A CONVEX FORMULATION FOR THE ROBUST ESTIMATION OF MULTIVARIATE EXPONENTIAL POWER MODELS

Nora Ouzir, Jean-Christophe Pesquet,[*]

University of Paris-Saclay,
Inria, CentraleSupélec,
Centre de Vision Numérique

Frédéric Pascal[†]

Université Paris-Saclay,
CNRS, CentraleSupélec,
Laboratoire des signaux et systèmes

ABSTRACT

The multivariate power exponential (MEP) distribution can model a broad range of signals. In noisy scenarios, the robust estimation of the MEP parameters has been traditionally addressed by a fixed-point approach associated with a nonconvex optimization problem. Establishing convergence properties for this approach when the distribution mean is unknown is still an open problem. As an alternative, this paper presents a novel convex formulation for robustly estimating MEP parameters in the presence of multiplicative perturbations. The proposed approach is grounded on a re-parametrization of the original likelihood function in a way that ensures convexity. We also show that this property is preserved for several typical regularization functions. Compared with the robust Tyler's estimator, the proposed method shows a more accurate precision matrix estimation, with similar mean and covariance estimation performance.

***Index Terms*—** Covariance Matrix Estimation, Convex Optimization, Multivariate Exponential Power distribution.

1. INTRODUCTION

The mean and covariance matrix are critical parameters in diverse applications that require data processing (*e.g.,* dimension reduction, detection, clustering, or classification). Since these parameters are usually unknown in practice, estimating them has been a long-standing problem in statistics. When adopting a model-based approach, the standard Gaussian assumption is frequently used for its simplicity and tractability in performance analysis. However, it is not the most suitable probabilistic model in the presence of data heterogeneity or outliers, which are common limitations in many domains such as radar signal processing or image processing [1, 2]. In contrast, robust estimation theory [3] is a powerful framework capable of accounting for the various perturbations affecting the data. In the robust framework, observations are modelled as following elliptically-contoured (EC) distributions, also known as complex or real Elliptically Symmetric distributions (see [4] for a review). The EC model is very general, encompassing well-known multivariate distributions such as the Gaussian distribution, t-distribution, or K-distribution.

In this paper, we focus on a popular subclass of EC models called MEP distributions [5], also referred to as Multivariate Generalised Gaussian Distributions (MGGD) [6]. More precisely, we address the complex problem of jointly estimating mean, covariance matrix, and other involved perturbation parameters. By analysing a MEP model perturbed in scale, the main contribution of this work is to introduce a convex formulation of the robust estimation problem, ensuring both the existence and uniqueness of parameter estimators. The proposed model is flexible, and accounts for diverse data structures, such as sparsity, which is especially relevant when the number of observations is relatively small, and the data dimension is large.

The paper is organized as follows. Section 2 introduces the underlying statistical framework, including the proposed perturbed MEP model. Section 3 presents the main contribution of this paper, *i.e.,* the convex parameter estimation approach. Section 4 then reports simulation results highlighting the interest of this approach. Finally, concluding remarks and perspectives are provided in Section 5.

Notation: $\mathcal{S}_K$ denotes the space of symmetric real matrices of size $K \times K$, $\mathcal{S}_K^+$ is the cone of positive semi-definite matrices, and $\mathcal{S}_K^{++}$ the cone of positive definite matrices. $\|\cdot\|$ denotes the Euclidean norm, and $\|\cdot\|_\tau$ with $\tau \in [1, +\infty[$ denotes the element-wise ℓ^τ norm (the same notation will be used for a matrix or for a vector whatever the dimension). $\iota_\mathcal{D}$ denotes the indicator function of $\mathcal{D} \subset \mathcal{H}$, which is equal to 0 on this set and $+\infty$ out of it.

2. PROBLEM FORMULATION

The MEP distribution [5] belongs to a broad subclass of elliptically-contoured distributions and can model various uni-modal probability density functions with heavier or lighter tails than the Gaussian distribution. The zero-mean MEP distribution of dimension K, denoted as $\text{MEP}_K(\beta, \mathbf{C})$,

[*]Part of this work was supported by BRIDGEABLE ANR Chair in IA.

[†]The work of F. Pascal has been partially supported by DGA under grant ANR-17-ASTR-0015.

is described by the following probability density function (PDF):

$$\mathsf{p}_{\mathsf{x}}(\cdot) = C_{K,\beta} (\det \mathbf{C})^{-1/2} \exp\left(-\frac{1}{2}\left[(\cdot)^\top \mathbf{C}^{-1}(\cdot)\right]^{\beta/2}\right), \tag{2.1}$$

with the constant $C_{K,\beta} = \dfrac{K\,\Gamma(K/2)}{\pi^{K/2}\Gamma(1+K/\beta)2^{1+K/\beta}}$, where $\Gamma(\cdot)$ denotes the Gamma function. $\mathbf{C} \in \mathcal{S}_K^{++}$ is up to a multiplicative factor the associated covariance matrix.[1] As a particular case, when $\beta = 2$, (2.1) corresponds to the classical multivariate Gaussian distribution. When $\beta < 2$, distributions with heavier tails than the Gaussian one are obtained. In the following, it will be assumed that the exponent $\beta > 1$ is known.

2.1. Observation Model

Let $(\mathbf{x}_n)_{1\leqslant n\leqslant N}$ be realizations of *i.i.d.* random K-dimensional vectors generated according to a MEP distribution with exponent β and zero-mean. Let us also consider a scenario where noisy observations $(\mathbf{y}_n)_{1\leqslant n\leqslant N}$ have been corrupted by a multiplicative perturbation $\boldsymbol{\tau}$. The resulting observation model is

$$(\forall n \in \{1,\dots,N\} \quad \mathbf{y}_n = \tau_n \mathbf{x}_n + \boldsymbol{\mu} \tag{2.2}$$

where $\boldsymbol{\tau} = (\tau_n)_{1\leqslant n\leqslant N} \in \,]0,+\infty[^N$ and $\beta \in]1,+\infty[$.

Starting from (2.2), our goal is to jointly estimate the unknown parameters $\boldsymbol{\mu}$, $\mathbf{C}$ and $(\tau_n)_{1\leqslant n\leqslant N}$[2]. It is worth pointing out that N scalar parameters need to be estimated in addition to the mean and covariance matrix parameters. (Note also that a particular case of this problem has been addressed in [7] for $\boldsymbol{\mu} = \mathbf{0}$ and $\beta = 2$.) Since model (2.2) accounts for possible outliers in the observations (inducing large values of τ_n's), it can be seen as a generalization of the classical stochastic representation for EC distributions. The existing and proposed approaches are discussed in the following.

2.2. Nonconvex Log-Likelihood

Estimating the model parameters ($\mathbf{C}$, $\boldsymbol{\mu}$, and $\boldsymbol{\tau}$) can be achieved by minimizing the negative log-likelihood function arising from model (2.2) , *i.e.,*

$$\mathcal{L}(\mathbf{C},\boldsymbol{\mu},\boldsymbol{\tau}) = \frac{1}{2}\sum_{n=1}^{N} \frac{\left[(\mathbf{y}_n-\boldsymbol{\mu})^\top \mathbf{C}^{-1}(\mathbf{y}_n-\boldsymbol{\mu})\right]^{\beta/2}}{\tau_n^\beta} + \frac{N}{2}\log\det\mathbf{C} + K\sum_{n=1}^{N}\log\tau_n. \tag{2.3}$$

[1] Precisely, the covariance matrix of the MEP distribution is equal to $2^{2/\beta}\Gamma((K+2)/\beta)(K\,\Gamma(K/\beta))^{-1}\,\mathbf{C}$.

[2] We consider that the shape parameter β is known. Note that the estimation of β using a Newton-Raphson procedure has been studied in [6].

Note that this function is nonconvex. Its minimization has been intensively investigated [8, 6]; the standard approach first aims to minimise with respect to $\boldsymbol{\tau}$, and then plug the optimal value into the negative log-likelihood $\mathcal{L}(\mathbf{C},\boldsymbol{\mu},\widehat{\boldsymbol{\tau}}(\mathbf{C},\boldsymbol{\mu}))$. More precisely, the first step yields the optimal value

$$\widehat{\boldsymbol{\tau}}(\mathbf{C},\boldsymbol{\mu}) = \left(\left(\frac{\beta[(\mathbf{y}_n-\boldsymbol{\mu})^\top \mathbf{C}^{-1}(\mathbf{y}_n-\boldsymbol{\mu})]^{\beta/2}}{2K}\right)^{\frac{1}{\beta}}\right)_{1\leqslant n\leqslant N}. \tag{2.4}$$

By assuming that none of the vectors $(\mathbf{y}_n)_{1\leqslant n\leqslant N}$ is equal to $\boldsymbol{\mu}$ (which is true with probability one for continuous random vectors), plugging (2.4) in (2.3) leads to

$$\mathcal{L}(\mathbf{C},\boldsymbol{\mu},\widehat{\boldsymbol{\tau}}(\mathbf{C},\boldsymbol{\mu})) = \frac{K}{2}\sum_{n=1}^{N}\log\left[(\mathbf{y}_n-\boldsymbol{\mu})^\top\mathbf{C}^{-1}(\mathbf{y}_n-\boldsymbol{\mu})\right] + \frac{N}{2}\log\det\mathbf{C} + \frac{KN}{\beta}\left[1-\log\left(\frac{2K}{\beta}\right)\right]. \tag{2.5}$$

For a given value of $\boldsymbol{\mu}$, minimizing function (2.5) either with respect to $\mathbf{C}$ or $\mathbf{C}^{-1}$ on $\mathcal{S}_K^{++}$ is a nonconvex problem. In the Gaussian case (with unknown variances τ_n), the optimal covariance matrix is Tyler's estimate and can be computed by a fixed point equation [9, 7]. Convergence properties of the related iterative scheme have been proved [7], but the generalization of these properties to the case when $\boldsymbol{\mu}$ is unknown is still an open issue.

3. CONVEX PARAMETER ESTIMATION

The objective of this work is to propose a convex alternative to (2.3). The key idea of the proposed approach is to transform the cost function by using appropriate variable changes as detailed in the following.

3.1. New Cost Function

Let us re-parameterize function $\mathcal{L}$ by setting

$$\mathbf{C}^{-1} = \mathbf{Q}^2 \tag{3.1}$$

$$\mathbf{m} = \mathbf{Q}\boldsymbol{\mu} \tag{3.2}$$

$$\boldsymbol{\theta} = (\theta_n)_{1\leqslant n\leqslant N} = (\tau_n^{\beta/(\beta-1)})_{1\leqslant n\leqslant N}, \tag{3.3}$$

where $\mathbf{Q} \in \mathcal{S}_K^{++}$. Then $\mathcal{L}(\mathbf{C},\boldsymbol{\mu},\boldsymbol{\tau}) = \widetilde{\mathcal{L}}(\mathbf{Q},\mathbf{m},\boldsymbol{\theta})$ where

$$\widetilde{\mathcal{L}}(\mathbf{Q},\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{n=1}^{N}\frac{\|\mathbf{Q}\mathbf{y}_n-\mathbf{m}\|^\beta}{\theta_n^{\beta-1}} - N\log\det\mathbf{Q} + K(1-1/\beta)\sum_{n=1}^{N}\log\theta_n. \tag{3.4}$$

It can be shown that if we exclude the last term, the obtained function is convex with respect to $(\mathbf{Q},\mathbf{m},\boldsymbol{\theta})$. In the following, we show that a careful choice of the regularization function will allow us to recover a fully convex formulation.

3.2. Including Prior Information

Prior information may be available about the sought variables and used to improve their estimation. For example, in graph processing applications it is known that the precision matrix $\mathbf{C}^{-1}$ is sparse [10]. Prior information on the nature of the perturbation $\boldsymbol{\tau}$ may also be accessible (*e.g.,* bounds), or one may seek to restrict the mean to a specific set (*e.g.,* known values for a restricted subset of components). To account for these prior information, we also introduce a regularized cost function

$$f(\mathbf{Q},\mathbf{m},\boldsymbol{\theta}) = \begin{cases} \widetilde{\mathcal{L}}(\mathbf{Q},\mathbf{m},\boldsymbol{\theta},\mathbf{d})+g_{\mathsf{Q}}(\mathbf{Q})+g_m(\mathbf{m})+g_\vartheta(\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta}\in\,]0,+\infty[^N \\ & \text{and } \mathbf{Q}\in\mathcal{S}_K^{++} \\ +\infty & \text{otherwise,} \end{cases} \tag{3.5}$$

where $g_{\mathsf{Q}}\colon \mathcal{S}_K \to\,]-\infty,+\infty]$, $g_{\mathsf{m}}\colon \mathbb{R}^K \to\,]-\infty,+\infty]$, and $g_\vartheta\colon \mathbb{R}^N \to\,]-\infty,+\infty]$ are regularization functions on $\mathbf{Q}$, $\mathbf{m}$, and $\boldsymbol{\theta}$, respectively. We allow these functions to take infinity values in order to model potential hard constraints on the variables. For example, if one seeks to restrict the vector $\mathbf{m}$ to some set $\mathcal{D}\subset\mathbb{R}^K$ (*e.g.*, some hypercube or some ball) a suitable choice for g_{m} is the indicator function $\iota_{\mathcal{D}}$ of $\mathcal{D}$.

The cost function f is suitable for different robust estimation problems depending on the choice of the regularizations. In this paper, we choose $g_{\mathsf{m}} = \iota_{\{\mathbf{0}\}}$ and focus on studying the regularizations of the variables $\mathbf{Q}$ and $\boldsymbol{\theta}$, for which prior information is more frequently available in practice. More precisely, we consider a typical example where the precision matrix $\mathbf{C}^{-1}$ is sparse. Note that in the particular case when $\beta = 2$, $g_{\mathsf{m}} = \iota_{\{\mathbf{0}\}}$ and $g_\vartheta = \iota_{\{\mathbf{1}\}}$ if we set $g_{\mathsf{Q}}\colon \mathbf{Q} \mapsto \lambda\|\mathbf{Q}^2\|_1$ with $\lambda\in\,]0,+\infty[$, minimizing $\widetilde{\mathcal{L}}$ is equivalent to the classical Graphical LASSO (GLASSO) problem [10]. However, with the variable changes we have performed, choosing $g_{\mathsf{Q}} = \lambda\|\cdot\|_1$ is a more natural way of imposing sparsity on the precision matrix $\mathbf{C}^{-1}$. We will use the latter function in the experiments in Section 4.

Let us now turn to the choice of the regularization on $\boldsymbol{\theta}$. The main idea is to choose a function that will ensure the convexity of the global cost function f. One regularization function allowing us to reach this goal is

$$(\forall\boldsymbol{\theta}\in\mathbb{R}^N)\quad g_\vartheta(\boldsymbol{\theta}) = \frac{1}{\eta^\alpha}\|\boldsymbol{\theta}\|_\alpha^\alpha + \kappa\sum_{n=1}^N \psi(\theta_n) \tag{3.6}$$

with

$$\psi\colon\mathbb{R}\to\,]-\infty,+\infty] : \xi\mapsto \begin{cases} -\log\xi & \text{if } \xi>0 \\ +\infty & \text{otherwise,}\end{cases} \tag{3.7}$$

and $\alpha\in[1,+\infty[$, $\eta\in\,]0,+\infty[$, and $\kappa\in]K(1-1/\beta),+\infty[$. It is worth noticing that (3.6) is the expression of the potential of a generalized Gamma distribution with scale parameter η, shape parameter $(\kappa+1)$, and exponent parameter α. The choice of the latter parameters will not be further investigated in this work. Note that, from a Bayesian perspective, the regularization functions introduced in (3.5) (and those used in this work specifically) can be viewed as the potentials associated with (possibly improper) prior distributions proportional to $\exp(-g_{\mathsf{Q}}(\cdot))$, $\exp(-g_{\mathsf{m}}(\cdot))$, and $\exp(-g_\vartheta(\cdot))$, respectively. Minimizing (3.5) then amounts to computing Maximum A Posteriori (MAP) estimates of $\mathbf{Q}$, $\mathbf{m}$, and $\boldsymbol{\theta}$.

With the choice of the regularization suggested above, the resulting cost function is convex and, under mild conditions, it has a unique minimizer. Due to the lack of space, a formal proof will be detailed in a forthcoming paper. Minimizing this function can be carried out by proximal primal-dual algorithms for which the proximity operators of all components of the function are available in closed form [11, 12, 13]. Convergence is guaranteed [14, 15] and reached in a few iterations in our experiments (see Section 4).

4. EXPERIMENTS

This section presents different examples of simulation results with varying N values. We also consider two simulation scenarios with dense and sparse precision matrices. The proposed method is compared to the robust Tyler's estimates and the empirical statistics in both the observed (*i.e.,* perturbed) and ideal non-perturbed ($(\forall n)$ $\tau_n = 1$) cases. Note that a fixed-point update is used for the estimation of the mean in Tyler's method. Although this approach has no convergence guarantees, it is frequently used in practice [16]. The comparison uses the estimated consistency with respect to the true parameters. More precisely, the estimated consistency of $\hat{\boldsymbol{A}}$ is defined as $\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_{\mathrm{F}}$, where $\boldsymbol{A}$ is the true parameter. Finally, note that all experiments are performed by averaging 1000 Monte Carlo runs.

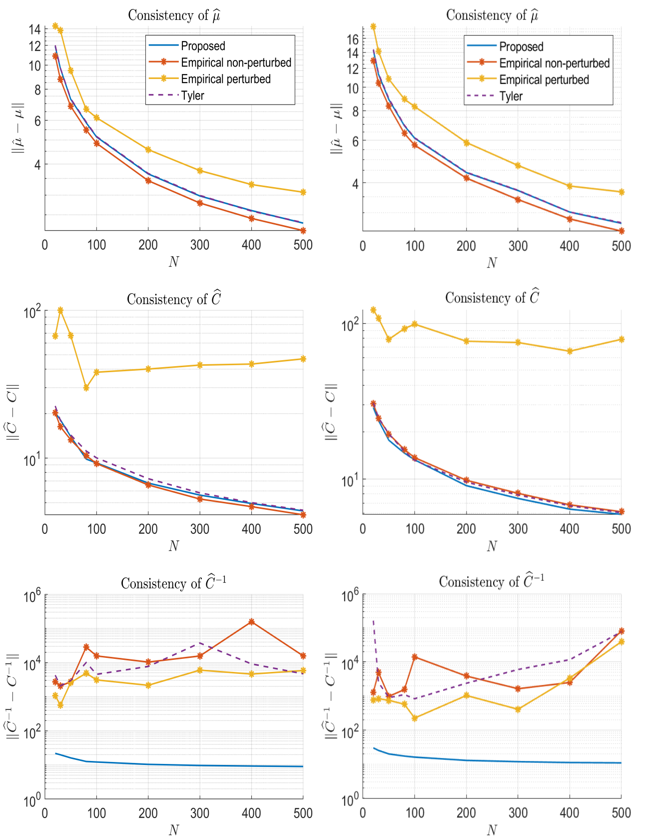
4.1. Simulation Parameters

In all experiments, $\beta = 1.5$, the correlation coefficient used to generate the precision matrices is $\rho = 0.5$, and the proportion of corrupted data is set to 10%. The sparse regularization parameter is estimated automatically by providing the desired sparsity level (*i.e.,* the input sparsity level of the generated matrices). The parameters of the gamma prior are set to $\kappa = 1.1\,K(1-1/\beta)$ and $\alpha = 1$ for all experiments.

4.2. Experiments with a Dense Precision Matrix

The first scenario considers *i.i.d.* $K = 3$ dimensional data vectors $(\mathbf{x}_n)_{1\leqslant n\leqslant N}$ distributed according to a $\mathrm{MEP}_K(\beta,\mathbf{C}_1)$ distribution with $\mathbf{C}_1$ defined as

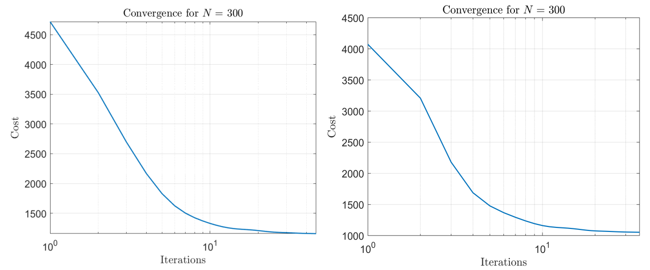
$$(\forall(i,j)\in\{1,\dots,K\}^2)\quad \mathbf{C}_1^{-1}(i,j) = \rho^{|i-j|}. \tag{4.1}$$



(a) (b)

Fig. 1. Estimated consistency of mean $\hat{\mu}$, covariance $\hat{C}$ and precision matrices $\hat{C^{-1}}$ for (a) the first and (b) second experiments. The Tyler's estimates and empirical statistics are shown for comparison.

As mentioned above, 10% of the data is corrupted according to Model (2.2). Fig. 1-(a) shows the estimated consistency of the obtained mean, covariance, and precision matrices compared to the empirical statistics and robust Tyler's estimates. One can see that for all methods the consistency of the mean decreases with increasing values of N. Two other main observations can be made: first, the estimated mean and covariance are very close for the proposed and Tyler's methods; secondly, the results for the precision matrix show a significant improvement of the proposed method over all other approaches. It is worth noting that, unlike Tyler's approach, the proposed method guarantees convergence in the unknown mean case. The much better results for the precision matrix can be explained by the fact that the proposed method estimates this quantity directly rather than performing a matrix inversion of the estimated covariance matrices (which may result in numerical issues). The convergence of the total cost is shown in Fig. 2-(a). This figure shows the average evolution of the cost function over all Monte Carlo experiments.



(a) (b)

Fig. 2. Evolution of the total cost *vs* iterations with $N = 300$ for (a) the first and (b) second experiments. Note that the plots show costs averaged between all Monte Carlo runs at each iteration.

4.3. Experiments with a Sparse Precision Matrix

In a second experiment, we consider a sparse set-up. The data samples are drawn from a $\mathrm{MEP}_K(\beta, \mathbf{C}_2)$ distribution with a sparse precision matrix $\mathbf{C}_2^{-1}$ modelled by an auto-regressive (AR) process of order 3. In this case $\mathbf{C}_2^{-1}$ is tri-diagonal and defined as

$$\mathbf{C}_2^{-1}(i,j) = \begin{cases} \rho^{|i-j|}, & \text{for } |i-j| = 0 \text{ and } i \in \{1,\dots,K\} \\ & \text{for } |i-j| = 1 \text{ and } i \in \{2,\dots,K-1\} \\ & \text{for } |i-j| = 2 \text{ and } i \in \{3,\dots,K-2\} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

The consistencies of the estimated mean, covariance and precision matrices are shown in Fig. 1-(b). One can see that the results are similar to the first experiment, with an improvement in the estimation of the covariance matrix for the proposed method. The evolution of the total cost is also shown in Fig. 2-(b).

5. CONCLUSION

This paper has introduced a novel convex formulation for the robust estimation of the MEP model parameters. The proposed method uses a re-parameterization and a judicious regularization leading to a convex estimation of the mean, covariance, and multiplicative perturbation parameters. The resulting cost function can be combined with useful regularizations on the sought parameters. It can be minimized by a standard proximal primal-dual algorithm with guaranteed convergence. The experiments show improvements for precision matrix estimation compared to non-robust empirical statistics and robust Tyler's estimators, which is interesting in many real-world applications. Compared to Tyler's estimates, the proposed approach seems appealing because of the convergence guarantees it provides for the mean estimate. In future work, different regularizations will be investigated in a high-dimensional context where $N \ll K$.

6. REFERENCES

[1] Gordana Drašković, Arnaud Breloy, and Frédéric Pascal, "On the performance of robust plug-in detectors using m-estimators," *Signal Processing*, vol. 167, pp. 107282, February 2020.

[2] Gordana Drašković, Frédéric Pascal, and Florence Tupin, "M-NL: Robust NL-Means Approach for PolSAR Images Denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 997–1001, June 2019.

[3] Ricardo A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Stat.*, vol. 5, no. 1, pp. 51–67, 1976.

[4] Esa Ollila, David E. Tyler, Visa Koivunen, and H. Vincent Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[5] E. Gómez, M.A. Gomez-Viilegas, 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.

[6] Frédéric Pascal, Lionel Bombrun, Jean-Yves Tourneret, and Yannick Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5960–5971, 2013.

[7] 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," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[8] Yilun Chen, Ami Wiesel, and Alfred O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, 2011.

[9] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Stat.*, vol. 15, no. 1, pp. 234–251, 1987.

[10] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, July 2008.

[11] Heinz Bauschke and Patrick Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Space*, CMS Books in Mathematics. Springer International Publishing, 2017.

[12] Patrick L. Combettes and Christian L. Müller, "Perspective functions: Proximal calculus and applications in high-dimensional statistics," *Journal of Mathematical Analysis and Applications*, vol. 457, no. 2, pp. 1283–1306, 2018, Special Issue on Convex Analysis and Optimization: New Trends in Theory and Applications.

[13] Patrick L. Combettes and Jean-Christophe Pesquet, *Proximal Splitting Methods in Signal Processing*, pp. 185–212, Springer New York, New York, NY, 2011.

[14] Nikos Komodakis and Jean-Christophe Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[15] Patrick L. Combettes and Jean-Christophe Pesquet, "Fixed point strategies in data science," *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[16] J. Frontera-Pons, M. Veganzones, F. Pascal, and J-P. Ovarlez, "Hyperspectral Anomaly Detectors using Robust Estimators," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS)*, vol. 9, no. 2, pp. 720–731, february 2016.