# An Efficient Multivariate Generalized Gaussian Distribution Estimator: Application to IVA

Zois Boukouvalas*, Geng-Shen Fu† and Tülay Adalı†
Email:{zb1, fugengs1, adali}@umbc.edu
*Department of Mathematics and Statistics
†Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250

***Abstract*—Due to its simple parametric form, multivariate generalized Gaussian distribution (MGGD) has been widely used for modeling vector-valued signals. Therefore, efficient estimation of its parameters is of significant interest for a number of applications. Independent vector analysis (IVA) is a generalization of independent component analysis (ICA) that makes full use of the statistical dependence across multiple datasets to achieve source separation, and can take both second and higher-order statistics into account. MGGD provides an effective model for IVA as well as for modeling the latent multivariate variables–sources–and the performance of the IVA algorithm highly depends on the estimation of the source parameters. In this paper, we propose an efficient estimation technique based on the Fisher scoring (FS) and demonstrate its successful application to IVA. We quantify the performance of MGGD parameter estimation using FS and further verify the effectiveness of the new IVA algorithm using simulations.**

***Index Terms*—** Multivariate generalized Gaussian distribution, Independent vector analysis, Fisher scoring.

## I. Introduction

Even though the MGGD has a simple parametric form, it provides a model sufficiently flexible for most applications, and hence it has been an attractive solution for many applications in signal processing such as in video coding [1], denoising [2] and biomedical signal processing [3]. The method of moments (MoM) and maximum likelihood (ML) estimation techniques have been proposed [4], [5] for estimating the scatter matrix and the shape parameter of a zero-mean MGGD. In [6], authors prove that the scatter matrix ML estimator exists and is unique for any value of the shape parameter that belongs to the interval $(0, 1)$, i.e., when the marginals have distributions ranging from very peaky and heavy tailed to Gaussian (for shape parameter 1). In addition, a fixed point algorithm (ML-FP) has been introduced for estimating the parameters of an MGGD. Simulation results reveal the unbiasedness and consistency of the ML estimators of the scatter matrix and the shape parameter of the distribution. However, since many applications require values of the shape parameter that are greater than one, i.e., flatter distributions as well, in our work we propose an effective method that preserves the consistency properties of the ML estimators when parameters are estimated jointly and for any value of the shape parameter.

IVA is an extension of ICA to multiple datasets, and achieves better performance than performing ICA separately on each dataset by exploiting the dependence across datasets. Among the solutions to IVA, IVA-Laplacian (IVA-L) [7], [8] does not exploit the linear dependencies expressed in the second-order statistics but only takes higher-order statistics into account and assumes a Laplacian distribution. Conversely, IVA-Gaussian (IVA-G) [9], [10] exploits linear dependencies but does not take higher-order statistics into account. Finally, IVA-GGD [11] is a more general IVA implementation where both second and higher order statistics are taken into account. IVA-GGD uses a fixed set of shape parameter values and only estimates the scatter matrix. In addition, IVA-GGD uses the MoM for the estimation of the scatter matrix, which does not possess the large sample optimality property. In our work, we use a ML estimation method to precisely estimate all the parameters of the MGGD sources simultaneously. Therefore, the corresponding IVA score and cost functions can be readily calculated, providing efficient performance for IVA.

In this paper, we introduce a maximum likelihood estimator based on FS (ML-FS) that estimates both the shape parameter and the scatter matrix. Using ML-FS we derive a new IVA algorithm, IVA with adaptive MGGD (IVA-A-GGD), that estimates shape parameter and scatter matrix jointly and takes into account both second and higher-order statistics. We show that IVA-A-GGD provides very desirable performance by comparing its performance with those of competing algorithms.

## II. Background

### A. MGGD

The probability density function of an MGGD is given by [12]

$$p(\mathbf{y};\boldsymbol{\Sigma},\beta,m)=\frac{\Gamma\left(\frac{K}{2}\right)}{\pi^{\frac{K}{2}}\Gamma\left(\frac{K}{2\beta}\right)2^{\frac{K}{2\beta}}}\frac{\beta}{m^{\frac{K}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\times\exp\left[-\frac{1}{2m^{\beta}}\left(\mathbf{y}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{y}\right)^{\beta}\right],$$

where $K$ is the dimension of the probability space, $\mathbf{y}\in\mathbb{R}^K$ is a random vector, $m$ is the scale parameter, $\beta>0$ is the shape parameter that controls the peakedness and the spread of the distribution and $\boldsymbol{\Sigma}$ is a $K\times K$ symmetric positive scatter matrix. If $\beta=1$, the MGGD is equivalent to the multivariate Gaussian and the matrix $\boldsymbol{\Sigma}$ becomes the covariance matrix. If $\beta<1$ the distribution of the marginals is more peaky and has heavier tails, while $\beta>1$ is less peaky and has lighter tails.

978-1-4799-8428-2/15/$31.00 ©2015 IEEE

### *B. IVA*

Let each dataset $\mathbf{x}^{[k]}$, $k = 1, ..., K$ be a linear mixture of $N$ statistically independent sources

$$\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, ..., K,$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}$, $k = 1, ..., K$ are invertible mixing matrices and $\mathbf{s}^{[k]} = [s_1^{[k]}, ..., s_N^{[k]}]^\top$ is the vector of latent sources for the $k$th dataset. The $n$th source component vector (SCV) $\mathbf{s}_n = [s_n^{[1]}, ..., s_n^{[k]}]^\top$, can be defined by concatenating the $n$th source from each of the $K$ data sets. The goal in IVA is to estimate $K$ demixing matrices to yield source estimates $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, such that each SCV is maximally independent of all other SCVs. This is achieved by minimizing the mutual information cost function

$$\mathcal{I}_{\text{IVA}} \;=\; \sum_{n=1}^{N} \mathcal{H}[\mathbf{y}_n] \;-\; \sum_{k=1}^{K} \log \left|\det\left(\mathbf{W}^{[k]}\right)\right| \;-\; C, \quad (1)$$

where $\mathcal{H}[\mathbf{y}_n]$ denotes the entropy of the $n$th SCV and $C$ is the constant term $\mathcal{H}[\mathbf{x}^{[1]}, ..., \mathbf{x}^{[K]}]$. Different algorithms for minimizing the IVA cost function have been described in [13], [9]. The gradient of the cost function in (1) is given by

$$\frac{\partial \mathcal{I}_{\text{IVA}}}{\partial \mathbf{W}^{[k]}} = -\sum_{n=1}^{N} E\left\{\frac{\partial \log p(\mathbf{y}_n)}{\partial y_n^{[k]}} \frac{\partial y_n^{[k]}}{\partial \mathbf{W}^{[k]}}\right\} - \left(\mathbf{W}^{[k]}\right)^{-\top}. \quad (2)$$

As observed in (2), the probability density function or its approximation for each estimated SCV plays an important role on the development of IVA algorithms.

## III. Approaches to MGGD parameter estimation

Let $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$ be a random sample of $T$ observation vectors of dimension $K$, drawn from a zero mean MGGD with parameters $m$, $\beta$, and $\mathbf{\Sigma}$. The ML estimators of $m$, $\beta$, and $\mathbf{\Sigma}$ are given by [6]

$$\hat{\mathbf{\Sigma}} = \sum_{i=1}^{T} \frac{p}{u_i + u_i^{1-\beta} \sum_{i\neq j} u_j^{\beta}} \mathbf{y}_i \mathbf{y}_i^\top, \quad (3)$$

$$\hat{m} = \left[\frac{1}{T}\sum_{i=1}^{T} (u_i)^{\beta}\right]^{\frac{1}{\beta}}, \quad (4)$$

and

$$\frac{KT}{2\sum_{i=1}^{T} u_i^{\beta}} \sum_{i=1}^{T}\left[u_i^{\beta}\ln(u_i)\right] - \frac{KT}{2\beta}\left[\Psi\left(\frac{K}{2\beta}\right) + \ln 2\right] - T - \frac{KT}{2\beta}\ln\left(\frac{\beta}{KT}\sum_{i=1}^{T} u_i^{\beta}\right) = 0, \quad (5)$$

where $u_i = \mathbf{y}_i^\top \mathbf{\Sigma}^{-1} \mathbf{y}_i$ and $\Psi(\cdot)$ is the digamma function. There is no closed form solution for the ML estimation of these parameters. Hence, we use an iterative scheme to simultaneously estimate $\mathbf{\Sigma}$ and $\beta$ using (3) and (5) respectively and introduce a stable descent algorithm that is based on the Fisher information matrix to estimate $\mathbf{\Sigma}$.

The likelihood function of $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$ where $m$ has been replaced by its estimate in (4) is given by

$$\mathcal{L}(\mathbf{\Sigma};\Theta) = \prod_{i=1}^{T} p(\mathbf{y}_i;\Theta) = \left[\frac{\beta\Gamma\left(\frac{K}{2}\right)}{\pi^{\frac{K}{2}}\Gamma\left(\frac{K}{2\beta}\right) 2^{\frac{K}{2\beta}}}\right]^{T} \left(\frac{KT}{\beta}\right)^{\frac{KT}{2\beta}} \exp\left(-\frac{KT}{2\beta}\right) \times \left[\frac{1}{|\mathbf{\Sigma}|}\left(\sum_{i=1}^{T} u_i^{\beta}\right)^{-\frac{K}{\beta}}\right]^{\frac{T}{2}}, \quad (6)$$

where $\Theta$ denotes the parameter space that contains the entries of the scatter matrix $\mathbf{\Sigma}$, $m$ and $\beta$. By fixing $\beta$, the likelihood function depends only on the entries of $\mathbf{\Sigma}$. Defining the gradient of the likelihood function similar to [6], we introduce the functional

$$F : \mathcal{S}_{++}^{K} \to \mathbb{R}^{+} \setminus \{0\}$$
$$\mathbf{\Sigma} \;\mapsto |\mathbf{\Sigma}|^{-1}\left(\sum_{i=1}^{T} u_i^{\beta}\right)^{-\frac{K}{\beta}},$$

where $\mathcal{S}_{++}^{K}$ is the space of all real $K\times K$ symmetric and positive definite matrices. By omitting the constant term from (6), the gradient of the likelihood can then be written as

$$\nabla\mathcal{L}(\mathbf{\Sigma};\Theta) = [F(\mathbf{\Sigma})]^{\frac{T-2}{2}}\; \nabla F(\mathbf{\Sigma}),$$

and the gradient of $F$ at a point $\mathbf{\Sigma} \in \mathcal{S}_{++}^{K}$ is given by [6]

$$\nabla F(\mathbf{\Sigma}) = F(\mathbf{\Sigma})\mathbf{\Sigma}^{-1}[f(\mathbf{\Sigma}) - \mathbf{\Sigma}]\mathbf{\Sigma}^{-1},$$

where

$$f : \mathcal{S}_{++}^{K} \to \mathcal{S}_{++}^{K}$$
$$\mathbf{\Sigma} \;\mapsto \sum_{i=1}^{T} \frac{K}{u_i + u_i^{1-\beta}\sum_{i\neq j} u_j^{\beta}} \mathbf{y}_i \mathbf{y}_i^\top.$$

To numerically maximize the likelihood function we use a variation of the Newton-Raphson optimization algorithm called the ML-FS algorithm. The main difference between the classical Newton-Raphson algorithm and ML-FS is that the negative inverse Hessian has been replaced by the Fisher information matrix.

To calculate the entries of the Fisher information matrix, we first define the manifold $\mathcal{M}$ of a zero-mean MGGD as well as an associated metric. The MGGD manifold is parameterized by $\beta$ and the matrix $\mathbf{\Sigma}$. In our particular case since $\beta$ is fixed, $\mathcal{M}$ is parameterized only by the entries of $\mathbf{\Sigma} \in \mathcal{S}_{++}^{K}$, so $\mathcal{M}$ is isomorphic to $\mathbb{R}^n$ where $n = \frac{K(K+1)}{2}$. Each point that lies on $\mathcal{M}$ is a probability density function. To measure the distance between two pdfs we need to calculate the length of the curve that connects those two points and has minimum length. This curve is called a geodesic path and is determined through the elements of the Fisher information matrix. Thus, if $\Theta = (\theta_1, \theta_2, ..., \theta_n)$ denotes the parameter space of $\mathcal{M}$, the Fisher metric is defined by the matrix elements

$$\mathbf{G}_{ij}(\theta) = -E\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\mathbf{y};\Theta)\right\}, \quad i,j = 1,...n,$$

where $\mathbf{G}$ is an $n \times n$ matrix. Trying to define geodesic paths within an MGGD manifold, [5] proposed a simpler way for the elements of the metric defined on a $K$-dimensional sub-manifold. Thus the entries of the Fisher information matrix are defined by

$$\mathbf{G}_{ii}(\beta) = \frac{1}{4}\left(\frac{3K+6\beta}{K+2} - 1\right), \tag{7}$$

and

$$\mathbf{G}_{ij}(\beta) = \frac{1}{4}\left(\frac{K+2\beta}{K+2} - 1\right), \ i \neq j, \tag{8}$$

for $i,j = 1,...,K$. From (7) and (8), we see that Fisher information matrix depends only on the fixed value of $\beta$ and the dimension $K$. This results in the reduction of the computational cost since the update of the inverse of the Fisher information matrix is not required. Fisher scoring iteration is defined as

$$\mathbf{\Sigma}^{(k+1)} = \mathbf{\Sigma}^{(k)} + \mathbf{G}^{-1}\nabla\mathcal{L}(\mathbf{\Sigma};\Theta).$$

MoM provides an effective and efficient initialization for the ML-FS algorithm with only a few steps of the algorithm being sufficient to obtain good estimates.

## IV. IVA-A-GGD Algorithm

Using the ML-FS algorithm as the main parameter estimation technique, we propose a new IVA algorithm with adaptive MGGD prior to estimate the shape parameter and scatter matrix jointly and exploit both second and higher-order statistics. Instead of minimizing (1) with respect to $\mathbf{W}^{[\mathbf{k}]}$ we use a decoupling procedure [14], [15] to minimize (1) with respect to each of the row vectors $\mathbf{w}_i^{[k]}$, $i = 1, ..., N$. Therefore by following [11], we can rewrite the cost function as

$$\mathcal{I}_{\text{IVA}} = \mathcal{H}[\mathbf{y}_n] - \log\left|\left(\mathbf{h}_n^{[k]}\right)^\top \mathbf{w}_n^{[k]}\right| - C_n^{[k]},$$

where $\mathbf{h}_n^{[k]}$ is the unit length vector, with the property that is perpendicular to all row vectors of the matrix $\mathbf{W}^{[k]}$ except of the vector $\mathbf{w}_n^{[k]}$. The differential entropy for an MGGD $\mathbf{y}_n$ is given by

$$\mathcal{H}[\mathbf{y}_n] = -E\left\{\log p(\mathbf{y}_n)\right\} = -\log\left(\frac{\beta\Gamma\left(\frac{K}{2}\right)}{\pi^{\frac{K}{2}}\Gamma\left(\frac{K}{2\beta}\right)2^{\frac{K}{2\beta}}}\right) + \frac{K}{2}\log m + \frac{1}{2}\log|\mathbf{\Sigma}| + \frac{1}{2m^\beta}E\left\{\left(\mathbf{y}_n^\top\mathbf{\Sigma}^{-1}\mathbf{y}_n\right)^\beta\right\},$$

and the gradient update rule by

$$\frac{\partial\mathcal{I}_{\text{IVA}}}{\partial\mathbf{w}_n^{[k]}} = E\left\{\phi_n^{[k]}(\mathbf{y}_n)\mathbf{x}^{[k]}\right\} - \frac{\mathbf{h}_n^{[k]}}{\left(\mathbf{h}_n^{[k]}\right)^\top\mathbf{w}_n^{[k]}},$$

where $\phi_n^{[k]}$ is the $k$th element of the multivariate score function

$$\phi(\mathbf{y}_n) = \frac{\beta}{m^\beta}\left(\mathbf{y}_n^\top\mathbf{\Sigma}^{-1}\mathbf{y}_n\right)^{\beta-1}\mathbf{\Sigma}^{-1}\mathbf{y}_n.$$

## V. Experimental Results

### A. MGGD estimators

For the first set of our experiments, we quantify the performance of the ML-FS, ML-FP, and MoM algorithms. Data were generated according to [16], [6] with $\mathbf{\Sigma}$ defined by

$$\mathbf{\Sigma}(i,j) = \sigma^{|i-j|}, \quad i,j = \{0,1,..,K-1\}$$

where $\sigma$ belongs to the interval $[0,1)$ and controls the correlation between the entries of the data. All results are averaged over 100 runs. For this experiment, we use $K = 3$, $T = 10000$ and $\sigma$ has ben uniformly selected from the range $(0.4, 0.6)$.

Fig.1 shows the difference between the estimated and the original scatter matrix as a function of different shape parameters values. We capture the difference between the original and the estimated scatter matrix using the Frobenius norm. From Fig.1 we observe that for values of $\beta < 1$, ML-FP and ML-FS provide better results than the MoM, while for values of $\beta > 3$ ML-FS performs the best among these three algorithms.

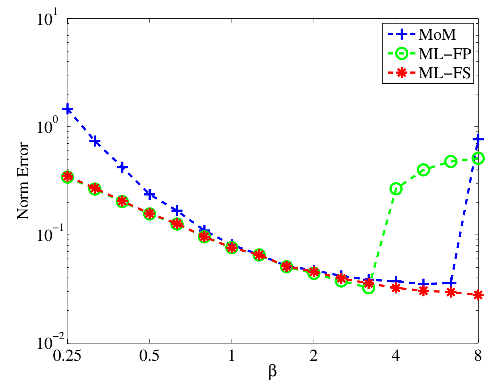


Fig. 1. Scatter matrix estimation performance for different values of shape parameter, for $T = 10000$, $\sigma \in (0.4, 0.6)$.

Fig.2 shows the comparison between the variances of the $\beta$ estimators generated from the MoM, ML-FP and ML-FS estimators as well as the Cramer-Rao lower bound (CRLB), as a function of different values of $\beta$. CRLB can be obtained by inverting the Fisher information matrix derived by [5]. From Fig.2, we observe that the performance of the ML-FS is very close to the CRLB, illustrating the MLE efficiency. On the other hand, the other two methods have issues when $\beta$ moves away from one.

### B. IVA results

To show the effectiveness of the IVA-A-GGD algorithm, we compare its performance with a number of widely used JBSS algorithms in terms of the joint inter-symbol-interference (ISI), defined in [17], [9]. In this set of expirements, we generate MGGD sources and consider two different cases for the shape parameter $\beta$. For the first case we generate ten MGGD sourses

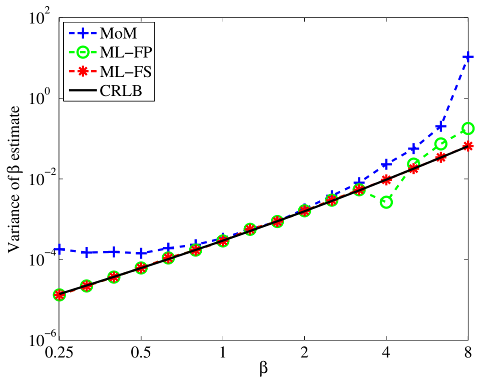


Fig. 2. Variance of $\beta$ estimate for different values of $\beta$. $T = 10000$, $\sigma \in (0.4, 0.6)$.

$N = 10$, $K = 3$, and $\beta$, $\sigma$ have been uniformly selected from the range $(0.25, 4)$ and $(0.4, 0.6)$ respectively. For the second case, we have $\beta \in (4, 8)$. From Fig.3 and Fig.4, we observe that IVA-A-GGD performs the best in terms of the joint ISI as function of sample sizes for any value of $\beta$, showing the effectiveness of IVA-A-GGD.

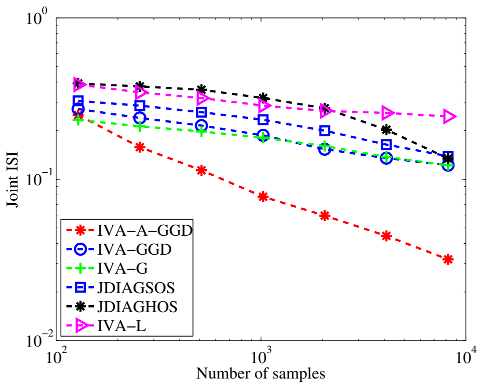


Fig. 3. Performance of IVA-A-GGD for different number of sample size. Data have been generated using $N = 10$ sources, $K = 3$ datasets, $\sigma \in (0.4, 0.6)$ and $\beta \in (0.25, 4)$.

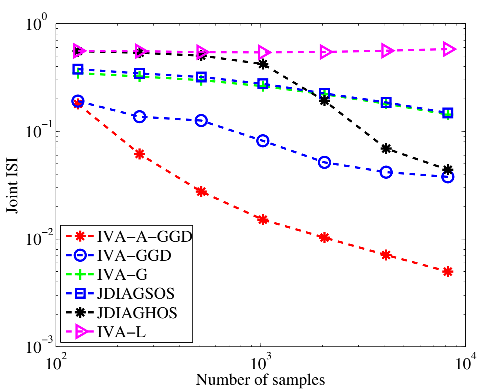


Fig. 4. Performance of IVA-A-GGD for different number of sample size. Data have been generated using $N = 10$ sources, $K = 3$ datasets, $\sigma \in (0.4, 0.6)$ and $\beta \in (4, 8)$.

## VI. Conclusion

We propose an effective ML estimation technique for MGGD by using the FS algorithm. Based on this technique, we introduce a new IVA algorithm, IVA-A-GGD that estimates shape parameter and scatter matrix jointly and exploits both second and higher-order statistics. Simulation results reveal the effectiveness of ML-FS algorithm as well as the desirable performance of IVA-A-GGD when compared with existing competing algorithms.

## References


[1] M. Z. Coban and R. Mersereau, "Adaptive subband video coding using bivariate generalized gaussian distribution model," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 4, pp. 1990–1993, IEEE, 1996.

[2] J. Yang, Y. Wang, W. Xu, and Q. Dai, "Image and video denoising using adaptive dual-tree discrete wavelet packets," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 19, no. 5, pp. 642–655, 2009.

[3] T. Elguebaly and N. Bouguila, "Bayesian learning of generalized gaussian mixture models on biomedical images," in *Artificial Neural Networks in Pattern Recognition*, pp. 207–218, Springer, 2010.

[4] 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*, vol. 95, no. 3, pp. 265–286, 2011.

[5] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, 2012.

[6] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized gaussian distributions," *Signal Processing, IEEE Transactions on*, vol. 61, pp. 5960–5971, Dec 2013.

[7] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ica to multivariate components," in *Independent Component Analysis and Blind Signal Separation*, pp. 165–172, Springer, 2006.

[8] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 70–79, 2007.

[9] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate gaussian model: Algorithms and performance analysis," *Signal Processing, IEEE Transactions on*, vol. 60, pp. 1672–1683, April 2012.

[10] J. Via, M. Anderson, X.-L. Li, and T. Adalı, "A maximum likelihood approach for independent vector analysis of gaussian data sets," in *Machine Learning for Signal Processing (MLSP), 2011 IEEE International Workshop on*, pp. 1–6, Sept 2011.

[11] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the kotz distribution, and performance bounds," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 3243–3247, May 2013.

[12] S. Kotz, *Multivariate distributions at a cross road*. Springer, 1975.

[13] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *Signal Processing Magazine, IEEE*, vol. 31, pp. 18–33, May 2014.

[14] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1803–1814, 2007.

[15] M. Anderson, X.-L. Li, P. Rodriguez, and T. Adali, "An effective decoupling method for matrix optimization and its application to the ica problem," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 1885–1888, IEEE, 2012.

[16] E. Gómez, M. Gomez-Viilegas, and J. Marin, "A multivariate generalization of the power exponential family of distributions," *Communications in Statistics-Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[17] Y.-O. Li, T. Adalı, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *Signal Processing, IEEE Transactions on*, vol. 57, no. 10, pp. 3918–3929, 2009.