Riemannian framework for robust covariance matrix estimation in spiked models

Florent Bouchard, Arnaud Breloy, Guillaume Ginolhac, Frédéric Pascal

RIEMANNIAN FRAMEWORK FOR ROBUST COVARIANCE MATRIX ESTIMATION IN SPIKED MODELS

Florent Bouchard[⋆] *Arnaud Breloy*[†] *Guillaume Ginolhac*[⋆] *Frédéric Pascal*[‡]

[⋆] LISTIC, Univ. Savoie Mont Blanc, Annecy, France
[†] LEME, Univ. Paris Nanterre, France
[‡] L2S, CentraleSupélec, Univ. Paris-Saclay, France

ABSTRACT

This paper aims at providing an original Riemannian geometry to derive robust covariance matrix estimators in spiked models (i.e. when the covariance matrix has a low-rank plus identity structure). The considered geometry is the one induced by the product of the Stiefel manifold and the manifold of Hermitian positive definite matrices, quotiented by the unitary group. One of the main contributions is to consider a Riemannian metric related to the Fisher information metric of elliptical distributions, leading to new representations for the tangent spaces and a new retraction. A new robust covariance matrix estimator is then obtained as the minimizer of Tyler's cost function, redefined directly on the set of low-rank plus identity matrices, and computed with the aforementioned tools. The main interest of this approach is that it appears well suited to the cases where the sample size is lower than the dimension, as illustrated by numerical experiments.

Index Terms— Covariance Matrices, Spiked Models, Robust Estimation, Riemannian Optimization

1. INTRODUCTION

Covariance matrix estimation is a crucial step in many machine learning and signal processing algorithms. This is still challenging at low/insufficient sample support ("small n large p" problems), where standard approches such as the traditional sample covariance matrix (SCM) fail to provide an accurate estimate.

In numerous applications, one can rightfully assume that the relevant information lies in a low dimensional subspace, leading to a low-rank structured covariance matrix (also referred to as spiked model [1]). This is the central idea of principal component analysis and related dimension reduction algorithms. In this context, a popular covariance matrix estimation process is to project the SCM onto the set of low-rank plus identity matrices (with pre-estimated rank k [2]), which has also a maximum likelihood (ML) interpretation for the Gaussian spiked model [3, 4]. This estimator offers a significant gain in terms of accuracy at low sample support, but it may still be sensitive to heavy-tailed distributed samples and outliers. This is due to the fact that the SCM is not robust to the aforementioned issues.

A general and elegant solution to this robustness issue is to turn to the framework of M-estimators [5]. To enjoy best of both worlds, one could transpose the previous approach and project an M-estimator (e.g., Tyler's estimator [6, 7]) on the set of interest. However, this plug-in 2-step procedure is not related to any ML formulation, and a more natural approach would be to directly derive an estimator as the minimizer of a robust cost function under a spiked structure constraint. This approach has been proposed in [8], where several majorization-minimization (MM) algorithms are proposed to minimize Tyler's cost function under various structure constraints (cf. section V.A. of [8] for the spiked one), that outperform the projection approach at low sample support.

However, several issues remain unanswered, notably for the insufficient sample support, because Tyler's estimator is not well defined for $n < p$, n being the sample size (number of observations) while p corresponds to the observation dimension. Furthermore, the tractability of constrained Tyler's estimators is also an opened question (cf. assumption 2 in [8]) in that case. Specifically for the spiked structure, the MM algorithm can present convergence issues in some practical case. To address these problems, we propose in this paper to leverage the Riemannian Geometry approach, in order to formulate an estimation procedure to optimize Tyler's cost function directly on the the manifold $\mathcal{H}_{p,k}^{+}$ ($p \times p$ Hermitian positive semi-definite matrices of rank k).

The manifold $\mathcal{H}_{p,k}^{+}$ has recently attracted much attention and different geometries have been proposed for it; see e.g. [9–13]. In this work, we consider the geometry induced by the quotient $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$, i.e. the product manifold of the complex Stiefel manifold $\mathrm{St}_{p,k}$ of orthogonal matrices in $\mathbb{C}^{p\times k}$ (with $p > k$) and the manifold $\mathcal{H}_k^{++}$ of Hermitian $k \times k$ positive definite (HPD) matrices, quotiented by the unitary group $\mathcal{U}_k$. The geometry of $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$ has already been studied in the context of low-rank matrices

This work was supported by MARGARITA ANR-17-ASTR-0015 and PHOENIX ANR-15-CE23-0012 grants of the French National Agency of Research.

in [9, 11]. However, our framework differs from these works: we propose a new Riemannian metric on $\mathrm{St}_{p,k} \times \mathcal{H}_k^{++}$ which has an interest when dealing with elliptical distributions. As a direct consequence, the representations of the tangent spaces of $(\mathrm{St}_{p,k} \times \mathcal{H}_k^{++})/\mathcal{U}_k$ and the Riemannian gradient used for optimization are different from the one previously considered. We also introduce a new retraction, which is obtained from a second order approximation of the geodesics.

2. RIEMANNIAN GEOMETRY AND OPTIMIZATION

We build a Riemannian framework to optimize cost functions defined on $\mathcal{H}_{p,k}^+$. To be able to perform Riemannian optimization of a criterion f on a manifold $\mathcal{M}$, we first need to characterize the *geometry* of $\mathcal{M}$ (tangent spaces and Riemannian metric). Then, given an iterate θ_i, the *Riemannian gradient* of f can be used to define a descent direction ξ_i. Given the step size t_i, the next iterate θ_{i+1} is obtained from $t_i\xi_i$ with a *retraction*, which is a mapping from the tangent spaces back onto $\mathcal{M}$. For a full review on this topic, see [14].

By exploiting a decomposition in $\overline{\mathcal{M}}_{p,k} = \mathrm{St}_{p,k} \times \mathcal{H}_k^{++}$ of every matrix in $\mathcal{H}_{p,k}^+$, we explain how $\mathcal{H}_{p,k}^+$ is isomorphic to the quotient manifold $\mathcal{M}_{p,k} = \overline{\mathcal{M}}_{p,k}/\mathcal{U}_k$. Thus, the geometry of $\mathcal{M}_{p,k}$ can be used in order to describe the one of $\mathcal{H}_{p,k}^+$, as proposed in [9, 11]. Then, the needed geometrical objects of $\mathcal{M}_{p,k}$ are studied. We introduce a new Riemannian metric of interest when dealing with elliptical distributions. We study the resulting representations of the tangent spaces of $\mathcal{M}_{p,k}$, which are subspaces of the tangent spaces of $\overline{\mathcal{M}}_{p,k}$; and the associated geodesics, which generalize the concept of straight lines on $\mathcal{M}_{p,k}$. Finally, the needed Riemannian optimization tools are provided: the Riemannian gradient resulting from our proposed geometry and a retraction, which is obtained by taking a second order approximation of the geodesics on $\mathcal{M}_{p,k}$. Due to space limitation, the proofs of the propositions will be presented in a forthcoming paper.

2.1. Quotient manifold $\mathcal{M}_{p,k}$

Every $\boldsymbol{H} \in \mathcal{H}_{p,k}^+$ can be decomposed as $\boldsymbol{H} = \boldsymbol{U\Sigma U}^H$, where $(\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k}$. Thus, $\overline{\varphi}(\boldsymbol{U}, \boldsymbol{\Sigma}) = \boldsymbol{U\Sigma U}^H$ defined from $\overline{\mathcal{M}}_{p,k}$ onto $\mathcal{H}_{p,k}^+$ is surjective. However, $\overline{\varphi}$ is not injective as the considered decomposition is not unique: given any $\boldsymbol{O} \in \mathcal{U}_k$, one has $\boldsymbol{H} = \overline{\varphi}(\boldsymbol{U}, \boldsymbol{\Sigma}) = \overline{\varphi}(\boldsymbol{UO}, \boldsymbol{O}^H\boldsymbol{\Sigma O})$.

As done in [9, 11], to account for the action of the unitary matrices, we define the quotient manifold

$$\mathcal{M}_{p,k} = \{\pi(\boldsymbol{U}, \boldsymbol{\Sigma}) : (\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k}\}, \tag{1}$$

where the equivalence class $\pi(\boldsymbol{U}, \boldsymbol{\Sigma})$ is

$$\pi(\boldsymbol{U}, \boldsymbol{\Sigma}) = \{(\boldsymbol{UO}, \boldsymbol{O}^H\boldsymbol{\Sigma O}) : \boldsymbol{O} \in \mathcal{U}_k\}. \tag{2}$$

As shown in [9, 11], it follows that the function φ on $\mathcal{M}_{p,k}$ induced by $\overline{\varphi}$ on $\overline{\mathcal{M}}_{p,k}$, *i.e.*, such that $\overline{\varphi} = \varphi \circ \pi$, is an isomorphism from $\mathcal{M}_{p,k}$ onto $\mathcal{H}_{p,k}^+$. Thus, the geometry of $\mathcal{M}_{p,k}$ can be exploited to treat problems defined on $\mathcal{H}_{p,k}^+$.

2.2. Riemannian geometry

To describe the geometry of the quotient $\mathcal{M}_{p,k}$, we exploit the submersion $\pi : \overline{\mathcal{M}}_{p,k} \to \mathcal{M}_{p,k}$ defined in (2). This allows to work with representatives of the geometrical objects of the quotient in $\overline{\mathcal{M}}_{p,k}$. In particular, $\theta \in \mathcal{M}_{p,k}$ is represented by any $(\boldsymbol{U}, \boldsymbol{\Sigma}) \in \overline{\mathcal{M}}_{p,k}$ such that $\theta = \pi(\boldsymbol{U}, \boldsymbol{\Sigma})$. The tangent space $T_\theta\mathcal{M}_{p,k}$ at $\theta = \pi(\boldsymbol{U}, \boldsymbol{\Sigma})$ in $\mathcal{M}_{p,k}$ is represented by a well chosen subspace of the tangent space $T_{(\boldsymbol{U},\boldsymbol{\Sigma})}\overline{\mathcal{M}}_{p,k}$ at $(\boldsymbol{U}, \boldsymbol{\Sigma})$ in $\overline{\mathcal{M}}_{p,k}$. Moreover, a Riemannian metric on $\mathcal{M}_{p,k}$ can be defined through a metric on $\overline{\mathcal{M}}_{p,k}$ that is invariant along the equivalence classes (2).

In the following, $\vartheta = (\boldsymbol{U}, \boldsymbol{\Sigma})$, $\xi = (\boldsymbol{\xi}_{\boldsymbol{U}}, \boldsymbol{\xi}_{\boldsymbol{\Sigma}})$ and $\eta = (\boldsymbol{\eta}_{\boldsymbol{U}}, \boldsymbol{\eta}_{\boldsymbol{\Sigma}})$. First recall that

$$T_\vartheta\overline{\mathcal{M}}_{p,k} = \{\xi \in \mathbb{C}^{p\times k} \times \mathcal{H}_k : \boldsymbol{U}^H\boldsymbol{\xi}_{\boldsymbol{U}} + \boldsymbol{\xi}_{\boldsymbol{U}}^H\boldsymbol{U} = \boldsymbol{0}\}.$$

We define the Riemannian metric $\langle\cdot,\cdot\rangle_\cdot$ on $\overline{\mathcal{M}}_{p,k}$ by

$$\begin{aligned}\langle\xi,\eta\rangle_\vartheta &= \mathfrak{Re}(\mathrm{tr}(\boldsymbol{\xi}_{\boldsymbol{U}}^H(\boldsymbol{I}_p - \frac{1}{2}\boldsymbol{UU}^H)\boldsymbol{\eta}_{\boldsymbol{U}})) \\ &+ \alpha\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}) + \beta\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}})\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Sigma}}),\end{aligned} \tag{3}$$

where $\alpha > 0$ and $\beta > -\frac{\alpha}{k}$. The part of the metric that concerns $\boldsymbol{U}$ is the so-called canonical metric on Stiefel [15], which is obtained by treating $\mathrm{St}_{p,k}$ as the quotient $\mathcal{U}_p/\mathcal{U}_{p-k}$. The one that concerns $\boldsymbol{\Sigma}$ corresponds to a class of affine invariant metrics on $\mathcal{H}_k^{++}$ that are of interest when dealing with elliptical distributions as they are related to the Fisher information metric [16][1]. It is readily checked that the metric (3) is invariant along the equivalence classes (2), *i.e.*, for all $\boldsymbol{O} \in \mathcal{U}_k$

$$\langle\xi,\eta\rangle_\vartheta = \langle\phi_{\boldsymbol{O}}(\xi), \phi_{\boldsymbol{O}}(\eta)\rangle_{\phi_{\boldsymbol{O}}(\vartheta)},$$

where $\phi_{\boldsymbol{O}}(Z) = (\boldsymbol{Z}_{\boldsymbol{U}}\boldsymbol{O}, \boldsymbol{O}^H\boldsymbol{Z}_{\boldsymbol{\Sigma}}\boldsymbol{O})$. Thus, metric (3) induces a Riemannian metric on the quotient $\mathcal{M}_{p,k}$.

The tangent space $T_\vartheta\overline{\mathcal{M}}_{p,k}$ can be decomposed into two complementary spaces: the vertical and horizontal spaces $\mathcal{V}_\vartheta$ and $\mathcal{H}_\vartheta$ [14]. The vertical space is the tangent space $T_\vartheta\pi(\vartheta)$ to the equivalence class (2), which is given by

$$\mathcal{V}_\vartheta = \{(\boldsymbol{U\Omega}, \boldsymbol{\Sigma\Omega} - \boldsymbol{\Omega\Sigma}) : \boldsymbol{\Omega}^H = -\boldsymbol{\Omega} \in \mathbb{C}^{k\times k}\},$$

as shown in [9, 11]. $\mathcal{H}_\vartheta$, which provides proper representatives for the elements of $T_{\pi(\vartheta)}\mathcal{M}_{p,k}$, is then defined as the orthogonal complement to $\mathcal{V}_\vartheta$ according to metric (3). The horizontal space $\mathcal{H}_\vartheta$ is given in proposition 1.

Proposition 1. *The horizontal space $\mathcal{H}_\vartheta$ at $\vartheta \in \overline{\mathcal{M}}_{p,k}$ is*

$$\mathcal{H}_\vartheta = \{\xi \in T_\vartheta\overline{\mathcal{M}}_{p,k} : \boldsymbol{U}^H\boldsymbol{\xi}_{\boldsymbol{U}} = 2\alpha(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Sigma}} - \boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})\}.$$

Finally, the geodesics on $\mathcal{M}_{p,k}$ associated with the metric (3) are given in proposition 2. This completes our study of the geometry of $\mathcal{M}_{p,k}$.

[1] For example, the Fisher information metric for the Gaussian distribution is obtained with $\alpha = 1$ and $\beta = 0$. Notice that these parameters are however simply treated as degrees of freedom in the present framework.

Proposition 2. *Representatives in $\overline{\mathcal{M}}_{p,k}$ of the geodesics on $\mathcal{M}_{p,k}$ associated with the metric* (3) *are, given $\theta = \pi(\vartheta) \in \mathcal{M}_{p,k}$ and $\xi \in \mathcal{H}_\vartheta$,*

$$\gamma(t) = \left([\boldsymbol{U}\,\boldsymbol{Q}] \exp t \begin{pmatrix} \boldsymbol{U}^H \boldsymbol{\xi}_{\boldsymbol{U}} & -\boldsymbol{R}^H \\ \boldsymbol{R} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{bmatrix}, \right.$$
$$\left. \boldsymbol{\Sigma} \exp(t \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) \right),$$

where $\boldsymbol{QR}$ is the QR decomposition of $(\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{\xi}_{\boldsymbol{U}}$.

2.3. Riemannian optimization

Let $\bar{f} : \overline{\mathcal{M}}_{p,k} \to \mathbb{R}$ be an objective function that induces a function f on the quotient $\mathcal{M}_{p,k}$, *i.e.*, such that, for all $\vartheta \in \overline{\mathcal{M}}_{p,k}$ and $\boldsymbol{O} \in \mathcal{U}_k$, $\bar{f}(\vartheta) = \bar{f}(\phi_{\boldsymbol{O}}(\vartheta))$, where $\phi_{\boldsymbol{O}}(\vartheta) = (\boldsymbol{U}\boldsymbol{O}, \boldsymbol{O}^H \boldsymbol{\Sigma} \boldsymbol{O})$. To perform Riemannian optimization, it remains to define the Riemannian gradient of f along with a retraction on $\mathcal{M}_{p,k}$. Proposition 3 provides a formula to compute the Riemannian gradient of f on $\mathcal{M}_{p,k}$ from the Euclidean gradient of $\bar{f}$ on $\overline{\mathcal{M}}_{p,k}$.

Proposition 3. *Given $\theta = \pi(\vartheta) \in \mathcal{M}_{p,k}$, the representative in $\mathcal{H}_\vartheta$ of the Riemannian gradient of f at θ is the Riemannian gradient of $\bar{f}$ at $\vartheta \in \overline{\mathcal{M}}_{p,k}$, which is*

$$\operatorname{grad} \bar{f}(\vartheta) = \Big(\nabla \bar{f}_{\boldsymbol{U}} - \boldsymbol{U} \nabla \bar{f}_{\boldsymbol{U}}^H \boldsymbol{U},$$
$$\frac{1}{\alpha} (\boldsymbol{\Sigma} \operatorname{herm}(\nabla \bar{f}_{\boldsymbol{\Sigma}}) \boldsymbol{\Sigma} - \lambda \boldsymbol{\Sigma}) + \frac{\lambda}{\alpha + k\beta} \boldsymbol{\Sigma} \Big),$$

where $\lambda = \frac{\operatorname{tr}(\operatorname{herm}(\nabla \bar{f}_{\boldsymbol{\Sigma}})\boldsymbol{\Sigma})}{p}$ and $\nabla \bar{f}(\vartheta) = (\nabla \bar{f}_{\boldsymbol{U}}, \nabla \bar{f}_{\boldsymbol{\Sigma}})$ is the Euclidean gradient of $\bar{f}$ in $\mathbb{C}^{p \times k} \times \mathbb{C}^{k \times k}$.

One can obtain a representative of a descent direction of f at $\theta = \pi(\vartheta) \in \mathcal{M}_{p,k}$ by selecting $\xi \in \mathcal{H}_\vartheta$ satisfying $\langle \operatorname{grad} \bar{f}(\vartheta), \xi \rangle_\vartheta < 0$. A new point on the manifold is then achieved by a retraction on $\mathcal{M}_{p,k}$. A natural choice is to take the Riemannian exponential map defined through the geodesics of proposition 2. However, it is also acceptable to choose a second order approximation of this exponential map, which, for $\theta = \pi(\vartheta) \in \mathcal{M}_{p,k}$ and $\xi \in \mathcal{H}_\vartheta$, is represented by

$$R_\vartheta(\xi) = \left([\boldsymbol{U}\,\boldsymbol{Q}] \operatorname{uf} \circ \Gamma \begin{pmatrix} \boldsymbol{U}^H \boldsymbol{\xi}_{\boldsymbol{U}} & -\boldsymbol{R}^H \\ \boldsymbol{R} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{bmatrix} \right.$$
$$\left. \boldsymbol{\Sigma} \Gamma(\boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}_{\boldsymbol{\Sigma}}) \right), \quad (4)$$

where uf returns the orthogonal factor of the polar decomposition and $\Gamma(\boldsymbol{X}) = \boldsymbol{I} + \boldsymbol{X} + \frac{1}{2}\boldsymbol{X}^2$ is a second order approximation of the matrix exponential. This approximation is essentially motivated for numerical stability reasons. The Riemannian gradient descent algorithm, which simply takes $-\operatorname{grad} \bar{f}(\vartheta_i)$ as a descent direction, is the

$$\vartheta_{i+1} = R_{\vartheta_i}(-t_i \operatorname{grad} \bar{f}(\vartheta_i)) \quad (5)$$

where t_i is the step size, which can for example be computed using a line search, or Armijo backtracking rule [14].

3. TYLER'S ESTIMATOR IN SPIKED MODELS : COST FUNCTION AND ALGORITHM

Given n observations $\{\boldsymbol{x}_i\}_{i=1}^n$, Tyler's M-estimator [6] is the minimizer of the cost function

$$f^{++}(\boldsymbol{C}) = n \log\det(\boldsymbol{C}) + p \sum_{i=1}^{n} \log(\boldsymbol{x}_i^H \boldsymbol{C}^{-1} \boldsymbol{x}_i). \quad (6)$$

on $\mathcal{H}_p^{++}$. For $n > p$ and when the samples span the whole space, this estimator exists and is unique (up to a scale factor) [7]. Moreover, it satisfies the following fixed-point equation:

$$\hat{\boldsymbol{R}}_{\mathrm{Ty}} = \frac{p}{n} \sum_{i=1}^{n} \frac{\mathbf{z}_i \mathbf{z}_i^H}{\mathbf{z}_i^H \hat{\boldsymbol{R}}_{\mathrm{Ty}}^{-1} \mathbf{z}_i} \triangleq \mathcal{H}_{\mathrm{Ty}}(\hat{\boldsymbol{R}}_{\mathrm{Ty}}), \quad (7)$$

and can be computed with the fixed-point algorithm

$$\boldsymbol{R}_{t+1} = \mathcal{H}_{\mathrm{Ty}}(\boldsymbol{R}_t). \quad (8)$$

This estimator presents good robustness properties [5–7], however it does not account for a spiked model, i.e. for the structure

$$\boldsymbol{R} = \boldsymbol{I}_p + \boldsymbol{H}, \ \boldsymbol{H} \in \mathcal{H}_{p,k}^+. \quad (9)$$

Imposing this structure in the estimation process can improve the estimation accuracy and/or deal with cases where $n < p$. For this purpose, [8, Sec. V.A.] proposed a MM algorithm to minimize (6) under the constraint (9). However, the tractability of this estimators is an open question for $n < p$ (cf. assumption 2 in [8]), and this algorithm can present a convergence issue in some practical case.

We aim at dealing with this issue by leveraging the presented Riemmanian framework. To that end, one needs a counterpart of Tyler's cost function that is properly defined directly on $\mathcal{M}_{p,k}$: consider the function $f : \mathcal{M}_{p,k} \to \mathbb{R}$ defined for all $\theta = \pi(\vartheta) \in \mathcal{M}_{p,k}$ by

$$f(\theta) = f^{++}(\boldsymbol{I}_p + \varphi(\theta)), \quad (10)$$

where $\varphi(\theta) = \bar{\varphi}(\vartheta) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$ as in Section 2.1 and $f^{++} : \mathcal{H}_p^{++} \to \mathbb{R}$ corresponds to the Tyler estimator's cost function (6).

As explained in Section 2, in order to minimize the cost function (10) within our Riemannian optimization framework, we simply need the Euclidean gradient of $\bar{f} = f \circ \pi$. It is provided in proposition 4.

Proposition 4. *The Euclidean gradient $\nabla \bar{f}(\vartheta)$ of $\bar{f}$ at ϑ is*

$$\nabla \bar{f}(\vartheta) = (2 \nabla f^{++}(\boldsymbol{I}_p + \bar{\varphi}(\vartheta)) \boldsymbol{U} \boldsymbol{\Sigma},$$
$$\boldsymbol{U}^H \nabla f^{++}(\boldsymbol{I}_p + \bar{\varphi}(\vartheta)) \boldsymbol{U}),$$

where $\nabla f^{++}(\boldsymbol{C})$ is the Euclidean gradient of (6)*, which is*

$$\nabla f^{++}(\boldsymbol{C}) = \boldsymbol{C}^{-1} \left[n\boldsymbol{C} - p \sum_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^H}{\boldsymbol{x}_i^H \boldsymbol{C}^{-1} \boldsymbol{x}_i} \right] \boldsymbol{C}^{-1}$$

Proof. One has

$$\begin{aligned} \mathrm{D}\,\bar{f}(\vartheta)[\xi] &= \mathrm{D}\,f^{++}(\boldsymbol{I}_p + \bar{\varphi}(\vartheta))[\mathrm{D}\,\bar{\varphi}(\vartheta)[\xi]] \\ &= \mathfrak{Re}(\mathrm{tr}(\nabla f^{++}(\boldsymbol{I}_p + \bar{\varphi}(\vartheta))^H\,\mathrm{D}\,\bar{\varphi}(\vartheta)[\xi])). \end{aligned}$$

The result is then obtained by using $\mathrm{D}\,\bar{\varphi}(\vartheta)[\xi] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{U}}^H + \boldsymbol{\xi}_{\boldsymbol{U}}\boldsymbol{\Sigma}\boldsymbol{U}^H + \boldsymbol{U}\boldsymbol{\xi}_{\boldsymbol{\Sigma}}\boldsymbol{U}^H$ and basic manipulations of the trace. □

Finally, applying the Riemmanian gradient descent algorithm (5) with this proposition allows to propose a procedure for robust covariance matrix estimation in the spiked model.

4. NUMERICAL EXPERIMENT

This section illustrates the theoretical results by performing covariance estimations of simulated data drawn from the multivariate Student t-distribution with $d = 3$ degree of freedom; see [5] for details. The covariance matrix of the simulated data follows the model $\boldsymbol{R} = \boldsymbol{I}_p + \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$, where $\boldsymbol{U}$ is a random matrix in $\mathrm{St}_{p,k}$ and $\boldsymbol{\Sigma}$ is a diagonal matrix whose elements are randomly drawn from the chi-squared distribution with expectation 1 and multiplied by 50. We generate sets $\{\boldsymbol{x}_i\}_{i=1}^n$, with $n \in \{10, 15, 17, 20, 30, 50\}$, from the multivariate Student t-distribution with covariance $\boldsymbol{R}$. For each value of n, 500 sets $\{\boldsymbol{x}_i\}_{i=1}^n$ are simulated and the aim is to estimate the structured covariance matrix $\boldsymbol{R}$ in each case.

Here are the considered estimators in the simulations: (a) projected sample covariance matrix $\hat{\boldsymbol{R}}_{\text{pSCM}} = \boldsymbol{I}_p + \hat{\boldsymbol{H}}_{\text{pSCM}}$, obtained by projecting $n^{-1}\sum_i \boldsymbol{x}_i\boldsymbol{x}_i^H$ on $\boldsymbol{I}_p + \mathcal{H}_{p,k}^+$ with [8, equation (53)]; (b) the MLE $\hat{\boldsymbol{R}}_{\text{T}_{LR}\text{-MM}} = \boldsymbol{I}_p + \hat{\boldsymbol{H}}_{\text{T}_{LR}\text{-MM}}$, solved with a block MM algorithm [8, algorithm 5]; (c) the proposed MLE $\hat{\boldsymbol{H}}_{\text{T}_{LR}\text{-RO}} = \boldsymbol{I}_p + \hat{\boldsymbol{H}}_{\text{T}_{LR}\text{-RO}}$, solved with a Riemannian gradient descent on $\mathcal{M}_{p,k}$ performed with the manopt toolbox [17]. We choose $\alpha = \frac{p+d}{p+d+1}$ and $\beta = \alpha - 1$ in the Riemannian metric (3)[2].

To measure the performance of the considered estimators, two different criteria are considered. The first one, which measures the error between the true covariance $\boldsymbol{R}$ and its estimate $\hat{\boldsymbol{R}}$, is the Riemannian distance on $\mathcal{H}_p^{++}$

$$\delta^2_{\mathcal{H}_p^{++}}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \alpha \sum_j \log(\lambda_j)^2 + \beta \left(\sum_j \log(\lambda_j) \right)^2, \quad (11)$$

where λ_j is the j^{th} eigenvalue of $\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}$. With $\alpha = \frac{p+d}{p+d+1}$ and $\beta = \alpha - 1$, it is the distance on $\mathcal{H}_p^{++}$ corresponding to the Fisher metric of the multivariate Student t-distribution [16]. The second one, which measures the error between the true subspace $\mathrm{span}(\boldsymbol{U})$ of $\boldsymbol{R} = \boldsymbol{I}_p + \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$ and its estimate $\mathrm{span}(\hat{\boldsymbol{U}})$, is the Riemannian distance on the Grassmann manifold [18]

$$\delta^2_{\mathcal{G}_{p,k}}(\mathrm{span}(\boldsymbol{U}), \mathrm{span}(\hat{\boldsymbol{U}})) = \|\boldsymbol{\Theta}\|_F^2, \quad (12)$$

where $\boldsymbol{\Theta}$ is the diagonal matrix obtained by the singular value decomposition $\boldsymbol{U}^H\hat{\boldsymbol{U}} = \boldsymbol{O}\cos(\boldsymbol{\Theta})\tilde{\boldsymbol{O}}^H$.

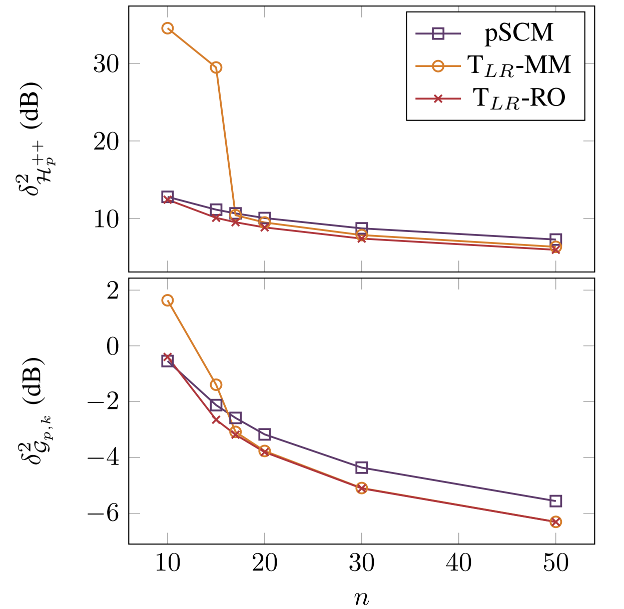

Fig. 1. Mean of performance measures (11) (top) and (12) (bottom) over 500 simulated sets $\{\boldsymbol{x}_i\}$ ($p = 16$ and $k = 4$) with respect to the number of samples n for the three considered estimators.

In figure 1, one can observe that for $n > p$, the performance of T_{LR}-MM and T_{LR}-RO are similar for both criteria and better than the one obtained with projected SCM. However, for $n < p$, T_{LR}-MM fails to converge toward a satisfying solution and is outperformed by the other methods. In contrast, our proposed algorithm, T_{LR}-RO, which minimizes the same criterion as T_{LR}-MM, remains competitive as compared to the projected SCM.

5. CONCLUSIONS AND PERSPECTIVES

This work focused on the problem of robust covariance matrix estimation in spiked models through the prism of Riemannian geometry. First, a Riemannian geometry for $\mathcal{H}_{p,k}^+$ involving the Fisher information metric of elliptical distributions has been proposed, leading to new representations for the tangent spaces, and a new retraction. Then, a new robust estimator has been formulated as the minimizer of Tyler's cost function, redefined directly on the set of low-rank plus identity matrices. A corresponding Riemannian gradient descent algorithm (using the aforementioned tools) has been derived in order to compute this estimator. Interestingly, the proposed approach is able to deal with insufficient sample support settings ($n < p$), as illustrated by numerical experiments and opens the way for more general estimation problems.

[2] These α and β correspond to the parameters of the Fisher information metric for the multivariate Student t-distribution in $\mathcal{H}_p^{++}$ (cf. [16] for details). However, notice that these parameters do not influence the obtained results in this context. This will be shown in a forthcoming paper.

6. REFERENCES

[1] I.M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” *The Annals of statistics*, vol. 29, no. 2, pp. 295–327, 2001.

[2] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[3] M.E. Tipping and C.M. Bishop, “Probabilistic principal component analysis,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[4] B. Kang, V. Monga, and M. Rangaswamy, “Rank-constrained maximum likelihood estimation of structured covariance matrices,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 501–515, January 2014.

[5] 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.

[6] D. E. Tyler, “A distribution-free m-estimator of multivariate scatter,” *The Annals of Statistics*, p. 234251, 1987.

[7] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, “Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, Jan 2008.

[8] 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.

[9] S. Bonnabel and R. Sepulchre, “Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1055–1070, 2009.

[10] B. Vandereycken and S. Vandewalle, “A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2553–2579, 2010.

[11] G. Meyer, S. Bonnabel, and R. Sepulchre, “Regression on fixed-rank positive semidefinite matrices: a Riemannian approach,” *Journal of Machine Learning Research*, vol. 12, pp. 593–625, 2011.

[12] B. Vandereycken, P.-A. Absil, and S. Vandewalle, “A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank,” *IMA Journal of Numerical Analysis*, vol. 33, no. 2, pp. 481–514, 2012.

[13] E. Massart and P.-A. Absil, “Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices,” *Technical Report UCL-INMA-2018.06*, 2018.

[14] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, USA, 2008.

[15] A. Edelman, T.A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[16] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard, “Intrinsic cramér–rao bounds for scatter and shape matrices estimation in ces distributions,” *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 262–266, 2018.

[17] 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.

[18] P.-A. Absil, R. Mahony, and R. Sepulchre, “Riemannian geometry of Grassmann manifolds with a view on algorithmic computation,” *Acta Applicandae Mathematica*, vol. 80, no. 2, pp. 199–220, 2004.