On-line Kronecker Product Structured Covariance Estimation with Riemannian geometry for t-distributed data

Florent Bouchard*, Arnaud Breloy†, Ammar Mian‡, Guillaume Ginolhac‡
* CentraleSupélec, CNRS, Univ. Paris Saclay, † LEME, University Paris Nanterre, ‡ LISTIC, University Savoie Mont-Blanc

***Abstract*—The information geometry of the zero-mean t-distribution with Kronecker-product structured covariance matrix is derived. In particular, we obtain the Fisher information metric which shows that this geometry is identifiable to a product manifold of $\mathcal{S}_p^{++}$ (positive definite symmetric matrices) and $s\mathcal{S}_p^{++}$ (positive definite symmetric matrices with unit determinant). From this result, we obtain the geodesics and the Riemannian gradient. Finally, this geometry makes it possible to propose an on-line covariance matrix estimation algorithm well adapted to large datasets. Numerical experiments show that optimal results are obtained for a reasonable number of data.**

***Index Terms*—covariance, robust estimation, kronecker product, Riemannian geometry, on-line estimation**

I. Introduction

Kronecker product structured covariance matrices occur in many applications such as MIMO communication [1], MEG/EEG data analysis [2], and separable statistical models [3]. Numerous works were thus devoted to estimate the factors of these Kronecker products from a given set of normally distributed measurement, see e.g. [4], [5] and reference therein. All these references assume a that data follow a Gaussian distribution. It is well known that this assumption is not realistic in many practical situations as RADAR [6], [7], SONAR [8] or in remote sensing [9]. Recent extensions considered the more general elliptical models [10] in order to derive estimation methods that are robust to samples possibly drawn from heavy-tailed distributions. Notably, [11], [12] studied the uniqueness/existence of the maximum likelihood estimator from normalized samples that follows a centered elliptical distribution, [13] proposed majorization-minimization algorithms for robust estimation of the Kronecker products factors, [14] proposed and studied the asymptotic the distribution of algorithms based on the extended invariance principle when these factors admit a linear structure.

Even if the obtained estimators have interesting performance, they suffer of a major practical issue: existing robust estimation methods in the considered context are computationally intensive. Indeed, these robust estimators satisfy intricate fixed-point equations, and thus, require iterative procedures to be computed. This can be a limitation, e.g. in large time-series analysis, where such computation cannot be be carried out each time a new sample is added to the batch [15]. In this configuration, a simple idea can be to update the robust covariance estimator as soon as a new data occurs without computing the fixed-point equations. A possible solution is given in the reference [16], which proposes a new on-line estimation algorithm to estimate the parameters of a CES distribution. Applying this approach requires to consider the information geometry (i.e., geometry of the parameter space induced by the Fisher information metric) of the model/parameters, which has not been derived for Kronecker product structured matrices. Therefore, in this paper, we propose to fill this gap and the contributions are summarized as follows:

- We focus on estimation procedures for Kronecker product structured covariance matrices formulated as the maximum likelihood estimator of the real multivariate t-distribution. This distribution family is especially interesting in robust estimation, as it can accounts for heavy tails, and includes both Cauchy and Gaussian distributions as special (or limit) case [10].
- We study the information geometry of the corresponding statistical model. This geometry turns out to be identifiable to the one of the product manifold of $\mathcal{S}_p^{++}$ (positive definite symmetric matrices) and $s\mathcal{S}_p^{++}$ (positive definite symmetric matrices with unit determinant) with a specific choice of (separable) affine invariant metric as in [17]. Thus well known results for these manifolds [18], allow us to obtain usual tools such as geodesics and Riemannian gradient.
- These tools are used to propose a new on-line estimation algorithm, relying on the methodology proposed in [16].
- Finally, the proposed algorithm is validated by a numerical experiment showing that the online procedure allows to obtain an optimal result in settings that are achievable in practical situations.

II. Model

The data set $\{\boldsymbol{x}_i\}_{i=1}^n \in (\mathbb{R}^p)^n$ is assumed to contain independent and identically distributed vectors drawn from the multivariate real Student t-distribution with unknown scatter matrix $\boldsymbol{\Sigma}$ and known $d \in \mathbb{N}^*$ degrees of freedom. The model, denoted $\boldsymbol{x} \sim \mathbb{R}t_d(\boldsymbol{0}, \boldsymbol{\Sigma})$, implies that the probability density function of $\boldsymbol{x}$ is of the form

$$f(\boldsymbol{x}) \propto |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}}{d}\right)^{-(d+p)/2}. \quad (1)$$

Furthermore, the scatter matrix $\boldsymbol{\Sigma}$ is assumed to admit a Kronecker product structure, *i.e.*, $\boldsymbol{\Sigma} = \boldsymbol{A} \otimes \boldsymbol{B}$, where $\boldsymbol{A} \in s\mathcal{S}_a^{++} = \{\boldsymbol{M} \in \mathcal{S}_a^{++} : |\boldsymbol{M}| = 1\}$ and $\boldsymbol{B} \in \mathcal{S}_b^{++}$. To avoid any scaling ambiguity on $\boldsymbol{\Sigma}$, an arbitrary normalization on $\boldsymbol{A}$

(or $\boldsymbol{B}$) is needed. We choose the unit determinant one, which, as noted in [15], [19], is particularly advantageous from a geometrical point of view.

The parameter of interest $\theta = (\boldsymbol{A}, \boldsymbol{B})$ thus lies in the product manifold $\mathcal{M}_{a,b} = s\mathcal{S}_a^{++} \times \mathcal{S}_b^{++}$ and the corresponding covariance matrix is obtained through the mapping $\varphi : \mathcal{M}_{a,b} \to \mathcal{S}_p^{++}$ such that

$$\varphi(\theta) = \boldsymbol{A} \otimes \boldsymbol{B}. \tag{2}$$

The negative log-likelihood on $\mathcal{M}_{a,b}$ corresponding to (1) is (up to an additive constant):

$$\mathcal{L}(\{\mathbf{x}_i\}_{i=1}^n; \theta) = \sum_{i=1}^n \rho(\mathbf{x}_i^T \varphi(\theta)^{-1} \mathbf{x}_i) + \frac{n}{2} \log |\varphi(\theta)|, \tag{3}$$

where $\rho(t) = \frac{(d+p)}{2} \log(d+t)$.

III. Information geometry of $\mathcal{M}_{a,b}$ induced by the t-distribution

In this section, we study the Riemannian geometry of $\mathcal{M}_{a,b}$ equipped with the Fisher information metric induced by the likelihood (3). By definition, the tangent space of $\mathcal{M}_{a,b}$ at θ is $T_\theta \mathcal{M}_{a,b} = T_{\boldsymbol{A}} s\mathcal{S}_a^{++} \times T_{\boldsymbol{B}} \mathcal{S}_b^{++}$. The tangent space $T_{\boldsymbol{B}} \mathcal{S}_b^{++}$ can be identified to $\mathcal{S}_b$ (symmetric matrices) while

$$T_{\boldsymbol{A}} s\mathcal{S}_a^{++} = \{\boldsymbol{\xi}_{\boldsymbol{A}} \in \mathcal{S}_a : \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}) = 0\}. \tag{4}$$

In the following, $\xi = (\boldsymbol{\xi}_{\boldsymbol{A}}, \boldsymbol{\xi}_{\boldsymbol{B}})$ and $\eta = (\boldsymbol{\eta}_{\boldsymbol{A}}, \boldsymbol{\eta}_{\boldsymbol{B}})$ denote two elements from $T_{\boldsymbol{B}} \mathcal{S}_b^{++}$. The Fisher information metric on $\mathcal{M}_{a,b}$ induced by the t-distribution is provided in Proposition 1.

Proposition 1. *Given $\theta \in \mathcal{M}_{a,b}$, ξ and $\eta \in T_\theta \mathcal{M}_{a,b}$, the Fisher information metric on $\mathcal{M}_{a,b}$ induced by the likelihood* (3) *is*

$$\begin{aligned}\langle \xi, \eta \rangle_\theta = \alpha b \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}\boldsymbol{A}^{-1}\boldsymbol{\eta}_{\boldsymbol{A}}) + \alpha a \operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}}\boldsymbol{B}^{-1}\boldsymbol{\eta}_{\boldsymbol{B}}) \\ + (\alpha - 1) a^2 \operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}}) \operatorname{tr}(\boldsymbol{B}^{-1}\boldsymbol{\eta}_{\boldsymbol{B}}),\end{aligned}$$

where $\alpha = {}^{(d+p)}/{}_{(d+p+1)}$.

Proof. From [20, Proposition 7], we have

$$\langle \xi, \eta \rangle_\theta = \langle \mathrm{D}\,\varphi(\theta)[\xi], \mathrm{D}\,\varphi(\theta)[\eta] \rangle_{\varphi(\theta)}^{\mathcal{S}_p^{++}},$$

where $\langle \cdot, \cdot \rangle_{\cdot}^{\mathcal{S}_p^{++}}$ is the Fisher metric of the t-distribution on $\mathcal{S}_p^{++}$ and the directional derivative of φ at θ in the direction ξ is

$$\mathrm{D}\,\varphi(\theta)[\xi] = \boldsymbol{A} \otimes \boldsymbol{\xi}_{\boldsymbol{B}} + \boldsymbol{\xi}_{\boldsymbol{A}} \otimes \boldsymbol{B}.$$

From [17], we know that

$$\begin{aligned}\langle \boldsymbol{\xi}_{\boldsymbol{\Sigma}}, \boldsymbol{\eta}_{\boldsymbol{\Sigma}} \rangle_{\boldsymbol{\Sigma}}^{\mathcal{S}_p^{++}} &= \alpha \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}) \\ &\quad + (\alpha - 1) \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}) \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}).\end{aligned}$$

The result is obtained by plugging $\mathrm{D}\,\varphi(\theta)[\cdot]$ into $\langle \cdot, \cdot \rangle_{\cdot}^{\mathcal{S}_p^{++}}$ and exploiting the relations $(\boldsymbol{M} \otimes \boldsymbol{N})(\boldsymbol{O} \otimes \boldsymbol{P}) = \boldsymbol{MO} \otimes \boldsymbol{NP}$, $\operatorname{tr}(\boldsymbol{M} \otimes \boldsymbol{N}) = \operatorname{tr}(\boldsymbol{M}) \operatorname{tr}(\boldsymbol{N})$ and $\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}) = \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{\eta}_{\boldsymbol{A}}) = 0$ (by definition of the tangent space at $\boldsymbol{A}$). □

One can notice that the metric is separable into two scaled Riemannian affine invariant metrics applying on $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The geometry resulting from such a product manifold metric is simply obtained by combining the geometries corresponding to each component. In particular, the Riemannian exponential mapping at $\theta \in \mathcal{M}_{a,b}$ is defined for $\xi \in T_\theta \mathcal{M}_{a,b}$ as

$$\exp_\theta^{\mathcal{M}_{a,b}}(\xi) = \left(\boldsymbol{A} \exp(\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}), \boldsymbol{B} \exp(\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}})\right). \tag{5}$$

This exponential mapping is very useful when it comes to optimizing a cost function on the manifold. However, for better numerical cost and stability, it might be advantageous to prefer a second order approximation, defined as

$$\begin{aligned}R_\theta(\xi) = \Bigg(&\boldsymbol{A} + \boldsymbol{\xi}_{\boldsymbol{A}} + \frac{1}{2}\boldsymbol{\xi}_{\boldsymbol{A}}\boldsymbol{A}^{-1}\boldsymbol{\xi}_{\boldsymbol{A}}, \\ &\boldsymbol{B} + \boldsymbol{\xi}_{\boldsymbol{B}} + \frac{1}{2}\boldsymbol{\xi}_{\boldsymbol{B}}\boldsymbol{B}^{-1}\boldsymbol{\xi}_{\boldsymbol{B}}\Bigg).\end{aligned} \tag{6}$$

IV. Online estimation

Given some data $\{\boldsymbol{x}_i\}_{i=1}^n$, the goal is to obtain the maximum likelihood estimator $\theta = (\boldsymbol{A}, \boldsymbol{B})$ in $\mathcal{M}_{a,b}$, which is solution to the optimization problem

$$\underset{\theta \in \mathcal{M}_{a,b}}{\operatorname{argmin}} \quad \mathcal{L}(\{\mathbf{x}_i\}_{i=1}^n; \theta) = \sum_i \ell_i(\theta), \tag{7}$$

where

$$\ell_i(\theta) = \frac{d+p}{2} \log(d + \boldsymbol{x}_i^T \varphi(\theta)^{-1} \boldsymbol{x}_i) + \frac{1}{2} \log |\varphi(\theta)|.$$

This problem can be solved on the manifold by means of Riemannian optimization [18]. In the simple case of the gradient descent, given iterate θ_k, one first needs to obtain the Riemannian gradient of the cost function $\sum_i \operatorname{grad} \ell_i(\theta_k)$. The next iterate is then obtained as

$$\theta_{k+1} = R_{\theta_k}(-t_k \textstyle\sum_i \operatorname{grad} \ell_i(\theta_k)), \tag{8}$$

where R_{θ_k} is given in (6) and t_k is a stepsize which can for instance be computed with a line-search.

In some scenarios, the whole data may not be available during the optimization process, either because of practical dimensionality issues, or because it has not been recorded yet. In such a case, one can turn to online optimization algorithms based on stochastic gradient descent, as for instance considered in [15], [16]. At each iteration, only one data $\boldsymbol{x}_i$ is exploited in order to update the estimated solution. For $1 \leq i \leq n$, given iterate θ_i, the next iterate is obtained as

$$\theta_{i+1} = R_{\theta_i}\left(-\frac{1}{i} \operatorname{grad} \ell_i(\theta_i)\right). \tag{9}$$

This procedure has the advantage of being lighter than the usual one in terms of computational cost while still providing some convergence properties [16]. In order to apply it to the optimization problem of interest, it remains to compute the Riemannian gradient of ℓ_i according to the metric of Proposition 1. This Riemannian gradient is given in Proposition 2.

Proposition 2. *The Riemannian gradient of ℓ_i at $\theta \in \mathcal{M}_{a,b}$ according to the metric of Proposition 1 is given by*

$$\operatorname{grad} \ell_i(\theta) = \left(\frac{1}{\alpha b} P_{\boldsymbol{A}}(\boldsymbol{A} \operatorname{sym}(\nabla_{\boldsymbol{A}} \ell_i(\theta)) \boldsymbol{A}), \right.$$
$$\left. \frac{1}{\alpha a} \boldsymbol{B} \operatorname{sym}(\nabla_{\boldsymbol{B}} \ell_i(\theta)) \boldsymbol{B} - \frac{(\alpha - 1) \operatorname{tr}(\boldsymbol{B} \nabla_{\boldsymbol{B}} \ell_i(\theta))}{\alpha(\alpha + (\alpha - 1)p)} \boldsymbol{B} \right),$$

where $\operatorname{sym}(\cdot)$ *returns the symmetrical part of its argument;* $P_{\boldsymbol{A}} : \mathcal{S}_a \to T_{\boldsymbol{A}} s\mathcal{S}_a^{++}$ *is the orthogonal projection map such that*

$$P_{\boldsymbol{A}}(\boldsymbol{\xi}_{\boldsymbol{A}}) = \boldsymbol{\xi}_{\boldsymbol{A}} - \frac{\operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}})}{a} \boldsymbol{A};$$

and $\nabla \ell_i(\theta) = (\nabla_{\boldsymbol{A}} \ell_i(\theta), \nabla_{\boldsymbol{B}} \ell_i(\theta))$ *is the Euclidean gradient of* ℓ_i *at* θ*, defined as*

$$\nabla_{\boldsymbol{A}} \ell_i(\theta) = \frac{1}{2} \boldsymbol{A}^{-1} \left(b\boldsymbol{A} - \frac{d+p}{d + Q_i(\theta)} \boldsymbol{M}_i^T \boldsymbol{B}^{-1} \boldsymbol{M}_i \right) \boldsymbol{A}^{-1},$$
$$\nabla_{\boldsymbol{B}} \ell_i(\theta) = \frac{1}{2} \boldsymbol{B}^{-1} \left(a\boldsymbol{B} - \frac{d+p}{d + Q_i(\theta)} \boldsymbol{M}_i \boldsymbol{A}^{-1} \boldsymbol{M}_i^T \right) \boldsymbol{B}^{-1},$$

with $\boldsymbol{M}_i$*, the* $b \times a$ *matrix such that* $\operatorname{vec}(\boldsymbol{M}_i) = \boldsymbol{x}_i$ *and* $Q_i(\theta) = \operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{M}_i^T \boldsymbol{B}^{-1} \boldsymbol{M}_i)$.

Proof. Recalling that $\boldsymbol{x}_i^T \varphi(\theta)^{-1} \boldsymbol{x}_i = \operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{M}_i^T \boldsymbol{B}^{-1} \boldsymbol{M}_i)$ and $|\boldsymbol{A} \otimes \boldsymbol{B}| = |\boldsymbol{A}|^b |\boldsymbol{B}|^a$ allows to write ℓ_i as

$$\ell_i(\theta) = \frac{(d+p)}{2} \log(d + Q_i(\theta)) + \frac{b}{2} \log |\boldsymbol{A}| + \frac{a}{2} \log |\boldsymbol{B}|.$$

Its directional derivative is

$$\mathrm{D}\, \ell_i(\theta)[\xi] = \frac{1}{2} \frac{(d+p)}{d + Q_i(\theta)} \mathrm{D}\, Q_i(\theta)[\xi]$$
$$+ \frac{b}{2} \operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}}) + \frac{a}{2} \operatorname{tr}(\boldsymbol{B}^{-1} \boldsymbol{\xi}_{\boldsymbol{B}}),$$

where

$$\mathrm{D}\, Q_i(\theta)[\xi] = -\operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{\xi}_{\boldsymbol{A}} \boldsymbol{A}^{-1} \boldsymbol{M}_i^T \boldsymbol{B}^{-1} \boldsymbol{M}_i)$$
$$- \operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{M}_i^T \boldsymbol{B}^{-1} \boldsymbol{\xi}_{\boldsymbol{B}} \boldsymbol{B}^{-1} \boldsymbol{M}_i).$$

The Euclidean gradient $\nabla \ell_i(\theta)$ of ℓ_i at θ is then given by identification

$$\mathrm{D}\, \ell_i(\theta)[\xi] = \operatorname{tr}(\nabla_{\boldsymbol{A}} \ell_i(\theta) \boldsymbol{\xi}_{\boldsymbol{A}}^T) + \operatorname{tr}(\nabla_{\boldsymbol{B}} \ell_i(\theta) \boldsymbol{\xi}_{\boldsymbol{B}}^T).$$

The Riemannian gradient according to the metric of Proposition 1 is obtained from the Euclidean one through identification

$$\langle \operatorname{grad} \ell_i(\theta), \xi \rangle_\theta = \operatorname{tr}(\nabla_{\boldsymbol{A}} \ell_i(\theta) \boldsymbol{\xi}_{\boldsymbol{A}}^T) + \operatorname{tr}(\nabla_{\boldsymbol{B}} \ell_i(\theta) \boldsymbol{\xi}_{\boldsymbol{B}}^T),$$

and projection onto the tangent space. □

V. Numerical results

In this section, we compare the performance of the maximum likelihood estimator obtained with a usual Riemannian optimization algorithm[1] and of the one obtained with the on-line procedure. In order to do so, we perform Kronecker product structured covariance estimation of simulated data drawn from the multivariate Student t-distribution with $d = 3$ (highly non-Gaussian) and $d = 100$ (almost Gaussian) degrees of freedom.

To generate a $p \times p$ ($p = 16$) covariance matrix, we compute

$$\boldsymbol{\Sigma} = \boldsymbol{A} \otimes \boldsymbol{B}, \qquad \boldsymbol{A} = \boldsymbol{U}_{\boldsymbol{A}} \boldsymbol{\Lambda}_{\boldsymbol{A}} \boldsymbol{U}_{\boldsymbol{A}}^T, \qquad \boldsymbol{B} = \boldsymbol{U}_{\boldsymbol{B}} \boldsymbol{\Lambda}_{\boldsymbol{B}} \boldsymbol{U}_{\boldsymbol{B}}^T, \tag{10}$$

where $a = b = 4$,

- $\boldsymbol{U}_{\boldsymbol{A}}$ and $\boldsymbol{U}_{\boldsymbol{B}}$ are random orthogonal matrices,
- $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{B}}$ are diagonal matrices whose minimal and maximal elements are $1/\sqrt{c}$ and $\sqrt{c}$ ($c = 10$ is the condition number with respect to inversion); their other elements are randomly drawn from the uniform distribution between $1/\sqrt{c}$ and $\sqrt{c}$; the determinant of $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ is then normalized.

100 sets $\{\boldsymbol{x}_i\}_{i=1}^n$ are drawn from the multivariate Student t-distribution with covariance $\boldsymbol{\Sigma}$ and $d \in \{3, 100\}$ degrees of freedom, where $n \in [\![1, 500]\!]$.

For this experiment, we consider the following estimators:

- the classical maximum-likelihood estimator obtained with Riemannian gradient descent (GD). Optimization for this estimator is performed with manopt toolbox [21].
- the online version obtained through stochastic gradient descent (SGD) presented in Section IV.

Both algorithms are initialized with $\theta_0 = (\boldsymbol{I}_a, \boldsymbol{I}_b)$.

In order to measure the performance of the estimators, we consider an error measure for each component $\boldsymbol{A}$ and $\boldsymbol{B}$, which are given by the usual Riemannian distances on $s\mathcal{S}_a^{++}$ and $\mathcal{S}_b^{++}$

$$\operatorname{err}(\widehat{\boldsymbol{A}}) = \|\log(\boldsymbol{A}^{-1/2} \widehat{\boldsymbol{A}} \boldsymbol{A}^{-1/2})\|_2^2, \qquad \operatorname{err}(\widehat{\boldsymbol{B}}) = \|\log(\boldsymbol{B}^{-1/2} \widehat{\boldsymbol{B}} \boldsymbol{B}^{-1/2})\|_2^2. \tag{11}$$

The results are presented in Figure 1. First, we notice that the on-line version converges to the classical estimate for both values of d which is expected as stated in [16]. For a small number of samples, the error for the quasi-Gaussian configuration is around 4-5 dB for the estimation of $\boldsymbol{A}$ and $\boldsymbol{B}$. However, we notice that the error is much smaller (around 1 dB) for the heavy-tailed configuration. This preliminary result shows that the on-line version can be used in real applications when the number of the samples of the dataset becomes too large to use classical robust estimators.

VI. Conclusions and perspectives

In this paper, we have developed a new on-line estimation algorithm for the Kronecker-product structured covariance matrix when data follow a real Student t-distribution. This algorithm is based on tools of the Riemannian geometry resulting from the Kronecker product of two symmetric positive definite matrices. Numerical experiments show achievable results for a small size of the data set.

Future works will be devoted to the extension of the approach to complex data and to the whole class of complex elliptically symmetric distributions. We will also test our

[1] Note that some other optimization method can still be envisioned to compute the MLE. However, the performance in terms of estimation accuracy is not expected to vary because the g-convexity of the objective function [11]

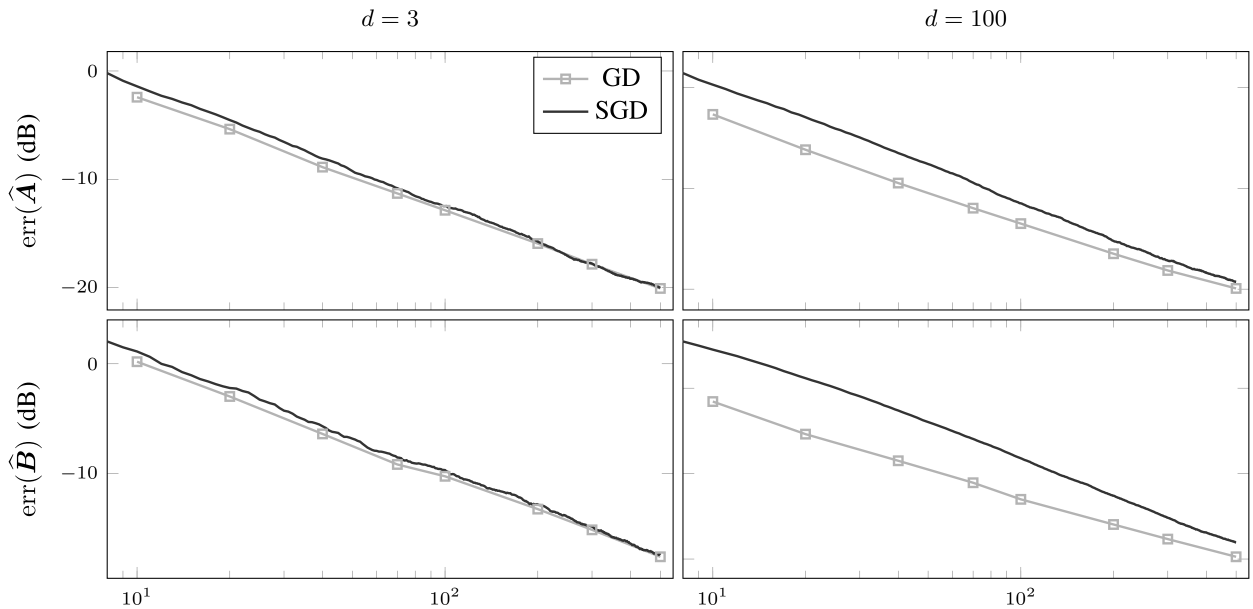

Fig. 1. Mean of error measures on $\boldsymbol{A}$ (top) and $\boldsymbol{B}$ (bottom) of the classical gradient descent method (GD) and its on-line counterpart (SGD) as functions of the number of samples n. Means are computed over 100 simulated sets $\{\boldsymbol{x}_i\}_{i=1}^n$ with $d = 3$ (left) and $d = 100$ (right). The data size is $p = 16$ and $a = b = 4$.

algorithm on different applications such as RADAR MIMO in order to detect targets [1] or to detect changes in large time series of SAR images [9]. In these applications the computational cost to obtain robust estimators is too expensive and the on-line procedure seems to be an interesting opening.

REFERENCES

[1] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 655–665, 2004.

[2] J. C. De Munck, H. M. Huizenga, L. J. Waldorp, and R. Heethaar, "Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1565–1572, 2002.

[3] N. Lu and D. L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statistics & probability letters*, vol. 73, no. 4, pp. 449–457, 2005.

[4] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with Kronecker product structure," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 478–491, 2008.

[5] M. S. Srivastava, T. von Rosen, and D. Von Rosen, "Models with a Kronecker product covariance structure: estimation and testing," *Mathematical Methods of Statistics*, vol. 17, no. 4, pp. 357–370, 2008.

[6] J. Billingsley, "Ground clutter measurements for surface-sited radar," MIT, Tech. Rep. 780, February 1993.

[7] M. Greco, F. Gini, and M. Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *Radar, Sonar and Navigation, IEE Proceedings -*, vol. 153, no. 6, pp. 473–481, December 2006.

[8] A. Saucan, T. Chonavel, C. Sintes, and J. Le Caillec, "CPHD-DOA tracking of multiple extended sonar targets in impulsive environments," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1147–1160, 2016.

[9] A. Mian, G. Ginolhac, J. Ovarlez, and A. M. Atto, "New robust statistics for change detection in time series of multivariate SAR images," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 520–534, 2019.

[10] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on signal processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[11] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE transactions on signal processing*, vol. 60, no. 12, pp. 6182–6189, 2012.

[12] I. Soloveychik and D. Trushin, "Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness," *Journal of Multivariate Analysis*, vol. 149, pp. 92–113, 2016.

[13] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3576–3590, 2016.

[14] B. Mériaux, C. Ren, A. Breloy, M. N. E. Korso, and P. Forster, "Matched and mismatched estimation of Kronecker product of linearly structured scatter matrices under elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 69, pp. 603–616, 2021.

[15] F. Bouchard, A. Mian, J. Zhou, S. Said, G. Ginolhac, and Y. Berthoumieu, "Riemannian geometry for compound Gaussian distributions: application to recursive change detection," *Signal Processing*, vol. 176, p. 107716, 2020.

[16] J. Zhou and S. Said, "Fast, asymptotically efficient, recursive estimation in a Riemannian manifold," *Entropy*, vol. 21, no. 10, p. 1021, 2019.

[17] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard, "Intrinsic Cramèr – Rao bounds for scatter and shape matrices estimation in CES distributions," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 262–266, 2019.

[18] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[19] D. Paindaveine, "A canonical definition of shape," *Statist. Probability Lett.*, vol. 78, no. 14, pp. 2240–2247, 2008.

[20] F. Bouchard, A. Breloy, G. Ginolhac, A. Renaux, and F. Pascal, "A Riemannian framework for low-rank structured elliptical models," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1185–1199, 2021.

[21] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014. [Online]. Available: http://www.manopt.org