Robust Rank Constrained Kronecker Covariance Matrix Estimation

A. Breloy, LEME, U. Paris X, France
Y. Sun, School of Industrial Engineering, Purdue University, USA
P. Babu, CARE, IIT Delhi, India
G. Ginolhac, LISTIC, U. SMB, France
D.P. Palomar, DECE, HKUST, Hong Kong

***Abstract*—In this paper, we consider the problem of robustly estimating a structured covariance matrix (CM). Specifically, we focus on CM structures that involve Kronecker products of low rank matrices, which often arise in the context of array processing (e.g. in MIMO-Radar, COLD array, and STAP). To tackle this problem, we derive a new Constrained Tyler's Estimators (CTE), which is defined as the minimizer of the cost function associated to Tyler's estimator under Kronecker structural constraint. Algorithms to compute these new CTEs are derived based on the Majorization-Minimization algorithmic framework.**

***Index Terms*—Adaptive signal processing, covariance matrix estimation, robust estimation, Majorization-Minimization, Kronecker product, low rank.**

I. Introduction

Covariance matrix (CM) estimation is a fundamental problem in adaptive signal processing. In terms of applications purposes, the accuracy of the CM estimate directly impacts the performance of the processes that rely on it. The most common estimator of the CM is the traditional Sample Covariance Matrix (SCM), which is the Maximum Likelihood Estimator (MLE) of the CM in a Gaussian context. Nevertheless, when the samples are heavy-tailed distributed or corrupted by outliers, the SCM fails to provide an accurate estimator [1].

To overcome this issue, non-Gaussian distributions and the robust estimation framework have lately attracted considerable interest [2]. Under this framework, a robust estimation of the CM can be performed using the M-estimators [3], such as Tyler's estimator [4, 5]. These estimators have been extensively studied and used in the modern detection/estimation literature due to their desirable robust properties (see [2] and the references therein).

Nevertheless, traditional M-estimators are not adapted to high dimensional CM estimation with low sample support. For instance, when number of samples K is less than the dimension of the data M, Tyler's M-estimator [4] is undefined. To solve this problem, the current approaches consider either regularize the M-estimator by shrinking it towards some given target, such as the identity matrix [6–9], or constraining the CM to have some structure known in a priori to reduce the numbers of parameters to be estimated [10–12].

In this paper, we follow the second approach and specifically focus on robustly estimating CM that can be expressed as the Kronecker product of (structured) low rank matrices. Indeed, this structure often arise in the context of array processing, such as in MIMO-Radar, COLD array and STAP: the Kronecker product structure generally comes from a redundancy induced by the multiplication of sensors and/or signal emissions [13, 14], while the low rank structure is induced by signals (or interference) being contained in a low dimensional subspace [15].

This work was supported by the Hong Kong RGC 16207814 research grant. This work was also supported by PHOENIX ANR-15-CE23-0012 grant of the French National Agency of Research.

Contribution: To force the rank-constrained Kronecker structure, we derive new Constrained Tyler's Estimators (CTE) as the minimizer of the cost function associated to Tyler's estimator under structural constraint. We consider two specific structure sets:

- $\mathcal{S}_{KPS}$: Matrices given as the Kronecker product of two structured (i.e. low rank plus identity) matrices.
- $\mathcal{S}_{KPLR}$: Matrices given as the sum of the Kronecker product of two low rank matrices and the identity matrix.

The problem is hard to solve since both the objective function and the constraint set are nonconvex. Following the lines of [11, 12], we derive iterative algorithms to compute these new CTE using the (block) Majorization-Minimization (MM) algorithmic framework [16]. The updates have closed-form expression, thus can be computed efficiently, and monotonically decrease the objective value.

Related works: Several MM algorithms have been proposed in [11] to compute CTEs for various structures, e.g., convex, low-rank plus identity, and Kronecker product structure. CTEs with group symmetry structure has been studied in [10], where the associated problem was shown to be geodesically convex and fixed-point iterative algorithms has been derived to compute the solution. In [12], the authors proposed MM algorithms to compute the CM estimator with a low rank Compound Gaussian plus white Gaussian noise structure (note these estimators are not CTE). Recently, structure set $\mathcal{S}_{KPLR}$ has been consider in [14], where an estimator was derived by projecting the SCM onto the set. As regard to this brief state of the art, CTEs were derived under either the Kronecker or low rank structure, but not under the set $\mathcal{S}_{KPS}$ which imposes them simultaneously. The set $\mathcal{S}_{KPLR}$ has been considered, but the derived estimator in [14] is not robust since SCM is known vulnerable to abnormal samples. This paper aims therefore at filling this gap and proposes algorithms that can exploit both structure priors $\mathcal{S}_{KPS}$ and $\mathcal{S}_{KPLR}$ in a robust estimation process.

II. Constrained Tyler's estimator (CTE)

Consider a set of K complex-valued M-dimensional i.i.d. samples $\{\mathbf{z}_k\}$. Tyler's CM estimator [4], also referred to as fixed point estimator [5], is the unique (up to a positive scaling factor) minimizer of the following negative log-likelihood function:

$$\mathcal{L}(\mathbf{\Sigma}) = \frac{M}{K}\sum_{k=1}^{K}\ln\left(\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k\right) + \ln|\mathbf{\Sigma}|. \tag{1}$$

Tyler's estimator can be computed using a fixed point algorithm that requires the number of samples K satisfying $K > M$ [4, 5]. This estimator has the desirable properties of being distribution-free over the class of CES distributions and robust to sample contamination by outliers [2]. However it requires a number of samples $K > M$, and the typical rule of thumb suggest that $K \simeq 2M$ is required in order to reach good estimation performance.

To overcome this issue, prior considerations on the model/system can provide some information about the CM structure. Such prior information can be exploited in the estimation process in order to improve performance at low sample support. A natural approach is to seek an estimator in the structural set that minimizes the cost function $\mathcal{L}(\boldsymbol{\Sigma})$. This leads to the CTE defined as the solution of the following problem:

$$\min_{\boldsymbol{\Sigma}\in\mathcal{S}} \quad \mathcal{L}(\boldsymbol{\Sigma}) \tag{2}$$

where $\mathcal{S}$ is a set of matrices possessing some prior structure (e.g. Toeplitz, persymmetric). While for some specific structures such as group symmetry, the existence and uniqueness of their CTE can be guaranteed [10], solving Problem (2) under a majority of the structures of practical interest remains a challenging task due to the nonconvexity of $\mathcal{L}$ and the possibly nonconvexity of $\mathcal{S}$. Therefore, instead of attempting to find the global optimal of (2), we focus on deriving efficient algorithms in computing at least its local solution or the two aforementioned structural sets $\mathcal{S}_{KPS}$ and $\mathcal{S}_{KPLR}$.

III. Block Majorization-Minimization framework

To solve further-coming optimization problems, we rely on the block majorization-minimization (MM) algorithmic framework, which is briefly stated below. For more complete information, we refer the reader to [16]. Consider the following optimization problem:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x}\in\mathcal{X}, \end{array} \tag{3}$$

where the optimization variable $\mathbf{x}$ can be partitioned into m blocks as $\mathbf{x}=\left(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(m)}\right)$, with each n_i-dimensional block $\mathbf{x}^{(i)}\in\mathcal{X}_i$ and $\mathcal{X}=\prod_{i=1}^{m}\mathcal{X}_i$. At the $(t+1)$-th iteration, the i-th block $\mathbf{x}^{(i)}$ is updated by solving the following problem:

$$\begin{array}{ll} \underset{\mathbf{x}^{(i)}}{\text{minimize}} & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \\ \text{subject to} & \mathbf{x}^{(i)}\in\mathcal{X}_i, \end{array} \tag{4}$$

with $i=(t \bmod m)+1$ (so blocks are updated in cyclic order) and the continuous surrogate function $g_i(\cdot|\mathbf{x}_t)$ satisfying the following properties:

$$\begin{array}{rcl} f(\mathbf{x}_t) & = & g_i\left(\mathbf{x}_t^{(i)}|\mathbf{x}_t\right), \\ f\left(\mathbf{x}_t^{(1)},\dots,\mathbf{x}^{(i)},\dots,\mathbf{x}_t^{(m)}\right) & \leq & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)\ \forall\mathbf{x}^{(i)}\in\mathcal{X}_i, \\ f'\left(\mathbf{x}_t;\mathbf{d}_i^0\right) & = & g_i'\left(\mathbf{x}_t^{(i)};\mathbf{d}_i|\mathbf{x}_t\right) \\ & & \forall\mathbf{x}_t^{(i)}+\mathbf{d}_i\in\mathcal{X}_i, \\ & & \mathbf{d}_i^0\triangleq(\mathbf{0};\dots;\mathbf{d}_i;\dots;\mathbf{0}), \end{array}$$

where $f'(\mathbf{x};\mathbf{d})$ stands for the directional derivative at $\mathbf{x}$ along $\mathbf{d}$. In short, at each iteration, the block MM algorithm updates the variables in one block by minimizing a tight upperbound of the function while keeping the other blocks fixed.

IV. Low Rank plus Identity Kronecker product

A. Problem statement

We consider the set of matrices that can be expressed as the Kronecker product of two structured (i.e. low rank plus identity) matrices, defined as:

$$\mathcal{S}_{KPS}=\left\{\boldsymbol{\Sigma}\in\mathbb{C}^{M^2}\left|\begin{array}{l}\boldsymbol{\Sigma}=\left(\mathbf{A}+\sigma^2\mathbf{I}\right)\otimes\left(\mathbf{B}+\sigma^2\mathbf{I}\right), \\ \mathbf{A}\in\mathbb{C}^{P^2},\ \mathbf{B}\in\mathbb{C}^{Q^2}\ , \\ \mathbf{A}\succeq\mathbf{0},\ \mathbf{B}\succeq\mathbf{0}\ , \\ \text{rank}(\mathbf{A})\leq R_A,\ \text{rank}(\mathbf{B})\leq R_B\end{array}\right.\right\}$$

Algorithm 1 "**KPS - MM**": Block MM algorithm for Robust estimation of KPS structured covariance matrix

1: Form a starting point $\left\{\boldsymbol{\Sigma}_A^{t=0},\boldsymbol{\Sigma}_B^{t=0}\right\}$.
2: **repeat**
3: $\quad t\leftarrow t+1$
4: $\quad$ Update $\boldsymbol{\Sigma}_A^t$ with (16).
5: $\quad$ Update $\boldsymbol{\Sigma}_B^t$ with (17).
6: **until** Some convergence criterion is met.

Note that an element of $\mathcal{S}_{KPS}$ is structured as a low rank plus identity matrix, but that the rank of its low rank component can not be inferred from R_A and R_B. The CTE corresponding to $\mathcal{S}_{KPS}$ is solution of the problem:

$$\underset{\boldsymbol{\Sigma}\in\mathcal{S}_{KPS}}{\text{min.}} \quad \mathcal{L}(\boldsymbol{\Sigma}). \tag{5}$$

Substituting

$$\begin{cases}\boldsymbol{\Sigma}_A=\mathbf{A}+\sigma^2\mathbf{I} \\ \boldsymbol{\Sigma}_B=\mathbf{B}+\sigma^2\mathbf{I} \\ \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_A\otimes\boldsymbol{\Sigma}_B \\ \mathbf{Z}_k=\text{uvec}(\mathbf{z}_k)\in Q\times P\end{cases} \tag{6}$$

into the objective function and applying the identity $\mathbf{z}_k^H\left(\boldsymbol{\Sigma}_A\otimes\boldsymbol{\Sigma}_B\right)^{-1}\mathbf{z}_k=\text{Tr}\left(\boldsymbol{\Sigma}_A^{-1}\mathbf{Z}_k^H\boldsymbol{\Sigma}_B^{-1}\mathbf{Z}_k\right)$ leads to

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Sigma})=\mathcal{L}(\boldsymbol{\Sigma}_A,\boldsymbol{\Sigma}_B)=&\frac{pq}{K}\sum_{k=1}^{K}\ln\left(\text{Tr}\left(\boldsymbol{\Sigma}_A^{-1}\mathbf{Z}_k^H\boldsymbol{\Sigma}_B^{-1}\mathbf{Z}_k\right)\right) \\ &+q\ln|\boldsymbol{\Sigma}_A|+p\ln|\boldsymbol{\Sigma}_B|.\end{aligned} \tag{7}$$

Hence to obtain the CTE, we aim at solving:

$$\begin{array}{ll}\underset{\boldsymbol{\Sigma}_A\succeq\mathbf{0},\boldsymbol{\Sigma}_B\succeq\mathbf{0}}{\min} & \mathcal{L}(\boldsymbol{\Sigma}_A,\boldsymbol{\Sigma}_B) \\ \text{s.t.} & \boldsymbol{\Sigma}_A=\mathbf{A}+\sigma^2\mathbf{I}\ ,\ \boldsymbol{\Sigma}_B=\mathbf{B}+\sigma^2\mathbf{I} \\ & \mathbf{A}\succeq\mathbf{0}\ ,\ \mathbf{B}\succeq\mathbf{0} \\ & \text{rank}(\mathbf{A})\leq R_A\ ,\ \text{rank}(\mathbf{B})\leq R_B\end{array} \tag{8}$$

Following the block MM methodology, we partition the variables as $\{\boldsymbol{\Sigma}_A,\boldsymbol{\Sigma}_B\}$ and derive an algorithm that updates these two blocks in cyclic order, by minimizing surrogates functions (upperbounds of the objective). This estimation procedure is referred to as "KPS - MM" with corresponding algorithm summed up in table Algorithm 1.

Note that this algorithm falls, as a special case, into the constrained iterations proposed in (64) of [11]. However we provide here some details since our case allows for obtaining closed form updates and that some of the results/methods are useful for the next sections.

B. Derivation of KPS - MM

1) Step 1: Update $\boldsymbol{\Sigma}_A$ for fixed $\boldsymbol{\Sigma}_B^t$:
For fixed parameter $\boldsymbol{\Sigma}_B^t$, we have the following objective function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Sigma}_A)=&\frac{pq}{K}\sum_{k=1}^{K}\ln\left(\text{Tr}\left(\boldsymbol{\Sigma}_A^{-1}\mathbf{Z}_k^H\boldsymbol{\Sigma}_B^{-t}\mathbf{Z}_k\right)\right) \\ &+q\ln|\boldsymbol{\Sigma}_A|+\text{const.}\end{aligned} \tag{9}$$

with shortened notation $\boldsymbol{\Sigma}_B^{-t}=\left(\boldsymbol{\Sigma}_B^t\right)^{-1}$ (the same convention applies hereafter). To obtain the update $\boldsymbol{\Sigma}_A^{t+1}$, we need to solve the problem

$$\begin{array}{ll}\underset{\boldsymbol{\Sigma}_A\succeq\mathbf{0}}{\min} & \mathcal{L}(\boldsymbol{\Sigma}_A) \\ \text{s.t.} & \boldsymbol{\Sigma}_A=\mathbf{A}+\sigma^2\mathbf{I} \\ & \mathbf{A}\succeq\mathbf{0}\ ,\ \text{rank}(\mathbf{A})\leq R_A\end{array} \tag{10}$$

This problem is nonconvex and has no closed form solution. We therefore minimize instead the surrogate function $g\left(\mathbf{\Sigma}_A|\mathbf{\Sigma}_A^t,\mathbf{\Sigma}_B^t\right)$ given in the following proposition.

Proposition 1 *(Eq. 20 of [16]) The objective function $\mathcal{L}\left(\mathbf{\Sigma}_A\right)$ can be upperbounded at $\mathbf{\Sigma}_A^t$ by the following surrogate function:*

$$\begin{aligned} g\left(\mathbf{\Sigma}_A|\mathbf{\Sigma}_A^t,\mathbf{\Sigma}_B^t\right) = & \frac{pq}{K}\sum_{k=1}^{K}\frac{Tr\left(\mathbf{\Sigma}_A^{-1}\mathbf{Z}_k^H\mathbf{\Sigma}_B^{-t}\mathbf{Z}_k\right)}{Tr\left(\mathbf{\Sigma}_A^{-t}\mathbf{Z}_k^H\mathbf{\Sigma}_B^{-t}\mathbf{Z}_k\right)} \\ & + q\ln|\mathbf{\Sigma}_A| + \text{const.} \end{aligned} \tag{11}$$

Equality is achieved at $\mathbf{\Sigma}_A=\mathbf{\Sigma}_A^t$. •

Denote the matrix

$$\mathbf{M}_A^t=\frac{P}{K}\sum_{k=1}^{K}\frac{\mathbf{Z}_k^H\mathbf{\Sigma}_B^{-t}\mathbf{Z}_k}{\text{Tr}\left(\mathbf{\Sigma}_A^{-t}\mathbf{Z}_k^H\mathbf{\Sigma}_B^{-t}\mathbf{Z}_k\right)} \tag{12}$$

To obtain the update $\mathbf{\Sigma}_A^{t+1}$, one has now to solve the following problem:

$$\begin{aligned} \min_{\mathbf{\Sigma}_A\succeq\mathbf{0}} \quad & \ln|\mathbf{\Sigma}_A|+\text{Tr}\left(\mathbf{\Sigma}_A^{-1}\mathbf{M}_A^t\right) \\ \text{s.t.} \quad & \mathbf{\Sigma}_A=\mathbf{A}+\sigma^2\mathbf{I} \\ & \text{rank}\left(\mathbf{A}\right)\le R_A, \end{aligned} \tag{13}$$

which is still nonconvex but admits an unique solution, as given in the following proposition.

Proposition 2 *([17, Section III.B.1.]) Define the SVD of the matrix $\mathbf{M}_A^t$ as:*

$$\mathbf{M}_A^t=\sum_{p=1}^{P}\lambda_p^A\ \mathbf{v}_p^A\ (\mathbf{v}_p^A)^H \tag{14}$$

and threshold $\tilde{\lambda}_p^A$ as

$$\tilde{\lambda}_p^A=\begin{cases}\max(\lambda_p^A,\sigma^2) & \text{for } p\in[\![1,R_a]\!]\\ \sigma^2 & \text{for } p\in[\![R_a+1,P]\!],\end{cases} \tag{15}$$

the unique solution of (13) yields the update $\mathbf{\Sigma}_A^{t+1}$ as:

$$\mathbf{\Sigma}_A^{t+1}=\sum_{p=1}^{P}\tilde{\lambda}_p^A\ \mathbf{v}_p^A\ (\mathbf{v}_p^A)^H. \tag{16}$$

2) Step 2: Update $\mathbf{\Sigma}_B$ for fixed $\mathbf{\Sigma}_A^{t+1}$: One can observe that $\mathbf{\Sigma}_B$ plays a similar role as $\mathbf{\Sigma}_A$ in the objective function (7). Apply Propositions 1 and 2 we obtain the update $\mathbf{\Sigma}_B^{t+1}$ as:

$$\mathbf{\Sigma}_B=\sum_{q=1}^{Q}\tilde{\lambda}_q^B\ \mathbf{v}_q^B\ (\mathbf{v}_q^B)^H \tag{17}$$

where we define matrix $\mathbf{M}_B^t$ and its SVD as:

$$\begin{aligned} \mathbf{M}_B^t \quad &= \quad \frac{Q}{K}\sum_{k=1}^{K}\frac{\mathbf{Z}_k\mathbf{\Sigma}_A^{-(t+1)}\mathbf{Z}_k^H}{\text{Tr}\left(\mathbf{\Sigma}_A^{-(t+1)}\mathbf{Z}_k^H\mathbf{\Sigma}_B^{-t}\mathbf{Z}_k\right)} \\ &\overset{SVD}{=} \quad \sum_{q=1}^{Q}\lambda_q^B\ \mathbf{v}_q^B\ (\mathbf{v}_q^B)^H \end{aligned}, \tag{18}$$

and the threshold $\tilde{\lambda}_q^B$ as

$$\tilde{\lambda}_q^B=\begin{cases}\max(\lambda_q^B,\sigma^2) & \text{for } q\in[\![1,R_b]\!]\\ \sigma^2 & \text{for } q\in[\![R_b+1,Q]\!].\end{cases} \tag{19}$$

Algorithm 2 "**KPLR - MM**": Block MM algorithm for Robust estimation of KPLR structured covariance matrix

1: Form a starting point $\left\{\mathbf{\Sigma}_A^{t=0},\mathbf{\Sigma}_B^{t=0}\right\}$.
2: **repeat**
3: $t\leftarrow t+1$
4: Update $\mathbf{D}_A^t$ with (31).
5: Update $\mathbf{D}_B^t$ with (32).
6: Update $\mathbf{U}_B^t$ with (37).
7: Update $\mathbf{U}_A^t$ with (39).
8: **until** Some convergence criterion is met.

V. Low Rank Kronecker Product plus Identity

In this section, we consider the set of matrices that are expressed as the sum of the Kronecker product of two low rank matrices and the identity matrix, defined as:

$$\mathcal{S}_{KPLR}=\left\{\mathbf{\Sigma}\in\mathbb{C}^{M^2}\left|\begin{array}{l}\mathbf{\Sigma}=\mathbf{A}\otimes\mathbf{B}+\sigma^2\mathbf{I},\\ \mathbf{A}\in\mathbb{C}^{P^2},\ \mathbf{B}\in\mathbb{C}^{Q^2}\ ,\\ \mathbf{A}\succeq\mathbf{0},\ \mathbf{B}\succeq\mathbf{0}\ ,\\ \text{rank}\left(\mathbf{A}\right)\le R_A,\ \text{rank}\left(\mathbf{B}\right)\le R_B\end{array}\right.\right\}.$$

The CTE corresponding to $\mathcal{S}_{KPLR}$ is solution of the problem

$$\min_{\mathbf{\Sigma}\in\mathcal{S}_{KPLR}}\quad\mathcal{L}\left(\mathbf{\Sigma}\right). \tag{20}$$

To solve this problem, we parameterize the matrices $\mathbf{A}$ and $\mathbf{B}$ by their SVD, i.e., their unitary eigenvectors basis $\mathbf{U}$ and diagonal matrix of eigenvalues $\mathbf{D}$ as:

$$\begin{cases}\mathbf{A}=\mathbf{U}_A\mathbf{D}_A\mathbf{U}_A^H & ,\ \mathbf{U}_A^H\mathbf{U}_A=\mathbf{I}_P\ ,\ \mathbf{D}_A=\text{diag}\{a_p\}\\ \mathbf{B}=\mathbf{U}_B\mathbf{D}_B\mathbf{U}_B^H & ,\ \mathbf{U}_B^H\mathbf{U}_B=\mathbf{I}_Q\ ,\ \mathbf{D}_B=\text{diag}\{b_q\}\end{cases}$$

Note that the low rank structure of both $\mathbf{A}$ and $\mathbf{B}$ impose $a_p=0\ \forall p\in[\![R_A+1,P]\!]$ and $b_q=0\ \forall q\in[\![R_B+1,Q]\!]$.

Substituting

$$\begin{cases}\mathbf{\Sigma}=\left(\mathbf{U}_A\otimes\mathbf{U}_B\right)\left(\mathbf{D}_A\otimes\mathbf{D}_B+\sigma^2\mathbf{I}\right)\left(\mathbf{U}_A^H\otimes\mathbf{U}_B^H\right)\\ \mathbf{\Sigma}^{-1}=\left(\mathbf{U}_A\otimes\mathbf{U}_B\right)\left(\mathbf{D}_A\otimes\mathbf{D}_B+\sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{U}_A^H\otimes\mathbf{U}_B^H\right)\end{cases}$$

into the objective function leads to

$$\begin{aligned} &\mathcal{L}\left(\mathbf{\Sigma}\right)=\mathcal{L}\left(\mathbf{D}_A,\mathbf{D}_B,\mathbf{U}_A,\mathbf{U}_B\right)=\ln|\ \mathbf{D}_A\otimes\mathbf{D}_B+\sigma^2\mathbf{I}\ |\ + \\ &\frac{M}{K}\sum_{k=1}^{K}\ln\left(\mathbf{z}_k^H\left(\mathbf{U}_A\otimes\mathbf{U}_B\right)\left(\mathbf{D}_A\otimes\mathbf{D}_B+\sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{U}_A^H\otimes\mathbf{U}_B^H\right)\mathbf{z}_k\right) \end{aligned} \tag{21}$$

Therefore, to obtain the CTE, we aim at solving:

$$\begin{aligned} \min_{\mathbf{D}_A,\mathbf{D}_B,\mathbf{U}_A,\mathbf{U}_B} \quad & \mathcal{L}\left(\mathbf{D}_A,\mathbf{D}_B,\mathbf{U}_A,\mathbf{U}_B\right) \\ \text{s.t.} \quad & \mathbf{D}_A=\text{diag}\{a_p\}\succeq\mathbf{0}, \\ & a_p=0\ \forall p\in[\![R_A+1,P]\!], \\ & \mathbf{D}_B=\text{diag}\{b_q\}\succeq\mathbf{0}, \\ & b_q=0\ \forall q\in[\![R_B+1,Q]\!], \\ & \mathbf{U}_A^H\mathbf{U}_A=\mathbf{I}_P\ ,\mathbf{U}_B^H\mathbf{U}_B=\mathbf{I}_Q. \end{aligned} \tag{22}$$

Following the block MM methodology, we derive an algorithm that updates the variables $\mathbf{D}_A$, $\mathbf{D}_B$, $\mathbf{U}_A$ and $\mathbf{U}_B$ in cyclic order, by minimizing surrogate functions (upperbounds) of the objective. This estimation procedure is referred to as "KPLR - MM" with corresponding algorithm summed up in table Algorithm 2.

A. Derivation of KPLR - MM

To lighten notation, we omit the reference on t for variables that are fixed in the considered block. For example, in the update of $\mathbf{D}_A$, we denote $\mathbf{U}_A = \mathbf{U}_A^t$, $\mathbf{U}_B = \mathbf{U}_B^t$ and $\mathbf{D}_B = \mathbf{D}_B^t$. Due to lack of space, the proof of some propositions is postponed to a forthcoming full length publication.

1) Step 1: Update $\mathbf{D}_A$ with other variables fixed: In addition, we denote the rotated samples

$$\tilde{\mathbf{z}}_k = \left(\mathbf{U}_A^H \otimes \mathbf{U}_B^H\right) \mathbf{z}_k. \tag{23}$$

The objective function then can be re-expressed as

$$\begin{aligned}\mathcal{L}\left(\{\alpha_w\}\right) = \quad & \frac{M}{K}\sum_{k=1}^{K} \ln\left(\tilde{\mathbf{z}}_k^H \text{diag}\left(\{\alpha_{(p,q)}^{-1}\}\right)\tilde{\mathbf{z}}_k\right) \\ & + \ln\left|\text{diag}\left(\{\alpha_{(p,q)}\}\right)\right|\end{aligned} \tag{24}$$

with shortened notation $\alpha_{(p,q)}$ referring to:

$$\begin{aligned}&\alpha_{(p,q)} = \alpha_{(p-1)\times Q+q} \triangleq a_p b_q + \sigma^2 \\ &\text{for } p \in [\![1, P]\!] \text{ and } q \in [\![1, Q]\!]\end{aligned} \tag{25}$$

This objective is expanded as ($[\ ]^2$ stands for "squared norm of"):

$$\begin{aligned}\mathcal{L}\left(\{a_p\}\right) = & \frac{M}{K}\sum_{k=1}^{K}\ln\left(\sum_{p=1}^{P}\sum_{q=1}^{Q}\frac{[\tilde{\mathbf{z}}_k]^2_{(p-1)\times Q+q}}{a_p b_q + \sigma^2}\right) \\ & + \sum_{p=1}^{P}\sum_{q=1}^{Q}\ln\left(a_p b_q + \sigma^2\right).\end{aligned} \tag{26}$$

The objective is separable in the a_p's and yields, for a given index $w \in [\![1, R_A]\!]$, $\mathcal{L}\left(\{a_p\}\right) = \sum_{w=1}^{R_A}\mathcal{L}\left(a_w\right)$ with

$$\begin{aligned}\mathcal{L}\left(a_w\right) = & \frac{M}{K}\sum_{k=1}^{K}\ln\left(\sum_{q=1}^{Q}\frac{[\tilde{\mathbf{z}}_k^{w,q}]^2}{a_w b_q + \sigma^2} + \gamma_w\right) \\ & + \sum_{q=1}^{Q}\ln\left(a_w b_q + \sigma^2\right) + \text{const.}\end{aligned} \tag{27}$$

where γ_w is a constant term, not depending on a_w, defined as

$$\gamma_w = \sum_{p=1, p\neq w}^{P}\ \sum_{q=1}^{Q}\frac{[\tilde{\mathbf{z}}_k]^2_{(p-1)\times Q+q}}{a_p b_q + \sigma^2} \tag{28}$$

and where we used the shortened notation

$$[\tilde{\mathbf{z}}_k^{w,q}]^2 = [\tilde{\mathbf{z}}_k]^2_{(w-1)\times Q+q} \tag{29}$$

Optimizing $\mathcal{L}\left(a_w\right)$ w.r.t. a_w is a nonconvex problem that has no closed form solution. To obtain an update for these parameters, we build an upperbound of $\mathcal{L}\left(a_w\right)$ as follows.

Proposition 3 *The objective $\mathcal{L}\left(a_w\right)$ can be upperbounded by the surrogate function*

$$g\left(a_w | a_w^t\right) = \mathcal{A}_w \ln\left(\omega_w a_w + \beta_w\right) - \mathcal{K}_w \ln\left(a_w\right) + \text{const.} \tag{30}$$

where $\mathcal{A}_w$, ω_w, β_w and $\mathcal{K}_q$ are functions of a_w^t defined in Appendix A. Equality is achieved at $a_w = a_w^t$. •

The update for the parameter a_w can be obtained thanks to the following proposition.

Proposition 4 *([12, Prop. 2]) The surrogate function $g\left(a_w | a_w^t\right)$ is quasiconvex and has a unique minimizer that provides the update*

$$a_w^{t+1} = \frac{\mathcal{K}_w \beta_w}{(\mathcal{A}_w - \mathcal{K}_w)\omega_w} \tag{31}$$

•

2) Step 2 : Update $\mathbf{D}_B$ with other variables fixed: Note the variables $\{a_p\}$ and $\{b_q\}$ play a similar role in the objective function. Let $w \in [\![1, R_B]\!]$, we can therefore adapt Proposition 3 and 4 to obtain the update of b_w as:

$$b_w^{t+1} = \frac{\mathcal{K}'_w \beta'_w}{(\mathcal{A}'_w - \mathcal{K}'_w)\omega'_w} \tag{32}$$

The constants $\mathcal{K}'_w$, $\mathcal{A}'_w$, β'_w and ω'_w are defined in Appendix A.

3) Update $\mathbf{U}_B$ with other variables fixed: To lighten notation, we omit the reference on t for fixed variables in this part. Hence we denote $\mathbf{U}_A = \mathbf{U}_A^t$, $\mathbf{D}_A = \mathbf{D}_A^{t+1}$ and $\mathbf{D}_B = \mathbf{D}_B^{t+1}$. Substituting

$$\begin{cases}\mathbf{Z}_k = \text{uvec}\left(\mathbf{z}_k\right) \in Q \times P \\ \left(\mathbf{U}_A^H \otimes \mathbf{U}_B^H\right)\mathbf{z}_k = \text{vec}\left(\mathbf{U}_B^H \mathbf{Z}_k \mathbf{U}_A\right)\end{cases} \tag{33}$$

into (21) gives the following objective:

$$\begin{aligned}&\mathcal{L}\left(\mathbf{U}_B\right) = \text{const.}+ \\ &\frac{M}{K}\sum_{k=1}^{K}\ln\left(\text{vec}\left(\mathbf{U}_B^H\mathbf{Z}_k\mathbf{U}_A\right)^H\left(\mathbf{D}_A\otimes\mathbf{D}_B + \sigma^2\mathbf{I}\right)^{-1}\text{vec}\left(\mathbf{U}_B^H\mathbf{Z}_k\mathbf{U}_A\right)\right)\end{aligned} \tag{34}$$

Once again, this objective can not be easily minimized directly w.r.t. $\mathbf{U}_B$ under unitary constraint. Following the block-MM methodology, we upperbound it using the following proposition.

Proposition 5 *The objective function $\mathcal{L}(\mathbf{U}_B)$ can be upperbounded at $\mathbf{U}_B^t$ by the surrogate function*

$$g\left(\mathbf{U}_B | \mathbf{U}_B^t\right) = Tr\left[(\mathbf{W}_B^t)^H\mathbf{U}_B\right] + Tr\left[\mathbf{U}_B^H\mathbf{W}_B^t\right] + \text{const.} \tag{35}$$

$\mathbf{W}_B^t$ is defined in Appendix A. Equality is achieved at $\mathbf{U}_B = \mathbf{U}_B^t$. •

The update for the parameter $\mathbf{U}_B$ is the solution of

$$\begin{aligned}\min_{\mathbf{U}_B}\quad & g\left(\mathbf{U}_B | \mathbf{U}_B^t\right) \\ \text{s.t.}\quad & \mathbf{U}_B^H\mathbf{U}_B = \mathbf{I}_Q,\end{aligned} \tag{36}$$

and is provided in following Proposition.

Proposition 6 *([18, Prop. 7]) The problem of minimizing the surrogate function $g\left(\mathbf{U}_B | \mathbf{U}_B^t\right)$ under constraint $\mathbf{U}_B^H\mathbf{U}_B = \mathbf{I}_Q$ has an optimal solution, that lead to the update*

$$\mathbf{U}_B^{t+1} = \mathbf{V}_L \mathbf{V}_R^H \tag{37}$$

where $\mathbf{V}_L$ and $\mathbf{V}_R^H$ are the left and right singular vectors of the matrix $-\mathbf{W}_B^t$ defined in Appendix A. •

4) Update $\mathbf{U}_A$ with other variables fixed:

The update of $\mathbf{U}_A$ is obtained in a similar way as the one of $\mathbf{U}_B$, thanks to the following propositions.

Proposition 7 *The objective function $\mathcal{L}(\mathbf{U}_A)$ can be upperbounded at $\mathbf{U}_B^t$ by the surrogate function*

$$g\left(\mathbf{U}_A | \mathbf{U}_A^t\right) = Tr\left[(\mathbf{W}_A^t)^H\mathbf{U}_A\right] + Tr\left[\mathbf{U}_A^H\mathbf{W}_A^t\right] + \text{const.} \tag{38}$$

$\mathbf{W}_A^t$ is defined in Appendix A. Equality is achieved at $\mathbf{U}_A = \mathbf{U}_A^t$. •

Proposition 8 *([18, Prop. 7]) The problem of minimizing the surrogate function $g\left(\mathbf{U}_A | \mathbf{U}_A^t\right)$ under constraint $\mathbf{U}_A^H\mathbf{U}_A = \mathbf{I}_P$ has an optimal solution, that lead to the update*

$$\mathbf{U}_A^{t+1} = \mathbf{V}_L \mathbf{V}_R^H \tag{39}$$

where $\mathbf{V}_L$ and $\mathbf{V}_R^H$ respectively the left and right singular vectors of the matrix $-\mathbf{W}_A^t$, defined in Appendix A. •

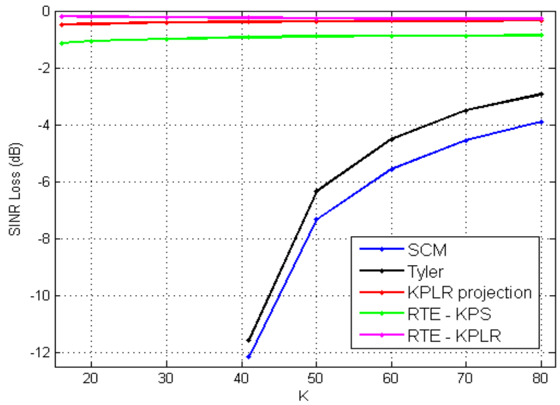

Fig. 1. SINR-Loss versus number of samples of adaptive filters build from various CM estimators. $\boldsymbol{\Sigma} \in \mathcal{S}_{KPLR}$. $\mathbf{A}$ and $\mathbf{B}$ are constructed by truncating the SVD of Toeplitz matrices of correlation parameter $\rho_a = 0.9$ and $\rho_b = 0.95$ with $P = 10$, $R_a = 4$, $Q = 4$, $R_b = 2$. $\sigma^2 = 1$ and matrices are scaled so that $\text{Tr}\left[\mathbf{A} \otimes \mathbf{B}\right] / R_a R_b = 10$dB. Noise is Compound Gaussian [2], i.e. $\mathbf{z} \stackrel{d}{=} \sqrt{\tau}\mathbf{n}$ where $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and τ follows a Gamma distribution of shape parameter $\nu = 1$ and scale parameter $1/\nu$.

VI. Validation simulation

We study the mean SINR-Loss [13]: the ratio between the output SNR of an adaptive filter $\hat{\mathbf{w}}_{lr} = \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{d}$ (build from any estimator $\hat{\boldsymbol{\Sigma}}$) and the output SNR of the optimal non-adaptive filter $\mathbf{w} = \boldsymbol{\Sigma}^{-1}\mathbf{d}$. We consider a scenario where the actual CM belongs to $\mathcal{S}_{KPLR}$ and we compare the performance of the following estimators: the SCM, Tyler's estimator, The projection of the SCM onto $\mathcal{S}_{KPLR}$ proposed in [14] and the two estimators proposed in this paper. Figure 1 presents the mean SINR-Loss versus the number of secondary data K. One can observe that estimators that do not exploit the structure prior reach the lowest performance. Tyler's estimator still performs better than the SCM since the simulated noise is not Gaussian. RTE-KPLR reaches the best performances, which was to be expected since it can both exploit the structure prior and handle non-Gaussian. The estimator from [14] reaches good performance (close to RTE-KPLR) in the considered context, even if it is build from the SCM. RTE-KPS also reaches acceptable performance even if it does not account for the actual structure prior.

Appendix A
Definition of constants of Algorithm KPLR - MM

A. Constants for steps 1 and 2

$\boldsymbol{\Sigma}^{-t}$ denotes the inverse estimated CM at the current step.

$$\zeta_k = \left(\mathbf{z}_k^H \boldsymbol{\Sigma}^{-t} \mathbf{z}_k\right)^{-1} \tag{40}$$

$$\begin{cases} \kappa_q^w = \dfrac{M}{K} \displaystyle\sum_{k=1}^{K} \zeta_k \dfrac{a_w^t b_q \left[\tilde{\mathbf{z}}_k^{w,q}\right]^2}{a_w^t b_q + \sigma^2} \\ \mathcal{K}_w = \displaystyle\sum_{q=1}^{Q} \kappa_q^w \\ \mathcal{A}_w = \displaystyle\sum_{q=1}^{Q} \left(\kappa_q^w + 1\right) = \mathcal{K}_w + Q \\ \omega_w = \dfrac{1}{\mathcal{A}_w} \displaystyle\sum_{q=1}^{Q} \dfrac{(\kappa_q^w + 1) b_q}{(a_w^t b_q + \sigma^2)} \\ \beta_w = \dfrac{1}{\mathcal{A}_w} \displaystyle\sum_{q=1}^{Q} \dfrac{(\kappa_q^w + 1)\sigma^2}{(a_w^t b_q + \sigma^2)} \end{cases} \qquad \begin{cases} \kappa_p^{w\prime} = \dfrac{M}{K} \displaystyle\sum_{k=1}^{K} \zeta_k \dfrac{a_p b_w^t \left[\tilde{\mathbf{z}}_k^{w,q}\right]^2}{a_p b_w^t + \sigma^2} \\ \mathcal{K}'_w = \displaystyle\sum_{p=1}^{P} \kappa_p^{w\prime} \\ \mathcal{A}'_w = \displaystyle\sum_{p=1}^{P} \left(\kappa_p^{w\prime} + 1\right) = \mathcal{K}'_w + P \\ \omega'_w = \dfrac{1}{\mathcal{A}'_w} \displaystyle\sum_{p=1}^{P} \dfrac{(\kappa_p^{w\prime} + 1) a_p}{(a_p b_w^t + \sigma^2)} \\ \beta'_w = \dfrac{1}{\mathcal{A}_w} \displaystyle\sum_{p=1}^{P} \dfrac{(\kappa_p^{w\prime} + 1)\sigma^2}{(a_p b_w^t + \sigma^2)} \end{cases}$$

B. Constants for steps 3 and 4

$$\begin{cases} \tilde{\mathbf{Z}}_k = \sqrt{\zeta_k}\mathbf{Z}_k \\ \boldsymbol{\Lambda}_p = a_p \mathbf{D}_B + \sigma^2 \mathbf{I} \end{cases}$$

$$\begin{cases} \tilde{\mathbf{Z}}_k^A = \tilde{\mathbf{Z}}_k \mathbf{U}_A \\ \mathbf{X}_p^A = \displaystyle\sum_{k=1}^{K} \left[\tilde{\mathbf{Z}}_k^A\right]_{:,p} \left[\tilde{\mathbf{Z}}_k^A\right]_{:,p}^H \\ \mathbf{M}_q = \displaystyle\sum_{p=1}^{P} \left[\boldsymbol{\Lambda}_p^A\right]_{q,q} \mathbf{X}_p^A \\ \mathbf{W}_B^t = \left[\mathbf{G}_1 \mathbf{u}_1^{B(t)} \ldots \mathbf{G}_Q \mathbf{u}_Q^{B(t)}\right] \\ \mathbf{G}_q = \mathbf{M}_q - \lambda_{max}^{(\mathbf{M}_q)} \mathbf{I} \end{cases} \qquad \begin{cases} \tilde{\mathbf{Z}}_k^B = \mathbf{U}_B^H \tilde{\mathbf{Z}}_k \\ \mathbf{M}_p = \displaystyle\sum_{k=1}^{K} (\tilde{\mathbf{Z}}_k^B)^H \boldsymbol{\Lambda}_p^{-1} \tilde{\mathbf{Z}}_k^B \\ \mathbf{W}_A^t = \left[\mathbf{G}_1 \mathbf{u}_1^{A(t)} \ldots \mathbf{G}_P \mathbf{u}_P^{A(t)}\right] \\ \mathbf{G}_p = \mathbf{M}_p - \lambda_{max}^{(\mathbf{M}_p)} \mathbf{I} \end{cases}$$

References

[1] P. Huber, *Robust Statistics*, ser. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series. Wiley, 2004.

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

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

[4] D. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[5] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, "Existence and characterization of the covariance matrix maximum likelihood estimate in Spherically Invariant Random Processes," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34 – 48, January 2008.

[6] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5640–5651, Nov 2014.

[7] Y. Sun, P. Babu, and D. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.

[8] E. Ollila and D. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[9] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, 2012.

[10] I. Soloveychik and A. Wiesel, "Group symmetric robust covariance estimation," *arXiv preprint arXiv:1412.2345*, 2014.

[11] Y. Sun, P. Babu, and D. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing, submited*.

[12] Y. Sun, A. Breloy, P. Babu, D. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *Signal Processing, IEEE Transactions on, to appear*.

[13] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., MIT, Lexington, Mass., USA, Tech. Rep., December 1994.

[14] K. Greenwald and A. Hero, "Kronecker pca-based robust sar stap," *to appear in Algorithms for Synthetic Aperture Radar Imagery XXIII, Baltimore, April 2016, http://arxiv.org/pdf/1501.07481*.

[15] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D. P. Palomar, and F. Pascal, "A robust signal subspace estimator," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–4.

[16] Y. Sun, P. Babu, and D. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.

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

[18] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 635–650, Mar 2002.

[19] K. Benidis, Y. Sun, P. Babu, and D. Palomar, "Orthogonal sparse pca and covariance estimation via procrustes reformulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6211–6226, 2016.