Parameter Estimation For Multivariate Generalized Gaussian Distributions

Frédéric Pascal, Lionel Bombrun, *Member, IEEE*, Jean-Yves Tourneret, *Senior Member, IEEE*, and Yannick Berthoumieu

Abstract—Due to its heavy-tailed and fully parametric form, the multivariate generalized Gaussian distribution (MGGD) has been receiving much attention in signal and image processing applications. Considering the estimation issue of the MGGD parameters, the main contribution of this paper is to prove that the maximum likelihood estimator (MLE) of the scatter matrix exists and is unique up to a scalar factor, for a given shape parameter $\beta \in (0, 1)$. Moreover, an estimation algorithm based on a Newton-Raphson recursion is proposed for computing the MLE of MGGD parameters. Various experiments conducted on synthetic and real data are presented to illustrate the theoretical derivations in terms of number of iterations and number of samples for different values of the shape parameter. The main conclusion of this work is that the parameters of MGGDs can be estimated using the maximum likelihood principle with good performance.

Index Terms—Covariance matrix estimation, fixed point algorithm, multivariate generalized Gaussian distribution.

I. Introduction

UNIVARIATE and multivariate generalized Gaussian distributions (GGDs) have received much attention in the literature. Historically, this family of distributions has been introduced in [1]. Some properties of these distributions have been reported in several papers such as [2]–[4]. These properties include various stochastic representations, simulation methods and probabilistic characteristics. GGDs belong to the family of elliptical distributions (EDs) [5], [6], originally introduced by Kelker in [7] and studied in [8], [9]. Depending on the value of the shape parameter β, multivariate GGDs (MGGDs) and the distributions of spherically invariant random vectors share common properties (see [10], [11] for more details).

MGGDs have been used intensively in the image processing community. Indeed, including Gaussian and Laplacian distributions as special cases, MGGDs are potentially interesting for modeling the statistical properties of various images or features extracted from these images. In particular, the distribution of wavelet or curvelet coefficients has been shown to be modeled accurately by GGDs [12]–[15]. This property has been exploited for many signal and image processing applications including image denoising [16]–[19], content-based image retrieval [20], [21], image thresholding [22] or texture classification in industrial problems [23]. Other applications involving GGDs include radar [24], video coding and denoising [25]–[27] or biomedical signal processing [26], [28], [29]. Finally, it is interesting to note that complex GGDs have been recently studied in [30], [31] and that multivariate regression models with generalized Gaussian errors have been considered in [32].

Considering the important attention devoted to GGDs, estimating the parameters of these distributions is clearly an interesting issue. Classical estimation methods that have been investigated for univariate GGDs include the maximum likelihood (ML) method [33] and the method of moments [34]. In the multivariate context, MGGD parameters can be estimated by a least-squares method as in [18] or by minimizing a χ^2 distance between the histogram of the observed data and the theoretical probabilities associated with the MGGD [35]. Estimators based on the method of moments and on the ML method have also been proposed in [36]–[38].

Several works have analyzed covariance matrix estimators defined under different modeling assumptions. On the one hand, fixed point (FP) algorithms have been derived and analyzed in [39], [40] for SIRVs. On the other hand, in the context of robust estimation, the properties of M-estimators have been studied by Maronna in [41]. Unfortunately, Maronna's conditions are not fully satisfied for MGGDs (see remark II.3). This paper shows that despite the non-applicability of Maronna's results, the MLE of MGGD parameters exists, is unique and can be computed by an FP algorithm. Although the methodology adopted in this paper has some similarities with the one proposed in [39], [40], there are also important differences which require a specific analysis (see for instance remark III.1). More precisely, the FP equation of [39] corresponds to an approximate MLE for SIRVs while in [40] the FP equation results from a different problem (see Eq. (14) in [40] compared to (15) of this paper).

The contributions of this paper are to establish some properties related to the FP equation of the MLE for MGGDs. More precisely, we show that for a given shape parameter β belonging to (0, 1), the MLE of the scatter matrix $\mathbf{M}$ exists and is unique up to a scalar factor.[1] An iterative algorithm based on a Newton-Raphson procedure is then proposed to compute the MLE of $\mathbf{M}$.

The paper is organized as follows. Section II defines the MGGDs considered in this study and derives the MLEs of their parameters. Section III presents the main theoretical results

---

[1]From the submission of this paper, another approach based on geodesic convexity was proposed in [38].

of this paper while a proof outline is given in Section IV. For presentation clarity, full demonstrations are provided in the Appendices. Section V is devoted to simulation results conducted on synthetic and real data. The convergence speed of the proposed estimation algorithm as well as the bias and consistency of the scatter matrix MLE are first investigated using synthetic data. Experimentations performed on real images extracted from the VisTex database are then presented. Conclusions and future works are finally reported in Section VI.

II. PROBLEM FORMULATION

A. Definitions

The probability density function of an MGGD in $\mathbb{R}^p$ is defined by [42]

$$p(\mathbf{x}|\mathbf{M}, m, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} h_{m,\beta}(\mathbf{x}^T \mathbf{M}^{-1} \mathbf{x}) \qquad (1)$$

for any $\mathbf{x} \in \mathbb{R}^p$, where $\mathbf{M}$ is a $p \times p$ symmetric real scatter matrix, $\mathbf{x}^T$ is the transpose of the vector $\mathbf{x}$, and $h_{m,\beta}(\cdot)$ is a so-called density generator defined by

$$h_{m,\beta}(y) = \frac{\beta \Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{p}{2\beta}\right) 2^{\frac{p}{2\beta}}} \frac{1}{m^{\frac{p}{2}}} \exp\left(-\frac{y^\beta}{2m^\beta}\right) I_{\mathbb{R}+}(y) \quad (2)$$

where $I_{\mathbb{R}+}(y)$ is the indicator function on $\mathbb{R}^+$ and m and β are the scale and shape parameters of the MGGD. The matrix $\mathbf{M}$ will be normalized in this paper according to $\mathrm{Tr}(\mathbf{M}) = p$, where $\mathrm{Tr}(\mathbf{M})$ is the trace of the matrix $\mathbf{M}$. It is interesting to note that letting $\beta = 1$ corresponds to the multivariate Gaussian distribution. Moreover, when β tends toward infinity, the MGGD is known to converge in distribution to a multivariate uniform distribution [42].

B. Stochastic Representation

Let $\mathbf{x}$ be a random vector of $\mathbb{R}^p$ distributed according to an MGGD with scatter matrix $\Sigma = m\mathbf{M}$ and shape parameter β. Gómez et al. have shown that $\mathbf{x}$ admits the following stochastic representation [2]

$$\mathbf{x} \stackrel{d}{=} \tau \Sigma^{\frac{1}{2}} \mathbf{u} \qquad (3)$$

where $\stackrel{d}{=}$ means equality in distribution, $\mathbf{u}$ is a random vector uniformly distributed on the unit sphere of $\mathbb{R}^p$, and τ is a scalar positive random variable such that

$$\tau^{2\beta} \sim \Gamma\left(\frac{p}{2\beta}, 2\right) \qquad (4)$$

where $\Gamma(a, b)$ is the univariate gamma distribution with parameters a and b (see [43] for definition).

C. MGGD Parameter Estimation for Known β

Let $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be N independent and identically distributed (i.i.d.) random vectors distributed according to an MGGD with parameters $\mathbf{M}$, m and β. This section studies estimators of $\mathbf{M}$ and m based on $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ for a known value

of $\beta \in (0,1)$.[2] The MGGD is a particular case of elliptical distribution that has received much attention in the literature. Following the results of [45] for real elliptical distributions, by differentiating the log-likelihood of vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ with respect to $\mathbf{M}$, the MLE of the matrix $\mathbf{M}$ satisfies the following FP equation

$$\mathbf{M} = \frac{2}{N} \sum_{i=1}^{N} \frac{-g_{m,\beta}\left(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i\right)}{h_{m,\beta}\left(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i\right)} \mathbf{x}_i \mathbf{x}_i^T \qquad (5)$$

where $g_{m,\beta}(y) = \partial h_{m,\beta}(y)/\partial y$. In the particular case of an MGGD with known parameters m and β, straightforward computations lead to

$$\mathbf{M} = \frac{\beta}{Nm^\beta} \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\left(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i\right)^{1-\beta}}. \qquad (6)$$

When the parameter m is unknown, the MLEs of $\mathbf{M}$ and m are obtained by differentiating the log-likelihood of $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ with respect to $\mathbf{M}$ and m yielding

$$\mathbf{M} = \frac{\beta}{Nm^\beta} \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\left(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i\right)^{1-\beta}}, \qquad (7)$$

$$m = \left[\frac{\beta}{pN} \sum_{i=1}^{N} \left(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i\right)^\beta\right]^{\frac{1}{\beta}}. \qquad (8)$$

After replacing m in (6) by its expression (8), the following result can be obtained

$$\mathbf{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{Np}{y_i + y_i^{1-\beta} \sum_{j \neq i} y_j^\beta} \mathbf{x}_i \mathbf{x}_i^T. \qquad (9)$$

As mentioned before and confirmed by (9), $\mathbf{M}$ can be estimated independently from the scale parameter m. Moreover, the following remarks can be made about (9).

Remark II.1: When $\beta = 1$, (9) is close to the sample covariance matrix (SCM) estimator (the only difference between the SCM estimator and (9) is due to the estimation of the scale parameter that equals 1 for the multivariate Gaussian distribution). For $\beta = 0$, (9) reduces to the FP covariance matrix estimator studied in [45]–[47].

Remark II.2: Equation (9) remains unchanged if $\mathbf{M}$ is replaced by $\alpha\mathbf{M}$ where α is any non-zero real factor. Thus, the solutions of (9) (when there exist) can be determined up to a scale factor α. The normalization $\mathrm{Tr}(\mathbf{M}) = p$ will be adopted in this paper and will be justified in the simulation section.

Remark II.3: Let us consider the function f_i defined by

$$f_i(y) = \frac{Np}{y + c_i y^{1-\beta}} I_{\mathbb{R}+}(y) \qquad (10)$$

where c_i is a positive constant independent of y (the index i is used here to stress the fact that $c_i = \sum_{j \neq i} y_j^\beta$ changes with i but does not depend on y_i). Equation (9) can be rewritten as

$$\mathbf{M} = \frac{1}{N} \sum_{i=1}^{N} f_i(y_i) \mathbf{x}_i \mathbf{x}_i^T. \quad (11)$$

Roughly speaking,[3] f_i satisfies Maronna's conditions (recapped below, see [41, p. 53] for more details) for any $\beta \in (0, 1)$ except the continuity at $y = 0$.

Maronna's conditions for a function $f : \mathbb{R} \to \mathbb{R}$
  (i) f is non-negative, non increasing, and continuous on $[0, \infty)$.
  (ii) Let $\psi(s) = s f(s)$ and $K = \sup_{s \geq 0} \psi(s)$. The function ψ is non decreasing and strictly increasing in the interval defined by $\psi < K$ with $p < K < \infty$.
  (iii) Let $P_N(\cdot)$ denotes the empirical distribution of $\mathbf{x}_1, \ldots, \mathbf{x}_N$. There exists $a > 0$ such that for all hyperplanes W with $\dim(W) \leq p - 1$

$$P_N(W) \leq 1 - \frac{p}{K} - a. \quad (12)$$

Because of non continuity of f_i around 0, the properties of M-estimators derived by Maronna cannot be applied directly to the estimators of the MGGD parameters. The objective of Section III is to derive similar properties for the estimator of $\mathbf{M}$ defined by the FP (9).

D. MGGD Parameter Estimation for Unknown β

When the shape parameter β of the MGGD is unknown, the MLE of $\mathbf{M}$, m and β is obtained by differentiating the log-likelihood of $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ with respect to $\mathbf{M}$ and m and β, i.e., by combining (7) and (8) with the following relation

$$\alpha(\beta) = \frac{pN}{2 \sum\limits_{i=1}^{N} y_i^\beta} \sum_{i=1}^{N} \left[y_i^\beta \ln(y_i) \right] - \frac{pN}{2\beta} \left[\Psi\left(\frac{p}{2\beta}\right) + \ln 2 \right]$$

$$-N - \frac{pN}{2\beta} \ln\left(\frac{\beta}{pN} \sum_{i=1}^{N} y_i^\beta\right) = 0 \quad (13)$$

where $\Psi(\cdot)$ is the digamma function. Equations (9) and (13) show that $\mathbf{M}$ and β can be estimated independently from the scale parameter m.

III. STATEMENTS OF THE MAIN RESULTS

As the estimation scenario presented in the previous section has some similarities with the FP estimator studied in [39], similar results about the estimator existence, uniqueness (up to a scale factor) and FP algorithm convergence are expected to be true. This section summarizes the properties of the FP estimator defined by (9) for a known value of $\beta \in (0, 1)$ (all proofs are provided in the Appendices to simplify the reading). The case of an unknown value of β will be discussed in the simulation section.

[3]Actually, Maronna's function depends only on the ith sample and not on all the samples as it is the case here!

A. Notations

For any positive integer n, $[\![1, n]\!]$ denotes the set of integers $\{1, \ldots, n\}$. For any vector $\mathbf{x} \in \mathbb{R}^p$, $\|\mathbf{x}\|$ denotes the Euclidean norm of $\mathbf{x}$ such as $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$, where $\mathbf{x}^T$ is the transpose of $\mathbf{x}$. Throughout the paper, we will use several basic results about square matrices, especially regarding the diagonalization of real symmetric and orthogonal matrices. We invite the reader to consult [48] for details about these standard results. Denote as $M_p(\mathbb{R})$ the set of $p \times p$ real matrices, $SO(p)$ the set of $p \times p$ orthogonal matrices and $\mathbf{M}^T$ the transpose of $\mathbf{M}$. The identity matrix of $M_p(\mathbb{R})$ will be denoted as $\mathbf{I}_p$. Several subsets of matrices used in the sequel are defined below
  • $\mathcal{D}$ is the subset of $M_p(\mathbb{R})$ defined by the symmetric positive definite matrices;
  • $\overline{\mathcal{D}}$ is the closure of $\mathcal{D}$ in $M_p(\mathbb{R})$, i.e., the subset of $M_p(\mathbb{R})$ defined by the symmetric non negative definite matrices;
  • For all $\alpha > 0$

$$\mathcal{D}(\alpha) = \{\mathbf{M} \in \mathcal{D} \mid \|\mathbf{M}\| = \alpha\},$$
$$\overline{\mathcal{D}}(\alpha) = \{\mathbf{M} \in \overline{\mathcal{D}} \mid \|\mathbf{M}\| = \alpha\}, \quad (14)$$

where $\overline{\mathcal{D}}(\alpha)$ is a compact subset of $M_p(\mathbb{R})$, $\| \cdot \|$ being the Frobenius norm.

For $\mathbf{M} \in \mathcal{D}$, we introduce the open-half line spanned by $\mathbf{M}$ defined by $\mathcal{L}_{\mathbf{M}} = \{\lambda \mathbf{M}, \lambda > 0\}$. Note that the order associated with the cone structure of $\mathcal{D}$ is called the Loewner order for symmetric matrices of $M_p(\mathbb{R})$ and is defined as follows: for any pair of two symmetric $p \times p$ real matrices $(\mathbf{A}, \mathbf{B})$, $\mathbf{A} \leq \mathbf{B}$ ($\mathbf{A} < \mathbf{B}$ respectively) means that the quadratic form defined by $\mathbf{B} - \mathbf{A}$ is non negative (positive definite respectively), i.e., for all non zero $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{x}^T(\mathbf{A} - \mathbf{B})\mathbf{x} \geq 0$, ($>0$ respectively). Using that order, one has $\mathbf{M} \in \mathcal{D}$ ($\in \overline{\mathcal{D}}$ respectively) if and only if $\mathbf{M} > \mathbf{0}$ ($\mathbf{M} \geq \mathbf{0}$ respectively).

This section will make use of the following two applications

$$F_\chi : \mathcal{D} \to \mathbb{R}^+ \setminus \{0\}$$
$$\mathbf{M} \mapsto |\mathbf{M}|^{-1} \left(\sum_{i=1}^{N} y_i^\beta\right)^{-p/\beta} \quad (15)$$

and

$$f_\chi : \mathcal{D} \to \mathcal{D}$$
$$\mathbf{M} \mapsto \frac{1}{N} \sum_{i=1}^{N} \frac{Np}{y_i + y_i^{1-\beta} \sum\limits_{j \neq i} y_j^\beta} \mathbf{x}_i \mathbf{x}_i^T \quad (16)$$

where $\chi = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, $y_i = \mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i$ and $\beta \in (0, 1)$. The function F_χ is the likelihood of $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ in which the parameter m has been replaced by its estimator (8), up to a multiplicative constant and a power factor. Indeed

$$\prod_{i=1}^{N} p(\mathbf{x}_i | \mathbf{M}, \hat{m}, \beta) = \left[\frac{\beta \Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{p}{2\beta}\right) 2^{\frac{p}{2\beta}}} \right]^N \exp\left(-\frac{pN}{2\beta}\right)$$

$$\times \left(\frac{pN}{\beta}\right)^{\frac{pN}{2\beta}} [F_\chi(\mathbf{M})]^{\frac{N}{2}}.$$

It is clear that F_χ is homogeneous of degree zero whereas f_χ is homogeneous of degree one, i.e., for all $\lambda > 0$ and $\mathbf{M} \in \mathcal{D}$, one has

$$F_\chi(\lambda \mathbf{M}) = F_\chi(\mathbf{M}), \quad f_\chi(\lambda \mathbf{M}) = \lambda f_\chi(\mathbf{M}).$$

In order to understand the relationships between the two functions F_χ and f_χ, we can compute the gradient of F_χ at $\mathbf{M} \in \mathcal{D}$. Straightforward computations lead to

$$\nabla F_\chi(\mathbf{M}) = F_\chi(\mathbf{M}) \mathbf{M}^{-1} \left[f_\chi(\mathbf{M}) - \mathbf{M} \right] \mathbf{M}^{-1}. \quad (17)$$

Clearly $\mathbf{M}$ is an FP of f_χ if and only if $\mathbf{M}$ is a critical point of the vector field defined by ∇F_χ on $\mathcal{D}$, i.e., $\nabla F_\chi(\mathbf{M}) = 0$.

Remark III.1: There are some close links between the MGGDs and the SIRV distributions (that are both specific elliptical distributions). However, all MGGDs are not SIRV distributions and conversely. As a consequence, the FP (9) associated with the MGGDs relies on the function f_χ which differs from the FP equation studied in [39] (which corresponds to the particular case $\beta = 0$) and from that of [40] which corresponds to SIRVs with random multipliers τ. Similarly, the shape of the function F_χ differs significantly from the likelihoods studied in [39] and [40] that are defined as products of integrals depending on the distribution of the unknown multipliers (see [39], [40] for more details).

In the sequel, we also use f^n for $n \geq 1$ to denote the n-th iterate of f, i.e., $f^n := f \circ \ldots \circ f$, where f is repeated n times. We also adopt the standard convention $f^0 := \mathrm{Id}_\mathcal{D}$, where $\mathrm{Id}_\mathcal{D}$ is the identity function defined in $\mathcal{D}$. To finish this section, we introduce an important assumption about the vectors $\mathbf{x}_i$ for $1 \leq i \leq N$

- (H): For any set of p indices belonging to $[\![1, N]\!]$ and satisfying $i(1) < \ldots < i(p)$, the vectors $\mathbf{x}_{i(1)}, \ldots, \mathbf{x}_{i(p)}$ are linearly independent.

This hypothesis is a key assumption for obtaining all our subsequent results. Hypothesis (H) has the following trivial but fundamental consequence that we state as a remark.

Remark III.2: For all n vectors $\mathbf{x}_{i(1)}, \ldots, \mathbf{x}_{i(n)}$ with $1 \leq n \leq p$, $1 \leq i \leq N$, the vector space generated by $\mathbf{x}_{i(1)}, \ldots, \mathbf{x}_{i(n)}$ has dimension n.

B. Contributions

The contributions of this paper are summarized in the following theorems with proofs outlined in the next section.

Theorem III.1: For a given value of $\beta \in (0, 1)$, there exists $\widehat{\mathbf{M}}_{FP} \in \mathcal{D}$ with unit norm such that, for all $\alpha > 0$, f_χ admits a unique FP of norm $\alpha > 0$ equal to $\alpha \widehat{\mathbf{M}}_{FP}$. Moreover, F_χ reaches its maximum in $\mathcal{L}_{\widehat{\mathbf{M}}_{FP}}$, the open half-line spanned by $\widehat{\mathbf{M}}_{FP}$.

Consequently, $\widehat{\mathbf{M}}_{FP}$ is the unique positive definite $p \times p$ matrix of norm one satisfying

$$\widehat{\mathbf{M}}_{FP} = p \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\hat{y}_i + \hat{y}_i^{1-\beta} \sum_{j \neq i} \hat{y}_j^\beta} \quad (18)$$

where $\hat{y}_i = \mathbf{x}_i^T \widehat{\mathbf{M}}_{FP}^{-1} \mathbf{x}_i$.

Remark III.3: Theorem III.1 relies on the fact that F_χ reaches its maximum in $\mathcal{D}$. In order to prove this result, the function F_χ is continuously extended by the zero function on the boundary of $\mathcal{D}$, except for the zero matrix. Since F_χ is positive and bounded in $\mathcal{D}$, we can conclude (see Appendix A for details).

As a consequence of Theorem III.1, the following result can be obtained.

Theorem III.2: Let S be the discrete dynamical system defined on $\mathcal{D}$ by the recursion

$$\mathbf{M}_{k+1} = f_\chi(\mathbf{M}_k). \quad (19)$$

Then, for all initial conditions $\mathbf{M}_0 \in \mathcal{D}$, the resulting sequence $(\mathbf{M}_k)_{k \geq 0}$ converges to an FP of f_χ, i.e., to a point where F_χ reaches its maximum.

Theorem III.2 can be used to characterize numerically the points where F_χ reaches its maximum and the value of this maximum. Note that the algorithm defined by (19) does not allow the norm of the FP to be controlled. Therefore, for practical convenience, a slightly modified algorithm can be used in which a $\mathbf{M}$-normalization is applied at each iteration. This modified algorithm is proposed in the following corollary

Corollary III.1: The recursion

$$\mathbf{M}'_{k+1} = \frac{f_\chi(\mathbf{M}'_k)}{\mathrm{Tr}\left[f_\chi(\mathbf{M}'_k) \right]} \quad (20)$$

initialized by

$$\mathbf{M}'_0 = \frac{\mathbf{M}_0}{\mathrm{Tr}(\mathbf{M}_0)} \quad (21)$$

yields a sequence of matrices $\{\mathbf{M}'_0, \ldots, \mathbf{M}'_k\}$ which converges to the FP $\widehat{\mathbf{M}}_{FP}$ up to a scaling factor. Moreover, the matrices $\{\mathbf{M}'_0, \ldots, \mathbf{M}'_k\}$ are related to $\{\mathbf{M}_0, \ldots, \mathbf{M}_k\}$ by

$$\mathbf{M}'_i = \frac{\mathbf{M}_i}{\mathrm{Tr}(\mathbf{M}_i)}, \quad 1 \leq i \leq k.$$

IV. PROOF OUTLINE

This section provides the proofs of Theorems III.1 and III.2. Each proof is decomposed into a sequence of lemmas and propositions whose arguments are postponed in the Appendices. For the proofs that can be directly obtained from those of [39], we refer to [39]. In these cases, the differences due to the definitions of the function f_χ and the MGGD model for the observed vectors $\mathbf{x}_i$, for $i = 1, \ldots, N$, imply only slight modifications.

A. Proof of Theorem III.1

The proof of Theorem III.1 is the consequence of several propositions whose statements are listed below. The first proposition shows the existence of an FP satisfying (9).

Proposition IV.1: The supremum of F_χ in $\mathcal{D}$ is finite and is reached at a point $\widehat{\mathbf{M}}_{FP} \in \mathcal{D}$ with $\|\widehat{\mathbf{M}}_{FP}\| = 1$. Therefore, f_χ admits the open-half line $\mathcal{L}_{\widehat{\mathbf{M}}_{FP}}$ as fixed points.

Proof: See Appendix A. ∎

It remains to show that there is no other FP of f_χ than those belonging to $\mathcal{L}_{\widehat{\mathbf{M}}_{FP}}$. For that purpose, it is sufficient to show that all FPs of f_χ are collinear. However, Corollary V.1 of [39] indicates that all FPs of f_χ are collinear if all the orbits of f_χ are bounded in $\mathcal{D}$. We recall here that the *orbit* of f_χ associated with $\mathbf{M} \in \mathcal{D}$ is the trajectory of the dynamical system S defined in (19) starting at $\mathbf{M}$ (See [49] for more details about orbits in dynamical systems). Moreover, according to [39], when a function f_χ admits an FP, every orbit of f_χ is bounded if the following proposition is verified.

Proposition IV.2: The function f_χ verifies the following properties

- (P1): For all $\mathbf{M}, \mathbf{Q} \in \mathcal{D}$, if $\mathbf{M} \leq \mathbf{Q}$, then $f_\chi(\mathbf{M}) \leq f_\chi(\mathbf{Q})$ (also true with strict inequalities);
- (P2): for all $\mathbf{M}, \mathbf{Q} \in \mathcal{D}$, then

$$f_\chi(\mathbf{M} + \mathbf{Q}) \geq f_\chi(\mathbf{M}) + f_\chi(\mathbf{Q}), \tag{22}$$

where equality occurs if and only if $\mathbf{M}$ and $\mathbf{Q}$ are collinear.

Proof: Since the function f_χ used in this paper differs from the one used in [39], a specific analysis is required. It is the objective of Appendix B. ∎

To summarize, Proposition IV.1 establishes the existence of matrices satisfying the FP (9) while Proposition IV.2 together with the results of [39] can be used to show that there is a unique matrix of norm 1 satisfying (9).

B. Proof of Theorem III.2

In order to prove Theorem III.2, we have to show that each orbit of f_χ converges to an FP of f_χ. For that purpose, we consider for all $\mathbf{M} \in \mathcal{D}$ the positive limit set $\omega(\mathbf{M})$ associated with $\mathbf{M}$, i.e., the set of cluster points of the sequence $(\mathbf{M}_k)_{k \geq 0}$ when k tends to infinity, where $\mathbf{M}_{k+1} = f_\chi(\mathbf{M}_k)$ and $\mathbf{M}_0 = \mathbf{M}$. Since the orbit of f_χ associated with $\mathbf{M}$ is bounded in $\mathcal{D}$, the set $\omega(\mathbf{M})$ is a compact of $\mathcal{D}$ and is invariant by f_χ: for all $\mathbf{P} \in \omega(\mathbf{M})$, $f_\chi(\mathbf{P}) \in \omega(\mathbf{M})$. It is clear that the sequence $(\mathbf{M}_k)_{k \geq 0}$ converges if and only if $\omega(\mathbf{M})$ reduces to a single point. According to [39], $\omega(\mathbf{M})$ reduces to a single point if the following proposition is satisfied.

Proposition IV.3: The function f_χ is eventually strictly increasing, i.e., for all $\mathbf{Q}, \mathbf{P} \in \mathcal{D}$ such that $\mathbf{Q} \geq \mathbf{P}$ and $\mathbf{Q} \neq \mathbf{P}$, then

$$f_\chi^p(\mathbf{Q}) > f_\chi^p(\mathbf{P}). \tag{23}$$

Proof: Since the function f_χ used in this paper differs from the one used in [39], a specific analysis is required. It is the objective of Appendix C. ∎

V. SIMULATIONS

This section presents simulation results to evaluate the performance of the MLE for the parameters of MGGDs. The first scenario considers i.i.d. p dimensional data vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ distributed according to an MGGD. These vectors have been generated using the stochastic representation (3) with a matrix $\mathbf{M}$ defined as

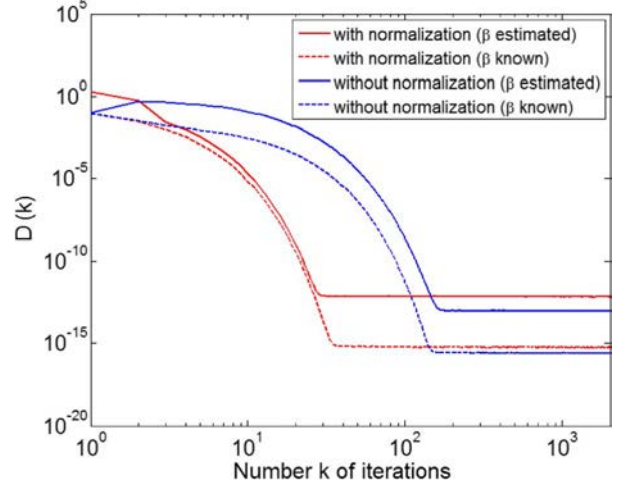$$\mathbf{M}(i, j) = \rho^{|i-j|} \quad \text{for} \quad i, j \in [\![0, p-1]\!]. \tag{24}$$

Fig. 1. Variations of $D(k)$ versus number of iterations for $p = 3$, $\beta = 0.2$ and $\rho = 0.8$.

In the following, 1000 Monte Carlo runs have been averaged in all experiments to evaluate the performance of the proposed estimation algorithms. Before analyzing the performance of the FP estimators based on (9), we illustrate the importance of the normalization $\mathrm{Tr}(\mathbf{M}) = p$ advocated in this paper.

A. Influence of the Normalization

The main advantage of the normalization (i.e., decomposition of $\boldsymbol{\Sigma}$ as the product $m \times \mathbf{M}$, and trace constraint for the matrix $\mathbf{M}$) concerns the convergence speed of the algorithm. To illustrate this point, Fig. 1 shows the evolution of the criterion $D(k)$

$$D(k) = \frac{\|\hat{\mathbf{A}}_{k+1} - \hat{\mathbf{A}}_k\|}{\|\hat{\mathbf{A}}_k\|}, \tag{25}$$

where $\hat{\mathbf{A}}_k = \hat{\boldsymbol{\Sigma}}_k$ for the blue curves and $\hat{\mathbf{A}}_k = \hat{m}_k \hat{\mathbf{M}}_k$ for the red curves. $\|\cdot\|$ is the Frobenius norm and $\hat{\mathbf{A}}_k$ is the estimator of $\mathbf{A}$ at step k. As observed, the convergence speed is significantly faster when a normalization condition is imposed at each iteration of the algorithm.

Fig. 2 shows the evolution of the estimated bias and consistency of $\hat{\mathbf{A}}$ (the plain curves correspond to $\hat{\mathbf{A}} = \hat{m}\hat{\mathbf{M}}$ whereas $\hat{\mathbf{A}} = \hat{\boldsymbol{\Sigma}}$ for the dotted lines) versus the number of samples when β is not estimated (the parameters are $\beta = 0.2$, $\rho = 0.8$ and $p = 3$). The estimated bias of $\hat{\mathbf{A}}$ is defined as $\|\overline{\mathbf{A}} - \mathbf{A}\|$ where the operator $\overline{\mathbf{A}}$ is the empirical mean of the estimated matrices

$$\overline{\mathbf{A}} = \frac{1}{I} \sum_{i=1}^{I} \hat{\mathbf{A}}(i). \tag{26}$$

For a given sample size, the experiments are performed I times ($I = 100$ in the following). Note that the bias criterion based on (26) was used in [47] for assessing the performance of matrix estimators. Note also that other approaches based on computing the mean on the manifold of positive definite matrices could also be investigated [50], [51]. The estimated consistency of $\hat{\mathbf{A}}$ is verified by computing $\|\hat{\mathbf{A}} - \mathbf{A}\|$. As observed, the estimation performance is the same when a normalization constraint for the scatter matrix is imposed or not.
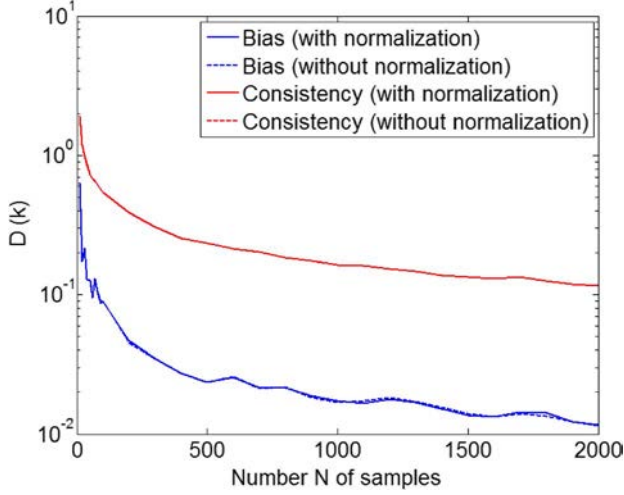
Fig. 2. Influence of the normalization of the scatter matrix on the estimation performance: estimated bias and consistency versus number of samples N.
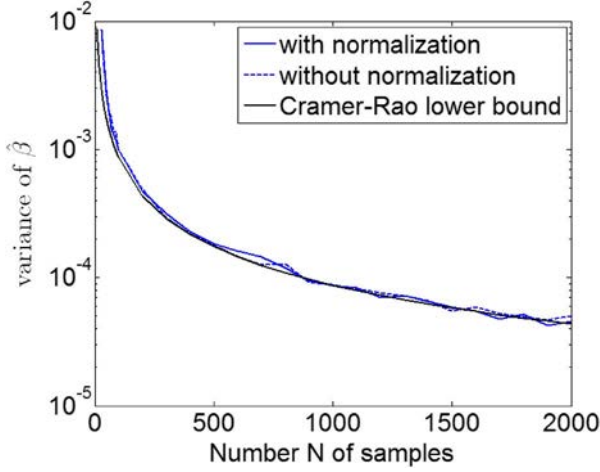

Fig. 3. Variance of $\hat{\beta}$ versus number of samples N.

A similar comment can be made for the shape parameter β when this parameter is estimated (see Fig. 3). The Fisher information matrix has been recently derived for the parameters of MGGDs [36]. It has been shown that this matrix only depends on the sample size N and the shape parameter β. The Cramér-Rao lower bounds (CRLBs) for the MGGD parameters can then be obtained by inverting the Fisher information matrix. These CRLBS provide a reference (in terms of variance or mean square error) for any unbiased estimator of the MGGD parameters. As observed in Fig. 3, the variance of $\hat{\beta}$ is very close to the Cramér-Rao lower bound for normalized or non-normalized scatter matrices.

To summarize, the normalization of the scatter matrix (decomposition of $\mathbf{\Sigma}$ as the product $m \times \mathbf{M}$, and trace constraint for the matrix $\mathbf{M}$) does not affect the statistical properties of the MLE. However, it ensures an increased convergence speed of the algorithm. Note also that a similar normalization was proposed in [46, Eq. (15)].

B. Known Shape Parameter

1) Convergence of the Scatter Matrix MLE: Fig. 4 shows some convergence results associated with the MLE of the scatter
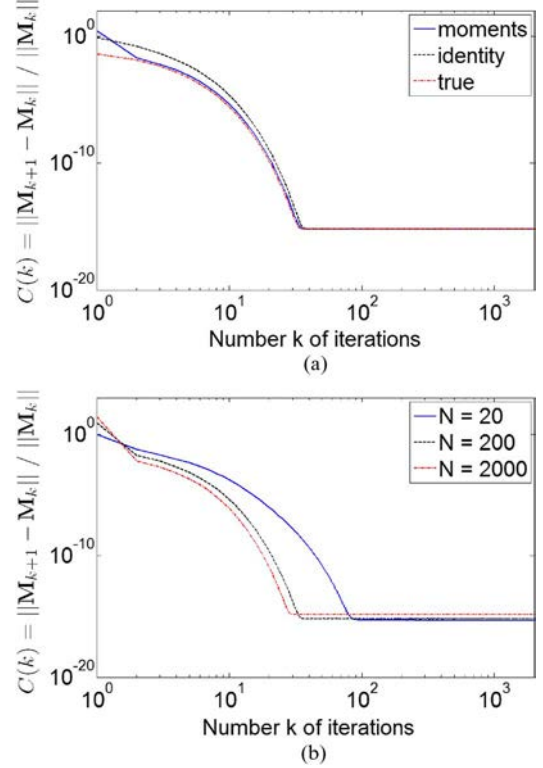

Fig. 4. Variations of $C(k)$ for $p = 3$, $\beta = 0.2$ and $\rho = 0.8$. (a) $C(k)$ versus number of iterations for different initializations ($N = 200$). (b) $C(k)$ versus number of iterations for various values of N.

matrix $\mathbf{M}$. These results have been obtained for $p = 3$, $\beta = 0.2$ (shape parameter) and $\rho = 0.8$. Convergence results are first analyzed by evaluating the sequence of criteria $C(k)$ defined as

$$C(k) = \frac{\|\widehat{\mathbf{M}}_{k+1} - \widehat{\mathbf{M}}_k\|}{\|\widehat{\mathbf{M}}_k\|}. \tag{27}$$

Fig. 4(a) shows examples of criteria $C(k)$ obtained for various initial matrices $\mathbf{M}_0$ ("moments" stands for $\mathbf{M}_0$ equal to the estimator of moments [36], "identity" stands for $\mathbf{M}_0 = \mathbf{I}_p$ and "true" corresponds to $\mathbf{M}_0 = \mathbf{M}$). After about 20 iterations, all curves converge to the same values. Hence, the convergence speed of the proposed algorithm seems to be independent of its initialization. Fig. 4(b) shows the evolution of criteria $C(k)$ for various values of N. It can be observed that the convergence speed increases with N as expected.

2) Bias and Consistency Analysis: Fig. 5(a) shows the estimated bias of $\hat{\mathbf{M}}$ for different values of β (precisely for $\beta \in \{0.2, 0.5, 0.8\}$). As observed, the bias converges very fast to a small value which is independent of β. Some consistency results for the proposed estimator are displayed in Fig. 5(b). Here, a plot of $\|\hat{\mathbf{M}} - \mathbf{M}\|$ as a function of the number of samples N is shown for different values of β (0.2, 0.5 and 0.8). It can be noticed that, for any value of β, $\|\hat{\mathbf{M}} - \mathbf{M}\|$ converges to a small value when N increases.

C. Unknown Shape Parameter

When β is unknown, the MLE of $\mathbf{M}$ and β is defined by (9) and (13). If $\mathbf{M}$ would be known, one might think of using a Newton-Raphson procedure to estimate β. The
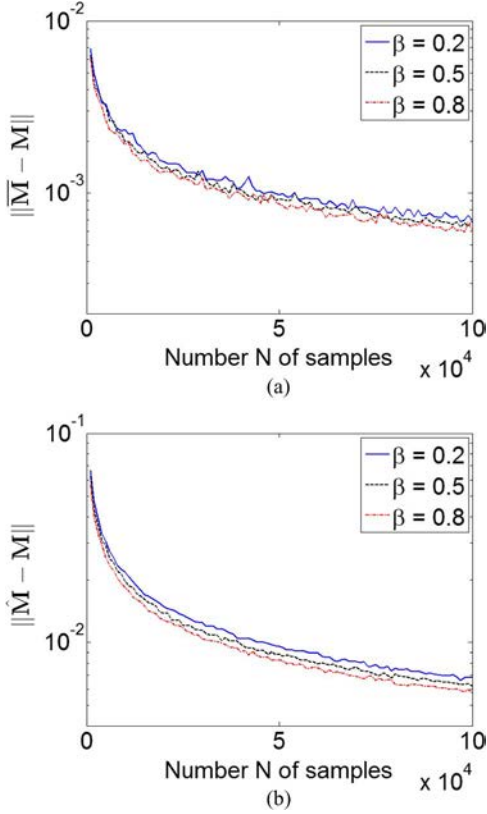
Fig. 5. (a) Estimated bias for different values of β, (b) estimated consistency for different values of β.

Newton-Raphson recursion based on (13) is defined by the following recursion

$$\hat{\beta}_{n+1} = \hat{\beta}_n - \frac{\alpha(\hat{\beta}_n)}{\alpha'(\hat{\beta}_n)} \qquad (28)$$

where $\hat{\beta}_n$ is an estimator of β at step n, and the function $\alpha(\beta)$ has been defined in (13). In practice, when the parameters $\mathbf{M}$ and β are unknown, we propose the following algorithm to estimate the MGGD parameters.

Algorithm 1: MLE for the parameters of MGGDs

1: Initialization of β and $\mathbf{M}$.

2: **for** $k = 1 : \text{N\_iter\_max}$ **do**

3:   Estimation of $\mathbf{M}$ using one iteration of (9) and normalization.

4:   Estimation of β by a Newton-Raphson iteration combining (13) and (28).

5: **end for**

6: Estimation of m using (8).

---

1) Bias and Consistency Analysis: Fig. 6 compares the algorithm performance when the shape parameter β is estimated (solid line) and when it is known (dashed line). As observed, the simulation results obtained with the proposed algorithm are very similar to those obtained for a fixed value of β.
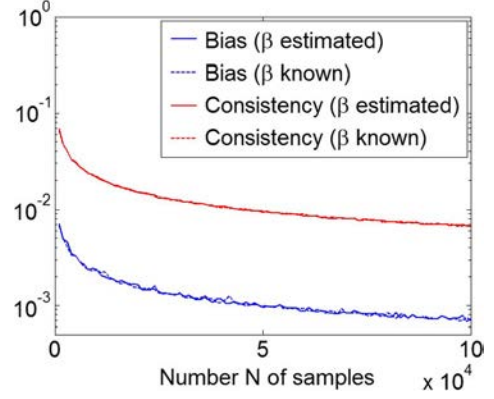

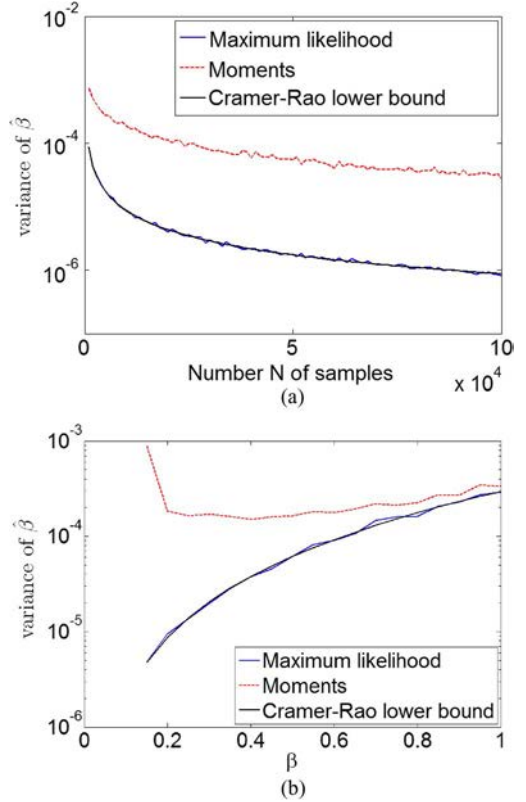
Fig. 6. Estimated bias and consistency for $\beta = 0.2$.

Fig. 7. Estimation performance for parameter β. (a) Variance of $\hat{\beta}$ versus number of samples N for $\beta = 0.2$, $\rho = 0.8$ and $p = 3$, (b) Variance of $\hat{\beta}$ versus β for $N = 10\,000$, $\rho = 0.8$ and $p = 3$.

2) Shape Parameter β: A comparison between the variances of estimators resulting from the method of moments and the ML principle as well as the corresponding CRLBs are depicted in Fig. 7 (versus the number of samples and versus β). Fig. 7(a) was obtained for $\beta = 0.2$, $\rho = 0.8$ and $p = 3$, while Fig. 7(b) corresponds to $N = 10000$, $\rho = 0.8$ and $p = 3$. The ML method yields lower estimation variances compared to the moment-based approach, as expected. Moreover, the CRLB of β is very close to the variance of $\hat{\beta}$ in all cases, illustrating the MLE's efficiency.

D. Experiments in a Real-World Setting

In this part, we propose to evaluate the performance of the MLE for the parameters of MGGDs encountered in a real-world
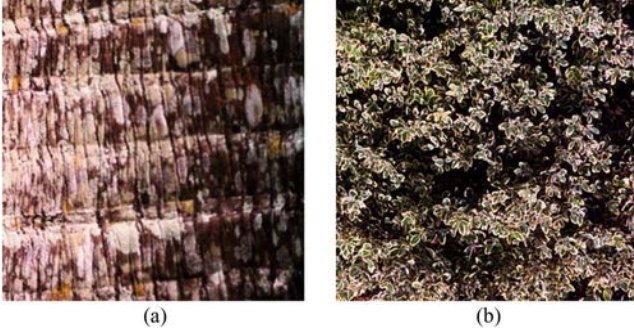
Fig. 8. Images from the VisTex database. (a) Bark.0000 and (b) Leaves.0008.

TABLE I
ESTIMATED MGGD PARAMETERS FOR THE FIRST SUBBAND OF
THE BARK.0000 AND LEAVES.0008 IMAGES

Image	$\hat{m}$	$\hat{\beta}$	$\hat{\mathbf{M}}$		
Bark 0000	0.036	0.328	0.988	0.992	0.883
			0.992	1.131	0.922
			0.883	0.922	0.881
Leaves 0008	0.054	0.265	0.935	0.966	0.871
			0.966	1.074	0.976
			0.871	0.976	0.991

application. MGGDs have been used successfully for modeling the wavelet statistics of texture images [37], [52]. In order to analyze the potential of MGGDs for texture modeling, we have considered two images from the VisTex database [53], namely the "Bark.0000" and "Leaves.0008" images displayed in Fig. 8. The red, green and blue channels of these images have been filtered using the stationary wavelet transform with the Daubechies db4 wavelet. For the first scale and orientation, the observed vector $\mathbf{x}$ (of size $p = 3$) contains the realizations of the wavelet coefficients for each channel of the RGB image. MGGD parameters have then been estimated using the proposed MLE for an unknown shape parameter (Algorithm 1), i.e., using the algorithm described in Section V-C. The results are reported in Table I. Fig. 9 compares the marginal distributions of the wavelet coefficients with the estimated MGGD and Gaussian distributions for the first subband of the red, green and blue channels (the top figures correspond to the image "Bark.0000" whereas the bottom figures are for the image "Leaves.0008"). These results illustrate the good potential of MGGDs for modeling color cue dependencies in texture images. For a more in depth analysis, the interested reader is referred to [37], [52] where goodness-of-fit results have been given for wavelet coefficients computed from a large database of natural texture images.

In the next experiments, we have generated 3-dimensional data vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ according to an MGGD with parameters given in Table I. Fig. 10 shows the MLE performance for these parameters resulting from real texture images. As observed in Fig. 10, the performance of the MLE of $\mathbf{M}$ is very similar when β is estimated or not (illustrating the unbiasedness and consistency properties of the scatter matrix estimator and the MLE efficiency of $\hat{\beta}$ that have also been observed for synthetic data).

VI. CONCLUSION

This paper has addressed the problem of estimating the parameters of multivariate generalized Gaussian distributions using the maximum likelihood method. For any shape parameter $\beta \in (0, 1)$, we have proved that the maximum likelihood estimator of the scatter matrix exists and is unique up to a scalar factor. By setting to zero the partial derivative with respect to the scale parameter of the likelihood associated with generalized Gaussian distributions, we obtain a closed form expression of the scale parameter as a function of the scatter matrix. The profile likelihood is then obtained by replacing this expression in the likelihood. The existence of the maximum likelihood estimator of the scatter matrix was proved by showing that this profile likelihood is positive, bounded in the set of symmetric positive definite matrices and equals zero on the boundary of this set. We have also proved that for any initial symmetric positive definite matrix, the sequence of matrices satisfying a fixed point equation converges to the unique maximum of this profile likelihood. Simulations results have illustrated the unbiasedness and consistency properties of the maximum likelihood estimator of the scatter matrix. Surprisingly, these unbiasedness and consistency properties are preserved when the shape parameter of the generalized Gaussian distribution is estimated jointly with the other parameters. Further works include the use of multivariate generalized Gaussian distributions for various remote sensing applications including change detection, image retrieval and image classification.

APPENDIX A
PROOF OF PROPOSITION IV.1

First, it is interesting to note that if $\widehat{\mathbf{M}}_{FP}$ is an FP of f_χ, $\lambda\widehat{\mathbf{M}}_{FP}$ is also an FP of f_χ for all $\lambda > 0$. This property is a direct consequence of the homogeneity of degree one of f_χ. We start by demonstrating the following lemma.

Lemma A.1: The function F_χ can be extended as a continuous function of $\overline{\mathcal{D}} \setminus \{\mathbf{0}\}$ such that $F_\chi(\mathbf{M}) = \mathbf{0}$ for all non invertible matrix $\mathbf{M} \in \overline{\mathcal{D}} \setminus \{\mathbf{0}\}$.

Proof: It is enough to show that, for all non invertible $\mathbf{M} \in \overline{\mathcal{D}} \setminus \{\mathbf{0}\}$, and all sequence $(\mathbf{Q}_k)_{k \geq 0}$ of $\overline{\mathcal{D}}$ converging to zero such that $\mathbf{M} + \mathbf{Q}_k$ is invertible, we have

$$\lim_{k \to \infty} F_\chi(\mathbf{M} + \mathbf{Q}_k) = 0.$$

Using the definition of F_χ in (15), the following result can be obtained for all $k \geq 0$

$$F_\chi(\mathbf{M}+\mathbf{Q}_k) = \left(\sum_{i=1}^{N} \left[|\mathbf{M}+\mathbf{Q}_k| \left(\mathbf{x}_i^T(\mathbf{M}+\mathbf{Q}_k)^{-1}\mathbf{x}_i \right)^p \right]^{\frac{\beta}{p}} \right)^{-p/\beta}.$$

Since $-p/\beta < 0$, the conclusion holds true if $\exists i^*, 1 \leq i^* \leq N$ such that

$$\lim_{k \to \infty} \frac{1}{|\mathbf{M} + \mathbf{Q}_k|} \frac{1}{\left[\mathbf{x}_{i^*}^T(\mathbf{M} + \mathbf{Q}_k)^{-1}\mathbf{x}_{i^*} \right]^p} = 0.$$

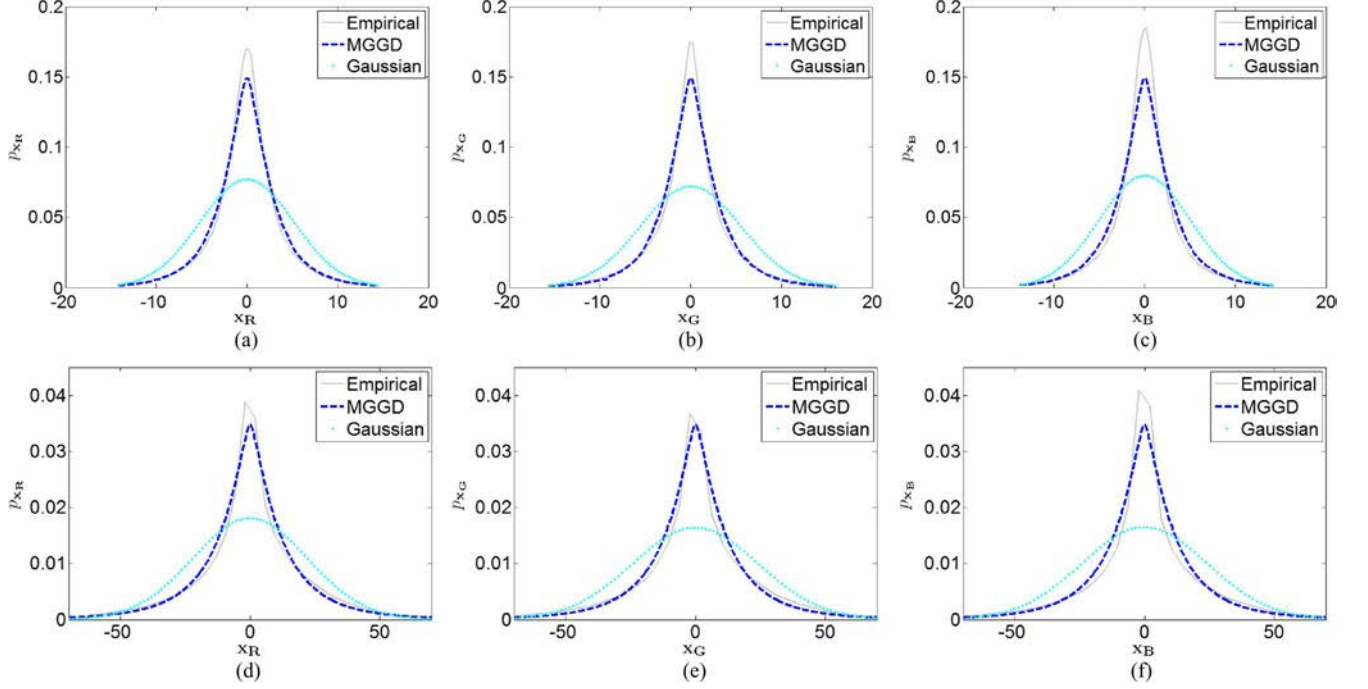which was demonstrated in [39] and concludes the proof. ∎

Fig. 9. Marginal distributions of the wavelet coefficients with the estimated MGGD and Gaussian distributions of the first subband for the red, green and blue channels of the Bark.0000 (a,b,c) and Leaves.0008 images (d,e,f).

End of the Proof of Proposition IV.1: The end of the proof of Proposition IV.1 is similar to the one given in [39]. Since F_χ is defined and continuous in the compact $\overline{\mathcal{D}}(1)$, this function reaches its maximum in $\overline{\mathcal{D}}(1)$ at a point denoted as $\widehat{\mathbf{M}}_{FP}$. Since F_χ is strictly positive in $\mathcal{D}(1)$ and equals 0 in $\overline{\mathcal{D}}(1) \setminus \mathcal{D}(1)$, the inequality $F_\chi(\widehat{\mathbf{M}}_{FP}) > 0$ leads to $\widehat{\mathbf{M}}_{FP} \in \mathcal{D}(1)$. In order to complete the proof of Proposition IV.1, we need to show the following lemma.

Lemma A.2: Let $\widehat{\mathbf{M}}_{FP} \in \mathcal{D}(1)$ maximizing the function F_χ. Then, $\nabla F_\chi(\widehat{\mathbf{M}}_{FP}) = \mathbf{0}$, which implies that $\widehat{\mathbf{M}}_{FP}$ is an FP of f_χ.

Proof: Since the function F_χ defined in (15) differs from the one used in [39], a specific analysis is required. By definition of $\widehat{\mathbf{M}}_{FP}$, one has

$$F_\chi(\widehat{\mathbf{M}}_{FP}) = \max_{\mathbf{M} \in \mathcal{D}(1)} F_\chi(\mathbf{M}).$$

By defining $\mathcal{N}(\mathbf{M}) = \|\mathbf{M}\|^2 - 1$, one has $\mathcal{N}(\widehat{\mathbf{M}}_{FP}) = 0$. The Lagrange theorem ensures that $\nabla F_\chi(\widehat{\mathbf{M}}_{FP}) = \lambda \nabla \mathcal{N}(\widehat{\mathbf{M}}_{FP}) = 2\lambda \widehat{\mathbf{M}}_{FP}$ for $\lambda \geq 0$. Straightforward computations lead to

$$\begin{aligned}
&\nabla F_\chi(\mathbf{M}).\nabla \mathcal{N}(\mathbf{M}) \\
&= \mathrm{Tr}\left[\nabla F_\chi(\mathbf{M})\nabla\mathcal{N}(\mathbf{M})\right] \\
&= 2F_\chi(\mathbf{M})\mathrm{Tr}\left[\mathbf{M}^{-1}\left(f_\chi(\mathbf{M}) - \mathbf{M}\right)\right] \\
&= 2F_\chi(\mathbf{M})\left(\mathrm{Tr}\left[\mathbf{M}^{-1}f_\chi(\mathbf{M})\right] - p\right) \\
&= 2F_\chi(\mathbf{M})\left(\frac{p}{N}\left(\frac{1}{N}\sum_{i=1}^{N}y_i^\beta\right)^{-1}\sum_{i=1}^{N}\frac{\mathrm{Tr}\left(\mathbf{M}^{-1}\mathbf{x}_i\mathbf{x}_i^T\right)}{y_i^{1-\beta}} - p\right) \\
&= 2F_\chi(\mathbf{M})\left(\frac{p}{N}\left(\frac{1}{N}\sum_{i=1}^{N}y_i^\beta\right)^{-1}\sum_{i=1}^{N}\frac{y_i}{y_i^{1-\beta}} - p\right) = 0.
\end{aligned}$$

Since $\nabla F_\chi(\widehat{\mathbf{M}}_{FP}) = 2\lambda\widehat{\mathbf{M}}_{FP}$, one has $2\lambda = 2\lambda\|\widehat{\mathbf{M}}_{FP}\|^2 = \nabla F_\chi(\widehat{\mathbf{M}}_{FP}).\widehat{\mathbf{M}}_{FP} = 0$ which completes the proof of Lemma A.2. ∎

APPENDIX B
PROOF OF PROPOSITION IV.2

We start by establishing $(P1)$. Let $\mathbf{M}, \mathbf{Q} \in \mathcal{D}$ with $\mathbf{M} \leq \mathbf{Q}$. Then, $\mathbf{M}^{-1} \geq \mathbf{Q}^{-1}$ and, for all $1 \leq i \leq N$, we have

$$\frac{1}{\mathbf{x}_i^T\mathbf{M}^{-1}\mathbf{x}_i + \left(\mathbf{x}_i^T\mathbf{M}^{-1}\mathbf{x}_i\right)^{1-\beta}\sum_{j\neq i}\left(\mathbf{x}_j^T\mathbf{M}^{-1}\mathbf{x}_j\right)^\beta}$$

$$\leq \frac{1}{\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i + \left(\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i\right)^{1-\beta}\sum_{j\neq i}\left(\mathbf{x}_j^T\mathbf{Q}^{-1}\mathbf{x}_j\right)^\beta},$$

which proves the property $(P1)$. The reasoning for the case with strict inequalities is identical.

We next turn to the proof of $(P2)$. Using the definition of f_χ in (16), the following result can be easily obtained

$$f_\chi(\mathbf{M}) = \frac{p}{N}\left(\frac{1}{N}\sum_{i=1}^{N}y_i^\beta\right)^{-1}\sum_{i=1}^{N}\frac{\mathbf{x}_i\mathbf{x}_i^T}{y_i^{1-\beta}}. \tag{B.29}$$

For all unit vector $\mathbf{x} \in \mathbb{R}^p$ such that $\|\mathbf{x}\| = 1$ and all $\mathbf{M} \in \mathcal{D}$, it is well known that

$$\frac{1}{\mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}} = \min_{\mathbf{z}^T\mathbf{x}\neq 0}\frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{x}^T\mathbf{z})^2}, \tag{B.30}$$
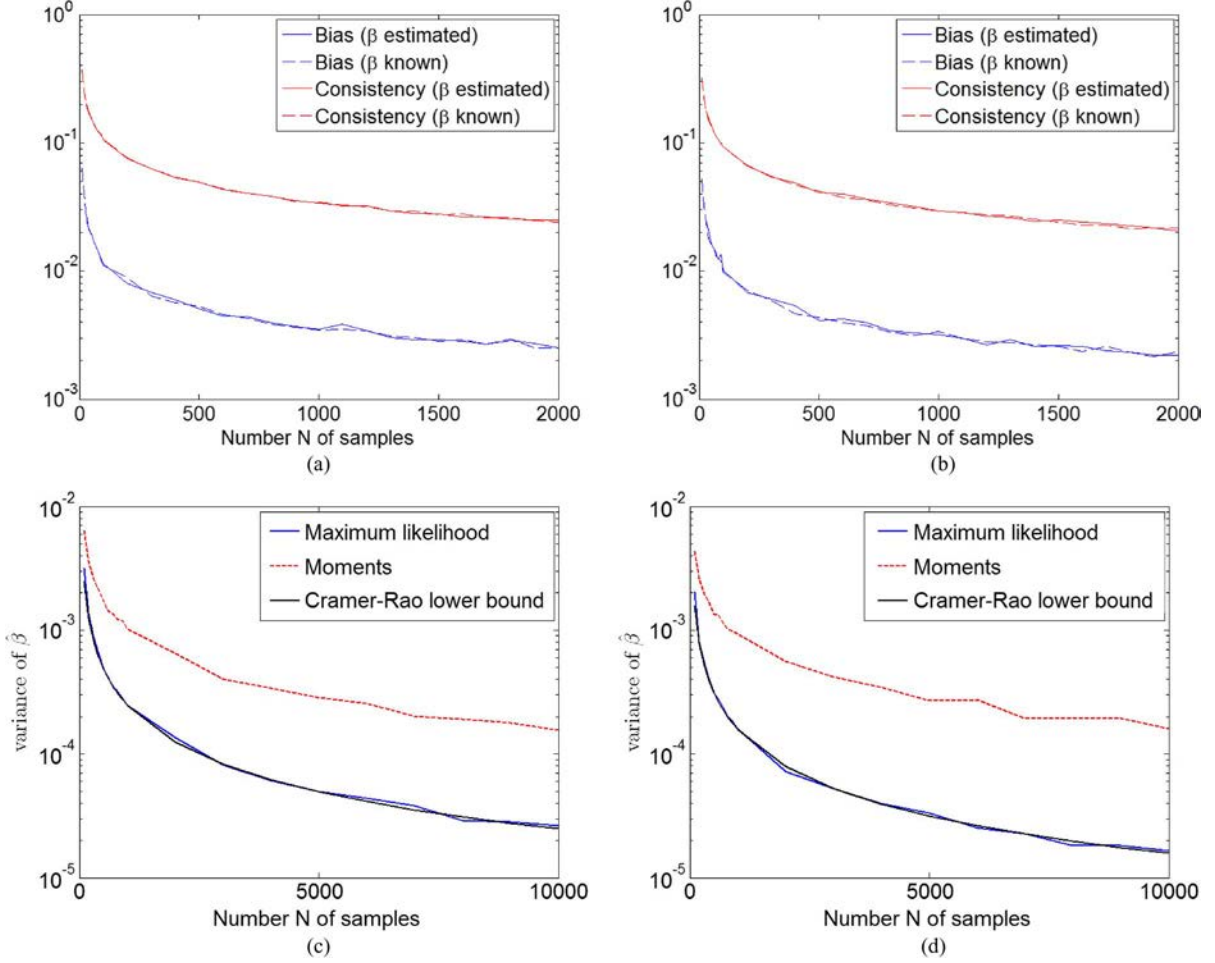
Fig. 10. Estimation performance in a real-world setting. Estimated bias and consistency for (a) the Bark.0000 and (b) the Leaves.0008 settings. Variance of $\hat{\beta}$ versus number of samples N for (c) the Bark.0000 and (d) the Leaves.0008 settings.

where the minimum is reached for the vectors $\mathbf{z}$ belonging to the line generated by $\mathbf{M}^{-1}\mathbf{x}$. Moreover, since $\beta \in (0,1)$, one has

$$\left(\frac{1}{\mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}}\right)^{1-\beta} = \left(\inf_{\mathbf{z}^T\mathbf{x}\neq 0} \frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{x}^T\mathbf{z})^2}\right)^{1-\beta}$$

$$\left(\frac{1}{\mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}}\right)^{\beta} = \left(\inf_{\mathbf{z}^T\mathbf{x}\neq 0} \frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{x}^T\mathbf{z})^2}\right)^{\beta}.$$

For $\mathbf{M}, \mathbf{Q} \in \mathcal{D}$, after noting that the function $f_\chi(\mathbf{M})$ is unchanged if we replace each vector $\mathbf{x_i}$ by the normalized vector $\mathbf{n_i} = \mathbf{x_i}/\|\mathbf{x_i}\|$, the following results can be obtained

$$f_\chi(\mathbf{M} + \mathbf{Q})$$

$$= \frac{p}{N} g_{\mathbf{n}}(\mathbf{M}+\mathbf{Q}) \sum_{i=1}^{N} \mathbf{n}_i \mathbf{n}_i^T \left[\min_{\mathbf{z}^T\mathbf{n}_i\neq 0} \frac{\mathbf{z}^T(\mathbf{M}+\mathbf{Q})\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2}\right]^{1-\beta}$$

$$= \frac{p}{N} g_{\mathbf{n}}(\mathbf{M}+\mathbf{Q}) \sum_{i=1}^{N} \mathbf{n}_i \mathbf{n}_i^T \left(\min_{\mathbf{z}^T\mathbf{n}_i\neq 0} \left[\frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2} + \frac{\mathbf{z}^T\mathbf{Q}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2}\right]\right)^{1-\beta},$$

where

$$g_{\mathbf{n}}(\mathbf{M}) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{n}_i^T\mathbf{M}^{-1}\mathbf{n}_i\right)^{\beta}\right]^{-1}.$$

More generally

$$\min_{\mathbf{z}\in\mathcal{A}}\left[f_1(\mathbf{z}) + f_2(\mathbf{z})\right] \geq \min_{\mathbf{z}\in\mathcal{A}} f_1(\mathbf{z}) + \min_{\mathbf{z}\in\mathcal{A}} f_2(\mathbf{z}),$$

for all functions f_1, f_2 and set $\mathcal{A}$ giving a sense to the previous inequality. The same reasoning can be applied to the function $g_{\mathbf{n}}(.)$ introduced above. Thus, $(P2)$ clearly holds true. It remains to study when equality occurs in $(P2)$. The property $(P2)$ becomes an equality if and only if, for all $1 \leq i \leq N$, one has

$$\min_{\mathbf{z}^T\mathbf{n}_i\neq 0}\left[\frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2} + \frac{\mathbf{z}^T\mathbf{Q}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2}\right]$$

$$= \min_{\mathbf{z}^T\mathbf{n}_i\neq 0} \frac{\mathbf{z}^T\mathbf{M}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2} + \min_{\mathbf{z}^T\mathbf{n}_i\neq 0} \frac{\mathbf{z}^T\mathbf{Q}\mathbf{z}}{(\mathbf{n}_i^T\mathbf{z})^2}, \quad \text{(B.31)}$$

which was shown in [39] to be true if and only if $\mathbf{M}$ and $\mathbf{Q}$ are collinear. ∎

APPENDIX C
PROOF OF PROPOSITION IV.3

The proof of Proposition IV.3 is similar to the proof of Proposition V.3 of [39], even if the function f_χ used in this paper differs from the one defined in [39]. We first show that for all $\mathbf{Q}, \mathbf{P} \in \mathcal{D}$, we have

$$\text{If } \mathbf{Q} \geq \mathbf{P} \text{ and } f_\chi(\mathbf{Q}) = f_\chi(\mathbf{P}), \quad \text{then } \mathbf{Q} = \mathbf{P}. \quad \text{(C.32)}$$

Since $\mathbf{Q} \geq \mathbf{P}$ implies $\mathbf{P}^{-1} - \mathbf{Q}^{-1} \geq \mathbf{0}$, for all $1 \leq i \leq N$, we have

$$\frac{1}{\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i} \geq \frac{1}{\mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i}.$$

Assuming $f_\chi(\mathbf{Q}) = f_\chi(\mathbf{P})$ and using hypothesis (H) yields for all $1 \leq i \leq N$,

$$\left(\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i\right)^{1-\beta} \left(\sum_{j=1}^{N} \mathbf{x}_j^T \mathbf{Q}^{-1} \mathbf{x}_j\right)$$
$$= \left(\mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i\right)^{1-\beta} \left(\sum_{j=1}^{N} \mathbf{x}_j^T \mathbf{P}^{-1} \mathbf{x}_j\right).$$

Moreover, assuming that there exists i such that $\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i \neq \mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i$, i.e., $\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i < \mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i$ implies

$$\left(\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i\right)^{1-\beta} \left(\sum_{j=1}^{N} \mathbf{x}_j^T \mathbf{Q}^{-1} \mathbf{x}_j\right)$$
$$< \left(\mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i\right)^{1-\beta} \left(\sum_{j=1}^{N} \mathbf{x}_j^T \mathbf{P}^{-1} \mathbf{x}_j\right)$$

which contradicts $f_\chi(\mathbf{Q}) = f_\chi(\mathbf{P})$. Thus, $f_\chi(\mathbf{Q}) = f_\chi(\mathbf{P})$ yields $\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i = \mathbf{x}_i^T \mathbf{P}^{-1} \mathbf{x}_i$, for all $1 \leq i \leq N$. As a consequence, for all $1 \leq i \leq N$

$$\mathbf{x}_i^T (\mathbf{P}^{-1} - \mathbf{Q}^{-1}) \mathbf{x}_i = 0.$$

Since $\mathbf{P}^{-1} - \mathbf{Q}^{-1} \geq \mathbf{0}$, the previous equality indicates that $(\mathbf{P}^{-1} - \mathbf{Q}^{-1}) \mathbf{x}_i = \mathbf{0}$, for all $1 \leq i \leq N$. Using hypothesis (H), the claim (C.32) is proved.

We now move to the proof of Proposition IV.3. We consider $\mathbf{Q}, \mathbf{P} \in \mathcal{D}$ such that $\mathbf{Q} \geq \mathbf{P}$ and $\mathbf{Q} \neq \mathbf{P}$. As shown above, we have $f_\chi(\mathbf{Q}) \geq f_\chi(\mathbf{P})$ and $f_\chi(\mathbf{Q}) \neq f_\chi(\mathbf{P})$, which implies the existence of an index $i_0 \in [\![1, N]\!]$ such that

$$\xi_{i_0} := \frac{p}{N} \left(\frac{1}{\left(\mathbf{x}_{i_0}^T \mathbf{Q}^{-1} \mathbf{x}_{i_0}\right)^{1-\beta}} - \frac{1}{\left(\mathbf{x}_{i_0}^T \mathbf{P}^{-1} \mathbf{x}_{i_0}\right)^{1-\beta}} \right) > 0.$$

Note that for $\beta = 0$, this result reduces to what was obtained in [39]. Up to a relabel, we may assume that $i_0 = 1$, hence

$$f_\chi(\mathbf{Q}) \geq f_\chi(\mathbf{P}) + \xi_1 \mathbf{x}_1 \mathbf{x}_1^T \qquad \text{(C.33)}$$

which is the same result as the one obtained in [39]. As a consequence, the end of the proof of Proposition V.3 derived in [39] can be applied to our problem without any change.

REFERENCES

[1] M. T. Subbotin, "On the law of frequency of error," *Matheticheskii Sbornik*, vol. 31, no. 2, pp. 296–301, 1923.

[2] E. Gomez, M. A. Gomez-Villegas, and J.-M. Marin, "A multivariate generalization of the power exponential family of distributions," *Commun. Statist. Theory Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[3] E. Gomez, M. A. Gomez-Villegas, and J.-M. Marin, "A matrix variate generalization of the power exponential family of distributions," *Commun. Statist. Theory Methods*, vol. 31, no. 12, pp. 2167–2182, 2002.

[4] K.-S. Song, "A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 510–527, Feb. 2006.

[5] K. Fang and Y. Zhang, *Generalized Multivariate Analysis*. Berlin, Germany: Springer, 1990.

[6] K. Fang, S. Kotz, and K. Ng, *Symmetric Multivariate and Related Distributions*. New York, NY, USA: Chapman & Hall, 1990.

[7] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian J. Statist., Ser. A*, vol. 32, no. 4, pp. 419–430, Dec. 1970.

[8] M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 111–124, Jan. 1993.

[9] M. Rangaswamy, D. Weiner, and A. Ozturk, "Computer generation of correlated non-Gaussian radar clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 106–116, Jan. 1995.

[10] E. Gómez-Sánchez-Manzano, M. Gómez-Villegas, and J. Marín, "Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications," *Commun. Statist.—Theory Methods*, vol. 37, no. 6, pp. 972–985, 2008.

[11] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, no. 11, pp. 5597–5625, Nov. 2012.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[13] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

[14] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[15] L. Boubchir and J. Fadili, "Multivariate statistical modeling of images with the curvelet transform," in *Proc. Int. Symp. Signal Process. Its Appl.*, Sydney, Australia, Aug. 2005, pp. 747–750.

[16] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.

[17] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.

[18] D. Cho and T. D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Process. Image Commun.*, vol. 20, no. 1, pp. 77–89, Jan. 2005.

[19] D. Cho, T. D. Bui, and G. Y. Chen, "Image denoising based on wavelet shrinkage using neighbour and level dependency," *Int. J. Wavelets Multi.*, vol. 7, no. 3, pp. 299–311, May 2009.

[20] M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[21] G. Verdoolaege, S. D. Backer, and P. Scheunders, "Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models," in *Proc. Int. Conf. Image Process (ICIP)*, San Diego, CA, USA, Oct. 2008, pp. 169–172.

[22] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recogn.*, vol. 40, no. 2, pp. 619–634, Feb. 2007.

[23] J. Scharcanski, "A wavelet-based approach for analyzing industrial stochastic textures with applications," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 1, pp. 10–22, Jan. 2007.

[24] M. N. Desai and R. S. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3115–3127, Dec. 2003.

[25] M. Z. Coban and R. M. Mersereau, "Adaptive subband video coding using bivariate generalized Gaussian distribution model," in *Proc. IEEE ICASSP*, Atlanta, GA, USA, May 1996, pp. 1990–1993.

[26] M. Bicego, D. Gonzalo-Jimenez, E. Grosso, and J. L. Alba-Castro, "Generalized Gaussian distributions for sequential data classification," in *Proc. Int. Conf. Pattern Recogn.*, Tampa, FL, USA, Dec. 2008, pp. 1–4.

[27] J. Yang, Y. Wang, W. Xu, and Q. Dai, "Image and video denoising using adaptive dual-tree discrete wavelet packets," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 642–655, May 2009.

[28] T. Elguebaly and N. Bouguila, "Bayesian learning of generalized Gaussian mixture models on biomedical images," in *Proc. ANNPR*, Cairo, Egypt, Apr. 2010, pp. 207–218.

[29] S. Le Cam, A. Belghith, C. Collet, and F. Salzenstein, "Wheezing sounds detection using multivariate generalized Gaussian distributions," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 541–544.

[30] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution—Characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.

[31] M. Novey, T. Adali, and A. Roy, "Circularity and Gaussianity detection using the complex generalized Gaussian distribution," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 993–996, Nov. 2009.

[32] M. Liu and H. Bozdogan, "Multivariate regression models with power exponential random errors and subset selection using genetic algorithms with information complexity," *Eur. J. Pure Appl. Math.*, vol. 1, no. 1, pp. 4–37, 2008.

[33] G. Agro, "Maximum likelihood estimation for the exponential power function parameters," *Commun. Statist.-Simul. Computat.*, vol. 24, no. 2, pp. 523–536, 1995.

[34] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *J. Acoust. Soc. Amer.*, vol. 86, no. 4, pp. 1404–1415, Oct. 1989.

[35] N. Khelil-Cherif and A. Benazza-Benyahia, "Wavelet-based multivariate approach for multispectral image indexing," presented at the SPIE—Conf. Wavelet Appl. Indust. Process., Philadelphia, PA, USA, Oct. 2004.

[36] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *J. Math. Imag. Vision*, vol. 43, no. 3, pp. 180–193, May 2012.

[37] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 265–286, May 2011.

[38] T. Zhang, A. Wiesel, and M. Greco, "Multivariate generalized Gaussian distribution: Convexity and graphical models," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4141–4148, 2013.

[39] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[40] Y. Chitour and F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4563–4573, Oct. 2008.

[41] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, Jan. 1976.

[42] S. Kotz, "Multivariate distributions at a cross road," in *Statistical Distributions in Scientific Work, I*. Dordrecht, The Netherlands: Reidel, 1968, pp. 247–270.

[43] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York, NY, USA: Wiley , 1994.

[44] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Proces.*, vol. 9, no. 1, pp. 55–68, Jan. 2000.

[45] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[46] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, Jan. 2012.

[47] F. Pascal, P. Forster, J. P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2206–2217, Jun. 2008.

[48] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[49] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York, NY, USA: Springer-Verlag, 1983, vol. 42.

[50] S. Sra, "A new metric on the manifold of kernel matrices with application to matrix geometric means," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, pp. 144–152.

[51] S. Fiori and T. Tanaka, "An algorithm to compute averages on matrix Lie groups," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4734–4743, Dec. 2009.

[52] R. Kwitt, P. Meerwald, A. Uhl, and G. Verdoolaege, "Testing a multivariate model for wavelet coefficients," in *Proc. Int. Conf. Image Process. (ICIP)*, Brussels, Belgium, Sep. 2011, pp. 1301–1304.

[53] MIT Vision and Modeling Group, Vision Texture [Online]. Available: http://vismod.media.mit.edu/pub/VisTex

Frédéric Pascal was born in Sallanches, France in 1979. He received the Master's degree ("Probabilities, Statistics and Applications: Signal, Image and Networks") with merit, in Applied Statistics from University Paris VII—Jussieu, Paris, France, in 2003. Then, he received the Ph.D. degree of Signal Processing, from University Paris X—Nanterre, advised by Pr. Philippe Forster: "Detection and Estimation in Compound Gaussian Noise" in 2006. This Ph.D. thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France.

From November 2006 to February 2008, he made a post doctoral position in the Signal Processing and Information team of the laboratory SATIE (Système et Applications des Technologies de l'Information et de l'Energie), CNRS, Ecole Normale Supérieure, Cachan, France. From March 2008, he is an Associate Professor in the SONDRA laboratory, SUPELEC. His research interests are estimation in statistical signal processing and radar detection. He is currently on a one-year sabbatical leave in the Electrical and Computer Engineering (ECE) department of the National University of Singapore (NUS).

Lionel Bombrun (S'06–M'09) was born in Tournon, France, in 1982. He received the M.S. and the Ph.D. degrees in signal, image, speech, and telecommunications from the Grenoble National Polytechnic Institute (INPG), Grenoble, France, in 2005 and 2008, respectively.

In 2008, he was a Teaching Assistant at Phelma, Grenoble, France. From October 2009 to September 2010, he was a Postdoctoral Fellow from the French National Council for Scientific Research (CNRS) between the Grenoble Image Speech Signal Automatics Laboratory and SONDRA, Gif-sur-Yvette, France. From October 2010 to August 2011, he was a Postdoctoral Fellow at IMS Bordeaux (UMR 5218). He is currently an Associate Professor at Bordeaux Sciences Agro and a member of the IMS laboratory. His research interests include image processing, stochastic modeling, and texture analysis.

Jean-Yves Tourneret (SM'08) received the ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a Professor with the University of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods.

Dr. Tourneret has been involved in the organization of several conferences including the European conference on signal processing EUSIPCO'02 (Program Chair), the International Conference ICASSP'06 (plenaries), the Statistical Signal Processing Workshop SSP'12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangement), and the Statistical Signal Processing Workshop SSP'2014 (special sessions). He has been the General Chair of the CIMI Workshop on Optimization and Statistics in Image Processing, Toulouse, in 2013. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2008–2011) and for the EURASIP *Journal on Signal Processing* (since July 2013).

Yannick Berthoumieu received the Ph.D. degree in January 1996 in signal processing from the University of Bordeaux I, France.

In 1998, he joined the Ecole Nationale Supérieure d'Electronique, Informatique et Radiocommunications de Bordeaux as an Assistant Professor. He is currently a Professor with the Department of Telecommunications, Poly-technique Institute of Bordeaux, and the University of Bordeaux. His major research interests are concerned with multi-dimensional signal analysis, stochastic modeling, texture analysis, and video processing. Since 2003, he has been with the Signal and Image Processing Research Group, IMS Lab.