MIMO Filters based on Robust Rank-Constrained Kronecker Covariance Matrix Estimation

Arnaud Breloy, Guillaume Ginolhac, Yongchan Gao, Frédéric Pascal

MIMO Filters based on Robust Rank-Constrained Kronecker Covariance Matrix Estimation

Arnaud Breloy[a], Guillaume Ginolhac[b], Yongchan Gao[c], Frédéric Pascal[d]

[a]*LEME, University Paris Nanterre, France*
[b]*LISTIC, University Savoie Mont-Blanc, France*
[c]*Xidian University, China*
[d]*Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.*

Abstract

In this paper, we propose a new estimator of the covariance matrix parameters when observations follow a mixture of a deterministic Compound-Gaussian (CG) and a white Gaussian noise. In particular, the covariance matrix of the CG contribution is assumed to be expressed as the Kronecker product of two low-rank matrices, which is a structure often involved in MIMO array processing. The proposed estimator is then obtained by maximizing the likelihood of the data with the use of a specifically tailored block Majorization-Minimization (MM) algorithm. Finally, the method is evaluated in terms of adaptive filtering on a MIMO-STAP radar setting, showing important improvements over standard processing.

Keywords: Adaptive signal processing, covariance matrix, Robust estimation, Majorization-Minimization, Kronecker product, low-rank filters.

1. Introduction

Within a statistical approach, the construction of adaptive filters usually requires the estimation of the interference-plus-noise covariance matrix as a preliminary step. This step represents a fundamental issue since the resulting processing performance strongly depends on the covariance matrix

*Corresponding author
Email address: abreloy@parisnanterre.fr (Arnaud Breloy)

estimation accuracy. The assumption that the disturbance follows a Gaussian distribution naturally points to the sample covariance matrix (SCM) for estimating this parameter. In this context it is well known that the corresponding adaptive filter reaches -3dB of expected Signal to Interference plus Noise Ratio (SINR) Loss (loss compared to the non-adaptive optimal filter) when the number of training samples (K) is equal to twice the size of the data (M) [1]. Nevertheless, the Gaussian assumption is not suited to impulsive measurements, *e.g.*, coming from high resolution systems. To model such observations, the general class of Compound-Gaussian (CG) distributions [2] (a subfamily of the complex elliptical symmetric distributions) has been shown to provide an accurate fit to empirical measurements in various RADAR and SONAR processing applications [3, 4, 5, 6, 7, 8, 9, 10, 11].

Additionally, the mixture of CG plus white Gaussian noise raised some interest in these applications [12, 13, 14, 15, 16, 17, 18], as it can account for the contribution of both the interference and the thermal noise through two separate and independent random processes. Within theses statistical models it is common to use a robust method to estimate the covariance matrix, such as M-estimators [19, 20]. These estimators have indeed been extensively leveraged in the modern detection/estimation literature due to their desirable robust properties (see [2, 21, 22] and references therein).

A major issue is that the general rule of thumb "$K = 2M$" is not always achievable in practice in modern systems (*e.g.*, due to a large number of sensors/pulses). This problem is even more critical at insufficient sample support since M-estimators require $K > M$ to be computed. In such cases, it is often possible to exploit some *a priori* knowledge on the covariance matrix structure in order to reduce the dimension of the estimation problem. This is specially relevant in array processing, where physical considerations on the system inherently imply that the covariance matrix belongs to a certain subset of the Positive Definite Hermitian (PDH) matrices manifold involving a particular structure, *e.g.,* Toeplitz, persymmetric, or low-rank. Robust estimation procedures for structured covariance matrices have recently been studied through various approaches, such as geodesic convexity [23], optimization [24] or 2-step approaches [25].

In this paper, we will focus on the matrices expressed as Kronecker product (KP) of Low-Rank (LR) matrices, which are of primary interest in Multiple Input Multiple Output (MIMO) systems [26, 27, 28, 29, 30, 31, 32]. Covariance matrix estimation processes for this structure have been proposed based on least square fitting in [28], and robust costs in [30]. This problem

has not been addressed within the context of mixture model, which is the gap we aim to fill. Hence, we formulate a new estimator as the maximum likelihood estimator (MLE) of a CG process with a structured covariance matrix (a Kronecker product (KP) of low-rank matrices), plus a white Gaussian noise. Inspired by [16], we derive a Majorization-Minimization (MM) algorithm [33] that cyclically updates the EigenValue Decomposition (EVD) parameters of the KP components, in order to evaluate this MLE. The proposed estimation method is then evaluated in terms of adaptive filtering on a MIMO-STAP radar setting.

The paper is organized as follows: Section 2 presents the general signal model while Section 3 concerns the derivation of the proposed estimation algorithm, which constitutes the main contribution. Finally, Section 4 illustrates the performance of the proposed approach in terms of adaptive filtering on a MIMO-STAP setting.

Notations: Vectors (resp. matrices) are denoted by bold-faced lowercase (resp. uppercase) letters. ^{T} and ^{H} respectively represent the transpose and the Hermitian operator. $\mathcal{CN}$ denotes the complex normal distributions. $\sim$ means "distributed as", $\overset{d}{=}$ stands for "shares the same distribution as". The operator $\otimes$ denotes the Kronecker product, and vec(.) is the operator which transforms a $m \times n$ matrix into a vector of length mn, using column-wise concatenation. $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ the matrix of zeroes with appropriate dimension.

2. MIMO-STAP formulation

2.1. Structure of the target signal

We consider a colocated MIMO radar on an airborne platform with M transmitting antennas and N receiving antennas [34, 35]. Both antenna arrays are side-looking and uniformly spaced linearly with the transmitted interspacing d_T and the received interspacing d_R, respectively. The radar transmits a coherent burst of L pulses in a coherent processing interval (CPI), where the pulse repetition interval (PRI) is T. For $m \in [\![0, M-1]\!]$ and $l \in [\![0, L-1]\!]$, the transmitted signal of the m^{th} element is expressed as

$$s_m(lT+\tau) = \sqrt{E/M}\phi_m(\tau)e^{j2\pi(lT+\tau)f} \tag{1}$$

where $\tau \in [0, T]$ represents the time within the pulse (fast time), E denotes the total transmitted energy, f is the carrier frequency, and $\phi_m(\tau)$ is the

transmitted waveform. The transmitted waveforms satisfy the orthogonality condition

$$\int_T \phi_m(\tau)\phi_j^*(\tau)d\tau = \delta_{mj}. \tag{2}$$

It is assumed that the transmitted waveform is narrow-band. For a target with relative velocity v_t toward the platform, the received echo of the n^{th} element after time delay τ_0 is given by

$$x_n(lT+\tau) \approx \sum_{m=0}^{M-1} \rho_t \phi_m(\tau) e^{j\frac{2\pi}{\lambda}[sin(\theta)(md_T+nd_R+2vTl)+2v_tTl]}, \tag{3}$$

where ρ_t is the complex-valued reflection coefficient of the target and the time delay $\tau_0 = 2R/c$ (where R is the distance between the target and the radar, and c is the speed of light) is absorbed in ρ_t, v is the velocity of the radar platform, $\lambda = c/f$ is the wavelength, and θ is the azimuth angle. Denote $f_s = d_R sin(\theta)/\lambda$ the received spatial frequency, $f_{T,s} = \alpha f_s$ the transmitted spatial frequency, and $f_D = 2(vsin(\theta)+v_t)T/\lambda$ the Doppler frequency, where $\alpha = d_T/d_R$. Note that $d_R = \lambda/2$ is chosen in order to avoid aliasing the spatial frequency. Thus, (3) can be rewritten as

$$x_n(lT+\tau) \approx \sum_{m=0}^{M-1} \rho_t \phi_m(\tau) e^{j2\pi(\alpha f_s m + f_s n)} e^{j2\pi f_D l}. \tag{4}$$

After down-conversion and matched-filtering, the output signal of the n^{th} element is given by

$$x_{m,n,l} = \int x_n(lT+\tau)\phi_m^*(\tau)d\tau = \rho_t e^{j2\pi(\alpha f_s m + f_s n)} e^{j2\pi f_D l}, \tag{5}$$

for $m \in [\![0, M-1]\!]$, $n \in [\![0, N-1]\!]$, and $l \in [\![0, L-1]\!]$. Then, stack the output signal yields the virtual snapshot of the target signal $\mathbf{x}_s \in \mathbb{C}^{MNL\times 1}$:

$$\mathbf{x}_s = \rho_t \mathbf{b}(f_D) \otimes \mathbf{a}(f_s) \otimes \mathbf{a}(f_{T,s}) = \rho_t \mathbf{s}(f_s, f_D), \tag{6}$$

where

- $\mathbf{a}(f_{T,s}) = [1, e^{j2\pi\alpha f_s}, \cdots, e^{j2\pi\alpha(M-1)f_s}]^T$ is the transmit spatial steering vector,

- $\mathbf{a}(f_s) = [1, e^{j2\pi f_s}, \cdots, e^{j2\pi(N-1)f_s}]^T$ is the receive spatial steering vector,
- $\mathbf{b}(f_D) = [1, e^{j2\pi f_D}, \cdots, e^{j2\pi(L-1)f_D}]^T$ is the Doppler steering vector,
- $\mathbf{s}(f_s, f_D) = \mathbf{b}(f_D) \otimes \mathbf{a}(f_s) \otimes \mathbf{a}(f_{T,s})$ is the MIMO space-time steering vector.

2.2. Structure and statistical model of the interference

The received target signal $\mathbf{x}_s$ is embedded in a disturbance $\mathbf{x}_d$ composed of the ground response $\mathbf{x}_i$, referred to as clutter, and additive white Gaussian noise $\mathbf{x}_n$ (WGN, *i.e.*, the thermal noise). Thus, the received signal $\mathbf{x}$ is written as follows:

$$\mathbf{x} = \mathbf{x}_s + \mathbf{x}_d = \mathbf{x}_s + \mathbf{x}_i + \mathbf{x}_n, \tag{7}$$

where the WGN $\mathbf{x}_n$ is assumed to be distributed as $\mathbf{x}_n \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{MNL})$[1].

Before specifying the distribution of the clutter $\mathbf{x}_d$, we can notice that its covariance matrix $\mathbf{R} = \mathbb{E}(\mathbf{x}_i\mathbf{x}_i^H)$ can be expressed in a Kronecker product form due to the system structure (cf. previous section). Thus

$$\mathbf{R} = \mathbf{A} \otimes \mathbf{B}. \tag{8}$$

In addition to this Kronecker product structure, many RADAR applications allow us to assume that the matrices $\mathbf{A}$ and $\mathbf{B}$ exhibit a low-rank structure, with ranks denoted R_A and R_B respectively. This assumption is generally related to dimension of the signal subspace implied by the underlying physics of the measurement system [36, 37]. In this paper, the ranks R_A and R_B are assumed to be known and fixed depending on the application (or previously estimated). This general covariance structure model is, e.g., involved in:

- classic side-looking STAP (see eq.(45) on page 19 in [26]) where $\mathbf{A} = \mathbf{I}_{ML}$, $\mathbf{B} = \sum_{j=1}^{J} \sigma_j^2 \mathbf{a}_j \mathbf{a}_j^H$, where J is the number of interferences (J is usually smaller than N), σ_j denotes the interference power of the jth and $\mathbf{a}_j$ denotes the corresponding interference steering vector;
- Synthetic Aperture Radar (SAR) STAP [28] where $\text{rank}(\mathbf{A}) = 1$;
- MIMO STAP where $\text{rank}(\mathbf{R}) < MNL$ [38], and where $\mathbf{A}$ and $\mathbf{B}$ can be assumed to be LR [29];

[1] In this paper, we will assume that σ^2 is known or previously estimated.

- and finally, general MIMO RADAR [27] where the rank of $\mathbf{A}$ and $\mathbf{B}$ depends on various assumptions on the array geometry [37].

Though it is not the main focus of this work, we also notice that the ranks can be still be assumed to be full ($R_A = ML$ and $R_B = N$) without loss of generality of the notation.

In conclusion, the covariance matrix of the total interference plus noise is thus expressed as $\mathbb{E}\left[\mathbf{x}_d \mathbf{x}_d^H\right] = \mathbf{\Sigma} = \mathbf{R} + \sigma^2 \mathbf{I}_{MNL} = \mathbf{A} \otimes \mathbf{B} + \sigma^2 \mathbf{I}_{MNL}$. Under standard assumptions, the interference $\mathbf{x}_i$ could be assumed to be Gaussian, which would yield $\mathbf{x}_d \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$. However, plugging the clutter covariance matrix expression in a Gaussian model is not always the most representative of clutter's behavior, in particular for high-resolution STAP and MIMO-STAP systems. In this context, the compound-Gaussian model can better reflect the clutter impulsiveness. A zero-mean compound-Gaussian distributed observation admits the following stochastic representation $\mathbf{x} \overset{d}{=} \sqrt{\tau}\mathbf{n}$, where $\mathbf{n}$ is a zero-mean complex Gaussian vector, and τ is an independent random positive scaling, referred to as *texture*[2]. The choice of the texture probability density function can lead to various well-known multivariate heavy-tailed distributions, such as t- and K- distributions [7]. Notice that this work will assume that $\mathbb{E}[\tau]$ exists, however this property is not always satisfied within the whole compound-Gaussian family (e.g., the Cauchy does not have finite second order moment). In this case, conditioning on the texture τ leads to the following representations

$$\begin{aligned} \mathbf{x}_k | \tau_k &\sim \mathcal{CN}(\mathbf{0}, \tau_k(\mathbf{A} \otimes \mathbf{B})), \\ \mathbf{x}_d | \tau_d &\sim \mathcal{CN}(\mathbf{0}, \tau_d(\mathbf{A} \otimes \mathbf{B}) + \sigma^2 \mathbf{I}_{MNL}), \end{aligned} \tag{9}$$

where it is recalled that we assume that the matrices $\mathbf{A}$ and $\mathbf{B}$ possibly have a low-rank structure (with ranks R_A and R_B). In the following, this model will be referred to as *Compound Kronecker Product of Low-Rank plus Identity* (CKPLR).

[2]Notice that the texture from the statistical model is not related to the time within the pulse of the section 2.1. Still, we kept the standard variable notations for both of these parameters as there will be no ambiguity in the remaining of the paper.

2.3. Adaptive MIMO-STAP filters

For the above signal model, the optimal filter (in terms of output SNR) is given by

$$\mathbf{w} = \frac{\mathbf{\Sigma}^{-1}\mathbf{s}(f_s, f_D)}{\mathbf{s}^H(f_s, f_D)\mathbf{\Sigma}^{-1}\mathbf{s}(f_s, f_D)} \tag{10}$$

where $\mathbf{\Sigma} = \mathbb{E}\left[\mathbf{x}_d\mathbf{x}_d^H\right] = \mathbb{E}\left[\tau\right](\mathbf{A}\otimes\mathbf{B}) + \sigma^2\mathbf{I}_{MNL}$. In practice, the disturbance covariance matrix is unknown and needs to be estimated from training data, denoted $\{\mathbf{x}_k\}_{k=1}^K$, which are assumed to be signal-free, as well as independent and identically distributed (i.i.d.). The obtained estimate $\hat{\mathbf{\Sigma}}$ is then plugged in (10) in order to build a so-called adaptive filter.

Of course, the performance of the adaptive filter strongly depends on the estimation process accuracy. In the classical Gaussian case, the sample covariance matrix (SCM) $\hat{\mathbf{\Sigma}}_{SCM} = \sum_{k=1}^K \mathbf{x}_k\mathbf{x}_k^H/K$ is generally used as estimator. In this case, it is well known that the expected SINR Loss is equal to -3dB when $K = 2MNL$. Nevertheless, the SCM is not an accurate estimate of the covariance matrix when samples are non-Gaussian distributed. Moreover, this estimator is not particularly suited to our context, as it does not exploit any prior knowledge on the structure of the covariance matrix. In order to respond to these two issues, we propose to derive a new robust estimator of the covariance matrix parameters suited to the CKPLR model in (9).

3. Robust filters for the CKPLR model

3.1. Problem formulation

The proposed estimator corresponds to the MLE of the CKPLR model when the textures are assumed to be unknown deterministic. This assumption on the textures parameters is made in order to be robust to any mismatch on their underlying distribution.

Define the set of structured matrices:

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

The evaluation of the MLE of the CKPLR model requires to solve the following problem:

$$\underset{\{\mathbf{\Sigma}_k\}_{k=1}^K \in \mathcal{S}_{CKPLR}}{\text{minimize}} \quad \mathcal{L}\left(\{\mathbf{\Sigma}_k\}_{k=1}^K\right), \tag{11}$$

where

$$\mathcal{L}\left(\{\mathbf{\Sigma}_k\}_{k=1}^K\right) = \sum_{k=1}^K \log|\mathbf{\Sigma}_k| + \sum_{k=1}^K \mathbf{x}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{x}_k, \tag{12}$$

is the negative Gaussian log-likelihood function (τ_k's being unknown but deterministic and contained in the covariance matrix).

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

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

Note that the LR 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 $\mathbf{A}$ and $\mathbf{B}$ in (12), the objective function becomes:

$$\begin{aligned} \mathcal{L}(\mathbf{D}_A, \mathbf{D}_B, \mathbf{U}_A, \mathbf{U}_B, \{\tau_k\}_{k=1}^K) &= \sum_{k=1}^K \sum_{p=1}^P \sum_{p=1}^Q \ln(\tau_k a_p b_q + \sigma^2) \\ &+ \sum_{k=1}^K \mathbf{x}_k^H (\mathbf{U}_A \otimes \mathbf{U}_B)(\tau_k \mathbf{D}_A \otimes \mathbf{D}_B + \sigma^2 \mathbf{I})^{-1} (\mathbf{U}_A \otimes \mathbf{U}_B)^H \mathbf{x}_k. \end{aligned} \tag{13}$$

This leads to the optimization problem

$$\begin{aligned} \underset{\{\tau_k\}_{k=1}^K, \mathbf{D}_A, \mathbf{D}_B, \mathbf{U}_A, \mathbf{U}_B}{\text{minimize}} \quad & \mathcal{L}(\mathbf{D}_A, \mathbf{D}_B, \mathbf{U}_A, \mathbf{U}_B, \{\tau_k\}_{k=1}^K) \\ \text{subject to} \quad & \mathbf{\Sigma}_k = \tau_k (\mathbf{A} \otimes \mathbf{B}) + \sigma^2 \mathbf{I} \\ & \tau_k \geq 0, \ \forall k \in [\![1, K]\!] \\ & \mathbf{U}_A^H \mathbf{U}_A = \mathbf{I}_P, \ \mathbf{U}_B^H \mathbf{U}_B = \mathbf{I}_Q \\ & \mathbf{D}_A = \text{diag}([a_1, \ldots, a_{R_A}, 0, \ldots, 0]) \\ & \mathbf{D}_B = \text{diag}([b_1, \ldots, b_{R_B}, 0, \ldots, 0]) \end{aligned}$$

This problem is too complex to be directly tackled. Hence, we propose to

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

Form a starting point $\{\mathbf{A}^{t=0}, \mathbf{B}^{t=0}, \{\tau_k\}^{t=0}\}$.
repeat
    $t \leftarrow t+1$
    Update $\{\tau_k^t\}$ with (17).
    Update $\mathbf{D}_A^t$ with (19).
    Update $\mathbf{D}_B^t$ with (22).
    Update $\mathbf{U}_A^t$ with (28).
    Update $\mathbf{U}_B^t$ with (34).
until Some convergence criterion is met.

leverage the block majorization-minimization (MM) framework to address it in the following sections. The resulting estimation procedure is referred to as "CKPLR - MM", which algorithm is summed up in the box Algorithm 1.

3.2. Majorization-Minimization algorithm to evaluate the MLE

At a given iteration step $\{\mathbf{D}_A^t, \mathbf{D}_B^t, \mathbf{D}_A^t, \mathbf{D}_B^t, \{\tau_k^t\}_{k=1}^K\}$, the block-MM algorithm consists in updating the variables cyclically, by minimizing surrogates functions (tight upper-bound of the objective), leading to a monotonic decrement of the objective. For more details on the MM algorithm, the reader is referred to [33]. These surrogates functions and corresponding updates for each variables are derived below. Note that, in order to lighten the notation, we omit the reference on t for variables that are fixed in each step.

3.2.1. Step 1: Update τ_k when fixed other parameters

From (13), it is easy to show that the objective is separable in τ_k and yields for a given index k

$$\mathcal{L}(\tau_k) = \sum_{p=1}^{P}\sum_{q=1}^{Q} \frac{[\tilde{\mathbf{x}}_k]^2_{(p-1)Q+q}}{\tau_k a_p b_q + \sigma^2} + \sum_{p=1}^{P}\sum_{p=1}^{Q} \ln(\tau_k a_p b_q + \sigma^2), \tag{14}$$

with $\tilde{\mathbf{x}}_k = (\mathbf{U}_A \otimes \mathbf{U}_B)^H \mathbf{x}_k$. This function has no closed-form minimizer, but it is possible to obtain a closed-form update that improves the value of the objective function thanks to the following propositions.

Proposition 1. *The objective function* (14) *can be upperbounded by the surrogate function*

$$g(\tau_k|\tau_k^t) = \mathcal{A}_k \ln(\omega_k \tau_k + \beta_k) - \mathcal{K}_k \ln(\tau_k) \tag{15}$$

with

$$\begin{array}{lll} \kappa_{pq}^k = \frac{\tau_k^t a_p b_q [\tilde{\mathbf{x}}_k]_{pq}^2}{\tau_k^t a_p b_q + \sigma^2}, & \mathcal{K}_k = \sum_{pq} \kappa_{pq}^k, & \mathcal{A}_k = \sum_{pq} (\kappa_{pq}^k + 1), \\ \beta_k = \frac{1}{\mathcal{A}_k} \sum_{pq} \frac{(\kappa_{pq}^k+1)\sigma^2}{\tau_k^t a_p b_q + \sigma^2}, & \omega_k = \frac{1}{\mathcal{A}_k} \sum_{pq} \frac{(\kappa_{pq}^k+1) a_p b_q}{\tau_k^t a_p b_q + \sigma^2}, & \end{array} \tag{16}$$

where we use aggregated indexes "pq" to lighten the summation notation.

Proof. See Appendix A. □

Proposition 2. *(Proposition 2 of [16]) The surrogate function which upperbounds* (14) *is quasiconvex and has the a unique closed-form minimizer, that provides the update:*

$$\tau_k^{t+1} = \frac{\mathcal{K}_k \beta_k}{(\mathcal{A}_k - \mathcal{K}_k)\omega_k}, \ \forall k \in [\![1, K]\!] \tag{17}$$

3.2.2. Step 2: Update $\mathbf{D}_A$ for fixed other parameters

As for the variables τ_k, the objective is separable in a_p and yields for a given index p

$$\mathcal{L}(a_p) = \sum_{k=1}^{K} \sum_{q=1}^{Q} \frac{[\tilde{\mathbf{x}}_k]_{(p-1)Q+q}^2}{\tau_k a_p b_q + \sigma^2} + \sum_{k=1}^{K} \sum_{q=1}^{Q} \ln(\tau_k a_p b_q + \sigma^2) \tag{18}$$

with $\tilde{\mathbf{x}}_k = (\mathbf{U}_A \otimes \mathbf{U}_B)^H \mathbf{x}_k$. Notice that the objective function (18) has a form similar to (14), and only differs from summation indexes. An appropriate adaptation of Propositions 1 and 2 (omitted for the sake of conciseness) leads to the following update:

$$\begin{array}{l} a_p^{t+1} = \frac{\mathcal{K}_p \beta_p}{(\mathcal{A}_p - \mathcal{K}_p)\omega_p}, \ \forall p \in [\![1, R_A]\!] \\ a_p = 0, \ \forall p \in [\![R_A + 1, P]\!] \end{array} \tag{19}$$

with

$$\begin{array}{lll} \kappa_{qk}^{p} = \frac{\tau_k a_p^t b_q [\tilde{\mathbf{x}}_k]_{pq}^2}{\tau_k a_p^t b_q + \sigma^2}, & \mathcal{K}_p = \sum_{qk} \kappa_{qk}^{p}, & \mathcal{A}_p = \sum_{qk} (\kappa_{qk}^{p} + 1), \\ \beta_p = \frac{1}{\mathcal{A}_p} \sum_{qk} \frac{(\kappa_{qk}^{p}+1)\sigma^2}{\tau_k a_p^t b_q + \sigma^2}, & \omega_p = \frac{1}{\mathcal{A}_p} \sum_{qk} \frac{(\kappa_{qk}^{p}+1) a_p^t b_q}{\tau_k a_p^t b_q + \sigma^2}, & \end{array} \tag{20}$$

where we use aggregated indexes "qk".

3.2.3. Step 3: Update $\mathbf{D}_B$ for fixed other parameters

As for the variables τ_k and a_p, the objective is separable in b_q and yields, for a given index q,

$$\mathcal{L}(b_q) = \sum_{p=1}^{P} \sum_{k=1}^{K} \frac{[\tilde{\mathbf{x}}_k]_{(p-1)Q+q}^2}{\tau_k a_p b_q + \sigma^2} + \sum_{p=1}^{P} \sum_{k=1}^{K} \ln(\tau_k a_p b_q + \sigma^2) \tag{21}$$

with $\tilde{\mathbf{x}}_k = (\mathbf{U}_A \otimes \mathbf{U}_B)^H \mathbf{x}_k$. The same reasoning as previously applies, leading to the following updates:

$$\begin{array}{l} b_q^{t+1} = \dfrac{\mathcal{K}_q \beta_q}{(\mathcal{A}_q - \mathcal{K}_q)\omega_q}, \ \forall q \in [\![1, R_B]\!] \\ b_q = 0, \ \forall q \in [\![R_B + 1, Q]\!] \end{array} \tag{22}$$

with

$$\begin{array}{lll} \kappa_{pk}^{q} = \frac{\tau_k a_p b_q^t [\tilde{\mathbf{x}}_k]_{pq}^2}{\tau_k a_p b_q^t + \sigma^2}, & \mathcal{K}_q = \sum_{pk} \kappa_{pk}^{q}, & \mathcal{A}_q = \sum_{pk} (\kappa_{pk}^{q} + 1), \\ \beta_q = (1/\mathcal{A}_q) \sum_{pk} \frac{(\kappa_{pk}^{q}+1)\sigma^2}{\tau_k a_p b_q^t + \sigma^2}, & \omega_q = (1/\mathcal{A}_q) \sum_{pk} \frac{(\kappa_{pk}^{q}+1) a_p b_q^t}{\tau_k a_p b_q^t + \sigma^2}, & \end{array} \tag{23}$$

where we use aggregated indexes "pk".

3.2.4. Step 4: Update $\mathbf{U}_A$ for fixed other parameters

Some matrix manipulations allow to rewrite (13) with respect to $\mathbf{U}_A = \left[\mathbf{u}_1^A, \ldots, \mathbf{u}_P^A\right]$ (with other fixed variables and ignoring constant terms) as

$$\begin{aligned}\mathcal{L}(\mathbf{U}_A) &= \sum_{k=1}^{K} \operatorname{vec}\left(\mathbf{U}_B^H \mathbf{X}_k \mathbf{U}_A\right)^H \left(\tau_k \mathbf{D}_A \otimes \mathbf{D}_B + \sigma^2 \mathbf{I}\right)^{-1} \operatorname{vec}\left(\mathbf{U}_B^H \mathbf{X}_k \mathbf{U}_A\right) \\ &= \sum_{k=1}^{K} \sum_{p=1}^{P} \left((\mathbf{u}_p^A)^H (\mathbf{X}_k^B)^H \mathbf{\Lambda}_{p,k}^{-1} \mathbf{X}_k^B \mathbf{u}_p^A\right) \\ &= \sum_{p=1}^{P} (\mathbf{u}_p^A)^H \mathbf{M}_p \mathbf{u}_p^A\end{aligned} \tag{24}$$

with

$$\begin{aligned}&\mathbf{X}_k = \operatorname{unvec}(\mathbf{x}_k) \in \mathbb{C}^{Q \times P}, \quad \mathbf{X}_k^B = \mathbf{U}_B^H \mathbf{X}_k, \\ &\mathbf{\Lambda}_{p,k} = \tau_k a_p \mathbf{D}_B + \sigma^2 \mathbf{I}, \qquad \mathbf{M}_p = \sum_{k=1}^{K} (\mathbf{X}_k^B)^H \mathbf{\Lambda}_{p,k}^{-1} \mathbf{X}_k^B,\end{aligned} \tag{25}$$

unvec(.) being the inverse operator of vec(.).

Again, this function admits no closed-form minimizer under the constraint $\mathbf{U}_A^H \mathbf{U}_A = \mathbf{I}_p$. However, it is possible to obtain a closed-form update that improves the value of the objective function thanks to the following propositions.

Proposition 3. *The objective function $\mathcal{L}(\mathbf{U}_A)$ can be upperbounded at $\mathbf{U}_A^t$ by the surrogate function*

$$g\left(\mathbf{U}_A | \mathbf{U}_A^t\right) = \operatorname{Tr}\left[(\mathbf{W}_A^t)^H \mathbf{U}_A\right] + \operatorname{Tr}\left[\mathbf{U}_A^H \mathbf{W}_A^t\right] + \text{const.} \tag{26}$$

with

$$\mathbf{W}_A^t = \left[\begin{array}{c} \mathbf{G}_1 \mathbf{u}_1^{A(t)}, \ \ldots, \ \mathbf{G}_P \mathbf{u}_P^{A(t)} \end{array}\right] \tag{27}$$

with $\mathbf{G}_p = \mathbf{M}_p - \lambda_{max}^{(\mathbf{M}_p)} \mathbf{I}$. Equality holds at $\mathbf{U}_A^t$.

Proof. See Appendix B. □

Proposition 4. *(Proposition 7 of [39]) The problem of minimizing the surrogate function $g\left(\mathbf{U}_A | \mathbf{U}_A^t\right)$ under orthonormality constraint has an optimal*

solution, that leads to the update

$$\mathbf{U}_A^{t+1} = \mathbf{V}_L \mathbf{V}_R^H \tag{28}$$

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 (27).*

3.2.5. Step 5: Update $\mathbf{U}_B$ *for fixed other parameters*

Some matrix manipulations allow to rewrite (13) with respect to $\mathbf{U}_B = \left[\mathbf{u}_1^B, \ldots, \mathbf{u}_Q^B\right]$ (with other fixed variables and ignoring constant terms) as:

$$\begin{aligned}\mathcal{L}(\mathbf{U}_B) &= \sum_{k=1}^{K} \operatorname{vec}\left(\mathbf{U}_B^H \mathbf{X}_k \mathbf{U}_A\right)^H \left(\tau_k \mathbf{D}_A \otimes \mathbf{D}_B + \sigma^2 \mathbf{I}\right)^{-1} \operatorname{vec}\left(\mathbf{U}_B^H \mathbf{X}_k \mathbf{U}_A\right) \\ &= \sum_{k=1}^{K} \sum_{p=1}^{P} \left(\left[\mathbf{X}_k^A\right]_{:,p}^H \mathbf{U}_B \mathbf{\Lambda}_{p,k}^{-1} \mathbf{U}_B^H \left[\mathbf{X}_k^A\right]_{:,p} \right)\end{aligned} \tag{29}$$

where we used $\left[\operatorname{vec}\left(\mathbf{U}_B^H \mathbf{X}_k^A\right)\right]^T = \left[\left(\mathbf{U}_B^H \left[\mathbf{X}_k^A\right]_{:,1}\right)^T \ \ldots \ \left(\mathbf{U}_B^H \left[\mathbf{X}_k^A\right]_{:,P}\right)^T \right]$, with the notation $\mathbf{X}_k^A = \mathbf{X}_k \mathbf{U}_A$. Now, recall that $\mathbf{\Lambda}_{p,k} = \tau_k a_p \mathbf{D}_B + \sigma^2 \mathbf{I}$. Since we have $\mathbf{D}_B = \operatorname{diag}([b_1, \ldots, b_{R_B}, 0, \ldots, 0])$, we can use the matrix inversion lemma to obtain

$$\mathbf{U}_B \mathbf{\Lambda}_{p,k}^{-1} \mathbf{U}_B^H = (\sigma^2 \mathbf{I} + \tau_k a_p \bar{\mathbf{U}}_B \bar{\mathbf{D}}_B \bar{\mathbf{U}}_B^H)^{-1} = \sigma^{-2} \mathbf{I} - \bar{\mathbf{U}}_B \mathbf{\Gamma}_{p,k} \bar{\mathbf{U}}_B^H \tag{30}$$

where we denoted the partition $\mathbf{U}_B = [\bar{\mathbf{U}}_B | \mathbf{U}_B^{\perp}]$, with $\bar{\mathbf{U}}_B \in \mathbb{C}^{Q \times R_B}$, and $\mathbf{\Gamma}_{p,k} = \operatorname{diag}(\{\frac{\tau_k a_p b_q}{\sigma^2(\sigma^2 + \tau_k a_p b_q)}\})$. Finally, the objective $\mathcal{L}(\bar{\mathbf{U}}_B)$ can be expressed

$$\mathcal{L}(\mathbf{U}_B) = \mathcal{L}(\bar{\mathbf{U}}_B) = -\sum_{k,p} \operatorname{Tr}\left\{ \mathbf{Z}_{p,k}^A \bar{\mathbf{U}}_B \mathbf{\Gamma}_{p,k} \bar{\mathbf{U}}_B^H \right\} + \text{const.} \tag{31}$$

with $\mathbf{Z}_{p,k}^A = \left[\mathbf{X}_k^A\right]_{:,p} \left[\mathbf{X}_k^A\right]_{:,p}^H$. As previously, it is again possible to obtain a closed-form update that improves the value of the objective function thanks to the following propositions.

Proposition 5. *(Extended from Proposition 5 in [16]) The objective function* $\mathcal{L}(\mathbf{U}_B)$ *can be upperbounded at* $\bar{\mathbf{U}}_B^t$ *by the surrogate function*

$$g\left(\mathbf{U}_B | \bar{\mathbf{U}}_B^t\right) = -\operatorname{Tr}\{(\bar{\mathbf{U}}_B^t)^H \mathbf{M}_q\} - \operatorname{Tr}\{\mathbf{M}_q^H \bar{\mathbf{U}}_B^t\} \tag{32}$$

with

$$\mathbf{M}_q = \sum_{k,p} \mathbf{X}^A_{p,k} \bar{\mathbf{U}}^t_B \mathbf{\Gamma}_{p,k} \tag{33}$$

Proposition 6. *(Extended from Proposition 7 of [39]) The problem of minimizing the surrogate function* $g\left(\mathbf{U}_B | \bar{\mathbf{U}}^t_B\right)$ *under orthonormality constraint on has an optimal solution, that leads to the update*

$$\mathbf{U}^{t+1}_B = [\bar{\mathbf{U}}^{t+1}_B | \mathbf{U}^{(t+1)\perp}_B] \tag{34}$$

with

$$\bar{\mathbf{U}}^{t+1}_B = \mathbf{V}_L \mathbf{V}^H_R \tag{35}$$

where $\mathbf{V}_L$ *and* $\mathbf{V}_R$ *are respectively the left and right singular vector of* $\mathbf{M}_q$ *in* (33)*, and* $\mathbf{U}^{(t+1)\perp}_B$ *is a orthogonal basis of the subspace spanned by* $\mathbf{I} - \bar{\mathbf{U}}^{t+1}_B \bar{\mathbf{U}}^{(t+1)H}_B$

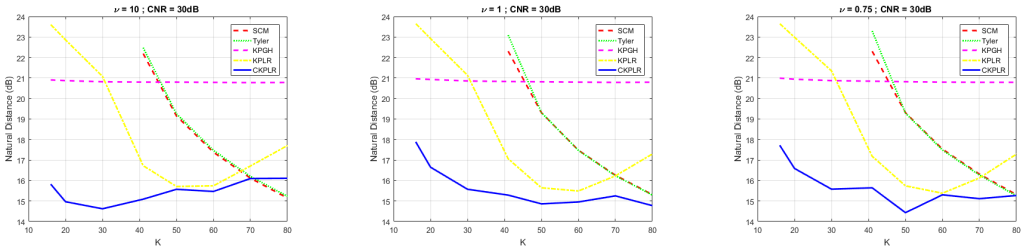
4. Simulations

In this section, we first validate our algorithm by computing the natural distance [40] between the obtained estimator and the true covariance matrix. In a second sub-section, we investigate the interest of our approach in a MIMO application.

4.1. Validation

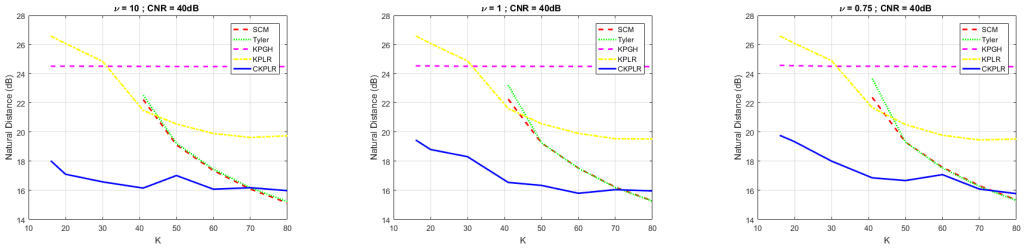
In this simulation, we set $\mathbf{A} \in \mathbb{C}^{P \times P}$ with $P = 10$ and $R_A = 4$, and $\mathbf{B} \in \mathbb{C}^{Q \times Q}$ with $Q = 4$ and $R_B = 2$. The matrix $\mathbf{A}$ (resp. $\mathbf{B}$) is constructed using the rank R_A (resp. R_B) truncated EVD of a Toeplitz matrix of correlation coefficient $\rho_A = 0.90$ (resp. $\rho_B = 0.95$). The texture τ is generated from a Gamma distribution $\tau \sim \Gamma(\nu, \frac{1}{\nu})$ (thus $\mathbb{E}[\tau] = 1$), yielding a K-distributed clutter. We will consider three values of ν: 0.75, 1 (two cases of heavy tailed clutter) and 10 (quasi-Gaussian clutter). The Clutter to Noise Ratio is defined by the ratio CNR=$\mathbb{E}[\tau]\,\mathrm{tr}(\mathbf{A} \otimes \mathbf{B})/(\sigma^2(R_A R_B))$, for which we will test the values 30 dB and 40 dB.

The following estimators are compared: 1) the SCM (MLE for unstructured Gaussian distributed); 2) Tyler's estimator [20, 21, 22, 41] (approximate MLE for unstructured CG distributions); 3) KPGH proposed in [28] (projection of the SCM on the set of LR Kronecker Product matrices); 4)



(a) $\nu = 10$ (b) $\nu = 1$ (c) $\nu = 0.75$

Figure 1: Natural distance between covariance estimators and the true covariance matrix w.r.t K. The total size is $PQ = 40$, the ranks are $R_A = 4, R_B = 2$ and the CNR is $30dB$.



(a) $\nu = 10$ (b) $\nu = 1$ (c) $\nu = 0.75$

Figure 2: Natural distance between covariance estimators and the true covariance matrix w.r.t K. The total size is $PQ = 40$, the ranks are $R_A = 4, R_B = 2$ and the CNR is $40dB$.

KPLR proposed in [30] (Tyler's estimator with appropriate structure constraint); 5) CKPLR, proposed in this paper. The estimation accuracy of these estimators is measured using Monte-Carlo simulations to evaluate the expected natural Riemannian distance between two positive definite matrices (shown to be a more discriminant metric than the standard mean squared error in [40]):

$$d_{nat}(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \|\text{logm}(\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2})\|_2^2, \tag{36}$$

where logm is the matrix logarithm.

Figures 1 and 2 display the simulation results under various ν as a function of K for CNR=30dB and for CNR=40dB respectively. In all cases, we notice that the proposed estimator reaches the best performance, and can interestingly be computed in undersampled scenarios ($K < PQ$). KPLR also achieves interesting results, but its performance decreases in high CNR scenarios. This is due to the fact that it is built upon Tyler's cost function, which is not well suited to (almost) rank deficient models. KPGH can also

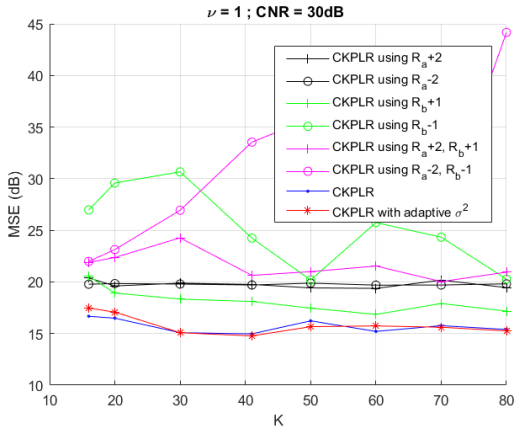


Figure 3: Comparison of various implementation of CKPLR-MM with model mismatch. The total size is $PQ = 40$, the ranks are $R_A = 4, R_B = 2$, $\nu = 1$ and the CNR is $30dB$.

be computed at low sample support, however it does not benefit from large sample support settings. This is probably due to its mismatch with the true model (the same behavior is observed in [28]).

Finally, we test the robustness of the proposed method to mismatches regarding the assumed known parameters. The samples are generated as previously, and we test the performance of: 1) CKPLR applied using different assumed ranks R_a and R_b; 2) CKPLR applied using data weighted by $\hat{\sigma}^{-1}$, where the estimate $\hat{\sigma}^2$ obtained as the mean of the last $PQ - R_a R_b$ eigenvalues of the SCM. Results are displayed in Figure 3. First, we can observe that the performance of the estimation algorithm is not dramatically impacted when the rank is slightly overevaluated. A possible explanation would be that it is more accurate to estimate eigenvalues that are actually zero than to fit the observations with an underdimensioned model. Such observations have also been made in [42, 43] (focused on non-Kronecker models and/or other applications). Second, we can also observe that the adaptive version (estimating the noise floor) of the algorithm reach performance close to the non-adaptive one, which illustrates that a slight mis-evaluation of σ^2 does not impact the proposed method.

4.2. MIMO application

In this section, we consider a colocated MIMO radar with $M = 2$ transmitting uniform antennas, $N = 4$ receiving uniform antennas, and $L = 10$ pulses, where $d_R = \lambda/2$ and $d_T = 8d_R$. Thus, the dimension of system is

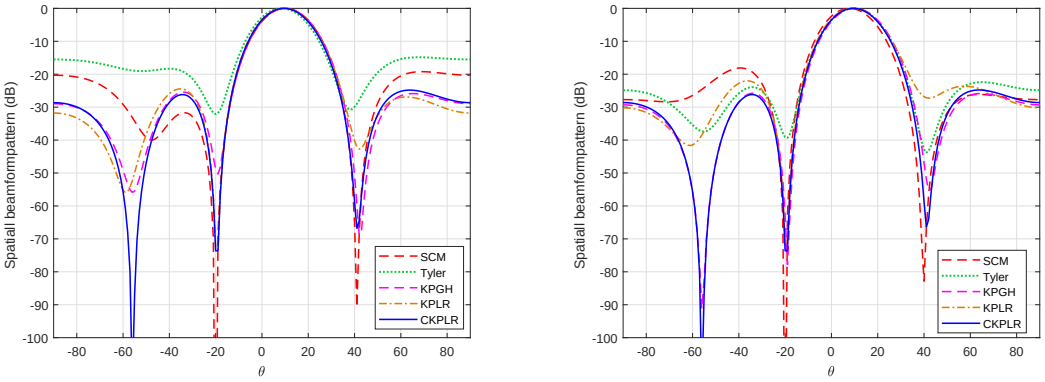
$MNL = 80$. The azimuth of the target is $10°$, and those of the two jammers are $-20°$ and $40°$. The noise floor is set to $\sigma^2 = 1$ and the Interference to Noise Ratio (INR) is set to respectively 50dB for the jammer at $-20°$, and 30dB for the one at $40°$. The wavelength is $\lambda = 0.32$m and the Pulse Repetition Frequency (PRF) is 1500 Hz. The platform velocity is set to 120m/s. The clutters follows a K-distribution as in the previous subsection , i.e. $\tau \sim \Gamma(\nu, \frac{1}{\nu})$. We will consider $\nu = 10$ to model a quasi-Gaussian clutter, and $\nu = 1$ for a heavy-tailed one. Let us denote $P = ML$. We set $\mathbf{A} = [\mathbf{I}_{R_A}, \mathbf{0}_{R_A \times (P-R_A)}; \mathbf{0}_{(P-R_A)\times R_A}, \mathbf{0}_{(P-R_A)\times(P-R_a)}]$ with $R_A = 9$. Finally $\mathbf{B}$ is set using the steering vector model below (6), and $R_B = 2$ since there are two jammers.

As in the previous subsection, we compare the following adaptive filters built from (10) and based on different estimators of the covariance matrix: SCM, Tyler's estimator, KPGH, KPLR, CKPLR.

Fig. 4 shows the spatial patterns for two different values of K and ν. It can be seen that the CKPLR maintains a better beamforming pattern with lower side lobe than all other methods. The main lobes of both the KPGH and the KPLR are wider than that of the CKPLR. For the SCM and Tyler's estimator, their side lobes are too high due to the effect of the small number of training samples, especially the main lobe of the SCM has deviated from the target direction.

Fig. 5 shows the Minimum Variance Distortionless Response (MVDR) spectrum [26] in the angle-frequency domain for $K = 60$ and $\nu = 10$. In Fig. 5, the KPLR, CKPLR and KPGH, after beamforming, can obtain better output power and the target can be distinguished clearly. Since the value of ν gives a quasi-Gaussian scenario, the KPLR shows a little better response performance than the CKPLR. Both the SCM and Tyler's estimator show the worse response performance since K is smaller than the system dimension 80. As seen in Fig. 3, one can notice that the response performance of the SCM improves as K increases to 120. In this case, the CKPLR shows the best response performance by exploiting the LR Kronecker structure.

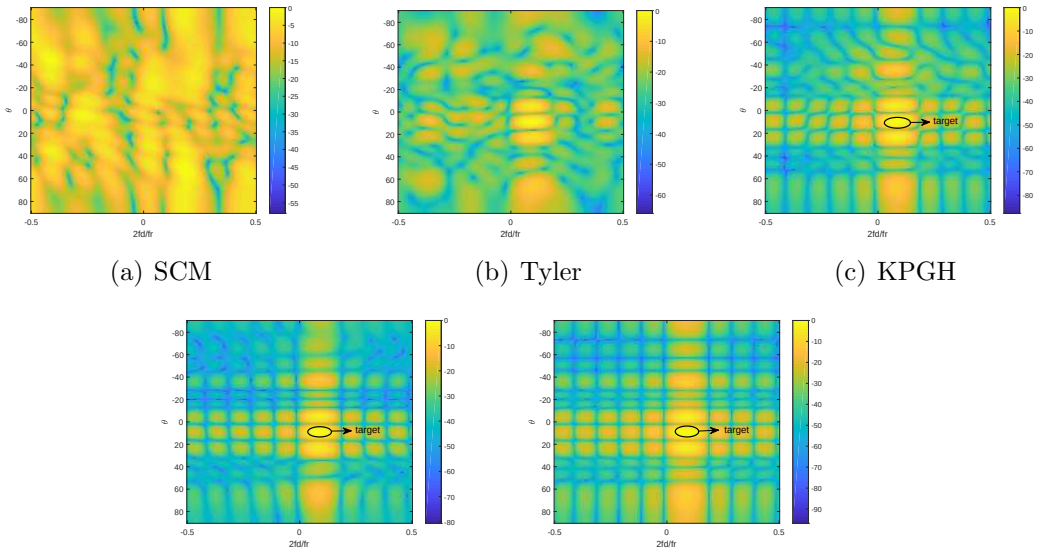
We now examine the performance of these methods for a heterogeneous case with $\nu = 1$. Fig. 7 and Fig. 8 show the MVDR spectra in angle-frequency domain for $K = 60$ and $K = 120$ respectively. For KPLR, CKPLR and KPGH, the target appears clearly. It is seen that the response performance of CKPLR significantly outperforms the other methods. This is as expected, since CKPLR utilizes the LR constrained Kronecker structure information and estimates the covariance matrix under the compound-



(a) $K = 60$, $\nu = 10$ (b) $K = 120$, $\nu = 10$

(c) $K = 60$, $\nu = 1$ (d) $K = 120$, $\nu = 1$

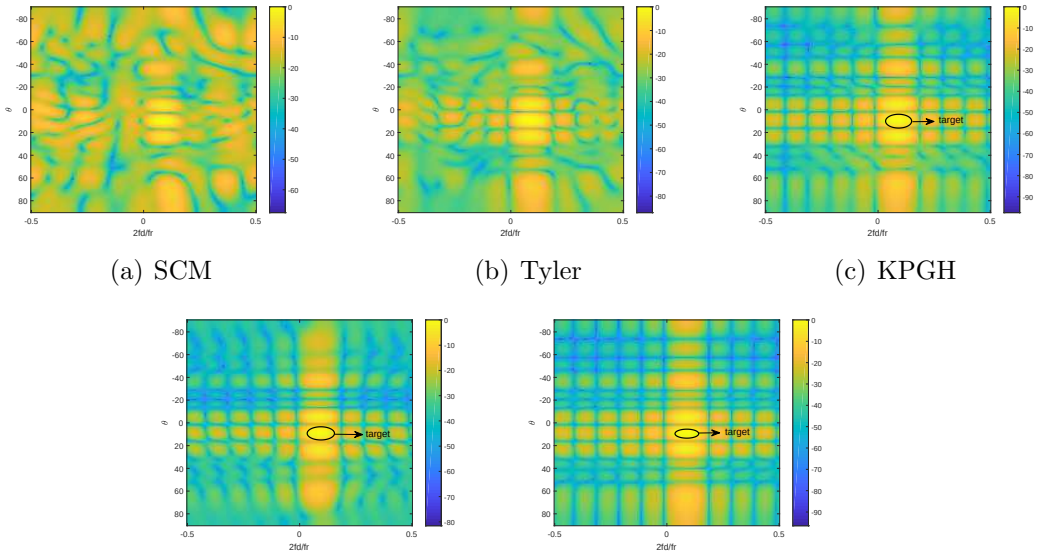
Figure 4: Spatial pattern for different parameters of K and ν. The total size is $MNL = 80$ and the ranks are $R_A = 9, R_B = 2$.



(a) SCM (b) Tyler (c) KPGH

(d) KPLR (e) CKPLR

Figure 5: Beampattern in angle-frequency domain for $K = 60$ and $\nu = 10$. The total size is $MNL = 80$ and the ranks are $R_A = 9, R_B = 2$.

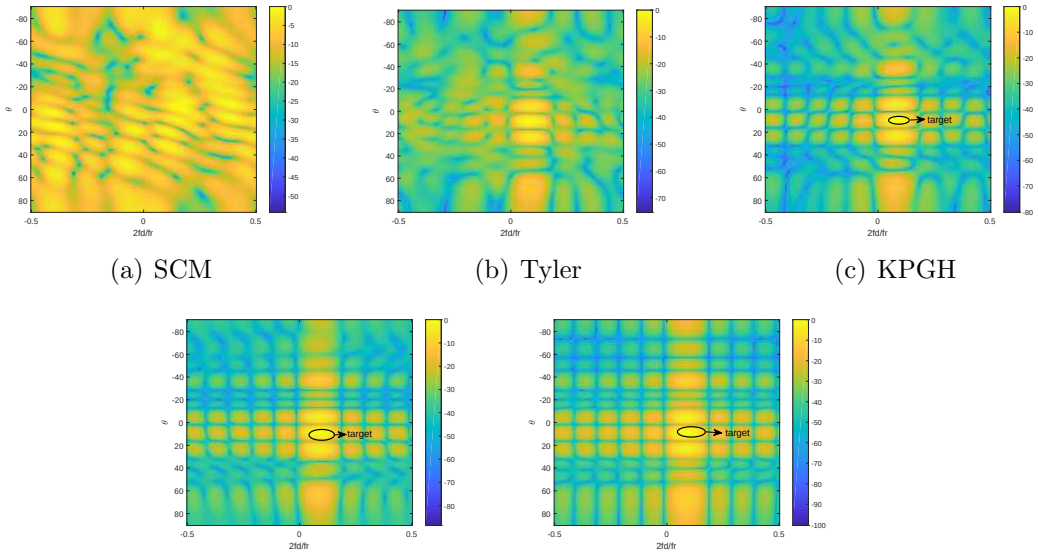


(a) SCM (b) Tyler (c) KPGH

(d) KPLR (e) CKPLR

Figure 6: Beampattern in angle-frequency domain for $K = 120$ and $\nu = 10$. The total size is $MNL = 80$ and the ranks are $R_A = 9, R_B = 2$.

Gaussian model.



(a) SCM (b) Tyler (c) KPGH

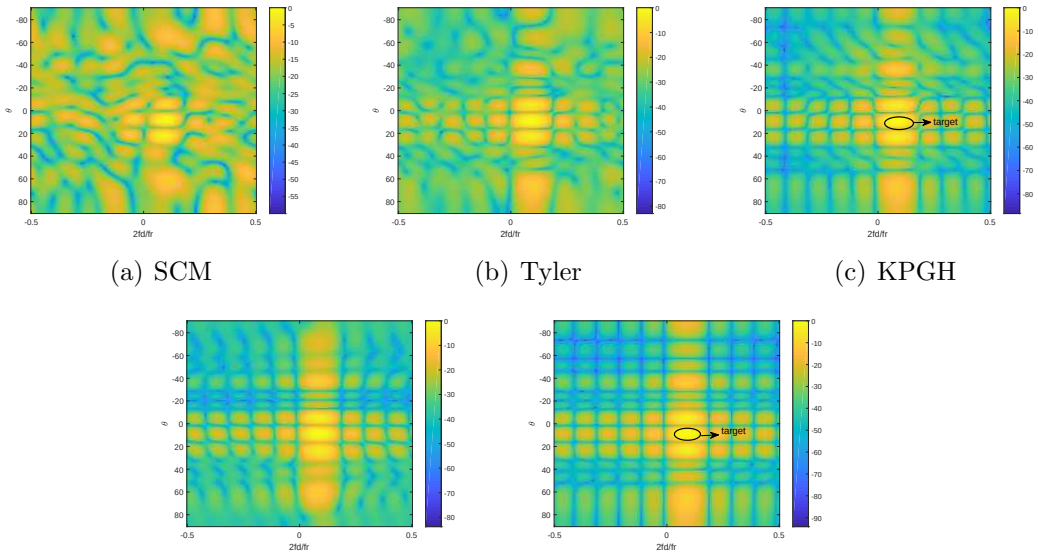
(d) KPLR (e) CKPLR

Figure 7: Beampattern in angle-frequency domain for $K = 60$ and $\nu = 1$. The total size is $MNL = 80$ and the ranks are $R_A = 9, R_B = 2$.

Figure 9 displays the SINR-Loss of the adaptive MIMO-STAP filters built with the different covariance matrix estimators (cf. section 2.3). Once again, we can notice that the proposed method achieves the best performance, as it is directly derived according to the considered noise model. Interestingly, the -3dB are even achieved at low sample support, which is due to the fact that the actual dimension of the estimation problem is greatly reduced compared to the ambient dimension when Kronecker-product and low rank models are combined.

5. Conclusion

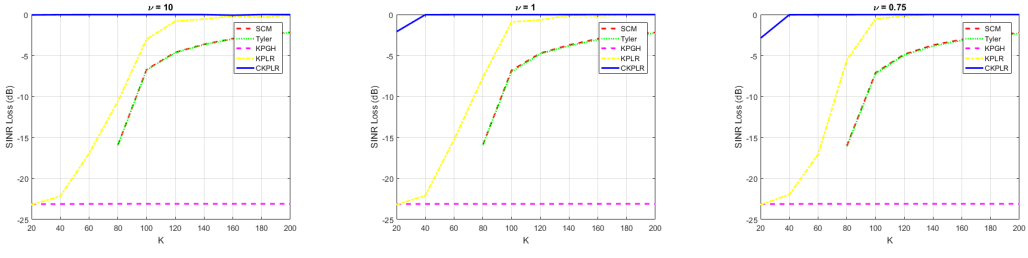
In this paper, we derived the MLE of the covariance matrix when the data follow a mixture of a Compound-Gaussian (CG) and a white Gaussian noise and when the covariance matrix of the CG contribution is assumed to be the Kronecker product of two low-rank matrices. To solve the corresponding optimization problem, we proposed an algorithm based on the block Majorization-Minimization framework. The proposed estimation method has



(a) SCM (b) Tyler (c) KPGH

(d) KPLR (e) CKPLR

Figure 8: Beampattern in angle-frequency domain for $K = 120$ and $\nu = 1$. The total size is $MNL = 80$ and the ranks are $R_A = 9, R_B = 2$.



(a) $\nu = 10$ (b) $\nu = 1$ (c) $\nu = 0.75$

Figure 9: SINR loss w.r.t K. The total size is $MNL = 80$, the ranks are $R_A = 9, R_B = 2$. The steering vector corresponds to the target parameters as in Fig. 5-8.

then been evaluated in terms of adaptive filtering performance on a MIMO-STAP radar setting. These simulations illustrated the interest of taking into account both the structure and the noise model in order to improve the performance of adaptive filters at low sample support.

Acknowledgment

This work was partially supported by ANR-ASTRID MARGARITA (ANR-17-ASTR-0015).

Appendix A. Proof of proposition 1

Before going to the formal proof of Proposition 1, we first state the following lemmas

Lemma 1. *(lemma 6 of [16]) The identity function can be upperbounded as:*

$$\begin{aligned} x \quad &\geq x^t \left(1 + \ln(x) - \ln(x^t)\right) \\ &\geq x^t \ln(x) + \text{const.} \end{aligned} \tag{A.1}$$

with equality achieved at $x = x^t$.

Lemma 2. *(Jensen's inequality, example 9 of [33]) The function $x \mapsto \sum_{q=1}^{Q} s_q \ln\left(f_q(x)\right)$ can be upperbounded as*

$$\begin{aligned} \sum_{q=1}^{Q} s_q \ln\left(f_q(x)\right) &\leq \sum_{q=1}^{Q} s_q \ln\left(f_q(x^t)\right) + \left(\sum_{q=1}^{Q} s_q\right) \ln\left(\frac{\sum_{q=1}^{Q} s_q \frac{f_q(x)}{f_q(x^t)}}{\sum_{q=1}^{Q} s_q}\right) \\ &\leq \left(\sum_{q=1}^{Q} s_q\right) \ln\left(\frac{\sum_{q=1}^{Q} s_q \frac{f_q(x)}{f_q(x^t)}}{\sum_{q=1}^{Q} s_q}\right) + \text{const.} \end{aligned} \tag{A.2}$$

with equality achieved at $x = x^t$.

We can now turn to the proof of Proposition 1:

Proof. By using $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ on the first term in (14), we have

$$\mathcal{L}\left(\tau_k\right) = -\sum_{pq=1} \frac{\tau_k a_p b_q \left[\tilde{\mathbf{x}}_k\right]_{pq}^2}{\tau_k a_p b_q + \sigma^2} + \sum_{pq=1} \ln\left(\tau_k a_p b_q + \sigma^2\right) + \text{const.} \tag{A.3}$$

We now use the inequality from lemma 1, with x parameterized as $\frac{\tau_k a_p b_q}{\tau_k a_p b_q + \sigma^2}$ to obtain the inequality

$$\begin{aligned}\sum_{pq=1} \frac{\tau_k a_p b_q \left[\tilde{\mathbf{x}}_k\right]_{pq}^2}{\tau_k a_p b_q + \sigma^2} &\geq \sum_{pq=1} \frac{\tau_k^t a_p b_q \left[\tilde{\mathbf{x}}_k\right]_{pq}^2}{\tau_k^t a_p b_q + \sigma^2} \times \left(\ln\left(\tau_k a_p b_q\right) - \ln\left(\tau_k a_p b_q + \sigma^2\right)\right) \\ &\quad + \text{const.} \\ &\geq \sum_{pq=1} \frac{\tau_k^t a_p b_q \left[\tilde{\mathbf{x}}_k\right]_{pq}^2}{\tau_k^t a_p b_q + \sigma^2} \times \left(\ln\left(\tau_k\right) - \ln\left(\tau_k a_p b_q + \sigma^2\right)\right) \\ &\quad + \text{const.}\end{aligned} \tag{A.4}$$

where const. absorbed the terms in $\ln(a_p b_q)$ to obtain the second inequality. Hence we upperbound the first term of the objective as

$$\begin{aligned}\mathcal{L}\left(\tau_k\right) \leq \sum_{pq=1} \frac{\tau_k^t a_p b_q \left[\tilde{\mathbf{x}}_k^{pq}\right]^2}{\tau_k^t a_p b_q + \sigma^2} \left(\ln\left(\tau_k a_p b_q + \sigma^2\right) - \ln\left(\tau_k\right)\right) \\ + \sum_{pq=1} \ln\left(\tau_k a_p b_q + \sigma^2\right) + \text{const.}\end{aligned} \tag{A.5}$$

which is compacted in

$$\mathcal{L}\left(\tau_k\right) \leq \sum_{pq=1} \left(\kappa_{pq}^k + 1\right) \ln\left(\tau_k a_p b_q + \sigma^2\right) - \mathcal{K}_p \ln\left(\tau_k\right) + \text{const.} \tag{A.6}$$

with constants in (16). The first term in this expression can be once again upperbounded using lemma 2, where we identify $s_{pq} = \left(\kappa_{pq}^k + 1\right)$ and $f_{pq}(\tau_k) = \tau_k a_p b_q + \sigma^2$. Hence we have

$$\sum_{pq=1} \left(\kappa_{pq}^k + 1\right) \ln\left(\tau_k a_p b_q + \sigma^2\right) \leq \mathcal{A}_k \ln\left(\omega_k \tau_k + \beta_k\right) + \text{const.} \tag{A.7}$$

with constants defined in (16). Finally, conclude with the inequality

$$\mathcal{L}\left(\tau_k\right) \leq \mathcal{A}_k \ln\left(\omega_k \tau_k + \beta_k\right) - \mathcal{K}_k \ln\left(\tau_k\right) + \text{const.} \tag{A.8}$$

which concludes the proof of the proposition. □

Appendix B. Proof of Proposition 3

First, we recall the following lemma:

Lemma 3. *(From proposition 4 of [44]) Let* $\mathbf{U} \in \mathbb{C}^{M\times R}$, $\mathbf{S} \in \mathbb{S}_+^{M\times M}$ *and* $\mathbf{\Lambda} \in \mathbb{S}_+^{R\times R}$. *Denote* $\lambda_{max}^{(\mathbf{S})}$ *the largest eigenvalue of* $\mathbf{S}$. *The function*

$$\mathcal{L}\left(\mathbf{U}\right) = \mathrm{Tr}\left[\mathbf{S}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\right] \tag{B.1}$$

can be upperbounded on the set $\{\mathbf{U} \in \mathbb{C}^{M\times R} | \mathbf{U}^H\mathbf{U} = \mathbf{I}_R\}$ *at point* $\mathbf{U}^t$ *by*

$$g\left(\mathbf{U}|\mathbf{U}^t\right) = \mathrm{Tr}\left[\mathbf{G}_t^H\mathbf{U}\right] + \mathrm{Tr}\left[\mathbf{U}^H\mathbf{G}_t\right] + \mathrm{const.} \tag{B.2}$$

with

$$\mathbf{G}_t = \left[\mathbf{\Lambda}\left(\mathbf{S} - \lambda_{max}^{(\mathbf{S})}\mathbf{I}\right)\mathrm{vec}\left(\mathbf{U}^t\right)\right]_{M,R} \tag{B.3}$$

More specifically, for a vector $\mathbf{u}$ *and for* $\mathbf{\Lambda} = 1$,

$$\begin{aligned}\mathbf{u}^H\mathbf{S}\mathbf{u} \quad \leq & \mathbf{u}_t^H\left(\mathbf{S} - \lambda_{max}^{(\mathbf{S})}\mathbf{I}\right)\mathbf{u} + \mathbf{u}^H\left(\mathbf{S} - \lambda_{max}^{(\mathbf{S})}\mathbf{I}\right)\mathbf{u}_t \\ & + \mathrm{const.}\end{aligned} \tag{B.4}$$

We can now turn to the proof of Proposition 3:

Proof. For each quadratic term in $\mathbf{u}_p$ in (24) we use the previous lemma to obtain a linear upper bound on the set $\{\mathbf{U} \in \mathbb{C}^{M\times R} | \mathbf{U}^H\mathbf{U} = \mathbf{I}_R\}$. This leads to the following surrogate function:

$$\mathcal{L}\left(\mathbf{U}_A\right) \leq \mathrm{Tr}\left[(\mathbf{W}_A^t)^H\mathbf{U}_A\right] + \mathrm{Tr}\left[\mathbf{U}_A^H\mathbf{W}_A^t\right] + \mathrm{const.} \tag{B.5}$$

with $\mathbf{W}_A^t$ defined in (27). □

References

[1] I. Reed, J. Mallett, L. Brennan, Rapid convergence rate in adaptive arrays, IEEE Transactions on Aerospace and Electronic Systems AES-10 (6) (1974) 853 – 863.

[2] E. Ollila, D. Tyler, V. Koivunen, H. Poor, Complex elliptically symmetric distributions: Survey, new results and applications, IEEE Transactions on Signal Processing 60 (11) (2012) 5597–5625.

[3] K. Ward, C. Baker, S. Watts, Maritime surveillance radar. part 1: Radar scattering from the ocean surface, Inst. Elect. Eng. Proc. F 137 (2) (1990) 51–62.

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

[5] M. Greco, F. Gini, M. Rangaswamy, Statistical analysis of measured polarimetric clutter data at different range resolutions, Radar, Sonar and Navigation, IEE Proceedings - 153 (6) (2006) 473–481.

[6] C. Chong, F. Pascal, J.-P. Ovarlez, M. Lesturgie, MIMO radar detection in non-gaussian and heterogeneous clutter, IEEE J. Sel. Top. Signal Process. 4 (1) (2010) 115–126.

[7] E. Ollila, D. Tyler, V. Koivunen, H. Poor, Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution, IEEE Signal Processing Letters 19 (12) (2012) 876–879.

[8] R. Palamà, M. Greco, F. Gini, Multistatic adaptive CFAR detection in non-gaussian clutter, EURASIP J. Adv. Signal Process. 2106 (107).

[9] A. Saucan, T. Chonavel, C. Sintes, J. Le Caillec, CPHD-DOA tracking of multiple extended sonar targets in impulsive environments, IEEE Transactions on Signal Processing 64 (5) (2016) 1147–1160. doi:10.1109/TSP.2015.2504349.

[10] C. Sintes, K. G. Foote, G. Llort-Pujol, J. Le Caillec, B. Solaiman, Performance prediction of a dual-baseline radar or sonar interferometer based on a vernier critical value concept, IEEE Journal of Oceanic Engineering 43 (4) (2018) 1114–1133. doi:10.1109/JOE.2017.2736220.

[11] A. Masnadi-Shirazi, B. D. Rao, A covariance-based superpositional CPHD filter for multisource DOA tracking, IEEE Transactions on Signal Processing 66 (2) (2018) 309–323. doi:10.1109/TSP.2017.2768025.

[12] M. Rangaswamy, F. Lin, K. Gerlach, Robust adaptive signal processing methods for heterogeneous radar clutter scenarios, Signal Processing 84 (2004) 1653 – 1665.

[13] R. Raghavan, Statistical interpretation of a data adaptive clutter subspace estimation algorithm, IEEE Transactions on Aerospace and Electronic Systems 48 (2) (2012) 1370 – 1384.

[14] G. Ginolhac, P. Forster, F. Pascal, J.-P. Ovarlez, Performance of two low-rank STAP filters in a heterogeneous noise, IEEE Transactions on Signal Processing 61 (1) (2013) 57–61.

[15] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, Clutter subspace estimation in low rank heterogeneous noise context, IEEE Transactions on Signal Processing 63 (9) (2015) 2173–2182.

[16] Y. Sun, A. Breloy, P. Babu, D. Palomar, F. Pascal, G. Ginolhac, Low-complexity algorithms for low rank clutter parameters estimation in radar systems, Signal Processing, IEEE Transactions on 64 (8) (2016) 1986 – 1998.

[17] Y. I. Abramovich, O. Besson, B. A. Johnson, Conditional expected likelihood technique for Compound Gaussian and Gaussian distributed noise mixtures, IEEE Transactions on Signal Processing 64 (24) (2016) 6640–6649. doi:10.1109/TSP.2016.2613073.

[18] O. Besson, Bounds for a mixture of low-rank compound-gaussian and white gaussian noises, IEEE Transactions on Signal Processing 64 (21) (2016) 5723–5732. doi:10.1109/TSP.2016.2603965.

[19] R. A. Maronna, Robust M-estimators of multivariate location and scatter, Annals of Statistics 4 (1) (1976) 51–67.

[20] D. Tyler, A distribution-free M-estimator of multivariate scatter, The Annals of Statistics 15 (1) (1987) 234–251.

[21] E. Conte, A. De Maio, G. Ricci, Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection, IEEE Transactions on Signal Processing 50 (8) (2002) 1908 – 1915.

[22] M. S. Greco, F. Gini, Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter, Signal Processing 82 (12) (2002) 1847–1859.

[23] A. Wiesel, T. Zhang, et al., Structured robust covariance estimation, Foundations and Trends® in Signal Processing 8 (3) (2015) 127–216.

[24] Y. Sun, P. Babu, D. Palomar, Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions, IEEE Transactions on Signal Processing 64 (14) (2016) 3576 – 3590.

[25] B. Meriaux, C. Ren, M. N. El Korso, A. Breloy, P. Forster, Robust estimation of structured scatter matrices in (mis) matched models, Signal Processing 165 (2019) 163–174.

[26] J. Ward, Space-Time Adaptive Processing for airborne radar, Tech. rep., Lincoln Lab., MIT, Lexington, Mass., USA (December 1994).

[27] S. Zhou, H. Liu, B. Liu, K. Yin, Adaptive MIMO radar target parameter estimation with kronecker-product structured interference covariance matrix, Signal Processing 92 (2012) 1177–1188. doi:10.1016/j.sigpro.2011.08.025.

[28] K. Greenwald, E. Zelnio, A. Hero, Robust SAR STAP via kronecker decomposition, IEEE Transactions on Aerospace and Electronic Systems 56 (2) (2016) 2612–2625.

[29] R. T. Suryaprakash, R. R. Nadakuditi, Algorithms for estimation of low-rank matrices with triple Kronecker structured singular vectors, in: IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015, pp. 1–4.

[30] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D. P. Palomar, Robust rank constrained kronecker covariance matrix estimation, in: 2016 50th Asilomar Conference on Signals, Systems and Computers, 2016, pp. 810–814. doi:10.1109/ACSSC.2016.7869159.

[31] W. Roberts, P. Stoica, J. Li, T. Yardibi, F. A. Sadjadi, Iterative adaptive approaches to MIMO Radar imaging, IEEE Journal of Selected Topics in Signal Processing 4 (1) (2010) 5–20. doi:10.1109/JSTSP.2009.2038964.

[32] D. Nion, N. D. Sidiropoulos, Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO Radar, IEEE Transactions on Signal Processing 58 (11) (2010) 5693–5705. doi:10.1109/TSP.2010.2058802.

[33] Y. Sun, P. Babu, D. Palomar, Majorization-Minimization algorithms in signal processing, communications, and machine learning, IEEE Transactions on Signal Processing PP (99) (2016) 1–1. doi:10.1109/TSP.2016.2601299.

[34] C.-Y. Chen, P. Vaidyanathan, A subspace method for mimo radar space-time adaptive processing, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Vol. 2, IEEE, 2007, pp. II–925.

[35] J. Xu, G. Liao, S. Zhu, L. Huang, H. C. So, Joint range and angle estimation using mimo radar with frequency diverse array, IEEE Transactions on Signal Processing 63 (13) (2015) 3396–3410.

[36] L. Brennan, F. Staudaher, Subclutter visibility demonstration, Tech. rep., Tech. Rep., RL-TR-92-21, Adaptive Sensors Incorporated (1992).

[37] N. A. Goodman, J. M. Stiles, On clutter rank observed by arbitrary arrays, IEEE Transactions on Signal processing 55 (1) (2006) 178–186.

[38] C. Chen, P. Vaidyanathan, MIMO radar space-time adaptive processing using prolate spheroidal wave functions, IEEE Transactions on Signal Processing 56 (2) (2008) 623 – 635.

[39] J. Manton, Optimization algorithms exploiting unitary constraints, IEEE Transactions on Signal Processing 50 (3) (2002) 635–650.

[40] S. T. Smith, Covariance, subspace, and intrinsic Cramér-Rao bounds, IEEE Transactions on Signal Processing 53 (5) (2005) 1610–1630.

[41] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, P. Larzabal, Existence and characterization of the covariance matrix maximum likelihood estimate in Spherically Invariant Random Processes, IEEE Transactions on Signal Processing 56 (1) (2008) 34 – 48.

[42] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, Robust covariance matrix estimation in heterogeneous low rank context, IEEE Transactions on Signal Processing 64 (22) (2016) 5794–5806. doi:10.1109/TSP.2016.2599494.

[43] A. Mian, A. Collas, A. Breloy, G. Ginolhac, J. P. Ovarlez, Robust low-rank change detection for multivariate sar image time series, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020) 3545–3556. doi:10.1109/JSTARS.2020.2999615.

[44] K. Benidis, Y. Sun, P. Babu, D. Palomar, Orthogonal sparse PCA and covariance estimation via procrustes reformulation, IEEE Transactions on Signal Processing 64 (23) (2016) 6211–6226.