Low-Rank Structured Covariance Matrix Estimation

Azer P. Shikhaliev, *Student Member, IEEE*, Lee C. Potter, *Senior Member, IEEE*, and Yuejie Chi, *Senior Member, IEEE*

***Abstract*—The covariance matrix estimation problem is posed in both the Bayesian and frequentist settings as the solution of a maximum a posteriori (MAP) or maximum likelihood (ML) optimization, respectively, when the true covariance consists of a known (or bounded) noise floor and a low-rank component. Persymmetric structure may also be assumed. The MAP and ML solutions with the non-convex rank constraint are shown to be a simple scalar thresholding of eigenvalues of a suitably translated and projected sample covariance matrix. No iterative optimization is required; therefore, the computation is suited to real-time applications. Our proof is short and elementary without resorting to the duality theory. Numerical results are presented to illustrate the improved estimation performance obtained by incorporating the structural constraints on the unknown covariance.**

***Index Terms*—Bayesian MAP, maximum likelihood, persymmetry, low rank.**

I. Introduction

The detection of a signal of interest among correlated interference is a problem of paramount importance in airborne radar systems [1], [2]. The canonical approach for the detection of such a signal is the one-step [3] or two-step [4] generalized likelihood ratio test (GLRT), which can similarly be thought of as the matched filter in colored Gaussian noise. However, this approach requires the knowledge of the covariance matrix of the interference signals, which consist of unwanted clutter returns, electronic jammers, and thermal noise. This covariance matrix is generally unknown and must be estimated from the nearby range cells in the radar datacube, assumed to be statistically homogeneous and target-free; however, in practice the secondary data are not ideal and often contaminated with clutter heterogeneities and/or targets, resulting in few homogeneous training samples being available for use in covariance matrix estimation.

Various approaches have been proposed to deal with this challenge, often by exploiting available knowledge about the covariance matrix in the adaptive detection scenario. For example, when a uniform linear array (ULA) is used and the clutter returns come from uncorrelated sources, the resulting clutter and noise covariance matrix is Toeplitz structured; the same result occurs when the detection problem is viewed from the perspective of a single antenna using uniformly spaced time samples in pulse-Doppler processing. In the space-time adaptive processing (STAP) scenario with multiple pulses and antennas, and in the absence of systems-level imperfections (array calibration errors, etc.), the resulting covariance matrix is Toeplitz-block-Toeplitz (TBT). Some of the prior works which invoke the Toeplitz constraint include [5], [6], and [7], while those invoking persymmetry include [8], [9], and [10]; furthermore, Toeplitz and TBT matrices may be embedded into circulant matrices [11]. In this work, in order to obtain closed-form solutions to the MAP and ML problems without resorting to computationally intensive iterative approaches, an approximation to the Toeplitz and TBT structures is used in the form of the persymmetric structure.

Another approach that has been considered for improving covariance matrix estimation in the low-snapshot regime is the Bayesian approach, most commonly invoking a complex inverse Wishart prior on the covariance matrix [12]–[14]. The physical meaning behind the use of this approach in STAP is that in many real-world scenarios, there is available knowledge of the parameters relevant to predicting the structure of the covariance matrix, such as platform velocity, crabbing angle, digital terrain data, etc., which can be used to construct a prior matrix $\mathbf{R_0}$ that characterizes the inverse Wishart distribution. For the Wishart prior, the resulting MAP estimate is simply a scaled sum of the unbiased sample covariance matrix and the prior matrix $\mathbf{R_0}$. This result can be interpreted as a "colored loading," namely, the linear combination of the sample covariance matrix with a non-diagonal matrix [15]–[17]. Likewise, colored loading itself can be viewed as a subset of "shrinkage," which is the linear combination of the sample covariance matrix with some positive definite matrix [17]–[20]. In a spirit similar to MAP estimation, convex programming may be used to determine a constrained covariance matrix closest to a specified positive semidefinite matrix [21]–[23]; adopting the spectral norm for distance, a lower bound on white thermal noise power, and an upper bound on condition number results in simple computation [23], rather than a computationally expensive iterative solution to the convex program.

A further approach has been to incorporate low-rankness, a non-convex constraint, into the covariance estimation problem. Due to the coupling between the azimuthal angle and the Doppler shift in airborne radar, the covariance of the clutter returns lies in a low-dimensional subspace whose rank can be predicted from system parameters (e.g., number of pulses and

Manuscript received January 9, 2019; revised March 3, 2019; accepted March 3, 2019. Date of publication March 27, 2019; date of current version April 1, 2019. This work was supported by industrial and government membership fees to the Center for Surveillance Research, a National Science Foundation I/UCRC. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mireille Boutin. *(Corresponding author: Lee C. Potter.)*

A. P. Shikhaliev and L. C. Potter are with the Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH 43210 USA (e-mail: shikhaliev.1@osu.edu; potter.36@osu.edu).

Y. Chi is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: yuejiechi@cmu.edu).

Digital Object Identifier 10.1109/LSP.2019.2906405

antennas, pulse repetition frequency, etc.) via Brennan's rule [2]. Thus, the interference covariance matrix can be decomposed into a low-rank clutter component and a scaled identity matrix representing the thermal noise, which is present in all electronic devices. Furthermore, the variance of the thermal noise itself can be estimated by running the radar system on receive-only mode when no active electronic interference is present.

Existing MAP formulations [12]–[14] admit a closed-form solution, but do not constrain the resulting estimate to conform to the physical constraints of low-rank clutter or an interval bound on thermal noise power. Likewise, closed-form solutions have been widely reported for maximum likelihood estimation for known thermal noise [15], [24], [25], or for both known thermal noise and low-rank clutter [26]–[28]; however, to the best of the authors' knowledge, no closed-form estimator has been reported when an additional symmetry constraint is enforced on the ML solution.

In this letter, we consider both MAP and ML estimation scenarios with simple, exact computation suitable for real-time implementation and yielding a performance improvement in many realistic scenarios relevant to airborne radar. We consider the structured covariance model, $\mathbf{R} = \mathbf{M} + \sigma^2\mathbf{I}$, with $\mathbf{M}$ low rank and an interval bound on σ^2. Estimators are also provided which additionally enforce persymmetry on $\mathbf{R}$. Our analysis is elementary and self-contained, and does not require duality theory, which might be of independent interest.

Notation: For matrices, we use boldface uppercase letters, and we use tr $(\mathbf{A})$ and $|\mathbf{A}|$ to denote the trace and determinant. We use superscripts ^{T}, ^{H}, and * for transpose, conjugate transpose, and conjugation, respectively; $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_2$ are the Frobenius and spectral norms. For vectors, we use boldface lower case letters; scalars are in non-bold font. We use $\mathrm{diag}(\mathbf{a})$ to denote a square diagonal matrix with entries given by $\mathbf{a}$, and $\mathrm{diag}(\mathbf{A})$ to denote the diagonal entries taken from a square matrix $\mathbf{A}$; $[\mathbf{A}]_{kl}$ is the (k,l)th entry of $\mathbf{A}$, while a_k is the k^{th} entry of a column vector, $\mathbf{a}$. Let $\mathbf{J}$ denote the p-by-p exchange matrix with entries $[\mathbf{J}]_{ij} = 1$ for $j = p - i + 1$ and zero otherwise, and let $\mathbf{I}$ denote the identity matrix, with the sizes of both matrices implied by context. Note that $\mathbf{J} = \mathbf{J}^T = \mathbf{J}^{-1}$. A matrix $\mathbf{M}$ is said to be persymmetric if $\mathbf{MJ} = \mathbf{JM}^T$. Finally, let $A \succeq B$ denote that $A - B$ is a positive semi-definite (psd) matrix.

II. Preliminaries

Consider iid, multivariate, complex circular Gaussian samples, $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ with zero mean and covariance matrix $\mathbf{R}$. Collect n samples (or, "snapshots") as columns in a data matrix, $\mathbf{X} \in \mathbb{C}^{p\times n}$. The joint density for $\mathbf{X}$ is

$$g(\mathbf{X};\mathbf{R}) = \pi^{-pn}|\mathbf{R}^{-1}|^n \exp\left\{-\text{tr}\ \left(\mathbf{R}^{-1}\mathbf{X}\mathbf{X}^H\right)\right\}. \quad (1)$$

The negative of the logarithm yields the ML risk

$$\text{tr}\ \left\{\mathbf{R}^{-1}\mathbf{X}\mathbf{X}^H\right\} - n\log|\mathbf{R}^{-1}| \quad (2)$$

where constants unrelated to $\mathbf{X}$ and $\mathbf{R}$ have been dropped. This risk function is convex in the precision matrix, $\mathbf{R}^{-1}$.

In contrast, a MAP estimator can be defined given a prior on the covariance, $\mathbf{R}$. To this end, we adopt a complex central inverse Wishart prior, which is a conjugate prior density for the covariance matrix in a multivariate normal model. For l samples, $\mathbf{z}_k \in \mathbb{C}^p$, drawn iid from $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma})$, the matrix $\mathbf{W} = \sum_{k=1}^{l}\mathbf{z}_k\mathbf{z}_k^H$ is complex central Wishart, $\mathcal{CW}(\boldsymbol{\Sigma}, l)$ [29], [30]. The matrix $\mathbf{A} = \mathbf{W}^{-1}$ has inverse Wishart distribution [31], $\mathcal{CIW}(\boldsymbol{\Sigma}^{-1}, l)$, with probability density proportional to $|\mathbf{A}|^{-(l+p)}\exp\{-\text{tr}\ (\boldsymbol{\Sigma}^{-1}\mathbf{A}^{-1})\}$. To incorporate the low-rank structure and an interval bound on the thermal noise power, we define a set $\mathcal{S} = \{\mathbf{R} : \mathbf{R} = \mathbf{M} + \sigma^2\mathbf{I},\ \mathbf{M} \succeq 0,\ \mathrm{rank}(\mathbf{M}) \le r\}$, with LB $\le \sigma^2 \le$ UB and $1 \le r \le p$; additionally, let $\mathcal{T}$ denote the set of persymmetric p-by-p matrices. We define a modified $\mathcal{CIW}(\mathbf{R}_0, l)$ distribution for the covariance matrix that is only nonzero on the set $\mathcal{S}\cap\mathcal{T}$:

$$f(\mathbf{R};\theta)\begin{cases}\propto |\mathbf{R}|^{-(l+p)}\exp\left\{-\text{tr}\ \left(\mathbf{R}_0\mathbf{R}^{-1}\right)\right\} & \mathbf{R}\in\mathcal{S}\cap\mathcal{T}\\ = 0, & \mathbf{R}\notin\mathcal{S}\cap\mathcal{T}\end{cases} \quad (3)$$

where $\theta = \{\mathbf{R}_0, l, r, \mathrm{LB}, \mathrm{UB}\}$. The MAP estimate of $\mathbf{R}$ is then given as the solution of the constrained optimization problem

$$\underset{\mathbf{R}\in\mathcal{S}\cap\mathcal{T}}{\arg\max}\quad g(\mathbf{X};\mathbf{R})f(\mathbf{R};\theta). \quad (4)$$

The negative of the logarithm yields the MAP risk

$$\text{tr}\ \left\{\mathbf{R}^{-1}(\mathbf{X}\mathbf{X}^H + \mathbf{R}_0)\right\} - (n+l+p)\log|\mathbf{R}^{-1}|. \quad (5)$$

The following lemma is an extension of the von Neumann trace inequality.

Lemma 1: [32], [33] Let $\mathbf{A}$ and $\mathbf{B}$ be psd Hermitian matrices, with $\mathbf{A} = \boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{\lambda}_{\boldsymbol{A}})\boldsymbol{\Phi}^H$ and $\mathbf{B} = \boldsymbol{\Psi}\,\mathrm{diag}(\boldsymbol{\lambda}_{\boldsymbol{B}})\boldsymbol{\Psi}^H$, where $\boldsymbol{\lambda}_{\boldsymbol{A}}$ ($\boldsymbol{\lambda}_{\boldsymbol{B}}$) are the eigenvalues, in non-increasing order, of $\mathbf{A}$ ($\mathbf{B}$), and $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ are unitary. Then,

$$\boldsymbol{\lambda}_{\mathbf{A}}^T\mathbf{J}\boldsymbol{\lambda}_{\boldsymbol{B}} \le \text{tr}\ (\mathbf{AB}) \le \boldsymbol{\lambda}_{\mathbf{A}}^T\boldsymbol{\lambda}_{\boldsymbol{B}} \quad (6)$$

with equality at the lower (upper) bound if and only if $\boldsymbol{\Psi} = \boldsymbol{\Phi}\mathbf{J}$ ($\boldsymbol{\Psi} = \boldsymbol{\Phi}$).

The trace inequality yields the following result for optimizing a cost function over the closed, but non-convex, set of p-by-p psd matrices with rank not exceeding $r \le p$ and with thermal noise power constrained to the interval LB $\le \sigma^2 \le$ UB.

Lemma 2: Let $\alpha > 0$, $\mathbf{S} = \mathbf{UDU}^H$ be psd, $\mathbf{D} = \mathrm{diag}\,(d_1,\ldots,d_p)$, $d_1 \ge d_2 \ge \cdots \ge d_p \ge 0$, and r an integer, $1 \le r \le p$. Then, the solution to the optimization problem

$$\begin{aligned}&\underset{\mathbf{M}\succeq 0,\,\sigma^2}{\arg\min}\,\text{tr}\ \left\{(\mathbf{M}+\sigma^2\mathbf{I})^{-1}\mathbf{S}\right\} - \alpha\log\left|\left(\mathbf{M}+\sigma^2\mathbf{I}\right)^{-1}\right|\\ &\qquad\text{s.t.}\quad \mathrm{rank}(\mathbf{M})\le r,\quad\text{and}\quad \mathrm{LB}\le\sigma^2\le\mathrm{UB}\end{aligned} \quad (7)$$

yields $(\widehat{\mathbf{M}} + \widehat{\sigma}^2\mathbf{I}) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \mathrm{diag}\,(\boldsymbol{\lambda})$ and

$$\lambda_k = \begin{cases}\max\left(\frac{d_k}{\alpha},\ \mathrm{LB}\right) & k = 1,\ldots,r\\ \min\left(\max\left(\frac{\overline{d}}{\alpha},\ \mathrm{LB}\right),\ \mathrm{UB}\right) & k = r+1,\ldots,p.\end{cases} \quad (8)$$

Here, $\overline{d}$ is the average of the noise space eigenvalues, $\overline{d} = \frac{1}{p-r}\sum_{k=r+1}^{p} d_k$.

Lemma 2 generalizes the results in [27], [28] to an interval bound on σ^2; for completeness and independent interest, we present a short and elementary proof here.

Proof: Define $\mathbf{P} = (\mathbf{M} + \sigma^2 \mathbf{I})^{-1}$ and rewrite the objective function as $\text{tr}\{\mathbf{PS}\} - \alpha \log |\mathbf{P}|$. Using eigendecompositions, let $\mathbf{S} = \mathbf{UDU}^H$ and $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^H$, where $\mathbf{D} = \text{diag}(d_1, \ldots, d_p)$ with $d_1 \geq \cdots \geq d_p$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \geq \cdots \geq \lambda_p$. Utilizing the trace inequality in Lemma 1, we can write

$$\text{tr}\{\mathbf{PS}\} - \alpha \log |\mathbf{P}| \geq \alpha \left(\sum_{k=1}^{p} \frac{d_k/\alpha}{\lambda_k} + \log \lambda_k \right), \quad (9)$$

where the equality holds if and only if $\mathbf{U} = \mathbf{V}$. Here, $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_p$ by hypothesis, and $\overline{d} = \frac{1}{p-r} \sum_{k=r+1}^{p} d_k$. Observe the sum of scalar functions, $h(x; c) = \frac{c}{x} + \log x$, inside the parentheses of the right-hand term, and note that $h(x; c)$ is increasing in c for $x > 0$. Thus, by the nonincreasing ordering of the d_k values and the iid thermal noise model, the optimization with rank constraint on $\mathbf{M}$ reduces to

$$\underset{\lambda_1, \ldots, \lambda_{r+1}}{\arg\min} \sum_{k=1}^{r} \left(\frac{d_k/\alpha}{\lambda_k} + \log \lambda_k \right) + \sum_{k=r+1}^{p} \left(\frac{d_k/\alpha}{\lambda_{r+1}} + \log \lambda_{r+1} \right)$$

$$= \underset{\lambda_1, \cdots, \lambda_{r+1}}{\arg\min} \sum_{k=1}^{r} h\left(\lambda_k; \frac{d_k}{\alpha}\right) + (p-r) h\left(\lambda_{r+1}; \frac{\overline{d}}{\alpha}\right) \quad (10)$$

subject to $\text{LB} \leq \lambda_k$, $k = 1, \ldots, r+1$ and $\lambda_{r+1} \leq \text{UB}$. Observe that $h(x; c)$ is strictly decreasing in x for $x < c$ and increasing for $x > c$. Thus, to minimize each $h(x; c)$ in Eq. (10) $\lambda_k = \max\left(\frac{d_k}{\alpha}, \text{LB}\right)$ for $k = 1, ..., r$ and $\lambda_{r+1} = \min(\max(\frac{\overline{d}}{\alpha}, \text{LB}), \text{UB})$. ■

Let $\phi(\mathbf{d}; \alpha, \text{LB}, \text{UB}, r)$, or the shorthand $\phi(\mathbf{d})$, denote the application of the thresholding in Eq. (8) to a vector $\mathbf{d}$.

As consequences of Lemma 2, we incorporate constraints into MAP and ML structured covariance estimation problems; the covariance is of the form $\mathbf{M} + \sigma^2 \mathbf{I}$ with low-rank $\mathbf{M}$ or low-rank persymmetric $\mathbf{M}$. In each case, the optimization admits a simple, closed-form solution despite the non-convexity of the rank constraint.

III. Rank-Constrained MAP

First consider constrained MAP estimators for a structured covariance matrix.

Proposition 1 (Rank-constrained MAP): Given iid, zero mean, multivariate complex circular Gaussian observations $\mathbf{X} \in \mathbb{C}^{p \times n}$, $\text{LB} \leq \sigma^2 \leq \text{UB}$, an inverse Wishart prior $\mathcal{CIW}(\mathbf{R}_0, l)$, $l \geq p$, and rank r, $1 \leq r \leq p$, the maximum a posteriori probability covariance estimate, $\widehat{\mathbf{R}} = \mathbf{M} + \sigma^2 \mathbf{I}$, with $\text{rank}(\mathbf{M}) \leq r$ and $\mathbf{M} \succeq 0$, is $\widehat{\mathbf{R}} = \mathbf{U}\,\text{diag}(\phi(\boldsymbol{\lambda}))\mathbf{U}^H$ with $\alpha = n + l + p$ and eigendecomposition $\left(\mathbf{XX}^H + \mathbf{R}_0\right) = \mathbf{U}\,\text{diag}(\boldsymbol{\lambda})\mathbf{U}^H$.

Proof: The MAP risk from Eq. (5) is equivalent to the cost in Eq. (7), with $\alpha = n + l + p$ and $\mathbf{S} = \mathbf{XX}^H + \mathbf{R}_0$. Thus, let $\mathbf{U}\,\text{diag}(\mathbf{d})\mathbf{U}^H$ be the eigendecomposition of the scaled and translated sample covariance matrix, $\mathbf{XX}^H + \mathbf{R}_0$. Then, by Eq. (8) in Lemma 2, the MAP estimate is $\mathbf{U}\,\text{diag}(\phi(\mathbf{d}))\mathbf{U}^H$. ■

Thus, the MAP covariance estimate, $\widehat{\mathbf{R}}$, is obtained by applying the thresholding operator of Eq. (8) to the eigenvalues of the scaled sample covariance matrix, $\mathbf{XX}^H$, translated by the inverse Wishart scale matrix, $\mathbf{R}_0$. Note that $\mathbf{XX}^H$ is n times the sample covariance matrix, and thus grows large as n increases, thereby increasing reliance of the MAP estimate on the data samples, rather than on the prior.

Proposition 2 (Rank-constrained Persymmetric MAP): Given iid, zero mean, multivariate complex circular Gaussian observations $\mathbf{X} \in \mathbb{C}^{p \times n}$, $\text{LB} \leq \sigma^2 \leq \text{UB}$, an inverse Wishart prior $\mathcal{CIW}(\mathbf{R}_0, l)$, $l \geq p$, and rank r, $1 \leq r \leq p$, the maximum a posteriori probability covariance estimate, $\widehat{\mathbf{R}} = \mathbf{M} + \sigma^2 \mathbf{I}$, with $\text{rank}(\mathbf{M}) \leq r$ and $\mathbf{M} \succeq 0$ persymmetric, is $\widehat{\mathbf{R}} = \mathbf{U}\,\text{diag}(\phi(\mathbf{d}))\mathbf{U}^H$ with $\alpha = n + l + p$ and eigendecomposition $\mathbf{S} = \mathbf{U}\,\text{diag}(\mathbf{d})\mathbf{U}^H$ for

$$\mathbf{S} = \tfrac{1}{2} \left\{ \left(\mathbf{XX}^H + \mathbf{R}_0\right) + \mathbf{J}\left(\mathbf{XX}^H + \mathbf{R}_0\right)^T \mathbf{J} \right\}. \quad (11)$$

Proof: An invertible persymmetric matrix has persymmetric inverse; so, let $\mathbf{P} = (\mathbf{M} + \sigma^2 \mathbf{I})^{-1}$ be a persymmetric matrix. Then, defining $\mathbf{S}_0 = \mathbf{XX}^H + \mathbf{R}_0$, we have

$$\begin{aligned} \text{tr}\,\{\mathbf{PS}_0\} &= \frac{1}{2}\left\{\text{tr}\,\{\mathbf{PS}_0\} + \text{tr}\,\left\{\mathbf{JP}^T\mathbf{JS}_0\right\}\right\} \\ &= \tfrac{1}{2}\text{tr}\,\left\{\mathbf{P}\left\{\mathbf{S}_0 + \mathbf{JS}_0^T\mathbf{J}\right\}\right\} = \text{tr}\,\{\mathbf{PS}\} \end{aligned}$$

where we utilized the invariance property of the trace operation to cyclic permutations and transposition. Let $\mathbf{U}\,\text{diag}(\mathbf{d})\mathbf{U}^H$ be the eigendecomposition of $\mathbf{S}$ from Eq. (11); by Lemma 2, the rank-constrained MAP risk is minimized by $\widehat{\mathbf{R}} = \mathbf{U}\,\text{diag}\{\phi(\mathbf{d})\}\mathbf{U}^H$. It remains only to show that the Hermitian matrix $\widehat{\mathbf{R}}$ is persymmetric. To this end, for distinct eigenvalues let $\{\mathbf{u}, d\}$ be an eigenvector/eigenvalue pair for $\mathbf{S}$ and note

$$d\mathbf{u} = \mathbf{Su} = \mathbf{JS}^*\mathbf{Ju} \Rightarrow \mathbf{S}^*\mathbf{Ju} = d\mathbf{Ju},$$

implying that $\{\mathbf{Ju}, d^*\}$ is also an eigenvector/eigenvalue pair. Because d is real-valued, we learn that $\mathbf{u}$ is either symmetric ($\mathbf{Ju} = \mathbf{u}^*$) or skew-symmetric ($\mathbf{Ju} = -\mathbf{u}^*$). Thus, $\mathbf{uu}^H = \mathbf{Ju}^*\mathbf{u}^T\mathbf{J} = \mathbf{J}(\mathbf{uu}^H)^*\mathbf{J}$ is persymmetric. The same conclusion holds for the case of repeated eigenvalues [34]. Thus, because $\phi(d_i)$ is real-valued, it follows that $\widehat{\mathbf{R}} = \sum_{i=1}^{p} \phi(d_i)\mathbf{u}_i\mathbf{u}_i^H$ is the sum of rank-1 persymmetric matrices and hence persymmetric. ■

Note that $\frac{1}{2n}\{\mathbf{XX}^H + \mathbf{J}(\mathbf{XX}^H)^T\mathbf{J}\}$ is the projection of the sample covariance matrix onto the closed, convex set of Hermitian persymmetric matrices [34]. So, the low-rank persymmetric MAP covariance estimate, $\widehat{\mathbf{R}}$, is obtained by applying the thresholding operator of Eq. (8) to the eigenvalues of the scaled sample covariance matrix, $\mathbf{XX}^H$, translated by the inverse Wishart scale matrix, $\mathbf{R}_0$, and projected to the set of persymmetric Hermitian matrices. Thus, fast computation is retained while leveraging prior knowledge of the covariance, its thermal noise floor, low-rank clutter, and covariance structure. Further, the persymmetry may be exploited for computational efficiency in computing STAP weight vectors using the estimated covariance, $\widehat{\mathbf{R}}$ [35]–[37].

Remark 1: The result for persymmetric structure immediately generalizes to $\mathbf{T}$-conjugate symmetry. Let $\mathbf{T} \in \mathbb{C}^{p \times p}$ be a nontrivial involution; i.e., $\mathbf{T} = \mathbf{T}^{-1} \neq \pm\mathbf{I}$. A matrix $\mathbf{R}$ is said to be $\mathbf{T}$-conjugate [38] if $\mathbf{R} = \mathbf{TR}^*\mathbf{T}$.

IV. Rank-Constrained ML

We next consider similar extension of ML estimators.

Proposition 3 (Rank-constrained ML): Given iid, zero mean, multivariate complex circular Gaussian observations $\mathbf{X} \in \mathbb{C}^{p \times n}$, LB $\leq \sigma^2 \leq$ UB, and rank r, $1 \leq r \leq p$, the maximum likelihood covariance estimate, $\widehat{\mathbf{R}} = \mathbf{M} + \sigma^2 \mathbf{I}$, with $\mathrm{rank}(\mathbf{M}) \leq r$ and $\mathbf{M} \succeq 0$, is $\widehat{\mathbf{R}} = \mathbf{U}\,\mathrm{diag}(\phi(\mathbf{d}))\mathbf{U}^H$ with $\alpha = n$ and eigendecomposition $\mathbf{X}\mathbf{X}^H = \mathbf{U}\,\mathrm{diag}(\mathbf{d})\mathbf{U}^H$.

Proof: The ML cost from Eq. (2) is the cost in Eq. (7), with $\alpha = n$ and $\mathbf{S} = \mathbf{X}\mathbf{X}^H$. The result then follows immediately from Lemma 2. ■

Proposition 3 has been previously shown using duality theory for special cases: $r = p$ and σ^2 known [24]; $r \leq p$ and σ^2 known [26]–[28]; $r \leq p$ and $\sigma^2 \geq$ LB [28]. Next, we extend the result to include the persymmetry constraint into the ML estimation.

Proposition 4 (Rank-constrained Persymmetric ML): Given iid, zero mean, multivariate complex circular Gaussian observations $\mathbf{X} \in \mathbb{C}^{p \times n}$, LB $\leq \sigma^2 \leq$ UB, and rank r, $1 \leq r \leq p$. The maximum likelihood covariance estimate, $\widehat{\mathbf{R}} = \mathbf{M} + \sigma^2 \mathbf{I}$, with $\mathrm{rank}(\mathbf{M}) \leq r$ and $\mathbf{M} \succeq 0$ persymmetric, is $\widehat{\mathbf{R}} = \mathbf{U}\,\mathrm{diag}(\phi(\mathbf{d}))\mathbf{U}^H$ with $\alpha = n$ and eigendecomposition $\mathbf{S} = \mathbf{U}\,\mathrm{diag}(\mathbf{d})\mathbf{U}^H$ for

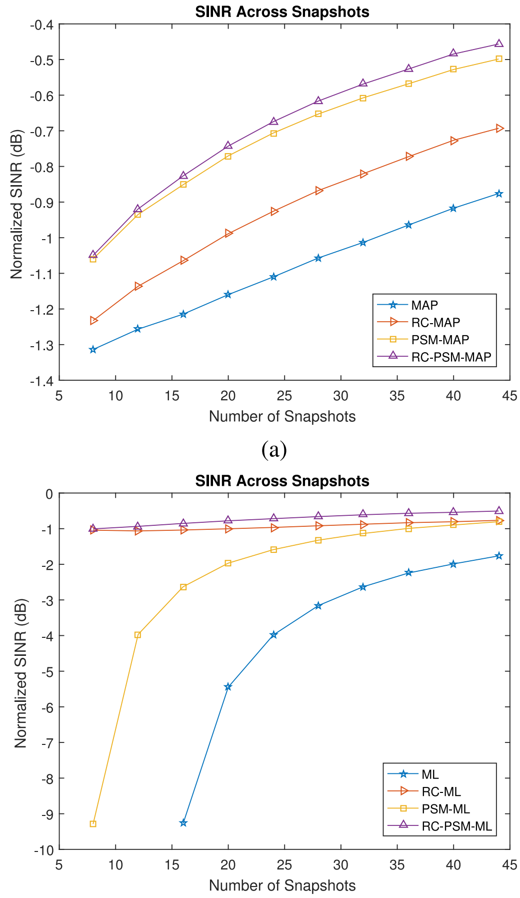
$$\mathbf{S} = \tfrac{1}{2}\left\{\mathbf{X}\mathbf{X}^H + \mathbf{J}\left(\mathbf{X}\mathbf{X}^H\right)^T \mathbf{J}\right\}. \tag{12}$$

Proof: The proof proceeds in the same manner as the proof of Proposition 2 with $\alpha = n$ and $\mathbf{S}$ as given in Eq. (12). ■

V. Numerical Experiments

For the MAP scenario, the true covariance $\mathbf{R}$ was drawn from f in Section II, which was approximated by drawing iid from $\mathcal{CIW}(\mathbf{R}_0, l = p+1)$, then projecting each random draw to the set $\mathcal{S} \cap \mathcal{T}$. The scale matrix was $\mathbf{R}_0 = \mathbf{M} + \sigma^2 \mathbf{I}$, with $\mathbf{M}$ constructed from the $r = 5$ largest eigenvectors and corresponding eigenvalues of the exponential tapered model [13], [14] with the i,jth element given as $\rho^{|i-j|}, i,j = 1,\ldots,p$, $\rho = 0.9$, $p = 16$, and $\sigma^2 = 1$ (i.e., LB $=$ UB for known thermal noise [15]). For the ML case, the true covariance was set equal to the scale matrix, $\mathbf{R} = \mathbf{R}_0$.

Eight covariance estimators were computed: the maximum a posteriori (MAP), rank-constrained MAP (RC-MAP, given in Prop. 1), persymmetric MAP (PSM-MAP, given in Eq. (11)), rank-constrained persymmetric MAP (RC-PSM-MAP, given in Prop. 2), and their ML counterparts: ML, RC-ML, given in Prop. 3, PSM-ML, given in Eq. (12), and RC-PSM-ML, given in Prop. 4, respectively. Of these eight estimators, MAP and ML are unconstrained (i.e., $r = p$, LB $= 0$, and UB $= \infty$), and they were computed as $\frac{1}{n+l+p}(n\mathbf{X}\mathbf{X}^H + l\mathbf{R_0})$ and $\frac{1}{n}\mathbf{X}\mathbf{X}^H$, respectively. For each estimated covariance matrix $\widehat{\mathbf{R}}$, the normalized signal-to-noise ratio (SINR) [1], [28] was computed as $\frac{(\mathbf{s}^H \widehat{\mathbf{R}}^{-1}\mathbf{s})^2}{(\mathbf{s}^H \widehat{\mathbf{R}}^{-1}\mathbf{R}\widehat{\mathbf{R}}^{-1}\mathbf{s})(\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s})}$ with the steering vector $\mathbf{s} = \frac{1}{\sqrt{p}}[1,\ldots,1]^T$. Results are omitted for the unconstrained ML estimator for $n < p$ because of its singularity. The results were averaged over 10000 Monte-Carlo simulations for each scenario.



(a)

(b)

Fig. 1. Normalized SINR loss versus number of samples, n, for (a) maximum a posteriori and (b) maximum likelihood covariance matrix estimators.

As expected, the results show increasing performance with increasing knowledge of the covariance matrix structure. In both the MAP and ML scenarios, the estimators employing persymmetry and/or rank constraints outperform their less constrained counterparts, demonstrating the efficacy of the simple, closed-form solutions given. The results further suggest that, given a reasonable prior, the MAP estimate can outperform ML. Modest SINR gains are observed in the ML case without additional computational complexity; significant relative SINR gains are observed in the MAP case. The intuitively expected SINR gain from persymmetry is observed in Figure 1(a); for example, comparing RC-PSM-MAP to RC-MAP at SINR $= -0.85$ shows a 50% reduction in the required number of snapshots.

VI. Conclusion

MAP and ML covariance estimators were extended to low-rank structured covariance matrices. Structure was a low-rank component plus a noise floor, $\mathbf{R} = \mathbf{M} + \sigma^2 \mathbf{I}$; additional structure of persymmetry was also considered. Estimates were obtained as simple, closed-form solutions, despite the non-convexity of the set of low-rank $\mathbf{M}$. Numerical simulations confirmed the intuition that rank and symmetry constraints improve the accuracy of a covariance matrix estimate, and allow for its estimation even when the number of homogeneous, target-free samples is less than the matrix dimension.

References

[1] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA, USA: Artech House, 2014.

[2] J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Lab., Lexington, MA, USA, Tech. Rep. TR-1015, 1994.

[3] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, Mar. 1986.

[4] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[5] D. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2194–2198, Oct. 1991.

[6] B. Kang, V. Monga, and M. Rangaswamy, "Computationally efficient Toeplitz approximation of structured covariance under a rank constraint," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 775–785, Jan. 2015.

[7] T. A. Barton and S. T. Smith, "Structured covariance estimation for space-time adaptive processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1997, pp. 3493–3496.

[8] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 124–127, Jan. 1980.

[9] A. De Maio, "Maximum likelihood estimation of structured persymmetric covariance matrices," *Signal Process.*, vol. 83, no. 3, pp. 633–640, 2003.

[10] G. Ginolhac, P. Forster, F. Pascal, and J. P. Ovarlez, "Exploiting persymmetry for low-rank space time adaptive processing," in *Proc. 20th Eur. Signal Process. Conf.*, Aug. 2012, pp. 81–85.

[11] D. R. Fuhrmann and T. Barton, "Estimation of block-Toeplitz covariance matrices," in *Proc. 24th Asilomar Conf. Signals, Syst. Comput.*, 1990, pp. 779–783.

[12] A. De Maio, A. Farina, and M. Wicks, "KB-GLRT: Exploiting knowledge of the clutter ridge in airborne radar," *IEE Proc.—Radar, Sonar Navig.*, vol. 152, no. 6, pp. 421–428, Dec. 2005.

[13] A. De Maio, A. Farina, and G. Foglia, "Knowledge-aided Bayesian radar detectors & their application to live data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 170–183, Jan. 2010.

[14] P. Wang, Z. Sahinoglu, M. O. Pun, H. Li, and B. Himed, "Knowledge-aided adaptive coherence estimator in stochastic partially homogeneous environments," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 193–196, Mar. 2011.

[15] K. Gerlach and M. L. Picciolo, "Airborne/spacebased radar STAP using a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 269–281, Jan. 2003.

[16] B. Tang, Y. Zhang, and J. Tang, "Robust fast maximum likelihood with assumed clutter covariance algorithm for adaptive clutter suppression," *IET Radar, Sonar Navig.*, vol. 8, no. 9, pp. 1184–1194, 2014.

[17] M. Riedl and L. C. Potter, "Multi-model shrinkage for knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 5, pp. 2601–2610, Oct. 2018.

[18] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

[19] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jul. 2008.

[20] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.

[21] A. De Maio, S. De Nicola, L. Landi, and A. Farina, "Knowledge-aided covariance matrix estimation: A MAXDET approach," *IET Radar, Sonar Navig.*, vol. 3, no. 4, pp. 341–356, 2009.

[22] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum likelihood estimation of a structured covariance matrix with a condition number constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, Jun. 2012.

[23] A. Aubry, A. De Maio, and L. Pallotta, "A geometric approach to covariance matrix estimation and its applications to radar problems," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 907–922, Feb. 2018.

[24] M. Steiner and K. Gerlach, "Fast converging adaptive processor for a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[25] A. Aubry, A. De Maio, and V. Carotenuto, "Optimality claims for the FML covariance estimator with respect to two matrix norms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 3, pp. 2055–2057, Jul. 2013.

[26] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 122–148, 1963.

[27] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 5, pp. 1235–1248, Oct. 1983.

[28] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–515, Jan. 2014.

[29] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–177, 2004.

[30] L. L. Scharf and P. J. Schreier, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[31] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Syst. Signal Process.*, vol. 5, pp. 263–279, 1994.

[32] L. Mirsky, "On the trace of matrix products," *Math. Nachrichten*, vol. 20, pp. 171–174, 1959.

[33] C. M. Theobald, "An inequality for the trace of the product of two symmetric matrices," *Math. Proc. Cambridge Philos. Soc.*, vol. 77, pp. 265–267, 1975.

[34] I. S. Pressman, "Matrices with multiple symmetry properties: applications of centrohermitian and perhermitian matrices," *Linear Algebra Appl.*, vol. 284, pp. 239–258, 1998.

[35] G. Heinig and K. Rost, "Fast algorithms for Toeplitz and Hankel matrices," *Linear Algebra Appl.*, vol. 435, pp. 1–59, 2011.

[36] H. Fassbender and K. Ikramov, "Computing matrixvector products with centrosymmetric and centrohermitian matrices," *Linear Algebra Appl.*, vol. 364, pp. 235–241, 2003.

[37] H. Fassbender and D. Kressner, "Structured eigenvalue problems," *GAMM-Mitteilungen*, vol. 29, no. 2, pp. 297–318, 2006. [Online]. Available: https://doi.org/10.1002/gamm.201490035

[38] W. F. Trench, "Characterization and properties of matrices with generalized symmetry or skew symmetry," *Linear Algebra Appl.*, vol. 377, pp. 207–218, 2004.