Advanced Methods for MLE of Toeplitz Structured Covariance Matrices With Applications to Radar Problems

Augusto Aubry, *Senior Member, IEEE*, Prabhu Babu, Antonio De Maio, *Fellow, IEEE*, and Massimo Rosamilia, *Member, IEEE*

***Abstract*— This work considers Maximum Likelihood Estimation (MLE) of a Toeplitz structured covariance matrix. In this regard, an equivalent reformulation of the MLE problem is introduced, and two iterative algorithms are proposed for the optimization of the equivalent statistical learning framework. Both strategies are based on the Majorization Minimization (MM) paradigm and hence enjoy nice properties such as monotonicity and ensured convergence to a stationary point of the equivalent MLE problem. The proposed framework is also extended to deal with MLE of other practically relevant covariance structures, namely, the banded Toeplitz, block Toeplitz, and Toeplitz-block-Toeplitz. Through numerical simulations, it is shown that the new methods provide excellent performance levels in terms of both mean square estimation error (which is very close to the benchmark Cramér-Rao Bound (CRB)) and signal-to-interference-plus-noise ratio, especially in comparison with state-of-the art strategies. Moreover, the estimation task is accomplished with a remarkable reduction in computational complexity compared with a standard approach relying on a Semidefinite Programming (SDP) solver.**

***Index Terms*— Toeplitz covariance matrix, maximum likelihood estimation, banded Toeplitz, block-Toeplitz, Toeplitz-block-Toeplitz, adaptive radar signal processing, array processing, spectral estimation.**

I. Introduction

ESTIMATION of the data covariance matrix has diverse applications in radar signal processing, such as direction of arrival estimation, target detection, adaptive beamforming, and sidelobe canceller design [1], [2], [3], [4]. In these situations, the interference covariance matrix is estimated from the secondary/training data, which are assumed target-free and collected from spatial and/or temporal returns corresponding to range cells close to the one of interest. When the data follows a complex, zero-mean, circular Gaussian distribution, it is well known that the Sample Covariance Matrix (SCM) is the unstructured Maximum Likelihood (ML) estimate of the covariance matrix. However, in the presence of a small number of training data and/or when mismatches in training data spectral properties occur, it does not always represent a reliable choice for the covariance inference [5], [6]. A well-known strategy, often discussed in the open literature to improve the performance of a covariance estimator, relies on the incorporation of some *a priori* knowledge about its underlying structure. For instance, in some radar/sensing applications, it is customary to suppose that data come from a stationary Gaussian random process, leading to a Hermitian symmetric Toeplitz Structured Covariance (TSC) matrix. Leveraging this information, one can obtain (under the design conditions) a more reliable estimator than the SCM [7]. Aside radar applications, the estimation of a TSC matrix is encountered in speech recognition [8], spectral estimation [2], gridless compressive sensing [9], [10], [11], and hyperspectral imaging [12].

So far, several algorithms have been proposed for estimating a TSC matrix. Let us first discuss those for ML Estimation (MLE). According to the Caratheodory parametrization [2], [13], [14] a Toeplitz covariance matrix $\mathbf{T} \in \mathbb{H}^{m\times m}$ can always be decomposed as[1]

$$\mathbf{T} = \mathbf{A}\tilde{\mathbf{P}}\mathbf{A}^H; \ [\tilde{\mathbf{P}}]_{k,k} \geq 0 \ , \tag{1}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_1} & \cdots & e^{j\omega_r} \\ \vdots & \ddots & \vdots \\ e^{j(m-1)\omega_1} & \cdots & e^{j(m-1)\omega_r} \end{bmatrix}, \tilde{\mathbf{P}} = \begin{bmatrix} \tilde{p}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{p}_r \end{bmatrix}, \tag{2}$$

ω_i and $\tilde{p}_i$, $i = 1, 2, \cdots, r \leq m$, denote some angular frequencies and their corresponding powers while r indicates the rank of $\mathbf{T}$. Capitalizing on this parametrization, Circulant Embedding (CE) of Toeplitz matrix ([16], [17], [18]) can

[1]Notice that the parametrization is unique provided that the rank of $\mathbf{T} < m$ [15].

Received 28 December 2023; revised 15 July 2024; accepted 26 September 2024. Date of publication 11 October 2024; date of current version 22 November 2024. The work of Augusto Aubry and Antonio De Maio was supported in part by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"); and in part by the European Union in the NextGenerationEU plan through the Italian program "Bando PRIN 2022," D.D. 104/2022 (PE7, project "CIRCE," code E53D23000410006). The work of Massimo Rosamilia was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"). An earlier version of this paper was presented in part at the 2022 IEEE Radar Conference (RadarConf22) [DOI: 10.1109/RadarConf2248738.2022.9764189] *(Corresponding author: Antonio De Maio.)*

Augusto Aubry, Antonio De Maio, and Massimo Rosamilia are with the Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli "Federico II," 80125 Naples, Italy, and also with National Inter-University Consortium for Telecommunications, 43124 Parma, Italy (e-mail: augusto.aubry@unina.it; ademaio@unina.it; massimo.rosamilia@unina.it).

Prabhu Babu is with CARE, IIT Delhi, New Delhi 110016, India (e-mail: prabhubabu@care.iitd.ac.in).

Communicated by A. Tajer, Associate Editor for Machine Learning and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3474977.

Digital Object Identifier 10.1109/TIT.2024.3474977

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

be used to compute approximately the ML estimate of $\mathbf{T}$. According to CE, a Positive SemiDefinite (PSD) $m \times m$ Toeplitz matrix is modeled as

$$\mathbf{T} = \tilde{\mathbf{F}}\mathbf{P}\tilde{\mathbf{F}}^H; \;\; \mathbf{P} = \text{diag}([p_1, p_2, \cdots, p_L]), p_k \geq 0, \tag{3}$$

where $\tilde{\mathbf{F}} = [\mathbf{I}_{m\times m}\, \mathbf{0}_{m\times L-m}]\mathbf{F}$, $\mathbf{I}_{m\times m}$ is the identity matrix of size $m\times m$, $\mathbf{0}_{m\times L-m}$ is the zero matrix of size $m\times L-m$, $\mathbf{F}$ is the normalized Discrete Fourier Transform (DFT) matrix of size $L \geq 2m-1$ and $\mathbf{P}$ is a diagonal matrix of size $L \times L$ with diagonal elements $p_k \geq 0$. Therefore, the matrix $\mathbf{T}$ is completely parameterized by the diagonal matrix $\mathbf{P}$. Although estimating the Toeplitz covariance matrix using CE seems attractive, the representation in (3) is valid only for a subset of Toeplitz covariance matrices. This can be intuitively justified because the Caratheodory parametrization in (1) does not give restrictions on the frequencies spacing, while the CE in (3) strictly requires the frequencies to lie on the Fourier grid. Hence, for some Toeplitz matrices, the parametrization in (3) is only approximated. Based on CE, [19] and [20] have proposed an iterative algorithm based on Expectation-Maximization (EM) for MLE of $\mathbf{T}$. By modifying the M step in the EM procedure, in [21] the technique has been extended to deal with the banded Toeplitz covariance case. In [22], still leveraging CE framework, a Majorization Minimization (MM) based optimization, with faster convergence than the EM of [19] and [20], has been introduced. In [23] a closed-form estimator has been designed by invoking the extended invariance principle to deal with the Toeplitz constraint. In [24], an efficient approximation of a Toeplitz covariance matrix under a rank constraint has been handled forcing the eigenvectors to be the same as those of the SCM whereas the Toeplitz constraint has been explicitly imposed while estimating the eigenvalues. Finally, some attempts to handle the MLE problem without frequencies restrictions have been pursued in [25], [26]. Other than the MLE, several other alternative paradigms have been considered for the problem at hand. Recently, in [27] the Toeplitz structure is forced together with a condition number constraint via SCM projection onto a suitable constraint set. Other geometric based approaches for the TSC estimation have also been proposed in [28] and [29].

In this work,[2] two iterative algorithms referred to as **A**lternating Projection Based **TO**eplitz Covariance **M**atrix Estimation 1 (ATOM1) and ATOM2 are devised leveraging a suitable reformulation of the MLE problem and the MM framework. Both ATOM1 and ATOM2 involve the construction of a bespoke surrogate function (s.f.) along with its optimization. Specifically, the two procedures construct distinct s.f. and therefore solve different surrogate minimization problems. While ATOM1 addresses the surrogate minimization problem using the Alternating Direction Method of Multipliers (ADMM), ATOM2 handles it either via alternating projection or Dykstra's algorithm. However, both the procedures directly estimate the Toeplitz covariance matrix without forcing a reparametrization via the CE. Moreover, ATOM2 is also extended to include other constraints, such as banded Toeplitz, block-Toeplitz, and Toeplitz-block-Toeplitz structures. This is among the most valuable contributions of this study since it addresses, via a unified framework, a quite general problem which notably enables the incorporation of convex constraints (in addition to the Toeplitz structure). The major contributions of this paper can be summarized as follows:

1) Two iterative algorithms ATOM1 and ATOM2 are proposed based on the MM framework to address MLE of a Toeplitz covariance matrix. Their computational complexities are thoroughly discussed. Also, the convergence of the procedures to a stationary point of the equivalent MLE problem is established.
2) The extensions of ATOM2 to handle additional covariance structures, such as banded Toeplitz, block-Toeplitz, and Toeplitz-block-Toeplitz, which is the main achievement of this study, being ATOM2 capable of including additional (other than Toeplitz) constraints (modeling convex sets) in the estimation process with convergence guarantees and a reasonable computational demand.
3) The derivation of the Cramér-Rao Bound (CRB) for the estimation of Toeplitz, banded Toeplitz, and Toeplitz-block-Toeplitz covariance matrices are provided.
4) Performance comparisons of the proposed algorithms (included their extensions) with some state-of-the-art procedures via numerical simulations are illustrated, using the Mean Square Error (MSE) and the Signal-to-Interference-plus-Noise Ratio (SINR) (for case studies related to radar applications) as performance metrics.

The organization of the paper is as follows. The MLE problem of Toeplitz covariance matrix for complex, zero-mean, circular Gaussian observations is formulated in Section II. In Section III, ATOM1 and ATOM2 algorithms are proposed, along with a discussion on their computational complexity and implementation aspects. Also, their convergence properties are studied. At the end of this section, the extension of ATOM2 to handle additional constraints along with the Toeplitz requirement is discussed too. In Section IV, the CRB for the estimation of Toeplitz, banded Toeplitz, and Toeplitz-block-Toeplitz covariance matrices is computed. In Section V, the proposed algorithms are compared with some state-of-the-art techniques, and finally, concluding remarks are given in Section VI.

A. Notation

Throughout the paper, bold capital and bold small letter denote matrix and vector, respectively. A scalar is represented by a small letter. The value taken by an optimization vector $\boldsymbol{x}$ at the t^{th} iteration is denoted by $\boldsymbol{x}_t$. Furthermore, $\mathbb{R}$ is used to denote the set of real numbers, $\mathbb{R}^m$ and $\mathbb{C}^m$ are used to represent the sets of m dimensional vectors of real and complex numbers, respectively, whereas $\mathbb{R}^{m\times m}$, $\mathbb{C}^{m\times m}$, and $\mathbb{H}^{m\times m}$ are used to represent the sets of $m\times m$ matrices of real numbers, $m \times m$ matrices of complex numbers, and $m \times m$ Hermitian matrices, respectively. Superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^{-1}$ indicate the transpose, complex conjugate, complex conjugate transpose, and inverse, respectively. For any $x \in \mathbb{R}$,

[2] A preliminary version of the methodology introduced here was presented in [30].

$\lceil x \rceil$ returns the least integer greater than or equal to x. The trace and the determinant of a matrix $\boldsymbol{X}$ are denoted by $\mathrm{Tr}(\boldsymbol{X})$ and $|\boldsymbol{X}|$, respectively. The notation $[\boldsymbol{X}]_i$ is used to represent the i^{th} column of the matrix $\boldsymbol{X}$. The symbol $\otimes$ indicates the Kronecker product while the gradient of a function f is denoted by ∇f. The symbol $\succeq$ (and its strict form $\succ$) is used to denote the generalized matrix inequality: for any $\boldsymbol{X} \in \mathbb{H}^{m\times m}$, $\boldsymbol{X} \succeq 0$ means that $\boldsymbol{X}$ is a PSD matrix ($\boldsymbol{X} \succ 0$ for positive definiteness). Besides, for any $\boldsymbol{X} \in \mathbb{H}^{m\times m}$, $\mathrm{eig}(\boldsymbol{X})$ is the vector collecting the eigenvalues of $\boldsymbol{X}$ (sorted in increasing order). The Euclidean norm of the vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|_2$, $|\boldsymbol{x}|$ indicates the element wise modulus of the vector $\boldsymbol{x}$. The notation $\mathbf{E}[\cdot]$ stands for statistical expectation. Finally, for any $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times m}$, $\max(\boldsymbol{X}, \boldsymbol{Y})$ refers to the matrix containing the element wise maximum between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

II. Problem Formulation

Let us assume the availability of n independent and identically distributed vectors $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n\}$, where[3] each $\mathbf{y}_i$ is of size m and follows a m-variate complex, zero-mean, circular Gaussian distribution with covariance matrix $\mathbf{R} \succ 0$. The maximum likelihood covariance estimation problem can be formulated as

$$\underset{\mathbf{R}\succ 0}{\text{minimize}}\ \bar{f}(\mathbf{R}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i^H \mathbf{R}^{-1}\mathbf{y}_i + \log|\mathbf{R}|. \tag{4}$$

If $n \geq m$, Problem (4) has a unique minimizer with probability one which is given by the SCM, i.e., $\mathbf{R}_{\text{SCM}} = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^H$. However, if the random process, where each observation is drawn, is stationary (at least in wide sense) then the covariance matrix also exhibits a Toeplitz structure which can be capitalized in the estimation process [2, Ch. 1], [34, Ch. 2]. By doing so, Problem (4) becomes

$$\text{MLE:}\quad \underset{\mathbf{R}\in Toep, \mathbf{R}\succ 0}{\text{minimize}}\ \bar{f}(\mathbf{R}), \tag{5}$$

where $Toep$ is used to denote the set of Hermitian Toeplitz matrices of size $m \times m$. The above problem has two constraints: a structural constraint and a positive definite constraint. Even though the structural constraint is convex, the non-convexity of the objective function makes Problem (5) challenging to solve and no analytical solution seems to be available. In the following two iterative solution procedures for (5) are designed exploiting the MM principle. Briefly, the MM technique mainly consists of two steps

1) constructing a s.f. $g(\mathbf{R}|\mathbf{R}_t)$ (where $\mathbf{R}_t$ is the estimate of $\mathbf{R}$ at the t^{th} iteration) for the objective function in (5), satisfying $g(\mathbf{R}_t|\mathbf{R}_t) = \bar{f}(\mathbf{R}_t), \forall\, \mathbf{R}_t \succ 0$ and $g(\mathbf{R}|\mathbf{R}_t) \geq \bar{f}(\mathbf{R}), \forall\, \mathbf{R}_t \succ 0$;
2) minimizing the resulting surrogate problem at each iteration.

For more details, [35], [36], [37] provide an in-depth discussion on MM based algorithms.

[3]Note that, from a practical point of view, a data selection scheme [31], [32], [33] can be employed for screening the available training data so as to excise possible outliers.

III. Algorithms for Toeplitz Covariance Matrix Estimation

In this section, ATOM1 and ATOM2 are proposed to tackle the MLE problem of TSC matrix. Both exploit the MM principle (applied to an equivalent reformulation of the MLE problem) and differ in the way they construct and handle the surrogate minimization problem. ATOM1 solves the surrogate optimization using ADMM while ATOM2 tackles it using either alternating projection or Dykstra's algorithm. Subsequently, the computational complexity and proof of convergence of the procedures are established. Finally, the extension of ATOM2 to deal with additional covariance constraints along with the Toeplitz structure is provided.

Before proceeding further, let us observe that the Hermitian Toeplitz matrices intrinsically endow the centro-Hermitian symmetry structure [38], i.e.,

$$\mathbf{R} = \boldsymbol{J}\mathbf{R}^*\boldsymbol{J} \tag{6}$$

with $\boldsymbol{J}$ the $m \times m$ permutation matrix given by

$$\boldsymbol{J} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{7}$$

As a consequence, Problem (5) is tantamount to (see also [2, Sec. 6.5.8])

$$\underset{\mathbf{R}\in Toep, \mathbf{R}\succ 0}{\text{minimize}}\ f(\mathbf{R}), \tag{8}$$

where

$$f(\mathbf{R}) = \mathrm{Tr}(\mathbf{R}_{FB}\mathbf{R}^{-1}) + \log|\mathbf{R}| \tag{9}$$

refers to the restriction of $\bar{f}(\cdot)$ to the centro-Hermitian covariance matrices, with $\mathbf{R}_{FB}$ the forward-backward (FB) averaged sample covariance matrix[4] given by $\mathbf{R}_{FB} = 1/2(\mathbf{R}_{\text{SCM}} + \boldsymbol{J}\mathbf{R}_{\text{SCM}}^*\boldsymbol{J})$ [39].

Now, decomposing $\mathbf{R}_{FB} = \boldsymbol{Y}\boldsymbol{Y}^H$, e.g., via LDL factorization [40], with $\boldsymbol{Y} \in \mathbb{C}^{m\times r}$, where $r = rank(\mathbf{R}_{FB}) \leq m$, Problem (8) can be equivalently cast as[5] (see Appendix A)

$$\begin{aligned} \min_{\mathbf{R}\in Toep, \boldsymbol{X}\in\mathbb{H}^{r\times r}}\ & \mathrm{Tr}(\boldsymbol{X}) + \log|\mathbf{R}| \\ \text{s.t.}\ & \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}^H \\ \boldsymbol{Y} & \mathbf{R} \end{pmatrix} \succeq \mathbf{0}, \end{aligned} \tag{10}$$

where the objective is a concave differentiable function of $\boldsymbol{X}$ and $\mathbf{R}$.

Before proceeding with the next important lemma, it is worth pointing out that Problem (10) holds true even if the Toeplitz structural constraint in Problem (5) and (10) is replaced by any set of positive definite (centro-Hermitian) matrices, provided that the estimation problem is solvable, i.e., the optimal solution exists.

[4]Hereafter, Problem (5) (and thus (8)) is assumed solvable, i.e., there exists a global optimizer $\mathbf{R}^* \succ 0$, as well as any limit point of a feasible sequence of matrices whose corresponding objectives converge to the optimal value is feasible to the optimization problem. As a consequence, without loss of generality, the constraint $\mathbf{R} \succ 0$ can be relaxed into $\mathbf{R} \succeq 0$. Notably, a sufficient condition to ensure the aforementioned properties is provided by $n \geq \lceil m/2 \rceil$, corresponding to $\mathbf{R}_{FB} \succ 0$ with probability one.

[5]A similar constraint reformulation is used in some studies involving atomic norm for sparse reconstruction [25], [41].

Lemma 1: Given a concave differentiable[6] function $h(\boldsymbol{K}) : \mathbb{H}^{r\times r} \rightarrow \mathbb{R}$, it can be majorized as

$$h(\boldsymbol{K}) \leq h(\boldsymbol{K}_t) + \mathrm{Tr}\left(\nabla h(\boldsymbol{K}_t)^H(\boldsymbol{K}-\boldsymbol{K}_t)\right), \tag{11}$$

where $\boldsymbol{K}_t \in \mathbb{H}^{r\times r}$. The upper bound to $h(\boldsymbol{K})$ is linear and differentiable with respect to (w.r.t.) $\boldsymbol{K}$.

Proof: Since $h(\boldsymbol{K})$ is a concave function w.r.t. $\boldsymbol{K}$, (11) stems from linearizing $h(\boldsymbol{K})$ via its first order Taylor expansion [42]. ∎

In order to tackle the challenging optimization problem (10), MM-based methods [43], [44], denoted ATOM1 and ATOM2, are now developed. To this end, let us observe that the term $\log|\mathbf{R}|$ in (10) is a concave function w.r.t. $\mathbf{R}$ [45]. Hence, it can be majorized using Lemma 1 (with $\boldsymbol{K}=\mathbf{R}$, $\boldsymbol{K}_t=\mathbf{R}_t$, $h(\boldsymbol{K})=\log|\mathbf{R}|$, and $\nabla h(\boldsymbol{K}_t)=\mathbf{R}_t^{-1}$) to get the following s.f.

$$g(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t) = \mathrm{Tr}(\boldsymbol{X}) + \mathrm{Tr}\big(\mathbf{R}_t^{-1}\mathbf{R}\big) + c_1 = \mathrm{Tr}(\mathbf{A}_t\boldsymbol{E}) + c_1, \tag{12}$$

where the constant $c_1 = \log|\mathbf{R}_t| - m$, $\mathbf{A}_t = \mathrm{diag}(\mathbf{I}, \mathbf{R}_t^{-1})$, whereas $\boldsymbol{E} = \mathrm{diag}(\boldsymbol{X},\mathbf{R})$ is the block-diagonal matrix with blocks $\boldsymbol{X}$ and $\mathbf{R}$ along the main diagonal. Given $\mathbf{R}_t$, which in our case is the value assumed by the variable $\mathbf{R}_t$ at the t-th iteration of the algorithm, the MM method demands for the solution of the following surrogate minimization task

$$\begin{aligned}\{\mathbf{R}_{t+1},\boldsymbol{X}_{t+1}\} \;=\; &\underset{\mathbf{R}\in Toep,\boldsymbol{X}\in\mathbb{H}^{r\times r}}{\arg\min}\; g(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t)\\ &\text{s.t.}\begin{pmatrix}\boldsymbol{X} & \boldsymbol{Y}^H\\ \boldsymbol{Y} & \mathbf{R}\end{pmatrix}\succeq\mathbf{0},\end{aligned} \tag{13}$$

which is a Semidefinite Programming (SDP) problem. Unfortunately, the computational complexity necessary to handle SDP using interior point methods is $O\left((r+m)^{4.5}\log(\frac{1}{\tilde{\eta}})\right)$ [46], [47], with $\tilde{\eta} > 0$ the desired solution accuracy. In order to alleviate the computational issue, two different approaches are pursued. The former directly handles Problem (13) via the iterative ADMM algorithm. The latter, by means of a suitable manipulation of (12), constructs a different s.f. for the objective function in Problem (10). By doing so, as clearly explained in the following, a computationally efficient and flexible estimation procedure capable of including additional constraints can be developed. To this end, let us observe that, adding and subtracting $\gamma\mathrm{Tr}(\boldsymbol{E}^2)$, (12) is equivalent to[7]

$$\mathrm{Tr}(\mathbf{A}_t\boldsymbol{E}) + \gamma\mathrm{Tr}(\boldsymbol{E}^2) - \gamma\mathrm{Tr}(\boldsymbol{E}^2) \tag{14}$$

with $\gamma > 0 \in \mathbb{R}$ a parameter of the surrogate construction stage (for $\gamma \downarrow 0$, the function in (14) reduces to (12)). Now, being $-\mathrm{Tr}(\boldsymbol{E}^2)$ a concave function of $\boldsymbol{E}$ and invoking Lemma 1 applied to the feasible solution $\boldsymbol{E}_t = \mathrm{diag}(\boldsymbol{X}_t;\mathbf{R}_t)$ with $\boldsymbol{X}_t = \boldsymbol{Y}^H\mathbf{R}_t^{-1}\boldsymbol{Y}$ and $\mathbf{R}_t$ provided by the t-th iteration step of the estimation process, it is possible to construct the s.f. for (14)

$$\tilde{g}(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t) = \mathrm{Tr}\left(\mathbf{A}_t\boldsymbol{E}\right) + \gamma\mathrm{Tr}(\boldsymbol{E}^2) - 2\gamma\mathrm{Tr}(\boldsymbol{E}\boldsymbol{E}_t) - \gamma\mathrm{Tr}(\boldsymbol{E}_t^2). \tag{15}$$

[6]For a non-differentiable function, the inequality in (11) can be cast as $h(\boldsymbol{K}) \leq h(\boldsymbol{K}_t) + \mathrm{Tr}\left(\boldsymbol{G}(\boldsymbol{K}_t)^H(\boldsymbol{K}-\boldsymbol{K}_t)\right)$, where $\boldsymbol{G}(\boldsymbol{K}_t)$ is the subgradient of the concave function $h(\boldsymbol{K})$ at $\boldsymbol{K}_t$ [35].

[7]Note that as γ approaches zero, the objective function (14) shrinks towards the original one in (12). Thereby, the smaller γ, the closer the ATOM2 surrogate objective function to the ATOM1 counterpart.

It is worth pointing out that $\tilde{g}(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t)$ represents a surrogate to a s.f.. Nonetheless, since $\tilde{g}(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t)$ is a tight approximation of $g(\boldsymbol{X},\mathbf{R}|\mathbf{R}_t)$, it is straightforward to show that (15) provides a direct surrogate for the objective function in Problem (12). Hence, given $\mathbf{R}_t$ and after some algebraic manipulations, the resulting surrogate minimization problem at the t-th iteration can be cast as

$$\begin{aligned}\{\mathbf{R}_{t+1},\boldsymbol{X}_{t+1}\} = &\underset{\mathbf{R}\in Toep,\boldsymbol{X}}{\arg\min}\; \|\boldsymbol{E}-\boldsymbol{B}_t\|_F^2\\ &\text{subject to}\quad \boldsymbol{E}+\boldsymbol{D}\succeq\mathbf{0},\end{aligned} \tag{16}$$

where $\boldsymbol{B}_t = \boldsymbol{E}_t - \gamma'\mathbf{A}_t$, with $\gamma' = \frac{0.5}{\gamma}$ and $\boldsymbol{D} = [\mathbf{0}, \boldsymbol{Y}^H; \boldsymbol{Y}, \mathbf{0}]$.

In the following subsections III-A and III-B two iterative methods, i.e., ATOM1 and ATOM2, are proposed to solve the surrogate minimization problems in (13) and (16), respectively.

A. ATOM1

The surrogate minimization problem in (13) is solved using ADMM [48], [49]. To this end, an auxiliary variable $\boldsymbol{U}\in\mathbb{H}^{(r+m)\times(r+m)}$ is introduced in (13) and the problem is framed in the equivalent form

$$\begin{aligned}\min_{\mathbf{R}\in Toep,\boldsymbol{U}\succeq\mathbf{0},\boldsymbol{X}\in\mathbb{H}^{r\times r}}\; &\mathrm{Tr}(\boldsymbol{X}) + \mathrm{Tr}\big((\mathbf{R}_t)^{-1}\mathbf{R}\big)\\ \text{s.t.}\quad &\begin{pmatrix}\boldsymbol{X} & \boldsymbol{Y}^H\\ \boldsymbol{Y} & \mathbf{R}\end{pmatrix} - \boldsymbol{U} = \mathbf{0}.\end{aligned} \tag{17}$$

The augmented Lagrangian [50, Ch. 2] associated with (17) is

$$\begin{aligned}&\mathcal{L}_\rho(\mathbf{R},\boldsymbol{X},\boldsymbol{U},\hat{\boldsymbol{\lambda}}) = \mathrm{Tr}(\boldsymbol{X}) + \mathrm{Tr}\big((\mathbf{R}_t)^{-1}\mathbf{R}\big) +\\ &\mathrm{Tr}\left[\hat{\boldsymbol{\lambda}}^H\left(\begin{pmatrix}\boldsymbol{X} & \boldsymbol{Y}^H\\ \boldsymbol{Y} & \mathbf{R}\end{pmatrix} - \boldsymbol{U}\right)\right] + \frac{\rho}{2}\left\|\begin{pmatrix}\boldsymbol{X} & \boldsymbol{Y}^H\\ \boldsymbol{Y} & \mathbf{R}\end{pmatrix} - \boldsymbol{U}\right\|_F^2,\end{aligned} \tag{18}$$

where $\rho > 0$ is the penalty parameter and $\hat{\boldsymbol{\lambda}}$ is the Lagrange multiplier of size $(r+m)\times(r+m)$. Problem (18) can be further rewritten as

$$\begin{aligned}&\mathcal{L}_\rho(\boldsymbol{E},\boldsymbol{U},\hat{\boldsymbol{\lambda}})\\ &= \mathrm{Tr}(\mathbf{A}_t\boldsymbol{E}) + \mathrm{Tr}\big(\hat{\boldsymbol{\lambda}}^H(\boldsymbol{E}+\boldsymbol{D}-\boldsymbol{U})\big) + \frac{\rho}{2}\|\boldsymbol{E}+\boldsymbol{D}-\boldsymbol{U}\|_F^2.\end{aligned} \tag{19}$$

The (inner) iterative steps of ADMM algorithm [48], [49] are

$$\begin{aligned}\boldsymbol{U}_{k+1}^t = &\underset{\boldsymbol{U}\succeq\mathbf{0}}{\arg\min}\quad \mathrm{Tr}\big((\hat{\boldsymbol{\lambda}}_k^t)^H(\boldsymbol{E}_k^t+\boldsymbol{D}-\boldsymbol{U})\big)\\ &+\frac{\rho}{2}\|\boldsymbol{E}_k^t+\boldsymbol{D}-\boldsymbol{U}\|_F^2\end{aligned} \tag{20}$$

$$\begin{aligned}\boldsymbol{E}_{k+1}^t = &\underset{\mathbf{R}\in Toep,\boldsymbol{X}}{\arg\min}\; \mathrm{Tr}(\mathbf{A}_t\boldsymbol{E}) + \mathrm{Tr}\big((\hat{\boldsymbol{\lambda}}_k^t)^H(\boldsymbol{E}+\boldsymbol{D}-\boldsymbol{U}_{k+1}^t)\big)\\ &+\frac{\rho}{2}\|\boldsymbol{E}+\boldsymbol{D}-\boldsymbol{U}_{k+1}^t\|_F^2\end{aligned} \tag{21}$$

$$\hat{\boldsymbol{\lambda}}_{k+1}^t = \hat{\boldsymbol{\lambda}}_k^t + \rho\big(\boldsymbol{E}_{k+1}^t + \boldsymbol{D} - \boldsymbol{U}_{k+1}^t\big), \tag{22}$$

where $(\cdot)_k^t$ is used to denote the k-th inner-iteration of the ADMM algorithm in correspondence of the t-th MM outer-loop. Problems (20) and (21) have closed-form solutions which

can be computed via the projection of appropriate matrices onto the respective feasible sets. Indeed, Problem (20) can be equivalently cast as

$$\boldsymbol{U}_{k+1}^{t} = \underset{\boldsymbol{U}\succeq 0}{\arg\min} \|\boldsymbol{U} - \boldsymbol{\Psi}_{k}^{t}\|_{F}^{2} \tag{23}$$

where $\boldsymbol{\Psi}_{k}^{t} = \boldsymbol{E}_{k}^{t} + \boldsymbol{D} + \frac{1}{\rho}\hat{\boldsymbol{\lambda}}_{k}^{t}$. Hence, solving (20) is tantamount to performing the orthogonal projection of the matrix $\boldsymbol{\Psi}_{k}^{t}$ onto the set of the PSD matrices which can be computed as $\boldsymbol{U}_{k+1}^{t} = \tilde{\boldsymbol{V}}_{k}^{t} \max(\mathrm{diag}(\tilde{\boldsymbol{U}}_{k}^{t}), \boldsymbol{0}) \tilde{\boldsymbol{V}}_{k}^{tH}$, where $\mathrm{diag}(\tilde{\boldsymbol{U}}_{k}^{t})$ and $\tilde{\boldsymbol{V}}_{k}^{t}$ are the matrices containing the eigenvalues and the corresponding orthonormal eigenvectors of $\boldsymbol{\Psi}_{k}^{t}$, respectively. Similarly, the update step of $\boldsymbol{E}$ in (21) can be rewritten as

$$\boldsymbol{E}_{k+1}^{t} = \underset{\mathbf{R}\in Toep, \boldsymbol{X}}{\arg\min} \quad \|\boldsymbol{E} - \boldsymbol{\Lambda}_{k}^{t}\|_{F}^{2}, \tag{24}$$

where $\boldsymbol{\Lambda}_{k}^{t} = \mathcal{P}_{D-Toep}\left(\boldsymbol{U}_{k+1}^{t} - \boldsymbol{D} - \frac{1}{\rho}(\hat{\boldsymbol{\lambda}}_{k}^{t} + \mathbf{A}_{t})\right)$, with $\mathcal{P}_{D-Toep}(\boldsymbol{\Psi})$ computed as follows: Partitioning the matrix $\boldsymbol{\Psi}$ as $\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^{H} & \boldsymbol{\Psi}_{22} \end{pmatrix}$ with $\boldsymbol{\Psi}_{12}$ of size $r \times m$, the orthogonal projection of interest amounts to set the upper diagonal block to $\boldsymbol{\Psi}_{11}$ whereas the second diagonal block is obtained by averaging the elements along each diagonal of $\boldsymbol{\Psi}_{22}$ and constructing the corresponding Toeplitz matrix.

Now, partitioning $\boldsymbol{\Lambda}_{k}^{t}$ as $\boldsymbol{\Lambda}_{k}^{t} = \begin{pmatrix} \boldsymbol{\Lambda}_{11,k}^{t} & \boldsymbol{\Lambda}_{12,k}^{t} \\ \boldsymbol{\Lambda}_{12,k}^{tH} & \boldsymbol{\Lambda}_{22,k}^{t} \end{pmatrix}$ with $\boldsymbol{\Lambda}_{11,k}^{t}$ and $\boldsymbol{\Lambda}_{22,k}^{t}$ being $r \times r$ and $m \times m$ matrices, respectively, it follows that $\boldsymbol{X}_{k+1}^{t} = \boldsymbol{\Lambda}_{11,k}^{t}$ and $\mathbf{R}_{k+1}^{t} = \boldsymbol{\Lambda}_{22,k}^{t}$. Before concluding, it is worth pointing out that since the surrogate minimization problem in (13) is convex and only an equality constraint is forced, it is guaranteed that ADMM converges to a supposed existing[8] optimal unique solution to (13) (see Section 3.2 in [50] and [51]). The pseudocode of the proposed algorithm is shown in **Algorithm 1**.

Algorithm 1 Pseudocode of ATOM1 algorithm

Input: Data-based matrix $\boldsymbol{Y}$ and ρ
Initialize: Set $t, k = 0$. Initialize $\mathbf{R}_0$, $\boldsymbol{X}_0$ and $\hat{\boldsymbol{\lambda}}_0$.
Repeat:
  $k \leftarrow 0$
  Compute $\mathbf{A}_t = \mathrm{diag}(\mathbf{I}, \mathbf{R}_t^{-1})$, $\boldsymbol{E}_k^t = \mathrm{diag}(\boldsymbol{X}_t, \mathbf{R}_t)$, $\hat{\boldsymbol{\lambda}}_k^t = \hat{\boldsymbol{\lambda}}_t$
  Repeat:
    1) Obtain $\boldsymbol{U}_{k+1}^t$ by projecting the matrix $\boldsymbol{\Psi}_k^t = \boldsymbol{E}_k^t + \boldsymbol{D} + \frac{1}{\rho}\hat{\boldsymbol{\lambda}}_k^t$ onto the set of PSD matrices.
    2) Compute $\boldsymbol{\Lambda} = \boldsymbol{U}_{k+1}^t - \boldsymbol{D} - \frac{1}{\rho}(\hat{\boldsymbol{\lambda}}_k^t + \mathbf{A}_t)$
    3) Set $\boldsymbol{X}_{k+1}^t$ equal to the first block $\boldsymbol{\Lambda}_{11}$ of $\boldsymbol{\Lambda}$
    4) Obtain $\mathbf{R}_{k+1}^t$ by projecting the second block $\boldsymbol{\Lambda}_{22}$ of $\boldsymbol{\Lambda}$ onto the set of Toeplitz matrices.
    5) Obtain $\boldsymbol{E}_{k+1}^t = \mathrm{diag}(\boldsymbol{X}_{k+1}^t, \mathbf{R}_{k+1}^t)$
    6) $\hat{\boldsymbol{\lambda}}_{k+1}^t = \hat{\boldsymbol{\lambda}}_k^t + \rho(\boldsymbol{E}_{k+1}^t + \boldsymbol{D} - \boldsymbol{U}_{k+1}^t)$
    7) $k \leftarrow k + 1$
  until convergence
  Set $\mathbf{R}_{t+1} = \mathbf{R}_k^t$, $\boldsymbol{X}_{t+1} = \boldsymbol{X}_k^t$, $\hat{\boldsymbol{\lambda}}_{t+1} = \hat{\boldsymbol{\lambda}}_k^t$
  $t \leftarrow t + 1$
until convergence
Output: $\mathbf{R}_{\mathrm{ATOM1}} = \mathbf{R}_t$.

From **Algorithm 1** it can be seen that ATOM1 requires initialization of the matrices $\mathbf{R}_0$, $\boldsymbol{X}_0^t$ and $\hat{\boldsymbol{\lambda}}_0^t$. $\mathbf{R}_0$ can be set using the initialization scheme discussed in [22] and, as $t = 0$, $\boldsymbol{X}_0^t$ can be set equal to $\boldsymbol{Y}^H \mathbf{R}_0^{-1} \boldsymbol{Y}$ while $\hat{\boldsymbol{\lambda}}_0^t$ can be constructed as $\hat{\boldsymbol{\lambda}}_0^t = \boldsymbol{V}\boldsymbol{V}^H$, where the elements of $\boldsymbol{V}$ are drawn randomly from a uniform distribution over $[0, 1]$. For $t \geq 1$, the matrices $\boldsymbol{E}_0^t$ and $\hat{\boldsymbol{\lambda}}_0^t$ can be initialized with their last value after convergence at the previous ADMM iteration, respectively. Another input parameter required by ATOM1 is the penalty weight ρ, introduced during the construction of the Augmented Lagrangian of the ADMM framework. It is shown in [50], that the ADMM algorithm converges for any value of $\rho > 0$. However, the numerical stability and the convergence rate depends on the choice of ρ. Simulation results have highlighted that for $\rho = 1$, the ADMM algorithm is stable for different values of n and m. Hence, unless otherwise stated, in all the numerical analysis $\rho = 1$ is used. Notably, in the open literature, [25], [26] addressed the Toeplitz estimation problem by devising optimization procedures similar to ATOM1, where in [26] a rank constraint is also considered in the estimation process. However, it is worth mentioning that ATOM1 optimizes a different surrogate function exploiting the persymmetric structure of the covariance matrix, so it represents a different implementation of the MM plus ADMM method to deal with the optimization problem at hand.

1) Computational Complexity and Discussion About ATOM1: ATOM1 is iterative in nature with two loops - the outer-loop updates the Toeplitz matrix $\mathbf{R}_t$ while the inner-loop solves the surrogate minimization problem using ADMM. Note that in the inner-loop, it is required to construct the data-based matrix $\boldsymbol{D} = \begin{pmatrix} \mathbf{0} & \boldsymbol{Y}^H \\ \boldsymbol{Y} & \mathbf{0} \end{pmatrix}$ - which is iteration independent and hence can be pre-computed and stored. Let us now discuss the complexity related to the outer and inner-loops of ATOM1. The inner-loop of ATOM1 requires the computation of the matrix $\mathbf{A}_t$ - which is outer-loop iteration dependent. Therefore, this matrix can be evaluated once in each outer-loop. Consequently, apart from the computations involved in the inner-loop, an outer-loop cycle just involves the evaluation of the matrix $\mathbf{R}_t^{-1}$. Since $\mathbf{R}_t$ is Toeplitz, its inverse can be efficiently computed with a complexity $\mathcal{O}(m\,\log m)$ [52]. The computational complexity of an inner-loop cycle is related to the projection of $\boldsymbol{\Psi}_k^t$ onto the set of PSD matrices and projection of $\boldsymbol{\Lambda}_k^t$ onto the set of block diagonal matrices where the upper part (of size $r \times r$) is unconstrained, whereas the lower block (of size $m \times m$) is Toeplitz structured. The cost of this latter operation mainly involves the projection of $\boldsymbol{\Lambda}_{22,k}^t$ onto the set of Toeplitz matrices; thus, it is substantially dictated by the computation of average of the elements along the diagonals of $\boldsymbol{\Lambda}_{22,k}^t$. Hence, the cost of the inner-step *4)* is $\mathcal{O}(m^2)$. Next, the projection of $\boldsymbol{\Psi}$ onto the set of PSD matrices mainly involves the computation of the eigenvalues and eigenvectors of the matrix $\boldsymbol{\Psi}_k^t$ - whose corresponding complexity is $\mathcal{O}((r+m)^3)$ [40]. Therefore, the per-outer-iteration computational complexity of ATOM1 is $\mathcal{O}(\eta(r+m)^3)$ where η is the total number of inner-loop iterations required by the algorithm to converge.

A drawback of ATOM1 is the lack of a theoretical quality guarantee when it has to handle additional constraints on the covariance matrix. This is because ATOM1 implements ADMM algorithm at each inner-iteration which requires (to

[8] A sufficient condition for the existence of the optimal solution to Problem (13) is provided by the solvability of (8).

endow convergence guarantees to the process) the optimization problem to exhibit the standard form [50], [53]

$$\begin{aligned} &\underset{\boldsymbol{Z},\boldsymbol{E}}{\text{minimize}}\; h_1(\boldsymbol{Z}_1)+h_2(\boldsymbol{Z}_2) \\ &\text{subject to}\;\; \mathbf{A}_1\boldsymbol{Z}_1+\mathbf{A}_2\boldsymbol{Z}_2=\boldsymbol{C} \end{aligned} \tag{25}$$

where $h_1(\boldsymbol{Z}_1)$, $h_2(\boldsymbol{Z}_2)$ are convex functions and $\mathbf{A}_1$, $\mathbf{A}_2$, $\boldsymbol{C}$ are matrices of appropriate dimensions, respectively. Therefore, to incorporate additional inequality constraints (such as those resulting from upper bound on the condition number of the matrix $\boldsymbol{Z}_1$ or a lower bound to the strength of diagonal elements, or more in general an intersection of closed convex sets that can be described by additional auxiliary variables), one needs to replace each inequality constraint with an appropriate equality constraint. This can be done by introducing a slack variable for each inequality constraint to the existing optimization variables $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$. However, there is no convergence guarantee of ADMM when there are more than two optimization variables [54]. This issue can be addressed by the low complexity algorithm, referred to as ATOM2, proposed to solve Problem (16).

B. ATOM2

Problem (16) is tantamount to seeking the block diagonal matrix $\boldsymbol{E}$ belonging to the intersection of the two sets - the former defined by block diagonal matrices with the lower diagonal block of size $m\times m$ fulfilling a Toeplitz structure and the latter given by the Linear Matrix Inequality (LMI) [55] $\boldsymbol{E}+\boldsymbol{D}\succeq 0$ - with minimum distance from $\boldsymbol{B}$. Being the feasible set of (16) characterized by the intersection of convex sets, a viable, even though heuristic, means to tackle Problem (16) is provided by the alternating projection or Projection Onto the Convex Sets (POCS) technique [56], [57], [58], which has already been successfully applied in the signal processing context, e.g., [59], [60].

Let us denote by $\mathcal{P}_{LMI}(\boldsymbol{\Psi})$ the orthogonal projection of an arbitrary matrix $\boldsymbol{\Psi}$ onto the set defined by $\boldsymbol{E}+\boldsymbol{D}\succeq\mathbf{0}$. Now, to proceed further and employ the POCS framework, $\mathcal{P}_{D-Toep}(\boldsymbol{\Psi})$ and $\mathcal{P}_{LMI}(\boldsymbol{\Psi})$ projections must be employed. Remarkably, both can be obtained in closed-form: the former is computed as described in subsection III-A; as to the latter, the orthogonal projection onto the set defined by LMI $\boldsymbol{E}+\boldsymbol{D}\succeq\mathbf{0}$ is computed by first evaluating the EigenValue Decomposition (EVD) of the matrix $\boldsymbol{\Psi}+\boldsymbol{D}$, i.e., obtaining $[\bar{\boldsymbol{U}},\bar{\boldsymbol{V}}]=\text{eig}(\boldsymbol{\Psi}+\boldsymbol{D})$, where $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are matrices containing the eigenvalues and eigenvectors of the spectral decomposition, respectively. Then, the orthogonal projection $\mathcal{P}_{LMI}(\boldsymbol{\Psi})$ is given by $\bar{\boldsymbol{V}}\max(\bar{\boldsymbol{U}},\mathbf{0})\bar{\boldsymbol{V}}^H-\boldsymbol{D}$.

According to POCS method, given an initial value $\mathbf{T}_0^t=\boldsymbol{B}_t$, at the k-th inner-iteration first compute $\boldsymbol{Y}_{k+1}^t=\mathcal{P}_{D-Toep}(\mathbf{T}_k^t)$ and then, using $\boldsymbol{Y}_{k+1}^t$, determine $\mathbf{T}_{k+1}^t=\mathcal{P}_{LMI}(\boldsymbol{Y}_{k+1}^t)$ which represents the starting point $\mathbf{T}_{k+1}^t$ of the next inner-iteration. Hence, the POCS-based solution approach finds a sequence of iterates $\{\mathbf{T}_k^t\}$ by alternatingly projecting between the two convex sets. Nevertheless, as reported in [61], POCS may suffer from slow convergence. Even more crucial, the convergence to the global optimal solution to (16) is, in general, not ensured [62], [63]. A possible solution to the aforementioned shortcoming is provided by Dykstra's projection algorithm [62], which is an iterative procedure aimed at minimizing the distance of a given point from the intersection of closed convex sets via appropriate projections on each single sets. Therefore this technique is extremely effective if the individual projections can be evaluated efficiently. Dykstra's method is thus a refinement of POCS capable of finding a point closest to $\boldsymbol{B}_t$ by adding correction matrices $\mathbf{P}_k$ and $\boldsymbol{Q}_k$ before each projection is performed, which in-turn ensures convergence of sequence $\{\mathbf{T}_{k+1}\}$ to the optimal solution $\mathbf{T}^*=\boldsymbol{E}^*$ [62]. In particular, let C_1 be one of the convex set involved in the optimization procedure, at each iteration, after performing the projection of a matrix $\boldsymbol{A}$ onto C_1 obtaining $\tilde{\boldsymbol{A}}=\mathcal{P}_{C_1}(\boldsymbol{A}+\boldsymbol{P}^{(A)})$, with $\boldsymbol{P}^{(A)}$ the corresponding correction matrix (initialized to $\mathbf{0}$), $\boldsymbol{P}^{(A)}$ is updated by computing the difference between the matrix $\boldsymbol{A}+\boldsymbol{P}^{(A)}$ and its projection as $\boldsymbol{P}^{(A)}=\boldsymbol{A}+\boldsymbol{P}^{(A)}-\tilde{\boldsymbol{A}}$.

The pseudocode of Dykstra's algorithm is shown in **Algorithm 2**. Once the optimal solution $\boldsymbol{E}^*$ is obtained via

Algorithm 2 Pseudocode of Dykstra's algorithm

Input: $\boldsymbol{B}_t$
Initialize: Set $\mathbf{T}_0^t=\boldsymbol{B}_t$, $\mathbf{P}_0^t=\mathbf{0}$ and $\boldsymbol{Q}_0^t=\mathbf{0}$, $k=0$
Repeat:
1) $\boldsymbol{Y}_k^t=\mathcal{P}_{D-Toep}(\mathbf{T}_k^t+\mathbf{P}_k^t)$
2) $\mathbf{P}_{k+1}^t=\mathbf{T}_k^t+\mathbf{P}_k^t-\boldsymbol{Y}_k^t$
3) $\mathbf{T}_{k+1}^t=\mathcal{P}_{LMI}(\boldsymbol{Y}_k^t+\boldsymbol{Q}_k^t)$
4) $\boldsymbol{Q}_{k+1}^t=\boldsymbol{Y}_k^t+\boldsymbol{Q}_k^t-\mathbf{T}_{k+1}^t$
5) $k\leftarrow k+1$
until convergence
Output: $\boldsymbol{E}^*=\mathbf{T}_k^t$.

Dykstra's projection, the matrix $\mathbf{R}_{t+1}$ can be constructed from its lower diagonal block of size $m\times m$. This process is repeated until the whole MM-procedure, i.e., including the outer-loop, converges. The complete ATOM2 is summarized in **Algorithm 3**. It requires the initialization of the matrix $\mathbf{R}$.

Algorithm 3 Pseudocode of ATOM2

Input: Data-based matrix $\boldsymbol{Y}$, surrogate parameter γ
Initialize: Set $t=0$. Initialize $\mathbf{R}_0$, $\boldsymbol{X}_0$.
Repeat:
1) Compute $\mathbf{A}_t=\text{diag}(\mathbf{I},\mathbf{R}_t^{-1})$, $\boldsymbol{E}_t=\text{diag}(\boldsymbol{X}_t,\mathbf{R}_t)$
2) Compute $\boldsymbol{E}^*$ from **Algorithm 2** execution with $\boldsymbol{B}_t=\boldsymbol{E}_t-\frac{0.5}{\gamma}\mathbf{A}_t$
3) Obtain $\mathbf{R}_{t+1}$ from the lower diagonal block of $\boldsymbol{E}^*$
4) Obtain $\boldsymbol{X}_{t+1}$ from the upper diagonal block of $\boldsymbol{E}^*$
5) $t\leftarrow t+1$
until convergence
Output: $\mathbf{R}_{\text{ATOM2}}=\mathbf{R}_t$

In this respect, a similar scheme as in ATOM1 is followed, i.e., at each outer-iteration, the initial guess required to determine $\mathbf{R}_{t+1}$ in the inner-loop is obtained starting from $\mathbf{R}_t$.

C. Computational Complexity of ATOM2

Like ATOM1, ATOM2 is an iterative algorithm with outer- and inner-loops. The outer-loop updates the Toeplitz matrix $\mathbf{R}_t$ and the inner-loop implements the Dykstra's algorithm - which requires the computation of the matrices $\boldsymbol{D}$ and $\mathbf{R}_t^{-1}$. The former is a iteration independent data matrix and therefore can be pre-constructed. The latter is outer-loop iteration dependent and therefore can be computed once in each outer-loop. Consequently, apart from the inner-loop computations, the

TABLE I
COMPARISON AMONG COMPUTATIONAL COMPLEXITY OF ATOM1 AND ATOM2 WITH OTHER STATE-OF-THE-ART ITERATIVE ALGORITHMS

Algorithm	Complexity
ATOM1	$\mathcal{O}(\eta(r+m)^3))$
ATOM2	$\mathcal{O}(\eta(r+m)^3))$
MELT [22]	$\mathcal{O}\left(\eta(m\log(m))\right)$
EM [19]	$\mathcal{O}\left(\eta(m\log(m))\right)$

outer-loop demands only the computation of $\mathbf{R}_t^{-1}$ - which can be computed efficiently with complexity $\mathcal{O}(m\,\log m)$. Meanwhile, the computational load of the inner-loop stems from the evaluation of EVD of the matrix $(\boldsymbol{Y}_k+\boldsymbol{Q}_k)$ plus a data matrix $\boldsymbol{D}$ - which has a complexity of about $\mathcal{O}((r+m)^3)$.

In Table I, the computational complexity of ATOM1 and ATOM2 is compared with that of the state-of-the-art iterative algorithms [19], [22]. Unlike the proposed algorithms, the state-of-the art methods are single loop iteration algorithms. Therefore, in the case of [22] and [19] η is used to represent the number of iterations required by the algorithm to converge. Inspection of Table I shows that ATOM1 and ATOM2 have the highest complexity when compared to MELT and EM. Nevertheless, it is worth anticipating that this complexity increase is complemented by a superior performance in terms of generality of the problem solved (ATOM1 and ATOM2 do not exploit the CE, ATOM2 permits to handle additional structural constraints with quality guarantee, as shown in subsection III-E), covariance matrix MSE, and achieved SINR.

D. Proof of Convergence

In this subsection, the proof of convergence of ATOM1 and ATOM2 is established. In this regard, it is worth pointing out that both the algorithms differ in the way they construct and optimize the s.f. for the Problem (5). Nonetheless, since ATOM1 and ATOM2 are based on the MM framework, the proof of convergence based on the following Theorem will hold for both algorithms.

Before stating the Theorem, let us first introduce the first-order optimality condition for minimizing a function over a convex constraint set. A point $\boldsymbol{X}$ is a stationary point of $f(\cdot)$ if $f'(\boldsymbol{X};\boldsymbol{D})\geq 0$ for all $\boldsymbol{D}$ such that $\boldsymbol{X}+\boldsymbol{D}\in\mathcal{C}$, where $\mathcal{C}$ is the convex constraint set and $f'(\boldsymbol{X};\boldsymbol{D})$ is the directional derivative of $f(\cdot)$ at point $\boldsymbol{X}$ in direction $\boldsymbol{D}$ and is defined as [36]

$$f'(\boldsymbol{X};\boldsymbol{D})=\lim_{\lambda\downarrow 0}\inf\frac{f(\boldsymbol{X}+\lambda\boldsymbol{D})-f(\boldsymbol{X})}{\lambda}\ . \tag{26}$$

Based on the following theorem, relying on the key results in [36] and assuming that the inner-loop achieves the global optimizer, both ATOM1 and ATOM2 are guaranteed to converge to a stationary point of Problem (5).

Theorem 1: Denoting by $\{\mathbf{R}_t\}$ the sequence of matrices generated by either ATOM1 or ATOM2, then the objective function of Problem (5) monotonically decreases along the iterations. Besides, any positive definite cluster point[9] to $\mathbf{R}_t$ is a stationary point to Problem (5).

Proof: See Appendix B for details. ∎

[9]Under the assumption $m\geq n/2$, all the cluster points are demanded to be positive definite.

E. Extensions of ATOM2

The augmentation of ATOM2 to handle additional constraints other than the Toeplitz structure in the covariance estimation process is now addressed. In particular, it is shown that ATOM2 can be generalized[10] to account for the following scenarios: Banded Toeplitz, block-Toeplitz, and Toeplitz-block-Toeplitz matrices. On the other side, as already mentioned in subsection III-A.1, ATOM1 cannot be directly extended to tackle the general constraints as for instance an upper bound requirement to the condition number [64], [65] or, in a "cognition-driven-processing" application, a similarity constraint [66], [67] to exploit some prior knowledge of the stationary process, whose statistical characteristics inference represent the task at hand.

1) MLE of Banded Toeplitz Covariance Matrix: The covariance matrix is constrained to exhibit a banded Toeplitz structure of bandwidth b (see [21], [68], [69] for relevant applications). For instance, assuming a bandwidth $b=2$ and dimension $m=5$ the covariance matrix enjoys the following structure

$$\mathbf{R}=\begin{bmatrix} r_1 & r_2 & r_3 & 0 & 0\\ r_2^* & r_1 & r_2 & r_3 & 0\\ r_3^* & r_2^* & r_1 & r_2 & r_3\\ 0 & r_3^* & r_2^* & r_1 & r_2\\ 0 & 0 & r_3^* & r_2^* & r_1 \end{bmatrix}.$$

Then, the MLE problem for banded Toeplitz covariance matrix can be formulated as

$$\underset{\mathbf{R}\in Band-Toep,\,\mathbf{R}\succ 0}{\text{minimize}}\ \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i^H\mathbf{R}^{-1}\mathbf{y}_i+\log|\mathbf{R}|\ , \tag{27}$$

where $Band-Toep$ is used to denote the set of banded Toeplitz matrices. Like in (10), the above problem can be cast in the following equivalent form

$$\begin{aligned}\underset{\mathbf{R}\in Band-Toep,\boldsymbol{X}}{\text{minimize}}\ & \mathrm{Tr}(\boldsymbol{X})+\log|\mathbf{R}|\\ \text{subject to}\ & \begin{pmatrix}\boldsymbol{X} & \boldsymbol{Y}^H\\ \boldsymbol{Y} & \mathbf{R}\end{pmatrix}\succeq\mathbf{0}.\end{aligned} \tag{28}$$

Hence, (28) is handled via MM framework solving the following surrogate minimization problem

$$\begin{aligned}\underset{\boldsymbol{E}}{\text{minimize}}\ & \|\boldsymbol{E}-\boldsymbol{B}\|_F^2\\ \text{subject to}\ & \boldsymbol{E}+\boldsymbol{D}\succeq\mathbf{0}\\ & \boldsymbol{E}=\mathrm{diag}(\boldsymbol{X},\mathbf{R})\ \text{with}\ \mathbf{R}\ \text{being a}\\ & \text{banded Toeplitz matrix}\end{aligned} \tag{29}$$

The above problem involves two convex sets: the set defined by the LMI $\boldsymbol{E}+\boldsymbol{D}\succeq\mathbf{0}$ and the set of block diagonal matrices where the second block has a banded Toeplitz structure with bandwidth b. Consequently, Dykstra's projection algorithm or POCS can be used to solve Problem (29). The projection of a matrix onto the LMI set can be calculated as discussed earlier in Subsection III-B. The projection of a matrix $\hat{\boldsymbol{\Psi}}=\begin{pmatrix}\hat{\boldsymbol{\Psi}}_{11} & \hat{\boldsymbol{\Psi}}_{12}\\ \hat{\boldsymbol{\Psi}}_{12}^H & \hat{\boldsymbol{\Psi}}_{22}\end{pmatrix}$ onto the set of block diagonal matrices with the second banded Toeplitz block can be obtained as follows. The first diagonal block is the same as $\hat{\boldsymbol{\Psi}}_{11}$ and the

[10]If it is not required that $\mathbf{R}$ satisfies the centro-Hermitian property, $\boldsymbol{Y}$ in (10) is obtained via the LDL factorization of $\mathbf{R}_{\text{SCM}}$ rather than $\mathbf{R}_{FB}$.

second diagonal block is constructed by averaging the entries of the main and the first b upper-diagonals of the matrix $\hat{\boldsymbol{\Psi}}_{22}$ and computing the corresponding Toeplitz matrix [68].

2) MLE of Block-Toeplitz or Toeplitz-Block-Toeplitz Covariance Matrix: In space-time adaptive processing radar applications, the covariance matrix exhibits a block-Toeplitz (BT) [70] or a Toeplitz-block-Toeplitz (TBT) structure. An example of a BT-structured covariance matrix with p blocks is shown below

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \dots & \mathbf{R}_{p-1} \\ \mathbf{R}_1^H & \mathbf{R}_0 & \dots & \mathbf{R}_{p-2} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{R}_{p-1}^H & \dots & \mathbf{R}_1^H & \mathbf{R}_0 \end{bmatrix}. \tag{30}$$

When each block exhibit a Toeplitz structure, then $\mathbf{R}$ is TBT [71], [72].

The MLE problem of a BT or a TBT covariance matrix is formulated as

$$\underset{\mathbf{R}\in BT(TBT),\mathbf{R}\succ 0}{\text{minimize}} \ \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i^H\mathbf{R}^{-1}\mathbf{y}_i + \log|\mathbf{R}|, \tag{31}$$

where the notation $BT(TBT)$ is used to indicate the set of $BT(TBT)$ matrices. A feasible solution to Problem (31) can be obtained by solving at any given step the following surrogate optimization problem

$$\begin{aligned} \underset{\boldsymbol{E}}{\text{minimize}} \quad & \|\boldsymbol{E}-\boldsymbol{B}\|_F^2 \\ \text{subject to} \quad & \boldsymbol{E}+\boldsymbol{D}\succeq \mathbf{0} \\ & \boldsymbol{E} \text{ is a block diagonal matrix with} \\ & \text{the second diagonal BT (TBT) block.} \end{aligned} \tag{32}$$

Problem (32) exhibits two constraints - 1) a LMI constraint and 2) a structural constraint - where the optimization variable $\boldsymbol{E}$ is confined to be a block diagonal matrix with the second block having a BT (TBT) structure. Since both the constraints are convex, Dykstra's projection or POCS can be applied to solve Problem (32). The projection of a matrix onto the LMI set can be calculated as discussed earlier in Section III-B. The projection of a given matrix $\bar{\boldsymbol{\Psi}}$ onto the set of matrices whose second diagonal block has the BT (TBT) constraint can be obtained as follows. For the first diagonal block, the submatrix $\bar{\boldsymbol{\Psi}}_{11}$ is directly used. Then, the second diagonal block is obtained following two (three) steps. First, p matrices are obtained by averaging the (upper-right) diagonal blocks of the matrix $\bar{\boldsymbol{\Psi}}_{22}$. Then, only for TBT, each of the p matrices are projected onto the Toeplitz set as described in subsection III-B. Finally, the resulting matrix is constructed according to (30).

IV. CRB Calculation

In this section, the CRB is derived[11] for the estimation of Toeplitz structured covariance matrix (the interested reader may refer to Appendix C with reference to the CRBs computation of Banded Toeplitz, BT, and TBT covariance model).

[11]Bespoke parametrization of the unknowns are exploited to compute CRBs. It is noteworthy to highlight that the general framework to handle CRB computation in the presence of parameters restrictions/relationships is provided by the constrained CRB (the interested reader may refer to [73], [74], [75], and [76]).

The CRB provides a lower bound on the variance of any unbiased estimator [77]. To proceed further, let $\boldsymbol{\theta}$ represent the real value vector parametrizing a given covariance matrix structure of interest. Then, the CRB is the inverse of the Fisher Information matrix (FIM) whose $(i,k)^{th}$ element is

$$[\mathbf{F}]_{i,k} = \mathbf{E}\left[\frac{\partial^2 \log \bar{f}(\mathbf{R})}{\partial\theta_i\partial\theta_k}\right], \tag{33}$$

where $\frac{\partial \log \bar{f}(\mathbf{R})}{\partial\theta_i}$ denotes the partial derivative of $\log \bar{f}(\mathbf{R})$ w.r.t. θ_i, with θ_i the i-th element of $\boldsymbol{\theta}$. Due to the Gaussian assumption, the $(i,k)^{th}$ element of the FIM can be computed using the *Slepian-Bangs formula* [2]

$$[\mathbf{F}]_{i,k} = n\text{Tr}\left(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_k}\right). \tag{34}$$

In the following subsection, the FIM is derived for the Toeplitz covariance structure.

A. Toeplitz Matrix

As the entries of the TSC matrix are completely characterized by its first row, i.e., $[r_1, r_2, \cdots r_m]^T$, the covariance matrix $\mathbf{R} \in \mathbb{H}^{m\times m}$ can be parameterized by $\boldsymbol{\theta} = [r_1, \Re(r_2), \cdots \Re(r_m), \Im(r_2), \dots, \Im(r_m)]^T \in \mathbb{R}^{2m-1}$ where $\Re(r_i)$ and $\Im(r_i)$ denotes the real and imaginary parts of r_i, respectively. Then, the covariance matrix $\mathbf{R}$ can be expressed in terms of $\boldsymbol{\theta}$ and basis matrices $\boldsymbol{B}_g^{\text{Toep}}$ (defined as in (36)), $g = 1, 2, \cdots, m$ [20]

$$\mathbf{R} = \sum_{g=1}^{m}\theta_g\Re(\boldsymbol{B}_g^{\text{Toep}}) + j\sum_{g=m+1}^{2m-1}\theta_g\Im(\boldsymbol{B}_{g-m+1}^{\text{Toep}}). \tag{35}$$

The $(i,k)^{th}$ element of the matrix $\boldsymbol{B}_g^{\text{Toep}}$ is given as

$$[\boldsymbol{B}_g^{\text{Toep}}]_{i,k} = \begin{cases} 1+j & i-k = g-1 = 0 \\ 1+j & k-i = g-1 \neq 0 \\ 1-j & i-k = g-1 \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{36}$$

Using (35), $\frac{\partial\mathbf{R}}{\partial\theta_i}$ can be obtained as

$$\frac{\partial\mathbf{R}}{\partial\theta_i} = \begin{cases} \Re(\boldsymbol{B}_i^{\text{Toep}}) & 1 \le i \le m \\ j\Im(\boldsymbol{B}_{i-m+1}^{\text{Toep}}) & m+1 \le i \le 2m-1 \end{cases}.$$

Substituting $\frac{\partial\mathbf{R}}{\partial\theta_i}$ in (34), yields the FIM for Toeplitz covariance matrix.

V. Numerical Simulations

In this section, the performance of the proposed covariance matrix estimators ATOM1 and ATOM2 is numerically analyzed in comparison with the following state-of-the-art algorithms: EM-based [19], [78], MELT [22], the SCM, and the FB estimators [39]. First, a convergence analysis of the derived methods is provided, also in comparison with the aforementioned counterparts. Then, the estimation capabilities are analyzed in three different scenarios, using the MSE as performance metric, defined as[12]

$$\text{MSE} = \mathbf{E}\left[\left\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right\|^2\right], \tag{37}$$

[12]In the following, (37) is computed via Monte Carlo technique.

where $\hat{\boldsymbol{\theta}}$ indicates the estimate of the unknown $\boldsymbol{\theta}$, obtained according to one of the aforementioned strategies. First of all, the covariance matrix is assumed to share the Toeplitz structure. Then, the banded Toeplitz, the BT, and the TBT constraints are considered. The CRB-based benchmark, computed as CRB $= \mathrm{Tr}(\mathbf{F}^{-1})$, is reported too, whereby, for each case study, the FIM is appropriately derived, see Section IV.

Furthermore, assuming a typical radar signal processing scenario, the performance is also evaluated in terms of average achievable SINR by an adaptive spatial filter.

It is also worth reporting that, in the aforementioned scenarios, ATOM1 and ATOM2 procedures are initialized using the FB estimate $\mathbf{R}_{FB}$, projected onto the set of Toeplitz matrices. Moreover, for the execution of ATOM2, the parameter γ is updated adaptively in each outer-loop iteration according to the following law[13]

$$\gamma = \gamma_0 (t \log t + k_1)^2. \tag{38}$$

To illustrate the role of γ in the optimization process performed by ATOM2, a notional representation of the objective function (conceptually depicted as a one-dimensional curve and corresponding to a specific portion of a restriction of the multivariate objective) and the s.f. of ATOM1 and ATOM2, is reported in Fig. 1. Remarkably, the value of γ affects the trade-off between performance and convergence speed of ATOM2. Indeed, while a smaller γ leads to a better performance (ATOM2 s.f. approaches the ATOM1 one as $\gamma \to 0$), it demands more inner-loop iterations to achieve convergence, due to the almost singular resulting metric. On the other hand, a larger γ reduces the overall computational cost, but introduces a growth in the approximation error. However, as the outer-loop iterations increase, the approximation error of the ATOM2 s.f. w.r.t. the objective function decreases as the updated point becomes closer and closer to a local minimum at which the sequence is "converging". That said, slowly increasing γ with the number of iterations allows to speed-up its computational burden without decreasing its performance.

A. Assessment of Iterative Algorithms Convergence for on-Grid and off-Grid Frequencies

In this simulation, the convergence of ATOM1 and ATOM2 (whose inner-loop was implemented via Dykstra's algorithm) is assessed in comparison with MELT and EM algorithms. To this end, each data snapshot $\mathbf{y}_k \in \mathbb{C}^m$ is modeled as

$$\mathbf{y}_k = \mathbf{R}^{\frac{1}{2}} \boldsymbol{n}_k, \; k = 1, 2, \cdots, n \tag{39}$$

where $\boldsymbol{n}_k \in \mathbb{C}^m$, $k = 1, \ldots, n$ are independent and identically distributed zero-mean circularly symmetric Gaussian random vectors with unit mean square value.

Two different experimental setups are considered, assuming $m = 6$ and $n = 20$. In the former, the true underlying Toeplitz covariance matrix $\mathbf{R}$ is constructed by choosing the

[13]As to the adaptive ATOM2 surrogate construction stage, it has been empirically shown that the updating rule (38), with $\gamma_0 = 10^{-4}$ and $k_1 = 5$, provides satisfactory performance in all the scenarios; therefore, unless otherwise stated, ATOM2 s.f. (and the subsequent processing) is constructed using (38) with the aforementioned values.

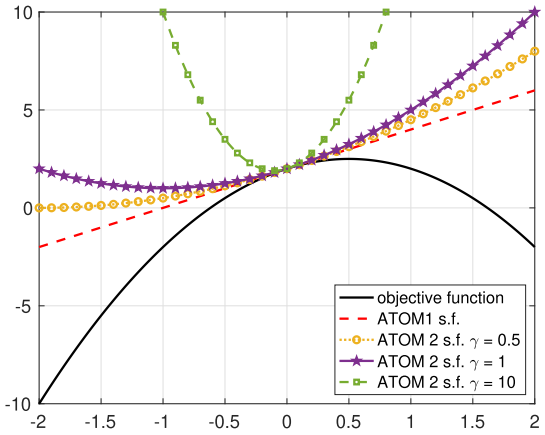
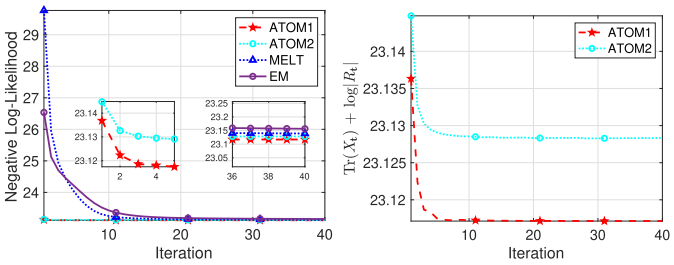


Fig. 1. A notional representation of the objective function of Problem (10) and the corresponding s.f. of ATOM1 and ATOM2, with the latter employing $\gamma \in \{0.5, 1, 10\}$, for a one-dimensional optimization problem.



(a) (b)

Fig. 2. Negative log-likelihood (9) and the objective function of (10) vs. outer-iterations for $m = 6$, $n = 20$, and on-grid frequencies scenario.

2-nd, 3-rd, 5-th, 7-th, 8-th and the 11-th column of the DFT matrix with $L = 2m - 1$ in (3), corresponding to the frequencies $[0.5712, 1.1424, 2.2848, 3.4272, 3.9984, 5.7120]$ rad, and as powers $[p_1, \ldots, p_6]^{\mathrm{T}} = [3, 6, 4, 1, 7, 5]^{\mathrm{T}}$, respectively. Figs. 2a and 2b show the negative log likelihood (9) and the objective function of problem (10) versus the number of iterations, respectively. It can be seen that all the algorithms numerically improve the negative log-likelihood as the number of iterations increases and almost converge to the same value, with negligible differences. Moreover, Fig. 2b indicates that the proposed algorithms monotonically decrease the problem objective function, which is expected since they optimize (10) using the MM framework.

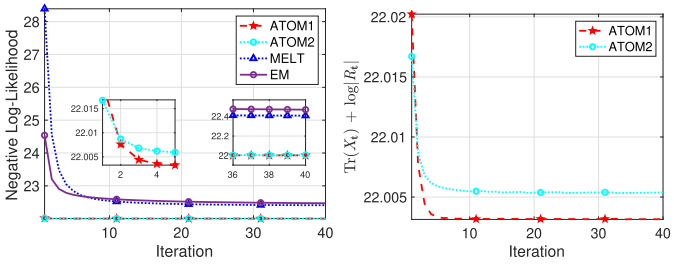
In the other experimental setup, the true underlying Toeplitz covariance matrix is constructed such that two of the frequencies are not on the Fourier grid. Therefore, the same parameters used in case study 1 are considered, with the exception that the Fourier frequencies 0.5712 rad and 3.9984 rad are replaced with 0.5 rad and 5.3 rad, respectively. For the case study at hand, the negative log-likelihood (9) and the objective function of (10) are reported in Figs. 3a and 3b versus the number of iterations, respectively. Inspection of Fig. 3a reveals that while MELT and EM converge to a value of ≈ 22.4, ATOM1 and ATOM2 converge to 22. Therefore, when two of the frequencies do not lie on the Fourier grid, the state-of-the-art iterative algorithms converge to a larger value of the negative log-likelihood than the proposed methods. This is due to the fact that unlike the counterparts, the proposed

TABLE II
COMPARISON OF THE AVERAGE RUN TIME (IN SECONDS) OF THE ITERATIVE ALGORITHMS

Dimension m	**ATOM1**	**ATOM2** ($\gamma_0 = 10^{-4}$)	**ATOM2** ($\gamma_0 = 10^{-2}$)	**ATOM2** ($\gamma_0 = 10^{-1}$)	**MELT**[22]	**EM**[19]
4	0.028	1.309	0.047	0.014	0.051	0.026
8	0.032	1.503	0.164	0.055	0.071	0.035
16	0.163	6.912	0.522	0.166	0.162	0.081
32	0.473	9.569	2.484	0.825	0.663	0.348

TABLE III
COMPARISON OF THE AVERAGE MSE OF THE ITERATIVE ALGORITHMS

Dimension m	**ATOM1**	**ATOM2** ($\gamma_0 = 10^{-4}$)	**ATOM2** ($\gamma_0 = 10^{-2}$)	**ATOM2** ($\gamma_0 = 10^{-1}$)	**MELT**[22]	**EM**[19]
4	42.48	38.12	45.08	47.88	45.04	44.64
8	22.80	19.92	23.04	23.92	82.48	82.32
16	30.88	26.40	32.80	35.36	93.60	91.36
32	20.16	20.16	20.16	20.48	112.96	107.21



(a) (b)

Fig. 3. Negative log-likelihood (9) and the objective function of (10) vs. outer-iterations for $m = 6$, $n = 20$, and off-grid frequencies scenario.

algorithms estimate the Toeplitz covariance matrix without reparametrizing it via the CE technique and thus they are able to cover the whole set of Toeplitz covariance matrices. Furthermore, remarks similar to those made for the on-grid case hold true with reference to the results depicted in Fig. 3b.

In the following, the mean computational time[14] (averaged over 1000 Monte Carlo trials) of the proposed techniques and the counterparts is examined. As case studies, four different values of m are considered, i.e., $m \in \{4, 8, 16, 32\}$. Moreover, the data samples $\mathbf{y}_k$ are generated as (39) using $n = 4m$ samples, with $\boldsymbol{R} = \boldsymbol{T} + \boldsymbol{I}$. The Toeplitz covariance matrix $\mathbf{T}$ is generated assuming 3 equal power sources, i.e., with $p = [5, 5, 5]$, whose frequencies are randomly selected (at each trial) such that two of them lie on the Fourier grid of the DFT matrix, with $L = 2m-1$, whereas the third one is drawn from a uniform distribution over $[0, 2\pi]$. The iterative algorithms have been run until the following condition is met[15]

$$p(\mathbf{R}_{t-1}, \boldsymbol{X}_{t-1}) - p(\mathbf{R}_t, \boldsymbol{X}_t) \leq 10^{-4} \tag{40}$$

with $p(\mathbf{R}, \boldsymbol{X}) = \mathrm{Tr}(\boldsymbol{X}) + \log|\mathbf{R}|$ the objective function of problem (10), or until the maximum number of iterations (set equal to 1000) is reached. The average computational time of the different algorithms (possibly with different values of the hyperparameters) are reported in Table II. The results show that ATOM2 has, in general, a longer execution time than ATOM1. This is because the inner-loop of ATOM2 (based on Dykstra's algorithm) requires an higher number of iterations and hence a longer run time to converge than ATOM1 inner-loop (implemented via ADMM), and similar to those of EM/MELT when γ_0 is small, where the distance is minimized in a metric space is ill defined more and more. However, when $\gamma_0 = 10^{-1}$, the run times of ATOM1 and ATOM2 are comparable and similar to those of MELT and EM. Interestingly, Table III pinpoints that, for γ_0 sufficiently small, i.e., 10^{-4}, ATOM2 is generally able to reach MSE values smaller than ATOM1, reasonably to its adaptive *step-size* strategy (38), which allows it to provide better quality estimates than ATOM1 as the outer-loop iteration increases. It can also be seen that EM has the least computational time (at large values of m). Nevertheless, as shown in Table III, although the proposed algorithms have a slight longer computational time, the obtained estimates are superior, in terms of MSE, to those provided by MELT and EM.

Interestingly, as the data dimension increases, the resulting average MSE values reached by the ATOM2 using different γ_0 parameters becomes closer and closer. Therefore, for a sufficient larger data size, i.e., $m \geq 32$, $\gamma_0 = 10^{-1}$ represents an appropriate choice for ATOM2 implementation, as it offers a good performance with a reduced computational burden.

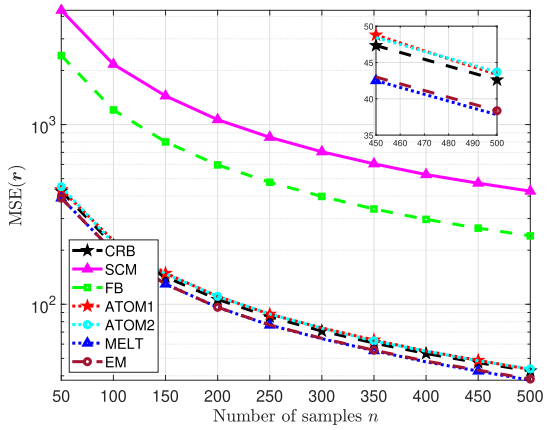
To further corroborate the computational efficiency of ATOM1 also from a practical point of view, several numerical examples are provided in the following by considering specific instances of problem (13), with $\boldsymbol{R}_t$ provided by the starting point of the estimation process. In particular, using the same setup parameters as in Tables II-III with $m = 8$ and $n = 4$, Table IV shows the average computational times and the average values of the objective function (12), computed over 500 Monte Carlo trials, achieved by either solving the SDP directly with the SeDuMi/SDPT3 solver or by employing ATOM1. Results reveal that the average value of the objective function achieved by ATOM1 substantially concides with that attained by the SDP solvers, but with a considerably faster processing time, reflecting its lower computational complexity as compared with the counterparts.

[14]The simulation has been executed using MATLAB R2020b on a desktop computer equipped with an Intel i5 processor and 16 GB of RAM.

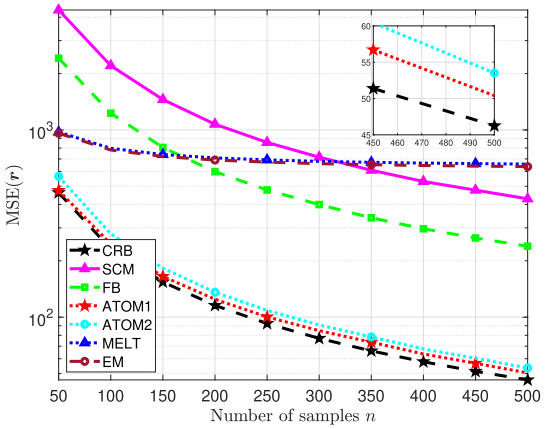
[15]For the execution of EM and MELT procedures, the exit condition is set as $f(\mathbf{R}_{t-1}) - f(\mathbf{R}_t) \leq 10^{-4}$.

TABLE IV
AVERAGE COMPUTATIONAL TIMES AND AVERAGE VALUES OF THE OBJECTIVE FUNCTION (12) ACHIEVED BY SEDUMI, SDPT3 AND ATOM1 FOR THE $m = 8,\ n = 4$ CASE

	SeDuMi	SDPT3	ATOM1
avg. obj. fun	19.425	19.425	19.447
avg. execution time	0.5099	0.5944	0.2073



(a)



(b)

Fig. 4. MSE vs. number of samples n for Toeplitz covariance matrix. a) on-grid frequencies; b) off-grid frequencies.

B. MSE Versus n for Toeplitz Covariance Matrix

For this case studies, it is assumed $m = 15$ and the number of samples n ranging between 50 and 500 in steps of 50. The data $\mathbf{y}_k \in \mathbb{C}^{15}$ are again simulated according to (39). Precisely, two different experiments are considered whereby the true Toeplitz covariance matrix is generated using on-grid[16] and off-grid frequencies,[17] respectively. The resulting MSE, computed over 1000 Monte Carlo trials, are illustrated in Fig. 4. Inspection of the curves depicted in Fig. 4a shows that, regardless of the number of samples n, in the first experiment ATOM1 and ATOM2 almost reach the CRB, whereas EM and MELT yield a slightly better performance, resulting in a deviation from the CRB. This can be explained observing that the derived CRB does not exploit the information that the frequencies lie on-grid. Fig. 4b highlight that in the second experiment, ATOM1 attain the best performance, with results quite close to the CRB and slightly better than ATOM2, with a limited gap between the corresponding curves. Furthermore, MELT and EM exhibit similar MSE values which seem to saturate as n increases. The performance behavior of Fig. 4b stems from the observation that, unlike MELT and EM, ATOM1 and ATOM2 are gridless methods, delivering the same performance regardless of the sources frequencies.

[16]The frequencies used in the first experiment are: $[0.2167, 0.6500, 1.0833, 1.3, 1.5166, 1.9500, 2.3833, 2.8166, 3.2499, 3.68324.1166, 4.5499, 4.9832, 5.4165, 5.8499]$ rad. Their corresponding powers increase linearly from 1 to 15 with a unit step.

[17]For the off-grid simulation, the frequencies $[1.3, 2.8166, 4.9832, 5.8499]$ rad are replaced with $[1.25, 3.01, 5.20, 5.8]$ rad, respectively.

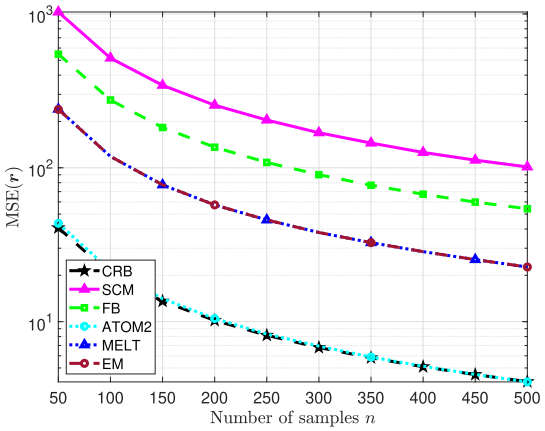


Fig. 5. MSE vs. number of samples n for banded Toeplitz covariance matrix.

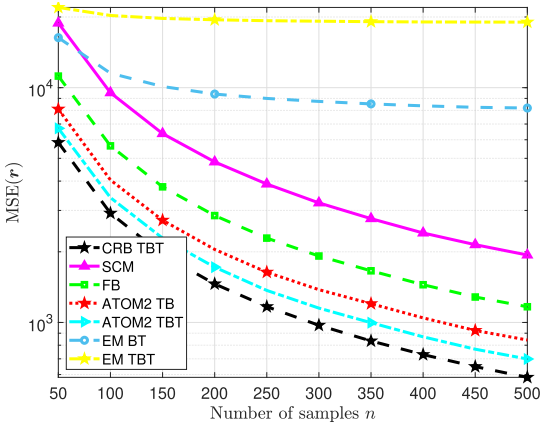


Fig. 6. MSE vs. number of samples n for TBT covariance matrix.

C. MSE Versus n for Banded Toeplitz Covariance Matrix

This subsection analyzes the performance in the case of covariance matrix belonging to the set of banded Toeplitz matrices. In particular, the same simulation setup as in Section V-B is considered, but enforcing the underlying covariance matrix to have a bandwidth $b = 6$. To this end, $\mathbf{R}$ is constructed by alternately projecting a random Hermitian matrix onto the set of banded Toeplitz matrices and the set of PSD matrices. Moreover, for this study case, ATOM2 is implemented according to the procedure described in Section III-E.1, namely explicitly including the banded Toeplitz structure in the constraint set.

Fig. 5 highlights that the bespoke implementation of ATOM2 delivers the best performance, with MSE values really

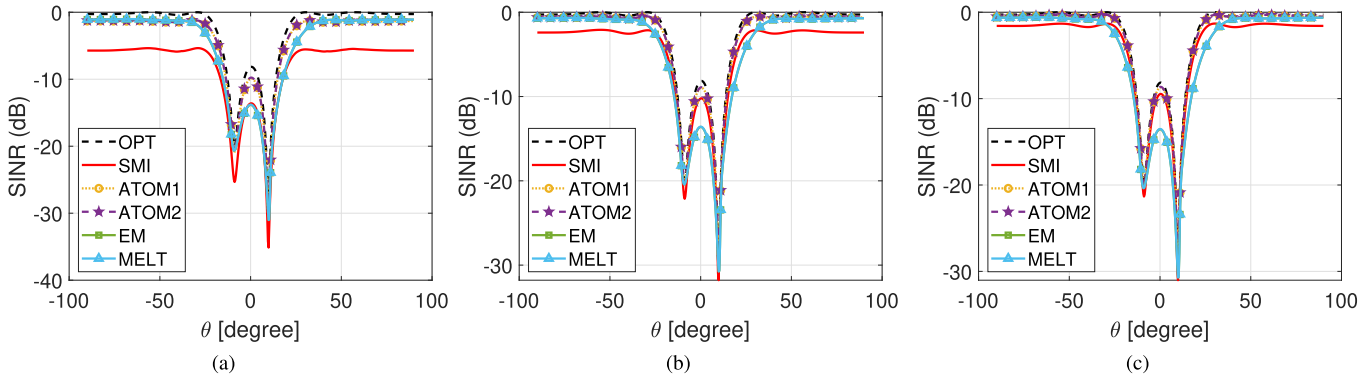


Fig. 7. Average SINR vs θ in the presence of two jammers, assuming $m = 6$ and a) $n = m$ b) $n = 2m$, and c) $n = 3m$.

close to the CRB. Furthermore, MELT and EM share the same performance with a noticeable gap w.r.t. ATOM2, which is expected since the aforementioned algorithms do not leverage the banded structure of the covariance matrix.

D. MSE Versus n for BT (TBT) Covariance Matrix

Here, the capabilities of ATOM2 are analyzed in the context of covariance matrix with TBT structure. To this end, assuming $m = 16$ and $p = 4$ blocks (each having block-size $l = 4$), the covariance matrix is modeled as $\mathbf{R} = \mathbf{T}_1 \otimes \mathbf{T}_1$, where $\mathbf{T}_1 \in \mathbb{C}^{l\times l}$ is a Toeplitz matrix constructed as in subsection V-A, with frequencies $[0.6, 1.4, 3.2, 5.1]$ rad and powers $[3, 6, 4, 1]$. Thus, each data snapshot $\mathbf{y}_k$ is drawn according to (39). The resulting MSE values (averaged over 1000 Monte Carlo trials) are displayed in Figure 6 versus the number of snapshots. Specifically, the performance of both the BT and the TBT extension of ATOM2 (described in Section III-E.2) are reported and compared with the CRB (see Appendix C) as well as with two EM-based estimators, tailored respectively for BT/TBT covariance matrix [78]. Inspection of the results reveals that ATOM2 TBT uniformly achieves the least MSE, with ATOM2 BT ranking second. As previously highlighted, the superior performance of the proposed method stems from the design criterion which does not require reparametrizing the covariance matrix using the CE.

E. Radar Application

In this subsection, the performance of the covariance estimation algorithms is evaluated with reference to the average achievable SINR in adaptive radar spatial processing context. To this end, let us consider a radar system equipped with a uniform linear array with $m = 6$ sensors, pointing toward the boresight direction. The inter-element distance between each sensor is set equal to $d = \lambda/2$, where λ is the radar operating wavelength.

For this simulation scenario, the interference covariance matrix is modeled as $\mathbf{R} = \mathbf{R}_s + \sigma_a^2\mathbf{I}$ where σ_a^2 is the power level of the white disturbance noise (assumed without loss of generality equal to 0 dB) and $\mathbf{R}_s$ is given by $\mathbf{R}_s = \sum_{l=1}^{J}\sigma_l^2\, \boldsymbol{s}(\phi_l)\boldsymbol{s}(\phi_l)^H$, where J is the number of uncorrelated narrow-band jammers and, for the l-th jammer,

$$\boldsymbol{s}(\phi_l) = \frac{1}{\sqrt{m}}[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi_l)}, \ldots, e^{j(m-1)\frac{2\pi}{\lambda}d\sin(\phi_l)}]^{\mathrm{T}} \quad (41)$$

is the steering vector in its direction-of-arrival ϕ_l, and σ_l^2 the corresponding interferer power.

The capabilities of the estimation methods are analyzed by means of the average SINR, computed as

$$\mathrm{SINR}_{\mathrm{avg}} = \frac{1}{K}\sum_{i=1}^{K}\frac{|\hat{\boldsymbol{w}}_i{}^H\boldsymbol{s}(\theta)|^2}{\hat{\boldsymbol{w}}_i^H\mathbf{R}\hat{\boldsymbol{w}}_i}, \quad (42)$$

where $K = 500$ is the number of Monte-Carlo trials and $\hat{\boldsymbol{w}}_i = \hat{\mathbf{R}}_i^{-1}\boldsymbol{s}(\theta)$ is the estimate of the optimal weight vector for adaptive spatial processing with $\hat{\mathbf{R}}_i$ the estimate of the interference-plus-noise covariance matrix for the i-th trial, computed either via the sample covariance matrix or enforcing the Toeplitz structure in the covariance matrix and employing the estimators ATOM1, ATOM2, EM, and MELT.

More precisely, $J = 2$ jammers, with powers $\sigma_1^2 = 30$ dB and $\sigma_2^2 = 20$ dB, respectively, impinging on the array from $\theta_1 = 9.8°$ and $\theta_2 = -8.8°$, is considered. As comparison terms, the optimum SINR, i.e., $\mathrm{SINR}_{\mathrm{OPT}} = \boldsymbol{s}(\theta)^H\mathbf{R}^{-1}\boldsymbol{s}(\theta)$ and the performance of the Sample Matrix Inversion (SMI) beamformer, are included too.

The average SINR versus $\theta \in \mathcal{T}$, with $\mathcal{T} = [-\pi/2, \pi/2]$ discretized with 500 equally-spaced points, is shown in Fig. 7, for $n \in \{m, 2m, 3m\}$. Inspection of the plots highlights that as the number of samples n increases, the results achieved by ATOM1 and ATOM2 gets closer and closer to the optimum, yielding superior performance w.r.t. the counterparts.

VI. Conclusion

In this paper, the MLE problem for TSC matrices has been addressed. Precisely, by reformulating appropriately the MLE optimization problem and leveraging the MM framework, two iterative algorithms ATOM1 and ATOM2 have been developed. Both inherit the key properties of MM i.e., they monotonically decrease the underlying cost function with guaranteed convergence to a stationary point of the equivalent MLE problem. Subsequently, ATOM2 has been extended to handle covariance matrix MLE forcing other Toeplitz-related structures, such as banded Toeplitz, BT, and TBT. Simulation

results have indicated that the proposed algorithms can perform better than some state-of-the-art techniques in terms of MSE and the SINR metrics.

Some of the possible future research directions are now outlined. In particular, ATOM2 could be further extended to include the cases of low rank TSC, with the rank assumed either known or unknown at the design stage, as well as covariance matrix with an upper bound to the condition number. Another possible extension of the proposed technique could be MLE of a Toeplitz covariance matrix assuming a compound Gaussian distribution for the underlining data which has a significant application in low-grazing angle target detection [79], [80]. Moreover, acceleration methods inspired for instance by the SQUAREd iterative Methods (SQUAREM) [81] could be investigated. In addition, the design of sub-optimal optimization strategies (e.g., based on the gradient projection method) with an improved computational burden (a valuable feature for real-time applications) is definitely worth to be pursued. Finally, it would be of great interest to apply the devised gridless framework to the problems of direction of arrival/frequency estimation [26], [82] [83], [84] and to compare the subsequent performance with existing methods already available in the open literature.

APPENDIX

A. Proof of Equivalence Between (8) and (10)

Let $\mathbf{R}^\star$ be an optimal solution to (8), then $(\boldsymbol{X}^\star, \mathbf{R}^\star)$, with $\boldsymbol{X}^\star = \boldsymbol{Y}^H \mathbf{R}^{\star -1} \boldsymbol{Y}$, is feasible for (10) and the two problems have the same objective values. This means that

$$v(8) \geq v(10), \tag{43}$$

where $v(\cdot)$ indicates the optimal value of the corresponding optimization problem.

Moreover, for any fixed $\mathbf{R}_1 \succ 0$, concentrating the objective function of (10) with respect to $\boldsymbol{X}$ (which is tantamount to placing $\boldsymbol{X} = \boldsymbol{Y}^H \mathbf{R}_1^{-1} \boldsymbol{Y}$), it follows that the concentrated optimization problem is

$$\underset{\mathbf{R}_1 \succeq 0}{\text{minimize}}\ \text{Tr}(\mathbf{R}_{FB} \mathbf{R}_1^{-1}) + \log |\mathbf{R}_1|, \tag{44}$$

due to Schur complement Theorem and the monotonicity of the trace operator with respect to generalized matrix inequality "$\succeq$". Finally, being by assumption (8) solvable, any minimizer of (44) satisfies $\mathbf{R}_1^\star \succ 0$ with a corresponding optimal solution to (10) given by $(\mathbf{R}_1^\star, \boldsymbol{Y}^H \mathbf{R}_1^{\star -1} \boldsymbol{Y})$. This implies that

$$v(8) \leq v(10). \tag{45}$$

Capitalizing on (43) and (45) as well as the above considerations, it follows that $v(8) = v(10)$ and given an optimal solution $(\mathbf{R}_1^\star, \boldsymbol{X}_1^\star)$ to (10), $\mathbf{R}_1^\star$ is also optimal to (8) and viceversa, given an optimal solution $\mathbf{R}^\star$ to (8) $(\boldsymbol{X}^\star, \mathbf{R}^\star)$ is an optimal point to (10).

B. Proof of Theorem 3.2

To begin with, let us denote by $h(\boldsymbol{E}|\boldsymbol{E}_t)$ either the objective function involved in the surrogate optimization problem of ATOM1 (12) or ATOM2 (15), where $\boldsymbol{E} = \text{diag}(\boldsymbol{X}, \mathbf{R})$. This function, regardless of the method, satisfies the following two inequalities

$$h(\boldsymbol{E}_t|\boldsymbol{E}_t) = l(\boldsymbol{E}_t) \tag{46}$$

$$h(\boldsymbol{E}_{t+1}|\boldsymbol{E}_t) \geq l(\boldsymbol{E}_{t+1}), \tag{47}$$

where $l(\boldsymbol{E}) = \text{Tr}(\boldsymbol{X}) + \log |\mathbf{R}|$. Leveraging the above inequalities, it follows that

$$l(\boldsymbol{E}_{t+1}) \overset{(a)}{\leq} h(\boldsymbol{E}_{t+1}|\boldsymbol{E}_t) \overset{(b)}{\leq} h(\boldsymbol{E}_t|\boldsymbol{E}_t) \overset{(c)}{=} l(\boldsymbol{E}_t). \tag{48}$$

In (48), the inequality (a) and equality (c) stem from (47) and (46), respectively; besides, the inequality (b) is obtained by exploiting the fact that ATOM1 and ATOM2 globally solve the corresponding convex surrogate optimization problem. Therefore, (48) implies that the sequence of objective value of Problem (16) generated by the proposed algorithms is monotonically decreasing, i.e.,

$$l(\boldsymbol{E}_0) \geq l(\boldsymbol{E}_1) \geq l(\boldsymbol{E}_2) \geq \cdots \tag{49}$$

Next, let us denote by $\boldsymbol{Z}$ a cluster point to $\{\boldsymbol{E}_t\}$ and let $\{\boldsymbol{E}_{r_t}\}$ be a subsequence of $\{\boldsymbol{E}_t\}$ converging to $\boldsymbol{Z}$. Then, from (46), (47), and (49)

$$\begin{aligned} h\left(\boldsymbol{E}_{r_{t+1}}|\boldsymbol{E}_{r_{t+1}}\right) &= l\left(\boldsymbol{E}_{t_{j+1}}\right) \leq l\left(\boldsymbol{E}_{r_t+1}\right) \\ &\leq h\left(\boldsymbol{E}_{r_t+1}|\boldsymbol{E}_{r_t}\right) \leq h\left(\boldsymbol{E}|\boldsymbol{E}_{r_t}\right), \forall \text{ feasible } \boldsymbol{E}. \end{aligned} \tag{50}$$

Thus, letting $t \to \infty$

$$h(\boldsymbol{Z}|\boldsymbol{Z}) \leq h(\boldsymbol{E}|\boldsymbol{Z}), \tag{51}$$

which implies that $h'(\boldsymbol{Z}|\boldsymbol{Z}; \boldsymbol{D}) \geq 0$, for any feasible direction $\boldsymbol{D}$ from any feasible $\boldsymbol{Z}$, where $h'(\boldsymbol{G}|\boldsymbol{Z}; \boldsymbol{D})$ denotes the directional derivative in a feasible direction $\boldsymbol{D}$ from $\boldsymbol{G}$ of the surrogate function at point $\boldsymbol{Z}$. Finally, by Proposition 1 in [36], the surrogate function $h(\boldsymbol{E}|\boldsymbol{Z})$ and the objective function $l(\boldsymbol{E})$ have the same first order behavior at $\boldsymbol{E} = \boldsymbol{Z}$ since both of them are differentiable at any feasible $\boldsymbol{E}$. Therefore, $h'(\boldsymbol{Z}|\boldsymbol{Z}; \boldsymbol{D}) \geq 0$ implies that $l'(\boldsymbol{Z}; \boldsymbol{D}) \geq 0$. Hence, $\boldsymbol{Z}$ is a stationary point of the objective function $l(\boldsymbol{E})$.

C. CRB of Banded Toeplitz and TBT Covariance Model

Herein, the CRB of Banded Toeplitz and TBT covariance model are provided.

1) Banded Toeplitz Matrix: In the case of banded Toeplitz matrix with bandwidth b, the first row of the covariance matrix $\mathbf{R} \in \mathbb{H}^{m \times m}$ has only $b+1$ non-zero terms. Therefore, $\mathbf{R}$ can be parameterized via $\boldsymbol{\theta} = [r_1, \Re(r_2), \cdots \Re(r_{b+1}), \Im(r_2), \ldots, \Im(r_{b+1})]^T \in \mathbb{R}^{2b+1}$. Besides $\mathbf{R}$ can be expressed in terms of basis matrices $\boldsymbol{B}_g^{\text{Toep}}$ and real coefficients $\boldsymbol{\theta}$

$$\mathbf{R} = \sum_{g=1}^{b+1} \theta_g \Re(\boldsymbol{B}_g^{\text{Toep}}) + j \sum_{g=b+2}^{2b+1} \theta_g \Im(\boldsymbol{B}_{g-b}^{Toep}) \tag{52}$$

and consequently

$$\frac{\partial \mathbf{R}}{\partial \theta_i} = \begin{cases} \Re(\boldsymbol{B}_i^{\text{Toep}}) & 1 \leq i \leq b+1 \\ j\Im(\boldsymbol{B}_{i-b}^{\text{Toep}}) & b+2 \leq i \leq 2b+1 \end{cases}.$$

Substituting $\frac{\partial \mathbf{R}}{\partial \theta_i}$ in (34), yields the FIM for banded Toeplitz covariance matrix.

2) Toeplitz-Block-Toeplitz Matrix: Before proceeding further, it is worth noting that a TBT matrix composed of p blocks of size l can be parameterized by the vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_0^T, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_{P-1}^T]^T \in \mathbb{R}^{2l-1+(p-1)(4l-2)}$ whereby $\boldsymbol{\theta}_0 = [r_{0,1}, \Re(r_{0,2}), \ldots, \Re(r_{0,l}), \Im(r_{0,2}), \ldots, \Im(r_{0,l})]^T \in \mathbb{R}^{2l-1}$ and $\boldsymbol{\theta}_p = [\Re(r_{p,1}), \ldots, \Re(r_{p,l}), \Im(r_{p,1}), \ldots, \Im(r_{p,l}), \Re(c_{p,2}), \ldots, \Re(r_{p,l}), \Im(r_{p,2}), \ldots, \Im(r_{p,l})]^T \in \mathbb{R}^{4l-2}$, $p = 1, \ldots, P-1$, with $r_{p,n}$ and $c_{p,n}$ the n-th row and n-th column of $\mathbf{R}_p$, respectively. Indeed, the TBT covariance matrix can be expressed as

$$\mathbf{R}^{\mathrm{TBT}} = \boldsymbol{C}_0 \otimes \mathbf{R}_0 + \sum_{w=1}^{p-1} \left((\boldsymbol{C}_w \otimes \mathbf{R}_w^H) + \left(\boldsymbol{C}_w^T \otimes \mathbf{R}_w \right) \right), \tag{53}$$

where

$$\mathbf{R}_0 = \sum_{g=1}^{l} \theta_{0,g} \Re(\boldsymbol{B}_g^{\mathrm{Toep}}) + j \sum_{g=l+1}^{2l-1} \theta_{0,g} \Im(\boldsymbol{B}_{g-l+1}^{Toep}) \tag{54}$$

and, for $w = 1, \ldots, p-1$,

$$\begin{aligned} \mathbf{R}_w = &\sum_{g=1}^{l} [\theta_{w,g} + j\theta_{w,g+l}] \Re(\boldsymbol{D}_g) \\ &+ \sum_{g=2l+1}^{3l-1} [\theta_{w,g} + j\theta_{w,g+l-1}] \Im(\boldsymbol{D}_{g-2l+1}) \end{aligned} \tag{55}$$

with $\theta_{w,g}$ the g-th element of $\boldsymbol{\theta}_w$, $\boldsymbol{D}_g = \boldsymbol{B}_g^{\mathrm{Toep}}$ as long as $g = 1$ and $1/2((\boldsymbol{B}_g^{\mathrm{Toep}})^T + j(\boldsymbol{B}_g^{\mathrm{Toep}})^T)$ elsewhere, whereas the $(i,k)^{th}$ element of the matrix $\boldsymbol{C}_w \in \mathbb{R}^{l \times l}$ is given by

$$[\boldsymbol{C}_w]_{i,k} = \begin{cases} 1 & i-k = w \\ 0 & \text{otherwise} \end{cases}.$$

That said, $\frac{\partial \mathbf{R}^{\mathrm{TBT}}}{\partial \theta_{w,g}}$ is given by

$$\frac{\partial \mathbf{R}^{\mathrm{TBT}}}{\partial \theta_{w,g}} = \begin{cases} \boldsymbol{C}_0 \otimes \Re(\boldsymbol{B}_g^{\mathrm{Toep}}) & 1 \le g \le l, w = 0 \\ \boldsymbol{C}_0 \otimes j\Im(\boldsymbol{B}_{g-l+1}^{\mathrm{Toep}}) & l+1 \le g \le 2l-1, w = 0 \\ \boldsymbol{C}_w \otimes \Re(\boldsymbol{D}_g)^T & \\ \quad + \boldsymbol{C}_w^T \otimes \Re(\boldsymbol{D}_g) & 1 \le g \le l, w > 0 \\ \boldsymbol{C}_w \otimes (-j)\Re(\boldsymbol{D}_{g-l})^T & \\ \quad + \boldsymbol{C}_w^T \otimes j\Re(\boldsymbol{D}_{g-l}) & l+1 \le g \le 2l, w > 0 \\ \boldsymbol{C}_w \otimes \Im(\boldsymbol{D}_{g-2l+1})^T & \\ \quad + \boldsymbol{C}_w^T \otimes \Im(\boldsymbol{D}_{g-2l+1}) & 2l+1 \le g \le 3l-1, w > 0 \\ \boldsymbol{C}_w \otimes (-j)\Im(\boldsymbol{D}_{g-3l+2})^T & \\ \quad + \boldsymbol{C}_w^T \otimes j\Im(\boldsymbol{D}_{g-3l+2}) & 3l \le g \le 4l-2, w > 0 \end{cases}$$

which, employed in (34), yields the FIM for TBT covariance matrix.

REFERENCES

[1] 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, May 1992.

[2] P. Stoica et al., *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[3] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[4] A. Farina, *Antenna-based Signal Processing Techniques for Radar Systems*. Norwood, MA, USA: Artech House, 1992.

[5] B. C. Armstrong, "Performance of adaptive optimal Doppler processors in heterogeneous clutter," *IEE Proc., Radar, Sonar Navigat.*, vol. 142, no. 4, p. 179, 1995.

[6] B. Himed and W. L. Melvin, "Analyzing space-time adaptive processors using measured data," in *Proc. Conf. Rec. 31st Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 1997, pp. 930–935.

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

[8] W. J. J. Roberts and Y. Ephraim, "Hidden Markov modeling of speech using Toeplitz covariance matrices," *Speech Commun.*, vol. 31, no. 1, pp. 1–14, May 2000.

[9] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4959–4973, Oct. 2014.

[10] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5565–5576, Nov. 2014.

[11] P. Pal and P. P. Vaidyanathan, "A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 737–741, Jun. 2014.

[12] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, vol. 1. Dordrecht, The Netherlands: Kluwer, 2003.

[13] C. Caratheodory and L. Fejér, "On the connection between the extremes of harmonic functions and their coefficients and on Picard–Landau's theorem," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, no. 1, pp. 218–239, 1884.

[14] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. New York, NY, USA: Chelsea, 1958.

[15] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, vol. 7. Amsterdam, The Netherlands: Elsevier, 2018, pp. 509–581.

[16] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Int.*, vol. 33, no. 3, pp. 347–366, Sep. 1973.

[17] A. Dembo, C. L. Mallows, and L. A. Shepp, "Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1206–1212, Nov. 1989.

[18] S. Haykin, *Nonlinear Methods of Spectral Analysis*, vol. 34. Berlin, Germany: Springer-Verlag, 2006.

[19] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and Toeplitz constrained covariances," *Proc. IEEE*, vol. 75, no. 7, pp. 892–907, Jul. 1987.

[20] M. J. Turmon and M. I. Miller, "Maximum-likelihood estimation of complex sinusoids and Toeplitz covariances," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1074–1086, May 1994.

[21] L. P. B. Christensen, "An EM-algorithm for band-Toeplitz covariance matrix estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2007, pp. III-1021–III-1024.

[22] P. Babu, "MELT—Maximum-likelihood estimation of low-rank Toeplitz covariance matrix," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1587–1591, Nov. 2016.

[23] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[24] 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.

[25] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.

[26] Z. Yang, X. Chen, and X. Wu, "A robust and statistically efficient maximum-likelihood method for DOA estimation using sparse linear arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 6798–6812, Oct. 2023.

[27] X. Du, A. Aubry, A. De Maio, and G. Cui, "Toeplitz structured covariance matrix estimation for radar applications," *IEEE Signal Process. Lett.*, vol. 27, pp. 595–599, 2020.

[28] F. Barbaresco, "Information geometry of covariance matrix: Cartan–Siegel homogeneous bounded domains, Mostow/Berger fibration and Frechet median," in *Matrix Information Geometry*. Berlin, Germany: Springer, 2013, pp. 199–255.

[29] B. Balaji, F. Barbaresco, and A. Decurninge, "Information geometry and estimation of Toeplitz covariance matrices," in *Proc. Int. Radar Conf.*, Oct. 2014, pp. 1–4.

[30] A. Aubry, P. Babu, A. De Maio, and R. Jyothi, "ATOM for MLE of Toeplitz structured covariance matrices for RADAR applications," in *Proc. IEEE Radar Conf. (RadarConf)*, Mar. 2022, pp. 1–6.

[31] P. Chen, W. L. Melvin, and M. C. Wicks, "Screening among multivariate normal data," *J. Multivariate Anal.*, vol. 69, no. 1, pp. 10–29, Apr. 1999.

[32] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Covariance matrix estimation via geometric barycenters and its application to radar training data selection," *IET Radar, Sonar Navigat.*, vol. 7, no. 6, pp. 600–614, Jul. 2013.

[33] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Median matrices and their application to radar training data selection," *IET Radar, Sonar Navigat.*, vol. 8, no. 4, pp. 265–274, Apr. 2014.

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1987.

[35] Y. Sun, P. Babu, and D. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.

[36] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.

[37] A. Aubry, A. De Maio, A. Zappone, M. Razaviyayn, and Z.-Q. Luo, "A new sequential optimization procedure and its applications to resource allocation for wireless systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6518–6533, Dec. 2018.

[38] A. Cantoni and P. Butler, "Properties of the eigenvectors of persymmetric matrices with applications to communication theory," *IEEE Trans. Commun.*, vol. COM-24, no. 8, pp. 804–809, Aug. 1976.

[39] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

[40] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[41] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[42] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. Hoboken, NJ, USA: Wiley, 1995.

[43] T. T. Wu and K. Lange, "The MM alternative to EM," *Stat. Sci.*, vol. 25, no. 4, pp. 492–505, 2010.

[44] W. J. Heiser, "Convergent computation by iterative majorization," in *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, Ed., Oxford, U.K.: Oxford Univ. Press, 1995, pp. 157–189.

[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[46] Z.-Q. Luo, W.-K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[47] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1389–1403, Mar. 2013.

[48] F. Xu and P. Pan, "A new algorithm for positive semidefinite matrix completion," *J. Appl. Math.*, vol. 2016, no. 1, pp. 1–5, 2016.

[49] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 946–977, Jan. 2013.

[50] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*. Norwell, MA, USA: Now Publishers, 2011.

[51] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 343–352.

[52] P. G. Martinsson, V. Rokhlin, and M. Tygert, "A fast algorithm for the inversion of general Toeplitz matrices," *Comput. Math. Appl.*, vol. 50, nos. 5–6, pp. 741–752, Sep. 2005.

[53] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented Lagrangian methods for semidefinite programming," *Math. Program. Comput.*, vol. 2, nos. 3–4, pp. 203–230, Dec. 2010.

[54] C. Chen et al., "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," *Math. Program.*, vol. 155, no. 1, pp. 57–79, 2016.

[55] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequal. in System and Control Theory*. Philadelphia, PA, USA: SIAM, 1994.

[56] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, Sep. 1996.

[57] W. Cheney and A. A. Goldstein, "Proximity maps for convex sets," *Proc. Amer. Math. Soc.*, vol. 10, no. 3, p. 448, Jun. 1959.

[58] C. L. Byrne, "Alternating minimization and alternating projection algorithms: A tutorial," *Sci. New York*, pp. 1–41, Mar. 2011. [Online]. Available: https://faculty.uml.edu/cbyrne/CandT.pdf

[59] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[60] D. Blatt and A. O. Hero, "Energy-based sensor network source localization via projection onto convex sets," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3614–3619, Sep. 2006.

[61] D. Henrion and J. Malick, "Projection methods for conic feasibility problems: Applications to polynomial sum-of-squares decompositions," *Optim. Methods Softw.*, vol. 26, no. 1, pp. 23–46, Feb. 2011.

[62] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," in *Advances in Order Restricted Statistical Inference*. New York, NY, USA: Springer, 1986, pp. 28–47.

[63] R. L. Dykstra, "An algorithm for restricted least squares regression," *J. Amer. Stat. Assoc.*, vol. 78, no. 384, p. 837, Dec. 1983.

[64] 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.

[65] 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.

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

[67] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Radar detection of distributed targets in homogeneous interference whose inverse covariance structure is defined via unitary invariant functions," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4949–4961, Oct. 2013.

[68] 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, Apr. 1997, pp. 3493–3496.

[69] X. Cui, Z. Li, J. Zhao, D. Zhang, and J. Pan, "Covariance matrix regularization for banded Toeplitz structure via Frobenius-norm discrepancy," in *Matrices, Statistics and Big Data*. Cham, Switzerland: Springer, 2016, pp. 111–125.

[70] K. Greenewald and A. O. Hero, "Regularized block Toeplitz covariance matrix estimation via Kronecker product expansions," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Jun. 2014, pp. 9–12.

[71] Y. I. Abramovich, B. A. Johnson, and N. K. Spencer, "Two-dimensional multivariate parametric models for radar applications—Part I: Maximum-entropy extensions for Toeplitz-block matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5509–5526, Nov. 2008.

[72] Y. I. Abramovich, B. A. Johnson, and N. K. Spencer, "Two-dimensional multivariate parametric models for radar applications—Part II: Maximum-entropy extensions for Hermitian-block matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5527–5539, Nov. 2008.

[73] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

[74] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[75] T. J. Moore, R. J. Kozick, and B. M. Sadler, "The constrained Cramér–Rao bound from the perspective of fitting a model," *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 564–567, Aug. 2007.

[76] E. Nitzan, T. Routtenberg, and J. Tabrikian, "Cramér–Rao bound for constrained parameter estimation using Lehmann-unbiasedness," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 753–768, Feb. 2019.

[77] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[78] D. R. Fuhrmann and T. A. Barton, "Estimation of block-Toeplitz covariance matrices," in *Proc. Conf. Rec. 24th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Los Alamitos, CA, USA: IEEE Computer Society, 1990, p. 779.

[79] K. Ward, C. Baker, and S. Watts, "Maritime surveillance radar. Part 1: Radar scattering from the ocean surface," *IEE Proc. F, Radar Signal Process.*, vol. 137, no. 2, pp. 51–62, 1990.

[80] F. Gini and M. V. Greco, "Suboptimum approach to adaptive coherent radar detection in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 1095–1104, Jul. 1999.

[81] Y. Du and R. Varadhan, "SQUAREM: An R package for off-the-shelf acceleration of EM, MM and other EM-like monotone algorithms," *J. Stat. Softw.*, vol. 92, no. 7, pp. 1–41, 2020.

[82] R. R. Pote and B. D. Rao, "Maximum likelihood-based gridless DoA estimation using structured covariance matrix recovery and SBL with grid refinement," *IEEE Trans. Signal Process.*, vol. 71, pp. 802–815, 2023.

[83] X. Wu, Z. Yang, P. Stoica, and Z. Xu, "Maximum likelihood line spectral estimation in the signal domain: A rank-constrained structured matrix recovery approach," *IEEE Trans. Signal Process.*, vol. 70, pp. 4156–4169, 2022.

[84] X. Wu, Z. Yang, and Z. Xu, "Multichannel frequency estimation in challenging scenarios via structured matrix embedding and recovery (StruMER)," *IEEE Trans. Signal Process.*, vol. 71, pp. 3242–3256, 2023.

Augusto Aubry (Senior Member, IEEE) received the Dr.Eng. degree (Hons.) in telecommunication engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2007 and 2011, respectively. From February 2012 to April 2012, he was a Visiting Researcher with Hong Kong Baptist University, Hong Kong. He is currently an Associate Professor with the University of Naples Federico II. His research interests include statistical signal processing and optimization theory, with an emphasis on MIMO communications and radar signal processing. He was a co-recipient of the 2013 Best Paper Award (titled to B. Carlton) of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter." He was a recipient of the 2022 IEEE Fred Nathanson Memorial Award as the Young (less than 40 years of age) AESS Radar Engineer 2022, with the following citation for "outstanding contributions to the application of modern optimization theory to radar waveform design and adaptive signal processing."

Prabhu Babu received the Ph.D. degree in electrical engineering from Uppsala University, Uppsala, Sweden, in 2012. From 2013 to 2016, he was a Post-Doctoral Fellow with The Hong Kong University of Science and Technology. He is currently an Associate Professor with the Centre for Applied Research in Electronics, Indian Institute of Technology Delhi, New Delhi, India.

Antonio De Maio (Fellow, IEEE) received the Dr.Eng. (Hons.) and Ph.D. degrees in information engineering from the University of Naples Federico II, Naples, Italy, in 1998 and 2002, respectively. From October 2004 to December 2004, he was a Visiting Researcher with U.S. Air Force Research Laboratory, Rome, NY, USA. From November 2007 to December 2007, he was a Visiting Researcher with The Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the University of Naples Federico II. His research interests include statistical signal processing, with an emphasis on radar detection, optimization theory applied to radar signal processing, and multiple-access communications. He was a recipient of the 2010 IEEE Fred Nathanson Memorial Award as the Young (less than 40 years of age) AESS Radar Engineer 2010, whose performance is particularly noteworthy as evidenced by contributions to the radar art over a period of several years, with the following citation for "robust CFAR detection, knowledge-based radar signal processing, and waveform design and diversity." He was a co-recipient of the 2013 Best Paper Award (titled to B. Carlton) of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter." He was also a recipient of the 2024 IEEE Warren White Award for outstanding achievements due to a major technical advance (or series of advances) in the art of radar engineering, with the citation for "Contributions to Radar Signal Processing Techniques for Target Detection, Waveform Design, and Electronic Protection."

Massimo Rosamilia (Member, IEEE) received the B.S. (Hons.) and M.S. degrees in computer engineering from the University of Salerno, Fisciano, Italy, in 2017 and 2019, respectively, and the Ph.D. degree (cum laude) in information technologies and electrical engineering from the University of Naples Federico II, Naples, Italy, in 2023. From September 2021 to November 2021, he was a Visiting Ph.D. Student with Cranfield University, Shrivenham, U.K. From September 2022 to December 2022, he was a Visiting Ph.D. Student with the University of Luxembourg, Luxembourg, and from June 2024 to August 2024, he was a Visiting Researcher with the University of Birmingham, U.K. He is currently an assistant professor with the University of Naples Federico II. His research interests include statistical signal processing, cognitive systems, waveform design, and optimization theory, with applications to radar problems. He ranked second in the Student Contest of the 1st International Virtual School on Radar Signal Processing, in 2020. In 2023, he co-authored the article winning the SET Panel Best Paper Award (Young Scientist) at the NATO SET-319 Specialists' Meeting on "New Mathematical Frontiers for Multi-Dimensional Radar Systems." He was a recipient of the 2023 GTTI Ph.D. Award for Ph.D. Theses in the field of communication technologies.

Open Access funding provided by 'Università degli Studi di Napoli "Federico II"' within the CRUI CARE Agreement