Robust signal and noise covariance matrix estimation using Riemannian optimization

Jesper Brunnström* , Marc Moonen* and Filip Elvander†
*Dept. of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium
†Department of Information and Communications Engineering, Aalto University, Espoo, Finland

***Abstract*—Covariance matrix estimation for a noise-contaminated signal is a common signal processing task, where the covariance matrix of the desired signal together with the noise covariance matrix are estimated from two sets of data, i.e., noise-only data and noise-contaminated signal data. The estimation problem can be posed as a constrained optimization problem, for which a closed-form solution exists in terms of the generalized eigenvalue decomposition if the cost function is chosen to be Euclidean distance with prewhitening. However, the problem is difficult to solve for other cost functions, which means it is difficult to incorporate prior knowledge or adapt to new applications. In this paper the optimization problem is recast as an unconstrained optimization problem on a Riemannian manifold, an approach that has not yet been investigated for this particular problem. Using the constructed manifold, a robust covariance matrix estimator based on Tyler's M-estimator is derived, which uses noise-only data and noise-contaminated signal data. In particular, the robust estimator is useful for impulsive noise, a situation where conventional estimators commonly perform poorly. The effectiveness of the proposed robust estimator under heavy-tailed noise is demonstrated, compared against conventional methods on synthesized data as well as an audio noise-reduction application.**

***Index Terms*—noise reduction, robust covariance matrix estimation, Riemannian optimization, Hermitian positive matrices, manifolds**

I. Introduction

Many signal processing applications critically rely on the estimation of the covariance matrix of a signal. When the signal is contaminated by noise, as often is the case, the covariance matrix estimation becomes difficult. However, for some signals which exhibit an on-off behaviour, it is possible to use noisy-only data together with the noise-contaminated signal data to obtain a better estimate of the covariance matrix of the desired signal together with the noise covariance matrix [1]. In addition, the signal covariance matrix is often low-rank, and so the estimate can be improved if a rank-constraint can be enforced.

An effective approach has been found in the use of the generalized eigenvalue decomposition (GEVD) [1]. The GEVD solution is the rank-constrained optimal solution with regards to a Euclidian distance with prewhitening [2]. Some additional prior knowledge can be imposed regarding the desired signal subspace [3], which has found use in spatial filtering applications [4], [5]. However, the approach does not easily generalize to other cost functions and additional constraints. Related problem statements can also be found in joint noise reduction with acoustic echo or feedback cancellation [6], as well as in sound field reproduction [7], [8].

For positive semi-definite matrices as considered in the covariance matrix estimation problem, Euclidian space is not a natural setting for analysis [9]. Instead, the matrices can be represented as a Riemannian manifold, and the resulting optimization problems can be solved using Riemannian optimization algorithms [10]. Optimization over positive definite matrices has attracted much attention [11], [12], over a variety of geometries [13]–[15]. The problem of signal and noise covariance matrix estimation from noise-only data and noise-contaminated signal data has not yet been considered in the context of Riemannian manifolds, which will be done in this paper.

The Riemannian manifold approach has earlier been utilized in the context of the robust covariance matrix estimation for a single covariance matrix. Arguably the most common robust estimator is Tyler's M-estimator, due to its desirable distribution-free properties, and good performance for heavy-tailed data distributions [16]. The construction and analysis of this type of estimator has been aided by the Riemannian manifold approach [17], [18]. In this paper, a related robust covariance matrix estimator is derived which uses noise-only data and noise-contaminated signal data, in a Riemannian setting.

The contribution of this paper is a flexible framework of covariance matrix estimation from noisy-only data and noise-contaminated signal data. The framework allows more flexibility with regards to cost functions, and therefore also allows for an easier incorporation of prior knowledge. In particular, an estimator based on Tyler's M-estimator is derived, with which the signal covariance matrix can be more effectively estimated under more heavy-tailed noise-contamination. The effectiveness of the proposed robust estimator under heavy-tailed noise is demonstrated, compared against conventional methods on synthesized data as well as an audio noise reduction application [1].

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: 'Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS'. The scientific responsibility is assumed by its authors.

[1]The code and data is available at github.com/sounds-research/riecovest

II. Problem statement

A. Signal model

It is assumed that there are M sensors available, each recording a noise-contaminated signal which is processed in the frequency domain. The sensor signal for each time-frequency bin can be represented as

$$\boldsymbol{y}_n = \boldsymbol{x}_n + \boldsymbol{v}_n \in \mathbb{C}^M, \tag{1}$$

which consists of the signal $\boldsymbol{x}_n$ contaminated by the noise $\boldsymbol{v}_n$. The time index is represented by n, and the frequency bin index is omitted for brevity.

The signal $\boldsymbol{x}_n$ is assumed to exhibit an on-off behaviour, allowing for the collection of N_y noise-contaminated signal samples, as well as N_v noise-only samples. The data available is

$$\begin{aligned} \boldsymbol{Y} &= \begin{bmatrix} \boldsymbol{y}_0 & \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_{N_y-1} \end{bmatrix} \in \mathbb{C}^{M \times N_y} \\ \boldsymbol{V} &= \begin{bmatrix} \boldsymbol{v}_{N_y} & \boldsymbol{v}_{N_y+1} & \cdots & \boldsymbol{v}_{N_y+N_v-1} \end{bmatrix} \in \mathbb{C}^{M \times N_v}, \end{aligned} \tag{2}$$

where the indices are offset to emphasize that the noise samples $\boldsymbol{V}$ and noise-contaminated signal samples $\boldsymbol{Y}$ are collected at different times. From here it will be assumed that $N = N_y = N_v$ for brevity, the extension to different number of samples is trivial.

B. Covariance matrices

The data described in (2) can be used to estimate the properties of the sensor signals, in particular its covariance matrix. The signal covariance matrix is $\boldsymbol{R}_x = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}] \in \mathcal{S}_+(M)$ and the noise covariance matrix is $\boldsymbol{R}_v = \mathbb{E}[\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}] \in \mathcal{S}_+(M)$ where $\mathbb{E}$ denotes the expectation operator and $\mathcal{S}_+(M)$ the set of Hermitian positive semi-definite matrices, i.e.,

$$\mathcal{S}_+(M) = \{\boldsymbol{R} \in \mathbb{C}^{M \times M} | \boldsymbol{R} \succcurlyeq 0, \boldsymbol{R} = \boldsymbol{R}^{\mathsf{H}}\}. \tag{3}$$

The signal and noise are assumed to be independent, implying that the covariance matrix of the noise-contaminated signal is

$$\boldsymbol{R}_y = \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}] = \boldsymbol{R}_x + \boldsymbol{R}_v. \tag{4}$$

The covariance matrices $\boldsymbol{R}_y$ and $\boldsymbol{R}_v$ can be estimated from the data $\boldsymbol{Y}$ and $\boldsymbol{V}$ respectively. There are many methods for estimating a single covariance matrix, taking the form of a function $C : \mathbb{C}^{M \times N} \to S_+(M)$. One of the most common estimator is the sample covariance matrix (SCM), defined as $C_{\text{SC}}(\boldsymbol{Y}) = \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$. Another useful estimator is Tyler's M-estimator $C_{\text{M}} : \mathbb{C}^{M \times N} \to \mathcal{S}_+(M)$, a robust estimator that can be effective for heavy-tailed distributions where the SCM performs poorly [16]. The estimate $\boldsymbol{R} = C_{\text{M}}(\boldsymbol{Y})$ can be defined as the solution to

$$\boldsymbol{R} = \frac{M}{N} \sum_{n=0}^{N-1} \frac{\boldsymbol{y}_n \boldsymbol{y}_n^{\mathsf{H}}}{\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{R}^{-1} \boldsymbol{y}_n}, \tag{5}$$

and can be computed through fixed-point iteration [19].

The problem of estimating the signal covariance matrix $\boldsymbol{R}_x$ is more difficult, as the signal $\boldsymbol{x}$ is never observed in isolation. The most straightforward estimator for the signal covariance matrix is to directly subtract the SCMs. Considering (4), the signal covariance matrix estimate could for instance be given as

$$\hat{\boldsymbol{R}}_x = C_{\text{SC}}(\boldsymbol{Y}) - C_{\text{SC}}(\boldsymbol{V}). \tag{6}$$

However, this estimate will be sensitive to estimation errors in the SCMs, and the resulting signal covariance matrix will in general not be positive semi-definite, and therefore not a valid covariance matrix.

Some prior knowledge may be considered to improve performance, namely that the signal covariance matrix is low-rank. This occurs when more sensors are used compared to the number of signal sources, in which case the rank is the number of independent signal sources. Assuming Q sources, the signal covariance matrix belongs to the set of rank-Q matrices

$$\mathcal{S}_+(M, Q) = \{\boldsymbol{R} \in \mathcal{S}_+(M) | \operatorname{rank} \boldsymbol{R} = Q\}. \tag{7}$$

The noise covariance matrix $\boldsymbol{R}_v$ is assumed to be full rank, i.e., $\boldsymbol{R}_v \in \mathcal{S}_+(M, M)$, implying that also the covariance matrix of the noise-contaminated signal is full-rank, i.e., $\boldsymbol{R}_y \in \mathcal{S}_+(M, M)$.

C. Signal and noise covariance matrix estimation

An effective estimator should guarantee that $\hat{\boldsymbol{R}}_x$ is Hermitian, positive semi-definite, rank-Q, in addition to fitting the collected data. The data fit is quantified by means of a cost function $\mathcal{D} : \mathbb{C}^{M \times N} \times \mathcal{S}_+(M, M) \to \mathbb{R}_+$. Estimation of the signal and noise covariance matrix can then be formulated as the optimization problem

$$\begin{aligned} \underset{\hat{\boldsymbol{R}}_x, \hat{\boldsymbol{R}}_v}{\text{minimize}} \quad & \mathcal{D}(\boldsymbol{Y}, \hat{\boldsymbol{R}}_v + \hat{\boldsymbol{R}}_x) + \mathcal{D}(\boldsymbol{V}, \hat{\boldsymbol{R}}_v) \\ \text{subject to} \quad & \hat{\boldsymbol{R}}_x, \hat{\boldsymbol{R}}_v \succcurlyeq 0, \ \operatorname{rank} \hat{\boldsymbol{R}}_x = Q. \end{aligned} \tag{8}$$

The optimization problem can be difficult to solve in general, depending on the choice of cost functions, in a large part due to the rank constraint, which is not convex [20].

A generally more effective method is based on the GEVD, which solves problem (8) using the cost function

$$\mathcal{D}(\boldsymbol{X}, \boldsymbol{R}) = \|\boldsymbol{Q}^{-1}(C_{\text{SC}}(\boldsymbol{X}) - \boldsymbol{R})\boldsymbol{Q}^{-\mathsf{H}}\|_F^2, \tag{9}$$

which is a Euclidean distance with prewhitening. The prewhitening matrix $\boldsymbol{Q}$ is defined by the GEVD of the matrix pencil $(C_{\text{SC}}(\boldsymbol{Y}), C_{\text{SC}}(\boldsymbol{V}))$, i.e.,

$$\begin{aligned} C_{\text{SC}}(\boldsymbol{Y}) &= \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathsf{H}} \\ C_{\text{SC}}(\boldsymbol{V}) &= \boldsymbol{Q}\boldsymbol{Q}^{\mathsf{H}}, \end{aligned} \tag{10}$$

where $\boldsymbol{Q} \in \mathbb{C}^{M \times M}$ is an invertible matrix containing the generalized eigenvectors, and $\boldsymbol{\Sigma}$ is a diagonal matrix containing the generalized eigenvalues. For the cost function (9), the optimization problem (8) has a closed form solution in terms of the GEVD, which is

$$\begin{aligned} \hat{\boldsymbol{R}}_x &= \boldsymbol{Q} \begin{bmatrix} \boldsymbol{\Sigma}_Q - \boldsymbol{I} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{Q}^{\mathsf{H}} \\ \hat{\boldsymbol{R}}_v &= \frac{1}{2}(C_{\text{SC}}(\boldsymbol{Y}) + C_{\text{SC}}(\boldsymbol{V}) - \hat{\boldsymbol{R}}_x), \end{aligned} \tag{11}$$

where $\boldsymbol{\Sigma}_Q \in \mathcal{S}_+(Q,Q)$ is a diagonal matrix containing the Q principal eigenvalues, and $\boldsymbol{I}$ is the identity matrix [1]. The expression in (11) assumes that $\boldsymbol{\Sigma}_Q \succcurlyeq \boldsymbol{I}$, which is not guaranteed. If that is not the case, the resulting negative values of $\boldsymbol{\Sigma}_Q - \boldsymbol{I}$ should be set to zero.

III. Covariance matrix estimation using Riemannian optimization

A. Optimization problem

The proposed approach is to recast the constrained optimization problem (8) into an equivalent unconstrained optimization problem on a Riemannian manifold. By constructing a Riemannian manifold $\mathcal{M}$ where any pair of covariance matrices $(\hat{\boldsymbol{R}}_x, \hat{\boldsymbol{R}}_v) \in \mathcal{M}$ satisfies the constraints in (8), the optimization problem could be recast as

$$\underset{(\hat{\boldsymbol{R}}_x, \hat{\boldsymbol{R}}_v) \in \mathcal{M}}{\text{minimize}} \quad \mathcal{D}(\boldsymbol{Y}, \hat{\boldsymbol{R}}_v + \hat{\boldsymbol{R}}_x) + \mathcal{D}(\boldsymbol{V}, \hat{\boldsymbol{R}}_v). \tag{12}$$

This allows for the low-rank constraint to be dealt with in a natural manner, and the resulting optimization problem is possible to solve for a wide range of cost functions. Such an optimization problem does not have an analytic solution in general, but can be solved using a number of Riemannian optimization algorithms, such as gradient descent or trust-region methods [10]. For the setting considered herein, i.e., joint estimation of signal and noise covariance matrices, the manifold $\mathcal{M}$ will be constructed in the following sections.

B. Riemannian manifolds

To model all possible noise covariance matrices $\boldsymbol{R}_v$, the Riemannian manifold of positive definite matrices $\mathcal{M}_+(M,M)$ is considered. This is defined as the set $\mathcal{S}_+(M,M)$ along with its associated Riemannian metric, defined on the tangent space $\mathcal{T}_{\boldsymbol{R}}\mathcal{M}_+(M,M)$, which can be identified with the set of Hermitian matrices $\mathcal{H}(M) = \{\boldsymbol{R} \in \mathbb{C}^{M\times M} | \boldsymbol{R} = \boldsymbol{R}^{\mathsf{H}}\}$ [12]. Specifically, the Riemannian metric is

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\boldsymbol{R}} := \operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{B}), \tag{13}$$

for $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{T}_{\boldsymbol{R}}\mathcal{M}_+(M,M)$ [12].

To model all possible signal covariance matrices $\boldsymbol{R}_x$, a fixed-rank Riemannian manifold $\mathcal{M}_+(M,Q)$ for $Q < M$ is also considered. Any rank-Q covariance matrix can be parametrized as $\boldsymbol{R} = \boldsymbol{L}\boldsymbol{L}^{\mathsf{H}}$ where $\boldsymbol{L} \in \mathbb{C}^{M\times Q}$. The parametrization is unique only up to a unitary transform, which leads to non-isolated optima, which can be difficult to find through numerical optimization [10]. To remedy this, $\mathcal{M}_+(M,Q)$ will be constructed as a quotient manifold, by considering the equivalence class

$$[\boldsymbol{L}] = \{\boldsymbol{L}\boldsymbol{U} : \boldsymbol{L} \in \mathbb{C}_*^{M\times Q}, \boldsymbol{U} \in \mathcal{U}(M)\} \tag{14}$$

as elements of the manifold, where $\mathbb{C}_*^{M\times Q}$ is the set of all full-rank $M \times Q$ matrices, and $\mathcal{U}(M)$ is the unitary group.

To define the Riemannian metric for the manifold, the total space $\overline{\mathcal{M}}_+(M,Q)$ will be defined first, defined on the set $\mathbb{C}_*^{M\times Q}$ and related to $\mathcal{M}_+(M,Q)$ through the canonical projection $\pi : \boldsymbol{L} \to [\boldsymbol{L}]$. The associated Riemannian metric of the total space is defined on the tangent space $\mathcal{T}_{\boldsymbol{L}}\overline{\mathcal{M}}_+(M,Q)$, which can be identified with $\mathbb{C}^{M\times Q}$. The metric is

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\boldsymbol{L}} := \frac{1}{2}\operatorname{tr}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{B} + \boldsymbol{B}^{\mathsf{H}}\boldsymbol{A}) \tag{15}$$

for $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{T}_{\boldsymbol{L}}\overline{\mathcal{M}}_+(M,Q)$. Finally, the quotient manifold $\mathcal{M}_+(M,Q)$ is endowed with the Riemannian metric

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{[\boldsymbol{L}]} := \langle \overline{\boldsymbol{A}}, \overline{\boldsymbol{B}} \rangle_{\boldsymbol{L}} \tag{16}$$

for $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{T}_{[\boldsymbol{L}]}\mathcal{M}_+(M,Q)$ and $\overline{\boldsymbol{A}}, \overline{\boldsymbol{B}} \in \mathcal{T}_{\boldsymbol{L}}\overline{\mathcal{M}}_+(M,Q)$. More details about the geometry of this manifold can be found in [20]–[22].

C. Joint signal and noise covariance matrix manifold

The pair of covariance matrices $(\boldsymbol{R}_x, \boldsymbol{R}_v)$ where $\boldsymbol{R}_x = \boldsymbol{L}\boldsymbol{L}^{\mathsf{H}}$ can be jointly considered as existing on the product manifold $\mathcal{M} := \mathcal{M}_+(M,Q) \times \mathcal{M}_+(M,M)$. It is defined as the set

$$\mathcal{S} := (\mathbb{C}_*^{M\times Q}/\mathcal{U}(M)) \times \mathcal{S}_+(M,M) \tag{17}$$

along with its associated Riemannian metric, defined in terms of the individual metrics as

$$\langle (\boldsymbol{A}_x, \boldsymbol{A}_v), (\boldsymbol{B}_x, \boldsymbol{B}_v) \rangle_{[\boldsymbol{L}],\boldsymbol{R}_v} = \langle \boldsymbol{A}_x, \boldsymbol{B}_x \rangle_{[\boldsymbol{L}]} + \langle \boldsymbol{A}_v, \boldsymbol{B}_v \rangle_{\boldsymbol{R}_v}, \tag{18}$$

where $(\boldsymbol{A}_x, \boldsymbol{A}_v)$ and $(\boldsymbol{B}_x, \boldsymbol{B}_v)$ are elements of the tangent space $\mathcal{T}_{([\boldsymbol{L}],\boldsymbol{R}_v)}\mathcal{M}$ [23, Example 3.57].

IV. Proposed robust covariance matrix estimator

When the data or noise contains considerable outliers, a robust covariance matrix estimator is preferred. This can occur in particular when the noise-contamination is impulsive, i.e., generated by a heavy-tailed distribution. A common approach is to consider the data as being generated from a complex elliptically symmetric (CES) distribution [19]. The maximum likelihood covariance matrix estimator for an elliptic distribution with regards to N independent and identically distributed samples $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_0 & \boldsymbol{x}_1 & \dots & \boldsymbol{x}_{N-1}\end{bmatrix} \in \mathbb{C}^{M\times N}$ is the estimator that maximizes

$$\ell(\boldsymbol{X}|\boldsymbol{R}) = -N\log(|\boldsymbol{R}|) + \sum_{n=0}^{N-1}\log(g(\boldsymbol{x}_n^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{x}_n)), \tag{19}$$

where the function $g : \mathbb{R}_+ \to \mathbb{R}_+$ is the characteristic generator function determining the particular CES distribution.

In most cases, the generator g is not known. Therefore, a particularly useful robust estimator is Tyler's M-estimator, which exhibits the same statistical properties over the family of CES distributions [16]. It can be derived from (19) by setting $g(x) = x^{-M}$, which gives the function that should be maximized as

$$\ell(\boldsymbol{X}|\boldsymbol{R}) = -N\log(|\boldsymbol{R}|) - M\sum_{n=0}^{N-1}\log(\boldsymbol{x}_n^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{x}_n). \tag{20}$$

Using the Riemannian manifold constructed in Section III, the proposed method is the solution to optimization problem (12) with $\mathcal{D}(\boldsymbol{X}, \boldsymbol{R}) = -\ell(\boldsymbol{X}|\boldsymbol{R})$.

The function (20) maximized by the matrix $\boldsymbol{R}$ is also maximized by $c\boldsymbol{R}$ for any $c \in \mathbb{R}_{++}$. This can be dealt with by normalizing the covariance matrix such that $\operatorname{tr} \boldsymbol{R} = M$ holds for $\boldsymbol{R} \in \mathcal{S}_+(M)$. The proposed method is invariant to the scaling of the summed covariance matrices, and the estimates are therefore normalized after optimization such that $\operatorname{tr}(\hat{\boldsymbol{R}}_x + \hat{\boldsymbol{R}}_v) = M$. However, the relative scaling of $\boldsymbol{R}_x$ compared to $\boldsymbol{R}_v$ is of importance, and will be determined through optimization of (12).

V. Evaluation

To demonstrate the properties of the proposed robust estimator, numerical experiments will be presented in this section. The Riemannian optimization is performed using a Riemannian trust region method implemented in [24]. The proposed estimator is compared against three reference methods. The first method is the subtraction method in (6). The second method is the conventional GEVD method in (11). The third method is a heuristic robust estimator referred to as Robust GEVD, given by (11) when the SCM C_{SC} is replaced by Tyler's M-estimator C_{M} defined in (5). Note that C_{M} is scale-invariant, while the relative scaling of the two matrices matters, so the estimates are scaled before computing the GEVD such that $\operatorname{tr}(C_{\text{M}}(\boldsymbol{V})) = \operatorname{tr}(C_{\text{SC}}(\boldsymbol{V}))$ and $\operatorname{tr}(C_{\text{M}}(\boldsymbol{Y})) = \operatorname{tr}(C_{\text{SC}}(\boldsymbol{Y}))$.

A. Synthesized data

The considered methods are evaluated on synthetic data. The desired signal $\boldsymbol{x}$ is sampled from a circularly symmetric complex Gaussian distribution. The noise $\boldsymbol{v}$ is sampled from a complex t-distribution, which is heavy-tailed for fewer degrees of freedom (DoF), and therefore requires a robust estimator [19]. As the DoF increases, the distribution converges to a Gaussian distribution. Note that the variance of the t-distribution increases as the DoF decreases, for a given trace-normalized covariance matrix.

A random rank-Q signal covariance matrix and full-rank noise covariance matrix are generated by sampling a unitary matrix, representing the eigenvectors, and then Q or M eigenvalues are sampled from a log-uniform distribution between $[10^{-2}, 1]$. The matrices are then normalized such that $\operatorname{tr} \boldsymbol{R}_y = M$, and the relative scaling is set according to $\operatorname{tr} \boldsymbol{R}_x = \operatorname{tr} \boldsymbol{R}_v$. Due to the inherent scale-invariance of the problem, the estimates are normalized such that $\operatorname{tr}(\hat{\boldsymbol{R}}_x + \hat{\boldsymbol{R}}_v) = M$, but their relative scaling is left unchanged. The dimension and rank are set to be $M = 10$ and $Q = 3$ respectively.

The estimation performance as a function of the DoF can be seen in Fig. 1, for the case of $N = 16$ and $N = 64$. The performance is measured as the normalized mean square error (NMSE) between the estimated and true signal covariance matrix, defined as $\|\boldsymbol{R}_x - \hat{\boldsymbol{R}}_x\|_F^2 / \|\boldsymbol{R}_x\|_F^2$. The subtraction method performs poorly for few DoF, and when only few samples are available. In addition, the estimates are not generally positive semi-definite, nor rank-Q. The GEVD method performs poorly for few DoF, but is effective when the distribution is closer to Gaussian. The Robust GEVD performs reasonably well for few samples, but falls behind other methods for many samples.

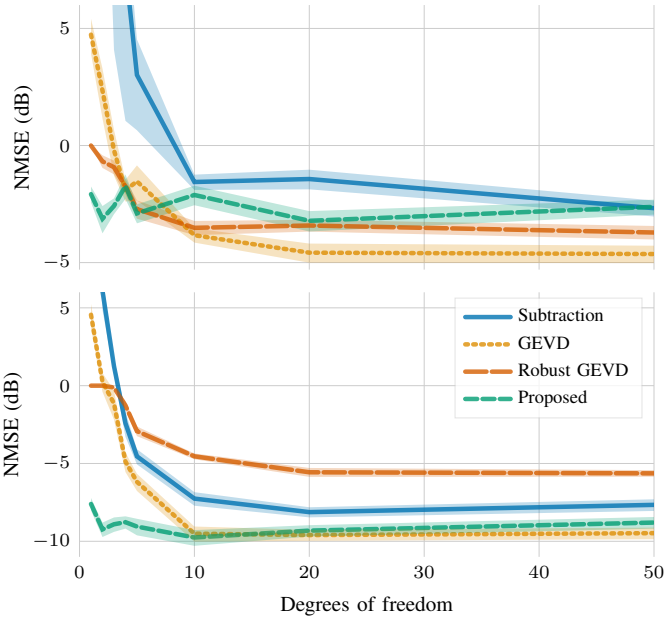

Fig. 1. The figure shows the performance in terms of normalized mean square error of the considered methods as a function of the degrees of freedom of the t-distribution. The top plot shows the result when few ($N = 16$) samples are available, and the bottom plot shows the result when many samples are available ($N = 64$). The result is calculated from 20 realizations, with the shaded area representing one standard deviation of the NMSE.

B. Spatial filtering for audio noise reduction

In this section the considered covariance matrix estimators will be compared using the multichannel Wiener filter (MWF) to remove unwanted impulsive noises from a multichannel audio signal. The MWF is given by the filter $\boldsymbol{W} \in \mathbb{C}^{M \times M}$, which is the optimal solution to the minimum mean square error estimation problem

$$\min_{\boldsymbol{W}} \mathbb{E}[\|\boldsymbol{W}\boldsymbol{y} - \boldsymbol{x}\|_2^2], \tag{21}$$

which can be obtained in closed form as

$$\boldsymbol{W} = \mathbb{E}[\boldsymbol{x}\boldsymbol{y}^{\mathsf{H}}]\,\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}]^{-1} = \boldsymbol{R}_x(\boldsymbol{R}_v + \boldsymbol{R}_x)^{-1}. \tag{22}$$

The filter $\boldsymbol{W}$ is used to obtain an estimate of the signal as $\hat{\boldsymbol{x}} = \boldsymbol{W}\boldsymbol{y}$. The filter is invariant to any scaling applied to both $\boldsymbol{R}_x$ and $\boldsymbol{R}_v$, and therefore the considered robust covariance matrix estimators can be used without modification.

A single source emits white Gaussian noise to $M = 5$ microphones, which represents the desired signal, and two noise sources emits white Gaussian noise to the same microphones. The propagation from sources to microphones is modelled by room impulse responses from the dataset in [25]. Each microphone signal is independently contaminated by impulsive microphone handling noise. Each microphone is contaminated by a different recording of the handling noise, leading to independent noise. The time-domain signals are sampled at 16 kHz, and then transformed into the short-time Fourier domain using a weighted overlap-add filter bank (WOLA) with a block size of 2048 samples, 50 % overlap, and a square-root Hamming window. The results for the frequency bin at

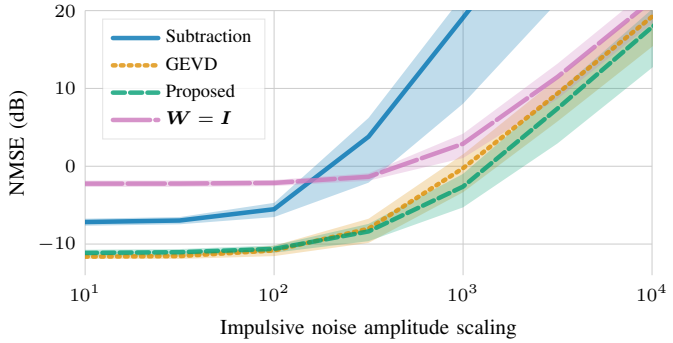

Fig. 2. The normalized mean square error (NMSE) in decibels between the filtered and true signal, for the frequency bin at 500 Hz. The plot for $\boldsymbol{W} = \boldsymbol{I}$ shows the NMSE without performing any filtering, equivalent to the inverse of the signal-to-noise ratio.

500 Hz is shown. A fully adaptive algorithm is outside the scope of this paper, so each set of consecutive 32 time steps are considered separately. There is only one sound source, so the signal covariance matrix is assumed to be rank-1.

The impulsive noise is multiplied by a scaling factor which controls the severity of the contamination, while the amplitude of the Gaussian noise is kept constant. The noise reduction performance is measured with the NMSE, defined as $\|\boldsymbol{W}\boldsymbol{y} - \boldsymbol{x}\|_2^2 / \|\boldsymbol{x}\|_2^2$ where $\boldsymbol{y}$ and $\boldsymbol{x}$ are the noise-contaminated and true signal vector respectively for a given block. In Fig. 2 the mean NMSE as a function of the scaling factor is shown, taken over 16 different segments of 32 time steps each. For this data, the Robust GEVD method consistently estimates the signal covariance matrix as zero, resulting in $\boldsymbol{W} = 0$, and therefore it is excluded from the figure. As the level of contamination from the impulsive noise increases, all considered methods perform worse in absolute terms, but the proposed method displays the least degradation.

VI. Conclusion

The problem of signal and noise covariance matrix estimation has been considered, where the problem was formulated as an unconstrained optimization problem over a Riemannian manifold. The framework allows for a more flexible selection of cost functions, while still enforcing constraints on the estimated covariance matrices, e.g. low-rank constraints. Using the proposed framework, a robust estimator based on Tyler's M-estimator has been derived. Through simulations both on synthetic data and audio signals, the derived robust covariance matrix estimator has been shown to be effective when impulsive noises are present, compared to for instance the GEVD based method.

References

[1] R. Serizel, M. Moonen, B. V. Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

[2] R. Van Rompaey and M. Moonen, "GEVD based speech and noise correlation matrix estimation for multichannel Wiener filter based noise reduction," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2018, pp. 2544–2548.

[3] R. Van Rompaey and M. Moonen, "Distributed adaptive node-specific signal estimation in a wireless sensor network with partial prior knowledge of the desired source steering vector," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2019, pp. 1–5.

[4] R. Ali, T. V. Waterschoot, and M. Moonen, "Using partial a priori knowledge of relative transfer functions to design an MVDR beamformer for a binaural hearing assistive device with external microphones," in *Proc. Int. Congr. Acoust. (ICA)*, Sept. 2019, pp. 6815–6822.

[5] E. Tengan, T. Dietzen, S. Ruiz, M. Alkmim, J. Cardenuto, and T. van Waterschoot, "Speech enhancement using ego-noise references with a microphone array embedded in an unmanned aerial vehicle," in *Proc. Int. Congr. Acoust. (ICA)*. Nov. 2022.

[6] S. Ruiz, T. van Waterschoot, and M. Moonen, "Distributed combined acoustic echo cancellation and noise reduction in wireless acoustic sensor and actuator networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 534–547, 2022.

[7] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, "Fast generation of sound zones using variable span trade-off filters in the DFT-domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 363–378, Dec. 2020.

[8] J. Brunnström, T. van Waterschoot, and M. Moonen, "Sound zone control for arbitrary sound field reproduction methods," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2023.

[9] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.

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

[11] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6182–6189, Dec. 2012.

[12] S. Sra and R. Hosseini, "Conic geometric optimization on the manifold of positive definite matrices," *SIAM J. Optim.*, vol. 25, no. 1, pp. 713–739, Jan. 2015.

[13] S. Sra, "Positive definite matrices and the S-divergence," *Proc. Amer. Math. Soc.*, vol. 144, no. 7, pp. 2787–2797, Oct. 2015.

[14] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *Proc. 32nd Int. Conf. Mach. Learn.* June 2015, pp. 720–729.

[15] A. Han, B. Mishra, P. K. Jawanpuria, and J. Gao, "On Riemannian optimization over positive definite matrices with the Bures-Wasserstein geometry," in *Adv. Neural Inf. Process. Syst.* 2021, vol. 34, pp. 8940–8953.

[16] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Stat.*, vol. 15, no. 1, pp. 234–251, 1987.

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

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

[19] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.

[20] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, "Low-rank optimization on the cone of positive semidefinite matrices," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2327–2351, Jan. 2010.

[21] S. Yatawatta, "Radio interferometric calibration using a Riemannian manifold," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013, pp. 3866–3870.

[22] K. Yang, Y. Shi, and Z. Ding, "Generalized low-rank optimization for topological cooperation in ultra-dense networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2539–2552, May 2019.

[23] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, 1 edition, Mar. 2023.

[24] J. Townsend, N. Koep, and S. Weichwald, "Pymanopt: A Python toolbox for optimization on manifolds using automatic differentiation," *J. Mach. Learn. Res.*, vol. 17, no. 137, pp. 1–5, 2016.

[25] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2021, pp. 151–155.