Journal of Multivariate Analysis 101 (2010) 1434–1444

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On the singularity of multivariate skew-symmetric models

Christophe Ley [a], Davy Paindaveine [b,*]

[a] E.C.A.R.E.S. and Département de Mathématique, Université Libre de Bruxelles, Boulevard du Triomphe, CP210, 1050 Brussels, Belgium

[b] E.C.A.R.E.S. and Département de Mathématique, Université Libre de Bruxelles, Avenue Roosevelt, 50, CP114, 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 6 May 2009
Available online 10 November 2009

AMS 2000 subject classifications:
60E05
62H05
62E10

Keywords:
Characterization property
Profile likelihood
Reparameterization
Singular Fisher information matrix
Skewness
Skew-normal distributions

ABSTRACT

In recent years, the skew-normal models introduced by Azzalini (1985) [1] – and their multivariate generalizations from Azzalini and Dalla Valle (1996) [4] – have enjoyed an amazing success, although an important literature has reported that they exhibit, in the vicinity of symmetry, singular Fisher information matrices and stationary points in the profile log-likelihood function for skewness, with the usual unpleasant consequences for inference. It has been shown (DiCiccio and Monti (2004) [23], DiCiccio and Monti (2009) [24] and Gómez et al. (2007) [25]) that these singularities, in some specific parametric extensions of skew-normal models (such as the classes of skew-*t* or skew-exponential power distributions), appear at skew-normal distributions only. Yet, an important question remains open: in broader semiparametric models of skewed distributions (such as the general skew-symmetric and skew-elliptical ones), which symmetric kernels lead to such singularities? The present paper provides an answer to this question. In very general (possibly multivariate) skew-symmetric models, we characterize, for each possible value of the rank of Fisher information matrices, the class of symmetric kernels achieving the corresponding rank. Our results show that, for strictly multivariate skew-symmetric models, not only Gaussian kernels yield singular Fisher information matrices. In contrast, we prove that systematic stationary points in the profile log-likelihood functions are obtained for (multi)normal kernels only. Finally, we also discuss the implications of such singularities on inference.

1. Introduction

Azzalini [1] introduced the so-called skew-normal model, which embeds univariate normal distributions into a flexible parametric class of (possibly) skewed distributions. More formally, a random variable X is said to be skew-normal with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_0^+$ and skewness parameter $\delta \in \mathbb{R}$ if it admits the pdf

$$x \mapsto 2\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\delta\left(\frac{x-\mu}{\sigma}\right)\right), \quad x \in \mathbb{R}, \tag{1.1}$$

where ϕ and Φ respectively denote the pdf and cdf of the standard normal distribution. A first intensive study of these distributions was provided by Azzalini himself in [1,2]. Besides quite appealing and nice stochastic properties, two closely related inferential problems appeared when dealing with such densities: at $\delta = 0$, corresponding to the symmetric situation, (i) the profile log-likelihood function for δ always admits a stationary point, and consequently, (ii) the Fisher information matrix for the three parameters in (1.1) is singular (typically, with rank 2). Thus, the skew-normal distributions happen

* Corresponding author.
E-mail addresses: chrisley@ulb.ac.be (C. Ley), dpaindav@ulb.ac.be (D. Paindaveine).
URLs: http://homepages.ulb.ac.be/~chrisley (C. Ley), http://homepages.ulb.ac.be/~dpaindav (D. Paindaveine).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.10.008

to be problematic from an inferential point of view, since such a singularity is incompatible with the assumptions needed for the standard asymptotic behavior of the maximum likelihood estimators. A situation of this kind has been studied in detail by Rotnitzky et al. [3], where it is shown that, in cases (as above) where the $p \times p$ Fisher information matrix has rank $p-1$, one component of the parameter cannot be estimated at the usual root-n rate, but only at a slower rate, and that the corresponding limiting distribution may be bimodal.

Despite these inferential drawbacks, the two papers by Azzalini had some sort of pioneering effect. In [4], Azzalini and Dalla Valle extended skew-normal distributions to the multivariate setup, while [5] studied further probabilistic properties of (multivariate) skew-normal distributions and investigated more statistical aspects. The growing interest for this flexible class of distributions led to a number of developments in various fields. For example, [6] applied the skew-normal model to psychometric real data, whereas [7] showed the connections with the problem of the selection of a sample.

The success of skew-normal distributions also gave rise to numerous further generalizations of the density in (1.1). To cite a few, [8] introduced a multivariate skew-Cauchy distribution, [9–11] proposed multivariate skew-t distributions, while [5,10] defined multivariate skew-elliptical distributions. In [12], Genton and Loperfido extended the latter into the so-called generalized skew-elliptical distributions, where asymmetry enters densities through very general skewing functions; most of the pre-cited examples are part of their broad framework. Finally, [13], in a further effort to introduce very general skew-symmetric distributions, proposed a class that is broader than the one from [12]: the skew-symmetric distributions defined there have a pdf of the form

$$x \mapsto 2|\Sigma|^{-1/2} f(\Sigma^{-1/2}(x-\mu))\, \Pi(\Sigma^{-1/2}(x-\mu)), \quad x \in \mathbb{R}^k,$$

where $\mu \in \mathbb{R}^k$ is a location parameter, $\Sigma \in \mathbb{R}^{k\times k}$ is a symmetric and positive definite scatter parameter, f (the *symmetric kernel*) is a centrally symmetric pdf (i.e., a pdf such that $f(-x) = f(x) \forall x \in \mathbb{R}^k$), and where the mapping $\Pi : \mathbb{R}^k \to [0, 1]$ satisfies $\Pi(-x) = 1 - \Pi(x) \forall x \in \mathbb{R}^k$. A particular subclass of these skew-symmetric densities is the class of the so-called *flexible skew-symmetric densities* [14], for which the skewing function Π takes the form of an arbitrary symmetric cdf evaluated at odd polynomials; see [14] or Section 5 for details.

Besides its generality, the class of multivariate skew-symmetric distributions is of high interest in diverse fields of statistics. Since the distribution of quadratic forms in skew-symmetric random vectors does not depend on the skewing function Π (see [13]), multivariate skew-symmetric distributions have potential applications in most domains where inference is based on quadratic statistics. In particular, they are of high relevance in multivariate analysis, spatial statistics, and time series, where the corresponding natural quadratic statistics are Mahalanobis distances, sample variograms, and sample autocovariances, respectively. Another advantage of those skewed distributions lies in their high flexibility, qualifying them as tools for shape analysis or for modeling random effects in linear mixed models. For extensive reviews about models of skewed distributions and related topics, we refer to the recent monograph [15] and to the review papers [16,17].

Parallel to the numerous extensions of skew-normal models described above, the aforementioned issue related to singularity of Fisher information matrices in the vicinity of symmetry has also attracted much attention. Besides [1] itself, this was investigated in [5,18–20] and [21]. Alternative parameterizations were proposed in [1,19] (for the univariate setup) and in [18] (for the multivariate one) in order to get rid of this singularity; the latter paper even is entirely dedicated to the so-called centered parameterization. The singularity result for the univariate skew-normal was extended in [22] by establishing the singularity of the Fisher information matrix, still in the vicinity of symmetry, for skewed distributions obtained by replacing, in (1.1), the standard normal cdf Φ with an arbitrary cdf H satisfying some mild regularity conditions.

All these papers share a common point: they show that, if a normal kernel ϕ is used, in some specific class of skewed densities similar to (1.1), then the Fisher information matrix is singular at $\delta = 0$. To the best of our knowledge, the only papers where such a result is turned into an "iff" statement are [23,24] and [25], where it is shown that, in the classes of skew-exponential power and skew-t distributions (which both contain the skew-normal as a special case), Fisher information matrices – in the vicinity of symmetry – actually are singular at skew-normal distributions only. These two classes, however, only constitute very specific *parametric* models of *univariate* skewed distributions, and a natural question is whether such "iff" results extend to much broader *semiparametric* models of *possibly multivariate* skewed distributions. On the basis of numerical work, Azzalini and Genton [26] conjecture that, among the class of multivariate skew-t distributions, only the skew-normal ones – which are obtained by letting the underlying number of degrees of freedom go to infinity – suffer from singular information matrices. These various findings naturally lead to the following general conjecture.

Conjecture 1.1. *In broad classes of semiparametric (possibly multivariate) skew-symmetric distributions, the only symmetric kernels leading, in the vicinity of symmetry, to singular Fisher information matrices are the (multi)normal ones.*

One of the main goals of this paper is to investigate the validity of this conjecture in the various models described above, and to determine, in cases where the conjecture fails to hold, the classes of symmetric kernels leading to singular information matrices.

Another problem, which is closely related (but not equivalent, as our results will show) to the one considered above, concerns the existence of a stationary point, in the vicinity of symmetry, of the skewness profile log-likelihood function for skew-normal models. This issue has also been extensively discussed; see, e.g., [5,18,26]. Of particular interest is the recent contribution of Azzalini and Genton [26]. Besides generalizing the results of [22] to the multivariate setup and showing that, for the general case of the (univariate) flexible skew-symmetric distributions introduced in [14], the Fisher information matrix is singular for normal kernels in the vicinity of symmetry, they address the "iff" problem. More precisely, they provide a heuristic argument showing that, in the univariate setup, the profile log-likelihood function should systematically present a stationary point at $\delta = 0$ for normal kernels only. It is also conjectured there that this result should carry on in higher dimensions.

Conjecture 1.2 (*Azzalini and Genton [26]*). *In broad classes of semiparametric (possibly multivariate) skew-symmetric distributions, the only symmetric kernels leading, for any sample of fixed size n (≥ 3) and in the vicinity of symmetry, to a stationary point of the profile log-likelihood function for skewness are the (multi)normal ones.*

Azzalini and Genton [26] clearly indicate that their proof (which is restricted to the univariate case) is not a formal one. Most importantly, they express the need for a clarification related to the results in Conjectures 1.1 and 1.2, which, they say, *is important to remove or at least alleviate the necessity of an alternative parameterization*. Accordingly, the second main goal of the present paper is to prove their conjecture.

The outline of the paper is as follows. In Section 2, we describe the class of skew-symmetric models we consider in what follows, and we solve Conjecture 1.1 by determining, for each possible value of the rank of the resulting Fisher information matrices (in the vicinity of symmetry), the class of symmetric kernels achieving the corresponding rank. We interpret the results and consider several important particular cases. In Section 3, we discuss implications of our results on inference in skew-symmetric models, with focus on optimal symmetry testing in relation with Le Cam's theory of asymptotic experiments. Section 4 shows that our results extend to very broad skew-symmetric models. We then turn to Conjecture 1.2 in Section 5. Finally, an Appendix collects the proofs.

2. Singularity of Fisher information matrices

As described in Section 1, there exist many distinct generalizations of the univariate skew-normal distributions described in [1]. In this section, we will drop the scale/scatter parameter and focus on a fixed (yet quite general) class of skew-symmetric densities (Section 4 will then restore the scale/scatter parameter and extend our results to even more general classes of skew-symmetric distributions). More precisely, we consider densities of the form

$$x \mapsto f^{\Pi}_{\mu,\delta}(x) := 2f(x-\mu)\,\Pi(\delta'(x-\mu)), \quad x \in \mathbb{R}^k, \tag{2.2}$$

where $\mu \in \mathbb{R}^k$ is a location parameter and $\delta \in \mathbb{R}^k$ is a skewness parameter, while f and Π satisfy

Assumption A. (i) The pdf f belongs to the collection $\mathcal{F}$ of a.e. positive, centrally symmetric ($f(-x) = f(x)$ for all $x \in \mathbb{R}^k$), and continuously differentiable densities for which both the covariance matrix $\Sigma_f := \int_{\mathbb{R}^k} xx'f(x)\mathrm{d}x$ and the Fisher information matrix (for location) $\mathcal{I}_f := \int_{\mathbb{R}^k} \varphi_f(x)(\varphi_f(x))'f(x)\,\mathrm{d}x$ (with $\varphi_f := -\nabla f/f$) are finite and invertible. (ii) The skewing function $\Pi : \mathbb{R} \to [0,1]$ is a continuously differentiable function that satisfies $\Pi(-x) = 1 - \Pi(x)$ for all $x \in \mathbb{R}$, and $\Pi'(0) \neq 0$.

It is common practice to use for Π the cdf of a symmetric (about the origin) univariate random variable, but we here work in the more general setup where Π might fail to be a cdf. Note that, as in [13], Π could very well depend on f, but since such a dependence will have no impact on our results, the skewing function Π will be regarded as fixed in what follows, and we will only stress dependence on $f \in \mathcal{F}$ for scores and Fisher information matrices.

Under Assumption A, the scores for location and skewness, in the vicinity of symmetry (that is, at any $(\mu', \delta')' = (\mu', 0')'$), are the quantities $m_{f;\mu}(x)$ and $d_{f;\mu}(x)$, respectively, in

$$\ell_{f;\mu}(x) := \begin{pmatrix} m_{f;\mu}(x) \\ d_{f;\mu}(x) \end{pmatrix} := \begin{pmatrix} (\nabla_\mu \log f^{\Pi}_{\mu,\delta}(x))_{|(\mu,\delta)=(\mu,0)} \\ (\nabla_\delta \log f^{\Pi}_{\mu,\delta}(x))_{|(\mu,\delta)=(\mu,0)} \end{pmatrix} = \begin{pmatrix} \varphi_f(x-\mu) \\ 2\Pi'(0)(x-\mu) \end{pmatrix}$$

(the factor 2 in the δ-score follows from the fact that $\Pi(0) = 1/2$). The corresponding Fisher information matrix is then given by $\Gamma_f = \int_{\mathbb{R}^k} \ell_{f;\mu}(x)\ell'_{f;\mu}(x)f(x-\mu)\mathrm{d}x$, which naturally partitions into

$$\Gamma_f := \begin{pmatrix} \Gamma_{f;\mu\mu} & \Gamma_{f;\mu\delta} \\ \Gamma_{f;\delta\mu} & \Gamma_{f;\delta\delta} \end{pmatrix} = \begin{pmatrix} \mathcal{I}_f & 2\Pi'(0)I_k \\ 2\Pi'(0)I_k & 4(\Pi'(0))^2\Sigma_f \end{pmatrix}, \tag{2.3}$$

where I_k stands for the k-dimensional identity matrix. The expression for $\Gamma_{f;\mu\delta} = \Gamma'_{f;\delta\mu}$ follows by integrating by parts in $\int_{\mathbb{R}^k} \varphi_f(x)\,x'f(x)\mathrm{d}x$. Note that, in view of (2.3), finiteness of $\mathcal{I}_f$ and Σ_f in Assumption A is necessary. Also, note that Γ_f does not depend on μ, hence the notation.

As mentioned in Section 1, one of the main goals of this paper is to investigate for which f the information matrix Γ_f is singular. Of course, we are interested in a possible singularity arising from the presence of skewness in the model. In particular, we do not want to investigate singularities coming from the location part of the model alone, which explains why Assumption A imposes the requirement (which is very standard in any location model) that the Fisher information for location $\mathcal{I}_f (= \Gamma_{f;\mu\mu})$ has full rank. Since $|\Gamma_f| = |\Gamma_{f;\mu\mu}||\Gamma_{f;\delta\delta.\mu}|$, where we let $\Gamma_{f;\delta\delta.\mu} := \Gamma_{f;\delta\delta} - \Gamma_{f;\delta\mu}\Gamma^{-1}_{f;\mu\mu}\Gamma_{f;\mu\delta} =$

$4(\Pi'(0))^2(\Sigma_f - \mathcal{I}_f^{-1})$ (throughout, $|A|$ denotes the determinant of the matrix A), it is clear that, under Assumption A, $\Gamma_{f;\delta\delta.\mu}$ is singular iff Γ_f is, hence potentially plays an important role in what follows.

We start our investigation of the possible singularity of Γ_f with the equivalence result of Lemma 2.1. Before stating that result, let us introduce the following notation, which we shall use in the proof of Lemma 2.1 (see the Appendix) and throughout the whole paper: for a given matrix A, we denote by $\ker(A)$ the kernel of A and by $\mathrm{Im}(A)$ its image (that is, the vector space spanned by the columns of A).

Lemma 2.1. *Let Assumption A hold and fix $m \in \{1, 2, \ldots, k\}$. Then* (i) $\mathrm{rank}(\Gamma_f) = 2k - m$ *iff* $\mathrm{rank}(\Gamma_{f;\delta\delta.\mu}) = k - m$, *and* (ii) $\mathrm{rank}(\Gamma_{f;\delta\delta.\mu}) = k - m$ *iff m is the largest integer $\ell \in \{1, 2, \ldots, k\}$ such that there exists a $k \times \ell$ matrix $V = (v_1, \ldots, v_\ell)$ with orthonormal columns satisfying $V'\Gamma_{f;\delta\delta.\mu}V = 0$.*

Define $d^*_{f;\mu}(x) := d_{f;\mu}(x) - \Gamma_{f;\delta\mu}\Gamma^{-1}_{f;\mu\mu}m_{f;\mu}(x) = 2\Pi'(0)[(x-\mu) - \mathcal{I}_f^{-1}\varphi_f(x-\mu)]$. Note that if X has pdf $f(.-\mu)$, we have $\mathrm{E}[d^*_{f;\mu}(X)] = 0$ and $\mathrm{Var}[d^*_{f;\mu}(X)] = \Gamma_{f;\delta\delta.\mu}$. Hence, the following result is a direct consequence of the previous lemma.

Lemma 2.2. *Let Assumption A hold and fix $m \in \{1, 2, \ldots, k\}$. Then $\mathrm{rank}(\Gamma_f) = 2k - m$ iff m is the largest integer $\ell \in \{1, 2, \ldots, k\}$ such that there exists a $k \times \ell$ matrix $V = (v_1, \ldots, v_\ell)$ with orthonormal columns satisfying $V'd^*_{f;\mu}(X) = 0$ a.s., where X has pdf $f(.-\mu)$.*

In order to fully exploit the necessary condition of Lemma 2.2, we translate it into a more analytical, easier to handle, setup. Let Assumption A hold, assume $\mathrm{rank}(\Gamma_f) = 2k - m$ for some $m \in \{1, 2, \ldots, k\}$, and consider a matrix $V = (v_1, \ldots, v_m)$ with orthonormal columns satisfying $V'd^*_{f;\mu}(X) = 0$ a.s. when X has pdf $f(.-\mu)$, or equivalently, such that $V'(x - \mathcal{I}_f^{-1}\varphi_f(x)) = 0$ a.e. in $\mathbb{R}^k$. Letting $g = \log f$ and $W = \mathcal{I}_f^{-1}$, this can be rewritten as

$$V'(x + W\nabla g(x)) = 0. \tag{2.4}$$

The problem of identifying the densities $f \in \mathcal{F}$ leading to singular Fisher information matrices Γ_f has been clearly transposed into a first-order partial differential equation problem, where the number of equations (namely, m) is determined by the rank of Γ_f. The following lemma, which is proved in the Appendix, provides the general solution of (2.4).

Lemma 2.3. *Let V and W be full-rank matrices, with dimensions $k \times m$ and $k \times k$, respectively ($m \leq k$). Then* (i) *Eq. (2.4) admits a solution g iff $V'(W - W')V = 0$;* (ii) *under the condition $V'(W - W')V = 0$, the general solution of (2.4) is*

$$g(x) = -\frac{1}{2}x'P_1W^{-1}P_1x - x'P_1W^{-1}P_2x + h(P_2x), \tag{2.5}$$

where h is an arbitrary function defined on $\ker(V'W)$ and where $P_1 := W'V(V'WW'V)^{-1}V'W$ and $P_2 := I_k - P_1$ are the matrices of the orthogonal projections from $\mathbb{R}^k$ onto $\mathrm{Im}(W'V)$ and its orthogonal complement, respectively.

We are now ready to state the following theorem, which is the main result of this section (see the Appendix for a proof).

Theorem 2.1. *Let Assumption A hold and fix $m \in \{1, 2, \ldots, k\}$. Denote by Σ_f and $\mathcal{I}_f$ the covariance matrix and Fisher information matrix (for location) associated with f, respectively. Then, Γ_f is singular with rank $2k - m$ iff*

$$f(x) = h(Px)\exp\Big[-\frac{1}{2}x'\mathcal{I}_f x\Big], \tag{2.6}$$

where P is the matrix of the orthogonal projection from $\mathbb{R}^k$ onto the $(k-m)$-dimensional subspace $\mathrm{Im}(\mathcal{I}_f\Sigma_f - I_k)$ and where h is an arbitrary function from $\mathrm{Im}(P) \subset \mathbb{R}^k$ to $\mathbb{R}^+$ such that the mapping $x \mapsto h(Px)\exp[-\frac{1}{2}x'\mathcal{I}_f x]$ belongs to $\mathcal{F}$ and has covariance matrix Σ_f and Fisher information matrix (for location) $\mathcal{I}_f$.

In the most singular case ($m = k$), we have $P = 0$, hence $h(Px) = h(0)$ for all x, so that Theorem 2.1 states that f must be the density of the k-variate normal distribution with mean zero and covariance matrix $\Sigma_f = \mathcal{I}_f^{-1}$ (recall that Lemma 2.1 indeed shows that, for $m = k$, the matrix $\Gamma_{f;\delta\delta.\mu} = 4(\Pi'(0))^2(\Sigma_f - \mathcal{I}_f^{-1})$ has rank zero). It is straightforward to check that, vice versa, if f is the density of a k-variate centered normal distribution with some positive definite covariance matrix, then Γ_f is singular with rank k. Theorem 2.1 therefore reveals that the only skew-symmetric distributions leading to such a maximal singularity ($m = k$) are those based on (multi)normal kernels, explaining why these distributions are particularly hard to deal with in inferential problems; see Section 3.

In the univariate setup ($k = 1$), Fisher information matrices – under Assumption A – are singular iff $m = k$, hence singularity occurs iff f is normal, which then proves Conjecture 1.1 for the whole class of skew-symmetric distributions. In the multivariate setup ($k > 1$), however, the situation is more complicated, as singularity with $m < k$ will *not* lead to multinormal distributions. Such cases call for some clearer interpretation, provided in the following result (see the Appendix for a proof).

Theorem 2.2. *Let X be a k-vector admitting pdf f, and denote by Σ_f and $\mathcal{I}_f$ the covariance matrix and Fisher information matrix (for location) associated with f, respectively. Define Y through $X = BY = (B_1|B_2)(Y_1', Y_2')'$, where B_2 is a full-rank $k \times (k-m)$ matrix such that $\mathrm{Im}(B_2) = \mathrm{Im}(\mathcal{I}_f \Sigma_f - I_k)$ and B_1 is such that B is invertible. Let Assumption A hold, fix $m \in \{1, 2, \ldots, k-1\}$, and partition the matrix $\mathcal{I}_f^Y := B'\mathcal{I}_f B$ into the blocks $\mathcal{I}_{f;11}^Y, \mathcal{I}_{f;12}^Y, \mathcal{I}_{f;21}^Y$, and $\mathcal{I}_{f;22}^Y$ with dimensions $m \times m$, $m \times (k-m)$, $(k-m) \times m$, and $(k-m) \times (k-m)$, respectively. Then, if Γ_f is singular with rank $2k - m$, we have that*

$$\text{(i)}\quad Y_1|Y_2 = y_2 \sim \mathcal{N}_m((\mathcal{I}_{f;11}^Y)^{-1}\mathcal{I}_{f;12}^Y y_2, (\mathcal{I}_{f;11}^Y)^{-1})$$

and (ii) *Y_2 is an arbitrary random vector such that the density f^Y of Y belongs to $\mathcal{F}$ and has covariance matrix $\Sigma_f^Y = B^{-1}\Sigma_f B'^{-1}$ and Fisher information matrix $\mathcal{I}_f^Y$.*

For the general class of skew-symmetric distributions considered above, Conjecture 1.1 thus holds in the univariate case only. A counterexample in the multivariate case, compatible with the distributions described in Theorem 2.2, is for instance obtained by considering the pdf f of a random vector $X = (X_1, \ldots, X_k)'$ with mutually independent marginals, where $X_1, \ldots, X_m$ (resp., $X_{m+1}, \ldots, X_k$) are standard Gaussian variables (resp., t_ν-distributed variables, with $\nu > 2$). It can easily be checked that Γ_f is then singular with rank $2k - m$, whereas, of course, X does not have a multinormal distribution.

In the class of skew-symmetric distributions, special attention has been paid to the family of generalized skew-elliptical distributions. These are obtained by restricting, in (2.2), to elliptically symmetric kernels f, that is, to kernels of the form

$$x \mapsto f(x) = |\Sigma|^{-1/2} f_1(\Sigma^{-1/2}x), \quad x \in \mathbb{R}^k,$$

where $x \mapsto f_1(x)$ is *spherically symmetric* (i.e., is a function of $\|x\|$ only) and Σ is some $k \times k$ symmetric and positive definite matrix; throughout, $A^{1/2}$, for a positive definite matrix A, denotes the symmetric square-root of A. The following result is a fairly direct consequence of our general results above (see the Appendix for a proof).

Theorem 2.3. *Let Assumption A hold, with the further assumption that the pdf f is elliptically symmetric, and fix $m \in \{1, 2, \ldots, k\}$. Then, Γ_f is singular iff f is the pdf of a (multi)normal distribution.*

This result shows that, in the class of generalized skew-elliptical distributions, only (multi)normal distributions yield a singular Fisher information matrix (which then has rank k), hence that Conjecture 1.1 holds in the class of generalized skew-elliptical densities. In that class, Fisher information matrices therefore either have maximal rank $2k$ or the lowest possible rank k. This is to be compared to the class of skew-symmetric distributions considered above where all intermediate rank values can be achieved.

3. Implications on inference

As mentioned in Section 1, the singularity of Fisher information matrices goes along with a certain number of inferential problems; see, e.g., [3,27]. We now illustrate this in the problem of testing for symmetry about an unspecified center, by using Le Cam's theory of asymptotic experiments (see [28]).

With the same notation as in Section 2, fix a couple (f, Π) that satisfies Assumption A, and denote by $\mathrm{P}_{f;\vartheta}^{(n)}$, $\vartheta = (\mu', \delta')'$, the hypothesis under which the observations $X_1, \ldots, X_n$ are i.i.d. with common density $f_{\mu,\delta}^{\Pi}$; see (2.2). Assume that, in the parametric model $\mathcal{P}_f^{(n)} := \{\mathrm{P}_{f;\vartheta}^{(n)} : \vartheta \in \mathbb{R}^{2k}\}$, we want to test the null hypothesis $\mathcal{H}_0 : \delta = 0$ under which the common density of the observations is *symmetric about an unspecified center* μ against the alternative $\mathcal{H}_1 : \delta \neq 0$.

By proceeding as in [29], it is easy to show that the family of distributions $\mathcal{P}_f^{(n)}$ is *uniformly locally asymptotically normal (ULAN)* in the vicinity of symmetry (that is, at any $\vartheta = (\mu', 0')'$), with *central sequence*

$$\Delta_f^{(n)}(\vartheta) = \begin{pmatrix} M_f^{(n)}(\vartheta) \\ D_f^{(n)}(\vartheta) \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} m_{f;\mu}(X_i) \\ d_{f;\mu}(X_i) \end{pmatrix}$$

and the same *Fisher information matrix* Γ_f as in (2.3). More precisely, this means that, for any $\vartheta^{(n)} = (\mu^{(n)\prime}, 0')' = \vartheta + O(n^{-1/2})$ and any bounded sequence $\tau^{(n)} = (\tau_1^{(n)\prime}, \tau_2^{(n)\prime})' \in \mathbb{R}^{2k}$, we have

$$\log\Big(d\mathrm{P}_{f;\vartheta^{(n)}+n^{-1/2}\tau^{(n)}}^{(n)} / d\mathrm{P}_{f;\vartheta^{(n)}}^{(n)}\Big) = \tau^{(n)\prime}\Delta_f^{(n)}(\vartheta^{(n)}) - \frac{1}{2}\,\tau^{(n)\prime}\,\Gamma_f\,\tau^{(n)} + o_{\mathrm{P}}(1)$$

and $\Delta_f^{(n)}(\vartheta^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma_f)$, both under $\mathrm{P}_{f;\vartheta^{(n)}}^{(n)}$ as $n \to \infty$.

Would the information be block-diagonal in the sense that $\Gamma_{f;\mu\delta} = 0 = \Gamma_{f;\delta\mu}'$, Le Cam optimal tests for $\mathcal{H}_0 : \delta = 0$, μ playing the role of an unspecified nuisance, would be based on the δ-part $D_f^{(n)}(\mu, 0)$ of the central sequence above (or more precisely, on statistics of the form $D_f^{(n)}(\hat{\mu}, 0)$ for some appropriate location estimator $\hat{\mu}$). In the present setup, however, the information is never block-diagonal (meaning that there is no f such that block-diagonality holds), which implies that a local perturbation of location has the same asymptotic impact on $D_f^{(n)}(\mu, 0)$ as some local perturbation of δ. This in turn implies that the performances of optimal tests for symmetry are affected by the non-specification of μ.

In this context where Γ_f is not block-diagonal, locally asymptotically optimal inference on δ, when μ remains unspecified, has to be based on the so-called *δ-efficient central sequence*

$$D_f^{*(n)}(\vartheta) := D_f^{(n)}(\vartheta) - \Gamma_{f;\delta\mu}\Gamma_{f;\mu\mu}^{-1}M_f^{(n)}(\vartheta).$$

Under $\mathrm{P}_{f;\vartheta}^{(n)}$, $D_f^{*(n)}(\vartheta)$ is clearly asymptotically (multi)normal with mean zero and covariance matrix $\Gamma_{f;\delta\delta.\mu}$. At asymptotic level α, the resulting Le Cam optimal test, $\phi_\alpha^{(n)}$ say, then rejects the null $\mathcal{H}_0$ whenever

$$\left[D_f^{*(n)}(\hat{\mu}, 0)\right]' \left(\Gamma_{f;\delta\delta.\mu}\right)^- D_f^{*(n)}(\hat{\mu}, 0)$$

exceeds the upper α-quantile of the chi-square distribution with ℓ degrees of freedom, where $\ell = \mathrm{rank}(\Gamma_{f;\delta\delta.\mu})$ and A^- stands for an arbitrary generalized inverse of the matrix A. If $\ell > 0$, asymptotic powers of $\phi_\alpha^{(n)}$, against local alternatives of the form $\vartheta^{(n)} = (\mu', n^{-1/2}\tau_2')'$, are given by $1 - \Psi_\ell(\tau_2'\Gamma_{f;\delta\delta.\mu}\tau_2)$, where Ψ_ℓ stands for the cdf of the chi-square distribution with ℓ degrees of freedom; if $\ell = 0$, the corresponding asymptotic powers are equal to the nominal level α.

It is intuitively clear that the more extreme the confounding between location and skewness parameters, the smaller the rank of Γ_f, and the poorer the performances achieved when testing symmetry about an unspecified center μ. The worst case is of course the one for which Γ_f would have rank k: the results of Section 2 indeed show that we would then have $\Gamma_{f;\delta\delta.\mu} = 0$, which would result into asymptotic local powers equal to the nominal level α, as for the trivial test. Thus we can state the following result.

Theorem 3.1. *Let Assumption A hold. Then, in the parametric family of skew-symmetric distributions $\mathcal{P}_f^{(n)}$, $\phi_\alpha^{(n)}$, the α-level Le Cam optimal test for symmetry about an unspecified center, has asymptotic local powers that are equal to α iff f is the pdf of a (multi)normal distribution.*

The result directly follows from Theorem 2.1 and the considerations above, hence the proof is omitted. To summarize, Theorem 3.1 states that, from an inferential point of view, the *skew-normal* distributions (as well as their extensions obtained by choosing a function Π different from the cdf of the standard normal distribution), which clearly are the most famous representatives of the class of skew-symmetric distributions, are the worst among that class when the focus lies on optimal testing for symmetry about an unknown center.

Finally, we note that it is sometimes (erroneously) thought that, since Γ_f in (2.3) does not depend on μ, the non-specification of μ will not have any cost when performing inference on δ. We stress that the loss of power associated with the non-specification of μ is entirely due to the fact that the Fisher information matrix Γ_f is not block-diagonal.

4. Possible extensions

The long-standing open problem of characterizing the symmetric kernels for which the resulting multivariate skew-symmetric distributions in (2.2) lead to singular Fisher information matrices in the vicinity of symmetry was solved in Section 2. Obviously, Gaussian kernels play a key role in the result, although they are not the unique ones leading to such a singularity. However, as shown in Theorem 2.2, each kernel leading to singularity has a "multinormal (conditional) component", whose dimension depends on the rank of the singular Fisher information matrix. Now, one might argue that, nice as they are, the results of Section 2 solve the problem of determining the kernels leading to singular Fisher information matrices for the class of multivariate skew-symmetric distributions in (2.2) only, whereas the problem makes sense for many other types of skew-symmetric distributions. The aim of this section therefore consists in showing that our results actually allow for solving the problem for much more general classes of densities than the one in (2.2).

We start with the case where the centrally symmetric pdf f in (2.2) is split into a properly standardized version of f and a scatter parameter Σ. More specifically, this consists in writing $f(x) := |\Sigma|^{-1/2} f_1(\Sigma^{-1/2}(x - \mu))$, where the scatter Σ is a symmetric and positive definite $k \times k$ matrix and the centrally symmetric pdf f_1 is standardized so that $\Sigma_{f_1} := \int_{\mathbb{R}^k} xx' f_1(x)\mathrm{d}x = I_k$; clearly, Σ is then the covariance matrix Σ_f of the corresponding distribution. Writing vech A for the vector obtained by stacking the upper-diagonal entries of a matrix A, the resulting skew-symmetric densities

$$x \mapsto 2\,|\Sigma|^{-1/2} f_1(\Sigma^{-1/2}(x - \mu))\,\Pi(\delta'(x - \mu)), \quad x \in \mathbb{R}^k, \tag{4.7}$$

are then indexed by the $K := 2k + k(k+1)/2$-dimensional parameter $(\mu', \delta', (\mathrm{vech}\,\Sigma)')'$. It is easy to check that, in the vicinity of symmetry, that is, at any parameter value $(\mu', \delta', (\mathrm{vech}\,\Sigma)')'$ with $\delta = 0$, the corresponding Fisher information matrix takes the form

$$\Gamma_{f_1}^{\mathrm{sc}} := \begin{pmatrix} \Gamma_{f_1;\mu\mu}^{\mathrm{sc}} & \Gamma_{f_1;\mu\delta}^{\mathrm{sc}} & 0 \\ \Gamma_{f_1;\delta\mu}^{\mathrm{sc}} & \Gamma_{f_1;\delta\delta}^{\mathrm{sc}} & 0 \\ 0 & 0 & \Gamma_{f_1;\Sigma\Sigma}^{\mathrm{sc}} \end{pmatrix} := \begin{pmatrix} G\Gamma_{f_1}G' & 0 \\ 0 & \Gamma_{f_1;\Sigma\Sigma}^{\mathrm{sc}} \end{pmatrix}, \tag{4.8}$$

where Γ_{f_1} is the matrix from (2.3) evaluated at $f = f_1$, and where

$$G := \begin{pmatrix} \Sigma^{-1/2} & 0 \\ 0 & \Sigma^{1/2} \end{pmatrix}$$

has full rank. The zero blocks in $\Gamma^{sc}_{f_1}$ result from symmetry arguments: at $\delta = 0$, the μ-score and δ-score are indeed antisymmetric in $x - \mu$, while the Σ-score is symmetric, so that the latter always is uncorrelated with the first two at any parameter value $(\mu', \delta', (\text{vech}\,\Sigma)')'$ with $\delta = 0$.

Under the very mild assumption that $\Gamma^{sc}_{f_1;\Sigma\Sigma}$ is of full rank (remember that we made similar assumptions on $\Gamma_{f;\mu\mu}$ and $\Gamma_{f;\delta\delta}$ in the previous sections), the structure of $\Gamma^{sc}_{f_1}$ in (4.8) clearly entails that the results of the previous sections extend naturally to the class of skew-symmetric densities in (4.7). For instance, if $\text{rank}(\Gamma^{sc}_{f_1}) = K - k$, then f_1 must be the pdf of the standard (multi)normal distribution (remember that $\Sigma_{f_1} = I_k$). Similarly, all intermediate situations of the form $\text{rank}(\Gamma^{sc}_{f_1}) = K - m$, with $m \in \{1, \ldots, k-1\}$, between this minimal rank case and the full-rank one, will give rise to a m-variate "multinormal (conditional) component" in f_1, in exactly the same fashion as in Theorem 2.2. Similarly, for generalized skew-elliptical densities (obtained when f_1 is spherically symmetric), Theorem 2.3 extends to this setup involving $(\text{vech}\,\Sigma)$ in Γ_f.

Another class of skew-symmetric densities of interest, which are again parameterized by a location parameter $\mu \in \mathbb{R}^k$, a skewness parameter $\delta \in \mathbb{R}^k$, and a scatter parameter $(\text{vech}\,\Sigma) \in \mathbb{R}^{k(k+1)/2}$, is the one associated with densities of the form

$$x \mapsto 2\,|\Sigma|^{-1/2} f_1(\Sigma^{-1/2}(x-\mu))\,\Pi(\delta'\Sigma^{-1/2}(x-\mu)), \quad x \in \mathbb{R}^k, \tag{4.9}$$

where the centrally symmetric pdf f_1 is still standardized so that $\Sigma_{f_1} = I_k$. If f_1 is spherically symmetric, this falls again under the class of generalized skew-elliptical densities. It is easy to check that the resulting Fisher information matrix, in the vicinity of symmetry, can be obtained by substituting

$$\tilde{G} := \begin{pmatrix} \Sigma^{-1/2} & 0 \\ 0 & I_k \end{pmatrix}$$

for G in (4.8). Since $\tilde{G}$ is also of full rank, we conclude that our results similarly apply for such skew-symmetric densities. Rather than restating all results, we only state that, in the particular case where f_1 is spherically symmetric (in (4.9)), the Fisher information matrix is singular iff f_1 is the pdf of the standard (multi)normal distribution.

Finally, we indicate that these results can partly be extended to setups where the skewing function does not simply involve a linear function of δ (such as in $(x, \delta) \mapsto \Pi(\delta'(x-\mu))$), but rather a more general (e.g., higher-order polynomial) function of δ; see for instance [13,14]. The exact structure of the corresponding "iff" results, however, does very much depend on the type of skewing functions used, and deriving results for specific classes of such skewing functions is beyond the scope of the present paper. On the contrary, as we will see in the next section, our treatment of the stationary point at $\delta = 0$ of the profile log-likelihood function will readily apply in such extremely general setups.

We stress, however, that all the asymmetric distributions considered above are obtained by transforming symmetric distributions by means of skewing mechanisms inherited from Azzalini [1], and that our results only apply for such distributions. Other classes of skewed distributions, like, e.g., the epsilon-skew-t ones, do not even suffer, in the vicinity of symmetry, from a singular Fisher information matrix at the normal density; see [30].

5. Stationary point of the profile log-likelihood function for skewness

In this section, we tackle the problem of a stationary point, in the vicinity of symmetry, of the profile log-likelihood function for skewness, and show that Conjecture 1.2 actually holds (in any dimension) for the broad class of skew-symmetric distributions considered in Section 2, as well as for its extensions from Section 4.

If we have a random sample $X^{(n)} := (X_1, \ldots, X_n)$ from (2.2), we define the profile log-likelihood function for skewness as

$$\tilde{L}^{\Pi}_{f;\delta}(X^{(n)}) := \sup_{\mu\in\mathbb{R}^k} L^{\Pi}_{f;\mu,\delta}(X^{(n)}), \quad \delta \in \mathbb{R}^k, \tag{5.10}$$

where $L^{\Pi}_{f;\mu,\delta}(X^{(n)}) := \sum_{i=1}^n \log f^{\Pi}_{\mu,\delta}(X_i)$ is the standard log-likelihood function associated with $X^{(n)}$. This expression can be rewritten under the more tractable form

$$\tilde{L}^{\Pi}_{f;\delta}(X^{(n)}) = L^{\Pi}_{f;\hat{\mu}_f(\delta),\delta}(X^{(n)}), \tag{5.11}$$

where $\hat{\mu}_f(\delta)$ stands for the MLE of μ at fixed δ.

Now, using (5.11) and denoting by $D_\delta\hat{\mu}_f(\delta) = (\partial_{\delta_j}(\hat{\mu}_f(\delta))_i)$ the Jacobian matrix of the mapping $\delta \mapsto \hat{\mu}_f(\delta)$, the chain rule leads to

$$\begin{aligned}\nabla_\delta \tilde{L}^{\Pi}_{f;\delta}(X^{(n)}) &= (D_\delta\hat{\mu}_f(\delta))'\,(\nabla_\mu L^{\Pi}_{f;\mu,\delta}(X^{(n)}))|_{(\mu,\delta)=(\hat{\mu}_f(\delta),\delta)} + (\nabla_\delta L^{\Pi}_{f;\mu,\delta}(X^{(n)}))|_{(\mu,\delta)=(\hat{\mu}_f(\delta),\delta)} \\ &= (\nabla_\delta L^{\Pi}_{f;\mu,\delta}(X^{(n)}))|_{(\mu,\delta)=(\hat{\mu}_f(\delta),\delta)},\end{aligned}$$

where the first term vanishes since $(\nabla_\mu L^{\Pi}_{f;\mu,\delta}(X^{(n)}))|_{(\mu,\delta)=(\hat{\mu}_f(\delta),\delta)} = 0$ for any $\delta \in \mathbb{R}^k$ (by definition of the MLE $\hat{\mu}_f(\delta)$). Therefore, a necessary condition for the profile log-likelihood function to always admit a stationary point at $\delta = 0$ is that

$$(\nabla_\delta L^{\Pi}_{f;\mu,\delta}(X^{(n)}))|_{(\mu,\delta)=(\hat{\mu}_f(0),0)} = 2\Pi'(0)\sum_{i=1}^{n}(X_i - \hat{\mu}_f(0)) = 0 \tag{5.12}$$

for any $X^{(n)}$. In other words, the maximum likelihood estimator for the location parameter μ, at $\delta = 0$, must coincide, for any $X^{(n)}$, with the sample average $\bar{X}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} X_i$. Remembering that $\hat{\mu}_f(0)$ is nothing but the MLE of μ at $\delta = 0$ (that is, in the location family of distributions with pdf $x \mapsto f(x-\mu)$), the following result directly follows from a well-known characterization property which can be traced back to Gauss (more precisely, from its version in [31], which is valid for any fixed sample size $n \geq 3$).

Theorem 5.1. *Let Assumption A hold. Then, the skewness profile log-likelihood function $\delta \mapsto \tilde{L}^{\Pi}_{f;\delta}(X^{(n)})$ admits, for any sample $X^{(n)}$ of a fixed sample size $n \geq 3$, a stationary point at $\delta = 0$ iff f is the pdf of a (multi)normal distribution.*

This theorem shows that, unlike for Conjecture 1.1, the result in Conjecture 1.2 holds in any dimension k. This clearly underlines that, for dimensions $k > 1$, no equivalence exists between the two problems considered in this paper: indeed, in the vicinity of symmetry, only multinormal kernels lead to stationary points of the profile log-likelihood function, whereas a much larger class of distributions causes Fisher information matrices to be singular. Further note that, of course, Theorem 5.1 can be regarded as a further characterization of the (multi)normal distribution.

Similarly to what has been done in Section 4 for the Fisher singularity problem, it is natural to investigate how far Theorem 5.1 extends to more general models of skew-symmetric distributions. Since the skewing function $\Pi(\delta'(. - \mu))$ is the same in (4.7) as in (2.2), the result trivially holds for the corresponding densities (note that, of course, the profile log-likelihood is then obtained by taking, for fixed δ, the supremum with respect to μ and Σ in (5.10), and that $\nabla_\delta \tilde{L}^{\Pi}_{f;\delta}(X^{(n)})$ remains unchanged despite extending the chain rule for differentiation to the parameter Σ following the formula in [32]). As for the ones in (4.9), the same argument as above shows that a necessary condition for the corresponding profile log-likelihood function to always admit a stationary point at $\delta = 0$ is that

$$2\Pi'(0)(\hat{\Sigma}_f(0))^{-1/2}\sum_{i=1}^{n}(X_i - \hat{\mu}_f(0)) = 0 \tag{5.13}$$

for any $X^{(n)}$, where $\hat{\Sigma}_f(0)$ stands for the MLE of Σ at $\delta = 0$; since $\hat{\Sigma}_f(0)$ is always positive definite by definition, the necessary condition in (5.13) is strictly equivalent to the one in (5.12), and we may conclude as above. Finally, one may also consider (multivariate) flexible skew-symmetric distributions, that is, skew-symmetric distributions based on skewing functions of the form $x \mapsto H(\sum_{j=1}^{D}\delta_j' P_{2j-1}(x))$, where H is an arbitrary cdf, $P_d(x)$ is a vector stacking all quantities $\Pi_{i=1}^{k} x_i^{r_i}$, with $r_i \in \mathbb{N}$ and $\sum_{i=1}^{k} r_i = d$, and where δ_j is a parameter with the same dimension as $P_{2j-1}(x)$; see [14]. Since, with obvious notation, $(\nabla_{\delta_1} L^{\Pi}_{f;\mu,\delta_1,\ldots,\delta_D}(X^{(n)}))|_{(\mu,\delta_1,\ldots,\delta_D)=(\hat{\mu}_f(0),0,\ldots,0)}$ is still of the same form as (5.12), Conjecture 1.2 trivially extends to this setup as well. Since the class of flexible skew-symmetric distributions is dense in the class of the skew-symmetric ones (see [14] for a precise statement), Conjecture 1.2 virtually applies for the latter class.

Acknowledgments

Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté française de Belgique, for support via a Mandat d'Aspirant FNRS. Davy Paindaveine, who is also a member of ECORE, the association between CORE and ECARES, is grateful to the Fonds National de la Recherche Scientifique, Communauté française de Belgique, for a Mandat d'Impulsion Scientifique. The authors would like to thank Nicolas Richard for his help in the proof of Lemma 2.3, Anna Clara Monti for the insightful comments on profile log-likelihood functions and for providing them a copy of the manuscript [24], Adelchi Azzalini and Antonella Capitanio for interesting suggestions, as well as the associate editor and two anonymous referees for their valuable comments that helped in improving the presentation and the quality of the paper.

Appendix

Proof of Lemma 2.1. (i) For any $x = (x_\mu', x_\delta')'$, with $x_\mu, x_\delta \in \mathbb{R}^k$, we clearly have that, under Assumption A, $\Gamma_f x = 0$ iff

$$\begin{cases} x_\mu = -\Gamma_{f;\mu\mu}^{-1}\Gamma_{f;\mu\delta}x_\delta \\ \Gamma_{f;\delta\delta.\mu}x_\delta = 0. \end{cases}$$

The latter system clearly implies that $\ker(\Gamma_f)$ and $\ker(\Gamma_{f;\delta\delta.\mu})$ have the same dimension. Part (i) of the result readily follows.

(ii) As a covariance (hence, symmetric and positive semidefinite) matrix, $\Gamma_{f;\delta\delta.\mu}$ can be diagonalized into $O\Lambda O'$, where O is a $k \times k$ orthogonal matrix and Λ is a diagonal matrix with nonnegative entries. Consider the symmetric square-root $(\Gamma_{f;\delta\delta.\mu})^{1/2} := O\Lambda^{1/2}O'$ of $\Gamma_{f;\delta\delta.\mu}$. Of course, $\mathrm{rank}((\Gamma_{f;\delta\delta.\mu})^{1/2}) = \mathrm{rank}(\Gamma_{f;\delta\delta.\mu})$, and the common value is the number of positive diagonal entries in Λ. This implies that $\mathrm{rank}(\Gamma_{f;\delta\delta.\mu}) = k - m$ iff $\mathrm{rank}((\Gamma_{f;\delta\delta.\mu})^{1/2}) = k - m$, which in turn holds iff $\ker((\Gamma_{f;\delta\delta.\mu})^{1/2})$ has dimension m. Equivalently, m is the largest integer $\ell \in \{1, 2, \ldots, k\}$ such that there exists a $k \times \ell$ matrix $V = (v_1, \ldots, v_\ell)$ with orthonormal columns satisfying $(\Gamma_{f;\delta\delta.\mu})^{1/2} V = 0$. This yields the result since we have $(\Gamma_{f;\delta\delta.\mu})^{1/2} V = 0$ iff $V'\Gamma_{f;\delta\delta.\mu} V = 0$. □

Proof of Lemma 2.3. If g is a solution of (2.4), then it must satisfy

$$z'V'W\nabla g(x) = -z'V'x, \quad \forall (x, z) \in \mathbb{R}^k \times \mathbb{R}^m. \tag{A.1}$$

Vice versa, any solution of (A.1) is also a solution of (2.4), so that (2.4) and (A.1) can be considered equivalent.

Let us first show that any solution of (2.4) (hence of (A.1)) is of the form given in (2.5). To this end, write

$$\begin{aligned} g(x) = g(P_1x + P_2x) &= g(P_2x) + \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} g(P_2x + tP_1x)\mathrm{d}t \\ &= g(P_2x) + \int_0^1 (P_1x)'\nabla g(P_2x + tP_1x)\mathrm{d}t. \end{aligned} \tag{A.2}$$

Since P_1 is the matrix of the orthogonal projection from $\mathbb{R}^k$ onto $\mathrm{Im}(W'V)$, there is, for any $x \in \mathbb{R}^k$, a unique $x_1 \in \mathbb{R}^m$ such that $P_1x = W'Vx_1$. Using this fact and (A.1) in (A.2) yields

$$\begin{aligned} g(x) &= g(P_2x) + \int_0^1 x_1'V'W\nabla g(P_2x + tW'Vx_1)\mathrm{d}t \\ &= g(P_2x) - \int_0^1 x_1'V'(P_2x + tW'Vx_1)\mathrm{d}t \\ &= g(P_2x) - x_1'V'P_2x - \frac{1}{2}x_1'V'W'Vx_1 \\ &= g(P_2x) - x'P_1W^{-1}P_2x - \frac{1}{2}x'P_1W^{-1}P_1x, \end{aligned}$$

which indeed confirms that any solution of (2.4) is as in (2.5); here, h is the restriction of g to $\ker(V'W)$.

Now, let us investigate under which conditions a function g as in (2.5) is a solution of (A.1) (hence of (2.4)). Using the facts that $P_1W'V = W'V$ and that $P_2W'V = 0$ yields

$$\begin{aligned} z'V'W\nabla g(x) &= \frac{\mathrm{d}}{\mathrm{d}t} g(x + tW'Vz)|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left[-\frac{1}{2}(P_1x + tW'Vz)'W^{-1}(P_1x + tW'Vz) - (P_1x + tW'Vz)'W^{-1}P_2x + h(P_2x)\right]\Bigg|_{t=0} \\ &= -\frac{1}{2}x'P_1W^{-1}W'Vz - \frac{1}{2}z'V'P_1x - z'V'P_2x. \end{aligned}$$

Hence, decomposing x into $P_1x + P_2x$ and using $P_1x = W'Vx_1$, we obtain

$$\begin{aligned} z'V'W\nabla g(x) &= -z'V'x - \frac{1}{2}x'P_1W^{-1}W'Vz + \frac{1}{2}z'V'P_1x \\ &= -z'V'x - \frac{1}{2}(x_1'V'W'Vz - x_1'V'WVz), \end{aligned}$$

which shows that, as announced, g in (2.5) is a solution to (A.1) iff $V'(W - W')V = 0$. This establishes the result. □

Proof of Theorem 2.1. Assume that $\mathrm{rank}(\Gamma_f) = 2k - m$. By Lemma 2.2, there exists a $k \times m$ matrix $V = (v_1, \ldots, v_m)$ with orthonormal columns such that $V'd^*_{f;\mu}(X) = 0$ a.s. when X has pdf $f(. - \mu)$. Hence, as in Section 2, we have that $V'(x + \mathcal{I}_f^{-1}\nabla \log f(x)) = 0$ a.e. in $\mathbb{R}^k$. Since $\mathcal{I}_f$ is symmetric, this system of PDEs admits at least a solution (Lemma 2.3(i)), and the general solution is (Lemma 2.3(ii))

$$\log f(x) = -\frac{1}{2}x'P_1\mathcal{I}_fP_1x - x'P_1\mathcal{I}_fP_2x + h(P_2x),$$

where $P_1 := \mathcal{I}_f^{-1}V(V'\mathcal{I}_f^{-2}V)^{-1}V'\mathcal{I}_f^{-1}$, $P_2 := I_k - P_1$, and where h is an arbitrary function defined on $P_2\mathbb{R}^k$. Equivalently,

$$\begin{aligned} \log f(x) &= -\frac{1}{2}\Big[x'P_1\mathcal{I}_fP_1x + 2x'P_1\mathcal{I}_fP_2x + x'P_2\mathcal{I}_fP_2x\Big] + \tilde{h}(P_2x) \\ &= -\frac{1}{2}x'\mathcal{I}_fx + \tilde{h}(P_2x), \end{aligned}$$

for an arbitrary function $\tilde{h}$ defined on $P_2\mathbb{R}^k$. Now, since the $k \times m$ matrix V satisfies $V'\Gamma_{f;\delta\delta.\mu}V = 0$, with $\Gamma_{f;\delta\delta.\mu} = 4(\Pi'(0))^2(\Sigma_f - \mathcal{I}_f^{-1})$, and since there is no $\ell > m$ for which this would hold for a $k \times \ell$ matrix V (Lemma 2.1), we obtain that $\ker(\Sigma_f \mathcal{I}_f - I_k) = \mathrm{Im}(\mathcal{I}_f^{-1}V)$. Hence, P_1 is the matrix of the orthogonal projection from $\mathbb{R}^k$ onto $\ker(\Sigma_f \mathcal{I}_f - I_k)$, and therefore P_2 is the matrix of the orthogonal projection from $\mathbb{R}^k$ onto $(\ker(\Sigma_f \mathcal{I}_f - I_k))^{\perp} = \mathrm{Im}(\mathcal{I}_f \Sigma_f - I_k)$, which establishes the necessity part.

For the sufficiency part, assume that f is given by (2.6). Then a direct computation yields that $\varphi_f(x) = P\varphi_h(Px) + \mathcal{I}_f x$, with $\varphi_h(x) := -\nabla h(x)/h(x)$. If C is a full-rank $k \times m$ matrix with $\mathrm{Im}(C) = \ker(P)$, we have $C'(\varphi_f(x) - \mathcal{I}_f x) = 0$, or equivalently, $(\mathcal{I}_f C)'(x - \mathcal{I}_f^{-1}\varphi_f(x)) = 0$. Since there is no full-rank $k \times \ell$ matrix V ($\ell > m$) for which we would have $V'(x - \mathcal{I}_f^{-1}\varphi_f(x)) = 0$ (if there was one, then P would not have rank $k - m$), Lemma 2.2 allows us to conclude that Γ_f is singular with $\mathrm{rank}(\Gamma_f) = 2k - m$. □

Proof of Theorem 2.2. For notational simplicity, we let in this proof $\Sigma = (\mathcal{I}_f^Y)^{-1} := (B'\mathcal{I}_f B)^{-1}$. Also, for any $k \times k$ matrix A, we will partition A into submatrices A_{ij}, $i, j = 1, 2$, in the same fashion as for $\mathcal{I}_f^Y$ in the statement of the theorem. The k-vector Y then has pdf

$$f^Y(y) = |B| h(PBy) \exp\Big[-\frac{1}{2}y'B'\mathcal{I}_f By\Big] = |B|\, h(B_2y_2) \exp\Big[-\frac{1}{2}y'\Sigma^{-1}y\Big],$$

so that, using the fact that $|\Sigma| = |\Sigma_{11.2}||\Sigma_{22}|$, we obtain that the pdf of Y_2 is given by

$$\begin{aligned} f^{Y_2}(y_2) &= \int_{\mathbb{R}^m} f^Y(y)\mathrm{d}y_1 \\ &= (2\pi)^{k/2}|\Sigma|^{1/2}|B|\, h(B_2y_2) \int_{\mathbb{R}^m} \frac{(2\pi)^{-k/2}}{|\Sigma|^{1/2}} \exp\Big[-\frac{1}{2}y'\Sigma^{-1}y\Big]\mathrm{d}y_1 \\ &= (2\pi)^{m/2}|\Sigma_{11.2}|^{1/2}|B|\, h(B_2y_2) \exp\Big[-\frac{1}{2}y_2'\Sigma_{22}^{-1}y_2\Big]. \end{aligned} \tag{A.3}$$

Now, the formula for inverses of partitioned matrices yields $\Sigma_{22} = ((\mathcal{I}_f^Y)^{-1})_{22} = (\mathcal{I}_{f;22.1}^Y)^{-1}$, so that $\Sigma_{22}^{-1} = (B'\mathcal{I}_f B)_{22.1} = B_2'MB_2$ for some $k \times k$ matrix M. Therefore $f^{Y_2}(y_2)$ is a symmetric density involving y_2 through B_2y_2 only. Since $h(.)$ is essentially arbitrary, so is $f^{Y_2}(.)$ (meaning that this density should just fulfill the conditions in the statement of Theorem 2.2).

Now, by using (A.3), we obtain that the conditional distribution of Y_1 given that $Y_2 = y_2$ is given by

$$f^{Y_1|Y_2=y_2}(y_1) = (2\pi)^{-m/2}|\Sigma_{11.2}|^{-1/2} \exp\Big[-\frac{1}{2}\big\{y'\Sigma^{-1}y - y_2'\Sigma_{22}^{-1}y_2\big\}\Big].$$

By using again partitioned inverses, one easily obtains that $y'\Sigma^{-1}y - y_2'\Sigma_{22}^{-1}y_2 = (y_1 - \Sigma_{11.2}(\Sigma^{-1})_{12}y_2)'(\Sigma_{11.2})^{-1}(y_1 - \Sigma_{11.2}(\Sigma^{-1})_{12}y_2)$, hence that $Y_1|Y_2 = y_2$ has an m-variate normal distribution with mean $\Sigma_{11.2}(\Sigma^{-1})_{12}y_2$ and covariance matrix $\Sigma_{11.2}$. The result then follows by noting that $\Sigma_{11.2} = ((\Sigma^{-1})_{11})^{-1} = \mathcal{I}_{f;11}^{-1}$ and $(\Sigma^{-1})_{12} = \mathcal{I}_{f;12}$. □

Proof of Theorem 2.3. Assume that $\mathrm{rank}(\Gamma_f) = 2k - m$, for some fixed $m \in \{1, 2, \ldots, k\}$. If f is elliptical with scatter matrix Σ, then $\mathcal{I}_f = i_{f_1}\Sigma^{-1}$ and $\Sigma_f = \sigma_{f_1}\Sigma$. Lemma 2.1(i) then yields that $k > k - m = \mathrm{rank}(\Gamma_{f;\delta\delta.\mu}) = \mathrm{rank}(\Sigma_f - \mathcal{I}_f^{-1}) = \mathrm{rank}((\sigma_{f_1} - i_{f_1}^{-1})\Sigma)$. Since Assumption A imposes that $\mathcal{I}_f$ is positive definite, Σ has full rank, so that the only way $\mathrm{rank}((\sigma_{f_1} - i_{f_1}^{-1})\Sigma) < k$ is to have $\sigma_{f_1} = i_{f_1}^{-1}$, which implies that $\Sigma_f = \mathcal{I}_f^{-1}$ (and that $m = k$). Theorem 2.1 then states that

$$f(x) = h(Px) \exp\Big[-\frac{1}{2}x'\mathcal{I}_f x\Big] = h(Px) \exp\Big[-\frac{1}{2}x'\Sigma_f^{-1}x\Big],$$

where P is the matrix of the orthogonal projection from $\mathbb{R}^k$ onto $\mathrm{Im}(\mathcal{I}_f\Sigma_f - I_k) = \{0\}$, that is, where P is the $k \times k$ zero matrix. We conclude that $f(x) = h(0) \exp[-\frac{1}{2}x'\Sigma_f^{-1}x] = (2\pi)^{-k/2}|\Sigma_f|^{-1/2} \exp[-\frac{1}{2}x'\Sigma_f^{-1}x]$, as was to be shown. □

References

[1] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.
[2] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 46 (1986) 199–208.
[3] A. Rotnitzky, D.R. Cox, M. Bottai, J. Roberts, Likelihood-based inference with singular information matrix, Bernoulli 6 (2000) 243–284.
[4] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[5] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew-normal distributions, J. R. Stat. Soc. B 61 (1999) 579–602.
[6] B.C. Arnold, R.J. Beaver, R.A. Groeneveld, W.Q. Meeker, The nontruncated marginal of a truncated bivariate normal distribution, Psychometrika 58 (1993) 471–478.
[7] J.B. Copas, H.G. Li, Inference for non-random samples (with discussion), J. R. Statist. Soc. B 59 (1997) 55–95.
[8] B.C. Arnold, R.J. Beaver, The skew-Cauchy distribution, Statist. Probab. Lett. 49 (2000) 285–290.
[9] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution, J. R. Stat. Soc. B 65 (2003) 367–389.

[10] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivariate Anal. 79 (2001) 99–113.
[11] A.K. Gupta, Multivariate skew-*t* distribution, Statistics 37 (2003) 359–363.
[12] M.G. Genton, N. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Ann. Inst. Statist. Math. 57 (2005) 389–401.
[13] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distribution, Statist. Sinica 14 (2004) 1259–1270.
[14] Y. Ma, M.G. Genton, Flexible class of skew-symmetric distributions, Scand. J. Statist. 31 (2004) 459–468.
[15] M.G. Genton, Skew-elliptical Distributions and their Applications: A Journey Beyond Normality, edited volume, Chapman and Hall/CRC, Boca Raton, FL, 2004.
[16] B.C. Arnold, R.J. Beaver, Skewed multivariate models related to hidden truncation and/or selective reporting, Test 11 (2002) 7–54.
[17] A. Azzalini, The skew-normal distribution and related multivariate families (with discussion), Scand. J. Statist. 32 (2005) 159–188.
[18] R.B. Arellano-Valle, A. Azzalini, The centred parametrization for the multivariate skew-normal distribution, J. Multivariate Anal. 99 (2008) 1362–1382.
[19] M. Chiogna, A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution, Stat. Methods Appl. 14 (2005) 331–341.
[20] A. Pewsey, Problems of inference for Azzalini's skew-normal distribution, J. Appl. Statist. 27 (2000) 859–870.
[21] H.S. Salinas, R.B. Arellano-Valle, H.W. Gómez, The extended skew-exponential power distribution and its derivation, Comm. Statist. Theory Methods 36 (2007) 1673–1689.
[22] A. Pewsey, Some observations on a simple means of generating skew distributions, in: N. Balakrishnan, E. Castillo, J.M. Sarabia (Eds.), Advances in Distribution Theory, Order Statistics, and Inference, Statistics for Industry and Technology, Birkhäuser, Boston, 2006, pp. 75–84.
[23] T.J. DiCiccio, A.C. Monti, Inferential aspects of the skew exponential power distribution, J. Amer. Statist. Assoc. 99 (2004) 439–450.
[24] T.J. DiCiccio, A.C. Monti, Inferential aspects of the skew *t*-distribution, 2009 (Manuscript in preparation).
[25] H.W. Gómez, O. Venegas, H. Bolfarine, Skew-symmetric distributions generated by the distribution function of the normal distribution, Environmetrics 18 (2007) 395–407.
[26] A. Azzalini, M.G. Genton, Robust likelihood methods based on the skew-*t* and related distributions, Internat. Statist. Rev. 76 (2008) 106–129.
[27] M. Bottai, Confidence regions when the Fisher information is zero, Biometrika 90 (2003) 73–84.
[28] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York, 1986.
[29] A.R. Swensen, The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend, J. Multivariate Anal. 16 (1985) 54–70.
[30] H.W. Gómez, F.J. Torres, H. Bolfarine, Large-sample inference for the Epsilon-Skew-*t* distribution, Comm. Statist. Theory Methods 36 (2007) 73–81.
[31] A. Azzalini, M.G. Genton, On Gauss's characterization of the normal distribution, Bernoulli 13 (2007) 169–174.
[32] W.M. Patefield, On the maximized likelihood function, Sankhyā Ser. B 39 (1977) 92–96.