

PERFORMANCE OF THE MAXIMUM LIKELIHOOD ESTIMATORS FOR THE PARAMETERS OF MULTIVARIATE GENERALIZED GAUSSIAN DISTRIBUTIONS

Lionel Bombrun¹, Frédéric Pascal², Jean-Yves Tournier³ and Yannick Berthoumieu¹

¹ : Université de Bordeaux, ENSEIRB-Matmeca, Laboratoire IMS, Groupe Signal et Image

{lionel.bombrun, yannick.berthoumieu}@ims-bordeaux.fr

² : SONDR, Supélec, Plateau du Moulon, frederic.pascal@supelec.fr

³ : Université de Toulouse, IRT/INP-ENSEEIH, jean-yves.tournier@enseeiht.fr

ABSTRACT

This paper studies the performance of the maximum likelihood estimators (MLE) for the parameters of multivariate generalized Gaussian distributions. When the shape parameter belongs to $]0, 1[$, we have proved that the scatter matrix MLE exists and is unique up to a scalar factor. After providing some elements about this proof, an estimation algorithm based on a Newton-Raphson recursion is investigated. Some experiments illustrate the convergence speed of this algorithm. The bias and consistency of the scatter matrix estimator are then studied for different values of the shape parameter. The performance of the shape parameter estimator is finally addressed by comparing its variance to the Cramér-Rao bound.

Index Terms— Multivariate generalized Gaussian distribution, Newton-Raphson recursion, M-estimators.

1. INTRODUCTION

In various signal and image processing applications, the univariate generalized Gaussian (UGG) distribution has been introduced due to its more peaky and heavy-tailed shape compared to the univariate Gaussian distribution [1]. In order to capture inter-band dependencies in a wide-sense (multiscale, multichannel, spatial dependencies), different multivariate distributions which generalize the UGG distribution have been proposed in the literature. A copula-based model with UGG distributed marginals has notably been proposed in [2] [3] to model the multichannel dependencies of wavelet coefficients. An extension of the UGG distribution to the multivariate case, referred to as anisotropic multivariate generalized Gaussian distribution, has also been proposed in [4]. In order to satisfy elliptically contoured distribution properties [5] [6], a natural extension of the UGG distribution is to consider the multivariate generalized Gaussian distribution (MGGD) introduced in [7], also known as the multivariate power exponential distribution [8]. The MGGD is completely characterized by its scatter matrix Σ and its shape parameter β . This model has recently shown good properties for several image processing applications such as multispectral image indexing [9], image denoising [10] and texture image retrieval [11] [12]. In these applications, the unknown parameters Σ and β have to be estimated from the observed images. These parameters can be estimated by minimizing a χ^2 distance as in [9], or by minimizing an L^2 -norm as in [10]. Estimators based on the method of moments and on the maximum likelihood method have also been proposed in [11] [12]. However, the performance of these estimators has not been investigated, which is the main objective of this work. More precisely, the main contribution of the paper is to show that for $\beta \in]0, 1[$

(which corresponds to most of the real-life problems), the maximum likelihood estimator (MLE) of the MGGD scatter matrix exists and is unique up to a scalar factor. An iterative algorithm based on a Newton-Raphson recursion is then proposed to compute the MLE of the normalized MGGD scatter matrix. Some experiments are then conducted to evaluate the convergence speed of the algorithm as well as the bias and the consistency of the scatter matrix estimator. Some results regarding the estimation of the MGGD shape parameter are finally presented. The paper is structured as follows. Section 2 introduces the MGGD and the MLEs of its parameters. Section 3 derives properties of the scatter matrix MLE. Some simulations results are presented in Section 4 to evaluate the performance of the MLEs of the MGGD parameters. Conclusions and future works are finally reported in Section 5.

2. THE MULTIVARIATE GENERALIZED GAUSSIAN DISTRIBUTION (MGGD)

2.1. Definition and stochastic representation

The probability density function of an MGGD is [7]

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{M}, m, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} h\left(\mathbf{x}^T \mathbf{M}^{-1} \mathbf{x}\right) \quad (1)$$

where $h(\cdot)$ is a so-called density generator defined by

$$h(v) = \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p}{2\beta}\right)} \frac{\beta}{\pi^{\frac{p}{2}} m^{\frac{p}{2}} 2^{\frac{p}{2\beta}}} \exp\left(-\frac{v^{\beta}}{2m^{\beta}}\right) \quad (2)$$

for any $v \in \mathbb{R}^+$ and $\mathbf{M} = \Sigma/m$ is a normalized matrix such that $\text{tr}(\mathbf{M}) = p$ (where $\text{tr}(\mathbf{M})$ is the trace of the matrix \mathbf{M} and p is the dimension of the vector \mathbf{x}). Note that $\beta = 0.5$ corresponds to the multivariate Laplace distribution, while $\beta = 1$ corresponds to the multivariate Gaussian distribution. When β tends toward infinity, the MGGD is also known to reduce to the multivariate uniform distribution.

Let \mathbf{x} be a random vector distributed according to an MGGD of scatter matrix $\Sigma = m\mathbf{M}$ and shape parameter β . Gómez *et al.* have shown in [8] that \mathbf{x} admits the following stochastic representation

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

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

random variable such that

$$\tau^{2\beta} \sim \mathcal{G}\left(\frac{p}{2\beta}, 2\right) \quad (4)$$

where $\mathcal{G}(a, b)$ is the univariate Gamma distribution with parameters a and b (see [13] for definition).

2.2. MGGD parameter estimation

Let $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ be N vectors independent and identically distributed according to an MGGD. Since an MGGD is a particular real elliptical distribution, the MLE of the matrix \mathbf{M} satisfies the following fixed point equation [14]

$$\mathbf{M} = \frac{2}{N} \sum_{i=1}^N \frac{-g(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i)}{h(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i)} \mathbf{x}_i \mathbf{x}_i^T \quad (5)$$

where $g(v) = \frac{\partial h(v)}{\partial v}$. By replacing $h(\cdot)$ by its definition (2), the following result can be obtained

$$\mathbf{M} = \frac{\beta}{Nm^\beta} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^T}{(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i)^{1-\beta}} \text{ with } \text{tr}(\mathbf{M}) = p. \quad (6)$$

Note that the first term outside the summation is a scalar that can be encompassed in the constraint. Thus (6) reduces to the following M-estimator fixed point equation

$$\mathbf{M} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^T}{(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i)^{1-\beta}} \text{ with } \text{tr}(\mathbf{M}) = p. \quad (7)$$

Note that (7) reduces to the sample covariance matrix (SCM) estimator when $\beta = 1$ and to the fixed point covariance matrix estimator derived in [14] [15] for $\beta = 0$.

Moreover, by differentiating the joint distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ with respect to m and β , the following results can be obtained

$$\begin{aligned} f(\beta) &= \frac{pN}{2 \sum_{i=1}^N u_i^\beta} \sum_{i=1}^N \ln u_i u_i^\beta - \frac{pN}{2\beta} \left[\Psi\left(\frac{p}{2\beta}\right) + \ln 2 \right] \\ &- N - \frac{pN}{2\beta} \ln \left(\frac{\beta}{pN} \sum_{i=1}^N u_i^\beta \right) = 0 \end{aligned} \quad (8)$$

and

$$m = \left(\frac{\beta}{pN} \sum_{i=1}^N u_i^\beta \right)^{\frac{1}{\beta}} \quad (9)$$

where $u_i = \mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i$, and $\Psi(\cdot)$ is the digamma function. The MLEs of parameters \mathbf{M} , β and m are the solutions of (6), (7) and (8) that will be denoted $\hat{\mathbf{M}}$, $\hat{\beta}$ and \hat{m} . Equations (7) and (8) show that \mathbf{M} and β can be estimated independently from the scale parameter m . To solve (8) for a known matrix \mathbf{M} , we consider a Newton-Raphson procedure defined by

$$\hat{\beta}_{n+1} = \hat{\beta}_n - \frac{f(\hat{\beta}_n)}{f'(\hat{\beta}_n)} \quad (10)$$

where $\hat{\beta}_n$ is the estimator of β at step n , and the function $f(\beta)$ is defined in (8). In practice, when the parameters \mathbf{M} and β are unknown, we propose the following algorithm to estimate the MGGD parameters:

- 1: Initialisation of β and \mathbf{M} .
- 2: **for** $k = 1 : N_iter_max$ **do**
- 3: Estimation of \mathbf{M} using one iteration of (7) and normalization.
- 4: Estimation of β by a Newton-Raphson iteration combining (8) and (10).
- 5: **end for**
- 6: Estimation of m using (9).

Note that we have observed that the algorithm convergence is significantly faster when the normalization constraint $\text{tr}(\mathbf{M}) = p$ is imposed at each iteration.

3. PROPERTIES OF THE M-ESTIMATOR

MGGDs belong to the general class of elliptical distributions (ED) denoted E_p . Maronna derived in [16] very useful results for estimating the parameters of EDs. Let $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ be a N -sample (N independent and identically distributed vectors) of p -dimensional real vectors, with zero mean and distributed according to an ED, i.e., $\mathbf{x}_i \sim E_p(\mathbf{0}, \Lambda)$ for $i = 1, \dots, N$. The M-estimator of Λ is defined as the solution of the following equation

$$\mathbf{V}_N = \frac{1}{N} \sum_{n=1}^N u \left(\mathbf{x}_n^T \mathbf{V}_N^{-1} \mathbf{x}_n \right) \mathbf{x}_n \mathbf{x}_n^T \quad (11)$$

where $u(\cdot)$ is a function satisfying a set of general assumptions provided in [16] and recalled below

- (i) u is non-negative, non increasing, and continuous in $[0, \infty)$.
- (ii) If $K = \sup_{s \geq 0} \psi(s)$ with $p < K < \infty$ and $\psi(s) = s u(s)$, ψ is non decreasing and strictly increasing in the interval where $\psi < K$.
- (iii) Let $P_N(\cdot)$ the empirical distribution of $\mathbf{x}_1, \dots, \mathbf{x}_N$, there exists $a > 0$ such that for every hyperplane H such that $\dim(H) \leq p - 1$ we have

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

Theorem 1. Under the assumptions (i), (ii) and (iii), a solution of (11) exists and is unique. Moreover, this solution is consistent up to a scale factor and a simple iterative procedure can be used to determine \mathbf{V}_N .

Let us now consider the following equation

$$\mathbf{M} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^T}{(\mathbf{x}_i^T \mathbf{M}^{-1} \mathbf{x}_i)^{1-\beta}} \quad (13)$$

which corresponds to (7) without the normalization part. It can be noticed that the function $u(s) = s^{\beta-1}$ with $\beta \in]0, 1[$ does not verify conditions (i) (absence of continuity in 0) and (ii) ($\psi(s) = s^\beta$ is not bounded on \mathbb{R}^+). Consequently, Theorem 1 cannot be applied to (13). To fill this gap, we have derived the following theorem

Theorem 2. Let $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ be N independent realizations of a p -dimensional random vector \mathbf{x} distributed according to an MGGD of parameters m , β and \mathbf{M} . For $\beta \in]0, 1[$,

1. the solution of the fixed point equation (13) exists and is unique up to a scalar factor,
2. Equation (7) admits a unique solution due to the normalization,

3. the recursive algorithms associated with equations (7) and (13) converge whatever the initialization.

Proof. Due to the lack of space, we will give an insight of the proof outline. Complete arguments can be found in [17]. Let us rewrite (13) as $\mathbf{M} = f(\mathbf{M})$ and let us consider a particular primitive F of f . To prove (1), we show that F admits a supremum on an open-half line of the set of positive-definite matrices and equivalently, that f admits $\alpha \mathbf{M}_{FP}$ as the unique solution of norm α , where \mathbf{M}_{FP} is the unit norm solution of (13). (2) is a straightforward corollary of (1). Then, to prove (3), we consider the discrete dynamical system $\mathbf{M}_{k+1} = f(\mathbf{M}_k)$ and we show that for every initialization \mathbf{M}_0 , the resulting trajectories (i.e., sequences of matrices \mathbf{M}_k) converge to the same point up to a scale factor. Imposing $\text{tr}(\mathbf{M}_k) = p$ at each iteration concludes the proof. \square

It is important to note that the condition $\beta \in]0, 1[$ corresponds to most applications including the modeling of spatial or color dependencies in wavelet coefficients for texture images [12].

4. SIMULATIONS

This section presents some simulation results to evaluate the performance of the MLEs for the parameters of MGGDs. The data vectors $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ have been generated as N independent and identically distributed random vectors distributed according to an MGGD by using the stochastic representation (3). All simulations have been conducted using the matrix $\mathbf{M} = (M_{i,j})$ with

$$M_{i,j} = \rho^{|i-j|}. \quad (14)$$

In all experiments, the unknown parameters of the MGGD (\mathbf{M} , β and m) have been jointly estimated by using the algorithm provided in section 2.2.

4.1. MLE of the scatter matrix

Fig. 1 shows some convergence results associated with the scatter matrix MLE $\hat{\mathbf{M}}$ when the length of each vector \mathbf{x}_i is $p = 3$. In this simulation, we have considered $\beta = 0.2$ and $\rho = 0.8$. Convergence results are analyzed by evaluating the criterion $C(k)$ defined as

$$C(k) = \frac{\|\hat{\mathbf{M}}_{k+1} - \hat{\mathbf{M}}_k\|}{\|\hat{\mathbf{M}}_k\|} \quad (15)$$

where $\|\cdot\|$ is the Frobenius norm and $\hat{\mathbf{M}}_k$ is the MLE of \mathbf{M} at step k . Note that more appropriate matrix norms could also be used to account for the geometrical structure of the set of positive definite matrices (see [11] [18] for more details).

Fig. 1.(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 [11], “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. 1.(b) shows the evolution of criteria $C(k)$ for various numbers N of secondary data. It can be observed that the convergence speed increases with N (as expected).

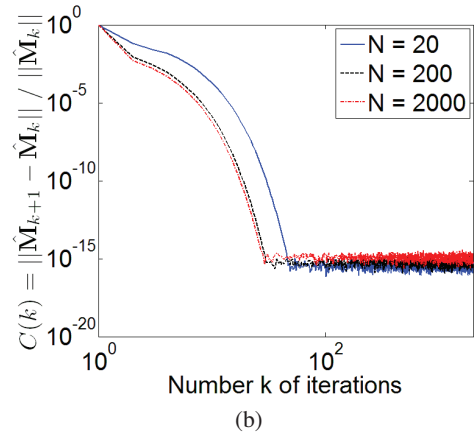
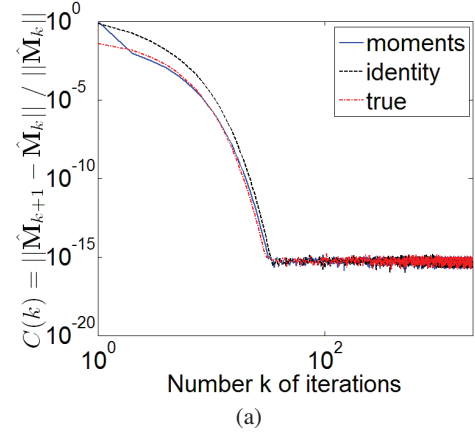


Fig. 1. Variations of $C(k)$ for $p = 3$, $\beta = 0.2$ and $\rho = 0.8$. (a) $C(k)$ versus number of iterations for different initializations ($N = 200$). (b) $C(k)$ versus number of iterations for various values of N .

4.2. Bias and consistency analysis

Fig. 2.(a) shows the estimate bias of $\hat{\mathbf{M}}$, for different values of β ($\beta \in \{0.2, 0.5, 0.8\}$), defined as $\|\bar{\mathbf{M}} - \mathbf{M}\|$ where the matrix $\bar{\mathbf{M}}$ is defined as the empirical mean of the estimated matrices

$$\bar{\mathbf{M}} = \frac{1}{I} \sum_{i=1}^I \hat{\mathbf{M}}(i). \quad (16)$$

All experiments presented in this paper have been obtained with $I = 1000$. Note that the bias criterion (16) was used in [19] for assessing the performance of matrix estimators. As observed in Fig. 2.(a), the bias converges very fast to a small value which is independent of β . Fig. 2.(b) presents some consistency results for the proposed estimator. Here, a plot of $\|\hat{\mathbf{M}} - \mathbf{M}\|$ is shown as a function of the number of samples N for different values of β . It can be noticed that this criterion tends to 0 when N approaches ∞ independently of β .

4.3. Parameter β

The Fisher information matrix has been recently derived for the parameters of MGGDs [11]. It has been shown that this matrix only depends on the number N of secondary data 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.

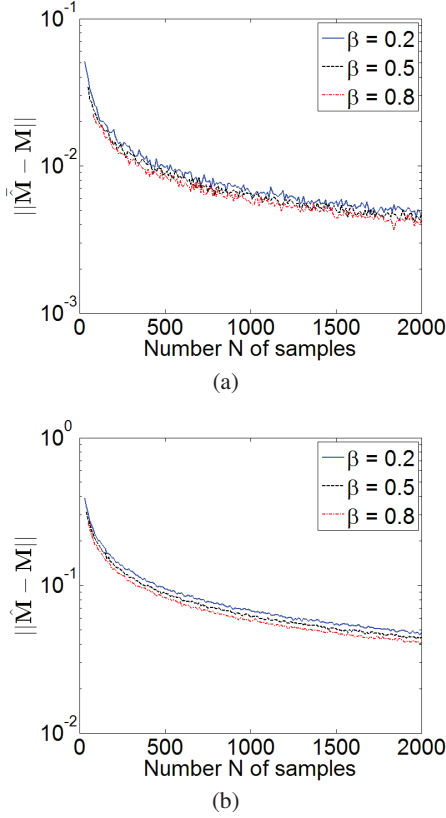


Fig. 2. (a) Estimated bias for different values of β , (b) estimated consistency for different values of β .

These CRLBS provide a reference (in terms of variance or mean square error) for any unbiased estimator of the MGD parameters. 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. 3 (versus the number of samples and the value of β). Fig. 3.(a) was obtained for $\beta = 0.2$, $\rho = 0.8$ and $p = 3$, while Fig. 3.(b) corresponds to $N = 10\,000$, $\rho = 0.8$ and $p = 3$. The ML method yields lower estimation variances compared to the moment-based approach, as expected. Moreover, the Cramér-Rao lower bound of $\hat{\beta}$ is very close to the variance of $\hat{\beta}$ in all cases illustrating the MLE efficiency.

5. 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. Simulation results have shown that this estimator is unbiased and consistent. The estimation of the shape parameter β was also investigated. The variance of its maximum likelihood estimator is very close to the corresponding Cramér-Rao bound illustrating its statistical efficiency. Further works include the use of multivariate generalized Gaussian distributions for remote sensing applications such as change detection or classification.

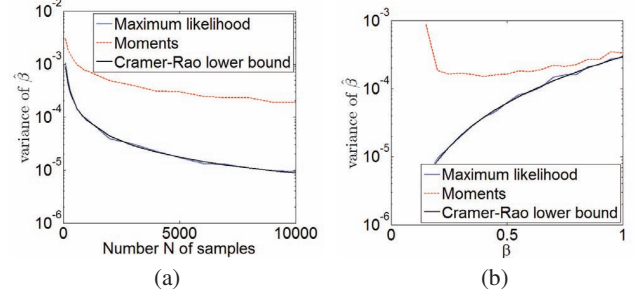


Fig. 3. Estimation performance for parameter β . (a) Variance of $\hat{\beta}$ versus number of samples N for $\beta = 0.2$, $\rho = 0.8$ and $p = 3$, (b) Variance of $\hat{\beta}$ versus β for $N = 10\,000$, $\rho = 0.8$ and $p = 3$.

6. REFERENCES

- [1] 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, 2010.
- [2] S. Sarra and A. Benazza-Benyahia, "Indexing of multichannel images in the wavelet transform domain," in *International Conference on Information and Communication Technologies: From Theory to Applications*, 2008, pp. 1–6.
- [3] Y. Stitou, N. Lasmar, and Y. Berthoumieu, "Copulas based multivariate gamma modeling for texture classification," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2009, pp. 1045–1048.
- [4] L. Boubchir and J. Fadili, "Multivariate statistical modeling of images with the curvelet transform," in *Proc. Int. Symp. Signal Process. and its Applications*, Aug. 2005, pp. 747–750.
- [5] K.-C. Chu, "Estimation and decision for linear systems with elliptical random processes," *IEEE Trans. Autom. Control*, vol. 18, no. 5, pp. 499 – 505, October 1973.
- [6] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 11, no. 3, pp. 368–385, 1981.
- [7] S. Kotz, *Statistical Distributions in Scientific Work, I.*, Dordrecht: Reidel, 1968, ch. Multivariate distributions at a cross road, pp. 247–270.
- [8] E. Gómez, M.A. Gomez-Villegas, and J.M. Marín, "A multivariate generalization of the power exponential family of distributions," *Communications in Statistics - Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.
- [9] N. Khelil-Cherif and A. Benazza-Benyahia, "Wavelet-based multivariate approach for multispectral image indexing," in *SPIE Conference on Wavelet Applications in Industrial Processing*, 2004.
- [10] D. Cho and T. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Processing: Image Communications*, vol. 20, no. 1, pp. 77–89, 2005.
- [11] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized gaussian models," *Journal of Mathematical Imaging and Vision*, 2011.
- [12] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized gaussian distributions with an application to multicomponent texture discrimination," *International Journal of Computer Vision*, 2011.
- [13] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley & Sons, 1994.
- [14] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [15] 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, 2008.
- [16] R.A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.
- [17] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "MGD covariance matrix estimation: existence, uniqueness, algorithm and statistical properties," Tech. Rep., 2012.
- [18] S.T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [19] 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. 2206–2216, 2008.