ISSN: 3066-1730

Research Article

Journal of Electrical and Computational Innovations

Properties of the Toeplitz Covariance Matrix Numerical Likelihood Ratio Maximization: Convex-Like Optimization Behavior

Yuri Abramovich*, **Victor Abramovich** and **Tanit Pongsiri**

WR Systems, Ltd., Fairfax, USA

***Corresponding Author**
Yuri Abramovich, WR Systems, Ltd., Fairfax, USA.

Submitted: 2026, Apr 06; **Accepted:** 2026, May 13; **Published:** 2026, May 28

Citation: Abramovich, Y., Abramovich, V., Pongsiri, T. (2026). Properties of the Toeplitz Covariance Matrix Numerical Likelihood Ratio Maximization: Convex-Like Optimization Behavior. *J Electr Comput Innov, 3*(2), 01-20.

Abstract

In this paper, we continue to investigate the well-known problem of the numerical likelihood maximization of the positive definite Toeplitz covariance matrix of complex Gaussian data. In our recent papers, we demonstrated that direct LR maximization, using Vandermonde parameterization, applied to initial Toeplitz matrices distant from the true Toeplitz covariance matrix, predominantly yields an inappropriate solution with negative eigenvalues [1,2]. Yet in all cases where the MATLAB fmincon routine converges to a positive definite Toeplitz matrix, the process converges to the same solution as if initiated by a true Toeplitz covariance matrix, irrespective of the initialization one. We also demonstrated that the optimized likelihood ratio (LR) exceeded that of the true Toeplitz covariance matrix and that, by starting fmincon iterations from the true covariance matrix $\mathbf{T}_N$*, we converged to the same solution as when starting from practical initializations. These two properties of the globally optimal solutions are crucial for practical Toeplitz matrix estimation. In this paper, we continue to explore these convex-like properties of the MATLAB fmincon LR maximization. In particular, to avoid generating large numbers of non-positive-definite solutions, we propose using the Carathéodory Toeplitz matrix representation in the initial step of the LR maximization with the fmincon routines. We demonstrate that this optimization exceeds the LR value of the true Toeplitz covariance matrix. Still, in most cases, it does not exceed the LR maximum obtained by optimizing the covariance lags of the optimized Toeplitz matrix. Therefore, the second stage of LR optimization, which we performed in covariance lags, drove the optimized LR to the maximum. Yet, we demonstrate that this second-stage LR maximization improves the LR but, in fact, degrades the proximity of the optimized solution to the true Toeplitz covariance matrix used instead of the sample matrix in the LR. We demonstrate that the proposed LR maximization allows for the essential reduction of the minimal input SNR or inter-source separation of the properly estimated sources' DOAs, compared to the classical MUSIC algorithm, applied to the sample matrix* $\hat{\mathbf{R}}_N$.

1. Introduction

The Maximum Likelihood (ML) estimation of the Toeplitz covariance matrix, given a set of T independent identically distributed (i.i.d.) complex Gaussian N-variate vectors, is one of the classical signal processing problems that has been under intensive investigation since at least the '80s, and yet it continues to attract attention nowadays [3-19]. One of the main problems is that the nature of this optimization problem remains unclear. The type of this optimization problem was not specified in the seminal paper by J.P. Burg, D.G. Luenberger, and D.L. Wenger in 1982 or in the papers by D.B. Fuhrmann, but, written in 2024, the problem was described as non-convex [7,20]. Therefore, the nature of this important optimization problem remained unclear despite a large number of efficient solutions proposed over the decades [3-17]. Correspondingly, the properties of the globally optimal maximum likelihood Toeplitz covariance matrix estimates have not been explored because of the expectation of non-convexity, with potentially many solutions and the inability to specify the globally optimal ones.

In our recent papers, we addressed this problem by modifying the maximized likelihood function into a likelihood ratio with a probability density function (pdf) for the true covariance matrix, which does not depend on the true matrix and is specified by a priori known parameters [1,2]. The application of the standard MATLAB fmincon optimization routine revealed that while the overwhelming majority of solutions were non-positive definite and therefore easily rejected, a small number of positive definite solutions possessed a very important property. Specifically, irrespective of the initial solution of this iterative optimization routine that ended up in a positive definite (p.d.) solution, this solution was the same. Moreover, the maximized likelihood ratio exceeded that of the true Toeplitz covariance matrix. In fact, all acceptable (i.e., positive definite) solutions of this optimization problem possessed these most important properties of the globally optimum solution.

In this paper, we continue to investigate this phenomenon. Specifically, by applying the Carathéodory description of the optimized Toeplitz matrix, we avoided non-positive definite solutions but did not achieve the LR maximum obtained by direct covariance lag optimization. Therefore, after fmincon convergence on the Carathéodory parameters, we switched to direct optimization of the covariance lags. The overwhelming majority of the solutions possessed the properties of the LRoptimum p.d. Toeplitz matrix. Yet, the "distance" between this solution with its improved LR and true covariance matrix $\mathbf{T}_N$ was proven to be greater than the p.d. solutions in Carathéodory parameterization. This "distance" was measured by the likelihood ratio with the sample covariance matrix $\hat{\mathbf{R}}_N$ replaced by the true Toeplitz covariance matrix $\mathbf{T}_N$.

Finally, we demonstrate that the application of this Toeplitz covariance matrix estimation to the problem of direction-of-arrival (DOA) estimation in a uniform linear antenna (ULA) array, allows it to significantly extend into the region of a smaller input SNR and/or smaller inter-source separation, the efficient resolution and DOA estimation of these sources, compared to MUSIC, applied to the sample covariance matrix $\hat{\mathbf{R}}_N$.

By focusing on these new results, the paper concentrates on the properties of the solutions provided by the MATLAB fmincon routine and the gains in the classical DOA estimation problem these properties offer. Specifically, we demonstrate that the p.d. solutions, delivered by the MATLAB fmincon routine, possess the following properties:

- the optimized LR value exceeds the LR value produced by the true Toeplitz covariance matrix,
- the converged positive definite Toeplitz matrices are the same, irrespective of the initial solution used. These tested initial solutions included the very far from the true covariance matrix initiations with LR < 10^{-20}, and the true covariance matrix $\mathbf{T}_N$.

These two properties belong to the truly globally optimal solution, but in our case, they describe only a subset of all possible solutions, including non-p.d. solutions. Since the non-p.d. solutions could be easily identified, the remaining p.d. solutions possess these important properties of the globally optimal ones.

The paper is focused on the two following goals:

- to describe a new technique for ML Toeplitz matrix estimation, which improves DOA estimation performance
- to investigate the properties of the quasi-global solution of the ML Toeplitz covariance matrix estimation.

Our main contributions are summarized as follows.

i We demonstrate that while the MATLAB fmincon routine optimizes the covariance lags of the Toeplitz matrix, the vast majority of solutions are not positive definite.

ii All positive definite solutions provided by the MATLAB fmincon routine are practically the same, irrespective of the p.d. Toeplitz matrix used for initialization. The same solutions are obtained both with initializations that differ substantially from the true matrix (with LR ≤ 10^{-20}) and with initializations utilizing the true Toeplitz covariance matrix $\mathbf{T}_N$. The optimized LR values for these solutions exceed those produced by the true covariance Toeplitz matrix. The independence of the result from the initialization matrix, including the true covariance matrix, together with exceeding the LR value of the true covariance matrix, are the most important properties of the globally optimum solution. Yet, the numerical methods highlight these properties but do not provide analytical proof of global optimality.

iii For avoiding the non-p.d. solutions generated by the MATLAB fmincon routine, we proposed starting LR maximization using the Carathéodory p.d. Toeplitz matrix representation, followed by LR maximization with further covariance lag updates (i.e., in Vandermonde parameterization).

iv We demonstrated that this second stage of the LR optimization of the covariance lags increases the LR but decreases the proximity to the true covariance matrix that replaces the sample matrix in the likelihood ratio.

Correspondingly, in Sec 2, we provide the analytical description of the optimization problem and its possible parameterizations. In Sec 3, we demonstrate the properties of the optimized solutions for data with a Toeplitz matrix and describe the clutter returns in the HF OTHR. In Sec 4, we demonstrate that applying the proposed Toeplitz matrix estimation technique yields significant gains in source DOA estimation for sources impinging upon a uniform linear array. In Sec 5, we conclude our paper.

2. Maximum Likelihood Covariance Toeplitz Matrix Optimization Algorithm

For the T i.i.d. Gaussian N-variate training data $\mathbf{X}_t \sim \mathbb{C}\mathcal{N}(0, \mathbf{T}_N), t = 1, \dots, T$, the likelihood function is [21]:

$$\mathrm{LF}[\mathbb{X}_T|\mathbf{T}_N] = \frac{\exp\left[-T\,\mathrm{Tr}\left(\hat{\mathbf{R}}_T\mathbf{T}_N^{-1}\right)\right]}{[\det \mathbf{T}_N]^T}, \quad (1)$$

where

$$\mathbb{X}_T = [\mathbf{X}_1, \dots, \mathbf{X}_T], \quad \hat{\mathbf{R}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t^{\mathrm{H}} = \frac{1}{T}\mathbb{X}_T\mathbb{X}_T^{\mathrm{H}}. \quad (2)$$

If we multiply this likelihood function (2) by the value c that does not depend on the optimized Toeplitz covariance matrix $\mathbf{T}_N$,

$$c = \det \widehat{\mathbf{R}}_T{}^T \exp(NT), \tag{3}$$

we get the well-known likelihood ratio [8]:

$$\mathrm{LR}\left[\mathbb{X}_T|\widehat{\mathbf{T}}_N\right] = \frac{\exp N \det\left(\widehat{\mathbf{R}}_N \widehat{\mathbf{T}}_N^{-1}\right)}{\exp \mathrm{Tr}\left(\widehat{\mathbf{R}}_N \widehat{\mathbf{T}}_N^{-1}\right)}. \tag{4}$$

Note that for the true covariance matrix $\hat{\mathrm{T}}_N = \mathbf{T}_N$ (true covariance matrix), the LR in (4) does not depend on $\mathbf{T}_N$ and is specified by the a priori known parameters (N, T):

$$\mathbf{X}_t = \mathbf{T}_N^{\frac{1}{2}} \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_N). \tag{5}$$

And with respect to (5), we have

$$\det\left[\widehat{\mathbf{R}}_N \mathbf{T}_N^{-1}\right] \equiv \det\left[\mathbf{T}_N^{-\frac{1}{2}} \widehat{\mathbf{R}}_N \mathbf{T}_N^{-\frac{1}{2}}\right] \equiv \det\left[\frac{1}{T}\sum_{t=1}^{N} \boldsymbol{\xi}_t \boldsymbol{\xi}_t^{\mathrm{H}}\right], \tag{6}$$

$$\mathrm{Tr}\left[\widehat{\mathbf{R}}_N \mathbf{T}_N^{-1}\right] = \mathrm{Tr}\left[\mathbf{T}_N^{-\frac{1}{2}} \widehat{\mathbf{R}}_N \mathbf{T}_N^{-\frac{1}{2}}\right] = \mathrm{Tr}\left[\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\xi}_t \boldsymbol{\xi}_t^{\mathrm{H}}\right].$$

Therefore, for the LR, we get

$$\mathrm{LR}[\mathbb{X}_T|\mathbf{T}_N] = \frac{\exp N \det\left(\frac{1}{T}\mathfrak{J}_T \mathfrak{J}_T^{\mathrm{H}}\right)}{\exp \mathrm{Tr}\left(\frac{1}{T}\mathfrak{J}_T \mathfrak{J}_T^{\mathrm{H}}\right)} \leq 1, \quad \mathfrak{J}_T = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_T] \quad . \tag{7}$$

The conditional (on $\mathbf{T}_N$) total power estimate $\hat{\sigma}_{\mathrm{ML}}^2$ is

$$\hat{\sigma}_{\mathrm{ML}}^2 = \arg_{\sigma^2}\left[\frac{\partial \mathrm{LF}[\mathbb{X}_T|\sigma^2]}{\partial \sigma^2}\right] = 0, \tag{8}$$

which leads to

$$\frac{\partial}{\partial \sigma^2}\left[-T\,\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right) - NT \log \sigma^2\right] = 0, \tag{9}$$

which in turn leads to

$$T\,\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right) - \frac{NT}{\sigma^2} = 0, \tag{10}$$

and to the ML power estimate:

$$\hat{\sigma}_{\mathrm{ML}}{}^{-2} = \frac{1}{N}\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right). \tag{11}$$

By substituting (11) into (4), we get:

$$\mathrm{LR}\left[\mathbb{X}_T|\widehat{\mathbf{T}}_N^{(0)}\right] = \frac{\exp N \det\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)}{\frac{1}{N}\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)^N \exp\left(N\frac{\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)}{\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)}\right)} = \frac{\det\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)}{\left[\frac{1}{N}\mathrm{Tr}\left(\widehat{\mathbf{R}}_N\left(\widehat{\mathbf{T}}_N^{(0)}\right)^{-1}\right)\right]^N}. \tag{12}$$

Note that the derived "sphericity test" (12) does not depend on the Toeplitz matrix's total power σ_T^2 and therefore, instead of $\widehat{\mathbf{T}}_N^{(0)}$ in (12), one can use $\widehat{\mathbf{T}}_N$ with its power σ_T^2. The maximum likelihood estimate $\widehat{\mathbf{T}}_N^{\mathrm{ML}}$ should generate a "sphericity test" value that exceeds the LR value generated by the true covariance matrix $\mathbf{T}_N$. The accurate pdf of the sphericity test for the true covariance matrix $\mathbf{T}_N$ of the complex data $\mathbf{X}_T$ and the oversampled training conditions ($T > N$) are introduced in [17]. We also analyzed the "undersampled" condition, in which the number of training vectors T is smaller than the Toeplitz matrix's dimension N; the corresponding likelihood ratio is introduced in [2,22].

For the numerical LR maximization, we adopted the MATLAB fmincon routine, which is the nonlinear programming solver specified in [23]:

$$\begin{gathered}\text{Find } \min_{\mathbf{T}_N} -\mathrm{LR}\ \left[\widehat{\mathbf{R}}_N\right] \text{ such that} \\ -\min \mathrm{eig}\, \mathbf{T}_N < 0,\end{gathered} \tag{13}$$

where $\mathbf{T}_N$ is a Hermitian Toeplitz matrix. Obviously, for the numerical Toeplitz matrix optimization, this matrix needs to be parameterized. One of the possible Toeplitz matrix parameterizations is the set of one positive (main diagonal) and (N - 1) complex-valued covariance lags of this Hermitian Toeplitz matrix. Since this parameterization may describe the non-positive definite Toeplitz matrix, an additional constraint (see (13)) on the positive definiteness of this matrix is required.

Unfortunately, the MATLAB fmincon routine does not always retain this introduced constraint. As one of the developers of this routine noted, "there is no absolute guarantee that by starting fmincon at a feasible point, you will see it terminate at a feasible point" [23]. Since a negative eigenvalue can be easily identified, it is possible to run this software again with a different p.d. Toeplitz matrix to initiate the fmincon routine and finally obtain a positive-definite solution. Obviously, this is not a good practical approach due to the indefinite execution time, but it could be used for problem investigation. Note that the eigenspectrum is not otherwise specified in this parameterization. Correspondingly, it is not easy to maintain equal minimal eigenvalues as required for the problem of direction of arrival (DOA) estimation of $m < N - 1$ point Gaussian sources in white Gaussian noise.

Yet, for the positive definite Hermitian Toeplitz matrices, there is an alternative to Vandermonde parameterization using the

Carathéodory decomposition of the p.d. Hermitian Toeplitz matrix. Indeed, let $\mathbf{T}_N$ be a p.d. Toeplitz Hermitian matrix with the eigendecomposition:

$$\mathbf{T}_N = \sum_{j=1}^{N} \lambda_j \mathbf{U}_j \mathbf{U}_j^{\mathrm{H}}; \quad \lambda_j \geq 0, \quad \mathbf{U}_j^{\mathrm{H}} \mathbf{U}_k = \delta_{jk}. \tag{14}$$

Then this matrix may be presented as:

$$\mathbf{T}_N = \lambda_{\min} \mathbf{I}_o + \sum_{j=1}^{N-1} \left(\lambda_j - \lambda_{\min}\right) \mathbf{U}_j \mathbf{U}_j^{\mathrm{H}}. \tag{15}$$

where the rank-deficient matrix $\mathbf{T}_{N-1}$,

$$\mathbf{T}_{N-1} = \sum_{j=1}^{N-1} \left(\lambda_j - \lambda_{\min}\right) \mathbf{U}_j \mathbf{U}_j^{\mathrm{H}}, \tag{16}$$

in accordance with Theorem 02 from [19], may be presented as a matrix of $m = N - 1$ independent plane waves. This theorem provides the alternative (Carathéodory) representation of an arbitrary p.d. Toeplitz Hermitian matrix:

$$\mathbf{T}_N = \lambda_{\min} \mathbf{I}_o + \sum_{j=1}^{N-1} \sigma_j^2 \mathbf{S}(\theta_j) \mathbf{S}(\theta_j)^{\mathrm{H}}, \quad \lambda_{\min}, \sigma_j^2 > 0. \tag{17}$$

Note that the covariance matrix (17) is described by one positive parameter ($\lambda_{\min}$) and $2(N-1)$ real-valued parameters (σ_j, θ_j) for $j = 1, \ldots, N-1$, similar to an arbitrary Toeplitz Hermitian matrix specified by its covariance lags. Despite the same number of "free" parameters that describe a Toeplitz Hermitian matrix, one (Carathéodory) set always describes a set of positive definite matrices. In contrast, the "covariance lags" set may describe an arbitrary Toeplitz Hermitian matrix. Therefore, the class of Toeplitz matrices described by (17) forms the inner set with respect to the set described by both positive and non-positive definite Hermitian Toeplitz matrices.

Yet, it is not self-evident that a particular optimization algorithm, such as MATLAB fmincon, should provide the same result if the non-negative definite solutions are of interest only. Apart from yielding a non-positive-definite solution, ignoring the introduced constraint ($\lambda_{\min} > 0$), this algorithm often converges to a solution with a higher LR value than when the admissible set is limited to non-negative-definite solutions only.

One should not expect significant differences in the optimization results. Still, the optimizations without a constraint, while often producing inappropriate results, in those rare cases of p.d. convergence to the proper solutions, may exceed the maximum LR value achieved by the algorithm with this additional constraint. Let us emphasize that this should be the property of the particular optimization algorithm, while ideally, we should get the same set of solutions. For the reasons described above, we will investigate the MATLAB fmincon routine to optimize the covariance lags directly, expecting that the constraint in (13) will often be violated. It would be more practical to perform the two-stage optimization described above by first using the Carathéodory decomposition (17) and then the covariance lag optimization in (13).

3. Numerical Likelihood Ratio Maximization for Models with Different Eigenvalues of the Toeplitz Covariance Matrix

Let us analyze the case in which the number of distinct positive definite eigenvalues of the optimized Toeplitz covariance matrix is not specified. In practice, this case describes the spatial covariance matrices of sky-wave-propagated clutter returns in HF OTH radars [24]. As an example of such Toeplitz matrices, we consider the following matrix $\mathbf{T}_N$:

$$\begin{aligned}\mathbf{T}_N = {} & q_m^{-2} \mathbf{I}_N + \operatorname{sinc}(W_1) \\ & + 0.5\, \mathbf{D}[\theta_o] \operatorname{sinc}(W_2) \mathbf{D}^{\mathrm{H}}[\theta_o],\end{aligned} \tag{18}$$

where

$$\operatorname{sinc}(W) = \left[\frac{\sin \pi W\,(i-j)}{\pi W (i-j)}\right], \quad i, j = 0, \ldots, N-1, \tag{19}$$

and

$$\begin{aligned}\mathbf{D}[\theta_o] = \operatorname{diag}\Big[& 1, \exp\left(i \frac{2\pi d}{\lambda} \sin\theta_o\right), \ldots, \\ & \exp\left(i (N-1) 2\pi \frac{d}{\lambda} \sin\theta_o\right)\Big].\end{aligned} \tag{20}$$

In numerical simulations, we use the following parameters:

$$\begin{gathered} N = 17, \theta_o = 20^{\circ}, W_1 = 0.2, W_2 = 0.4, \\ q_m^{-2} = 10^{-4}, T = 85\ (T = 5N). \end{gathered} \tag{21}$$

In the oversampled ($T \geq N$) case, we use the sphericity likelihood ratio test. Let us now specify the goals of the numerical investigations.

A. LR Optimization of the Covariance Lags Mostly Results in Inappropriate (Non-Positive Definite) Solutions, But When It Converges to a Positive Definite One, It Is the Same Solution

The problem of MATLAB fmincon LR maximization by optimization of the matrix's covariance lags may be formulated as follows:

$$\text{Find } \min_{\mathbf{T}_N} -\mathrm{LR}[\mathbb{X}_T | \mathbf{T}_N] \ (\coloneqq f(x)) \tag{22}$$

subject to

Istat	Nfailures	Niterations	Min Eigenvalue	Ideal LR	Initial LR	"DoAs" LR	Tn2opt LR	Final LR
1	178	481	6.4012E-04	1.4939E-01	3.6260E-27	1.7131E-01	1.8529E-01	1.8529E-01
2	89	544	6.4040E-04	1.4939E-01	3.9225E-26	1.7180E-01	1.8529E-01	1.8529E-01
3	959	527	6.2834E-04	1.4939E-01	1.3085E-19	1.7388E-01	1.8529E-01	1.8529E-01
4	100	630	6.2830E-04	1.4939E-01	2.7265E-19	1.7388E-01	1.8529E-01	1.8529E-01
5	397	S25	6.2832E-04	1.4939E-01	5.3184E-29	1.7388E-01	1.8519E-01	1.8519E-01
6	234	375	6.2832E-04	1.4939E-01	1.9070E-23	1.7388E-01	1.8529E-01	1.8529E-01
7	209	566	6.5517E-04	1.4939E-01	8.8923E-19	1.7078E-01	1.8529E-01	1.8529E-01
8	700	483	6.2834E-04	1.4939E-01	3.0410E-19	1.7388E-01	1.8529E-01	1.8529E-01
9	2,932	498	6.5490E-04	1.4939E-01	1.9461E-25	1.7076E-01	1.8529E-01	1.8529E-01
10	288	347	5.6374E-04	1.4939E-01	9.9890E-28	1.7894E-01	1.8529E-01	1.8529E-01

Table 1: Parameters of the LR Optimization in 10 Trials of the "Natural" Parameterization (*Different Initial Matrices for the Same Sample Matrix Are Used*)

$$\mathbf{T}_N = x_o\mathbf{I}_o + \sum_{l=1}^{N-1} x_l \begin{bmatrix} 0 & \cdots & 0 & \overbrace{1}^{l} & 0 & \cdots & 0 \\ \vdots & & & & 1 & & \\ 0 & & & & & & \\ \overbrace{1}^{l} & & & & & \ddots & \\ 0 & 1 & & & & & \overbrace{1}^{l} \\ & & \ddots & & & & \\ 0 & \cdots & 0 & \overbrace{1}^{l} & 0 & \cdots & 0 \end{bmatrix} \tag{23}$$

$$+ \sum_{l=1}^{N-1} jx_{N+l} \begin{bmatrix} 0 & \cdots & 0 & \overbrace{1}^{l} & 0 & \cdots & 0 \\ \vdots & & & & 1 & & \\ 0 & & & & & & \\ \overbrace{-1}^{l} & & & & & \ddots & \\ 0 & -1 & & & & & \overbrace{1}^{l} \\ & & \ddots & & & & \\ 0 & \cdots & 0 & \overbrace{-1}^{l} & 0 & \cdots & 0 \end{bmatrix}$$

$$x_o > 0, \quad \lambda_{\min}[\mathbf{T}_N] > 0. \tag{24}$$

The expected behavior of this optimization routine, with a high probability of convergence to non-p.d. solutions, forces us to prepare a large class of p.d. Toeplitz Hermitian matrices that can be used for the initialization of this fmincon optimization routine. Obviously, a covariance matrix of $N-1$ point sources with random-like powers and directions of arrival can be used. An elegant procedure that generates a random-like symmetric (i.e., real-valued) p.d. Hermitian Toeplitz matrix was introduced and used below to initialize the LR maximization routine [25]. Obviously, the redundancy-averaged sample Hermitian matrix with additional diagonal loading that "kills" all negative eigenvalues was also used to generate the initial Toeplitz matrices.

Next, the "failed" fmincon solutions with negative eigenvalues were used after the appropriate diagonal loading to generate the set of appropriate p.d. initial solutions. Finally, the true covariance matrix $\mathbf{T}_N$ was also used to initiate the routine. Obviously, such an extremely large class of initial solutions was used to obtain all possible fmincon optimization results. From a practical point of view, we should be looking for the initialization that provides the fastest convergence. The results of 10 successful trials that converged to a p.d. solution is presented in Table 1. The sample covariance matrix $\hat{\mathbf{R}}_N$ was the same for all trials, with only the initial matrices changed.

One can see that the number of unsuccessful trials, before the trial with the positive definite Toeplitz matrix was generated, ranged from 89 to 2932. The likelihood of a successful initial Toeplitz matrix producing the final p.d. Toeplitz matrix ranges from $2.3067{\cdot}10^{-12}$ to $1.9070{\cdot}10^{-30}$, which means that even the "successful" initializations were extremely far from the true Toeplitz matrix. The "expected" likelihood of the true covariance used is obviously a single number for this single sample matrix, equal to LR = $1.4939{\cdot}10^{-1}$. The fmincon optimization that used the true covariance matrix $\mathbf{T}_N$ as the initial matrix provided for this true Toeplitz matrix, with LR = $1.4939{\cdot}10^{-1}$, while the optimization result gave LR = $1.8529{\cdot}10^{-1}$.

The same result was achieved in all optimizations that converged to a p.d. Toeplitz matrix (where one result was 1.8519·10−1). As stated above, in all ten successful trials, which required 347 to 630 fmincon iterations, the results were the same: LR$\left[\hat{\mathbf{T}}_N^{\mathrm{ML}}\right]$ = $1.8529{\cdot}10^{-1}$. As demonstrated above, it required up to several thousand different random-like initial p.d. Toeplitz matrices to reach the initialization that resulted in this optimum solution. The trial was dismissed only if it converged to a solution with a negative minimum eigenvalue.

Despite the very small variations in the optimized results, potentially due to finite numerical accuracy, the achieved LR = = $1.8529{\cdot}10^{-1}$ may be treated as the global LR maximum.

Simulations were naturally conducted for 1,000 trials using the various initializations discussed above, searching for results markedly different from those reported above. Yet, in all trials that converged to a positive definite Toeplitz matrix, the optimized LR value was the same.

B. LR Optimization Conducted in the Carathéodory Parameters Result in a P.D. Toeplitz Matrix with a Slightly Smaller LR Value

While the analysis of the fmincon optimization results with direct covariance lag optimization provided very important insight into the nature of this optimization problem, it may not be considered a practical technique due to its high failure rate. As an alternative, one may use the Carathéodory parameterization of the optimized Toeplitz matrix. Let us consider an arbitrary p.d. Toeplitz Hermitian matrix $\mathbf{T}_N$ with the eigendecomposition:

$$\mathbf{T}_N = \mathbf{U}_N \mathbf{\Lambda}_N \mathbf{U}_N^{\mathrm{H}} = \lambda_{\min} \mathbf{I}_N + \sum_{j=1}^{N-1} (\lambda_j - \lambda_{\min}) \mathbf{U}_j \mathbf{U}_j^{\mathrm{H}}. \quad (25)$$

Since the rank-deficient second matrix in the decomposition (25) is also a Toeplitz Hermitian matrix, we can apply Theorem 02 from [19]:

Theorem 02 [19]

Let $\mathbf{T}_N$ be an $N \; x \; N$ positive semidefinite Toeplitz matrix with rank m. Then

$$\mathbf{T}_N = \mathbf{R}\mathbf{D}\mathbf{R}^{\mathrm{H}}, \quad (26)$$

where $\mathbf{D}$ is the $m \; x \; m$ diagonal matrix with positive diagonal entries, and $\mathbf{R}$ is the $N \; x \; m$ Carathéodory matrix with $\left(a_k^{j-1}\right)_{1 \le j \le N, i \le k < m}$, and $\left|a_k^{j-1}\right| = 1$ for $1 \le k \le m$ ∎

According to this theorem, any rank-deficient non-negative Toeplitz Hermitian matrix may be presented as the matrix of $m < N$ independent plane waves. Therefore, an alternative parameterization of the p.d. Toeplitz Hermitian matrix is the following:

$$\mathbf{T}_N = \lambda_{\min} \mathbf{I}_N + \sum_{j=1}^{N-1} \sigma_j^2 \mathbf{S}(\theta_j) \mathbf{S}(\theta_j)^{\mathrm{H}}, \quad \lambda_{\min}, \sigma_j^2 > 0. \quad (27)$$

Note that in $(2N - 1)$ real-valued parameters that describe the Toeplitz matrix (27), N parameters ($\lambda_{\min}$, σ_j^2, $j = 1, \dots, N$), are positive and $(N - 1)$ parameters θ_j, $j = 1, \dots, N$, are the angles of arrival; $|\theta_j| < \pi/2, j = 1, \dots, N$. The total number of these parameters is $2N - 1$ and equals the Toeplitz matrix's description in terms of its covariance lags. But the restrictions of these parameters are very different for these two sets of the $(2N - 1)$ real-valued parameters.

In the Carathéodory parameters in (27), N powers ($\lambda_{\min}$, σ_j^2, $j = 1, \dots, N$) are positive, and $(N - 1)$ angles of arrival parameters θ_j are limited to the half-plane:

$$\lambda_{\min}, \sigma_j^2 > 0; \quad |\theta_j| < \frac{\pi}{2}, \quad i = 1, \dots, N-1. \quad (28)$$

In the Toeplitz matrix optimization via its covariance lags (Vandermonde description), only the central element is positive, while the remaining $(N - 1)$ complex-valued covariance lags t_j, $j = 1, \dots, N - 1$, should be represented by the $2(N - 1)$ real-valued parameters with no simple restriction on their values. Potentially, one can also use the polar description of the covariance lag,

$$\mathrm{t}_j = |t_j| \exp(j \arg t_j). \quad (29)$$

The number of real-valued parameters is the same, equal to $2(N - 1)$. Yet, while the moduli $|t_j|$ could be only positive, the arg t_j may be an arbitrary number within the entire range

$$0 < \arg t_j < 2\pi. \quad (30)$$

Therefore, despite the same number of $(2N - 1)$ real valued degrees of freedom, the Carathéodory set has the maximum number of constraints on these parameters and therefore describes a different set of N-variate Toeplitz matrices. In particular, the Carathéodory parameters describe only the positive definite matrices, given $\lambda_{\min}$, $\sigma_j^2 > 0$. Not surprisingly, the arbitrary covariance lag t_j describes both positive definite and non-positive definite Toeplitz matrices.

The least expected result was that the Carathéodory parameters did not reach the same LR maximum as the optimized covariance lags. While the reason may be the limitations of the Carathéodory set, the specifics of the particular LR maximization fmincon routine and finite computational accuracy may also be to blame.

Since, in most cases, the Carathéodory representation (27) did not reach the maximum likelihood achieved by the covariance lags optimization, we had to introduce the two-step LR optimization. We proposed starting LR maximization with the Carathéodory parameterization (27) and, upon convergence, continuing it with direct covariance lag optimization via the MATLAB fmincon routine.

Let us now compare the optimization results for the covariance lags with those of the LR maximization in the Carathéodory parameterization for the same sample matrix $\widehat{\mathbf{R}}_N$ and the same successful initialization solutions for the covariance lags parameterization. We demonstrate below the behavior of the maximized likelihood ratio for the two different initializations. Note that for the Carathéodory initialization, there is no need for multiple different initializations to avoid solutions with negative eigenvalues.

Yet, in the vast majority of the 1,000 conducted trials, fmincon with the Carathéodory parameterization did not reach the maximum LR values, which were delivered for the same input data by the fmincon optimization in the covariance lags parameterization. Specifically, in only 50 out of 1,000 trials, the LR values, obtained using Carathéodory parameterization, exceeded—insignificantly (by no more than 0.0015%)—the LR values obtained by maximizing the LR under covariance lag parameterization.

Given the different calculations for these two parameterizations, this insignificant difference could be attributed to the finite accuracy of the calculations. Out of 1,000 trials with a different sample matrix $\widehat{\mathbf{R}}_N$ per trial, in 950 trials, the LR optimized by the covariance lag parameterization exceeded the LR from the Carathéodory parameterization alone by a few percent (up to 15%, Figure 1).

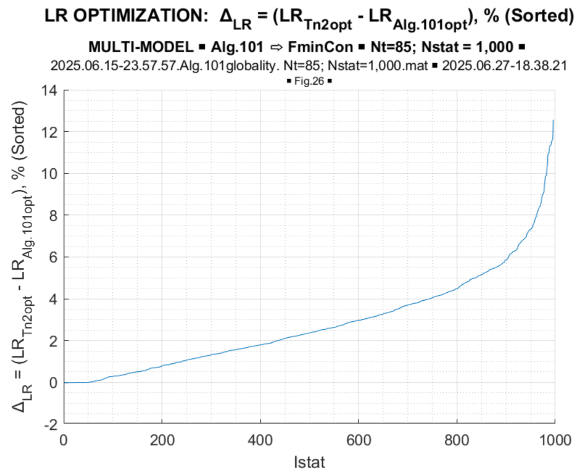

Figure 1: LR Gain Provided by the Final "Natural" Parameterization with Respect to the Initial "DoAs" Parameterization (1,000 trials)

Based on these numerical results, we may conclude that the proposed combination of parameterizations, starting with Carathéodory fmincon LR maximization and followed by the covariance lags parameterization, allowed us to mostly avoid solutions with negative eigenvalues, thus making an arbitrary p.d. Toeplitz matrix appropriate for fmincon initialization. By optimizing the covariance lags, we reach the LR maximum while mostly avoiding solutions with negative eigenvalues. Since the LR maximum is the same as the one achieved by fmincon when starting from the true covariance matrix $\mathbf{T}_N$, we may treat these solutions as globally optimum. Moreover, the number of these additional iterations for the covariance lag parameters is negligible.

In our first example (Figure 2), after 925 iterations with the Carathéodory parameters and LR = $1.708 \cdot 10^{-1}$, it required only 44 iterations using the Vandermonde covariance lags to reach the global LR maximum of LR = $1.853 \cdot 10^{-1}$.

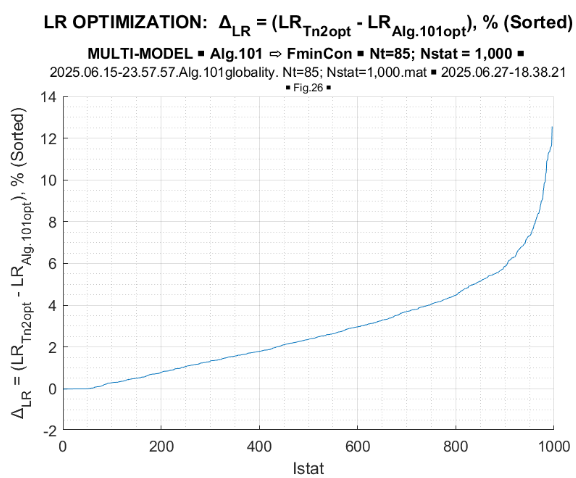

Figure 2: LR Gain Provided by the Final "Natural" Parameterization with Respect to the Initial "DoAs" Parameterization (1,000 trials)

In Figure 3, we show the percentage improvement in LR achieved by this additional optimization of covariance lags.

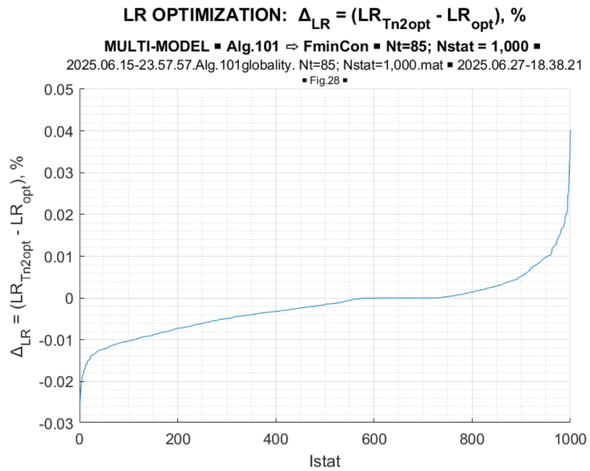

Figure 3: Equal LRs for Optimization, Started from the True Matrix (LR_{Tn2opt}) and a Random Matrix (LR_{opt}) (1,000 trials)

One can see that in only one out of 1,000 conducted trials, the LR maximized over Carathéodory parameters exceeded the LR value from direct covariance lags optimization by no more than 0.01%. In all other trials, the direct optimization of covariance lags provided further LR improvement over the Carathéodory parameterization, with gains of less than 15% and an average of 6%. Therefore, starting the LR fmincon maximization from the Carathéodory matrix description, we avoid solutions with negative eigenvalues for a practically arbitrary p.d. Toeplitz matrix used to initialize the fmincon routine. The following LR maximization using the covariance lags description yields a globally optimal solution in all cases, with the achieved LR value equal to that initiated by the true covariance matrix $\mathbf{T}_N$.

Let us remind the reader that the comparative analysis of two different problem formulations in covariance lags and Carathéodory parameters was performed for the same sample covariance matrix $\hat{\mathbf{R}}_N$. For this reason, in our next simulation series, we investigated 1,000 different sample matrices generated from the same data with the same true Toeplitz covariance matrix. The results of these simulations are illustrated in Figure 4 and Figure 5, where the dependence of the maximized LR on the number of iterations is provided.

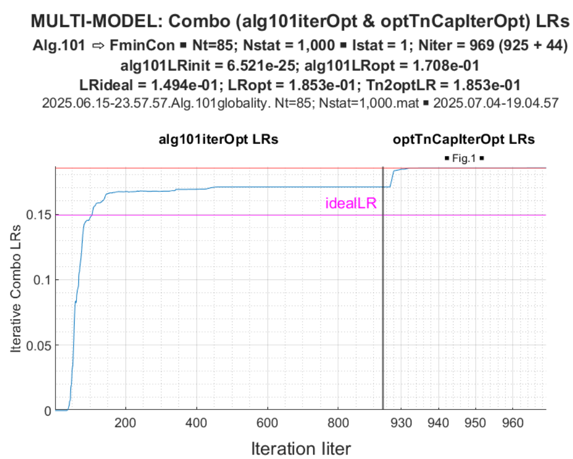

Figure 4: fmincon LR Progression During the Initial Stage of Optimization in the “DoAs” Parameterization, Followed by the Optimization in “Natural” Parameterization

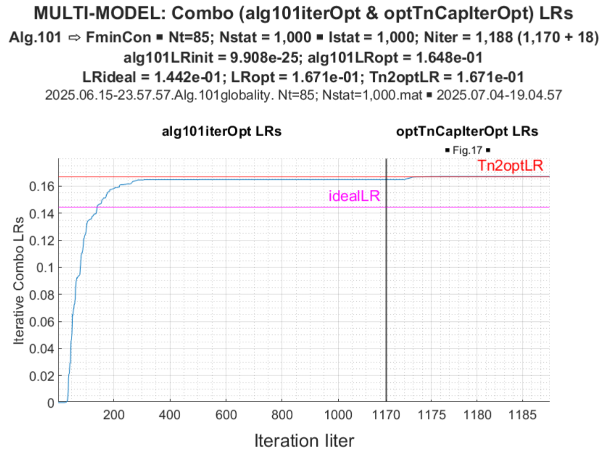

Figure 5: Another Example of the Two-Stage fmincon LR Maximization

The first part of the curve in Figure 4 corresponds to LR maximizations performed in Carathéodory parameters (925 iterations), followed by 44 iterations using covariance lag optimization by the fmincon routine. From ~150 iterations to the final 925 in the Carathéodory parameters, the LR value did not change much, finally reaching LR = $1.708 \cdot 10^{-1}$. The next 44 iterations in covariance lags values, conducted by covariance lags trimming, reached the global maximum LR = $1.853 \cdot 10^{-1}$, which significantly exceeded the LR value produced by the true covariance matrix, LR = $1.494 \cdot 10^{-1}$.

The second example, illustrated by the Figure 5, differs only in LR values, with the minimal gain provided by the second level LR maximization, reaching LR = $1.671 \cdot 10^{-1}$ compared with the first-order LR optimization in Carathéodory parameters, reaching LR = $1.648 \cdot 10^{-1}$. Note that the LR values generated by the true covariance matrix were equal to LR = $1.442 \cdot 10^{-1}$, versus LR = $1.494 \cdot 10^{-1}$ in our first example.

The sample pdf of the LR gains (in percentage) achieved by the second-level LR maximization using the covariance lags optimization, compared with the LR achieved as a result of the first-level LR maximization in the Carathéodory parameters, is presented in Figure 6.

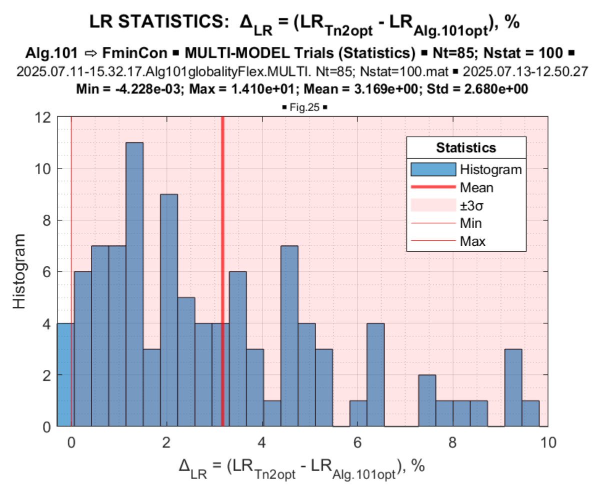

Figure 6: Distribution of the LR Gains Provided by the Second Stage LR Maximization in "Natural" Parameters

One can see that this gain can reach 10%, with an average of 3.2%. Only in very rare cases did the combined LR maximization slightly exceed the LR values obtained with the covariance lags parameterization alone.

These improvements do not exceed 0.005% and are therefore attributed to finite calculation accuracy. It seems acceptable that for quite complicated LR optimization algorithms, we may treat the results of our two-step LR maximization as globally optimal values, despite the negligible discrepancy caused by finite computational accuracy. In Figure 7, we provide the results of LR optimization, sorted in increasing order of the first step of LR maximization in Carathéodory parameters.

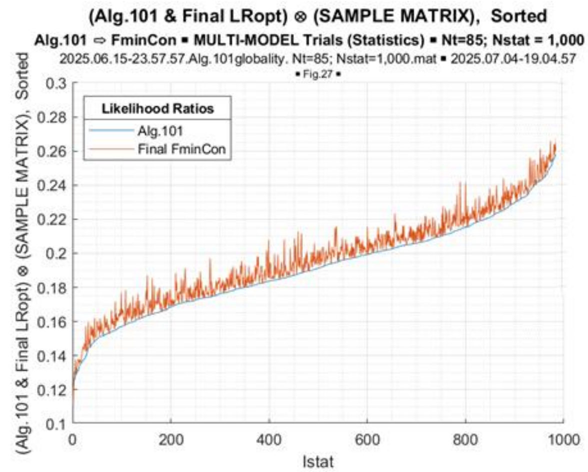

Figure 7: Maximized LR Values by "DoAs" Parameterization (blue line) and by the "Natural" Parameters Optimization (red line)

Then, for every LR value obtained from the first-step LR maximization over Carathéodory parameters, we provide the LR value achieved by the second-step LR maximization via direct trimming of the covariance lags. One can see that the initial LR maximization over Carathéodory parameters ranges from 0.12 to 0.28, while the second-stage LR maximization using the covariance lags adds, on average, 0.01 to the global LR maximum.

The conducted trials provide a detailed enough description of the statistical nature of the LR maximum using the proposed two-step numerical optimization.

C. Second Stage LR Maximization of the Covariance Lags Improves LR Values but Degrades the Proximity to the True Toeplitz Covariance Matrix

In statistical signal processing, it is well known that the more specific the a priori knowledge about the estimated set, the smaller the excess optimized likelihood ratio is and the closer the estimation result is to the true parameters. The comparison of the Hermitian and Toeplitz Hermitian covariance matrices' ML estimates is one such example, where the maximized LR value for the Hermitian matrix is always 1, whereas a priori knowledge of its Toeplitz structure leads to a significant degradation in the maximized LR value.

The discrepancy between the optimized LR values for the Carathéodory parameters and the covariance lag parameters should not arise, since we are only interested in positive definite matrices. Yet, the demonstrated superior performance of the fmincon LR optimization for covariance lag parameters is beyond any doubt and is not associated with a more specific (more accurate) model of the estimated Toeplitz matrix. Therefore, the optimized Toeplitz matrices with covariance lags are closer to the sample matrices $\hat{\mathbf{R}}_N$, since the LR is larger. But are they closer to the true covariance Toeplitz matrix as well? To assess the achieved proximity of the optimized Toeplitz matrix to this true covariance matrix $\mathbf{T}_N$ rather than to the sample matrix $\hat{\mathbf{R}}_N$, let us introduce the "proximity ratio":

$$\mathrm{PR}\left[\mathbf{T}_N|\hat{\mathbf{T}}_N^{\mathrm{ML}}\right] = \frac{\det\left(\mathbf{T}_N\left(\hat{\mathbf{T}}_N^{\mathrm{ML}}\right)^{-1}\right)}{\left[\frac{1}{N}\mathrm{Tr}\left(\mathbf{T}_N\left(\hat{\mathbf{T}}_N^{\mathrm{ML}}\right)^{-1}\right)\right]^N}. \tag{31}$$

Note that $\mathrm{PR}\left[\mathbf{T}_N|\hat{\mathbf{T}}_N^{\mathrm{ML}}\right]$ reaches a maximum (equal to one) only when $\hat{\mathbf{T}}_N^{\mathrm{ML}}$ is proportional to the true matrix, i.e., when

$$\mathbf{T}_N = c\hat{\mathbf{T}}_N^{\mathrm{ML}}, \quad c > 0, \tag{32}$$

which means that PR may be treated as the level of proximity of the optimized estimate $\hat{\mathbf{T}}_N^{\mathrm{ML}}$ to the true Toeplitz covariance matrix $\mathbf{T}_N$.

Let us check the "proximity ratio" (31) analysis for each step of the LR[$\hat{\mathbf{T}}_N^{\mathrm{ML}}$]. In Figure 8, we illustrate an analysis of the data of Figure 7.

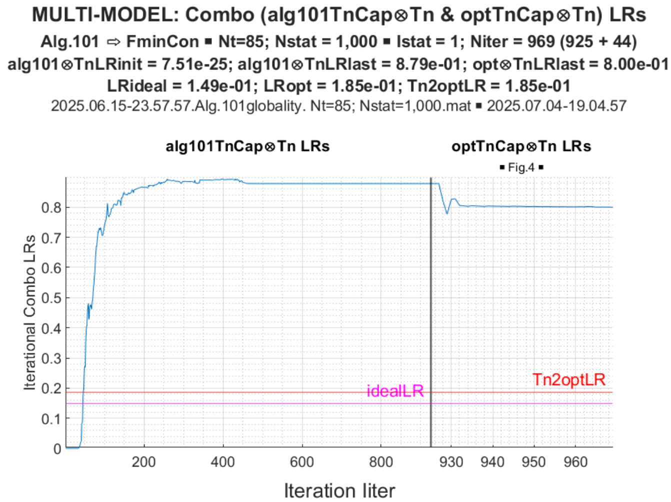

Figure 8: fmincon PR Progression with Respect to the True Covariance Matrix $\mathbf{T}_N$

The proximity ratio, calculated over 955 iterations for the LR maximization interval in Carathéodory parameters, is followed by 40 iterations of LR maximization over covariance lags. While the LR maximization is performed on the sample matrix, the resulting value is the proximity ratio. Also surprising is the proximity ratio, calculated for the interval with Carathéodory parameters optimization, which remains practically equal to the LR values over this interval.

More surprising are the results of the "proximity ratio" analysis over the second step of the LR maximization, performed over the covariance lags of the optimized Toeplitz matrix. Indeed, the results of this second-step LR maximization led to a degradation of the "proximity ratio", with the losses in PR being practically equal to the LR gain achieved in the second step of LR maximization of covariance lag parameters.

It seems rather natural that all the further gains in LR values, provided by fmincon using the covariance lags for the Toeplitz matrix representation, were lost when the received solution was compared with the original true Toeplitz covariance matrix $\mathbf{T}_N$ and not with the sample matrix $\hat{\mathbf{R}}_N$, as per the likelihood ratio. This property evokes certain doubts in the application of the "second step" LR maximization using the covariance lag parameters for the optimized Toeplitz matrix parameterization. This concern is illustrated by another example of a successful trial (Figure 9), which converged to the solution with LR = $1.8529 \cdot 10^{-1}$ after 481 iterations.

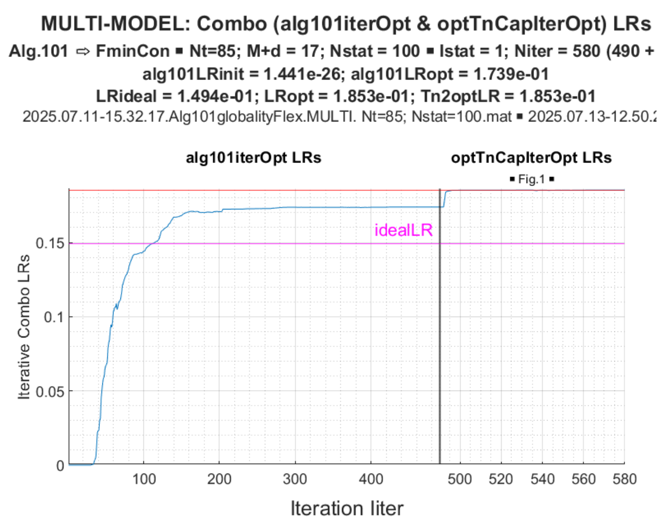

Figure 9: fmincon LR Progression During the Initial Stage of Optimization in "DoA" Parameters, Followed by Optimization in "Natural" Parameters

The true Toeplitz covariance matrix generated the likelihood ratio LR = 1.494 · 10^{-1}. When the true covariance matrix was used to initialize the next fmincon optimization, the LR remained the same, LR = 1.8529 · 10^{-1}. Therefore, the 180th initialization was successful, following 179 previous unsuccessful ones, that ended with negative minimal eigenvalues but reached the same Maximum Likelihood value (LR = 1.8529 · 10^{-1}) as the fmincon routine when launched from the true covariance matrix, which had LR = 1.494 · 10^{-1}. This experiment was then repeated with different random-like initializations of the fmincon routine.

The Carathéodory parameterization did not produce solutions with negative eigenvalues, but in the overwhelming majority of conducted trials, it did not achieve LR values comparable to those from a successful (i.e., with no negative eigenvalues) LR maximization that exploited the covariance lags parameterization. Therefore, for the class of Toeplitz covariance matrices with an unspecified number of equal minimal eigenvalues, the proposed optimization of the "oversampled" ($T \geq N$) LR by initial optimization in Carathéodory parameters with $m = N - 1$ point sources, followed by optimization over the Toeplitz matrix's covariance lags, provides the best convergence rate to the LR maximum, measured by the number of required initializations.

4. Numerical Likelihood Ratio Maximization for the Models with Limited Signal Subspace Dimension

In the important class of direction of arrival estimation problems, the covariance matrix has a finite dimension $m \ll N$ of its signal subspace. The number of active sources m (signal subspace dimension) is usually estimated before DOA estimation and is treated as a known parameter. Therefore, for the class of uniform linear arrays (ULAs), the DOA estimation of $m \ll N$ sources is practically identical to the problem of Hermitian Toeplitz p.d. matrix reconstruction, with the Toeplitz matrix parameterized as in (27) with $m \ll N$.

For this type of problem of DOA estimation of $m \ll N$ sources, the MUSIC solutions are often produced for a sample volume T that is smaller than the matrix dimension N and exceeds the number of sources m:

$$m \leq T < N. \tag{33}$$

Therefore, the Maximum Likelihood solution should be developed for the "undersampled" training data ($T < N$) as well. The requirement to expand the maximum likelihood Toeplitz matrix estimation over the class of Toeplitz matrices with a finite signal subspace dimension $m \ll N$ is an important distinction with respect to the problems addressed in the previous chapter.

Another important distinction with the previous problem is the known limited signal subspace dimension m. Since this type of Toeplitz covariance matrix is very similar to (27):

$$\mathbf{T}_N = \sigma_n^2 \mathbf{I}_N + \sum_{j=1}^{m} \sigma_j^2 \mathbf{S}(\theta_j)\mathbf{S}(\theta_j)^{\mathrm{H}}, \quad m \ll N, \tag{34}$$

it is quite natural to use this model instead of (27) with $m = N - 1$ for LR maximization. Yet, the model (27) with its $m = N - 1$ describes the full class of p.d. Toeplitz Hermitian matrices, as demonstrated above. Therefore, the reduction of the "full" model (27) to the model (34) with the smaller $m \ll N$ may affect our ability to calculate the maximum likelihood Toeplitz matrix estimate. The Toeplitz matrix parameterization based on its elements is hardly applicable in this case.

Indeed, the constraints on equality of $(N - m)$ minimal eigenvalues of the matrix $\mathbf{T}_N$ may be as difficult to retain, if not more difficult, than the condition on the positive definiteness of the optimized Toeplitz matrix. Our attempts to use the model (34) with $m \ll N$ for the MATLAB fmincon optimization failed, since the iterations converged to an LR value that remained smaller than the LR value of the true covariance matrix $\mathbf{T}_N$. A mathematically rigorous explanation of this phenomenon is still missing, but a few negative examples suffice to demonstrate this property.

Note that for the "oversampling" training conditions ($T \geq N$), the eigenvalues of the sample covariance matrix $\hat{\mathbf{R}}_N$ are all different with probability one. Therefore, a different approach can be applied to the problem of maximum likelihood Toeplitz matrix estimation with a finite dimension $m < N$ of the matrix signal subspace. For $T > N$, the sample matrix $\hat{\mathbf{R}}_N$ has all different eigenvalues, and it is possible to look for the maximum likelihood Toeplitz matrix with all different eigenvalues as well.

It is clear that the global maximum likelihood of this solution should exceed the global maximum likelihood for the finite signal subspace dimension. Since such an ML Toeplitz matrix estimate may always be presented as the covariance matrix of $(N - 1)$ sources plus white noise, one can present the received solution in this format and then select the m most powerful sources. The likelihood of this solution must exceed that of the true covariance matrix, which can be checked during Monte Carlo simulations. This approximation can be found by alternating projections over the "full" rank Toeplitz solutions presented in format (25), by the straightforward exclusion of the $(N - 1 - m)$ weakest sources.

For applications with the "undersampled" training condition ($T < N$), the optimization of the Toeplitz matrix should be different. One approach is to find the "diagonally loaded" Toeplitz matrix:

$$\mathbf{T}_N = \alpha \mathbf{I}_N + \mathbf{T}(m), \tag{35}$$

where $\mathbf{T}(m)$ is the Toeplitz matrix of rank m. For this matrix, we may first try to find the Toeplitz p.d. matrix that maximizes the LR ratio modified in [1,16]:

$$\mathrm{LR}\left(\hat{\mathbf{R}}_N \middle| \hat{\mathbf{T}}_N\right) = \frac{\det\left[\,\mathbb{X}_T^{\mathrm{H}}\hat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right]}{\left[\frac{1}{T}\ \mathrm{Tr}\left(\mathbb{X}_T^{\mathrm{H}}\hat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right)\right]^{\mathrm{T}}}, \qquad (36)$$

and then find the best approximation of this Toeplitz matrix by a p.d. Toeplitz matrix with a smaller signal subspace dimension $m < N$. A special investigation is required to determine the optimal m for signal subspace dimension estimation. It is very likely that for applications with a ULA receiver array, better performance may be achieved if the signal subspace dimension m is identified after the maximum likelihood Toeplitz matrix is estimated. While the classical MUSIC may be applied to this ML Toeplitz matrix estimate, in this study, we tested alternating projections to find the p.d. Toeplitz matrix of dimension m for the signal subspace, avoiding the "full" Toeplitz matrix transformation to the "Carathéodory" format (25).

Let us first test the "oversampled" case (T = 85) of DOA estimation by comparing the results of the classical MUSIC applied to the sample matrix $\hat{\mathbf{R}}_N$ to the ML Toeplitz matrix full rank solution. The ML Toeplitz matrix estimate was obtained by first optimizing the Carathéodory representation (25) with the ultimate number of ($N-1$) sources, followed by ML optimization of the Toeplitz matrix's covariance lags.

We consider the scenario with m = 3 independent sources acting upon an N=17-element uniform linear array with d/λ = 0.5. The three sources have DOAs:

$$\theta_1 = -30^o, \theta_2 = 0^o, \theta_3 = 1^o, T = 85. \qquad (37)$$

The results of the two-step LR optimization are illustrated in Figure 10, which demonstrates the worst estimated DOA error over the three sources. Recall that in the first step, we used the Carathéodory representation, and in the second step, the Vandermonde representation. The analysis of the "proximity ratio" with respect to the true covariance matrix (instead of the sample matrix) is presented on Figure 11. It demonstrates a much higher value of the "proximity" to the true covariance matrix $\mathbf{T}_N$ than to the sample matrix $\hat{\mathbf{R}}_N$ in the maximized likelihood ratio.

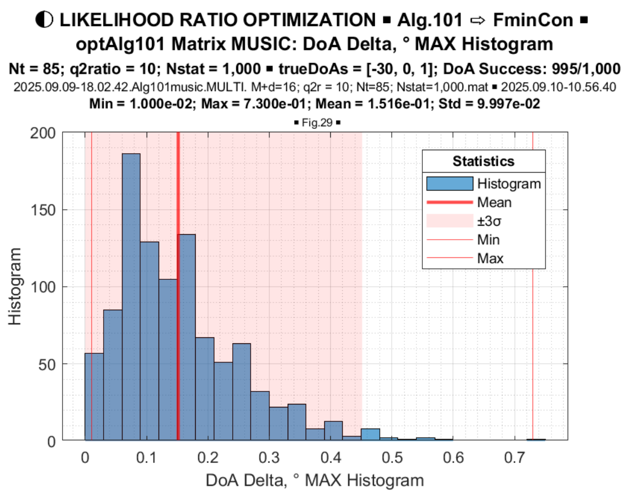

Figure 10: Maximal “DoAs” Estimation Error PDF

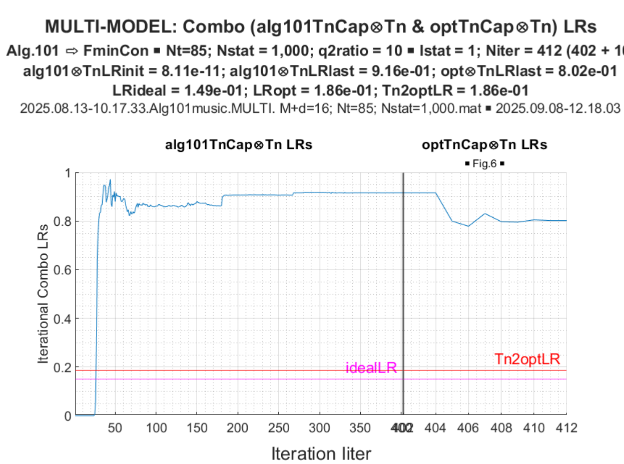

Figure 11: fmincon Proximity Ratio Progression for LR-optimized “DoAs” Reconstructed Toeplitz Covariance Matrix

Even though the final step of LR maximization over the covariance lags of the optimized Toeplitz matrix increased the likelihood ratio from LR = $1.630 \cdot 10^{-1}$ to LR = $1.863 \cdot 10^{-1}$, the "proximity ratio" in fact degraded as a result of this optimization from LR = 0. 916, achieved by the Carathéodory parameters optimization, to LR = 0.802 produced by it due to the second-stage LR maximization in covariance lags.

This phenomenon was observed across all conducted trials, suggesting that further LR enhancement beyond the level achieved by the Carathéodory parameterization is not helpful for DOA estimation. Note that for the selected parameters, the classical MUSIC applied to the sample matrix $\hat{\mathbf{R}}_N$ does not resolve the second and third sources in any of the 1,000 trials, while in the proposed Maximum Likelihood Toeplitz matrix estimation, all three sources were resolved in every trial.

In Figure 12, we provide the conventional MUSIC pseudospectrum demonstrating this phenomenon.

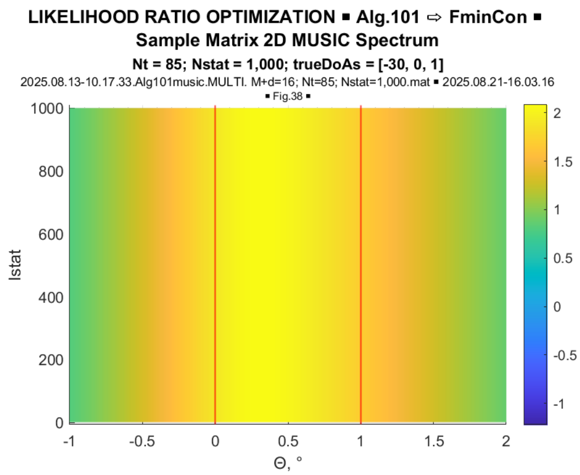

Figure 12: 1,000 MUSIC Pseudo-Spectra of the Second and Third (unresolved) Sources of the Original Sample Matrix $\hat{\mathbf{R}}_N$

Let us continue by presenting the results of maximum likelihood Toeplitz matrix estimation by analyzing the maximum (over the three acting sources) estimation error. Since LR optimization over the covariance lags increased the likelihood ratio but decreased the "proximity" of the solution to the true Toeplitz matrix $\mathbf{T}_N$, let us separately analyze the results of the first stage of LR maximization using the Carathéodory parameterization and the results of the following LR optimization of the covariance lags. As expected, despite the second stage of LR maximization in covariance lags increasing the LR values, the DOA estimation accuracy did not improve.

This result is expected because the "proximity factor" above degraded during the second stage of LR maximization by fmincon, which used the covariance lags as the optimized parameters. The average over all three acting sources RMSE is equal to RMSE = 0.1617° with the standard deviation σ = 0.1006 for the first-stage LR optimization in Carathéodory parameters, and RMSE = 0.1625° and σ = 0.1006 for the second stage of LR improvement via covariance lag optimization. One can see that the improvement in the LR value at the second stage of optimization did not translate into improved DOA estimation accuracy.

Obviously, Monte-Carlo simulation results should be compared with the associated Cramér-Rao lower bounds. The Fisher information matrix (FIM) is calculated using the formula [26]:

$$\mathrm{FIM} = \mathrm{Tr}\left[\mathbf{T}_N^{-1}\frac{\partial \mathbf{T}_N}{\partial \gamma_l}\mathbf{T}_N^{-1}\frac{\partial \mathbf{T}_N}{\partial \gamma_k}\right], \tag{38}$$

where γ_j are the estimated parameters, equal in our case to the DOA parameters (σ_j^2, θ_j), $j = 1, \ldots, 3$. For the simulated parameters $N = 17$, $T = 85$, $q^2 = 10$ dB, we got RMSE = 0.0093 − 0.05673 for the scenario with three true DOAs, $\theta = 30^o, 0^o$, and 1^o, which are in good agreement with the results of the Monte-Carlo simulations.

With respect to the conventional "CBF + MUSIC" inability to resolve the second and third sources, the ML treatment of this sample covariance matrix $\hat{\mathbf{R}}_N$ provided the practical utility of MUSIC DOA estimation when applied to the ML Toeplitz covariance matrix.

In Figure 13, we provide a sample distribution for the LR gain $(\mathrm{LR}[\mathbf{T}_{\mathrm{ML}}] - \mathrm{LR}[\mathbf{T}_N])$ and $(\mathrm{LR}[\mathrm{T}_{(m)}^{\mathrm{ML}}] - \mathrm{LR}[\mathbf{T}_m])$, and with the equalized $(N - m)$ noise subspace eigenvalues.

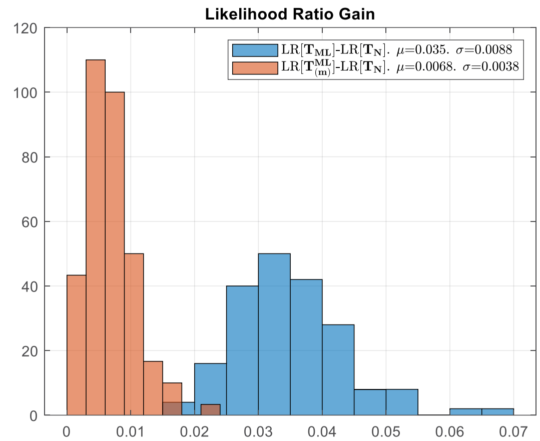

Figure 13: Sample distribution of the LR gain of ($\mathrm{LR}[\mathbf{T}_{\mathrm{ML}}] - \mathrm{LR}[\mathbf{T}_N]$) and $\mathrm{LR}[\mathbf{T}^{\mathrm{ML}}_{(\mathrm{m})}] - \mathrm{LR}[\mathbf{T}_{\mathrm{N}}]$), where $\mathbf{T}_{\mathrm{ML}}$= maximized LR with no eigenvalues equalization, $\mathbf{T}^{\mathrm{ML}}_{(\mathrm{m})}$ = ML Toeplitz matrix with equalized $(N{-}m)$ noise subspace eigenvalues, $\mathbf{T}_N$ = true Toeplitz covariance matrix

While the noise subspace eigenvalues equalization somewhat degraded the LR of the ML Toeplitz matrix, the likelihood ratio of the Toeplitz matrix $\mathbf{T}^{\mathrm{ML}}_{(\mathrm{m})}$ still exceeded the LR value of the true Toeplitz matrix $\mathbf{T}_m$. Despite these differences, the DOA estimation accuracy provided by both Toeplitz matrices is practically the same and consistent with the CRB.

Recall that the traditional MUSIC, applied to the sample covariance matrix, failed to resolve these close-in sources. In Figure 14 and Figure 15, we introduce 1,000 MUSIC pseudospectra, calculated for Toeplitz matrices with maximized LR, first using the Carathéodory Toeplitz matrix parameterization and then maximizing LR using the covariance lags directly.

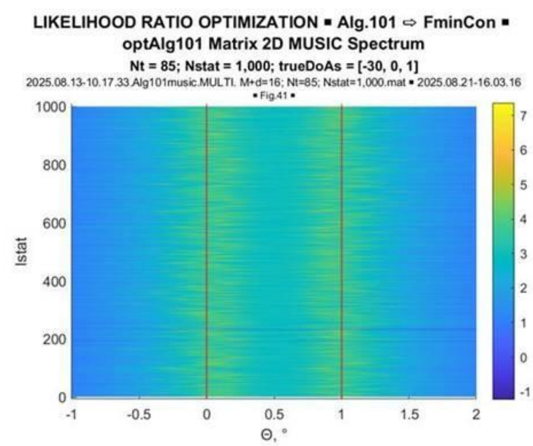

Figure 14: 1,000 MUSIC Pseudo-Spectra of the First (top figure) and Second and Third Sources (bottom figure) after LR Maximization in “DoAs”-Only Parameters

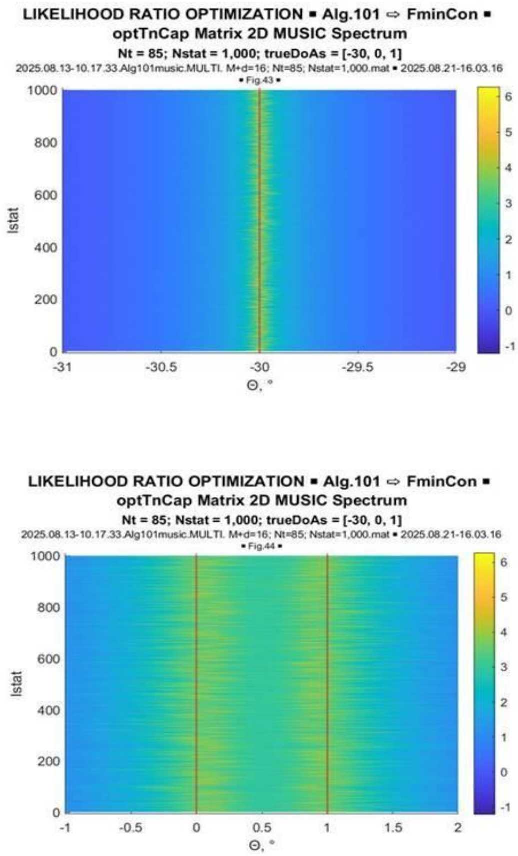

Figure 15: 1,000 MUSIC Pseudo-Spectra of the First (top figure) and Second and Third Sources (bottom figure) after Two-Stage LR Maximization in the “DoAs” and “Traditional” Parameters

Despite the second step of LR maximization using covariance lags slightly increasing the optimized LR values, the DOA estimation accuracy in both steps remains largely unchanged. Note also that the restriction of the signal subspace dimension by alternating projections sufficiently increased the dynamic range of the MUSIC pseudo-spectrum, but did not improve the spectral peak positions and associated DOAs estimation errors.

In Figure 16, we compare the MUSIC pseudo-spectra of the original ML Toeplitz matrix with the results of alternating projections that left only three signal subspace eigenvalues above the noise eigenvalue floor in the modified p.d. Toeplitz matrix.

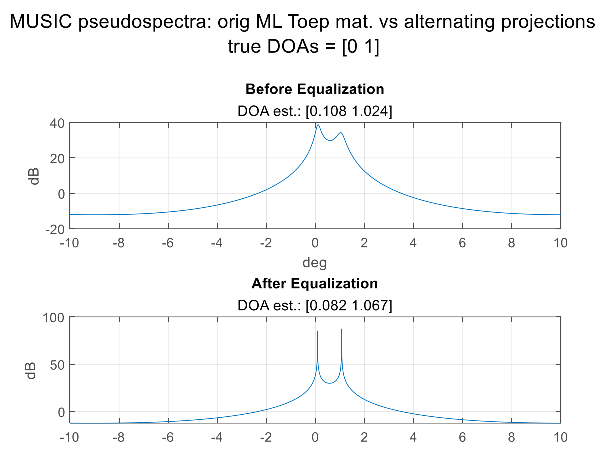

Figure 16: MUSIC Spectrum: Original ML Toeplitz Matrix vs. Alternating Projections

5. LR Maximization in the “Undersampled” Training Conditions

Obviously, if the MUSIC DOA estimates exist for a small training sample support ($T < N$), it would be prudent to extend the Maximum Likelihood methodology to this class of problems and treat it as the "benchmark" for the provided accuracy. For example, we may be looking for the maximum of the "modified sphericity" test (38):

$$\underbrace{\mathrm{LR}\left(\mathbb{X}_T|\hat{\mathbf{T}}_N\right)}_{T<N} = \frac{\det\left[\mathbb{X}_T^{\mathrm{H}}\hat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right]}{\left[\frac{1}{T}\operatorname{Tr}\left(\mathbb{X}_T^{\mathrm{H}}\hat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right)\right]^{\mathrm{T}}}, \tag{39}$$

introduced in, where the pdf of this test for the case when $\hat{\mathbf{T}}_N$ is equal to the true covariance matrix $\mathbf{T}_N$ is also introduced [16]. Note that this ratio maximization requires some care. Indeed, the important properties of the maximum likelihood are proven for an asymptotically large i.i.d. sample volume $T \to \infty$. For a finite T, none of these properties are accurate, while for a small T, they may not hold. Indeed, the LR (39) gets smaller for a growing i.i.d. sample support T. Indeed, since

$$\mathbf{X}_t = \mathbf{T}_N^{1/2}\boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_N), \tag{40}$$

$$\mathrm{LR}(\mathbf{X}_T|\mathbf{T}_N) = \frac{\det[\mathfrak{J}_T^{\mathrm{H}}\mathfrak{J}_T]}{\left[\frac{1}{T}\operatorname{Tr}(\mathfrak{J}_T^{\mathrm{H}}\mathfrak{J}_T)\right]^{\mathrm{T}}}, \tag{41}$$

where

$$\mathfrak{J}_T^{\mathrm{H}}\mathfrak{J}_T = \sum_{t=1}^{N}\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{\mathrm{H}}, \quad \boldsymbol{\xi}_t \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_T). \tag{42}$$

Obviously, for $N > T$ with a smaller T, the random matrix gets closer to the diagonal matrix, which increases the LR value [27]. Indeed, for $N = 17$ and $T = 5$, the sample LR average is $\overline{\mathrm{LR}} = 3.514\cdot 10{-}1$, while for $T = 17$ (T = N), this likelihood ratio has a minimal mean value $\overline{\mathrm{LR}[17]} = 2.478\cdot 10{-}8$. Therefore, for the "undersampled" training conditions, the growth of the training sample volume reduces the "expected" likelihood of the true covariance matrix, while for the "oversampled" scenarios ($T \geq N$), the expected likelihood "grows" with the sample support. Therefore, since some DOA estimation methods for $m \ll N$ exist for $T < N$, finding the "benchmark" is not that straightforward due to the discussed properties of the likelihood ratio test.

Let us introduce another special case for the data, described by the p.d. Toeplitz covariance matrix. This case is based on the fundamental property of the multivariate vector $\mathbf{X}_t$, described by the p.d. Hermitian Toeplitz matrix $\mathbf{T}_N$:

$$\mathrm{E}[\mathbf{X}_t\mathbf{X}_t^{\mathrm{H}}] = \mathbf{T}_N, \tag{43}$$

which suggests that the inverted and complexly conjugated vector $\mathcal{J}\mathbf{X}^*$ is described by the same covariance matrix $\mathbf{T}_N$. Using this property of the Toeplitz covariance matrix data set, for the sample volume T that exceeds half of the antenna dimension:

$$T > \frac{N}{2}, \tag{44}$$

we may introduce the sample matrix $\hat{\mathbf{T}}_N$

$$\hat{\mathbf{T}}_N = \frac{1}{2T}\left[\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t^{\mathrm{H}} + \sum_{t=1}^{T}\mathcal{J}\mathbf{X}_t^{*}\mathbf{X}_t^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\right], \tag{45}$$

that should have the same mean value

$$\mathrm{E}\left[\hat{\mathbf{T}}_N\right] = \mathbf{T}_N, \tag{46}$$

and a total number of distinct training samples exceeding N. In this case, we may use the LR expressions derived for the "superior" case, bearing in mind that the statistical dependence of the training data should reduce the "expected" likelihood ratio of the true Toeplitz covariance matrices.

Let us now introduce an example with $m = 3$ independent sources in an N = 17-element antenna array, where the i.i.d. sample volume $T = 10$ allows us to exceed the antenna dimension of $N = 17$ by "doubling" the sample volume, as described by (45). For comparison, we also simulated a scenario using T = 20 truly independent Gaussian training vectors to evaluate the losses associated with the statistical dependence of the training data. As with the previous example, the input SNR was selected in a way that makes MUSIC incapable of resolving the close sources, and for $T = 10$, the standard MUSIC applied to the same sample matrix $\hat{\mathbf{R}}_N$ failed in all 1,000 trials. The training sample volume "doubled" by (45) in the Carathéodory parameterization in the MATLAB fmincon routine, allowed us to increase the probability of correct initializations to 822 (out of 1,000 trials) for $T = 20$ i.i.d. training samples and 289 (out of 1,000 trials) for the "doubled" training samples ($N = 10 \times 2 = 20$) using the Toeplitz property (45)-(46).

Finally, note that our second LR maximization using the covariance lags (Vandermonde) parameterization followed the previously established trend and increased the success rate of the Carathéodory parameterization from 820 to 822 successful trials (for $T = 20$ i.i.d. training samples).

These examples were provided for the minimal "superior" $T = 20$ ($T \geq N$) number of training samples that demonstrated that it has practically the same performance as the original ($T = 20$) and the "doubled" by (45) number of training data. For the truly i.i.d. training samples, the gains are more significant. Yet, the proposed technique allows for the LR maximizing the conventional likelihood for scenarios with $T \geq N/2$ using the same LR maximization technique and growing the maximum with an increased sample volume $T > N/2$. For $T > N$, this way of doubling the volume of training data does not improve ML estimation performance. This was demonstrated in the scenario with the N = 17 ULA, three sources, and N = 85.

Finally, let us get back to the properties of the likelihood ratio (12) for the case of a very small sample support ($T \ll N$). The expected likelihood ratio of the true covariance matrix $\left(\widehat{\mathbf{T}}_N = \mathbf{T}_N\right)$ is equal to

$$\mathrm{LR}\left(\widehat{\mathbf{R}}_N|\widehat{\mathbf{T}}_N\right) = \frac{\det\left[\,\mathbb{X}_T^{\mathrm{H}}\widehat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right]}{\left[\frac{1}{T}\,\mathrm{Tr}\left(\mathbb{X}_T^{\mathrm{H}}\widehat{\mathbf{T}}_N^{-1}\mathbb{X}_T\right)\right]^{\mathrm{T}}},$$
$$\mathbb{X}_T^{\mathrm{H}}\widehat{\mathbf{T}}_N^{-1}\mathbb{X}_T = \mathfrak{J}_T\mathfrak{J}_T^{\mathrm{H}} = \sum_{t=1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{\mathrm{H}},\quad \mathfrak{J}_T \in \mathbb{C}^{TxN},$$
$$\mathbb{X}_T^{\mathrm{H}}\widehat{\mathbf{T}}_N^{-1}\mathbb{X}_T = \mathfrak{J}_T\mathfrak{J}_T^{\mathrm{H}} = CW(N'). \tag{47}$$

$\mathbb{X}_T^{\mathrm{H}}\widehat{\mathbf{T}}_N^{-1}\mathbb{X}_T$ is a (TxT)-variate random matrix with a complex Wishart distribution averaged over N/T vectors. The distribution of this "expected likelihood" was found by R.C.S. Pillai and B.N. Nagarsenker [27]. While analytical formulas could be applied for the calculations, the existence of the zonal polynomial and Meijer's G-function $W_{N,N}^{N,0}\left(W\left|\begin{matrix}a_1,\ldots,a_N\\ NT+k+Nh\end{matrix}\right.\right)$ significantly complicates such computation.

Finally, let us explore the limits of the efficient ML DOA estimation. Let us introduce an example with the same three sources, directions of arrivals (-30°, 0°, 1°), and the same SNR per source of 10 dB, with a very limited training sample support T = 7 such that even its "doubling" does not exceed the antenna dimension N = 17 but by a very small margin. In general, this example illustrates the decay in maximum likelihood performance, despite the LR values exceeding those of the true covariance matrix $\mathbf{T}_N$ in every trial.

Of course, the conventional MUSIC, applied to the "doubled" sample matrix, failed to resolve the DOAs of the second and third sources in any of the 1,000 trials. The LR maximization in this scenario provided a successful resolution of the three acting sources in 300 out of 1,000 trials. While only in 300 trials were the close-in sources successfully resolved, in the remaining 700 trials, the close-in second and third DOAs were not resolved. For the 300 resolved sources, the standard deviation was 1.08° and the mean error modulus was 0.299°.

Let us once again state that the likelihood exceeded that of the true covariance matrix in all 1,000 trials, but this did not imply proper DOA estimation.

6. Conclusions and Recommendations

The analysis of the numerical maximization of the likelihood ratio for a p.d. Toeplitz covariance matrix revealed several properties that we believe are important from both theoretical and practical perspectives.

In particular, we demonstrated that numerical LR maximization, utilizing Vandermonde parameterization, applied to random initial Toeplitz matrices, can often yield an inappropriate solution with negative eigenvalues. Yet, in all cases where the MATLAB fmincon routine converges to a positive definite Toeplitz matrix, the process converges to the same solution as if initiated by a true Toeplitz covariance matrix, irrespective of the initial matrix. Our numerous attempts to initiate the fmincon routine with matrices far from the true covariance matrix with LR values < 10^{-20} (!), nevertheless, finally converged to the same solution as the one initiated by the true covariance matrix.

The obtained calculation results cannot be treated as proof of global optimality. Yet, the substantial volume of optimization results—obtained from a variety of random initial approximations that converged to the same solution—allows for the cautious conjecture that the optimization problem, as formulated here, is convex and that the identified optimum is global.

Naturally, this assertion requires rigorous mathematical proof. Indeed, the fact that—starting from the true covariance matrix—we converge to the very same solution serves as a significant argument in support of such a proof. A further argument in favor of this proof is that this solution yields a likelihood function value that exceeds that of the true covariance matrix and coincides with the value attained when the process is initialized with the true Toeplitz covariance matrix.

Note that the finite calculation accuracy leads to some insignificant variations of the optimum solution, and this effect needs to be closely monitored for larger ULA dimensions (N > 17) and eigenvalue spread (> 10^{10}).

The proposed two-step optimization starts from the Carathéodory parameterization and, upon convergence, switches to the Vandermonde parameterization. While the maximized LR increases during this second stage of optimization, the proximity ratio relative to the true matrix decreases.

The numerical analysis discussed above was conducted only for N = 17; therefore, the revealed properties should be closely monitored for uniform linear arrays with larger apertures and larger dynamic range, λ_1/λ_N. This particular sequence of Toeplitz matrix parameterizations was shown to be appropriate for the LR maximization of Toeplitz matrices when the Toeplitz matrix has a known a priori number of noise subspace eigenvalues. We demonstrated that the Toeplitz p.d. matrices with a known number of noise subspace eigenvalues cannot be optimized by the fmincon routine.

Specifically, such an LR maximization stops before the LR of the true covariance matrix is reached. For this reason, the problem of LR maximization of the p.d. Toeplitz matrix with a priori known finite signal subspace dimension is proposed to be solved in two steps. First, the problem of maximum likelihood Toeplitz matrix estimation is resolved with no control over the reduction of the signal subspace dimension. In the second stage, for the resulting p.d. Toeplitz matrix with a (globally) maximum-likelihood ratio and distinct eigenvalues, the second problem of noise subspace eigenvalue equalization is addressed using the method of alternating projections. For example, we demonstrated that noise eigenvalue equalization somewhat reduces the maximum likelihood ratio of the final optimized Toeplitz matrices. However, the remaining likelihood ratio still exceeds that of the true Toeplitz covariance matrix.

Moreover, the proposed two-stage version of the likelihood ratio maximization procedure enabled us to significantly expand the method's scope of applicability— encompassing sources situated closer to one another, as well as weaker sources—and to successfully perform their separation and the optimal estimation of their directions of arrival, whereas the standard MUSIC algorithm, when applied to the Hermitian (sample) maximum likelihood matrix, fails to handle this task [28-52].

References

1. Abramovich, Y., Abramovich, V., & Pongsiri, T. (2025). Maximum Likelihood Toeplitz Covariance Matrix Estimation: Is It a Convex Optimization Problem?. *Authorea Preprints*.
2. Abramovich, Y., Abramovich, V., & Pongsiri, T. (2026). Numerical Techniques for the Maximum Likelihood Toeplitz Covariance Matrix Estimation: Part II. Hermitian Toeplitz Matrices. *IEEE Transactions on Aerospace and Electronic Systems, 62,* 3971-3990.
3. Burg, J. P., Luenberger, D. G., & Wenger, D. L. (1982). Estimation of structured covariance matrices. *Proceedings of the IEEE, 70*(9), 963-974.
4. Fuhrmann, D. R. (1991). Application of Toeplitz covariance estimation to adaptive beamforming and detection. *IEEE Transactions on Signal Processing, 39*(10), 2194-2198.
5. Fuhrmann, D. R., & Barton, T. A. (1990, October). Estimation of block-Toeplitz covariance matrices. In *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers, 1990.* (Vol. 2, p. 779). IEEE.
6. Robey, F. C. (1990). *A covariance matrix modeling approach to adaptive beamforming and detection* (Doctoral dissertation, PhD dissertation, Washington Univ., St. Louis, MO).
7. Fuhrmann, D. R., Turmon, M. J., & Miller, M. I. (1988, March). Efficient implementation of the EM algorithm for Toeplitz covariance estimation. In Proceedings of the Twenty-Second Annual Conference on Information Sciences and Systems.
8. Miller, M. I., Fuhrmann, D. R., O'Sullivan, J. A., & Snyder, D. L. (1991). Maximum-likelihood methods for Toeplitz covariance estimation and radar imaging. *Advances in Spectrum Analysis and Array Processing, 2*, 145-172.
9. Turmon, M. J., Miller, M. I., Snyder, D. L., & O'Sullivan, J. A. (1988, August). Performance evaluation of maximumlikelihood Toeplitz covariance estimates generated using the EM algorithm. In Proc. Fourth ASSP Workshop on Spect. Est. and Modeling (pp. 182-185).
10. Turmon, M. J., & Miller, M. I. (2002). Maximum-likelihood estimation of complex sinusoids and Toeplitz covariances. *IEEE transactions on signal processing, 42*(5), 1074-1086.
11. Gray, D. A., Anderson, B. D. O., & Sim, P. K. (1987). Estimation of structured covariances with application to array beamforming. *Circuits, Systems and Signal Processing, 6*(4), 421-447.
12. Vinogradova, J., Couillet, R., & Hachem, W. (2015). Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection. *IEEE Transactions on Signal Processing, 63*(18), 4903-4913.
13. Cai, T. T., Ren, Z., & Zhou, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probability Theory and Related Fields, 156*(1), 101-143.
14. Vallet, P., & Loubaton, P. (2014, May). Toeplitz rectification and DOA estimation with music. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2237-2241). IEEE.
15. Abramovich, Y. I., & Spencer, N. K. (2005, May). Expected-likelihood covariance matrix estimation for adaptive detection. In *IEEE International Radar Conference, 2005*. (pp. 623-628). IEEE.
16. Abramovich, Y. I., & Johnson, B. A. (2007, December). Expected likelihood estimation: Asymptotic properties for" stochastic" complex Gaussian models. In *2007 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (pp. 33-36). IEEE.
17. Abramovich, Y. I., Arov, D. Z., Kachur, V. G. (1988). Adaptive cancellation filters for stationary interference with a Toeplitz correlation matrix. *J. Comm. Tech. Elect., 33*(4), 54-61.
18. Aubry, A., Babu, P., De Maio, A., & Rosamilia, M. (2024). Advanced methods for MLE of Toeplitz structured covariance matrices with applications to RADAR problems. *IEEE Transactions on Information Theory, 70*(12), 9277-9292.
19. Yang, Z., Chen, X., & Wu, X. (2023). A robust and statistically efficient maximum-likelihood method for DOA estimation using sparse linear arrays. *IEEE Transactions on Aerospace and Electronic Systems, 59*(5), 6798-6812.
20. Cederberg, D. (2024). Toeplitz covariance estimation with applications to MUSIC. *Signal Processing, 221*, 109506.
21. Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*. John Wiley & Sons.
22. Abramovich, Y. I., Spencer, N. K., & Gorokhov, A. Y. (2004). Bounds on maximum likelihood ratio-Part I: Application to antenna array detection-estimation with perfect wavefront coherence. *IEEE transactions on signal processing, 52*(6), 1524-1536.
23. MathWorks. (2024). *Problem when using fmincon: The solver starts from a feasible point but converges to an infeasible point.*
24. Kolosov, A. B. (1987). *Over-the-Horizon Radar.* MA, Artech House.
25. Valmarde, J. (2015). *Generating symmetric positive definite Toeplitz matrices.*
26. Skolnik, M., Nemhauser, G., & Sherman, J. (1964). Dynamic

programming applied to unequally spaced arrays. *IEEE Transactions on antennas and propagation, 12*(1), 35-43.
27. Nagarsenker, B. N., & Pillai, K. C. S. (1973). The distribution of the sphericity test criterion. *Journal of Multivariate Analysis, 3*(2), 226-235.
28. Horn, R. A., Johnson, C. R. (1985). C. Johnson, Matrix Analysis. Cambridge.
29. Abramovich, Y., Danilov, B.G. (1976). The use of dynamic programming for the synthesis of array antennas with discrete phase shifters. *Radio Eng. Elect. Phys., 21*(1), 69-74.
30. Paulraj, A., & Kailath, T. (1985, April). Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase. In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 10, pp. 640-643). IEEE.
31. Weiss, A., & Yeredor, A. (2020). Asymptotically optimal blind calibration of uniform linear sensor arrays for narrowband Gaussian signals. *IEEE Transactions on Signal Processing, 68,* 5322-5333.
32. Weiss, A., & Yeredor, A. (2021, June). Enhanced blind calibration of uniform linear arrays with one-bit quantization by kullback-leibler divergence covariance fitting. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4625-4629). IEEE.
33. Chu, M. T. (1993). The stability group of symmetric Toeplitz matrices. *Linear algebra and its applications, 185,* 119-123.
34. MathWorks. *fmincon - Nonlinear Optimization.*
35. Chu, M. T., Golub, G. H. (2002). Inverse eigenvalue problems. Oxford Science Publications, 1-71.
36. Shan, T. J., & Kailath, T. (1985). Adaptive beamforming for coherent signals and interference. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 33*(3), 527-536.
37. Paulraj, A., Reddy, V. U., Shan, T. J., & Kailath, T. (1986, October). Performance analysis of the MUSIC algorithm with spatial smoothing in the presence of coherent sources. In *MILCOM 1986-IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's* (Vol. 3, pp. 41-45). IEEE.
38. Linebarger, D. A., & Johnson, D. H. (2002). The effect of spatial averaging on spatial correlation matrices in the presence of coherent signals. *IEEE transactions on acoustics, speech, and signal processing, 38*(5), 880-884.
39. Indukumar, K. C., & Reddy, V. U. (1992). A note on redundancy averaging. *IEEE Transactions on signal processing, 40*(2), 466-469.
40. Shan, T. J., Wax, M., & Kailath, T. (1985). On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 33*(4), 806-811.
41. Doron, M. A., & Weiss, A. J. (2002). Performance analysis of direction finding using lag redundancy averaging. *IEEE transactions on signal processing, 41*(3), 1386-1391.
42. Turmon, M. J. (1994, June). *Cramér-Rao Bounds for Toeplitz Covariance Estimation.*
43. Consul, P. C. (1969). The exact distribution of likelihood criterion of different hypotheses. In Multivariate Analysis II, P.R. Krishnain, Ed., Academic, 171-181.
44. Gradshteyn, L. S., Ryzhik, L. M. (2000). Tables of integrals, series, and products. Sixth Edition, Academic Press.
45. Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory, 54*(11), 5113-5129.
46. MathWorks. *linprog – solve linear programming problems*.
47. Wu, W. B., & Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica,* 1755-1768.
48. Chun, J., & Kailath, T. (1989). A constructive proof of the Gohberg-Semencul formula. *Linear Algebra and its Applications, 121,* 475-489.
49. Evans, J. E. (1981). High resolution angular spectrum estimation technique for terrain scattering analysis and angle of arrival estimation. In 1st *IEEE ASSP Workshop Spectral Estimat., McMaster Univ., Hamilton, Ont., Canada, 1981* (pp. 134-139).
50. Greenbaum, A., Li, R. C., & Overton, M. L. (2020). First-order perturbation theory for eigenvalues and eigenvectors. *SIAM review, 62*(2), 463-482.
51. Abramovich, Y., Abramovich, V., Pongsiri, T. (2025). Numerical Techniques for the Maximum Likelihood Toeplitz Covariance Matrix Estimation: Part I. Symmetric Toeplitz Matrices. *J Electr Comput Innov, 2*(2), 01-21.
52. Abramovich, Y. I., & Johnson, B. A. (2008). GLRT-based detection-estimation for undersampled training conditions. *IEEE Transactions on Signal Processing, 56*(8), 3600-3612.