MELT—Maximum-Likelihood Estimation of Low-Rank Toeplitz Covariance Matrix

Prabhu Babu

***Abstract*—In this letter, we develop a low-complexity algorithm named maximum-likelihood estimation of low-rank Toeplitz covariance matrix (MELT) to solve the maximum-likelihood estimation of a low-rank Toeplitz covariance matrix. Our derivation of MELT is based on the technique of majorization–minimization (MM), in which we design and optimize a novel tight upper-bound function. MELT is an iterative algorithm, and its each iterative step is a closed-form update, which can be implemented efficiently by fast Fourier transforms. As MELT is based on MM, it enjoys nice properties such as monotonicity and guaranteed convergence to a stationary point. Finally, we numerically show that the performance of MELT is much better than some of the algorithms currently available in the literature.**

***Index Terms*—Majorization–minimization (MM), maximum-likelihood estimation (MLE), Toeplitz covariance matrix.**

I. Data Model and Problem Formulation

Let $\{\mathbf{y} \in \mathbb{C}^{M\times 1}\}$ denote an M-variate random process following complex Gaussian distribution with zero mean and covariance matrix $\mathbf{R}$, and $\{\mathbf{y}_k\}_{k=1}^{N}$ denote the N independent samples of the underlying random process. Given that $N > M$, the maximum-likelihood (ML) estimation problem for estimating $\mathbf{R}$ would be

$$\text{MLE}: \quad \underset{\mathbf{R}}{\text{minimize}} \left\{ f(\mathbf{R}) \triangleq \text{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right) + \log|\mathbf{R}| \right\} \quad (1)$$

where $\hat{\mathbf{R}} \triangleq \frac{1}{N}\sum_{k=1}^{N} \mathbf{y}_k\mathbf{y}_k^H$ denotes the sample covariance matrix. Problem (1) has a unique minimizer, which is given by $\mathbf{R}_{\text{ML}} = \hat{\mathbf{R}}$. However, if the random process is also assumed to be stationary, which leads to $\mathbf{R}$ being a Hermitian Toeplitz matrix, no analytical solution for (1) is available. Instead, only iterative techniques are available to arrive at a minimum of (1) (see, e.g., [1]–[3]). Now, on top of that if additional rank constraints are imposed in (1), there are not many computational methods available to directly solve the problem in (1). Some of the applications where we encounter this kind of covariance models with some *a priori* structures are knowledge-aided radars [4], hyperspectral imaging [5], direction of arrival estimation [6], and adaptive beamforming [7]. Li *et al.* [8] have proposed a closed-form estimator for Hermitian Toeplitz covariance matrices by invoking the extended invariance principle. Similarly, Steiner and Gerlach [9] have proposed an estimator for a covariance matrix having the form of the identity matrix plus an unknown positive semidefinite Hermitian matrix; some optimal properties of the estimator proposed in [9] were proved in [10]. Recently, Aubry *et al.* [11], have proposed an ML estimator for a structured covariance matrix with a condition number constraint. For covariance matrices taking the form of identity plus a positive semidefinite matrix, the problem of shrinkage parameter selection is addressed in [12]. Recently, the idea of robust estimation of covariance matrices under different structural constraints was explored in [13]. For our problem in (1), most of the methods available in the literature either solve the problem in (1) under only the low-rank constraint neglecting the Toeplitz constraint [14] or relax the Toeplitz constraint at the start and obtain an intermediate solution and then tweak the intermediate solution to satisfy the Toeplitz constraint [15].

In this letter, we directly try and solve the ML estimation problem in (1) under the low-rank Toeplitz constraint. The covariance matrix is assumed to have the following form:

$$\mathbf{R} = \mathbf{T} + \sigma\mathbf{I} \quad (2)$$

where $\mathbf{T}$ denotes a low-rank Toeplitz matrix and $\sigma > 0$ is a known scalar. In applications such as radar, the signal part is captured in $\mathbf{T}$ and $\sigma\mathbf{I}$ would be coming from the noise, and the noise floor level σ can be computed by some standard procedures [16]. The rank of $\mathbf{T}$ denoted by $\rho < M$ is assumed to be known, thus the optimization problem to be solved can be given as

$$\begin{aligned} \underset{\mathbf{R}}{\text{minimize}} \quad & \text{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right) + \log|\mathbf{R}| \\ \text{subject to} \quad & \mathbf{R} = \mathbf{T} + \sigma\mathbf{I} \\ & \text{rank}(\mathbf{T}) \le \rho, \mathbf{T} \text{ is a Hermitian Toeplitz matrix.} \end{aligned} \quad (3)$$

In the next section, we will develop a computationally less intensive algorithm named maximum-likelihood estimation of low-rank Toeplitz covariance matrix (MELT) to solve (3), and we will use the approach of majorization–minimization (MM) to tackle the problem. In Section II, we will start with a brief description of the MM procedure and derive the MELT algorithm. In Section III we will show some numerical results, and Section IV concludes the letter.

II. Low-Rank Toeplitz Covariance Matrix Estimation

In this section, we start with a brief description of the MM procedure, interested readers may refer [17] for more details on MM.

A. MM

Consider the following optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S} \end{aligned} \quad (4)$$

where $\mathcal{S}$ denotes the constraint set. With an initial point $\mathbf{x}_0 \in \mathcal{S}$, the MM algorithm generates a sequence of feasible points

Manuscript received July 13, 2016; revised September 8, 2016 and September 9, 2016; accepted September 9, 2016. Date of publication September 13, 2016; date of current version September 28, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Augusto Aubry.

P. Babu is with Centre for Applied Research in Electronics, IIT Delhi, New Delhi 110016, India (e-mail: prabhubabu@care.iitd.ac.in).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2608845

$\{\mathbf{x}_{t+1}\}$ by solving the following optimization problem:

$$\mathbf{x}_{t+1} \stackrel{\Delta}{=} \arg\min_{\mathbf{x}\in\mathcal{S}} g\left(\mathbf{x}|\mathbf{x}_t\right) \tag{5}$$

where the surrogate function defined by $g(\mathbf{x}|\mathbf{x}_t)$ required to satisfy the following properties:

(P1) $f(\mathbf{x}_t) = g(\mathbf{x}_t|\mathbf{x}_t)$;

(P2) $f(\mathbf{x}) \leq g(\mathbf{x}|\mathbf{x}_t) \quad \forall \mathbf{x} \in \mathcal{S}$

The sequence $\{f(\mathbf{x}_t)\}$ is nonincreasing since

$$f\left(\mathbf{x}_t\right) \stackrel{(\mathrm{P1})}{=} g\left(\mathbf{x}_t|\mathbf{x}_t\right) \stackrel{(5)}{\geq} g\left(\mathbf{x}_{t+1}|\mathbf{x}_t\right) \stackrel{(\mathrm{P2})}{\geq} f\left(\mathbf{x}_{t+1}\right). \tag{6}$$

Under some mild conditions on $f(\cdot)$, $g(\cdot)$, and $\mathcal{S}$, the sequence $f(\mathbf{x}_t)$ can be shown to be converging to a stationary point of the problem in (4).

B. Circulant Embedding

Before developing the MELT algorithm, in this section we will discuss the circulant embedding of Toeplitz matrices, which will be useful when we apply the MM procedure on the problem in (4). Circulant embedding of Toeplitz matrices is an old technique that has been successfully applied by many people in the literature. For instance, [2] and [3] discuss the ML estimation of a covariance matrix with just the Toeplitz constraint, in which the underlying Toeplitz covariance matrix was considered to be embedded in the upper-left block (of size $M \times M$) of a circulant covariance matrix of dimension $L \geq 2M-1$, and an expectation-maximization (EM) algorithm was proposed to estimate the covariance matrix. In addition to the Toeplitz structure on the covariance matrix, using the circulant embedding technique, in [18] a band structure was imposed on the covariance matrix and an EM-based algorithm with some appropriate modifications in the M stage to impose the band structure was proposed to solve the problem. Also here, we will try and use the circulant embedding for the Toeplitz matrix and later show how this technique can be exploited easily to handle the low-rank constraint.

As we know, any circulant matrix is characterized completely by its first row (or, its first column) and it can be diagonalized by a discrete Fourier transform (DFT) matrix with the entries in the diagonal matrix nothing but the DFT of the entries of the first row of the circulant matrix. Thus, a Hermitian positive semidefinite Toeplitz covariance matrix[1] $\mathbf{T}$ can be written as follows [19]:

$$\mathbf{T} = \tilde{\mathbf{F}}\mathbf{P}\tilde{\mathbf{F}}^H; \quad [\mathbf{P}]_{k,k} = p_k \geq 0 \tag{7}$$

where

$$\tilde{\mathbf{F}} = [\mathbf{I}\ \mathbf{0}]\,\mathbf{F}, [\mathbf{F}]_{k,l} = \tfrac{1}{\sqrt{L}} e^{\frac{i2\pi(k-1)(l-1)}{L}}, \tag{8}$$

$\mathbf{I}$ and $\mathbf{0}$ denote the identity matrix of size $M \times M$ and all zero matrix of size $M \times L-M$, respectively. The quantities $\{p_k\}$ and $\{r_k\}$ (the elements of $\mathbf{T}$) in (7) are related via

$$\mathbf{p} \stackrel{\Delta}{=} \mathbf{F}\mathbf{r} \tag{9}$$

where

$$\begin{aligned} \mathbf{p} &= [p_1, \ldots, p_L]^T \\ \mathbf{r} &= [r_0, \ldots, r_{M-1}, r_{M-1}^*, r_{M-2}^*, \ldots, r_1^*]^T. \end{aligned} \tag{10}$$

[1]Here we would like to mention that the set of positive semidefinite Toeplitz matrices that can be represented via a positive semidefinite circulant matrices does not include the entire set of positive semidefinite Toeplitz matrices. Please see [19] for some counter examples. So, the representation in (7) is valid only for some subset of Toeplitz matrices; however, the subset can be made large enough by increasing L.

Here the length L is chosen to be equal to $2M-1$. Thus, the Toeplitz matrix $\mathbf{T}$ can be completely parametrized by $\{p_k\}$s. In fact, the number of unknowns in this new parameterization remains the same as in the original Toeplitz matrix. However, we would like to mention that the parameterization in (7) has some restrictions, which can be explained through the following lemma.

Lemma 2.1: Caratheodory parametrization of a Toeplitz covariance matrix: Any rank r positive semidefinite Toeplitz matrix $\mathbf{T}$ can be represented as

$$\mathbf{T} = \mathbf{A}\mathbf{P}\mathbf{A}^H \tag{11}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \ldots & e^{j\omega_r} \\ e^{j2\omega_1} & e^{j2\omega_2} & \ldots & e^{j2\omega_r} \\ \vdots & \cdots & \cdots & \vdots \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_1} & \ldots & e^{j(M-1)\omega_r} \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \alpha_1^2 & 0 & \cdots & 0 \\ 0 & \alpha_2^2 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \alpha_r^2 \end{bmatrix} \tag{12}$$

and $\{\omega_k\}$ and $\{\alpha_k\}$ denote some relevant frequencies and their corresponding amplitudes.

Proof: See [Sec. 4.9.2, p. 172] ■

These frequencies $\{\omega_k\}$ are not constrained to lie on the Fourier grid, thus the parameterization defined via (7), which strictly requires the frequencies to lie on the Fourier grid, is slightly approximate. Or, in other words, for some Toeplitz covariance matrices, the parameterization in (7) will not be accurate. Nonetheless, for those Toeplitz matrices, the accuracy of the parameterization in (7) can be improved by increasing the dimension L.

C. MELT

Without loss in generality, hereafter, we will use $\mathbf{F}$ instead of $\tilde{\mathbf{F}}$. Using the parameterization in (7), let us consider the following problem:

$$\begin{aligned} \underset{\{p_k\}}{\text{minimize}} \quad & f(\mathbf{p}) \stackrel{\Delta}{=} \mathrm{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right) + \log|\mathbf{R}| \\ \text{subject to} \quad & \mathbf{R} = \mathbf{F}\left(\mathbf{P} + \sigma\mathbf{I}\right)\mathbf{F}^H \\ & [\mathbf{P}]_{k,k} = p_k \geq 0,\ \|\mathbf{p}\|_0 \leq \rho. \end{aligned} \tag{13}$$

Lemma 2.2: The optimization problems defined in (3) and (13) are equivalent.

Proof: Using the parameterization in (7), the constraint in (2) can be expressed as $\mathbf{T} = \mathbf{F}\mathbf{P}\mathbf{F}^H$, and the rank constraint on $\mathbf{T}$ can be expressed as $\|\mathbf{p}\|_0 \leq \rho$, where $\|\mathbf{p}\|_0$ denotes the number of nonzero elements in $\mathbf{p}$, and the rank constraint on $\mathbf{T}$ can be expressed as $\|\mathbf{p}\|_0 \leq \rho$. We have used the semiunitary property of $\mathbf{F}$, i.e., $\mathbf{F}\mathbf{F}^H = \mathbf{I}$ in expressing $\mathbf{R} = \mathbf{F}(\mathbf{P} + \sigma\mathbf{I})\mathbf{F}^H$. ■

Hereafter, we will work only with the problem in (13) for which we will try and apply the MM procedure. Before giving the main theorem, in the following we will introduce some lemmas that will be needed in proving the main theorem.

Lemma 2.3: Given a estimate of $\mathbf{R}$ denoted by $\mathbf{R}_t$, function $\log\det(\mathbf{R})$ can be upper-bounded as

$$\log\det|\mathbf{R}| \leq \log|\mathbf{R}_t| + \mathrm{Tr}\left(\mathbf{R}_t^{-1}(\mathbf{R}-\mathbf{R}_t)\right) \tag{14}$$

with equality achieved at $\mathbf{R}=\mathbf{R}_t$.

Proof: As the function $\log|\mathbf{R}|$ is concave in $\mathbf{R}$ [20], given $\mathbf{R}_t$, a tighter upper-bound is given by its tangent hyperplane passing through $\mathbf{R}_t$. ■

Lemma 2.4: Assuming $\mathbf{P}\succ\mathbf{0}$ and given a estimate of $\mathbf{P}$ denoted by $\mathbf{P}_t$ (and the corresponding $\mathbf{R}_t=\mathbf{A}\mathbf{P}_t\mathbf{A}^H$), matrix $(\mathbf{A}\mathbf{P}\mathbf{A}^H)^{-1}$ can be upper-bounded as

$$\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1} \succeq \left(\mathbf{A}\mathbf{P}\mathbf{A}^H\right)^{-1}. \tag{15}$$

Equality is achieved at $\mathbf{P}=\mathbf{P}_t$.

Proof: See [13, Proposition 4] for the detailed proof. ■

Theorem 2.5: Given a current estimate $\mathbf{p}_t$ and its corresponding $\mathbf{R}_t$, the objective function $f(\mathbf{p})$ in (13) can be majorized by the surrogate function $g(\mathbf{p}|\mathbf{p}_t)=\mathbf{w}_t^H\mathbf{p}+\mathbf{d}_t^H(\mathbf{p}+\sigma\mathbf{1})^{-1}+\text{const.}$, where $\mathbf{w}_t=\mathrm{diag}(\mathbf{F}^H\mathbf{R}_t^{-1}\mathbf{F})$ and $\mathbf{d}_t=\mathrm{diag}((\mathbf{P}_t+\sigma\mathbf{I})\mathbf{F}^H\mathbf{R}_t^{-1}\hat{\mathbf{R}}\mathbf{R}_t^{-1}\mathbf{F}(\mathbf{P}_t+\sigma\mathbf{I}))$.

Proof:

$$\begin{aligned}
f(\mathbf{p}) &= \log|\mathbf{R}| + \mathrm{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right)\\
&\overset{\text{Lemma2.3)}}{\leq} \mathrm{Tr}\left(\mathbf{R}_t^{-1}\mathbf{R}\right) + \mathrm{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right) + \log|\mathbf{R}_t| - M\\
&\overset{\text{Lemma2.4)}}{\leq} \mathrm{Tr}\left(\mathbf{R}_t^{-1}\mathbf{R}\right) + \mathrm{Tr}(\hat{\mathbf{R}}\mathbf{R}_t^{-1}\mathbf{F}(\mathbf{P}_t+\sigma\mathbf{I})(\mathbf{P}+\sigma\mathbf{I})^{-1}\\
&\qquad\times(\mathbf{P}_t+\sigma\mathbf{I})\mathbf{F}^H\mathbf{R}_t^{-1}) + \log|\mathbf{R}_t| - M\\
&= \mathbf{w}_t^H\mathbf{p} + \mathbf{d}_t^H(\mathbf{p}+\sigma\mathbf{1})^{-1} + \text{const.}\\
&= g(\mathbf{p}|\mathbf{p}_t).
\end{aligned}$$

■

The resulting problem is given by

$$\begin{aligned}
&\underset{\{p_k\}}{\text{minimize}} \quad \mathbf{w}_t^H\mathbf{p} + \mathbf{d}_t^H(\mathbf{p}+\sigma\mathbf{1})^{-1}\\
&\text{subject to} \quad p_k\geq 0,\ \|\mathbf{p}\|_0\leq\rho.
\end{aligned} \tag{16}$$

The surrogate function in (16) is separable in the p_js, therefore can be minimized separately. For any index j, $g(p_j|\mathbf{p}_t)=w_jp_j+\frac{d_j}{p_j+\sigma}$, and the minimum (and the corresponding minimizer) would be

$$\begin{aligned}
&g^\star(p_j|\mathbf{p}_t)\\
&= \begin{cases} \dfrac{d_j}{\sigma}\ (p_j^\star=0) & \text{if } d_j\leq\sigma w_j\\ 2\sqrt{d_jw_j}-\sigma w_j\left(p_j^\star=\sqrt{\dfrac{d_j}{w_j}}-\sigma\right) & \text{if } d_j>\sigma w_j. \end{cases}
\end{aligned} \tag{17}$$

Obviously for the entries $j: d_j\leq\sigma w_j$, we just set p_j to be zero. As for the entries with $d_j>\sigma w_j$, we compute the difference

$$e_j = \frac{d_j}{\sigma} - \left(2\sqrt{d_jw_j}-\sigma w_j\right) \tag{18}$$

which is the reduction of the objective value by allowing p_j to be nonzero. Naturally, we will allow the entries with the ρ largest e_js to be nonzero.

The steps of MELT are summarized in the following algorithm:

MELT Algorithm

Require: Data samples $\{\mathbf{y}_n\}$, ρ and σ
Pre compute: $\hat{\mathbf{R}}\overset{\Delta}{=}\frac{1}{N}\sum_{k=1}^{N}\mathbf{y}_k\mathbf{y}_k^H$
Set $t=0$, initialize $\mathbf{p}_0$
Repeat:
1) Compute $\mathbf{R}_t^{-1}$
2) Compute $\mathbf{w}_t=\mathrm{diag}(\mathbf{F}^H\mathbf{R}_t^{-1}\mathbf{F})$
and $\mathbf{d}_t=\mathrm{diag}((\mathbf{P}_t+\sigma\mathbf{I})\mathbf{F}^H\mathbf{R}_t^{-1}\hat{\mathbf{R}}\mathbf{R}_t^{-1}\mathbf{F}(\mathbf{P}_t+\sigma\mathbf{I}))$
3) Compute $\mathbf{p}_t$ according to eqs (17) and (18)
$t\leftarrow t+1$
until the convergence criterion : $\|\mathbf{p}_t-\mathbf{p}_{t-1}\|_2<10^{-5}$ is met.

Several remarks are in order:

Remarks

1) The initial point for MELT should be carefully chosen, one choice could be $\mathbf{p}_0=$ Choose the r maximum values of $\mathrm{diag}(\mathbf{F}^H\hat{\mathbf{R}}\mathbf{F})$, which is nothing but the periodogram of the data samples [6].
2) The sequence of estimates $\{\mathbf{p}_t\}$ generated by MELT would monotonically decrease the objective function in (13) as mentioned in (6). Moreover, the limit point of the sequence $\mathbf{p}^*=\lim_{t\leftarrow\infty}\mathbf{p}_t$ can be shown to be a local minimizer of the problem (13).
3) The main computational complexity of MELT is in computing the inverse of $\mathbf{R}_t$ at every iteration. As $\mathbf{R}_t$ is Toeplitz, its inverse can be efficiently computed using the Gohberg–Semencul formula [6], which has a computational complexity of $O(M\log M)$. Additionally, computing the quantities $\mathbf{w}_t$ and $\mathbf{d}_t$ require evaluating the matrix–vector multiplications involving the columns of $\mathbf{F}$ (or, $\mathbf{F}^H$), which in turn can efficiently be implemented via FFTs. Thus, the computational complexity of MELT would be $O(M\log M)$.
4) Sometimes the true underlying Toeplitz covariance matrix may not be representable via the parameterization in (7) as discussed in Section; this can happen for instance when the frequencies of the Caratheodory parameterization of the underlying true Toeplitz matrix do not lie on the Fourier grid frequencies $\{\frac{2\pi k}{L}\}_{k=0}^{L-1}$. However, any off-grid frequency can be closely represented by a linear combination of its adjacent Fourier grid frequencies. So, to counter the off-grid issue we need to increase the sparsity (i.e., $\|\mathbf{p}\|_0$) by one for every off-grid frequency. In the scenarios where the matrix rank ρ is unknown, order selection rules such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) can be used to choose ρ.
5) Since we are solving an ML estimation problem, the natural question here would be can we derive Cramer-Rao bound here? However, handling a low-rank constraint (or, the equivalent l_0 norm constraint) in the Cramer–Rao bound derivation is tricky and not clear to us. Moreover, it is not in the scope of this letter. Instead, in the following numerical section, we will use the mean square error as a performance metric.

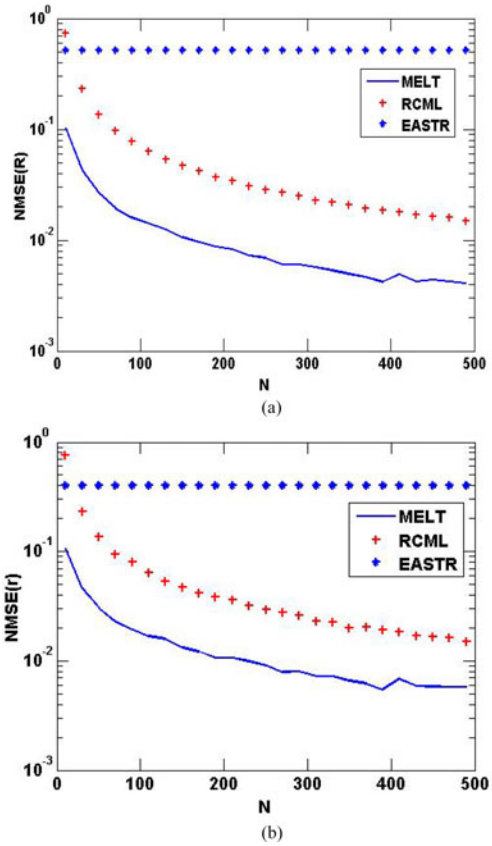

Fig. 1. NMSE versus N for experiment 1 (a) NMSE for the entire $\mathbf{R}$ and (b) NMSE of $\mathbf{r}$.

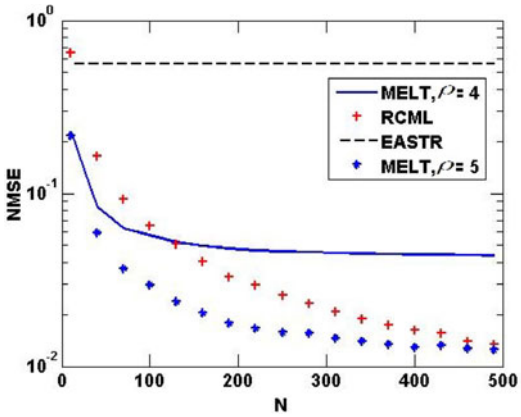

Fig. 2. NMSE of $\mathbf{r}$ versus N for experiment 2.

III. Numerical Simulations

In this section, we demonstrate the numerical performance of the algorithm developed in the last section. In the simulations, we generated the data samples according to the model

$$\mathbf{y}_k = \mathbf{R}^{1/2}\boldsymbol{n}_k \quad k = 1, \ldots, N \tag{19}$$

with $M = 10$ and $\{\boldsymbol{n}_k\}$s drawn randomly from complex Gaussian distribution with zero mean and covariance matrix $\mathbf{I}$. We generated $\mathbf{R}$ as in (2) with the rank r and σ^2 chosen accordingly. Some of the competing methods we chose are as follows:

1) Rank-constrained maximum likelihood estimator (RCML), proposed originally in [14], where MLE of the covariance matrix is computed only by imposing the low-rank structure neglecting the Toeplitz structure.
2) Efficient approximation of structured Toeplitz covariance under a rank constraint (EASTR), proposed recently in [15], in which the eigenvectors of the Toeplitz matrix are chosen to be the same as the eigenvectors of the sample covariance matrix, and in the estimation of the eigenvalues, they try and impose the Toeplitz constraints.

We have chosen the above-mentioned two methods as they were shown to outperform most of the methods available in the literature [15]. As mentioned in the previous section, the performance metric chosen here for comparison would be the normalized mean square error (NMSE), which was computed as follows:

$$\mathrm{NMSE}(\mathbf{r}) = \boldsymbol{E}\left(\frac{\|\mathbf{r}-\mathbf{r}_{\mathrm{ML}}\|^2}{\|\mathbf{r}\|^2}\right) \tag{20}$$

where $\mathbf{r}$ and $\mathbf{r}_{\mathrm{ML}}$ denote the first row of the true $\mathbf{R}$ and $\mathbf{R}_{\mathrm{ML}}$, respectively, and $\boldsymbol{E}$ denotes the expectation operator. The number of Monte Carlo simulations used in the NMSE calculations was 500. The NMSE of the entire covariance matrix ($\mathbf{R}$) is defined similar to (20). In the first experiment, we constructed a low-rank Toeplitz matrix of dimension $M = 10$ by choosing $\rho = 4$ columns of a DFT as in (8), the corresponding frequencies and the amplitudes are $[0.6283, 2.5133, 3.1416, 5.0265]$ rad/s and $[10, 4, 3, 1]$, respectively. Fig. 1(a) shows the NMSE plots of $\mathbf{R}$ and $\mathbf{r}$ versus N for the three different methods. As can be seen from the figure, MELT performs much better than RCML and EASTR. EASTR performs the worst and in fact, it is heavily biased; although, RCML does not yield a Toeplitz matrix estimate, its performance is much better than EASTR. In experiment 2, we intentionally constructed the underlying true Toeplitz matrix such that one of its underlying frequencies does not lie on the Fourier grid. The experimental settings here are the same as in experiment 1 except that the frequency at 2.5133 rad/s was replaced by 2.1991 rad/s. Fig. 2 shows the simulation results for experiment 2, and as can be seen from the figure, MELT with the choice of $\rho = 4$ is biased for large values of N, which is due to the presence of one off-grid frequency; but nonetheless, MELT with the choice of $\rho = 5$ tackles the problem and performs better than the other methods.

IV. Conclusion and Future Work

In this letter, we developed a computationally efficient low-cost algorithm named MELT to solve the ML covariance estimation problem under the given low-rank Toeplitz constraint. MELT is based on the MM technique and has monotonic convergence to a local minimum. Finally, we showed numerically that MELT performs better than the techniques currently available in the literature. The following list of items are potential problems related to the low-rank Toeplitz covariance matrix estimation problem that we would like to address as future works:

1) As mentioned earlier that any off-grid frequency can be closely represented by a linear combination of its adjacent Fourier grid frequencies, and to counter the effect of off-grid frequencies one need to accordingly increase the sparsity of $\mathbf{p}$. Moreover, in the scenarios where the rank ρ is unknown, order selection rules such as AIC or BIC should be used to choose ρ. We would like address this issue in a more detailed manner in our future work.
2) In this letter, we assumed the noise variance (σ) and the matrix rank (ρ) to be known; however, in the practical scenarios both of them are not known. As a part of future work we would like to analyze the impact of mismatches in the rank and the noise level on our method.
3) As a part of future work, we would like to derive the Cramer–Rao bound for the low-rank Toeplitz covariance matrix estimation problem.

REFERENCES

[1] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, Sep. 1982.

[2] 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.

[3] 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.

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

[5] C. -I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, vol. 1. New York, NY, USA: Springer Science & Business Media, 2003.

[6] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson/Prentice-Hall, 2005.

[7] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Signal Process.*, vol. 39, pp. 2194–2198, Oct. 1991.

[8] 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.

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

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

[11] 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.

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

[13] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, Jul. 2015.

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

[15] 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.

[16] M. I. Skolnik, "Radar handbook," 1970.

[17] D. R. Hunter and K. Lange, "A tutorial on mm algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, 2004.

[18] L. P. Christensen, "An em-algorithm for band-Toeplitz covariance matrix estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3 IEEE, 2007, pp. III–1021.

[19] 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.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2003.