# Maximum Likelihood Structured Covariance Matrix Estimation and connections to SBL: A Path to Gridless DoA Estimation

Rohan R. Pote
*ECE Department, UC San Diego*

Bhaskar D. Rao
*ECE Department, UC San Diego*

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10051945

***Abstract*—In this paper, we revisit the framework for maximum likelihood estimation (MLE) as applied to parametric models with an aim to estimate the parameter of interest in a *gridless* manner. The approach has inherent connections to the sparse Bayesian learning (SBL) formulation, and naturally leads to the problem of *structured* matrix recovery (SMR). We therefore pose the parameter estimation problem as a SMR problem, and recover the parameter of interest by appealing to the Carathéodory-Fejér result on decomposition of positive semi-definite Toeplitz matrices. We propose an iterative algorithm to estimate the structured covariance matrix; each iteration solves a semi-definite program. We numerically compare the performance with other gridless schemes in literature and demonstrate the superior performance of the proposed technique.**

***Index Terms*—Maximum likelihood, structured matrix recovery, sparse Bayesian learning, coarrays, direction-of-arrival estimation, superresolution**

## I. Introduction

Consider the following parametric data model

$$\mathbf{y}_l = \mathbf{\Phi}_{\boldsymbol{\theta}}\mathbf{x}_l + \mathbf{n}_l, \qquad 0 \leq l < L, \tag{1}$$

where $\mathbf{y}_l \in \mathbb{C}^M$ denotes the measurements, and $L$ denotes the total number of snapshots available. The $k$th column of $\mathbf{\Phi}_{\boldsymbol{\theta}} \in \mathbb{C}^{M\times K}$ is a known vector function of the parameter $\theta_k, k \in \{1,\ldots,K\}$, and the parameter itself lies in some known continuous domain. $K$ denotes the number of sources. The sources' signal $\mathbf{x}_l \in \mathbb{C}^K$ and noise $\mathbf{n}_l \in \mathbb{C}^M$ are independent of each other, and i.i.d. over time. The noise, $\mathbf{n}_l$, is distributed as $\mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$. In (1), the parameters $(\boldsymbol{\theta}, \mathbf{x}_l, \sigma_n^2)$ are the unknowns. The model parameters affect the measurements in a non-linear manner, which makes the inverse problem extremely difficult to solve, even in the absence of noise. The above problem is ubiquitous, with applications including biomagnetic imaging [1], functional approximations [2], and echo cancellation [3]. In this work we are concerned with problems such as in line spectral estimation and direction-of-arrival (DoA) estimation [4] for narrowband signals; we emphasize the latter as means for exposition.

Approaches to solve (1) have rich history and can broadly be classified as traditional vs. modern, both significant in insights and contributions. Traditional approaches can be further classified into spectral based [5]–[7] and parametric methods [4]. The typical ingredients to solve (1) include geometrical[1] and statistical properties of the model in (1). Table-I summarizes

| | Methods | Primary Bottleneck [4] |
|---|---|---|
| (a) | i. Spatial filtering (beamforming) | Aperture/ degrees of freedom |
| | ii. Subspace based methods | *Number of snapshots* |
| (b) | Deterministic/ Stochastic MLE | Model & computational complexity |

(a) Spectral based methods (b) Parametric methods
TABLE I

these methods [4]. Most modern techniques to solve (1), under the rubric of sparse signal recovery (SSR), explicitly impose sparsity and recover the parameter of interest in either grid-based or grid-*less* manner. The emphasis there is on optimizing an appropriate fit to the measurements with an additional (sparsity) regularizer [8]–[11]. These methods are therefore sensitive to setting the regularization parameter properly. An exception to the regularization based methods includes sparse Bayesian learning (SBL) [12]–[14] which formulates the recovery problem under the MLE framework. The approach recovers sparse solutions via implicit regularization [15].

In this work, we revisit the traditional and modern MLE based approaches, with an aim to recover $\boldsymbol{\theta}$ in a *gridless* manner. We identify the following contributions: a) Reformulating the SBL problem as a novel *structured matrix recovery* (SMR) problem *under the MLE framework*; the structure is influenced by appropriate *geometry* and *prior* b) Propose an iterative algorithm to optimize the novel SBL cost function using a semi-definite program (SDP) c) Providing perspectives to understand the proposed approach and connect with the traditional MLE framework and the modern SBL formulation. We provide numerical results to further elucidate the impact of the proposed reformulation, namely the ability to go *gridless*, and compare the proposed technique with other gridless approaches in literature. Notations: $\odot$ denotes Khatri-Rao product, $(.)^c$ denotes element-wise complex conjugate, $(.)^T$ denotes transpose, $(.)^H$ denotes the Hermitian operation.

## II. Reformulating SBL as a SMR Problem

We begin by highlighting a simple insight for SBL when applied to (1) in applications such as line spectral or DoA estimation. The goal is to take this insight further and enable gridless DoA estimation. For the purpose of simplicity, we

This research was supported by NSF under Grant CCF-2124929 and Grant CCF-2225617, and ONR under Grant No. N00014-18-1-2038.

[1] E.g., subspace orthogonality in MUltiple SIgnal Classification (MUSIC).

focus on the Uniform Linear Array (ULA) geometry in this section, and postpone the general case of ULAs with *missing sensors* until next section. However, the insights presented here are applicable to the general case as well.

Consider a ULA with $M$ sensors and $d = \bar{\lambda}/2$ distance between adjacent sensors to prevent ambiguity in DoA estimation; $\bar{\lambda}$ denotes the wavelength of the incoming narrowband source signals. The array manifold matrix for $K$ source signals incoming at angle $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^T, \theta_k \in [-\frac{\pi}{2}, \frac{\pi}{2})$, is given by $\boldsymbol{\Phi}_{\boldsymbol{\theta}} = [\boldsymbol{\phi}(\theta_1), \ldots, \boldsymbol{\phi}(\theta_K)]$, where $\boldsymbol{\phi}(\theta_k) = [1, \exp(-j\pi \sin\theta_k), \ldots, \exp(-j(M-1)\pi \sin\theta_k)]^T$.

For the problem at hand, SBL first discretizes the possible values of $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$ and introduces the measurement matrix $\boldsymbol{\Phi} \in \mathbb{C}^{M \times G}$, where $G$ denotes the grid size. It further imposes a *parameterized* Gaussian prior on the source signal $\bar{\mathbf{x}}_l \in \mathbb{C}^G$ as $\bar{\mathbf{x}}_l \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma})$. Note that SBL explicitly imposes an *uncorrelated sources* prior, and thus $\boldsymbol{\Gamma}$ is a diagonal matrix; let $\mathrm{diag}(\boldsymbol{\Gamma}) = \boldsymbol{\gamma}$. Under the SBL formulation, the original problem in (1) now becomes

$$\mathbf{y}_l = \boldsymbol{\Phi}\bar{\mathbf{x}}_l + \bar{\mathbf{n}}_l, \qquad 0 \leq l < L, \tag{2}$$

and $\mathbf{y}_l \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I})$; $\lambda$ denotes the estimate for noise variance. Under the MLE framework, the hyperparameter $\boldsymbol{\Gamma}$ and $\lambda$ can be estimated as [14]

$$\min_{\substack{\boldsymbol{\Gamma}\succeq\mathbf{0},\\ \lambda\geq 0}} \log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right) + \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right), \tag{3}$$

where $\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l \mathbf{y}_l^H$ denotes the sample covariance matrix (SCM). Choices for solving the above problem for SBL include the Tipping iterations [12], EM iterations [13], etc.

*Remark* 1. Note that if the number of sources $K$ is known exactly in (1), such model order information is not used in the SBL formulation. Instead, the $\log\det$ penalty in (3) helps to promote sparsity, and deal with small but unknown number of sources. If there is prior knowledge on $K$, then $\|\boldsymbol{\gamma}\|_0 = K$ would have to be imposed on the objective function.

We now present the following useful insight.

*Proposition* 1. $\forall \boldsymbol{\gamma} \geq \mathbf{0}$ such that $(\boldsymbol{\Phi} \odot \boldsymbol{\Phi}^c)\boldsymbol{\gamma} = \mathbf{w}$, for some fixed $\mathbf{w} \in \mathbb{C}^{M^2}$, the SBL cost is a constant i.e.,

$$\log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right) + \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) = C(\lambda),$$

where $C(\lambda)$ is some constant.

*Proof.* The proof follows simply by observing that $(\boldsymbol{\Phi} \odot \boldsymbol{\Phi}^c)\boldsymbol{\gamma} = \mathbf{w}$ implies $\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H$ is a fixed *structured* matrix with entries dictated by components of $\mathbf{w}$. □

The above result demonstrates that, the hyperparameter $\boldsymbol{\gamma}$ affects the SBL cost function only through the entries of the *structured covariance matrix* of the measurements.

The structure for $\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H$ in the case of ULA is a Toeplitz matrix, and is informed by the array geometry and the uncorrelated sources' correlation prior. In other words, SBL attempts to find the 'best' positive semidefinite (PSD) Toeplitz matrix approximation to the SCM $\hat{\mathbf{R}}_{\mathbf{y}}$. We use this insight and reparameterize the SBL cost function to directly estimate the entries of the Toeplitz covariance matrix. Let $\mathbf{v}$ denote the first row of such a Toeplitz matrix, denoted by $\mathrm{Toep}(\mathbf{v})$. We reformulate the SBL optimization problem to get

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^M \text{ s.t.}\\ \mathrm{Toep}(\mathbf{v})\succeq\mathbf{0}, \lambda\geq 0}} \log\det\left(\mathrm{Toep}\left(\mathbf{v}\right) + \lambda\mathbf{I}\right) + \mathrm{tr}\left(\left(\mathrm{Toep}(\mathbf{v}) + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{4}$$

Once the solution $\mathbf{v}^*$ is obtained, we estimate the DoAs by decomposing the Toeplitz matrix, $\mathrm{Toep}(\mathbf{v}^*)$. In our simulations we use root-MUSIC to estimate the DoAs [16].

*Remark* 2. It is known that a low rank ($D < M$) PSD Toeplitz matrix such as $\mathrm{Toep}(\mathbf{v}^*)$ can be uniquely decomposed as $\mathrm{Toep}(\mathbf{v}^*) = \sum_{i=1}^{D} p_i \boldsymbol{\phi}(\theta_i)\boldsymbol{\phi}(\theta_i)^H, p_i > 0$ and $\theta_i$ are distinct [17]. In (4), a low-rank solution is encouraged by the $\log\det$ term [18], while its effect is being moderated by the additional noise variance term, '$+\lambda\mathbf{I}$'.

Note that the SBL formulation in (3) not only finds a structured matrix fit to the measurements, it also *factorizes* it. The same is true with the classical MLE approach as well, and is discussed in Section IV. The structured matrix factorization is a crucial step. In the proposed approach, we find a structured matrix in the MLE sense. The factorization, if unique, yields the MLE estimate of $\boldsymbol{\theta}$ under the invariance property of the MLE. We therefore refer to the proposed approach in this paper as 'StructCovMLE'. The problem in (4) is non-convex. We discuss an iterative algorithm to solve (4) and extension to allow ULAs with missing sensors, in the next section.

## III. Proposed Algorithm

We assume that the noise variance is known and set $\lambda = \sigma_n^2$ in (4), but it can be estimated as well, similar to $\mathbf{v}$ in this section.

### A. Uniform Linear Array Geometry

We majorize the $\log\det$ term in (4) and replace it with a linear term using its Taylor expansion

$$\log\det(\mathrm{Toep}(\mathbf{v}) + \lambda\mathbf{I}) \leq \log\det\left(\mathrm{Toep}(\mathbf{v}^{(k)}) + \lambda\mathbf{I}\right) + \mathrm{tr}\left((\mathrm{Toep}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathrm{Toep}(\mathbf{v} - \mathbf{v}^{(k)})\right), \tag{5}$$

where $\mathbf{v}^{(k)}$ denotes the iterate value at the $k$th iteration. Note that the linear term from Taylor expansion provides a supporting hyperplane to the hypograph $\{(\mathbf{v}, t) : t <= \log\det(\mathrm{Toep}(\mathbf{v}) + \lambda\mathbf{I})\}$. We ignore the constant terms above and get the following majorized objective function

$$\mathrm{tr}\left((\mathrm{Toep}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\,\mathrm{Toep}(\mathbf{v})\right) + \mathrm{tr}\left((\mathrm{Toep}(\mathbf{v}) + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{6}$$

The resulting optimization problem is convex and can be formulated as a SDP using Schur complement lemma as

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^M,\\ \mathbf{U}\in\mathbb{C}^{M\times M}}} \mathrm{tr}\left((\mathrm{Toep}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathrm{Toep}(\mathbf{v})\right) + \mathrm{tr}\left(\mathbf{U}\,\hat{\mathbf{R}}_{\mathbf{y}}\right)$$

$$\text{subject to} \begin{bmatrix} \mathbf{U} & \mathbf{I}_M \\ \mathbf{I}_M & \mathrm{Toep}(\mathbf{v}) + \lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \mathrm{Toep}(\mathbf{v}) \succeq \mathbf{0}, \tag{7}$$

and can be solved using any standard solvers (e.g. CVX solvers such as SDPT3, SeDuMi). It can be solved iteratively and we summarize the proposed steps in Algorithm 1.

**Algorithm 1:** Proposed 'StructCovMLE' Algorithm

**Result:** $\mathbf{v}^*$
**Input:** $\mathbf{Y} = [\mathbf{y}_0, \ldots, \mathbf{y}_{L-1}], \lambda = \sigma_n^2, \text{ITER}$

```
Initialize: R̂_y = YY^H/L, v* = e_1 = [1, 0, ..., 0]^T
for k := 1 to ITER do
    v^(k) ← v*
    v* ← Solve the problem in (7)
end
```

### B. ULA with Missing Sensors

We begin by identifying the relevant *structure* for the more general case of ULAs with missing sensors. Consider a linear array with $M$ sensors on a grid with minimum inter-element spacing $d = \bar{\lambda}/2$. Let $\mathbb{P} = \{p_i \mid p_i \in \mathbb{Z}, 0 \leq i < M\}$ denote the set of normalized (w.r.t. $d$) sensor positions. We assume $p_0 = 0$ without loss of generality. The array manifold vector is given by $\boldsymbol{\phi}(\theta) = [1, \exp(-jp_1\pi\sin\theta), \ldots, \exp(-jp_{M-1}\pi\sin\theta)]^T$. The difference coarray is given by $\mathbb{D} = \{z \mid z = r - s,\ r, s \in \mathbb{P}\}$. The concept of difference coarray influences the structure we seek to identify, and also arises naturally when computing the received signal covariance matrix. The latter under the SBL formulation is given by $\mathbf{\Phi\Gamma\Phi}^H + \lambda\mathbf{I}$, as discussed previously. The $(m, n)$ entry in $\mathbf{\Phi\Gamma\Phi}^H$ is given by $[\mathbf{\Phi\Gamma\Phi}^H]_{m,n} = \sum_{i=1}^{G} \gamma_i \exp(-j(p_m - p_n)\pi\sin\theta_i)$, and $[\mathbf{\Phi\Gamma\Phi}^H]_{m,n} = [\mathbf{\Phi\Gamma\Phi}^H]_{n,m}^c$. Thus, $[\mathbf{\Phi\Gamma\Phi}^H]_{m,n} = [\mathbf{\Phi\Gamma\Phi}^H]_{m',n'}, \forall$ tuples $(m, n)$ and $(m', n')$ such that $p_m - p_n = p_{m'} - p_{n'}$. In other words, the entries in $\mathbf{\Phi\Gamma\Phi}^H$ can be distinct only corresponding to distinct elements in $\mathbb{D}$. $\mathbf{\Phi\Gamma\Phi}^H$ is Hermitian symmetric, which further restricts the number of distinct entries. This reveals the underlying *structure* that the model $\mathbf{\Phi\Gamma\Phi}^H$ satisfies, and we formalize it below.

Let $M_{\text{apt}}$ denote the aperture of the array. We can write $M_{\text{apt}} = \max_{d\in\mathbb{D}} d + 1$. We define a linear mapping $\mathbf{T}(\mathbf{v}) : \mathbb{C}^{M_{\text{apt}}} \to \mathbb{C}^{M\times M}$ given by

$$[\mathbf{T}(\mathbf{v})]_{i,j} = \begin{cases} v_{|p_i - p_j|} & j \geq i \\ v^c_{|p_i - p_j|} & \text{otherwise} \end{cases}, 0 \leq i, j < M. \tag{8}$$

Note that the mapping $\mathbf{T}(\mathbf{v})$ in general is many-to-one. It is only when the difference coarray has no holes, the mapping is one-to-one. For such cases we define $\mathbf{T}^{-1}(\mathbf{R}) : \mathbb{C}^{M\times M} \to \mathbb{C}^{M_{\text{apt}}}$ as a function that extracts the entries of a given structured matrix $\mathbf{R}$, formed using (8), to form a column vector. For the ULA with no missing sensors' case, we have $\mathbf{T}(\mathbf{v}) = \text{Toep}(\mathbf{v})$.

Thus, for the general case, (3) can be reformulated as:

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^{M_{\text{apt}}} \text{ s.t.} \\ \text{Toep}(\mathbf{v})\succeq\mathbf{0}, \lambda\geq 0}} \log\det(\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I}) + \text{tr}\left((\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{9}$$

*Remark* 3. We would like to highlight a non-trivial choice made above of imposing $\text{Toep}(\mathbf{v}) \succeq \mathbf{0}$, instead of only requiring $\mathbf{T}(\mathbf{v}) \succeq \mathbf{0}$. Note that the former constraint ensures that the latter is satisfied. The choice imposes a relevant constraint and is an important aspect of the model we wish to fit to the data in MLE sense. It also helps to connect the proposed reformulation to the traditional and modern MLE approaches, and is discussed in Section IV.

*Remark* 4. As in the case for SBL, if the number of sources, $K$, is known, a rank constraint $\text{rank}(\text{Toep}(\mathbf{v})) = K$ should be imposed. Since imposing a rank constraint is difficult, surrogate measures like in compressed sensing may be used, such as '$+\beta\log\det(\text{Toep}(\mathbf{v}) + \epsilon\mathbf{I})$' as a regularizer in (9) to further promote sparse solutions. In this work, we do not exploit knowledge of $K$ to solve (9).

Like in the previous case, we majorize the cost function in (9) to get a convex function and rewrite it as a SDP, assuming knowledge of noise variance and setting $\lambda = \sigma_n^2$. The majorized objective is given by

$$\text{tr}\left((\mathbf{T}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathbf{T}(\mathbf{v})\right) + \text{tr}\left((\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{10}$$

The resulting SDP is given below

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^{M_{\text{apt}}}, \\ \mathbf{U}\in\mathbb{C}^{M\times M}}} \text{tr}\left((\mathbf{T}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathbf{T}(\mathbf{v})\right) + \text{tr}\left(\mathbf{U}\,\hat{\mathbf{R}}_{\mathbf{y}}\right) \tag{11}$$

$$\text{subject to} \begin{bmatrix} \mathbf{U} & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{T}(\mathbf{v}) + \lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \text{Toep}(\mathbf{v}) \succeq \mathbf{0},$$

where $\mathbf{v}^{(k)}$ denotes the value at the $k$th iteration. Steps similar to Algorithm 1 can be followed to find the optimal point $\mathbf{v}^*$. To estimate the DoAs we perform root-MUSIC on $\mathbf{T}(\mathbf{v}^*)$.

*Remark* 5. It was shown in [19] that sparse arrays with a larger number of consecutive lags than the number of sensors, $M$, can identify more sources than $M$. Under the proposed approach, a similar higher identifiability can be achieved by instead performing root-MUSIC on $\text{Toep}(\mathbf{v}^*)$, and we numerically verify this in section V.

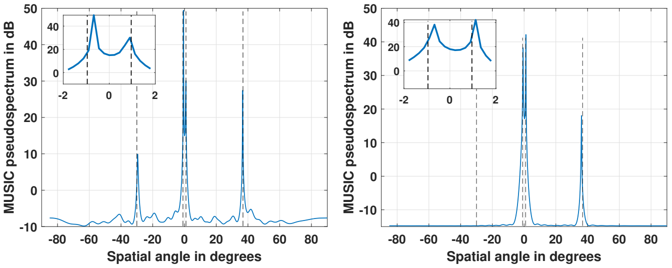
## IV. On Proposed Method: From MLE to Sparse Bayesian Learning

In this section, we aim at connecting the classical MLE framework and the grid SBL formulation with the proposed technique. The hope is to answer the following question: *how has the reparameterization affected the original problem in (1) of solving for $\boldsymbol{\theta}$?*

### A. Connection with the classical MLE formulation

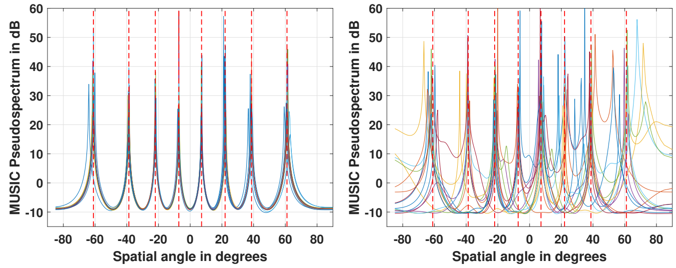
We begin by first stating the traditional MLE formulation. In this approach, we impose a *parametrized* Gaussian prior on $\mathbf{x}_l$ i.e., $\mathbf{x}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{P})$. Note that an explicit knowledge of model order information is a requisite here. We further assume that the sources are uncorrelated, and thus $\mathbf{P}$ is a diagonal matrix. The resulting optimization problem is given by

$$\min_{\substack{\boldsymbol{\theta}\in[-\frac{\pi}{2},\frac{\pi}{2})^K, \mathbf{P}\succ\mathbf{0}, \\ \lambda\geq 0}} \log\det\left(\mathbf{\Phi}_{\boldsymbol{\theta}}\mathbf{P}\mathbf{\Phi}_{\boldsymbol{\theta}}^H + \lambda\mathbf{I}\right) + \text{tr}\left((\mathbf{\Phi}_{\boldsymbol{\theta}}\mathbf{P}\mathbf{\Phi}_{\boldsymbol{\theta}}^H + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right), \tag{12}$$




(a) StructCovMLE (b) RAM

Fig. 1. Nested Array: Sensor locations, $\mathbb{P} = \{0, 1, 2, 3, 4, 5, 11, 17, 23, 29\}$.




(a) StructCovMLE (b) RAM

Fig. 2. Nested Array: $M = 6, K = 8, L = 4$, SNR= 20 dB, 15 realizations

The model is also referred to as the *unconditional model* in the DoA literature [20], compared to the *conditional model* where $\mathbf{x}_l$ is assumed deterministic. Consider the following updated MLE optimization problem:

$$\min_{\substack{K\in\mathbb{Z}^+\\0<K<M_{\text{apt}}}} \min_{\substack{\boldsymbol{\theta}\in[-\frac{\pi}{2},\frac{\pi}{2})^K,\\ \mathbf{P}\succ\mathbf{0},\lambda\geq 0}} \text{logdet}\left(\boldsymbol{\Phi}_{\boldsymbol{\theta}}\mathbf{P}\boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda\mathbf{I}\right) + \text{tr}\left((\boldsymbol{\Phi}_{\boldsymbol{\theta}}\mathbf{P}\boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \quad (13)$$

The difference with the traditional MLE formulation is that, in the above we consider all non-zero model orders such that $K < M_{\text{apt}}$, to optimize the cost function. We then have the following result.

*Theorem* 1. The problem in (9) and in (13) are equivalent, in that they achieve the same globally minimum cost.

*Proof.* Proof is provided in the appendix. □

### *B. Connection with SBL in (3)*

Consider the following updated SBL optimization problem:

$$\min_{\boldsymbol{\Phi}} \min_{\boldsymbol{\Gamma}\succeq\mathbf{0},\lambda\geq 0} \log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right) + \text{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \quad (14)$$

Like in the MLE case, here for SBL we allow all possible dictionaries $\boldsymbol{\Phi}$ with array manifold vectors as columns, to optimize the cost function. The following result follows similarly.

*Theorem* 2. The problem in (9) and in (14) are equivalent, in that they achieve the same globally minimum cost.

*Proof.* The proof follows similarly as for Theorem 1. □

The above results help to understand the proposed approach in (9): (9) estimates a structured covariance matrix fit to the measurements in the MLE sense over all model orders for classical MLE or all appropriate dictionaries for SBL.

The entries of a structured matrix and noise variance may be combined as presented in [21]. However, the choice of explicitly involving $\lambda$ parameter has two important consequences: a) In the case when $\sigma_n^2$ is known, the proposed approach allows a mechanism to feed this information which is not presented in [21] b) In the absence of such knowledge, a better learning strategy to estimate the noise variance and then feeding it as part of the model may result in better DoA estimates than jointly estimating $\boldsymbol{\theta}$ and $\sigma_n^2$. Finally, note also that, although the optimization problem in [21] and the proposed are similar, an algorithm for solving it is missing in [21].

## V. Simulation Results

We compare the proposed technique with other gridless schemes in literature, namely ANM [8], RAM [10], gridless SPARROW [9], and gridless SPICE [11]. Since RAM is an iterative algorithm as well, we compare with it in the first two experiments. We run 20 iterations for the proposed algorithm and RAM. We also provide the noise variance, $\sigma_n^2$, to these algorithms, and set $\lambda = \sigma_n^2$ for the proposed algorithm. Note that except gridless SPICE, all other algorithms use the knowledge of the noise variance, $\sigma_n^2$.

**Experiment 1:** We evaluate the performance of the proposed technique for resolution and compare it with RAM. We also compare the two algorithms for the case when sources have different SNRs. We consider a nested array [22] with $M = 10$ sensors, and allow $K = 4$ sources incoming at angles $\{-0.5, -1/2M_{\text{apt}}, 1/2M_{\text{apt}}, 0.6\}$ in $u$-space ($u = \sin\theta$), where $M_{\text{apt}} = 30$. The corresponding SNR for sources is $\{5, 20, 20, 10\}$ dB and only a single snapshot ($L = 1$) is available. The two sources near broadside are $1/M_{\text{apt}}$ apart, or equivalently $0.5/M_{\text{apt}}$ apart in normalized frequencies, which is a challenging scenario. As seen in Fig. 1, both the proposed algorithm and RAM are able to resolve the two sources. The proposed algorithm is able to identify all 4 sources, but RAM misses the weakest source. This behavior for RAM comes from the fact that the model is matched to an estimate of *noiseless* data. In an attempt to construct such a noiseless estimate of measurement, the algorithm effectively suppressed the weakest source. It was observed that setting $\eta = 0$ helped to identify all sources for RAM. This indicates that RAM is highly sensitive to setting the parameter $\eta$ appropriately.

**Experiment 2:** We compare the proposed algorithm with RAM when the number of sources is greater than the number of sensors. We consider a nested array with $M = 6$ sensors at locations $\{0, 1, 2, 3, 7, 11\}$, and $K = 8$ sources incoming at angles uniformly in $u$-space. Their locations in MATLAB notation are $\{-1+1/K : 2/K : 1-1/K\}$. As seen in Fig. 2, the proposed algorithm is able to localize all the 8 sources, whereas the RAM algorithm suffers from poor identifiability.

**Experiment 3:** We focus on ULA, and compare the performance of StructCovMLE with other gridless techniques, including performance using the sample covariance matrix (SCM) directly, and the Cramér-Rao lower bound (CRLB). We plot the RMSE in degrees (averaged over 50 realizations) as a function of $L$ in Fig. 3. It is observed that the pro-

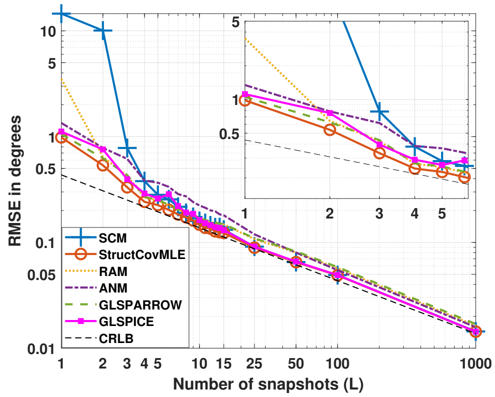



Fig. 3. $M = 6, K = 3, u_k = \sin\theta_k = \{-2/3, 0, 2/3\}$, SNR=20 dB

posed technique outperforms other techniques. With a single snapshot it identifies the three sources, whereas SCM requires three snapshots to satisfy the rank condition. Moreover, as $L$ increases, the performance of the proposed algorithm coincides with using SCM directly. This is expected as with large number of snapshots, the SCM is approximately a structured matrix and the proposed technique converges to it.

## VI. Conclusion

We proposed a novel reformulation of the SBL optimization problem and recover the underlying parameter of interest (e.g., DoAs, source frequencies) in a gridless manner. This approach naturally leads to estimating a structured covariance matrix in the MLE sense. We optimize the cost function iteratively; each iteration involving a SDP. We also provide perspectives to relate the new approach with the traditional MLE framework and the modern SBL formulation. Future directions include more theoretical analysis and lower complexity implementations to solve the proposed optimization problem.

## VII. Appendix: Proof of Theorem 1

*Proof.* It is clear that the cost functions in (9) and (13) are identical, except for the model for the received signal covariance matrix. The optimization variables for the two respective problems affect their cost only through the final covariance matrix. Thus, the two problems are equivalent if the effective matrix search domains, up to an additional '$+\tilde{\lambda}\mathbf{I}$' ($\tilde{\lambda} \geq 0$) term, are same. Let $\mathcal{D}_1$ denote the matrix search space spanned by $\mathbf{T}(K, \boldsymbol{\theta}, \mathbf{P}) = \mathbf{\Phi}_{\boldsymbol{\theta}}\mathbf{P}\mathbf{\Phi}_{\boldsymbol{\theta}}^H$ in (13), and $\mathcal{D}_2$ for $\mathbf{T}(\mathbf{v})$ in (9), where the domain for the parameters are indicated in the respective problems. To prove $\mathcal{D}_1 \subseteq \mathcal{D}_2$: Let $\mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}') \in \mathcal{D}_1$ for some $(K', \boldsymbol{\theta}', \mathbf{P}')$, then the construction $\mathbf{v}' = \mathbf{T}^{-1}(\mathbf{\Phi}_{\boldsymbol{\theta}',\mathrm{ULA}}\mathbf{P}'\mathbf{\Phi}_{\boldsymbol{\theta}',\mathrm{ULA}}^H)^2$ ensures that $\mathrm{Toep}(\mathbf{v}') \succeq \mathbf{0}$ and $\mathbf{T}(\mathbf{v}') = \mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}')$, i.e., $\mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}') \in \mathcal{D}_2$. This concludes $\mathcal{D}_1 \subseteq \mathcal{D}_2$. To prove $\mathcal{D}_2 \subseteq \mathcal{D}_1$: Let $\mathbf{T}(\mathbf{v}'') \in \mathcal{D}_2$ for some $\mathbf{v}''$, then we have $\mathrm{Toep}(\mathbf{v}'') \succeq \mathbf{0}$. We skip the case when $\mathrm{Toep}(\mathbf{v}'')$ is low rank as it follows simply from unique Vandermonde decomposition. If $\mathrm{Toep}(\mathbf{v}'')$ is full rank, then it uniquely decomposes as $\mathbf{\Phi}_{\boldsymbol{\theta}'',\mathrm{ULA}}\mathbf{P}''\mathbf{\Phi}_{\boldsymbol{\theta}'',\mathrm{ULA}}^H + \lambda''\mathbf{I}$, for some $(\boldsymbol{\theta}'', \mathbf{P}'', \lambda'' > 0)$, where the corresponding $K'' < M_{\mathrm{apt}}$ [23]. This ensures that $\mathbf{\Phi}_{\boldsymbol{\theta}''}\mathbf{P}''\mathbf{\Phi}_{\boldsymbol{\theta}''}^H + \lambda''\mathbf{I} = \mathbf{T}(\mathbf{v}'')$, which are equal up to the additional '$+\lambda''\mathbf{I}$' term. This concludes that $\mathcal{D}_2 \subseteq \mathcal{D}_1$. □

[2] $\mathbf{\Phi}_{\boldsymbol{\theta}',\mathrm{ULA}}$ denotes the array manifold matrix for a ULA of size $M_{\mathrm{apt}}$.

## References


[1] I.F. Gorodnitsky, J.S. George, and B.D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm.," *J. Electroencephalog. Clinical Neurophysiol.*, vol. 95, no.4, pp. 231–251, 1995.

[2] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[3] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[5] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[7] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[8] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[9] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1483–1497, 2018.

[10] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 995–1006, 2016.

[11] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[12] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[13] D. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[14] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[15] D. Wipf and S. Nagarajan, "Iterative reweighted $\ell_1$ and $\ell_2$ methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.

[16] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1983, vol. 8, pp. 336–339.

[17] C. Carathéodory and L. Fejér, "Uber den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und uber den Picard-Landau'schen Satz," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 218–239, 1911.

[18] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proceedings of the 2003 American Control Conference, 2003.*, 2003, vol. 3, pp. 2156–2162 vol.3.

[19] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proceedings of the IEEE*, vol. 73, no. 10, pp. 1522–1524, 1985.

[20] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.

[21] H. Qiao and P. Pal, "On maximum-likelihood methods for localizing more sources than sensors," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 703–706, 2017.

[22] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[23] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice-Hall, Upper Saddle River, NJ, USA, 2005.