I}_p)\boldsymbol{\lambda}'] + s\boldsymbol{\lambda} \otimes \boldsymbol{\lambda}' \otimes \boldsymbol{\lambda}.$$

On the other hand, using the standard properties of the kronecker product, the following mathematical relationships can be derived after simple calculations:

$$(\mathbf{I}_p \otimes \boldsymbol{\lambda}')\mathbf{K}_3(\mathbf{U}) = (\boldsymbol{\lambda}' \otimes \mathbf{I}_p)\mathbf{K}_3(\mathbf{U}) = r\boldsymbol{\lambda}'\boldsymbol{\lambda}\mathbf{I}_p + (2r + \boldsymbol{\lambda}'\boldsymbol{\lambda}s)(\boldsymbol{\lambda}\boldsymbol{\lambda}'),$$

$$\boldsymbol{\lambda} \text{vec}(\mathbf{I}_p)'\mathbf{K}_3(\mathbf{U}) = (r(p + 2) + \boldsymbol{\lambda}'\boldsymbol{\lambda}s)(\boldsymbol{\lambda}\boldsymbol{\lambda}'),$$

$$(\boldsymbol{\lambda}' \otimes \boldsymbol{\lambda} \otimes \boldsymbol{\lambda}')\mathbf{K}_3(\mathbf{U}) = (\boldsymbol{\lambda}'\boldsymbol{\lambda})(3r + \boldsymbol{\lambda}'\boldsymbol{\lambda}s)(\boldsymbol{\lambda}\boldsymbol{\lambda}'),$$

from which we obtain

$$\mathbf{K}'_3(\mathbf{U})\mathbf{K}_3(\mathbf{U}) = 2r^2(\boldsymbol{\lambda}'\boldsymbol{\lambda})\mathbf{I}_p + [r^2(p + 6) + 6rs(\boldsymbol{\lambda}'\boldsymbol{\lambda}) + s^2(\boldsymbol{\lambda}'\boldsymbol{\lambda})^2]\boldsymbol{\lambda}\boldsymbol{\lambda}',$$

and as a result:  $\mathbf{K}'_3(\mathbf{U})\mathbf{K}_3(\mathbf{U})\boldsymbol{\lambda} = \rho_\lambda\boldsymbol{\lambda}$ , where  $\rho_\lambda = \boldsymbol{\lambda}'\boldsymbol{\lambda}[r^2(p + 8) + 6rs\boldsymbol{\lambda}'\boldsymbol{\lambda} + s^2(\boldsymbol{\lambda}'\boldsymbol{\lambda})^2]$ . This finding implies that  $\boldsymbol{\lambda}$  is an eigenvector of  $\mathbf{K}'_3(\mathbf{U})\mathbf{K}_3(\mathbf{U})$  with  $\rho_\lambda$  its associated eigenvalue. Hence,  $\boldsymbol{\lambda}$  is a right singular vector of  $\mathbf{K}_3(\mathbf{U})$ .

From the well-known moment inequality  $c_3 \geq c_1c_2$  we conclude that  $r \geq 0$ . Thus, taking into account the condition for the moments of the mixing variable of the statement, we can assert that

$$s = \sqrt{\frac{2}{\pi}} \frac{1}{c_2} \left( \frac{6}{\pi} c_1^2 c_3 - \frac{2}{\pi} c_1^3 c_2 - c_2 c_3 \right) \geq \sqrt{\frac{2}{\pi}} \frac{1}{c_2} \left( \frac{6}{\pi} c_1^2 c_3 - \frac{2}{\pi} c_1^3 c_2 - \frac{4}{\pi} c_1^2 c_3 \right) = \sqrt{\frac{2}{\pi}} \frac{1}{c_2} \frac{2}{\pi} c_1^2 (c_3 - c_1 c_2) \geq 0,$$

which implies that  $\rho_\lambda \geq 0$ .

It can also be shown that  $\lambda$  is the right dominant eigenvector of  $K_3(\mathbf{U})$  using the following simple argument: If  $\mathbf{v}$  is another right eigenvector orthogonal to  $\lambda$  then  $K_3'(\mathbf{U})K_3(\mathbf{U})\mathbf{v} = 2r^2(\lambda'\lambda)\mathbf{v}$  with  $2r^2(\lambda'\lambda) < \rho_\lambda$ . Therefore,  $\lambda$  is the right dominant eigenvector of  $K_3(\mathbf{U})$ . □

The result given by Theorem 1 gives rise to the following meaningful remarks:

First of all, we know that the direction attaining the maximal skewness projection for the scaled vector  $\mathbf{U}$  is proportional to  $\lambda$  [22]; henceforth, the vector driving the maximal skewness projection for  $\mathbf{X}$  is  $\Sigma^{-1}\gamma$  which in turn lies on the direction of  $\eta$ , being also an eigenvector of  $\Sigma^{-1}\Omega$  ([22] Lemma 1). This fact enhances the role of  $\eta$  as a shape vector that regulates the asymmetry of the multivariate model in a directional fashion. Furthermore, the previous disquisition also provides a grounded theoretical basis for the choice of the right dominant eigenvector of  $K_3(\mathbf{U})$  to set the initial direction of the HOPM algorithm, an issue advocated by De Lathauwer et al. [14] without theoretical justification.

Secondly, we can also note that the expression

$$K_3'(\mathbf{U})K_3(\mathbf{U}) = 2r^2(\lambda'\lambda)I_p + [r^2(p + 6) + 6rs(\lambda'\lambda) + s^2(\lambda'\lambda)^2]\lambda\lambda',$$

from which the right dominant singular vector of  $K_3(\mathbf{U})$  is derived, reduces to that one found by (Loperfido [11] Proposition 2) for the case of the multivariate skew-t distribution when the moments of the skew-t are inserted into the quantities  $r$  and  $s$  above. Actually, our result is an extension of (Loperfido [11] Theorem 2) to the SMSN family.

Thirdly, coming back to expression Equation (6), recall that a unit length vector  $\mathbf{v}$  satisfies  $\gamma_{1,p}^D(\mathbf{X}) = \gamma_1(\mathbf{v}'\mathbf{U})$  if and only if  $K_3'(\mathbf{U})(\mathbf{v} \otimes \mathbf{v}) = \rho\mathbf{v}$  with  $\rho$  the largest eigenvalue [11,51]. Thus, we can put  $\mathbf{v} = \lambda/\sqrt{\lambda'\lambda}$ , with  $\lambda = \Sigma^{-1/2}\gamma$  from Theorem 1, to show that

$$K_3'(\mathbf{U})(\mathbf{v} \otimes \mathbf{v}) = [3r(\lambda'\lambda) + s(\lambda'\lambda)^2] \frac{1}{\sqrt{\lambda'\lambda}} \mathbf{v}.$$

Therefore,  $\gamma_{1,p}^D(\mathbf{X}) = \mathbf{v}'K_3'(\mathbf{U})(\mathbf{v} \otimes \mathbf{v}) = \frac{3r(\lambda'\lambda) + s(\lambda'\lambda)^2}{\sqrt{\lambda'\lambda}}$  which, taking into account that  $\lambda'\lambda = \gamma'\Sigma^{-1}\gamma$  with  $\Sigma^{-1}\gamma = \frac{\Omega^{-1}\gamma}{c_2 - \frac{2}{\pi}c_1^2\gamma'\Omega^{-1}\gamma}$ , reduces to the expression derived by [22] using another approach. Moreover, for the case of a SN vector we know that  $S$  is degenerate at  $S = 1$ ; so  $r = 0$ ,  $s = \sqrt{\frac{2}{\pi}} \left( \frac{4}{\pi} - 1 \right)$  and  $\lambda'\lambda = \frac{\eta'\Omega\eta}{1 + (1 - \frac{2}{\pi})\eta'\Omega\eta'}$  which gives  $[\gamma_{1,p}^D(\mathbf{X})]^2 = s^2(\lambda'\lambda)^3 = 2(4 - \pi)^2 \left( \frac{\eta'\Omega\eta}{\pi + (\pi - 2)\eta'\Omega\eta'} \right)^3$  as previously stated by [18].

Finally, it is worthwhile noting that the moment condition given in the statement of Theorem 1 is fulfilled by the most popular subfamilies of SMSN distributions such as those described in Section 2.2.

### 3.2. Projection Pursuit from Scatter Matrices

In the previous section, projection pursuit under SMSN multivariate distributions is formulated as an eigenvector problem which stems from the third order cumulant matrix. In this section, the approach is formulated through an eigenvector problem concerned with



the simultaneous diagonalization of the scatter matrices  $\Sigma$  and  $\Omega$ ; in fact, such formulation is related to the canonical transformation of an SMSN vector [23]. The next theorem provides the theoretical foundation for the approach.

**Theorem 2.** *Let  $\mathbf{X}$  be a vector such that  $\mathbf{X} \sim \text{SMSN}_p(\xi, \Omega, \alpha, H)$  with mixing variable having finite first and second order moments. Then,  $\eta$  is the dominant eigenvector of  $\Sigma^{-1}\Omega$ .*

**Proof.** The proof follows pieces of previous work [22,23], so we sketch it. From (Arevalillo and Navarro [22] [Lemma 1]), we know that the shape vector  $\eta$  is an eigenvector of  $\Sigma^{-1}\Omega$ ; it is straightforward to show that it has eigenvalue

$$\rho = \frac{1}{c_2 - \frac{2}{\pi}c_1^2 \frac{\eta'\Omega\eta}{1+\eta'\Omega\eta}} = \frac{1}{c_2 - \frac{2}{\pi}c_1^2 \gamma'\Omega^{-1}\gamma}.$$

Now, let us consider another eigenvector  $\mathbf{u}$  of  $\Sigma^{-1}\Omega$ . Then, it holds that

$$c_2\Sigma^{-1}\Omega\mathbf{u} = \mathbf{u} + \frac{\eta\eta'\Omega\mathbf{u}/(1+\eta'\Omega\eta)}{\frac{\pi}{2}\frac{c_2}{c_1^2} - \gamma'\Omega^{-1}\gamma} = \mathbf{u},$$

since the vector  $\mathbf{u}$  is orthogonal to  $\gamma$  [23]. As a result,  $\mathbf{u}$  has associated eigenvalue equal to  $1/c_2$  which satisfies that  $1/c_2 \leq \rho$ ; hence, we conclude that  $\eta$  is the dominant eigenvector of  $\Sigma^{-1}\Omega$ .  $\square$

Theorem 2 establishes that the dominant eigenvector of  $\Sigma^{-1}\Omega$  lies on the direction of  $\eta$ , which in turn gives the direction at which the maximal skewness projection is attained [22]; this finding points out the parametric interpretation of the dominant eigenvector of  $\Sigma^{-1}\Omega$  under SMSN distributions as a shape parameter accounting for the directional asymmetry of the model. On the other hand, it is well-known that the canonical transformation of a SMSN vector can be derived by the simultaneous diagonalization of the pair of scatter matrices  $(\Sigma, \Omega)$ , with the only skewed canonical variate being obtained by the projection onto the direction of the shape vector  $\eta$  [23]; therefore, the computation of the skewness-based PP under the SMSN family would only require the estimation of both scatter matrices. Another alternative to derive the canonical transformation relies on the pair of scatter matrices  $(\Sigma, \mathcal{K})$ , where  $\mathcal{K}$  is the scatter kurtosis matrix given by  $\mathcal{K} = E[(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)(\mathbf{X} - \mu)(\mathbf{X} - \mu)']$  [52]; such alternative has appealing properties for SMSN distributions [23]. Hence, we can work with either  $(\Sigma, \Omega)$  or  $(\Sigma, \mathcal{K})$ , in order to derive the canonical form of a SMSN vector, and compute the corresponding dominant eigenvectors in their simultaneous diagonalizations to get the vector driving the projection pursuit direction.

It is worthwhile noting that the ideas behind the canonical form of SMSN vectors are inspired in the ICS method for representing multivariate data through the comparison of two scatter matrices [53]. Although the choice of the scatter pair is of key relevance,  $(\Sigma, \mathcal{K})$  has become a standard choice which works well in the statistical practice. The idea behind the ICS method is the search of a new representation coordinate system of the data by sequential maximization of the quotient  $\kappa_1(\mathbf{a}) = \frac{\mathbf{a}'\mathcal{K}\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}}$ . For SMSN distributions, a natural alternative is the scatter pair  $(\Sigma, \Omega)$ , given by the covariance and scale matrices of the model; therefore, the projection index for maximization would be  $\kappa_2(\mathbf{a}) = \frac{\mathbf{a}'\Omega\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}}$ . The study and comparison of projection pursuit and the ICS method has been previously discussed in the literature ascertaining the problem of which method is better suited to uncover structure in multivariate data [54–56]; the SMSN family is a flexible and wide class of distributions that allows to settle the problem within a parametric framework so that we can deep dive into the investigation by means of a simulation study at controlled parametric scenarios.

The next section tackles several issues concerned with the estimation of the dominant eigenvector from the formulations we have studied so far: on the one hand, the PP from the third cumulant matrix and, on the other hand, the projection pursuit from scatter matrices with the indices  $\kappa_1$  and  $\kappa_2$ .

### 3.3. Estimation and Computational Issues

The previous theoretical material has outstanding practical implications, as they point to different estimation methods for computing the skewness PP direction for SMSN vectors. In this section, we discuss the computational side of the issue.

Let us denote by  $\mathbf{X}_n$  the  $n \times p$  data matrix and let  $\mathbf{S}$  be the regular covariance matrix. Furthermore, let  $\mathbf{U}_n = (\mathbf{X}_n - \mathbf{1}_n \bar{\mathbf{x}}') \mathbf{S}^{-1/2}$  be the scaled data matrix, where  $\mathbf{1}_n$  is a  $n \times 1$  vector,  $\bar{\mathbf{x}}$  is the  $p \times 1$  sample mean vector and  $\mathbf{S}^{-1/2}$  the symmetric positive definite square root of  $\mathbf{S}^{-1}$ . Then, the third cumulant matrix  $\mathbf{K}_3(\mathbf{U})$  can be estimated by  $\hat{\mathbf{K}}_{3,n} = \frac{1}{n} \sum_{i=1}^n \mathbf{u}_i \otimes \mathbf{u}_i' \otimes \mathbf{u}_i$ , where  $\mathbf{u}_i'$  is the  $i$ -th row of  $\mathbf{U}_n$ .

The result given by Theorem 1 gives theoretical support for the empirical conjecture that proposes the right dominant eigenvector of the cumulant matrix  $\hat{\mathbf{K}}_{3,n}$  as the starting direction of the HOPM iterative method [14], an issue also ascertained by Loperfido [11, Proposition 1] under the rank-one assumption for the left dominant singular vector of the third cumulant matrix, as well as in the context of the multivariate skew-t distribution with a reference to Arevalillo and Navarro [19]. When the input vector  $\mathbf{X}$  follows a SMSN distribution, we know that the vector  $\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}$ , with  $\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\gamma}$  the right dominant eigenvector of  $\mathbf{K}_3(\mathbf{U})$ , is proportional to the shape vector  $\boldsymbol{\eta}$  that yields the maximal skewness projection [22]. Then, it stands to reason that the skewness PP direction can be estimated by  $\hat{\boldsymbol{e}}_{cum} = \mathbf{S}^{-1/2} \hat{\boldsymbol{e}}_{1,\mathbf{K}}$ , where  $\hat{\boldsymbol{e}}_{1,\mathbf{K}}$  is the right dominant eigenvector of  $\hat{\mathbf{K}}_{3,n}$ ; therefore, the first method we propose to obtain the maximal skewness direction for SMSN vectors is non-parametric in nature and relies on the computation of the right dominant eigenvector of the third cumulant sample matrix.

**Method 1 (M1).** Estimate the skewness PP direction by  $\hat{\boldsymbol{e}}_{cum} = \mathbf{S}^{-1/2} \hat{\boldsymbol{e}}_{1,\mathbf{K}}$ .

Another approach to compute the PP direction arises from the scatter matrices  $\boldsymbol{\Sigma}$  and  $\mathbf{K}$  which can also be estimated from data. In this case the PP direction can be derived by computing the dominant eigenvector of the matrix  $\mathbf{S}^{-1} \hat{\mathbf{K}}$ , where  $\hat{\mathbf{K}}$  is a sample estimate of the population kurtosis matrix  $\mathbf{K}$ . The ICS R package [57] has functionalities for the estimation of  $\mathbf{K}$  and the computation of the dominant eigenvector as a result. Hence, the second approach could be stated as follows.

**Method 2 (M2).** Estimate the skewness PP direction by the dominant eigenvector,  $\hat{\boldsymbol{e}}_{1,\mathbf{K}}$ , of  $\mathbf{S}^{-1} \hat{\mathbf{K}}$ .

Theorem 2 suggests an alternative approach to the problem, which brings it to the parametric arena: Let us denote by  $\hat{\boldsymbol{\Sigma}}$  and  $\hat{\boldsymbol{\Omega}}$  the maximum likelihood estimators of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Omega}$ ; then the dominant eigenvector of  $\hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Omega}}$  is the estimator of the shape vector  $\boldsymbol{\eta}$  derived from the maximum likelihood (ML) method. The ML approach has obvious practical limitations as it requires knowing in advance the multivariate distribution of the underlying model that generated the data. In order to overcome this limitation, we propose an approach aimed at balancing the accuracy and the computational cost of the estimation procedure; a varied bunch of simulation assays—not reported here for the sake of space—suggests, as a good tradeoff, the estimation of the dominant eigenvector of  $\boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}$  by the dominant eigenvector of  $\mathbf{S}^{-1} \hat{\boldsymbol{\Omega}}_{SN}$ , where  $\mathbf{S}$  is the regular covariance matrix and  $\hat{\boldsymbol{\Omega}}_{SN}$  is the estimation of the scale matrix  $\boldsymbol{\Omega}$ , calculated by the ML method under the multivariate SN model. The proposal can be stated as follows:



**Method 3 (M3).** Estimate the skewness PP direction by  $\hat{e}_{1,SN}$  which is the dominant eigenvector of  $S^{-1}\hat{\Omega}_{SN}$ .

In the next section, a simulation study with artificial data is carried out in order to investigate the performance of the proposed methods.

#### 4. Simulation Experiment

Here, we consider the multivariate ST and SSL models; for the sake of simplicity, we take  $\xi = 0$  for the location vector. The assays of the simulation study are carried out by drawing 1000 samples from a multivariate ST or SSL distribution under the experimental sampling conditions specified as follows. The dimension of the input vector is set at the values  $p = 2, 4, 8$ ; the parameters of the underlying model are determined by the following settings: the correlation matrix is defined by a Toeplitz matrix  $\bar{\Omega} = (w_{i,j})_{1 \leq i, j \leq p}$ , where  $w_{i,j} = \rho^{|i-j|}$  :  $1 \leq i \leq j \leq p$ , with  $\rho = 0.25, 0.75$ , and the entries of the diagonal matrix  $\omega$  are generated at random from the set of integer values between 5 and 25; the shape vector is given by  $\alpha = \tau e$  with  $e$  a unit-norm vector chosen at random and  $\tau$  the magnitude, which is set at the values  $\tau = 2, 5, 10$ ; the tail weight parameters are set at the values  $\nu = 5, 10$  for the ST model and  $q = 4, 10$  for the SSL model. Finally, the sample size is taken in accordance with the dimension of the input vector at the values  $n = 60, 120, 240$ . Under these experimental settings, the accuracy of the proposed estimation methods is assessed using the mean square error (MSE), which is computed through the comparison of the unit-norm vectors corresponding to the directions estimated by each method with the exact theoretical direction of the multivariate model, using the square of the Euclidean distance.

The next section gives the results of the simulations for the bidimensional case, they are illustrated by fancy plots of directions; on the other hand, Section 4.2 shows the results when the dimension is  $p > 2$ .

##### 4.1. Simulation Study for Bidimensional SMSN Distributions

This section addresses the bidimensional case; the outputs of the simulations are summarized by means of plots of directions arranged in tables of figures that display the following information: the red and black lines represent the directions of the theoretical shape vectors  $\eta$  and  $\alpha$ , respectively. On the other hand, the points in the coloured circular crowns correspond to the locations of the directions estimated by the proposed estimation methods, with the blue crown showing the estimations provided by M1, the red crown by M2 and the green crown by M3; their MSE values are displayed at the bottom right legend of each figure. Finally, the clouds of points for the observations obtained by the simulations are also depicted for completeness.

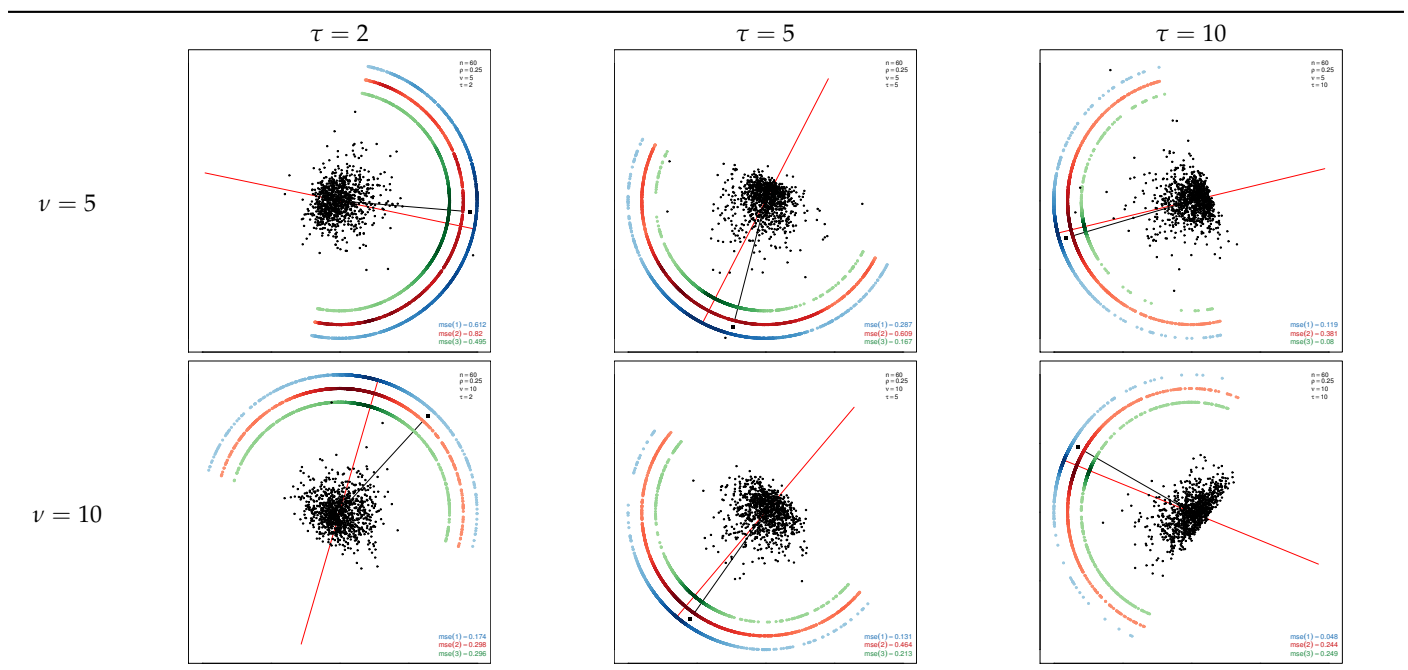
In order to assess the performance of the estimation methods for the bidimensional case, the simulation assays are carried out taking a sample size  $n = 60$  which may be a representative choice of the small data sets arising in real studies. The plots depicted in Table 1 show the MSEs obtained for samples coming from a ST distribution when  $\rho = 0.25$ . Overall, we can observe that methods M1 and M3 are more accurate than M2. The outputs show that the errors are quite similar for all the methods in the scenario closer to the normal model, provided by the pair  $(\tau, \nu) = (2, 10)$ ; starting from this scenario, we can see that, as we inject asymmetry into the multivariate model, M1 outperforms methods M2 and M3 but its accuracy gets worse when we move to the heavy tail scenarios for  $\nu = 5$ ; in such a case, M3 provides the most accurate results ( $MSE = 0.495, 0.167, 0.080$ ). On the other hand, the results displayed by the plots in Table 2 for the coefficient  $\rho = 0.75$  show that M3 is the most accurate method in all the situations, with its best performance ( $MSE = 0.036$ ) in the scenario which is furthest away from the normal distribution.

The results of the simulations under the SSL distribution are given by the plots displayed in Tables 3 and 4. In this case, the accuracy of the estimation methods presents similar patterns as those ones observed under the ST model, with the exceptions of the

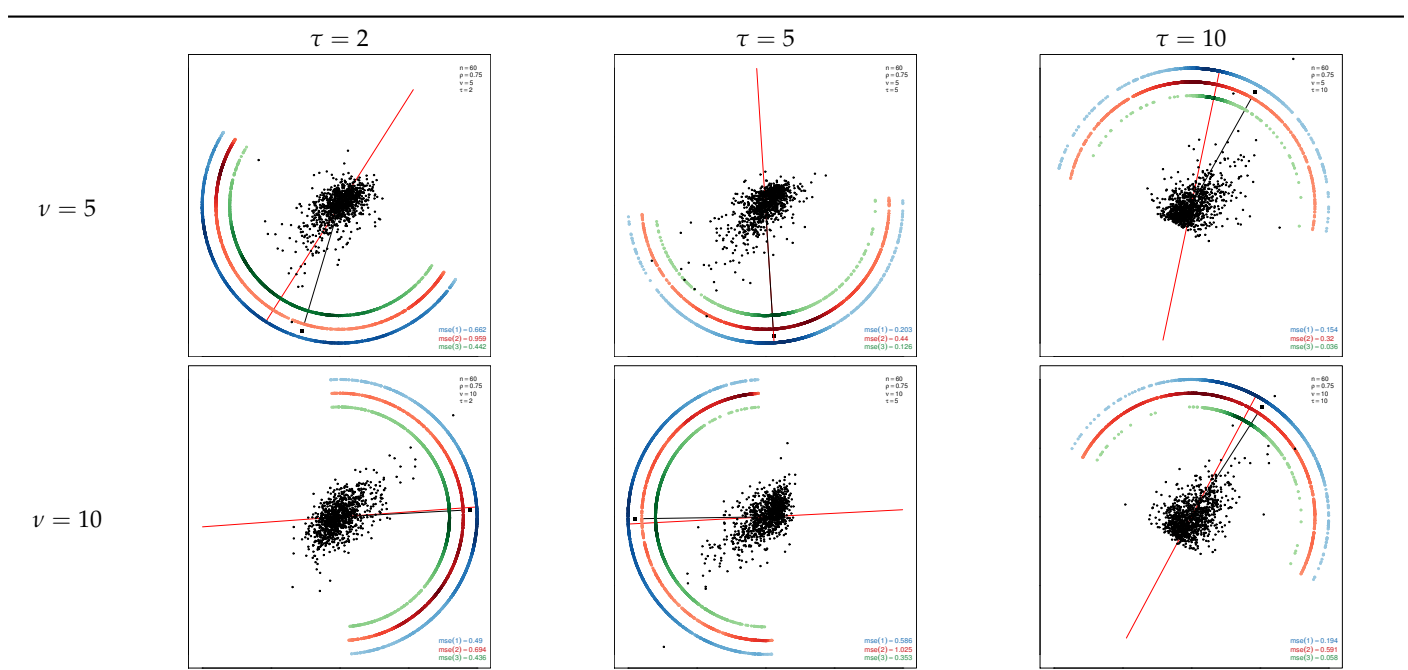
non-normality pair  $(\tau, q) = (10, 10)$  at the bottom right corner, where M3 outperforms M1 and M2, and the pair  $(\tau, q) = (2, 4)$ , at the top right corner, where M2 and M3 are outperformed by M1. When the coefficient  $\rho$  is equal to 0.75 (see Table 4), we can observe once again that M3 gives more accurate results than M1 and M2.

We have deployed additional experiments for other sample sizes, not reported here for the sake of space. In order to illustrate their results, we provide a representative case when  $n = 200$  for the scenario with the highest asymmetry and tail weight under both multivariate models ST and SSL. The results are displayed by the plots appearing in the figures of Table 5; they show the remarkably superiority of method M3.

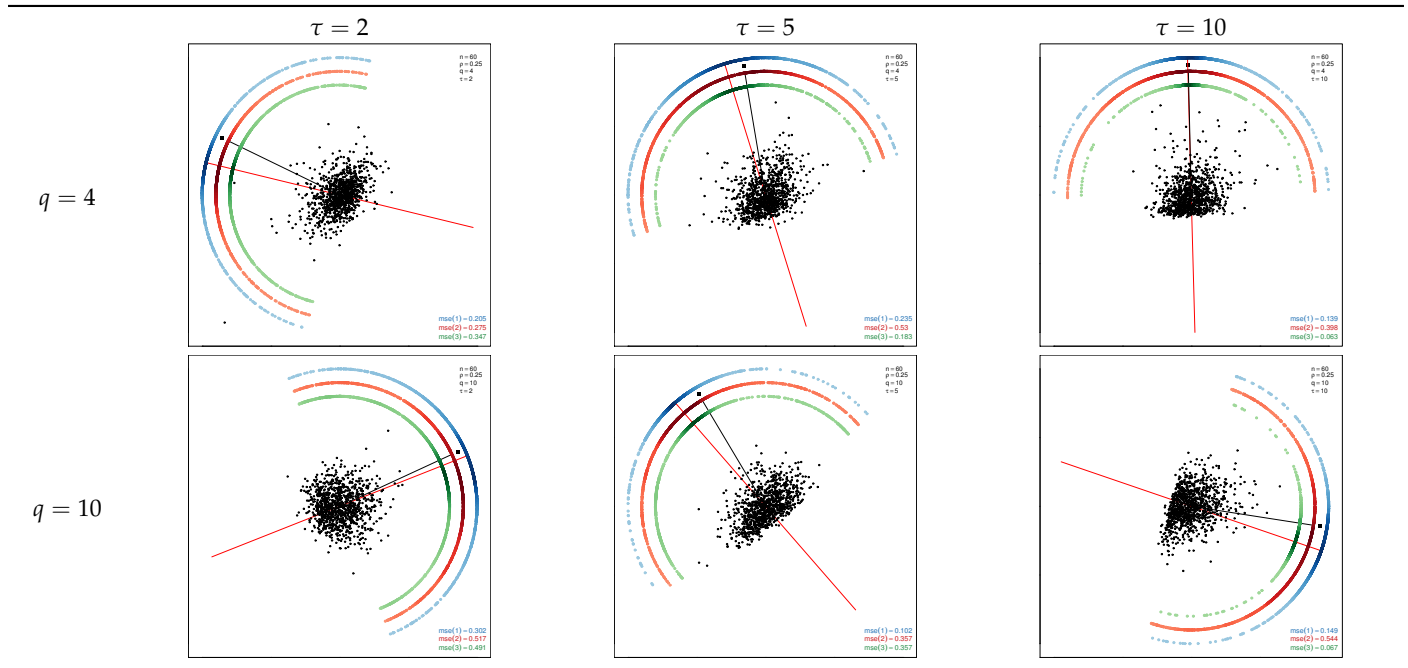
**Table 1.** Plots with the theoretical and estimated directions under the bidimensional ST distribution when  $\rho = 0.25$ .



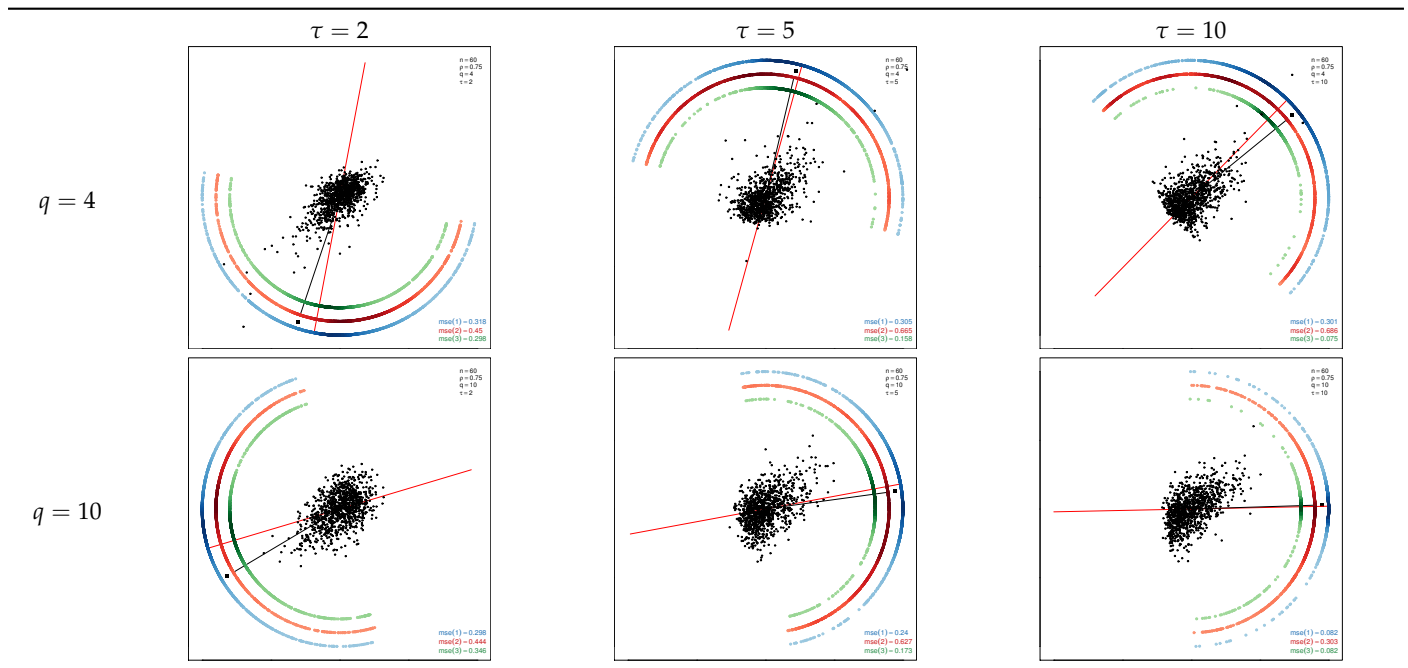
**Table 2.** Plots with the theoretical and estimated directions under the bidimensional ST distribution when  $\rho = 0.75$ .



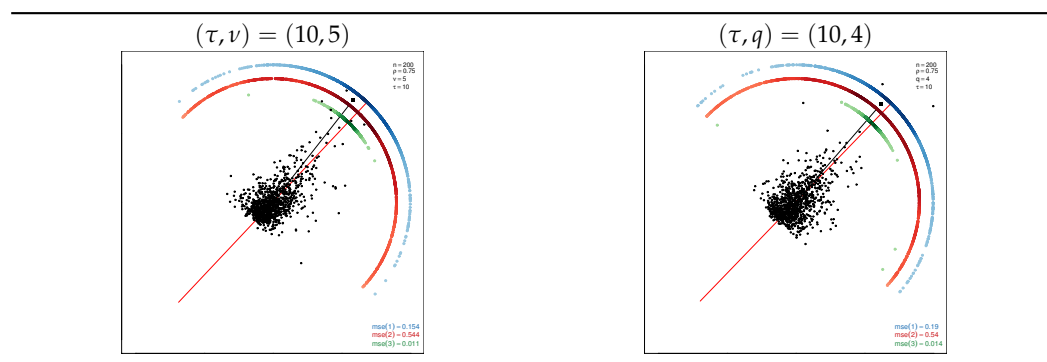
**Table 3.** Plots with the theoretical and estimated directions under the bidimensional SSL distribution when  $\rho = 0.25$ .



**Table 4.** Plots with the theoretical and estimated directions under the bidimensional SSL distribution when  $\rho = 0.75$ .



**Table 5.** Plots with the theoretical and estimated directions under the bidimensional ST (left) and SSL (right) distributions for the highest asymmetry and tail weight levels when  $\rho = 0.75$ .



4.2. Simulation Study for SMSN Distributions when  $p > 2$

We have carried out a wide range of experimental trials for dimensions greater than two. Here, we just provide the results obtained from a few simulation assays—for dimensions  $p = 4, 8$  and sample sizes  $n = 60, 120, 240$ —which summarize and illustrate our main findings (see Tables 6 and 7). As could be expected, we observe that, as the dimension increases, the accuracy of all the estimation methods deteriorates, with M3 being the most resistant method to the curse of dimensionality.

**Table 6.** Results under the ST distribution. For each non-normality pair  $(\tau, \nu)$ , dimensions of the data frame  $(p, n)$  and coefficient  $\rho$ , the cells contain the MSEs given by method 1 (first row), method 2 (second row) and method 3 (third row).

$(p, n)$	$(4, 60)$		$(4, 120)$		$(8, 120)$		$(8, 240)$	
$\rho$	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75
$(\tau, \nu)$								
$(2, 5)$	0.646	0.916	0.500	0.584	1.306	1.090	1.284	1.472
	0.772	0.975	0.677	0.675	1.383	1.155	1.521	1.632
	0.654	0.875	0.435	0.528	1.090	0.931	0.826	0.903
$(2, 10)$	0.610	0.932	0.518	0.553	0.873	1.182	1.195	1.280
	0.760	1.075	0.865	0.665	1.029	1.260	1.525	1.580
	0.668	0.831	0.542	0.579	0.938	1.097	0.888	0.838
$(5, 5)$	0.258	0.581	0.402	0.403	0.757	1.303	0.634	0.768
	0.443	0.801	0.794	0.629	0.956	1.413	0.971	1.048
	0.338	0.355	0.167	0.105	0.640	0.526	0.244	0.187
$(5, 10)$	0.362	0.738	0.299	0.477	0.536	0.918	0.369	0.550
	0.622	1.048	0.769	0.953	0.733	1.157	0.854	1.107
	0.469	0.428	0.198	0.195	0.557	0.785	0.227	0.254
$(10, 5)$	0.382	0.496	0.456	0.573	0.696	0.749	0.632	0.944
	0.618	0.725	0.894	0.977	0.938	0.899	0.987	1.263
	0.305	0.333	0.128	0.067	0.550	0.315	0.101	0.107
$(10, 10)$	0.495	0.549	0.167	0.644	0.657	0.731	0.265	0.460
	0.776	0.888	0.620	1.165	1.006	0.950	0.837	0.966
	0.426	0.319	0.118	0.179	0.656	0.453	0.145	0.106

**Table 7.** Results under the SSL distribution. For each non-normality pair  $(\tau, q)$ , dimensions of the data frame  $(p, n)$  and coefficient  $\rho$ , the cells contain the MSEs given by method 1 (first row), method 2 (second row) and method 3 (third row).

$(p, n)$	(4, 60)		(4, 120)		(8, 120)		(8, 240)	
$\rho$	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75
$(\tau, q)$								
(2, 4)	0.817	0.854	0.963	1.104	1.406	1.553	0.893	1.528
	0.929	0.900	1.209	1.332	1.491	1.632	1.020	1.652
	0.732	0.796	0.735	0.748	1.144	1.039	0.741	0.933
(2, 10)	0.739	0.776	0.678	0.843	1.306	1.410	1.039	1.249
	0.851	0.844	1.026	1.279	1.458	1.595	1.423	1.625
	0.852	0.803	0.698	0.604	1.273	1.246	0.956	0.988
(5, 4)	0.347	0.667	0.225	0.567	1.243	1.510	1.078	1.356
	0.516	0.841	0.462	0.763	1.402	1.634	1.351	1.582
	0.455	0.439	0.232	0.254	0.767	0.676	0.347	0.328
(5, 10)	0.413	0.650	0.125	0.289	1.057	1.284	0.598	0.738
	0.784	1.005	0.439	0.813	1.455	1.579	1.379	1.516
	0.655	0.648	0.162	0.200	0.804	0.889	0.392	0.293
(10, 4)	0.383	0.720	0.392	0.463	1.209	1.439	1.104	1.275
	0.574	0.941	0.735	0.830	1.396	1.564	1.420	1.572
	0.401	0.332	0.162	0.103	0.565	0.574	0.260	0.160
(10, 10)	0.222	0.588	0.163	0.207	0.905	1.202	0.550	0.902
	0.524	1.004	0.657	0.826	1.379	1.598	1.391	1.500
	0.430	0.452	0.318	0.131	0.537	0.775	0.224	0.280

Under the ST distribution, the results of Table 6 report the following findings: when  $p = 4$  and  $\rho = 0.25$ , the estimations given by methods M1 and M3 show similar competitive accuracies provided that  $n = 60$ , whereas the superiority of M3 can be observed for the sample size  $n = 120$ , especially in the case of a heavier tail weight  $\nu = 5$ . When we take  $\rho = 0.75$ , the superiority of M3 becomes more remarkable, as M1 tends to get worse while M3 gains in accuracy, with the lowest MSE appearing when  $n = 120$  in the furthest scenario from the normal model  $(\tau, \nu) = (10, 5)$ . When we move to the higher dimension  $p = 8$ , the errors increase to values greater than 1 in many cases, specifically in the scenarios closer to the normality; we can also see that, as the asymmetry and tail weight increase, the accuracy of M1 and M3 improves, with the latter giving acceptable errors ( $MSE = 0.101, 0.107$ ) at the sample size  $n = 240$ .

For the SSL distribution, the general patterns reported for the case of the ST model come up once again: all the estimation methods give similar accuracies in the scenarios close to normality, M1 and M3 have a good performance as the asymmetry increases, with a better behavior of M3 for the heavier tail weight; note that this pattern becomes more remarkable for the larger coefficient  $\rho = 0.75$ . Nevertheless, in this case, we can observe the negative impact of the dimension on the performance of M1, and particularly M2, methods which in most cases yield very large MSEs. Unlike them, M3 gives very good estimations with the minimum errors ranging in the interval  $(0.150, 0.300)$ .

Finally, we would like to stress that the curse of dimensionality has a smaller impact on the performance of method M3 than some other simulation experiments, not reported here for the sake of space, have shown. Here, we provide an illustrative example: when we consider a ST vector with dimension  $p = 15$  and parametric settings given by  $\rho = 0.25$  and  $(\tau, \nu) = (10, 5)$ , the values of the MSE for the sample size  $n = 200$  are 1.587 (M1), 1.684 (M2) and 0.615 (M3).

## 5. Summary and Concluding Remarks

This paper has addressed the skewness-based PP problem when the underlying multivariate distribution of the input vector belongs to the family of SMSN distributions. We have presented skewness PP as an eigenvector problem based on either the third order cumulant matrix or well-known scatter matrices of the model; the connection between the resulting dominant eigenvectors and the shape vector that regulates the asymmetry of the multivariate model has been established, such a connection also sheds light on the interpretability of PP and the ICS method when approached from the parametric framework. The theoretical results point out three methods for estimating the skewness PP direction; the application to artificial data has allowed us to evaluate their performance and has also provided useful insights for comparing them.

This paper has also gone through the role played by the shape vector of the model to assess the multivariate asymmetry in a directional way. However, it is not clear to us how the tail weight and shape vector of the SMSN distribution act together to account for the multivariate non-normality. We conjecture that the issue should be addressed by examining the role of both parameters as indicators of non-normality in accordance to well-defined multivariate stochastic orderings along the lines of previous work [58–60]. To the best of the authors' knowledge, the issue is still unexplored and may deserve future theoretical research to achieve a deeper understanding about their role as non-normality indicators in the SMSN family.

**Author Contributions:** Conceptualization, J.M.A. and H.N.; methodology, J.M.A. and H.N.; validation, J.M.A. and H.N.; formal analysis, J.M.A. and H.N.; investigation, J.M.A. and H.N.; writing—original draft preparation, J.M.A. and H.N.; writing—review and editing, J.M.A. and H.N.; visualization, J.M.A. and H.N.; funding acquisition, J.M.A. and H.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Universidad Nacional de Educación a Distancia (UNED). Spain.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors wish to thank the reviewers for their helpful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kruskal, J.B. Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new "index of condensation". In *Statistical Computation*; Elsevier: Amsterdam, The Netherlands, 1969; pp. 427–440.
2. Friedman, J.H.; Tukey, J.W. A Projection Pursuit Algorithm for Exploratory Data Analysis. *IEEE Trans. Comput.* **1974**, *23*, 881–890. [[CrossRef](#)]
3. Huber, P.J. Projection Pursuit. *Ann. Stat.* **1985**, *13*, 435–475. [[CrossRef](#)]
4. Jones, M.C.; Sibson, R. What is Projection Pursuit? *J. R. Stat. Society. Ser. A* **1987**, *150*, 1–37. [[CrossRef](#)]
5. Friedman, J.H. Exploratory Projection Pursuit. *J. Am. Stat. Assoc.* **1987**, *82*, 249–266. [[CrossRef](#)]
6. Hall, P. On Polynomial-Based Projection Indices for Exploratory Projection Pursuit. *Ann. Stat.* **1989**, *17*, 589–605. [[CrossRef](#)]
7. Cook, D.; Buja, A.; Cabrera, J. Projection Pursuit Indexes Based on Orthonormal Function Expansions. *J. Comput. Graph. Stat.* **1993**, *2*, 225–250. [[CrossRef](#)]
8. Peña, D.; Prieto, F.J. Cluster Identification Using Projections. *J. Am. Stat. Assoc.* **2001**, *96*, 1433–1445. [[CrossRef](#)]
9. Rodriguez Martinez, E.; Goulermas, J.; Mu, T.; Ralph, J. Automatic Induction of Projection Pursuit Indices. *Neural Netw. IEEE Trans.* **2010**, *21*, 1281–1295. [[CrossRef](#)]
10. Hou, S.; Wentzell, P. Re-centered kurtosis as a projection pursuit index for multivariate data analysis. *J. Chemom.* **2014**, *28*, 370–384. [[CrossRef](#)]
11. Loperfido, N. Skewness-Based Projection Pursuit: A Computational Approach. *Comput. Stat. Data Anal.* **2018**, *120*, 42–57. [[CrossRef](#)]



12. Loperfido, N. Kurtosis-based projection pursuit for outlier detection in financial time series. *Eur. J. Financ.* **2020**, *26*, 142–164. [[CrossRef](#)]
13. Posse, C. Projection pursuit exploratory data analysis. *Comput. Stat. Data Anal.* **1995**, *20*, 669–687. [[CrossRef](#)]
14. De Lathauwer, L.; De Moor, B.; Vandewalle, J. On the best rank-1 and rank- $(R_1, R_2, \dots, R_N)$  approximation of higher-order tensor. *SIAM J. Matrix Anal. Appl.* **2000**, *21*, 1324–1342. [[CrossRef](#)]
15. Kofidis, E.; Regalia, P.A. On the Best Rank-1 Approximation of Higher-Order Supersymmetric Tensors. *SIAM J. Matrix Anal. Appl.* **2002**, *23*, 863–884. [[CrossRef](#)]
16. Loperfido, N. A new kurtosis matrix, with statistical applications. *Linear Algebra Its Appl.* **2017**, *512*, 1–17. [[CrossRef](#)]
17. Franceschini, C.; Loperfido, N. An Algorithm for Finding Projections with Extreme Kurtosis. In *Studies in Theoretical and Applied Statistics*; Perna, C., Pratesi, M., Ruiz-Gazen, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 61–70.
18. Loperfido, N. Canonical transformations of skew-normal variates. *TEST* **2010**, *19*, 146–165. [[CrossRef](#)]
19. Arevalillo, J.M.; Navarro, H. A note on the direction maximizing skewness in multivariate skew-t vectors. *Stat. Probab. Lett.* **2015**, *96*, 328–332. [[CrossRef](#)]
20. Arevalillo, J.M.; Navarro, H. Skewness-Kurtosis Model-Based Projection Pursuit with Application to Summarizing Gene Expression Data. *Mathematics* **2021**, *9*, 954. [[CrossRef](#)]
21. Kim, H.M.; Kim, C. Moments of scale mixtures of skew-normal distributions and their quadratic forms. *Commun. Stat. Theory Methods* **2017**, *46*, 1117–1126. [[CrossRef](#)]
22. Arevalillo, J.M.; Navarro, H. Data projections by skewness maximization under scale mixtures of skew-normal vectors. *Adv. Data Anal. Classif.* **2020**, *14*, 435–461. [[CrossRef](#)]
23. Capitanio, A. On the canonical form of scale mixtures of skew-normal distributions. *Statistica* **2020**, *80*, 145–160.
24. Malkovich, J.F.; Afifi, A.A. On Tests for Multivariate Normality. *J. Am. Stat. Assoc.* **1973**, *68*, 176–179. [[CrossRef](#)]
25. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602. [[CrossRef](#)]
26. Azzalini, A. The Skew-normal Distribution and Related Multivariate Families. *Scand. J. Stat.* **2005**, *32*, 159–188. [[CrossRef](#)]
27. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler Divergence Measure for Multivariate Skew-Normal Distributions. *Entropy* **2012**, *14*, 1606–1626. [[CrossRef](#)]
28. Balakrishnan, N.; Scarpa, B. Multivariate measures of skewness for the skew-normal distribution. *J. Multivar. Anal.* **2012**, *104*, 73–87. [[CrossRef](#)]
29. Balakrishnan, N.; Capitanio, A.; Scarpa, B. A test for multivariate skew-normality based on its canonical form. *J. Multivar. Anal.* **2014**, *128*, 19–32. [[CrossRef](#)]
30. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; IMS monographs; Cambridge University Press: Cambridge, UK, 2014.
31. González-Farías, G.; Domínguez-Molina, A.; Gupta, A.K. Additive properties of skew normal random vectors. *J. Stat. Plan. Inference* **2004**, *126*, 521–534. [[CrossRef](#)]
32. González-Farías, G.; Domínguez-Molina, A.; Gupta, A.K. The closed skew-normal distribution. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; Chapter 2; pp. 25–42.
33. Loperfido, N. Generalized skew-normal distributions. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; Chapter 4, pp. 65–80.
34. Lachos, V.H.; Labra, F.V.; Bolfarine, H.; Ghosh, P. Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics* **2010**, *44*, 541–556. [[CrossRef](#)]
35. Vilca, F.; Balakrishnan, N.; Zeller, C.B. Multivariate Skew-Normal Generalized Hyperbolic distribution and its properties. *J. Multivar. Anal.* **2014**, *128*, 73–85. [[CrossRef](#)]
36. Kahrari, F.; Rezaei, M.; Yousefzadeh, F.; Arellano-Valle, R. On the multivariate skew-normal-Cauchy distribution. *Stat. Probab. Lett.* **2016**, *117*, 80–88. [[CrossRef](#)]
37. Kahrari, F.; Arellano-Valle, R.; Rezaei, M.; Yousefzadeh, F. Scale mixtures of skew-normal-Cauchy distributions. *Stat. Probab. Lett.* **2017**, *126*, 1–6. [[CrossRef](#)]
38. Arellano-Valle, R.B.; Ferreira, C.S.; Genton, M.G. Scale and shape mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **2018**, *166*, 98–110. [[CrossRef](#)]
39. Ogasawara, H. A non-recursive formula for various moments of the multivariate normal distribution with sectional truncation. *J. Multivar. Anal.* **2021**, *183*, 104729. [[CrossRef](#)]
40. Azzalini, A.; Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [[CrossRef](#)]
41. Branco, M.D.; Dey, D.K. A General Class of Multivariate Skew-Elliptical Distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [[CrossRef](#)]
42. Kim, H.M. A note on scale mixtures of skew normal distribution. *Stat. Probab. Lett.* **2008**, *78*, 1694–1701. [[CrossRef](#)]
43. Basso, R.M.; Lachos, V.H.; Cabral, C.R.B.; Ghosh, P. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Comput. Stat. Data Anal.* **2010**, *54*, 2926–2941. [[CrossRef](#)]
44. Zeller, C.B.; Cabral, C.R.B.; Lachos, V.H. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *TEST* **2016**, *25*, 375–396. [[CrossRef](#)]

45. Lachos, V.H.; Garay, A.M.; Cabral, C.R.B. Moments of truncated scale mixtures of skew-normal distributions. *Braz. J. Probab. Stat.* **2020**, *34*, 478–494. [[CrossRef](#)]
46. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 367–389. [[CrossRef](#)]
47. Wang, J.; Genton, M.G. The multivariate skew-slash distribution. *J. Stat. Plan. Inference* **2006**, *136*, 209–220. [[CrossRef](#)]
48. Loperfido, N. Skewness and the linear discriminant function. *Stat. Probab. Lett.* **2013**, *83*, 93–99. [[CrossRef](#)]
49. Kollo, T.; von Rosen, D. *Advanced Multivariate Statistics with Matrices*; Mathematics and Its Applications; Springer: Dordrecht, The Netherlands, 2005.
50. Christiansen, M.C.; Loperfido, N. Improved Approximation of the Sum of Random Vectors by the Skew Normal Distribution. *J. Appl. Probab.* **2014**, *51*, 466–482. [[CrossRef](#)]
51. Qi, L. Rank and eigenvalues of a supersymmetric tensor, the multivariate homogeneous polynomial and the algebraic hypersurface it defines. *J. Symb. Comput.* **2006**, *41*, 1309–1327. [[CrossRef](#)]
52. Cardoso, J.F. Source separation using higher order moments. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Glasgow, UK, 23–26 May 1989; pp. 2109–2112.
53. Tyler, D.E.; Critchley, F.; Dümbgen, L.; Oja, H. Invariant co-ordinate selection. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 549–592. [[CrossRef](#)]
54. Kent, J.T. Discussion on the paper by Tyler, Critchley, Dümbgen & Oja: Invariant co-ordinate selection. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 575–576.
55. Peña, D.; Prieto, F.J.; Viladomat, J. Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. *J. Multivar. Anal.* **2010**, *101*, 1995–2007. [[CrossRef](#)]
56. Alashwali, F.; Kent, J.T. The use of a common location measure in the invariant coordinate selection and projection pursuit. *J. Multivar. Anal.* **2016**, *152*, 145–161. [[CrossRef](#)]
57. Nordhausen, K.; Oja, H.; Tyler, D.E. Tools for Exploring Multivariate Data: The Package ICS. *J. Stat. Softw.* **2008**, *28*, 1–31. [[CrossRef](#)]
58. Wang, J. A family of kurtosis orderings for multivariate distributions. *J. Multivar. Anal.* **2009**, *100*, 509–517. [[CrossRef](#)]
59. Arevalillo, J.M.; Navarro, H. A study of the effect of kurtosis on discriminant analysis under elliptical populations. *J. Multivar. Anal.* **2012**, *107*, 53–63. [[CrossRef](#)]
60. Arevalillo, J.M.; Navarro, H. A stochastic ordering based on the canonical transformation of skew-normal vectors. *TEST* **2019**, *28*, 475–498. [[CrossRef](#)]