

Exact maximum likelihood estimates for sirv covariance matrix: Existence and algorithm analysis

Yacine Chitour, Frédéric Pascal

▶ **To cite this version:**

Yacine Chitour, Frédéric Pascal. Exact maximum likelihood estimates for sirv covariance matrix: Existence and algorithm analysis. IEEE Transactions on Signal Processing, 2008, 56 (10), pp.4563-4573. 10.1109/TSP.2008.927464 . hal-00353594

HAL Id: hal-00353594

https://centralesupelec.hal.science/hal-00353594

Submitted on 4 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exact Maximum Likelihood Estimates for SIRV Covariance Matrix: Existence and Algorithm Analysis

Yacine Chitour, Frédéric Pascal

***Abstract*—In this paper, we investigate the existence and the algorithm analysis of an adaptive scheme [1], [2] which has been introduced for covariance structure matrix estimation in the context of adaptive radar detection under non-Gaussian noise. This latter has been modeled by Spherically Invariant Random Vector (SIRV), which is the product $\mathbf{c}$ of the square root of a positive unknown random variable τ and an independent Gaussian vector $\mathbf{x}$, $\mathbf{c} = \sqrt{\tau}\,\mathbf{x}$. A similar line of work was undertaken in the context of compound Gaussian noise [3], [4] and the present paper extends the previous results in the case of SIRV modeled noise. More precisely, the Fixed Point estimate to be studied verifies a non-linear algebraic equation $(E)\ \ x = f(x)$. The aim of this paper is twofold. First, we prove that (E) admits a unique solution $\bar{x}$ and secondly, we show that the corresponding iterative algorithm $x_{n+1} = f(x_n)$ converges to $\bar{x}$ for every admissible initial condition.**

***Index Terms*—Maximum Likelihood estimate, SIRV model, Fixed Point estimate, iterative algorithm convergence, adaptive detection.**

I. Introduction

THE basic problem of detecting a complex signal embedded in an additive Gaussian noise has been extensively studied these last decades. In these contexts, adaptive detection schemes required an estimate of the noise covariance matrix generally obtained from signal-free data traditionally called secondary data or reference data. The resulting adaptive detectors, as those proposed by [5] and [6], are all based on the Gaussian assumption for which the Maximum Likelihood (ML) estimate of the covariance matrix is given by the sample covariance matrix. However, these detectors may exhibit poor performance when the additive noise is non-Gaussian [7].

When this additive noise is non-Gaussian, one of the most general and elegant non-Gaussian noise model is provided by the so-called *Spherically Invariant Random Vectors* (SIRV). These processes encompass a large number of non-Gaussian distributions thanks to the random variable τ which has an unknown Probability Density Function (PDF). Detectors resulting of such a model require an estimate of the covariance matrix of the Gaussian component. In this context, ML estimates based on secondary data have been introduced in [8], [9], together with a numerical procedure supposed to obtain them. However, as noticed in [9] p.1852, "*existence of the ML estimate and convergence of iteration [...] is still an open problem*".

To the best of our knowledge, the proofs of existence, uniqueness of the ML estimate and convergence of the algorithm proposed in [1] have not yet been established. The main purpose of this paper is to fill these gaps.

A first work [3] began to answer this problem but only, in the case where the texture τ has been assumed to be deterministic. The present paper provides results on the exact ML estimate in the general SIRV case, i.e., when the random variable τ has an unknown PDF. Therefore, the present paper is indeed a continuation of [3], as regards the issues which are addressed. Even though the present work uses some of the results from [3], one must stress that the two papers are technically disjoint, in the sense that every detailed argument provided here does not appear in [3]. We however postpone to Remark IV.1 below the description of the differences between the two papers because it requires several definitions introduced later on.

The main results of this paper is to provide the existence and the uniqueness of a ML estimate, because in the case of SIRV data, the estimate is defined as the solution of an implicit equation. Several estimation procedures lead to similar problems of solving implicit equations. This is the case for the expectation-maximisation (EM) algorithm useful for ML estimation with missing data [10], developed in [11]. This can be explained by the fact that missing data leads to Likelihood function conditioned to the parameter of interest. Moreover, elliptically contoured distributions [12] are very closed to symmetrically distributions which encompass SIRV distributions. Finally, this work is also tightly related to the **M**-estimates [13], [14], due to the particular expression of SIRV distribution.

The paper is organized as follows. In Section II, we provide the statistical framework. Section III presents the main results of the paper in the complex case and in the (more specific) real case. Section IV gives proofs outline for results in the real case and we gather in Section V complete arguments.

Acknowlegdments. The authors would like to thank F. Gini

Y. Chitour is with the Laboratoire des Signaux et Systèmes, Supelec, Plateau du Moulon, 3 rue Joliot Curie, F-91192 Gif-sur-Yvette Cedex, France (e-mail: yacine.chitour@lss.supelec.fr)

F. Pascal is with SONDRA, Supelec, Plateau du Moulon, 3 rue Joliot Curie, F-91192 Gif-sur-Yvette Cedex, France (e-mail: frederic.pascal@supelec.fr).

for useful clarifications and S. Gaubert for bringing to our attention the reference [15].

II. Problem formulation

A SIRV $\mathbf{c}$ [16], [17] is defined as the product

$$\mathbf{c} = \sqrt{\tau}\,\mathbf{x}\,, \tag{1}$$

where the positive random variable τ is called the *texture*, having unknown Probability Density Function (PDF) $p_\tau(\tau)$, and $\mathbf{x}$ is an m-dimensional zero-mean complex Gaussian vector with covariance matrix $\mathbf{M} = E[\mathbf{x}\mathbf{x}^H]$ usually normalized according to $\mathrm{Tr}(\mathbf{M}) = m$, cf. [9]. Such a normalization is referred to as the $\mathbf{M}$-normalization. The symbol H denotes the conjugate transpose operator, $E[.]$ stands for the expectation of a random variable and $\mathrm{Tr}(.)$ stands for the trace operator.

Here, in this general model, we follow the well-known SIRV modeling where the texture is considered to be a random variable with unknown PDF (see [1], [2], [18], [19]). Generally, the covariance matrix $\mathbf{M}$ is not known and an estimate $\widehat{\mathbf{M}}$ is required for the Likelihood Ratio (LR) computation. Classically, such an estimate is obtained from Maximum Likelihood (ML) theory, well-known for its good statistical properties. In this problem, estimation of $\mathbf{M}$ has to respect the $\mathbf{M}$-normalization, $\mathrm{Tr}(\widehat{\mathbf{M}}) = m$. This estimate $\widehat{\mathbf{M}}$ will be built using K independent realizations of $\mathbf{c}$ denoted $\mathbf{c}_k = \sqrt{\tau_k}\,\mathbf{x}_k$ for $k = 1, \ldots, K$.

1) Notations: In this paragraph, we introduce the main notations of the paper for the real case. Notations already defined in the complex case are translated in the real one. Moreover, real results will be valid for every positive integer m. We use $\mathbb{C}$ (respectively $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_+^*$) to denote the set of complex (resp. real, non negative real and positive real) numbers, while for any integer m, $\mathbb{C}^m$ (resp. $\mathbb{R}^m$) represents the set of m-vectors with complex (resp. real) elements. For vectors of $\mathbb{R}^m$, the norm used is the Euclidean one. Throughout the paper, we will use several basic results on square matrices, especially regarding spectral properties of real symmetric and orthogonal matrices. We refer to [20] and/or [21] for such standard results.

We use $\mathcal{M}_m(\mathbb{R})$ to denote the set of $m \times m$ real matrices, $\mathcal{SO}(m)$ to denote the set of $m \times m$ orthogonal matrices and $\mathbf{M}^\top$, the transpose of $\mathbf{M}$. We denote the identity matrix of $\mathcal{M}_m(\mathbb{R})$ by $\mathbf{I}_m$.

We next define and list the several sets of matrices used in the sequel:

- $\mathcal{D}$, the subset of $\mathcal{M}_m(\mathbb{R})$ defined by the symmetric positive definite matrices;
- $\overline{\mathcal{D}}$, the closure of $\mathcal{D}$ in $\mathcal{M}_m(\mathbb{R})$, i.e., the subset of $\mathcal{M}_m(\mathbb{R})$ defined by the symmetric non negative matrices.

For $\mathbf{M} \in \mathcal{D}$, we use $\mathcal{L}_{\mathbf{M}}$ to denote the open-half line spanned by $\mathbf{M}$ in the cone $\mathcal{D}$, i.e., the set of points $\lambda\,\mathbf{M}$, with $\lambda > 0$. Let us recall that the order associated with the cone structure of $\mathcal{D}$ is called the Loewner order for symmetric matrices of $\mathcal{M}_m(\mathbb{R})$ and is defined as follows. Let $\mathbf{A}, \mathbf{B}$ be two symmetric $m \times m$ real matrices. Then $\mathbf{A} \leq \mathbf{B}$ (resp. $\mathbf{A} < \mathbf{B}$) means that the quadratic form defined by $\mathbf{B} - \mathbf{A}$ is non negative (resp. positive definite), i.e., for every non zero $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x}^\top (\mathbf{A} - \mathbf{B})\,\mathbf{x} \leq 0$, (resp. < 0). Using that order, one has $\mathbf{M} \in \mathcal{D}$ (resp. $\in \overline{\mathcal{D}}$) if and and only if $\mathbf{M} > \mathbf{0}$ (resp. $\mathbf{M} \geq \mathbf{0}$).

For $\mathbf{M} \in \mathcal{D}$, we define $\mathcal{D}(\mathbf{M})$ as the subset of $\mathcal{D}$ given by

$$\mathcal{D}(\mathbf{M}) = \{\mathbf{P} \in \mathcal{D} |\ \mathbf{P} \leq \mathbf{M}\}\,. \tag{2}$$

In the sequel, if $f : \mathcal{D} \to \mathcal{D}$, we use f^n, $n \geq 1$, to denote the n^{th} iterate of f i.e., $f^n := f \circ ... \circ f$, where f is repeated n times. We also adopt the standard convention that $f^0 := Id_{\mathcal{D}}$.

We also need some basic notations. The integer m is always positive and the integer K is always larger than m. Let $\mathcal{G}$ be the set of $m \times m$ Hermitian positive definite matrices. For any $m \times m$ matrix $\mathbf{M}$, its determinant is denoted by $|\mathbf{M}|$. Given a mapping $f : \mathcal{G} \to \mathcal{G}$, the iterative algorithm associated to f is the procedure which associates to any $\mathbf{M} \in \mathcal{G}$ the sequence $(\mathbf{M}_n)_{n \geq 0}$, where $\mathbf{M}_0 = \mathbf{M}$ and $\mathbf{M}_{n+1} = f(\mathbf{M}_n)$.

$[\![1, K]\!]$ denotes the set of the K first integers $\{1, \ldots K\}$. A K-tuple $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_K)$ made of vectors of $\mathbb{C}^m$ verifies Hypothesis $(H1)$ if

$$(H1) : \begin{cases} \text{For any } m \text{ two by two distinct indices} \\ k(1) < ... < k(m) \text{ chosen in } [\![1, K]\!], \text{ the vectors} \\ \mathbf{c}_{k(1)}, \ldots, \mathbf{c}_{k(m)} \text{ are linearly independent.} \end{cases} \tag{3}$$

Given a C^1 PDF p_τ (i.e., a continuously derivable Probability Density Function), let us consider the function P defined on $\mathbb{R}_+^*$ by $P(\tau) = \tau \dfrac{p'_\tau(\tau)}{p_\tau(\tau)}$. Then, the PDF p_τ verifies Hypothesis $(P1)$ if

$$(P1) \quad P \text{ is a strictly decreasing function.} \tag{4}$$

Notice that Hypothesis $(P1)$ is verified for every texture PDF having closed-form expression.

Let us recall that the SIRV PDF expression [16], [17] is

$$p(\mathbf{c}) = \frac{1}{\pi^m\,|\mathbf{M}|} \int_0^{+\infty} \frac{1}{\tau^m} \exp\left(-\frac{\mathbf{c}^H\,\mathbf{M}^{-1}\,\mathbf{c}}{\tau}\right) p_\tau(\tau)\,\mathrm{d}\tau\ .$$

To obtain the ML estimate of $\mathbf{M}$, with no proofs of existence and uniqueness, Gini *et al.* derived in [9] an Approximate Maximum Likelihood (AML) estimate $\widehat{\mathbf{M}}$ as follows.

Let us first consider the Likelihood function given by

$$p_{\mathbf{C}}(\mathbf{c}_1, \ldots, \mathbf{c}_K; \mathbf{M}) = \prod_{k=1}^{K} \int_0^{+\infty} \frac{1}{(\pi\,\tau_k)^m\,|\mathbf{M}|} \times \exp\left(-\frac{\mathbf{c}_k^H\,\mathbf{M}^{-1}\,\mathbf{c}_k}{\tau_k}\right) p_\tau(\tau_k)\,\mathrm{d}\tau_k\,. \tag{5}$$

As a preliminary step, let us mention that, in the deterministic case, Conte and Ricci (cf. [22]) showed that $p_{\mathbf{C}}$ was finitely upper bounded over $\mathcal{G}$. However, it was not proved if that upper bound was reached or not. Anyhow, in order to

maximize the above function, one needs to look for critical points of $p_{\mathbf{C}}$, i.e., points $\mathbf{M} \in \mathcal{D}$ which annihilate the gradient of $p_{\mathbf{C}}$.

After some computations, Gini *et al.* obtained initially the following equation for $\mathbf{M}_{MV}$

$$\mathbf{M}_{MV} = \frac{m}{K}\sum_{k=1}^{K} c_m(\mathbf{c}_k^H\,\mathbf{M}_{MV}^{-1}\,\mathbf{c}_k)\mathbf{c}_k\,\mathbf{c}_k^H, \tag{6}$$

where the real-valued function c_m is defined on $\mathbb{R}_+^*$ by

$$c_m(q) := \frac{h_{m+1}(q)}{h_m(q)}, \tag{7}$$

and the real-valued function h_m is defined on $\mathbb{R}_+^*$ by

$$h_m(q) := \int_0^{+\infty} \tau^{-m}\exp(-q/\tau)p_\tau(\tau)d\tau. \tag{8}$$

Let us define the matrix-valued function $f_{\mathbf{C}}(\mathbf{M})$ as the right-hand side of Eq. (6), i.e., for $\mathbf{M}$ symmetric positive definite

$$f_{\mathbf{C}}(\mathbf{M}) := \frac{1}{K}\sum_{k=1}^{K} c_m(\mathbf{c}_k^H\,\mathbf{M}^{-1}\,\mathbf{c}_k)\mathbf{c}_k\,\mathbf{c}_k^H. \tag{9}$$

Unfortunately, the fixed points of $f_{\mathbf{C}}(\mathbf{M})$ (if it exists) do not naturally verify the $\mathbf{M}$-normalization in general. This is a fatal drawback for the fixed point of $f_{\mathbf{C}}(\mathbf{M})$ to be an estimate of $\mathbf{M}$ since it results a biased estimate. This is why Gini *et al.* chose finally $\mathbf{M}_{MV}$ as a fixed point of the following mapping

$$g_{\mathbf{C}}(\mathbf{M}) := \frac{m}{\mathrm{Tr}(f_{\mathbf{C}}(\mathbf{M}))}f_{\mathbf{C}}(\mathbf{M}), \tag{10}$$

i.e., the $\mathbf{M}_{MV}$ solution of $\mathbf{M} = g_{\mathbf{C}}(\mathbf{M})$. Numerically, that estimate was shown to exist as the limit of the iterative scheme.

To derive the exact ML estimate of the covariance matrix, in the case of SIRV modeling, we assume that the PDF p_τ verifies Hypothesis $(P1)$. We then prove that, for every K-tuple $\mathbf{C}$ verifying Hypothesis $(H1)$, the mapping $g_{\mathbf{C}}$ admits a unique fixed point denoted $\mathbf{M}_g(\mathbf{C})$ and the iterative algorithm associated to $g_{\mathbf{C}}$ converges to $\mathbf{M}_g(\mathbf{C})$ for every initial condition. In the course of the argument, we will need to establish first that these results hold for the mapping $f_{\mathbf{C}}$.

III. Statement of the main results

A. The complex case

We first provide additional notations. Let m and K be positive integers such that $m < K$. We use $\mathcal{M}_m(\mathbb{C})$ to denote the set of $m \times m$ complex matrices. For $\mathbf{M} \in \mathcal{M}_m(\mathbb{C})$ the Frobenius norm of $\mathbf{M}$ is defined as $\mathrm{Tr}\left(\mathbf{M}^H\mathbf{M}\right)^{1/2}$ and we use $||\mathbf{M}||$ to denote it. Moreover, from the statistical independence hypothesis of the K complex m-vectors $\mathbf{x}_k$, it is natural to assume the following Hypothesis $(H1')$.
Let us set $\mathbf{x}_k = \mathbf{x}_k^{(1)} + j\mathbf{x}_k^{(2)}$,

$$(H1'): \begin{cases} \text{Any } 2m \text{ distinct vectors taken in} \\ \left\{\begin{pmatrix}\mathbf{x}_1^{(1)}\\ \mathbf{x}_1^{(2)}\end{pmatrix},\ldots,\begin{pmatrix}\mathbf{x}_K^{(1)}\\ \mathbf{x}_K^{(2)}\end{pmatrix},\begin{pmatrix}-\mathbf{x}_1^{(2)}\\ \mathbf{x}_1^{(1)}\end{pmatrix},\ldots,\begin{pmatrix}-\mathbf{x}_K^{(2)}\\ \mathbf{x}_K^{(1)}\end{pmatrix}\right\} \\ \text{are linearly independent.} \end{cases} \tag{11}$$

Theorem III.1 (Existence and algorithm analysis)
Let $\mathbf{C}$ be a K-tuple verifying Hypothesis $(H1)$.

(i) The mapping $g_{\mathbf{C}}$ admits a unique fixed point $\mathbf{M}_g(\mathbf{C}) \in \mathcal{G}$ with $Tr(\mathbf{M}_g(\mathbf{C})) = m$;

(ii) Let $(S)_{dis}$ be the discrete dynamical system defined on $\mathcal{G}$ by

$$(S)_{dis}\ : \mathbf{M}_{n+1} = g_{\mathbf{C}}(\mathbf{M}_n). \tag{12}$$

Then, for every initial condition $\mathbf{M}_0 \in \mathcal{G}$, the resulting sequence $(\mathbf{M}_n)_{n\geq 0}$ converges to $\mathbf{M}_g(\mathbf{C})$.

The same problem and the same result can be formulated with real numbers instead of complex numbers and symmetric matrices instead of Hermitian matrices, while hypothesis $(H1)$ becomes hypothesis $(H2)$ stated below (just before Remark III.1). The proof of Theorem III.1 breaks up into several steps. The way to derive Theorem III.1 from the corresponding real results has been shown in [3]. Then, the rest of the paper is devoted to the study of the real case.

B. The real case

1) Preliminaries: A K-tuple $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_K)$ of vectors of $\mathbb{R}^m$ verifies Hypothesis $(H2)$ if

$$(H2): \begin{cases} \text{For any } m \text{ two by two distinct indices} \\ k(1) < \ldots < k(m) \text{ chosen in } [\![1, K]\!], \text{ the vectors} \\ \mathbf{c}_{k(1)}, \ldots, \mathbf{c}_{k(m)} \text{ are linearly independent.} \end{cases} \tag{13}$$

Let us already emphasize that hypothesis $(H2)$ is the key assumption for getting all our subsequent results. Hypothesis $(H2)$ has the following trivial but fundamental consequence that we state as a remark.

Remark III.1
For every n vectors $\mathbf{x}_{k(1)}, \ldots, \mathbf{x}_{k(n)}$ (resp. $\mathbf{c}_{k(1)}, \ldots, \mathbf{c}_{k(n)}$) with $1 \leq n \leq m$, $1 \leq k \leq K$, the vector space generated by $\mathbf{x}_{k(1)}, \ldots, \mathbf{x}_{k(n)}$ (resp. $\mathbf{c}_{k(1)}, \ldots, \mathbf{c}_{k(n)}$) has dimension n.

We also need the next definition.

Definition III.1
Let us consider a continuous mapping $f : \mathcal{D} \to \mathcal{D}$. Then, f is said to be

- **strictly increasing** *if, for every $\mathbf{M} < \mathbf{Q}$ in $\mathcal{D}$, one has $f(\mathbf{M}) < f(\mathbf{Q})$;*
- **eventually completely increasing of order p** *if, there exists a positive integer p such that, for every $\mathbf{M} \leq \mathbf{Q}$ in $\mathcal{D}$ with $\mathbf{M} \neq \mathbf{Q}$, one has $f^p(\mathbf{M}) < f^p(\mathbf{Q})$;*
- **subhomogeneous** *if, for every $\mathbf{M}$ in $\mathcal{D}$ and $\lambda \in (0,1)$, one has $f(\lambda\mathbf{M}) > \lambda f(\mathbf{M})$.*

Given a K-tuple $\mathbf{C}$, define the map $F_{\mathbf{C}}$ as

$$F_{\mathbf{C}}: \begin{cases} \mathcal{G} \longrightarrow \mathbb{R}_+^* \\ \mathbf{M} \longrightarrow F_{\mathbf{C}}(\mathbf{M}) = \dfrac{1}{|\mathbf{M}|^K}\displaystyle\prod_{k=1}^{K} h_m(\mathbf{c}_k^H\,\mathbf{M}^{-1}\,\mathbf{c}_k). \end{cases} \tag{14}$$

Then, from (5) and (8), the two functions $F_{\mathbf{C}}$ and $f_{\mathbf{C}}$ are related by the following relation, which is obtained after an easy computation. For every $\mathbf{M} \in \mathcal{D}$, let $\nabla F_{\mathbf{C}}(\mathbf{M})$ be the

gradient of $F_{\mathbf{C}}$ at $\mathbf{M} \in \mathcal{D}$ [21], i.e., the unique symmetric matrix verifying, for every matrix $\mathbf{M} \in \mathcal{D}$,

$$\nabla F_{\mathbf{C}}(\mathbf{M}) = K\, F_{\mathbf{C}}(\mathbf{M})\, \mathbf{M}^{-1}\, \left(f_{\mathbf{C}}(\mathbf{M}) - \mathbf{M}\right)\, \mathbf{M}^{-1}. \quad (15)$$

Clearly $\mathbf{M}$ is a fixed point of $f_{\mathbf{C}}$ if and only if $\mathbf{M}$ is a critical point of the vector field defined by $\nabla F_{\mathbf{C}}$ on $\mathcal{D}$.

2) Statements of the results in the real case:

Theorem III.2
Assume that the PDF p_τ verifies Hypothesis $(P1)$. Then, given a K-tuple $\mathbf{C}$ verifying Hypothesis $(H2)$,

- (a) *the mapping $g_{\mathbf{C}}$ admits a unique fixed point $\mathbf{M}_g(\mathbf{C})$;*
- (b) *Let $(S)_{dis}$ be the discrete dynamical system defined on $\mathcal{D}$ by*

$$(S)_{dis}\ : \mathbf{M}_{n+1} = g_{\mathbf{C}}(\mathbf{M}_n). \quad (16)$$

Then, for every initial condition $\mathbf{M}_0 \in \mathcal{D}$, the resulting sequence $(\mathbf{M}_n)_{n\geq 0}$ converges to $\mathbf{M}_g(\mathbf{C})$.

In order to prove the above theorem, one has first to prove a similar (and richer, cf. Item (d) below on gradient systems) theorem when $g_{\mathbf{C}}$ is replaced by $f_{\mathbf{C}}$. More precisely, we have the following.

Theorem III.3
Assume that the PDF p_τ verifies Hypothesis $(P1)$. Then, given a K-tuple $\mathbf{C}$ verifying Hypothesis $(H2)$,

- (a) *the mapping $f_{\mathbf{C}}$ admits a unique fixed point $\mathbf{M}_f(\mathbf{C})$;*
- (b) *the map $F_{\mathbf{C}}$ reaches its maximum only at $\mathbf{M}_f(\mathbf{C})$ and $Hess(\mathbf{M}_f(\mathbf{C}))$, the Hessian of $F_{\mathbf{C}}$ at $\mathbf{M}_f(\mathbf{C})$, is negative definite;*
- (c) *Let $(S)_{dis}$ be the discrete dynamical system defined on $\mathcal{D}$ by*

$$(S)_{dis}\ : \mathbf{M}_{n+1} = f_{\mathbf{C}}(\mathbf{M}_n). \quad (17)$$

Then, for every initial condition $\mathbf{M}_0 \in \mathcal{D}$, the resulting sequence $(\mathbf{M}_n)_{n\geq 0}$ converges to $\mathbf{M}_f(\mathbf{C})$;

- (d) *Let $(S)_{cont}$ be the continuous dynamical system defined on $\mathcal{D}$ by*

$$(S)_{cont}\ : \dot{\mathbf{M}} = \quad \nabla F_{\mathbf{C}}(\mathbf{M}). \quad (18)$$

Then, for every initial condition $\mathbf{M}(0) = \mathbf{M}_0 \in \mathcal{D}$, the resulting trajectory $\mathbf{M}(t)$, $t \geq 0$, converges, when t tends to $+\infty$, to $\mathbf{M}_f(\mathbf{C})$.

Both proofs are outlined in the next section. Proof of Theorem III.2 is detailed in section V while the complete proof of Theorem III.3 is postponed in Appendix A.

IV. Proofs outline

First of all, we will rely on several results of [3] and more precisely on properties established for the maps

$$B_{\mathbf{C}}: \quad \begin{cases} \mathcal{D} \longrightarrow \mathbb{R}_+^* \\ \mathbf{M} \longrightarrow \dfrac{1}{|\mathbf{M}|^K} \displaystyle\prod_{k=1}^{K} \frac{1}{\left(\mathbf{x}_k^\top\, \mathbf{M}^{-1}\, \mathbf{x}_k\right)^m}, \end{cases} \quad (19)$$

and

$$b_{\mathbf{C}}: \quad \begin{cases} \mathcal{D} \longrightarrow \mathcal{D} \\ \mathbf{M} \longrightarrow \dfrac{m}{K} \displaystyle\sum_{k=1}^{K} \frac{\mathbf{x}_k\, \mathbf{x}_k^\top}{\mathbf{x}_k^\top\, \mathbf{M}^{-1}\, \mathbf{x}_k}, \end{cases} \quad (20)$$

where $\mathbf{C}$ is a K-tuple of vectors of $\mathbb{R}^m$ verifying Hypothesis $(H2)$. Similarly to Eq. 15, one has

$$\nabla B_{\mathbf{C}}(\mathbf{M}) = K\, B_{\mathbf{C}}(\mathbf{M})\, \mathbf{M}^{-1}\, \left(b_{\mathbf{C}}(\mathbf{M}) - \mathbf{M}\right)\, \mathbf{M}^{-1}. \quad (21)$$

Let us recall that, in [3], it was proved that $B_{\mathbf{C}}$ is homogeneous of degree zero, uniformly bounded over $\mathcal{D}$ and it can be continuously extended to $\overline{D}/\{0\}$ by zero on $\overline{D}/(\mathcal{D}\cup\{0\})$. Moreover, $B_{\mathbf{C}}$ reaches its maximum over a unique half-line $L_{\mathbf{P}_B}$, with $\mathrm{Tr}(\mathbf{P}_B) = m$. Finally, it was proved that $b_{\mathbf{C}}$ is eventually completely increasing of order m.

Remark IV.1
We can explain the differences between the present paper and [3]. In term of problem formulation, [3] assumes that τ is deterministic while in this paper, τ is a random variable with unknown PDF. The main difference stems from the fact that $g_{\mathbf{C}}$ is not homogeneous of degree one (as $b_{\mathbf{C}}$ is) and that creates other difficulties with respect to [3]. The only way we were able to find to overcome these obstacles consisted in relating $g_{\mathbf{C}}$ with a family of maps $f_{\mathbf{C},\mu}$ (see below in (25)), which turn out to be only sub-homogeneous. As explained later, one must first study the $f_{\mathbf{C},\mu}$'s to deduce information for $g_{\mathbf{C}}$: each $f_{\mathbf{C},\mu}$ shares properties with $f_{\mathbf{C},1} = f_{\mathbf{C}}$ and this is the reason why we start with the study of $f_{\mathbf{C}}$. It takes all the present work to follow the steps of the above described program and most of the related arguments are new with respect to [3].

A. Proof outline of Theorem III.3

We start by proving some facts on the functions h_m and c_m.

Proposition IV.1
With the notations above, we have, for every $q > 0$ and positive integer m,

- (i) *$h'_m = -h_{m+1}$, which implies that h_m is strictly decreasing;*
- (ii) *$\lim_{q\to 0} q^m h_m(q) = \lim_{q\to\infty} q^m h_m(q) = 0$. As a consequence, the function $H_m(q) := q^m h_m(q)$ is uniformly bounded over R_+^*;*
- (iii) *$c_m(q) < c_{m+1}(q)$, which implies that c_m is strictly decreasing;*
- (iv) *$g_m(q) := q c_m(q)$ has a monotony inverse with respect to that of $P(q) = q\dfrac{p'(q)}{p(q)}$. As a consequence, for all known PDF's, g_m is strictly increasing. In that case, define*

$$g_- := g_m(0), \quad g_+ := \lim_{q\to\infty} g_m(q). \quad (22)$$

Clearly, $0 \leq g_-$ and either $g_- < g_+$ or $g_+ = \infty$.

Taking into account the previous proposition, we show item (a) of Theorem III.3 as the consequence of the next proposition.

Proposition IV.2
Let $\mathbf{C}$ be a K-tuple of vectors of $\mathbb{R}^m$ verifying Hypothesis $(H2)$. One has

(A) *$F_{\mathbf{C}}$ reaches its maximum at some point $\mathbf{M}_f(\mathbf{C}) \in \mathcal{D}$;*
(B) *Assume that g_m is strictly increasing and $\mathbf{M} \in \mathcal{D}$ is a critical point of $F_{\mathbf{C}}$ (or equivalently a fixed point of $f_{\mathbf{C}}$). Then, $Hess(\mathbf{M})$, the Hessian of $F_{\mathbf{C}}$ at $\mathbf{M}$, is negative definite, implying that $\mathbf{M}$ a strict local maximum of $F_{\mathbf{C}}$.*

Then, we turn to an argument for item (c) of Theorem III.3. It first requires to study the mapping $f_{\mathbf{C}}$.

Proposition IV.3
Let $\mathbf{C}$ be a K-tuple of vectors of $\mathbb{R}^m$ verifying Hypothesis $(H2)$. Then, $f_{\mathbf{C}}$ verifies the following properties.

(i) *$f_{\mathbf{C}}$ is strictly increasing and eventually completely increasing of order m;*
(ii) *If g_m is strictly increasing, then, $f_{\mathbf{C}}$ is subhomogeneous (cf. [15]), i.e., for every $\lambda \in (0,1)$ and $\mathbf{M} \in \mathcal{D}$,*

$$f_{\mathbf{C}}(\lambda \mathbf{M}) > \lambda f_{\mathbf{C}}(\mathbf{M}). \tag{23}$$

As a consequence, for every $\lambda > 1$ and $\mathbf{M} \in \mathcal{D}$,

$$f_{\mathbf{C}}(\lambda \mathbf{M}) < \lambda f_{\mathbf{C}}(\mathbf{M}). \tag{24}$$

Finally, we combine the above proposition with the uniqueness of the fixed point to derive Item (c). Item (b) of Theorem III.3 follows at once, as well as item (d), since $(S)_{cont}$, defined in (18), is a gradient system (cf. [23] for references on gradient systems).

It is important to notice that several statements of the theorems could be derived by using the results of [15] but it requires to study the linear map $Df_{\mathbf{C}}(\cdot)$ (defined as the differential of $f_{\mathbf{C}}$), which leads to more involved computations than those of the present paper. In addition, the convergence results of Corollary 2.2 in [15] are weaker than those obtained here but one can adapt the proof of Corollary 2.2 to the present situation. We chose to provide a more direct argument in order to be self-contained. As a last remark, it is not clear to us how to obtain our results from those of [24] since we are not able to check the following condition

$$\begin{cases} \text{there exists } 0 < \gamma < 1, \text{ such that } f(t\mathbf{M}) \geq t^\gamma f(\mathbf{M}), \\ \text{for all } 0 < t < 1, \quad \mathbf{M} \in \mathcal{D} \end{cases}$$

for $f_{\mathbf{C}}$.

B. Proof outline of Theorem III.2

Even though the statements of the theorems are similar, dealing with $g_{\mathbf{C}}$ presents new difficulties with respect to $f_{\mathbf{C}}$. First of all, $g_{\mathbf{C}}$ is not related to the gradient of a real-valued function defined on $\mathcal{D}$, and thus it is difficult to relate the existence of a fixed point of $g_{\mathbf{C}}$ to that of a critical point of some real-valued function. More importantly, $g_{\mathbf{C}}$ does not have useful properties shared by $f_{\mathbf{C}}$ such as monotonicity or subhomogeneity. As a consequence, the study of $g_{\mathbf{C}}$ requires other ingredients.

The starting point of the analysis consists of the next remark. A fixed point $\mathbf{M}_g$ of $g_{\mathbf{C}}$ (if any) verifies the equation $f_{\mathbf{C}}(\mathbf{M}_g) = \bar{\mu}\mathbf{M}_g$ for some $\bar{\mu} > 0$ i.e., $\mathbf{M}_g$ is the fixed point of the mapping

$$f_{\mathbf{C},\mu} := \frac{f_{\mathbf{C}}}{\mu}, \tag{25}$$

with $\mu = \bar{\mu}$. This suggests to consider the whole family of mappings $f_{\mathbf{C},\mu}$, for an arbitrary $\mu > 0$, as defined in (25). Indeed, these mappings retain all the useful properties of $f_{\mathbf{C}}$: if one considers the real-valued map $F_{\mathbf{C},\mu}$ defined on $\mathcal{D}$ by

$$F_{\mathbf{C},\mu}(\mathbf{M}) := \frac{1}{|\mathbf{M}|^K} \prod_{k=1}^{K} h_{m,\mu}(\mathbf{c}_k^H \mathbf{M}^{-1} \mathbf{c}_k), \tag{26}$$

where $h_{m,\mu}(q) := [h_m(q)]^{1/\mu}$, then

$$\nabla F_{\mathbf{C},\mu}(\mathbf{M}) = N\, F_{\mathbf{C},\mu}(\mathbf{M})\, \mathbf{M}^{-1} \left(f_{\mathbf{C},\mu}(\mathbf{M}) - \mathbf{M}\right) \mathbf{M}^{-1}, \tag{27}$$

and, in particular, every fixed point of $f_{\mathbf{C},\mu}$ is a critical point of $F_{\mathbf{C},\mu}$ (if any). Moreover, it is trivial to see that Item (B) of Proposition IV.2 and Proposition IV.3 hold true when $f_{\mathbf{C}}$ is replaced by $f_{\mathbf{C},\mu}$. However, for a general pdf, $F_{\mathbf{C},\mu}$ does not admit critical points for every $\mu > 0$ but we can show that Item (A) of Proposition IV.2 holds true for $\mu \in I_{g_m}$, where $I_{g_m} \subset \mathbb{R}_+^*$ is an open (in $\mathbb{R}_+^*$) interval containing 1. As a consequence, $f_{\mathbf{C},\mu}$ admits a unique fixed point $\mathbf{M}_f(\mathbf{C},\mu)$ for $\mu \in I_{g_m}$. More precisely, we show the following proposition.

Proposition IV.4
Let $\mathbf{C}$ be a K-tuple of vectors of $\mathbb{R}^m$ verifying Hypothesis $(H2)$ and assume that g_m is strictly increasing. Then, Theorem III.3 holds true when $f_{\mathbf{C}}$ and $F_{\mathbf{C}}$ are respectively replaced by $f_{\mathbf{C},\mu}$ and $F_{\mathbf{C},\mu}$ for and only for $\mu \in I_{g_m}$, where $I_{g_m} := \left(\frac{g_-}{m}, \frac{g_+}{m}\right)$. For $\mu \in I_{g_m}$, we use $\mathbf{M}_f(\mathbf{C},\mu)$ to denote the unique fixed point (resp. global maximum) of $f_{\mathbf{C},\mu}$ (resp. $F_{\mathbf{C},\mu}$). In addition, the following mappings, defined on I_{g_m},

$$\mu \mapsto \mathbf{M}_f(\mathbf{C},\mu), \quad \mu \mapsto \mu\mathbf{M}_f(\mathbf{C},\mu) \text{ are strictly decreasing}, \tag{28}$$

and

$$\lim_{\mu \to \frac{g_-}{m}} \|\mathbf{M}_f(\mathbf{C},\mu)\| = \infty, \quad \lim_{\mu \to \frac{g_+}{m}} \mathbf{M}_f(\mathbf{C},\mu) = 0. \tag{29}$$

It follows that the real-valued function $\mu \mapsto \mathrm{Tr}(\mathbf{M}_f(\mathbf{C},\mu))$ is a bijection from I_{g_m} to $\mathbb{R}_+^*$. Since $g_{\mathbf{C}}$ takes values in the matrices of $\mathcal{D}$ with trace equal to m, we easily deduce that $g_{\mathbf{C}}$ admits a unique fixed point $\mathbf{M}_g(\mathbf{C}) \in \mathcal{D}$. As regards the convergence of the iterative scheme, it can be directly deduced from Corollary 2.5 of [15] since the second point of the orbit starting at any $\mathbf{M} \in \mathcal{D}$ is already normalized.

V. Proof of Theorem III.2

Before starting the argument let us add the notation

$$c_{m,\mu}(q) := c_m(q)/\mu, \quad g_{m,\mu}(q) := q c_{m,\mu}(q),$$

for $q > 0$ and $\mu > 0$. In the next proposition, we gather facts on $F_{\mathbf{C},\mu}$, $f_{\mathbf{C},\mu}$, $h_{m,\mu}$ and $c_{m,\mu}$.

Proposition V.1
With the notations above, we have, for every $q > 0$, $\mu > 0$ and $\mathbf{M} \in \mathcal{D}$,

- $(a1)$ $c_{m,\mu}(q) = -\dfrac{h'_{m,\mu}(q)}{h_{m,\mu}(q)}$; *$c_{m,\mu}$ is strictly decreasing and $g_{m,\mu}$ is strictly increasing;*
- $(a2)$ $\dfrac{\nabla F_{\mathbf{C},\mu}(\mathbf{M})}{NF_{\mathbf{C},\mu}(\mathbf{M})} = \mathbf{M}^{-1}\left(f_{\mathbf{C},\mu}(\mathbf{M}) - \mathbf{M}\right)\mathbf{M}^{-1}$. *In particular, $\mathbf{M}$ is a critical point of $F_{\mathbf{C},\mu}$ if and only if $\mathbf{M}$ is a fixed point of $f_{\mathbf{C},\mu}$;*
- $(a3)$ *$f_{\mathbf{C},\mu}$ is strictly increasing and eventually completely increasing of order m.*
- $(a4)$ *If $f_{\mathbf{C},\mu}$ admits a fixed point, then it is unique (denoted $\mathbf{M}_f(\mathbf{C},\mu)$) and the iterative scheme associated to $f_{\mathbf{C},\mu}$ converges to $\mathbf{M}_f(\mathbf{C},\mu)$ for every initial condition. Moreover, $\mathbf{M}_f(\mathbf{C},\mu)$ is a strict maximum for $f_{\mathbf{C},\mu}$ with negative definite Hessian at $\mathbf{M}_f(\mathbf{C},\mu)$.*

One can derive the above proposition by easily adapting the arguments provided for Theorem III.3, i.e., for the case $\mu = 1$.

We next establish a technical lemma, which is crucial for the rest of the paper.

Lemma V.1
Assume that $f_{\mathbf{C},\bar{\mu}}$ admits a fixed point $\mathbf{M}_f(\mathbf{C},\bar{\mu})$ for some $\bar{\mu} > 0$. Then, there exists a open neighborhood V of $\bar{\mu}$ in $\mathbb{R}_+^*$ such that, for every $\mu \in V$, $f_{\mathbf{C},\mu}$ admits a fixed point $\mathbf{M}_f(\mathbf{C},\mu)$.

Proof of Lemma V.1: This is a consequence of the implicit function theorem applied to the mapping

$$\Phi(\mu,\mathbf{M}) = f_{\mathbf{C}}(\mathbf{M}) - \mu\mathbf{M},$$

defined on $\mathbb{R}_+^* \times \mathcal{D}$. We only have to show that, at a point $(\mu,\mathbf{M})$ where $\Phi(\mu,\mathbf{M}) = 0$, the differential of Φ with respect to $\mathbf{M}$ is invertible as an endomorphism of the vector space of the symmetric matrices, i.e.,

$$\mathbf{Q} \mapsto D_{\mathbf{M}}\Phi(\mu,\mathbf{M}).\mathbf{Q}$$

is invertible. For every symmetric matrix $\mathbf{Q}$, a simple computation gives

$$D_{\mathbf{M}}\Phi(\mu,\mathbf{M}).\mathbf{Q} = -\frac{1}{K}\sum_{k=1}^{K} c'_m(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k) \times (\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\mathbf{c}_k)\mathbf{c}_k\mathbf{c}_k^T - \mu\mathbf{Q}.$$

We now proceed as in the proof of Item (B) of Proposition IV.2. Set $\mathbf{R} := \mathbf{M}^{-1/2}\mathbf{Q}\mathbf{M}^{-1/2}$ and $\mathbf{d}_k := \mathbf{M}^{-1/2}\mathbf{c}_k$ for $k = 1,\cdots,K$. Using the fact that $\Phi(\mu,\mathbf{M}) = 0$, one has

$$\mu\,\mathrm{Tr}(\mathbf{R}^2) = \frac{1}{K}\sum_{k=1}^{K} c_m(\|\mathbf{d}_k\|^2)\|\mathbf{R}\mathbf{d}_k\|^2.$$

Then,

$$\langle \mathbf{R}, \mathbf{M}^{-1/2}D_{\mathbf{M}}\Phi(\mu,\mathbf{M}).\mathbf{Q}\mathbf{M}^{-1/2}\rangle \leq -\frac{1}{K}\sum_{k\in I_{\mathbf{R}}} \|\mathbf{R}\mathbf{d}_k\|^2 r_k g'_m(\|\mathbf{d}_k\|^2), \quad (30)$$

where $I_{\mathbf{R}}$ is the set of indices k for which $\mathbf{R}\mathbf{d}_k \neq 0$ and $r_k = \left(\dfrac{\mathbf{d}_k^T\mathbf{R}\mathbf{d}_k}{\|\mathbf{d}_k\|\|\mathbf{R}\mathbf{d}_k\|}\right)^2$. The last inequality shows that $D_{\mathbf{M}}\Phi(\mu,\mathbf{M})$ is clearly injective and the lemma is proved. ∎

We next study the subset $I \in \mathbb{R}_+$ defined as the set of $\mu's$ such that $f_{\mathbf{C},\mu}$ admits a fixed point. By Item $(a4)$ of Proposition V.1, if $\mu \in I$, then $f_{\mathbf{C},\mu}$ admits a unique fixed point $\mathbf{M}_f(\mathbf{C},\mu)$. Since $1 \in I$, the latter is non empty and, thanks to the lemma, it is open (in $\mathbb{R}_+$).

We now prove that I is an interval of $\mathbb{R}_+^*$.

Lemma V.2
With the notations above, the C^1 maps defined on I $\phi_1 : \mu \mapsto \mathbf{M}_f(\mathbf{C},\mu)$ and $\phi_2 : \mu \mapsto \mu\mathbf{M}_f(\mathbf{C},\mu)$ are decreasing and then I is convex, which implies that I is an interval of $\mathbb{R}_+^*$.

Proof of Lemma V.2: Consider $\mu_1 < \mu_2$ in I. Then

$$f_{\mathbf{C},\mu_1}(\mathbf{M}_f(\mathbf{C},\mu_2)) = \frac{\mu_2}{\mu_1}\mathbf{M}_f(\mathbf{C},\mu_2) > \mathbf{M}_f(\mathbf{C},\mu_2).$$

Then the orbit of $\mathbf{M}_f(\mathbf{C},\mu_2)$ associated to $f_{\mathbf{C},\mu_1}$ defines a strictly increasing sequence in $\mathcal{D}$, which converges to $\mathbf{M}_f(\mathbf{C},\mu_1)$, according to Item $(a4)$ of Proposition V.1. We deduce that $\mathbf{M}_f(\mathbf{C},\mu_2) < \mathbf{M}_f(\mathbf{C},\mu_1)$. Moreover, $f_{\mathbf{C},\mu_1}(\mathbf{M}_f(\mathbf{C},\mu_2))$ is also strictly smaller than $\mathbf{M}_f(\mathbf{C},\mu_1)$ since it is also a point of the orbit. We deduce at once that ϕ_1 and ϕ_2 are strictly decreasing.

Let us pick $\mu \in (\mu_1,\mu_2)$ and show that $\mu \in I$, that $\mathbf{M}_f(\mathbf{C},\mu) \in \mathcal{D}$ is well-defined. For that, consider the orbit $((\mathbf{M}_n)_{n\geq 0})$ of $\mathbf{M}_f(\mathbf{C},\mu_2)$ associated to $f_{\mathbf{C},\mu}$. By a simple inductive argument, one shows that $\mathbf{M}_n < \mathbf{M}_f(\mathbf{C},\mu_1)$ and

$$\mathbf{M}_0 := \mathbf{M}_f(\mathbf{C},\mu_2) < \frac{\mu_2}{\mu}\mathbf{M}_f(\mathbf{C},\mu_2) = \mathbf{M}_1.$$

Therefore, $((\mathbf{M}_n)_{n\geq 0})$ defines an upper-bounded increasing sequence, which implies that it is converging to a fixed point of $f_{\mathbf{C},\mu}$. ∎

It immediately follows that I is an interval (μ_-,μ_+) with $\mu_- \geq 0$ and μ_+ possibly infinite.

Lemma V.3
With the notations above, $\mu_- = \dfrac{g_m(0)}{m}$ and

$$\lim_{\mu\to\mu_-^+} \|\mathbf{M}_f(\mathbf{C},\mu)\| = \infty. \quad (31)$$

Proof of Lemma V.3: We first prove the last part of the lemma. Reasoning by contradiction, we would have a finite limit, as μ tends to μ_-, for $\mathbf{M}_f(\mathbf{C},\mu)$ since ϕ_1 is monotone. We could therefore extend by continuity ϕ_1 at $\mu = \mu_-$, i.e., define $\mathbf{M}_f(\mathbf{C},\mu_-) \in \mathcal{D}$. Note, in that case, that $\mu_- > 0$. Applying Lemma V.1 at μ_- allows one to extend I on the left of μ_- (recall that, if $f_{\mathbf{C},\mu}$ admits a fixed point, it is unique) and then we would contradict the definition of μ_- as the infimum of the μ's for which $\mathbf{M}_f(\mathbf{C},\mu)$ is well defined in $\mathcal{D}$.

Next we prove that $\mu_- \geq g_m(0)/m$. From the definition of $\mathbf{M}_f(\mathbf{C},\mu)$, one deduces that

$$\mu\mathbf{I}_m = \frac{1}{K}\sum_{k=1}^{K} c_m(\|\mathbf{d}_k\|^2)\mathbf{d}_k\mathbf{d}_k^T = \frac{1}{K}\sum_{k=1}^{K} g_m(\|\mathbf{d}_k\|^2)\frac{\mathbf{d}_k\mathbf{d}_k^T}{\|\mathbf{d}_k\|^2}, \quad (32)$$

where

$$\mathbf{d}_k(\mu) := \mathbf{M}_f(\mathbf{C},\mu)^{-1/2}\mathbf{c}_k. \tag{33}$$

Taking the trace yields

$$\mu\, m = \frac{1}{K}\sum_{k=1}^{K} g_m(\|\mathbf{d}_k(\mu)\|^2). \tag{34}$$

Since g_m is strictly increasing, it implies the required inequality.

To prove equality, we first assume that $c_m(0)$ is finite. Then $g_m(0) = 0$ and $g'_m(0) = c_m(0) = 0$. By looking carefully at the argument of Lemma V.1, one can see from (30) that if $\mu_- > 0$, then the differential $D_{\mathbf{M}}\Phi(\mu, \mathbf{M}_f(\mathbf{C},\mu))$ is uniformly invertible in an open neighborhood of μ_-. Then $\mathbf{M}_f(\mathbf{C},\cdot)$ can be extended at $\mu = \mu_-$. Since it is not possible, we conclude that $\mu_- = 0$.

Assume now that $\lim_{q\to 0} c_m(q) = \infty$. We first establish the following lemma.

Lemma V.4
For every $k = 1, \cdots, K$,

$$\lim_{\mu\to\mu_-} \|\mathbf{d}_k(\mu)\| = 0, \tag{35}$$

Proof of Lemma V.4: We may assume that $\mu_- > 0$, otherwise the conclusion follows. Finally, notice that Lemma V.3 follows readily from Lemma V.3 by letting μ tend to μ_- in (34).

The argument goes by contradiction. Since $\mu \to \mathbf{M}_f(\mathbf{C},\mu)^{-1}$ is increasing, we can assume that there exists $\delta > 0$ such that, for every $\mu \in I$, $\|\mathbf{d}_1(\mu)\| \geq \delta$. Let $J := \{1, \cdots, l\}$, $1 \leq l \leq K$ of all indices k such that $\lim_{\mu\to\mu_-} \|\mathbf{d}_k\| > 0$, i.e., there exists $\delta > 0$ such that for every $k \in J$,

$$\|\mathbf{d}_k(\mu)\| \geq \delta. \tag{36}$$

We use J_K to denote the set of the others indices $l+1 \leq t \leq K$ for which $\lim_{\mu\to\mu_-} \|\mathbf{d}_t\| = 0$.

Up to a subsequence of $\mu's$, we may assume that $\mathbf{d}_k(\mu)$ converges to a non zero vector $\mathbf{d}_k$, for $k \in J$. For $k \in J$ and $\mu \in I$, one has

$$\mu\|\mathbf{c}_k\|^2 = \mu\mathbf{d}_k(\mu)^T\mathbf{M}_f(\mathbf{C},\mu)\mathbf{d}_k(\mu) = \frac{1}{K}\sum_{s=1}^{K} c_m(\|\mathbf{d}_s\|^2)(\mathbf{d}_k(\mu)^T\mathbf{c}_s)^2. \tag{37}$$

Since $\lim_{q\to 0} c_m(q) = \infty$, one deduces, by letting μ tend to $\mu_- > 0$, that $\mathbf{d}_k^T\mathbf{c}_t = 0$ for $k \in J$ and $t \in J_K$. It implies that the vector space generated by the $\mathbf{c}_t$'s, $t \in J_K$, is of dimension at most $m - l$. Thanks to Remark III.1, the cardinality of J_k must be lower than $m-l$, and thus $K-l \leq m-l$ i.e., $K \leq m$, which is a contradiction. ∎

It remains to study the behavior of ϕ_1 as μ tends to μ_+. It is described in the next lemma.

Lemma V.5
With the notations above, one has $\mu_+ = \dfrac{g_+}{m}$ and

$$\lim_{\mu\to\mu_+^-} \mathbf{M}_f(\mathbf{C},\mu) = 0. \tag{38}$$

Proof of Lemma V.5: Assume first that $\lim_{q\to\infty} g_m(q) = \infty$ and that μ_+ is finite. Then, according to (34), $\|\mathbf{d}_k\|$ is uniformly bounded above over $k = 1, \cdots, K$ and as μ tends to μ_+. Applying Hadamard's inequality (cf. [20]), one deduces that $\left|\mathbf{M}_f(\mathbf{C},\mu)^{-1}\right|$ is also uniformly bounded above as μ tends to μ_+ and then that $|\mathbf{M}_f(\mathbf{C},\mu)|$ is uniformly bounded from below, as μ tends to μ_+. Since ϕ_1 is decreasing, we get that $\mathbf{M}_f(\mathbf{C},\mu_+)$ is invertible and belongs to $\mathcal{D}$. We would be therefore able to extend ϕ_1 on the right of μ_+ and reach a contradiction. Therefore, $\mu_+ = \infty$. Since ϕ_2 is decreasing, one has $\mu\mathbf{M}_f(\mathbf{C},\mu) \leq \mathbf{M}_f(\mathbf{C},1)$ if $\mu \geq 1$. By letting μ tend to ∞, we have (38).

Assume now that $\lim_{q\to\infty} g_m(q) = l$. Since $l > 0$ and according to Item (ii) of Proposition IV.1, we necessarily have $l > m$. Coming back to the definition of g_m, one gets that there exists a positive constant C_{max} such that

$$q^l h_m(q) \sim_{q\to\infty} C_{max}.$$

For $q > 0$, define

$$H^l(q) := q^l h_m(q).$$

We have $(H^l)'(q) = q^{l-1}h_m(q)(l - g_m(q)) > 0$. Then, H^l is strictly increasing and bounded over $\mathbb{R}^+$.

Let us first prove that $\mu_+ \geq \dfrac{l}{m}$. Use $\mathbf{M}(1)$ to denote $\mathbf{M}_f(\mathbf{C},1)$. For $1 < \mu < \dfrac{l}{m}$, let us majorize $F_{\mathbf{C},\mu}(\mathbf{M})$ for $\mathbf{M} \in \mathcal{D}(\mathbf{M}(1))$. One has

$$F_{\mathbf{C},\mu}(\mathbf{M}) = B_{\mathbf{C}}(\mathbf{M})\left(\prod_{k=1}^{K} H^l(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k)\right)^{\frac{1}{\mu}} \times \prod_{k=1}^{K}\left(\frac{1}{\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k}\right)^{(\frac{l}{\mu}-m)}. \tag{39}$$

The right-hand side of the above inequality is the product of three terms. The first one, $B_{\mathbf{C}}(\mathbf{M})$, is bounded over $\mathcal{D}$. The second one is also bounded over $\mathcal{D}$. For the third term, notice that, for $\mathbf{M} \in \mathcal{D}(\mathbf{M}(1))$ and $k = 1, \cdots, K$,

$$\frac{1}{\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k} \leq \frac{\mathbf{c}_k^T\mathbf{M}\mathbf{c}_k}{\|\mathbf{c}_k\|^4} \leq \frac{\mathbf{c}_k^T\mathbf{M}(1)\mathbf{c}_k}{\|\mathbf{c}_k\|^4}.$$

We immediately deduce that $F_{\mathbf{C},\mu}$ is bounded over $\mathcal{D}(\mathbf{M}(1))$. Since $B_{\mathbf{C}}(\mathbf{M})$ tends to zero as $\mathbf{M}$ gets close to the set of non invertible matrices, we deduce that $F_{\mathbf{C},\mu}$ reaches its maximum at some point $\mathbf{M}_* \in \mathcal{D}$ such that $\mathbf{M}_* \leq \mathbf{M}(1)$. We will prove, by contradiction, that $\mathbf{M}_* < \mathbf{M}(1)$. Otherwise, there exists $\bar{\mathbf{x}} \in \mathbb{R}^m$, of unit norm such that

$$\mathbf{M}_*\bar{\mathbf{x}} = \mathbf{M}(1)\bar{\mathbf{x}}. \tag{40}$$

Let us evaluate the gradient of $F_{\mathbf{C},\mu}$ at $\mathbf{M}_*$ in the direction $\mathbf{Q} := -\mathbf{x}\mathbf{x}^T$ for $\mathbf{x} \in \mathbb{R}^m$ to be fixed later. One has

$$\nabla F_{\mathbf{C},\mu}(\mathbf{M}_*)(-\mathbf{x}\mathbf{x}^T) = -F_{\mathbf{C},\mu}(\mathbf{M}_*)N \times \mathrm{Tr}(\mathbf{M}_*^{-1}(f_{\mathbf{C},\mu}(\mathbf{M}_*) - \mathbf{M}_*)\mathbf{M}_*^{-1}\mathbf{x}\mathbf{x}^T).$$

Choose $\mathbf{x} = \mathbf{M}_*\bar{\mathbf{x}}$. Then,

$$\nabla F_{\mathbf{C},\mu}(\mathbf{M}_*)(-\mathbf{x}\mathbf{x}^T) = -F_{\mathbf{C},\mu}(\mathbf{M}_*)N\bar{\mathbf{x}}^T\Big(f_{\mathbf{C},\mu}(\mathbf{M}_*) - \mathbf{M}_*\Big)\bar{\mathbf{x}}.$$

Using the fact that $\mathbf{M}_* \leq \mathbf{M}(1)$ and (40), we get

$$\nabla F_{\mathbf{C},\mu}(\mathbf{M}_*)(-\mathbf{x}\mathbf{x}^T) \geq -F_{\mathbf{C},\mu}(\mathbf{M}_*)N(\frac{1}{\mu}-1)\bar{\mathbf{x}}^T\mathbf{M}(1)\bar{\mathbf{x}} > 0.$$

Proceeding as above in every direction $\mathbf{x}$ verifying (40), we can prove that $F_{\mathbf{C},\mu}$ takes larger values than $F_{\mathbf{C},\mu}(\mathbf{M}_*)$ in the interior of $\mathcal{D}(\mathbf{M}(1))$, contradicting the fact that $F_{\mathbf{C},\mu}$ reaches its maximum at $\mathbf{M}_*$.

Since $\mathbf{M}_*$ belongs to the interior of $\mathcal{D}(\mathbf{M}(1))$, we get that $\mathbf{M}_* = \mathbf{M}_f(\mathbf{C},\mu)$. Then $\mu_+ \geq \dfrac{l}{m}$. Moreover, applying the previous reasoning to $A\mathbf{M}(1)$, $A > 1$, instead of $\mathbf{M}(1)$ shows that $F_{\mathbf{C},\mu}$ reaches its maximum over $\mathcal{D}(A\mathbf{M}(1))$ at $\mathbf{M}_f(\mathbf{C},\mu)$ and then, by letting A tend to ∞, we obtain that $F_{\mathbf{C},\mu}$ admits a unique maximum over $\mathcal{D}$ at $\mathbf{M}_f(\mathbf{C},\mu)$.

We finally show that

$$\lim_{\mu \to l/m^-} \mathbf{M}_f(\mathbf{C},\mu) = 0,$$

which will imply Item (ii) of the Lemma and (38) in that case. We reason by contradiction and assume that $\mu_+ > l/m$.

Recall that $\mathbf{P}_B$ is the unique fixed point of $B_{\mathbf{C}}$ of trace equal to m. For $\mu \in (l/m, \mu_+)$, we will estimate $F_{\mathbf{C},\mu}(\mathbf{M}_f(\mathbf{C},\mu))$ in two different ways. First, we evaluate $F_{\mathbf{C},\mu}(\rho\mathbf{P})$ as ρ tends to zero. Applying (39) and taking into account the fact that $B_{\mathbf{C}}$ is homogeneous of degree zero, we get

$$F_{\mathbf{C},\mu}(\rho\mathbf{P}) = B_{\mathbf{C}}(\mathbf{P})\rho^{l/\mu - m}A_1(\mu)A_2(\rho,\mu),$$

with

$$A_1(\rho) := \Big(\prod_{k=1}^K H^l(\mathbf{c}_k^T\mathbf{P}^{-1}\mathbf{c}_k/\rho)\Big)^{1/\mu},$$
$$A_2(\rho,\mu) := \prod_{k=1}^K \Big(\frac{1}{\mathbf{c}_k^T\mathbf{P}^{-1}\mathbf{c}_k}\Big)^{(l/\mu - m)}.$$

Clearly, $A_1(\rho)$ tends to $C_{max}^{K/\mu}$ as ρ tends to zero and $A_2(\rho,\mu)$ tends to one as (μ,ρ) tends to $(l/m^+, 0)$. Let us now choose ρ as

$$\rho_a(\mu) := \exp(\frac{a}{l/\mu - m}),$$

where $a > 0$ is arbitrary. Let μ tends to l/m^+. Then $\rho_a(\mu)$ tends to zero and we have

$$\lim_{\mu \to (l/m)^+} F_{\mathbf{C},\mu}(\rho_a(\mu)\mathbf{P}) = B_{\mathbf{C}}(\mathbf{P})\exp(a)C_{max}^{Km/l}.$$

Since $F_{\mathbf{C},\mu}$ admits a global maximum over $\mathcal{D}$, we have

$$\limsup_{\mu \to (l/m)^+} F_{\mathbf{C},\mu}(\mathbf{M}_f(\mathbf{C},\mu)) \geq B_{\mathbf{C}}(\mathbf{P})\exp(a)C_{max}^{Km/l},$$

and, since $a > 0$ is arbitrary, we finally get

$$\limsup_{\mu \to l/m} F_{\mathbf{C},\mu}(\mathbf{M}_f(\mathbf{C},\mu)) \geq B_{\mathbf{C}}(\mathbf{P})C_{max}^{Km/l}. \quad (41)$$

On the other hand, a direct evaluation of $F_{\mathbf{C},\mu}(\mathbf{M}_f(\mathbf{C},\mu))$ using (39) yields

$$F_{\mathbf{C},\mu}(\mathbf{M}_f(\mathbf{C},\mu)) = B_{\mathbf{C}}(\mathbf{M}_f(\mathbf{C},\mu))$$
$$\Big(\prod_{k=1}^K H^l(\mathbf{c}_k^T\mathbf{M}_f(\mathbf{C},\mu)^{-1}\mathbf{c}_k)\Big)^{\frac{1}{\mu}} \prod_{k=1}^K \big(\mathbf{c}_k^T\mathbf{M}_f(\mathbf{C},\mu)\mathbf{c}_k\big)^{(\frac{l}{\mu}-m)}. \quad (42)$$

For $l/m =< \mu_+$, $\mathbf{M}_f(\mathbf{C}, l/m) \in \mathcal{D}$ and we would have, by gathering (41) and (42) evaluated at $\mu = l/m$,

$$C_{max}^{Km/l}B_{\mathbf{C}}(\mathbf{P})$$
$$\leq B_{\mathbf{C}}(\mathbf{M}_f(\mathbf{C},l/m))\Big(\prod_{k=1}^K H^l(\mathbf{c}_k^T\mathbf{M}_f(\mathbf{C},l/m)^{-1}\mathbf{c}_k)\Big)^{m/l}$$
$$\leq C_{max}^{Km/l}B_{\mathbf{C}}(\mathbf{M}_f(\mathbf{C},l/m)).$$

This implies that we have equalities in all the inequalities and in particular that

$$H^l(\mathbf{c}_k^T\mathbf{M}_f(\mathbf{C},l/m)^{-1}\mathbf{c}_k) = C_{max},$$

for all $k = 1, \cdots, K$. Since H^l is strictly increasing, we reached a contradiction. Therefore, $\mu_+ = l/m$ and thus $\mathbf{M}_f(\mathbf{C},l/m) \notin \mathcal{D}$.

It remains to prove (38). Notice that all the terms in (42) except $B_{\mathbf{C}}(\mathbf{M}_f(\mathbf{C},l/m))$ remain clearly bounded as μ tends to μ_+ from below. If $\mathbf{M}_f(\mathbf{C},l/m) \neq 0$, then $B_{\mathbf{C}}(\mathbf{M}_f(\mathbf{C},\mu))$ tends to zero as μ tends to μ_+ since $B_{\mathbf{C}}$ can be continuously extended by zero on the boundary of $\mathcal{D}$ minus 0. From (41), it would result $B_{\mathbf{C}}(\mathbf{P}) = 0$, which is impossible. That contradiction ends the proof of Lemma V.5. ∎

We now have enough material to provide an argument for Theorem III.2. Thanks to the previous results, it turns out that the map $\mu \mapsto \mathrm{Tr}\Big((\mathbf{M}_f(\mathbf{C},\mu))\Big)$, defined on I, is is bijection between I and $\mathbb{R}_+^*$. Therefore, there exists a unique $\mu_m \in I$ such that $\mathrm{Tr}\Big((\mathbf{M}_f(\mathbf{C},\mu_m))\Big) = m$. A simple computation yields that $\mathbf{M}_f(\mathbf{C},\mu_m)$ is a fixed point of $G_{\mathbf{C}}$. Let us prove that it is the only one.

Indeed, let $\mathbf{M}$ be a fixed point of $G_{\mathbf{C}}$. Then,

$$f_{\mathbf{C}}(\mathbf{M}) = \xi\mathbf{M},$$

with $\xi := \dfrac{\mathrm{Tr}(f_{\mathbf{C}}(\mathbf{M}))}{m}$, i.e., $\mathbf{M}$ is the fixed point of $f_{\mathbf{C},\xi}$. Then $\mathbf{M} = \mathbf{M}_f(\mathbf{C},\xi)$ and $\mathrm{Tr}(\mathbf{M}) = m$. We deduce that $\xi = \mu_m$ hence $\mathbf{M} = \mathbf{M}_f(\mathbf{C},\mu_m)$.

Remark V.1

Following [15], one could derive similar results by replacing the normalization $Tr(\mathbf{M}) = 1$ by another increasing and homogeneous of degree one real-valued function over $\mathcal{D}$ such as $|\mathbf{M}|^{1/m}$.

VI. Conclusion

In this paper, the problem of covariance matrix estimation in impulsive noise modeled by Spherically Invariant Random Vectors was considered. The exact maximum likelihood estimate, defined as the solution of a fixed point equation and

denoted $\widehat{\mathbf{M}}_f(\mathbf{C})$ has been studied; we first demonstrate its existence and its uniqueness for the chosen normalization on the real covariance matrix, $\mathrm{Tr}(\mathbf{M}) = m$, i.e., $\widehat{\mathbf{M}}_f(\mathbf{C})$ is the unique solution of the corresponding fixed point equation which verifies $\mathrm{Tr}(\widehat{\mathbf{M}}_f(\mathbf{C})) = m$. The second main result consists in showing that the associated algorithm, which has been proposed in [9], is indeed convergent for any initial condition in $\mathcal{G}$.

The next step of the estimation analysis regards the statistical performance of $\widehat{\mathbf{M}}_f(\mathbf{C})$: bias, consistency and asymptotic distribution. This will be the object of a future work.

APPENDIX A
PROOF OF THEOREM III.3

We provide additional notation. If $\mathbf{T} \in \mathcal{M}_m(\mathbb{R})$ is invertible and $\mathbf{C}$ is a K-tuple of vectors of $\mathbb{R}^m$, then $\mathbf{T}\cdot\mathbf{C}$ denotes the K-tuple of vectors of $\mathbb{R}^m$ given by

$$\mathbf{T}\cdot\mathbf{C} := (\mathbf{T}\mathbf{c}_1, \cdots, \mathbf{T}\mathbf{c}_K). \tag{43}$$

Moreover, one has, for every $\mathbf{M} \in \mathcal{D}$,

$$\begin{cases} B_{\mathbf{C}}(\mathbf{M}) = \dfrac{1}{|\mathbf{T}|^K} B_{\mathbf{T}\cdot\mathbf{C}}(\mathbf{T}^{-1}\mathbf{M}\mathbf{T}), \\ F_{\mathbf{C}}(\mathbf{M}) = \dfrac{1}{|\mathbf{T}|^K} F_{\mathbf{T}\cdot\mathbf{C}}(\mathbf{T}^{-1}\mathbf{M}\mathbf{T}). \end{cases} \tag{44}$$

A. Proof of Proposition IV.1

Item (i) follows from a trivial computation. For $q > 0$, write $H_m(q) = q^m h_m(q)$ as

$$H_m(q) = \int_0^\infty s_m(\tau/q)p(\tau)d\tau,$$

where $s_m(t) := t^{-m}\exp(-1/t)$. Set $t_m := 1/m$. After some computations, one has

$$\lim_{t\to 0} s_m(t) = \lim_{t\to\infty} s_m(t) = 0,$$

and s_m is increasing on $(0, t_m)$, decreasing on (t_m, ∞) with unique maximum at t_m.

Let $\mathcal{P}(X) := \int_0^X p(\tau)d\tau$ for $X > 0$ and let $A > 1$ be a parameter to be fixed later. Then, by cutting $\mathbb{R}_+$ in the three intervals $[0, t_m/A]$, $[t_m/A, At_m]$ and $[At_m, \infty)$, one has

$$\begin{aligned} H_m(q) \le\ & s_m(t_m/A)\mathcal{P}(qt_m/A) + s_m(At_m)(1 - \mathcal{P}(qAt_m)) \\ & + s_m(t_m)(\mathcal{P}(qAt_m) - \mathcal{P}(qt_m/A)). \end{aligned}$$

One immediately deduces that

$$\limsup_{q\to 0} H_m(q) \le s_m(At_m),$$

and the conclusion by letting A tend to ∞. Similarly,

$$\limsup_{q\to\infty} H_m(q) \le s_m(t_m/A),$$

and the conclusion by letting A tend to ∞. As for Item (iii), an easy computation yields

$$c_m'(q) = -c_m(q)(c_{m+1}(q) - c_m(q)),$$

and

$$c_{m+1}(q) - c_m(q) = \frac{h_{m+2}(q)h_m(q) - h_{m+1}^2(q)}{h_m(q)^2}.$$

Writing the numerator as a double integral leads to

$$\begin{aligned} h_{m+2}(q)h_m(q) - h_{m+1}^2(q) = & \iint_{(\mathbb{R}_+)^2} p(\tau)p(\mu) \\ & \exp(-q(\frac{1}{\tau}+\frac{1}{\mu}))(\tau\mu)^{-m}(\frac{1}{\tau^2} - \frac{1}{\tau\mu})d\tau d\mu. \end{aligned}$$

By exchanging τ and μ in the above integral and taking the arithmetic mean, we get

$$\begin{aligned} h_{m+2}(q)h_m(q) - h_{m+1}^2(q) = & \frac{1}{2}\iint_{(\mathbb{R}_+)^2} p(\tau)p(\mu) \\ & \exp(-q(\frac{1}{\tau}+\frac{1}{\mu}))(\tau\mu)^{-m}(\frac{1}{\tau} - \frac{1}{\mu})^2 d\tau d\mu > 0. \end{aligned}$$

We proceed similarly for Item (iv). We start from

$$g_m(q) = \frac{H_{m+1}(q)}{H_m(q)},$$

and get $g_m'(q) = \dfrac{H_{m+1}'(q)H_m(q) - H_{m+1}(q)H_m'(q)}{H_m^2(q)}$. In all the integrals, we make the change of variable $\tau \to \tau/q$ and we write the numerator $\mathcal{H}$ of the last fraction as a double integral,

$$\begin{aligned} \mathcal{H} = & \iint_{(\mathbb{R}_+)^2} (\tau\mu)^{-m}\exp(-q(\frac{1}{\tau}+\frac{1}{\mu})) \\ & \Big(p'(q\tau)p(q\mu) - \frac{\mu}{\tau}p(q\tau)p'(q\mu)\Big)d\tau d\mu. \end{aligned}$$

Finally, one has

$$p'(q\tau)p(q\mu) - \frac{\mu}{\tau}p(q\tau)p'(q\mu) = \frac{p(q\tau)p(q\mu)}{q\tau}(P(q\tau) - P(q\mu)),$$

where $P(\tau) = \tau\dfrac{p'(\tau)}{p(\tau)}$. By exchanging τ and μ in the above integral and taking the arithmetic mean, we get

$$\begin{aligned} \mathcal{H} = -\frac{1}{2q}\iint_{(\mathbb{R}_+)^2} & (\tau\mu)^{-(m+1)}\exp(-q(\frac{1}{\tau}+\frac{1}{\mu}))p(q\tau)p(q\mu) \\ & \times(\tau-\mu)^2\frac{P(q\tau) - P(q\mu)}{q\tau - q\mu}d\tau d\mu. \end{aligned}$$

Therefore, if the sign of the derivative of P is constant, so is the sign of the derivative of g_m and the two signs are opposite. A simple computation on known PDF's shows that g_m is thus strictly increasing.

B. Proof of Proposition IV.2

For $\mathbf{M} \in \mathcal{D}$, write

$$F_{\mathbf{C}}(\mathbf{M}) = B_{\mathbf{C}}(\mathbf{M})\prod_{k=1}^K H_m(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k).$$

In [3], it was proved that $B_{\mathbf{C}}$ is homogeneous of degree zero, uniformly bounded over $\mathcal{D}$ and it can be continuously extended to the boundary of $\mathcal{D}$ minus the zero matrix by zero.

Moreover, $B_{\mathbf{C}}$ reaches its maximum over a unique half-line, supported by a well-defined matrix $\mathbf{P}_B$ of trace m.

Combined with Item (ii) of Proposition IV.1, we deduce from the above that $F_{\mathbf{C}}$ is uniformly bounded over $\mathcal{D}$. Set

$$\max(F) = \sup_{\mathbf{M}\in\mathcal{D}} F_{\mathbf{C}}(\mathbf{M}), \quad \max(B) = \sup_{\mathbf{M}\in\mathcal{D}} B_{\mathbf{C}}(\mathbf{M}),$$

and define

$$\mathcal{K} := \{\mathbf{M}\in\mathcal{D} \ \mid \ F_{\mathbf{C}}(\mathbf{M}) \geq \max(F)/2\}.$$

Proving Item (A) amounts to show that $\mathcal{K}$ is a compact set of $\mathcal{D}$, i.e., there exists two positive real numbers k_1, k_2 so that, for every $\mathbf{M}\in\mathcal{K}$,

$$k_1\mathbf{I}_m \leq \mathbf{M}^{-1} \leq k_2\mathbf{I}_m.$$

We argue by contradiction. Then, there exists a sequence of unit norm vectors $\mathbf{x}_n$, $n\geq 0$, and a sequence of matrices $\mathbf{M}_n$ in $\mathcal{D}$ so that

$$\text{either } \lim_{n\to\infty} \mathbf{x}_n^T\mathbf{M}_n^{-1}\mathbf{x}_n = 0 \text{ or } \lim_{n\to\infty} \mathbf{x}_n^T\mathbf{M}_n^{-1}\mathbf{x}_n = \infty. \quad (45)$$

Note that, if $\mathbf{U}\in\mathcal{SO}(m)$, Eq. (44) reduces to

$$B_{\mathbf{U}.\mathbf{C}}(\mathbf{U}^{-1}\mathbf{M}\mathbf{U}) = B_{\mathbf{C}}(\mathbf{M}), \quad F_{\mathbf{U}.\mathbf{C}}(\mathbf{U}^{-1}\mathbf{M}\mathbf{U}) = F_{\mathbf{C}}(\mathbf{M}).$$

In particular, $\max(F_{\mathbf{U}.\mathbf{C}})$ and $\max(B_{\mathbf{U}.\mathbf{C}})$ only depend on $\mathbf{C}$ and not on $\mathbf{U}\in\mathcal{SO}(m)$. Consequently, if $(\mathbf{U}_n)_{n\geq 0}$ is a sequence of orthogonal matrices so that $\mathbf{U}_n\mathbf{c}_1 = \|\mathbf{c}_1\|\mathbf{x}_n$, then

$$\begin{aligned}\max(F_{\mathbf{C}})/2 \leq F_{\mathbf{C}}(\mathbf{M}_n) &= F_{\mathbf{U}_n.\mathbf{C}}(\mathbf{U}_n^{-1}\mathbf{M}_n\mathbf{U}_n)\\ &\leq \max(B_{\mathbf{C}})s_m(t_m)^{N-1}H_m(\|\mathbf{c}_1\|^2\mathbf{x}_n^T\mathbf{M}_n^{-1}\mathbf{x}_n),\end{aligned}$$

and the last term tends to zero as n tends to ∞, according to (45). Since $\max(F_{\mathbf{C}})>0$, we reached a contradiction and Item (A) is proved.

Let $\mathbf{M}$ be a critical point of $F_{\mathbf{C}}$ or equivalently a fixed point of $f_{\mathbf{C}}$. A simple computation of the Hessian of $F_{\mathbf{C}}$ at $\mathbf{M}$ yields

$$\begin{aligned}Hess\ F_{\mathbf{C}}(\mathbf{M})(\mathbf{Q},\mathbf{Q}) = K(&\operatorname{Tr}(\mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q})+\\ &\frac{1}{K}\sum_{k=1}^K c_m'(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k)(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\mathbf{c}_k)^2),\end{aligned}$$

where $\mathbf{Q}$ is an arbitrary symmetric $m\times m$ matrix. Set $\mathbf{d}_k := \mathbf{M}^{-1/2}\mathbf{c}_k$ and $\mathbf{R} := \mathbf{M}^{-1/2}\mathbf{Q}\mathbf{M}^{-1/2}$. Then, one has

$$\begin{aligned}&Hess\ F_{\mathbf{C}}(\mathbf{M})(\mathbf{Q},\mathbf{Q}) =\\ &\quad -K\left[\operatorname{Tr}(\mathbf{R}^2)+\frac{1}{K}\sum_{k=1}^K c_m'(\|\mathbf{d}_k\|^2)(\mathbf{d}_k^T\mathbf{R}\mathbf{d}_k)^2\right]. \quad (46)\end{aligned}$$

Since $\mathbf{M}$ is a fixed point of $f_{\mathbf{C}}$, one has,

$$\mathbf{I}_m = \frac{1}{K}\sum_{k=1}^K c_m(\|\mathbf{d}_k\|^2)\mathbf{d}_k\mathbf{d}_k^T.$$

Multiplying the previous equation on the left and on the right by $\mathbf{R}$, then taking the trace and inserting the result in (46), we get

$$\begin{aligned}&Hess\ F_{\mathbf{C}}(\mathbf{M})(\mathbf{Q},\mathbf{Q}) =\\ &\quad -\sum_{k\in I_R}\|\mathbf{R}\mathbf{d}_k\|^2[c_m(\|\mathbf{d}_k\|^2)+\mathbf{r}_k\|\mathbf{d}_k\|^2c_m'(\|\mathbf{d}_k\|^2)],\end{aligned}$$

where I_R is the set of indices k such that $\mathbf{R}\mathbf{d}_k\neq 0$ and

$$\mathbf{r}_k := \left(\frac{\mathbf{d}_k^T\mathbf{R}\mathbf{d}_k}{\|\mathbf{d}_k\|\|\mathbf{R}\mathbf{d}_k\|}\right)^2 \in [0,1].$$

Then clearly,

$$Hess\ F_{\mathbf{C}}(\mathbf{M})(\mathbf{Q},\mathbf{Q}) \leq -\sum_{k\in I_R}\mathbf{r}_k\|\mathbf{R}\mathbf{d}_k\|^2g_m'(\|\mathbf{d}_k\|^2),$$

which is strictly negative if $\mathbf{Q}\neq 0$.

C. Proof of Proposition IV.3

Items (i) is proved exactly as Item $(P1)$ of Proposition $V.2$ and Proposition $V.3$ of [3]. For Item (ii), consider $\lambda\in(0,1)$ and $\mathbf{M}\in\mathcal{D}$. Write $f_{\mathbf{C}}(\lambda\mathbf{M})$ as

$$f_{\mathbf{C}}(\lambda\mathbf{M}) = \frac{\lambda}{K}\sum_{k=1}^K g_m(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k/\lambda)\frac{\mathbf{c}_k\mathbf{c}_k^T}{\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k}.$$

Since g_m is increasing, one has

$$f_{\mathbf{C}}(\lambda\mathbf{M}) > \frac{\lambda}{K}\sum_{k=1}^K g_m(\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k)\frac{\mathbf{c}_k\mathbf{c}_k^T}{\mathbf{c}_k^T\mathbf{M}^{-1}\mathbf{c}_k} = \lambda f_{\mathbf{C}}(\mathbf{M}).$$

Suppose now that $\lambda>1$. Apply the previous result to $\lambda\mathbf{M}$ and $1/\lambda$ respectively instead of $\mathbf{M}$ and λ.

D. Proof of Item (c) of Theorem III.3

Let $\mathbf{M}\in\mathcal{D}$ and $(\mathbf{M}_n)_{n\geq 0}$ be the orbit of $\mathbf{M}$ by the iterative scheme associated to $f_{\mathbf{C}}$. Then $\mathbf{M}_0 = \mathbf{M}$ and $\mathbf{M}_{n+1} = f_{\mathbf{C}}(\mathbf{M}_n)$ for $n\geq 0$. There exists two positive real numbers $\lambda_1<1$ and $\lambda_2>1$ such that

$$\lambda_1\mathbf{M}_f(\mathbf{C}) < \mathbf{M} < \lambda_2\mathbf{M}_f(\mathbf{C}).$$

Define $\mathcal{K}_{\mathbf{M}}$ as the compact subset of $\mathcal{D}$ given by

$$\mathcal{K}_{\mathbf{M}} := \{\mathbf{Q}\in\mathcal{D} \ \mid \ \lambda_1\mathbf{M}_f(\mathbf{C}) \leq \mathbf{Q} \leq \lambda_2\mathbf{M}_f(\mathbf{C})\}.$$

By a trivial inductive argument which uses the subhomogeneity of $f_{\mathbf{C}}$, we show that the orbit of $\mathbf{M}$ remains in $\mathcal{K}_{\mathbf{M}}$. We will prove a more precise statement. For that purpose, define, for $\varepsilon>0$ small enough,

$$M(\varepsilon) := \min\frac{g_m(\mathbf{c}_k^T\mathbf{Q}^{-1}\mathbf{c}_k/\lambda)}{g_m(\mathbf{c}_k^T\mathbf{Q}^{-1}\mathbf{c}_k)},$$

over all $\lambda\in[\lambda_1, 1-\varepsilon]$, $k\in\{1,\cdots,K\}$ and $\mathbf{Q}\in\mathcal{K}_{\mathbf{M}}$ and

$$m(\varepsilon) := \max\frac{g_m(\mathbf{c}_k^T\mathbf{Q}^{-1}\mathbf{c}_k/\lambda)}{g_m(\mathbf{c}_k^T\mathbf{Q}^{-1}\mathbf{c}_k)},$$

over all $\lambda\in[1+\varepsilon,\lambda_2]$, $k\in\{1,\cdots,K\}$ and $\mathbf{Q}\in\mathcal{K}_{\mathbf{M}}$. It is easy to see that $M(\varepsilon)>1$ and $m(\varepsilon)<1$. Then, one clearly has

$(i1)$ If $\mathbf{Q}\in\mathcal{K}_{\mathbf{M}}$ and $\lambda\in[\lambda_1,1-\varepsilon]$, then $f_{\mathbf{C}}(\lambda\mathbf{Q}) \geq M(\varepsilon)\lambda f_{\mathbf{C}}(\mathbf{Q})$;

$(i2)$ if $\mathbf{Q}\in\mathcal{K}_{\mathbf{M}}$ and $\lambda\in[1+\varepsilon,\lambda_2]$, then $f_{\mathbf{C}}(\lambda\mathbf{Q}) \leq m(\varepsilon)\lambda f_{\mathbf{C}}(\mathbf{Q})$.

Assume now that ε is very small. Define the integers m_1, m_2 and n_ε as follows that

$$m_1 := \left[\frac{\ln((1-\varepsilon)/\lambda_1)}{\ln(M(\varepsilon))}\right], \quad m_2 := \left[\frac{\ln((1+\varepsilon)/\lambda_2)}{\ln(m(\varepsilon))}\right],$$
$$n_\varepsilon = \max(m_1, m_2),$$

where $[\cdot]$ denotes the integer part. We first claim that

$$\lambda_1 M(\varepsilon)^{m_1}\mathbf{M}_f(\mathbf{C}) \leq \mathbf{M}_{n_\varepsilon} \leq \lambda_2 m(\varepsilon)^{m_2}\mathbf{M}_f(\mathbf{C}). \quad (47)$$

Indeed, if for some integers n and $l \leq n_\varepsilon - 1$, one has $\lambda_1 M(\varepsilon)^l \mathbf{M}_f(\mathbf{C}) \leq \mathbf{M}_n$, then, by Item $(i1)$, we have $\lambda_1 M(\varepsilon)^{l+1}\mathbf{M}_f(\mathbf{C}) \leq \mathbf{M}_{n+1}$, and that goes on as long as $\lambda_1 M(\varepsilon)^l < 1$. The argument is identical for the other inequality of (47).

With an easy inductive argument using the subhomogeneity of $f_{\mathbf{C}}$, we show that, for every $n \geq n_\varepsilon$,

$$\lambda_1 M(\varepsilon)^{m_1}\mathbf{M}_f(\mathbf{C}) \leq \mathbf{M}_n \leq \lambda_2 m(\varepsilon)^{m_2}\mathbf{M}_f(\mathbf{C}).$$

This clearly implies that, for every $n \geq n_\varepsilon$,

$$\frac{1-\varepsilon}{M(\varepsilon)}\mathbf{M}_f(\mathbf{C}) \leq \mathbf{M}_n \leq \frac{1+\varepsilon}{m(\varepsilon)}\mathbf{M}_f(\mathbf{C}).$$

Since both $M(\varepsilon)$ and $m(\varepsilon)$ tend to one as ε tends to zero, we deduce that $\mathbf{M}_n$ tends to $\mathbf{M}_f(\mathbf{C})$ as n tends to ∞.

REFERENCES

[1] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-gaussian clutter," *IEEE Trans.-AES*, vol. 31, no. 2, pp. 617–625, April 1995.

[2] F. Gini, "Sub-optimum coherent radar detection in a mixture of k-distributed and gaussian clutter," *IEE Proc. Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 39–48, February 1997.

[3] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum likelihood estimates in compound gaussian noise : Existence and algorithm analysis," *IEEE Trans.-SP*, vol. 56, no. 1, pp. 34–48, January 2008.

[4] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[5] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans.-AES*, vol. 23, no. 1, pp. 115–127, November 1986.

[6] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A cfar adaptive matched filter detector," *IEEE Trans.-AES*, vol. 28, no. 1, pp. 208–216, January 1992.

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

[8] D. R. Raghavan, R S and N. B. Pulsone, "A generalization of the adaptive matched filter receiverfor array detection in a class of a non-gaussian interference," *Proc. ASAP Workshop, Lexinton, MA*, pp. 499–517, March 1996.

[9] F. Gini and M. V. Greco, "Covariance matrix estimation for cfar detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.

[10] R. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. John Wiley & Sons, New York, 1987.

[11] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-gaussian interference backgrounds," *IEEE Trans.-SP*, vol. 53, no. 6, pp. 2101–2111, June 2005.

[12] K.-T. Fang and T. W. Anderson, *Statistical Inference in Elliptically Contoured Related Distributions*. Allerton Press Inc., New York, 1990.

[13] R. A. Maronna, "Robust m-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.

[14] J. T. Kent and D. E. Tyler, "Redescending m-estimates of multivariate location and scatter," *Annals of Statistics*, vol. 19, no. 4, pp. 2102–2119, December 1991.

[15] R. D. Nussbaum, "Hilbert's projective metric and iterated nonlinear maps," *Mem. Amer. Math. Soc.*, vol. 75, no. 391, 1988.

[16] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *IEEE Trans.-IT*, vol. 19, no. 5, pp. 600–608, September 1973.

[17] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Non-gaussian vector identification using spherically invariant random processes," *IEEE Trans.-AES*, vol. 29, no. 1, pp. 111–124, January 1993.

[18] E. Jay, J.-P. Ovarlez, D. Declercq, and P. Duvaut, "Bord: Bayesian optimum radar detector," *Signal Processing*, vol. 83, no. 6, pp. 1151–1162, June 2003.

[19] E. Jay, "Détection en environnement non-gaussien"," Ph.D. dissertation, University of Cergy-Pontoise / ONERA, France, June 2002.

[20] R. A. Horn and J. C. R, *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.

[21] K. B. Petersen and P. M. S, "The matrix cookbook," http://matrixcookbook.com, Tech. Rep., February 2008.

[22] E. Conte and G. Ricci, "Performance prediction in compound-gaussian clutter," *IEEE Trans.-AES*, vol. 30, no. 2, pp. 611–616, April 1994.

[23] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential equations, dynamical systems, and an introduction to chaos*, 2nd ed., ser. Pure and Applied Mathematics (Amsterdam). Elsevier/Academic Press, Amsterdam, 2004, vol. 60.

[24] A. J. B. Potter, "Applications of hilbert's projective metric to certain classes of non-homogeneous operators," *Quart. J. Math. Oxford Ser.*, vol. 28, no. 2, pp. 93–99, 1977.



Yacine Chitour was born in Algiers, Algeria, in 1968. He graduated from Ecole Polytechnique, Palaiseau, France, in 1990 and received the Ph.D. degree in mathematics from Rutgers University, New Brunswick, NJ, in 1996. Currently, he is the Professor of Control Theory at UniversitÈ Paris-Sud, where he is a member of the Laboratoire des Signaux et SystËmes, Gif-sur-Yvette, France. His research interests are in nonlinear control theory.



Frédéric Pascal was born in Sallanches, France in 1979. He received the Master's degree ("Probabilities, Statistics and Applications : Signal, Image et Networks") with merit, in Applied Statistics from University Paris VII - Jussieu, Paris, France, in 2003. Then, he received the Ph.D. degree of Signal Processing, from University Paris X - Nanterre, advised by Pr. Philippe Forster : "Detection and Estimation in Compound Gaussian Noise" in 2006. This Ph.D thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France. From november 2006 to February 2008, he made a post doctoral position in the Signal Processing and Information team of the laboratory SATIE (Système et Applications des Technologies de l'Information et de l'Energie), CNRS, Ecole Normale Supérieure, Cachan, France. From march 2008, he is an Assistant Professor in the SONDRA laboratory, Supelec. His research interests are estimation in statistical signal processing and radar detection.