A New Riemannian Averaged Fixed-Point Algorithm for MGGD Parameter Estimation

Zois Boukouvalas, Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Tülay Adalı

***Abstract*—Multivariate generalized Gaussian distribution (MGGD) has been an attractive solution to many signal processing problems due to its simple yet flexible parametric form, which requires the estimation of only a few parameters, i.e., the scatter matrix and the shape parameter. Existing fixed-point (FP) algorithms provide an easy to implement method for estimating the scatter matrix, but are known to fail, giving highly inaccurate results, when the value of the shape parameter increases. Since many applications require flexible estimation of the shape parameter, we propose a new FP algorithm, Riemannian averaged FP (RA-FP), which can effectively estimate the scatter matrix for any value of the shape parameter. We provide the mathematical justification of the convergence of the RA-FP algorithm based on the Riemannian geometry of the space of symmetric positive definite matrices. We also show using numerical simulations that the RA-FP algorithm is invariant to the initialization of the scatter matrix and provides significantly improved performance over existing FP and method-of-moments (MoM) algorithms for the estimation of the scatter matrix.**

***Index Terms*—Fixed-point algorithm, maximum likelihood estimation, multivariate generalized Gaussian distribution, Riemannian geometry, symmetric positive definite matrix.**

I. Introduction

MULTIVARIATE generalized Gaussian distributions belong to the family of elliptical distributions [1]. They are defined by their probability density functions,

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

where $\mathbf{x}\in\mathbb{R}^p$, $m>0$ is the scale parameter, $\beta>0$ is the shape parameter, and $\boldsymbol{\Sigma}\in\mathbb{R}^{p\times p}$ is a symmetric positive definite matrix, called the scatter matrix. In the case $\beta=1$, the MGGD is a multivariate Gaussian distribution, with $\boldsymbol{\Sigma}$ its covariance matrix. In general, the shape parameter β controls the peakedness and spread of the distribution. If $\beta<1$ the distribution of the marginals is more peaky than Gaussian with heavier tails, and if $\beta>1$, it is less peaky with lighter tails.

Recently, the estimation of the parameters of MGGD has received significant attention, due to the fact that MGGD has numerous applications including those in video coding, image denoising, and medical image analysis [2]–[6]. Existing approaches to this problem, see e.g., [7]–[13], attempt to estimate $\boldsymbol{\Sigma}$ for a given value of β. However, their accuracy suffers when the value of β becomes large, which makes them unsuitable for many applications. Our main contribution is the presentation of an effective method that yields accurate estimates of $\boldsymbol{\Sigma}$ for any value of β.

In [7], [8], method of moments (MoM) and maximum likelihood (ML) techniques were explored for the estimation of $\boldsymbol{\Sigma}$. With regard to ML estimators, it has become clear, from [9]–[13], that they can be computed using FP algorithms. These algorithms are relatively easy to implement, but numerical results show that they provide highly inaccurate results when $\beta\geq 2$.

In this letter, we present a new FP algorithm, called Riemannian averaged FP (RA-FP) that accurately estimates $\boldsymbol{\Sigma}$ for any positive value of β. The basic idea of the RA-FP algorithm is to implement successive Riemannian averages of fixed-point iterates, in order to prevent them from diverging away from the true value of $\boldsymbol{\Sigma}$. Moreover, we present a theoretical justification of the convergence of RA-FP, and using numerical experiments, we verify that the basic assumptions of the main proposition hold.

Section II describes the ML equations, which are to be solved in order to estimate $\boldsymbol{\Sigma}$. Section III defines the new RA-FP algorithm, used to solve these equations. Section IV presents numerical experiments, using synthetic data, to show the improved performance of the RA-FP algorithm in comparison with existing methods. Finally, Section V provides the proof of convergence of the RA-FP algorithm.

II. ML Equations

Let $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$ be a random sample of N observation vectors of dimension p, which are drawn from an MGGD with parameters $\boldsymbol{\Sigma}$, β, and m. The corresponding ML estimates $\hat{\beta}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{m}$ are found by solving the ML equations, described next.

Manuscript received July 14, 2015; accepted September 04, 2015. Date of publication September 15, 2015; date of current version September 21, 2015. This work was supported by the Investments for the Future Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02) of the French National Research Agency (ANR), and by the National Science Foundation (NSF) under Grant NSF-1117056. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tolga Tasdizen.

Z. Boukouvalas is with the Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: zb1@umbc.edu).

S. Said, L. Bombrun and Y. Berthoumieu are with Laboratoire IMS, Universit de Bordeaux, Bordeaux, France (e-mail: firstname.lastname@u-bordeaux.fr; lionel.bombrun@u-bordeaux.fr; yannick.berthoumieu@u-bordeaux.fr).

T. Adalı is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: adali@umbc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2478803

Assume first β is known. The ML estimate $\hat{\mathbf{\Sigma}}$ is the solution of the following equation [9]

$$\mathbf{\Sigma} = \sum_{i=1}^{N} \frac{p}{u_i + u_i^{1-\beta} \sum\limits_{i \neq j} u_j^{\beta}} \mathbf{x}_i \mathbf{x}_i^{\top}, \tag{1}$$

for unknown $\mathbf{\Sigma}$, where $u_i = \mathbf{x}_i^{\top} \mathbf{\Sigma}^{-1} \mathbf{x}_i$. Once $\hat{\mathbf{\Sigma}}$ has been computed, $\hat{m}$ is immediately given by

$$\hat{m} = \left(\frac{1}{N} \sum_{i=1}^{N} \hat{u}_i^{\beta} \right)^{\frac{1}{\beta}}, \tag{2}$$

where $\hat{u}_i = \mathbf{x}_i^{\top} \hat{\mathbf{\Sigma}}^{-1} \mathbf{x}_i$.

In the general case where β is unknown, $\hat{\mathbf{\Sigma}}$ and $\hat{\beta}$ are found by solving (1), along with

$$\gamma(\beta) = \frac{pN}{2 \sum_{i=1}^{N} u_i^{\beta}} \sum_{i=1}^{N} \left[u_i^{\beta} \ln(u_i) \right] - \frac{pN}{2\beta} \left[\Psi\left(\frac{p}{2\beta} \right) + \ln 2 \right] - N - \frac{pN}{2\beta} \ln \left(\frac{\beta}{pN} \sum_{i=1}^{N} u_i^{\beta} \right) = 0, \tag{3}$$

whose solution is $\hat{\beta}$. Here, Ψ is the digamma function. Once the solutions $\hat{\mathbf{\Sigma}}$ and $\hat{\beta}$ of (1) and (3) have been found, $\hat{m}$ is computed directly from (2).

It is seen from the above that the main difficulty, in the computation of $\hat{\beta}$, $\hat{\mathbf{\Sigma}}$ and $\hat{m}$, lies in solving (1). This is a nonlinear equation in the space of symmetric positive definite matrices. As in [9], [13], it is possible to formulate (1) as a fixed point equation. To do so, let $\mathcal{S}_+^p$ denote the space of $p \times p$ symmetric positive definite matrices. Consider the function $f : \mathcal{S}_+^p \to \mathcal{S}_+^p$, given by

$$f(\mathbf{\Sigma}) = \sum_{i=1}^{N} \frac{p}{u_i + u_i^{1-\beta} \sum\limits_{i \neq j} u_j^{\beta}} \mathbf{x}_i \mathbf{x}_i^{\top}, \tag{4}$$

Clearly, $f(\mathbf{\Sigma})$, as given above, is just the right hand side of equation (1). Therefore, this equation can be written as

$$\mathbf{\Sigma} = f(\mathbf{\Sigma}), \tag{5}$$

which is indeed a fixed point equation. In other words, the ML estimate $\hat{\mathbf{\Sigma}}$ is the solution of the fixed point equation (5) associated with the function f defined in (4). It is well-known that the solution of a fixed point equation, such as (5) may be attempted using an FP algorithm, which gives successive fixed point iterates

$$\mathbf{\Sigma}_{k+1} = f(\mathbf{\Sigma}_k) \qquad k = 0, 1, 2, \ldots \tag{6}$$

Indeed, this algorithm was used in [9], [13]. Concretely, it consists of repeating (6) until the iterates $\mathbf{\Sigma}_k$ stabilize, i.e., until there is no sensible difference between $\mathbf{\Sigma}_k$ and $\mathbf{\Sigma}_{k+1}$.

The convergence of the FP algorithm (6) depends on the function f being contractive, (in a sense to be made precise in Section V). In the present context, numerical experiments show that the function f, (which depends on β as can be seen in (4)), is not contractive when $\beta \geq 2$. The new RA-FP algorithm, presented in the following section, overcomes this difficulty.

III. RA-FP Algorithm

It has been shown in [9], [13], that the FP algorithm in (6) gives accurate estimates of $\mathbf{\Sigma}$ when $\beta < 2$. The main contribution of the present letter is to describe the new RA-FP algorithm, which is a generalization of the FP algorithm (6), and is capable of producing accurate estimates of $\mathbf{\Sigma}$ when $\beta \geq 2$.

The RA-FP algorithm uses the Riemannian geometry of the space $\mathcal{S}_+^p$. Precisely, it implements successive Riemannian averages of fixed point iterates. The definition of the Riemannian average of $\mathbf{P}, \mathbf{Q} \in \mathcal{S}_+^p$ is the following, (a more detailed discussion is given in Section V). For $t \in [0, 1]$, the Riemannian average with ratio t of $\mathbf{P}$ and $\mathbf{Q}$ is $\mathbf{P} \#_t \mathbf{Q}$, given as in [13]

$$\mathbf{P} \#_t \mathbf{Q} = \mathbf{P}^{1/2} (\mathbf{P}^{-1/2} \mathbf{Q} \mathbf{P}^{-1/2})^t \mathbf{P}^{1/2}, \tag{7}$$

where, on the right hand side, the exponent $(\cdot)^t$ denotes elevation of a symmetric matrix to the power t. Note that

$$\mathbf{P} \#_0 \mathbf{Q} = \mathbf{P} \qquad \mathbf{P} \#_1 \mathbf{Q} = \mathbf{Q}. \tag{8}$$

The RA-FP algorithm is defined as follows. When $\mathbf{\Sigma}_k$ is given, instead of defining $\mathbf{\Sigma}_{k+1}$ by (6), let

$$\mathbf{\Sigma}_{k+1} = \mathbf{\Sigma}_k \#_{t_k} f(\mathbf{\Sigma}_k), \tag{9}$$

where $t_k \in [0, 1]$. The RA-FP algorithm given in (9) is indeed a generalization of the FP algorithm in (6), since putting $t_k = 1$ in (9) yields (6), as can be seen from (8). In our work, we set

$$t_k = \frac{1}{k+1}. \tag{10}$$

A pseudo-code description of the RA-FP algorithm is given in Algorithm 1 below. The main part of this algorithm is the loop described in lines 4-11. The algorithm exits this loop when $D(k) < \mathrm{tol}$, where $D(k)$ is the relative difference between two successive estimates, and tol is a tolerance bound, chosen by the user. The loop is also terminated whenever the number of iterations exceeds a pre-defined upper bound $N_{\max}$.

Algorithm:1 RA-FP

1: **Input**: $\mathbf{X} \in \mathbb{R}^{p \times N}$, optionally β
2: Initialize $\mathbf{\Sigma}$ using either MoM or $\mathbf{\Sigma} = \mathbf{I}_p$
3: If β is not given initialize both $\mathbf{\Sigma}$ and β using MoM
4: Initialize $k = 0$
5: **while** $(D(k) > \mathrm{tol})$ and $(k < N_{\max})$ **do**
6: $k = k + 1$
7: Estimate $\mathbf{\Sigma}$ using one iteration of (9)
8: Normalize $\mathbf{\Sigma}$ so that $\mathrm{Tr}(\mathbf{\Sigma}) = p$
9: **if** β is not given **then**
10: Estimate β by applying Newton-Raphson to (3)
11: **else**
12: Go to step 5
13: **end**
14: **end**
15: Using $\mathbf{\Sigma}$ and β, estimate m with (2)
16: **Output**: $\mathbf{\Sigma}$, β, m

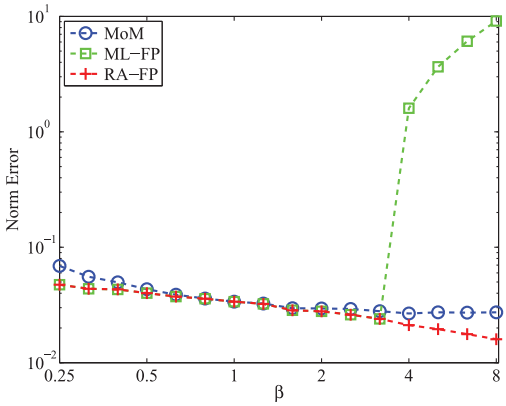

Fig. 1. Scatter matrix estimation performance for different values of the shape parameter, for $N = 10000$, $\sigma \in (0.4, 0.6)$.

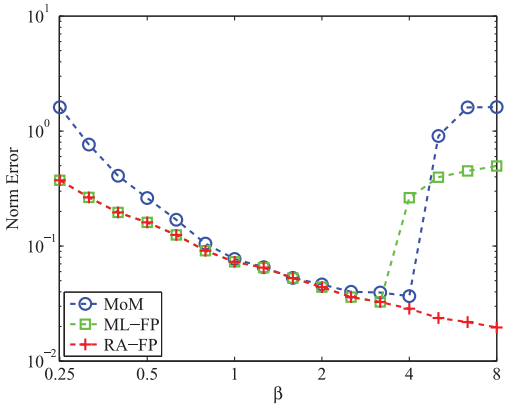

Fig. 2. Scatter matrix estimation performance for different values of the shape parameter when $\mathbf{\Sigma}$ and β have been jointly estimated. $N = 10000$, $\sigma \in (0.4, 0.6)$.

IV. Experimental Results

To quantify the performance of RA-FP, we generate data according to [9], [14] with $\mathbf{\Sigma}$ defined by

$$\mathbf{\Sigma}(i,j) = \sigma^{|i-j|},\ i,j = \{1,2,\dots p\}, \tag{11}$$

where σ belongs to the interval [0,1) and controls the correlations between the entries of the data. For these experiments, we used $p = 3$, $N = 10000$, and σ uniformly selected from the range $(0.4, 0.6)$. Similar results are obtained for different values of σ. All results are averaged over 500 runs.

Fig. 1 shows the Frobenius norm of the difference between the estimated and the original scatter matrix as a function of β. It can be observed that for $\beta < 1$, RA-FP and ML-FP provide better results than the MoM, while for $\beta \geq 2$ RA-FP performs the best among the three algorithms.

Fig. 2 displays the Frobenius norm of the difference between the estimated and the original scatter matrices, when $\mathbf{\Sigma}$ and β are jointly estimated. When $\beta < 1$, the two ML techniques perform better than MoM, and for $\beta > 4$, again RA-FP provides the best performance.

Finally, Fig. 3 shows the number of iterations for RA-FP to successfully converge to the true value as a function of β for different initializations. Here, the tolerance parameter, called tol in Algorithm 1, is chosen as 0.05. As observed in the figure, RA-FP converges for any value of $\beta \in (0.25, 8)$, and remains invariant to several choices of initialization. When $\beta = 1$, the MGGD reduces to the Gaussian distribution, where the ML estimator

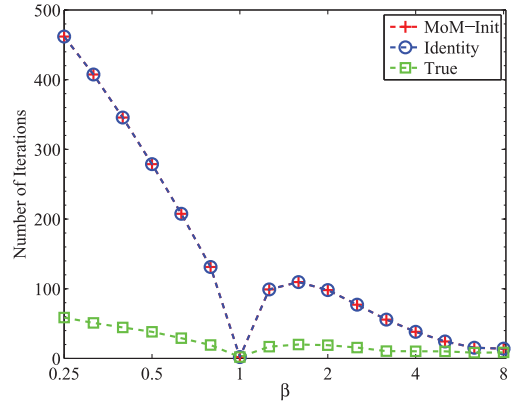

Fig. 3. Number of iterations needed for RA-FP to converge as a function of β. $N = 10000$, $\sigma \in (0.4, 0.6)$.

of the scatter becomes the covariance matrix, hence, only one iteration of RA-FP is sufficient. In addition, it is worth noting that the number of iterations depends on the value of β, and decreases as β becomes large.

V. Proof of Convergence

This section provides the proof of convergence of the RA-FP algorithm, which was presented in Section III.

The proof essentially relies on the Riemannian geometry of the space $\mathcal{S}_+^p$, the space of symmetric positive definite, $p \times p$ real matrices [15], [16]. The main geometric property to be used is the *strong convexity of Riemannian distance*[17], which is now explained.

To begin, the length of a differentiable curve $c : [0,1] \to \mathcal{S}_+^p$ is defined as [15]

$$L(c) = \int_0^1 \|c^{-1}(t)\dot{c}(t)\|_F \times dt, \tag{12}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Let $\mathbf{P}$ and $\mathbf{Q}$ be two points in $\mathcal{S}_+^p$. A curve c is said to connect $\mathbf{P}$ and $\mathbf{Q}$ if $c(0) = \mathbf{P}$ and $c(1) = \mathbf{Q}$. Among all curves connecting $\mathbf{P}$ and $\mathbf{Q}$, there exists a unique curve γ, whose length is minimum, (recall length is defined by (12)). This curve γ is called the *geodesic* connecting $\mathbf{P}$ and $\mathbf{Q}$. Its equation, in the notation of (7), is [16], [13]

$$\gamma(t) = \mathbf{P}\#_t\mathbf{Q}. \tag{13}$$

In particular, this exhibits the geometric meaning of the Riemannian average of $\mathbf{P}$ and $\mathbf{Q}$, defined in Section III. The Riemannian average with ratio t of $\mathbf{P}$ and $\mathbf{Q}$ is the point $\gamma(t)$ lying on the geodesic γ connecting $\mathbf{P}$ and $\mathbf{Q}$.

Riemannian distance between $\mathbf{P}$ and $\mathbf{Q}$, denoted $d(\mathbf{P},\mathbf{Q})$ is the length of the geodesic curve γ, defined by (13). Using (12), it can be found analytically [15],

$$d(\mathbf{P},\mathbf{Q}) = \|\log(\mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2})\|_F, \tag{14}$$

The main property of Riemannian distance, used in the proof of convergence of the RA-FP algorithm is its strong convexity [17]. This is defined as follows. Let $\mathbf{R},\mathbf{P},\mathbf{Q} \in \mathcal{S}_+^p$ and $\gamma : [0,1] \to \mathcal{S}_+^p$ the geodesic connecting $\mathbf{P}$ and $\mathbf{Q}$, given by (13). Then,

$$d^2(\mathbf{R},\gamma(t)) \leq td^2(\mathbf{R},\mathbf{Q}) + (1-t)d^2(\mathbf{R},\mathbf{P}) - t(1-t)d^2(\mathbf{P},\mathbf{Q}). \tag{15}$$

This inequality simply means the function $t \mapsto d^2(\mathbf{R}, \gamma(t))$, which is a real-valued function of the real variable t, is a strongly convex function.

Consider now, once more, the fixed point equation (5). The FP algorithm (6), produces iterates $\mathbf{\Sigma}_k$ which converge to the unique fixed point $\hat{\mathbf{\Sigma}}$ of the function f, whenever f is contractive. That is, whenever [18]

$$d(f(\mathbf{P}), f(\mathbf{Q})) \leq \lambda \times d(\mathbf{P}, \mathbf{Q}) \lambda < 1 \tag{16}$$

for all $\mathbf{P}, \mathbf{Q} \in \mathcal{S}_+^p$. On the other hand, the FP algorithm (6) has no guarantee of convergence when $\lambda = 1$, in which case f is said to be non-expansive. Precisely, in this case [18],

$$d(f(\mathbf{P}), f(\mathbf{Q})) \leq d(\mathbf{P}, \mathbf{Q}), \tag{17}$$

for all $\mathbf{P}, \mathbf{Q} \in \mathcal{S}_+^p$. For function f as defined in (4), numerical experiments have shown that, in a neighborhood of the true value $\mathbf{\Sigma}$, this function is contractive when $\beta < 2$, but only non expansive, when $\beta \geq 2$. In this case, as shown in Section IV, the FP algorithm (6) fails to converge to $\mathbf{\Sigma}$, while the RA-FP algorithm (9) converges systematically.

The mathematical explanation of this convergence is given in the following proposition.

Proposition 1: Let $f : \mathcal{S}_+^p \to \mathcal{S}_+^p$ be a function, which has a fixed point $\hat{\mathbf{\Sigma}}$. Assume there exists a neighborhood U of $\hat{\mathbf{\Sigma}}$, such that $\hat{\mathbf{\Sigma}}$ is the unique fixed point of f in U. Assume also f is non-expansive in U. That is, for $\mathbf{P}, \mathbf{Q} \in U$, inequality (17) holds. If $\mathbf{\Sigma}_0 \in U$ and, $\mathbf{\Sigma}_{k+1}$ is defined by the RA-FP algorithm (9), for $k = 0, 1, 2, \ldots$, then the sequence $\{\mathbf{\Sigma}_k\}$ remains in U and converges to $\hat{\mathbf{\Sigma}}$, as $k \to \infty$.

Proof: Assume $\mathbf{\Sigma}_k \in U$. Since $\hat{\mathbf{\Sigma}} \in U$, it follows from (17),

$$d(f(\hat{\mathbf{\Sigma}}), f(\mathbf{\Sigma}_k)) \leq d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k).$$

But $\hat{\mathbf{\Sigma}}$ is a fixed point of f, so $f(\hat{\mathbf{\Sigma}}) = \hat{\mathbf{\Sigma}}$. Replacing (18) in the above inequality, it follows that

$$d(\hat{\mathbf{\Sigma}}, f(\mathbf{\Sigma}_k)) \leq d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k). \tag{18}$$

Now, apply the strong convexity property (15), with $\mathbf{R} = \hat{\mathbf{\Sigma}}$, $\mathbf{P} = \mathbf{\Sigma}_k$, $\mathbf{Q} = f(\mathbf{\Sigma}_k)$, and $t = t_k$. Using (9) and (13), this gives

$$\begin{aligned} d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_{k+1}) \leq\ & t_k d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k) + (1 - t_k) d^2(\hat{\mathbf{\Sigma}}, f(\mathbf{\Sigma}_k)) \\ & - t_k(1 - t_k) d^2(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)). \end{aligned}$$

Replacing (18) in this last inequality, it follows after a short calculation

$$d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k) - d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_{k+1}) \geq t_k(1 - t_k) d^2(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)). \tag{19}$$

This shows that $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_{k+1}) \leq d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k)$. So if $\mathbf{\Sigma}_k$ belongs to U, so does $\mathbf{\Sigma}_{k+1}$. Thus, if $\mathbf{\Sigma}_0 \in U$, then the sequence $\{\mathbf{\Sigma}_k\}$ remains in U. To prove this sequence converges to $\hat{\mathbf{\Sigma}}$, sum (19) over $k = 0, \ldots, n-1$. This gives,

$$d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_0) - d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_n) \geq \sum_{k=0}^{n-1} t_k(1 - t_k) d^2(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)) \tag{20}$$

The right hand side of this inequality is bounded above by $d^2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_0)$, which does not depend on n. Therefore,

$$\sum_{k=0}^{\infty} t_k(1 - t_k) d^2(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)) < +\infty. \tag{21}$$

To complete the proof, take the neighborhood U of $\hat{\mathbf{\Sigma}}$ to be compact. This can be done without any loss of generality.

The sequence $\mathbf{\Sigma}_k$ converges to $\hat{\mathbf{\Sigma}}$ if and only if $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k) \to 0$. It is now shown that assuming this is not true would lead to a contradiction.

By (19), the sequence of distances $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k)$ is decreasing. Therefore, if it does not converge to 0, there exists a positive number δ such that $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k) \geq \delta$ for all k.

Let C be the set of matrices $\mathbf{\Sigma}$ such that $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \geq \delta$. This is a closed set. Therefore, the set $U \cap C$ is compact. Note the function $\mathbf{\Sigma} \mapsto d(\mathbf{\Sigma}, f(\mathbf{\Sigma}))$ is continuous. Therefore, this function reaches its minimum, say c, over $U \cap C$. Since $U \cap C$ does not contain any fixed points of f, it follows that $c > 0$.

It has been proved that $\mathbf{\Sigma}_k \in U$ for all k, and that, assuming $\mathbf{\Sigma}_k$ does not converge to $\hat{\mathbf{\Sigma}}$, $\mathbf{\Sigma}_k \in C$ for all k. In this case, $\mathbf{\Sigma}_k \in U \cap C$ for all k. This implies $d(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)) \geq c$ for all k. Replacing in the right hand side of (21),

$$\sum_{k=0}^{\infty} t_k(1 - t_k) d^2(\mathbf{\Sigma}_k, f(\mathbf{\Sigma}_k)) \geq c^2 \sum_{k=0}^{\infty} t_k(1 - t_k).$$

Since $t_k = \frac{1}{1+k}$, this sum is infinite, which contradicts (21).

Since the assumption that $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k)$ does not converge to zero has lead to a contradiction, it follows that $d(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}_k) \to 0$, which means that $\mathbf{\Sigma}_k$ converges to $\hat{\mathbf{\Sigma}}$. ■

Recall that function f is defined by (4) within the ML framework for the estimation of MGGD parameters. As discussed right after (5), the maximum likelihood estimate $\hat{\mathbf{\Sigma}}$ of the scatter matrix $\mathbf{\Sigma}$ is a fixed point of this function. Moreover, as discussed after (17), numerical experiments have shown that this function verifies the assumption of non-expansivity in a small neighborhood of the true value, and since for sufficiently large sample size the maximum likelihood estimate $\hat{\mathbf{\Sigma}}$ is expected to be close to the true value, Proposition 1 asserts that the RA-FP algorithm (9) applied to function f converges to $\hat{\mathbf{\Sigma}}$, if it is initialized in a small neighborhood of the true value. This is in full agreement with the numerical results of Section IV.

VI. Conclusion

This letter presented a new FP algorithm, for the estimation of MGGD parameters, $\mathbf{\Sigma}$, β and m. This new algorithm, called RA-FP, unlike the ones existing in current literature, is able to estimate $\mathbf{\Sigma}$ for any value of β. It is based on the idea of implementing Riemannian averages of successive fixed point iterates, preventing them from diverging when the value of β increases. Numerical results show that for any value of β, and any initialization of the RA-FP algorithm, this algorithm converges to the true value of $\mathbf{\Sigma}$. Finally, we have proved that RA-FP converges when initialized in the neighborhood of any fixed point, assuming that the mapping f is non-expansive in that neighborhood. Numerical simulations provide empirical support of this assumption, and enable the mathematical proof of the convergence of the RA-FP algorithm.

REFERENCES

[1] S. Kotz, *Multivariate Distributions at a Cross Road*. Berlin, Germany: Springer, 1975.

[2] M. Z. Coban and R. Mersereau, "Adaptive subband video coding using bivariate generalized Gaussian distribution model," in *IEEE International Conf. Acoustics, Speech and Signal Processing (ICASSP), 1996*, Mar. 1996, vol. 4, pp. 1990–1993.

[3] 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, 2009.

[4] M. S. Allili, N. Bouguila, and D. Ziou, "Finite general Gaussian mixture modeling and application to image and video foreground segmentation," *J. Electron. Imag.*, vol. 17, no. 1, pp. 1–13, 2008.

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

[6] J. Laney, K. P. Westlake, S. Ma, E. Woytowicz, V. D. Calhoun, and T. Adalı, "Capturing subject variability in fMRI data: A graph-theoretical analysis of GICA vs. IVA," *J. Neurosci. Meth.*, vol. 247, pp. 32–40, 2015.

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

[8] 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, 2011.

[9] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Trans. Signal Process.*, vol. 61, pp. 5960–5971, Dec. 2013.

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

[11] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, pp. 5597–5625, Nov. 2012.

[12] L. Bombrun, F. Pascal, J.-Y. Tourneret, and Y. Berthoumieu, "Performance of the maximum likelihood estimators for the parameters of multivariate generalized Gaussian distributions," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2012*, Mar. 2012, pp. 3525–3528.

[13] S. Sra and R. Hosseini, "Geometric optimisation on positive definite matrices for elliptically contoured distributions," *Adv. Neural Inf. Process. Syst.*, pp. 2562–2570, 2013.

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

[15] A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications*. New York, NY, USA: Springer-Verlag, 1988, vol. II.

[16] R. Bhatia, *Positive Definite Matrices*. Princeton, NJ, USA: Princeton Univ. Press, 2007.

[17] K. Sturm, "Probability measures on metric spaces of nonpositive curvature," *Contemp. Math.*, vol. 338, pp. 357–390, 2003.

[18] D. Smart, *Fixed Point Theorems*. Cambridge, U.K.: Cambridge Univ. Press, 1974.