2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) | 979-8-3503-4452-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAMSAP58249.2023.10403415

Joint Sparse Estimation with Cardinality Constraint via Mixed-Integer Semidefinite Programming

Tianyi Liu[1], Frederic Matter[2], Alexander Sorg[1], Marc E. Pfetsch[2], Martin Haardt[3], Marius Pesavento[1]
[1]Communication Systems Group, Technical University of Darmstadt, Germany
[2]Research Group Optimization, Technical University of Darmstadt, Germany
[3]Communications Research Laboratory, Ilmenau University of Technology, Germany

***Abstract*—The multiple measurement vectors (MMV) problem refers to the joint estimation of multiple signal realizations where the signal samples share a common sparse support over a known dictionary, which is a fundamental challenge in various applications in signal processing, e.g., direction-of-arrival (DOA) estimation. We consider the maximum a posteriori (MAP) estimation of an MMV problem, which is classically formulated as a regularized least-squares (LS) problem with an $\ell_{2,0}$-norm constraint and derive an equivalent mixed-integer semidefinite program (MISDP) reformulation, which can be solved by state-of-the-art numerical MISDP solvers at an affordable computation time. Numerical simulations in the context of DOA estimation demonstrate the improved error performance of our proposed method in comparison to several popular DOA estimation methods.**

***Index Terms*—DOA estimation, multiple measurement vectors, joint sparsity, $\ell_{2,0}$-mixed-norm constraint, mixed-integer semidefinite program, maximum a posteriori estimation**

I. Introduction

The multiple measurement vectors (MMV) problem is a fundamental challenge in signal processing and compressed sensing. It involves the joint estimation of multiple signals that share a common sparse support over a known dictionary. The MMV problem arises in various applications, e.g., imaging [1], communications [2], [3], and signal processing [4].

Similar to the classical sparse signal recovery from a single measurement vector (SMV), the MMV problem is NP-hard due to the combinatorial nature of the cardinality constraint [5], [6]. Hence, approximate procedures are conventionally applied. Many existing approximate approaches for the SMV case have been extended to the MMV case, which can be roughly divided into greedy methods [3], [7]–[9], convex relaxation approaches based on mixed-norm minimization [3], [10]–[13], and sparse Baysian learning methods [14], [15]. In particular, as an extension of basis pursuit [16] or LASSO [17] for the SMV case, the $\ell_{2,1}$-mixed-norm minimization is investigated in [3], [11]. An equivalent compact reformulation of the $\ell_{2,1}$-mixed-norm minimization, named SPARROW, is proposed in [12], which can be solved at a greatly reduced running time. Recovery guarantees of several methods for the MMV problem are established in [9], [18]–[22].

This work was supported in part by the DFG PRIDE Project PE 2080/2-1 and the EXPRESS II Project within the DFG priority program CoSIP (DFG-SPP 1798).

Inspired by the capacity of compressed sensing [23], sparsity-based DOA estimation methods have been developed, where the DOA estimation from multiple snapshots is modeled as an MMV problem by introducing a predefined dictionary that samples the complete field-of-view (FOV) [24]–[28]. The sparsity-based approach often exhibits excellent estimation performance in several demanding scenarios at affordable running time. A comprehensive review of the sparsity-based DOA estimation methods can be found in [27].

In this paper, we consider the MAP estimation for joint sparse signal recovery from multiple measurement vectors, with application to DOA estimation. This MAP estimation is typically formulated as a regularized LS problem with $\ell_{2,0}$-norm constraint, which can be viewed as a generalization of the ℓ_0-norm constrained LS problem investigated in [29] for the regression from a single measurement to the MMV case. By the reformulation techniques in [12], [29], the $\ell_{2,0}$-norm constrained LS problem can be exactly reformulated as a mixed-integer semidefinite program (MISDP), which can be solved by state-of-the-art numerical MISDP solvers at an affordable computation time. Simulation results demonstrate the efficiency of our proposed methods in comparison to several widely used DOA estimation methods. In particular, compared to the deterministic maximum likelihood (DML) estimator obtained by brute-force search over a multidimensional grid, which is considered to be statistically optimal, the proposed MISDP-based method with the SCIP-SDP solver [30] exhibits a superior error performance at a considerably reduced running time in difficult scenarios, e.g., in the case with a limited number of snapshots.

The paper is organized as follows. The sensor array signal model is presented in Section II. In Section III, we briefly review the DML estimator and the MAP estimator established in the Bayesian framework, as two classical multi-source estimation methods. In Section IV, the DOA estimation task is modeled as an MMV problem and the equivalent MISDP reformulation is established. Simulation results are presented in Section V, and conclusions are drawn in Section VI.

II. Signal Model

Consider a linear array of M omnidirectional sensors. Assume that L narrowband far-field source signals are impinge from distinct directions $\theta_1, \ldots, \theta_L \in [0, 180°]$. The corresponding spatial frequencies are defined as

$$\mu_l = \pi \cos\theta_l \in [-\pi, \pi) \tag{1}$$

for $l = 1, \ldots, L$ and summarized in the vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_L]^\mathsf{T}$. We consider the DOA estimation problem with multiple snapshots, where the array output provides measurements recorded at N time instants. We assume that the spatial frequencies in $\boldsymbol{\mu}$ remain constant within the entire observation time. Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N] \in \mathbb{C}^{M\times N}$ be the matrix that contains the N snapshots and, specifically, the (m, n)th entry $y_{m,n}$ is the output of sensor m at time instant n. The measurement matrix is modeled as

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{\Psi} + \boldsymbol{N}, \tag{2}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N] \in \mathbb{C}^{L\times N}$ is the source waveform matrix with $\psi_{l,n}$ being the signal emitted by source l at time instant n. The matrix $\boldsymbol{A}(\boldsymbol{\mu})$ collects the L steering vectors as

$$\boldsymbol{A}(\boldsymbol{\mu}) = \begin{bmatrix}\boldsymbol{a}(\mu_1) & \ldots & \boldsymbol{a}(\mu_L)\end{bmatrix} \in \mathbb{C}^{M\times L}, \tag{3}$$

where $\boldsymbol{a}(\mu) = [\mathrm{e}^{\mathrm{j}\mu\zeta_1}, \ldots, \mathrm{e}^{\mathrm{j}\mu\zeta_m}]^\mathsf{T}$ is the steering vector corresponding to the frequency μ and $\xi_1, \ldots, \xi_M$ denote the sensor locations in the linear array measured in half-wavelength. Furthermore, the matrix $\boldsymbol{N} = [\boldsymbol{n}_1, \ldots, \boldsymbol{n}_N] \in \mathbb{C}^{M\times N}$ represents independent and identically distributed (i.i.d.) circular and spatio-temporal white Gaussian noise with σ^2 being the variance of each noise entry $n_{m,n}$.

III. Deterministic Maximum Likelihood and Maximum A Posteriori Estimators

In this section we briefly review the DML estimator and the MAP estimator established in the Bayesian framework, as two classical multi-source estimation methods. As those methods are often computationally demanding, e.g., if the number of sources is large, we propose an equivalent reformulation of the MAP estimation problem in Section IV. The resulting MISDP reformulation enables a computationally efficient solution to the MAP estimation problem using state-of-the-art numerical MISDP solvers.

In the deterministic maximum likelihood (DML) approach, the source waveform matrix $\boldsymbol{\Psi}$ in (2) is considered to be deterministic and unknown. According to the signal model in (2), the snapshots $\boldsymbol{y}_n$ are statistically independent and follow the complex normal distribution

$$\boldsymbol{y}_n | \boldsymbol{\psi}_n \sim \mathcal{CN}(\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{\psi}_n, \sigma^2 \boldsymbol{I}_M). \tag{4}$$

Thus, the DML estimator for the frequencies $\boldsymbol{\mu}$ and the source waveforms $\boldsymbol{\Psi}$ is obtained as the solution of the following nonlinear LS problem [25]:

$$\min_{\boldsymbol{\mu}\in[-\pi,\pi)^L,\, \boldsymbol{\Psi}\in\mathbb{C}^{L\times N}} \quad \|\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{\Psi} - \boldsymbol{Y}\|_\mathsf{F}^2, \tag{5}$$

where $\|\cdot\|_\mathsf{F}$ is the Frobenius norm. As we are mainly interested in estimating the DOA parameters $\boldsymbol{\mu}$, the objective function in (5) can be concentrated with respect to the nuisance parameters $\boldsymbol{\Psi}$. That is, for each solution of $\boldsymbol{\mu}$, the minimizer of the nuisance parameters $\boldsymbol{\Psi}$ can be expressed in closed form, which can then be substituted into the original objective function to obtain the concentrated optimization problem. Particularly, the DML estimation problem in (5) can be concentrated as

$$\min_{\boldsymbol{\mu}\in[-\pi,\pi)^L} \quad \mathrm{tr}\left(\boldsymbol{Y}^\mathsf{H}\boldsymbol{\Pi}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})}\boldsymbol{Y}\right), \tag{6}$$

where $\boldsymbol{\Pi}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})} = \boldsymbol{I}_M - \boldsymbol{A}\left(\boldsymbol{\mu}\right)\left(\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H}\boldsymbol{A}(\boldsymbol{\mu})\right)^{-1}\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H}$ denotes the orthogonal projector onto the nullspace of $\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H}$.

The maximum a posteriori (MAP) estimator [26], [31] developed in the Bayesian framework is another widely used estimation method that is closely related to the ML estimation. In this approach, only the DOAs are considered to be deterministic, whereas the source waveforms are assumed to be stochastic. In particular, we consider the spatio-temporal i.i.d. assumption that the signal waveforms $\psi_{l,n}$ are statistically independent for different sources and snapshots. They follow the same circularly-symmetric complex Gaussian distribution

$$\boldsymbol{\psi}_n \sim \mathcal{CN}(\boldsymbol{0}, \gamma\boldsymbol{I}_L), \tag{7}$$

where γ is the source power that is assumed to be known a priori. By the Bayes' rule, the MAP estimator for the uncorrelated Gaussian prior in (7) is given by the solution of the following regularized LS problem [31]:

$$\min_{\boldsymbol{\mu}\in[-\pi,\pi)^L,\, \boldsymbol{\Psi}\in\mathbb{C}^{L\times N}} \quad \|\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{\Psi} - \boldsymbol{Y}\|_\mathsf{F}^2 + \rho\|\boldsymbol{\Psi}\|_\mathsf{F}^2 \tag{8}$$

$$\text{with} \quad \rho = \sigma^2/\gamma. \tag{9}$$

The first LS data fitting term in (8) resulting from the likelihood is identical to the DML cost function in (5), whereas the prior, according to the Gaussian assumption in (7), introduces the Tikhonov regularization term in (8). With intermediate derivations omitted, the MAP estimation can similarly be concentrated with respect to the nuisance parameters $\boldsymbol{\Psi}$ as

$$\min_{\boldsymbol{\mu}\in[-\pi,\pi)^L} \quad \mathrm{tr}\left(\boldsymbol{Y}^\mathsf{H}\widetilde{\boldsymbol{\Pi}}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})}\boldsymbol{Y}\right) \tag{10}$$

with $\widetilde{\boldsymbol{\Pi}}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})} = \boldsymbol{I}_M - \boldsymbol{A}\left(\boldsymbol{\mu}\right)\left(\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H}\boldsymbol{A}(\boldsymbol{\mu}) + \rho\boldsymbol{I}_K\right)^{-1}\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H}$. Moreover, by using the matrix inversion lemma, the matrix $\widetilde{\boldsymbol{\Pi}}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})}$ can be rewritten as $\widetilde{\boldsymbol{\Pi}}^\perp_{\boldsymbol{A}(\boldsymbol{\mu})} = (\frac{1}{\rho}\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H} + \boldsymbol{I}_M)^{-1}$, which leads to the following equivalent expression of the concentrated MAP estimation in (10):

$$\min_{\boldsymbol{\mu}\in[-\pi,\pi)^L} \quad \mathrm{tr}\left(\boldsymbol{Y}^\mathsf{H}(\tfrac{1}{\rho}\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{A}(\boldsymbol{\mu})^\mathsf{H} + \boldsymbol{I}_M)^{-1}\boldsymbol{Y}\right). \tag{11}$$

IV. A MISDP Reformulation of MAP Estimation for the MMV Problem

Due to the quadratic term in the matrix inversion, both the DML estimation in (6) and the MAP estimation in (11) are nonconvex and multimodal with a large number of local minima. Hence, the corresponding optimization procedure is computationally demanding and generally requires a multi-dimensional grid search to find the exact solution. Inspired by the concept of compressed sensing [23], the above DOA estimation problem can be modeled as an MMV problem by introducing a predefined dictionary that samples the complete FOV [27]. In this section, we first introduce the MMV-based model for DOA estimation. For the MMV problem, a dictionary-based MAP estimation is derived according to (8), which is reformulated as a MISDP problem by the reformulation techniques in [29] and [12]. The problem can then be solved by state-of-the-art MISDP solvers, e.g., SCIP-SDP [30].

The problem of recovering the frequencies in $\boldsymbol{\mu}$ from the measurements $\boldsymbol{Y}$ can be formulated as an MMV problem by using the following sparse representation for the model in (2):

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{\nu})\boldsymbol{X} + \boldsymbol{N}, \tag{12}$$

where $\boldsymbol{A}(\boldsymbol{\nu}) = [\boldsymbol{a}(\nu_1), \ldots, \boldsymbol{a}(\nu_K)] \in \mathbb{C}^{M\times K}$ is an overcomplete dictionary constructed by sampling the FOV in $K \gg L$ directions with spatial frequencies $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_K]^\mathsf{T}$ and $\boldsymbol{X} \in \mathbb{C}^{K\times N}$ is a sparse representation of the source signal matrix $\boldsymbol{\Psi}$. Specifically, provided that the true frequencies $\boldsymbol{\mu}$ are contained in the frequency grid, i.e.,

$$\{\mu_l\}_{l=1}^L \subset \{\nu_k\}_{k=1}^K, \tag{13}$$

then $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]^\mathsf{T}$ admits a row-sparse structure, which has only L nonzero rows corresponding to the signal waveforms of the L sources, i.e., $\boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{\Psi} = \boldsymbol{A}(\boldsymbol{\nu})\boldsymbol{X}$. Thus, the considered DOA estimation problem can be described as an MMV problem that aims at jointly recovering a set of signal samples in $\boldsymbol{X}$ that have a common sparse support over the fixed dictionary $\boldsymbol{A}(\boldsymbol{\nu})$. The spatial frequencies are then estimated from the support of the recovered row-sparse signal matrix $\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{x}}_1, \ldots, \widehat{\boldsymbol{x}}_K]^\mathsf{T}$ by $\{\hat{\mu}_l\}_{l=1}^L = \{\nu_k \mid \|\widehat{\boldsymbol{x}}_k\|_0 > 0, k = 1, \ldots, K\}$, where the ℓ_0-pseudo-norm $\|\widehat{\boldsymbol{x}}_k\|_0$ counts the number of nonzeros entries in $\widehat{\boldsymbol{x}}_k$. For simplicity, in the rest of the paper, the dictionary is referred to as $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{\nu})$.

The $\ell_{p,q}$-mixed-norms are commonly used to enforce the row-sparsity assumption in sparse recovery problems [13]. The $\ell_{p,q}$-norm for a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]^\mathsf{T}$ is defined as $\|\boldsymbol{X}\|_{p,q} = \|\boldsymbol{x}^{(\ell_p)}\|_q$ with $\boldsymbol{x}^{(\ell_p)} = [\|\boldsymbol{x}_1\|_p, \ldots, \|\boldsymbol{x}_K\|_p]^\mathsf{T}$. In particular, the $\ell_{p,0}$-pseudo-norm represents the exact number of nonzero rows of the matrix, which, however, typically leads to an NP-hard problem due to its nonconvexity. The SPARROW method in [12] utilizes an $\ell_{2,1}$-norm regularization as a convex approximation of the $\ell_{2,0}$-norm, to address the MMV problem described above. In contrast, in this paper, we consider the exact MAP estimation for the sparse model in (12), which, similar to (8), is formulated as the following regularized LS problem with $\ell_{2,0}$-norm constraint:

$$\min_{\boldsymbol{X}\in\mathbb{C}^{K\times N},\, \|\boldsymbol{X}\|_{2,0}\le L} \quad \|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Y}\|_\mathsf{F}^2 + \rho\|\boldsymbol{X}\|_\mathsf{F}^2. \tag{14}$$

The DML approach for the sparse model in (12) is obtained from (14) by choosing the parameter ρ to be zero. However, compared to the DML approach, the MAP estimation in (14) can be equivalently reformulated as a MISDP problem due to the additional Tikhonov regularization, and then, its global optimum can be conveniently obtained by a state-of-the-art MISDP solver. Problem (14) with the $\ell_{2,0}$-norm constraint can be viewed as a generalization of the ℓ_0-norm constrained LS regression problem for a single measurement, as considered by Pilanci et al. in [29], to the MMV case. In this paper, we provide a nontrivial extension of Pilanci's MISDP reformulation for the SMV case to the MMV problem (14). The regularization parameter ρ is chosen according to (9) if the prior information of the expected power of the source waveforms is known and satisfies the assumption in (7). Otherwise, if training data are available, a suitable value of ρ may be obtained through cross-validation or, more efficiently, with the help of the algorithm unrolling procedure [32].

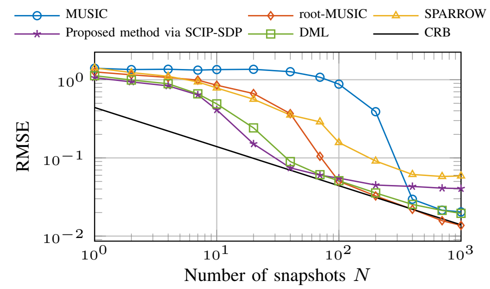

Fig. 1. RMSE vs. number of snapshots for $L = 3$ uncorrelated sources, $M = 8$ sensors, SNR $= -5$ dB, and $K = 100$ grid points.

In the following, we present a simplified derivation of the MISDP reformulation for the MMV case, which, unlike that in [29], does not involve the dual problem constructed with the Legendre-Fenchel conjugate. First, by introducing binary variables $\boldsymbol{u} \in \{0,1\}^K$, the $\ell_{2,0}$-norm constrained problem in (14) can be equivalently represented as the lifted problem

$$\min_{\substack{\boldsymbol{u}\in\{0,1\}^K\\ \boldsymbol{u}^\mathsf{T}\boldsymbol{1}\le L}} \; \min_{\boldsymbol{X}\in\mathbb{C}^{K\times N}} \; \|\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{X} - \boldsymbol{Y}\|_\mathsf{F}^2 + \rho\|\boldsymbol{X}\|_\mathsf{F}^2, \tag{15}$$

where $\boldsymbol{1}$ is an all-ones vector. The matrix $\boldsymbol{D}(\boldsymbol{u})$ in (15) is a diagonal matrix with $\boldsymbol{u}$ on its diagonal, which determines the directions with nonzero source signals. Note that $\boldsymbol{D}(\boldsymbol{u})$ is not required in the regularization in (15) because the rows of $\boldsymbol{X}$ that are not selected by $\boldsymbol{D}(\boldsymbol{u})$ are not involved in the data fitting term and, hence, enforced to all-zero by the minimization of $\|\boldsymbol{X}\|_\mathsf{F}^2$.[1] Like the MAP estimation in (8), (15) can be concentrated with respect to $\boldsymbol{X}$ and then reformulated by the matrix inversion lemma as the integer program

$$\min_{\boldsymbol{u}\in\{0,1\}^K,\, \boldsymbol{u}^\mathsf{T}\boldsymbol{1}\le L} \; \mathrm{tr}\left(\boldsymbol{Y}^\mathsf{H}(\tfrac{1}{\rho}\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{A}^\mathsf{H} + \boldsymbol{I}_M)^{-1}\boldsymbol{Y}\right). \tag{16}$$

Next, by applying the same SDP reformulation technique as in [12], [29], the integer program in (16) can be further written as the following MISDP problem with a slack variable $\boldsymbol{T}$:

$$\min_{\boldsymbol{u}\in\{0,1\}^K,\boldsymbol{T}\in\mathbb{S}_+^N} \quad \mathrm{tr}(\boldsymbol{T}) \tag{17a}$$

$$\text{s.t.} \quad \begin{bmatrix} \frac{1}{\rho}\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{A}^\mathsf{H} + \boldsymbol{I}_M & \boldsymbol{Y} \\ \boldsymbol{Y}^\mathsf{H} & \boldsymbol{T} \end{bmatrix} \succeq 0, \tag{17b}$$

$$\boldsymbol{u}^\mathsf{T}\boldsymbol{1} \le L, \tag{17c}$$

where $\mathbb{S}_+^N$ is the set of $N\times N$ positive semidefinite matrices. The positive semidefiniteness of the matrix $\boldsymbol{T}$ is required by the PSD constraint (17b). The equivalence between (16) and (17) is shown as follows. Since $\frac{1}{\rho}\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{A}^\mathsf{H} + \boldsymbol{I}_M$

[1] Note that the MAP estimation in (8) is often solved by grid search in practice. By adding $\boldsymbol{D}(\boldsymbol{u})$ into the Tikhonov regularization in (15), problem (15) can also be interpreted as a discretized version of the MAP estimation in (8). That is, problem (15) is equivalent to solving the MAP estimation in (8) via a brute-force search over the grid $\{\nu_k\}_{k=1}^K$ in (13).

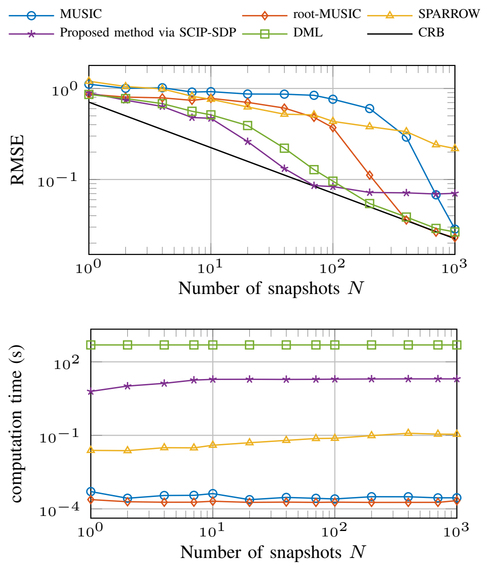

Fig. 2. Performance vs. the number of snapshots for $L = 5$ uncorrelated sources, $M = 8$ sensors, SNR $= -5$ dB, and $K = 100$ grid points.[3]

is positive definite, by the Schur complement formula, the constraint (17b) is equivalent to [33]

$$\boldsymbol{T} \succeq \boldsymbol{Y}^{\mathsf{H}}(\tfrac{1}{\rho}\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{A}^{\mathsf{H}} + \boldsymbol{I}_M)^{-1}\boldsymbol{Y}.$$

Therefore, for every given $\boldsymbol{u}$, the minimum of $\mathrm{tr}(\boldsymbol{T})$ in (17) is achieved at $\boldsymbol{T} = \boldsymbol{Y}^{\mathsf{H}}(\frac{1}{\rho}\boldsymbol{A}\boldsymbol{D}(\boldsymbol{u})\boldsymbol{A}^{\mathsf{H}} + \boldsymbol{I}_M)^{-1}\boldsymbol{Y}$.

V. Simulation Results

We conduct numerical experiments on synthetic data to evaluate and analyze the performance of the proposed method via the SCIP-SDP solver [30] of version 4.1.0.[4] The nonlinear branch-and-bound approach is chosen in the SCIP-SDP solver, where the relaxed continuous SDP subproblems are solved by MOSEK [34]. The proposed method is compared to the stochastic Cramér-Rao Bound (CRB) [35] and several widely used approaches for DOA estimation, namely, MUSIC [36], root-MUSIC [37], the SPARROW method with coordinate descent implementation, and the DML estimator. The DML estimator is obtained via a brute-force search over the same grid as in (12), which is equivalent to the solution of problem (14) with the regularization parameter ρ being zero. The results are averaged over $N_R = 200$ Monte-Carlo trials. In particular, the quality of the estimated spatial frequencies $\widehat{\boldsymbol{\mu}}(n) = [\hat{\mu}_1(n), \ldots, \hat{\mu}_L(n)]^{\mathsf{T}}$ for $n = 1, \ldots, N_R$ are measured by the root-mean-square error (RMSE) with respect to the ground-truth $\boldsymbol{\mu}$ defined as

$$\mathrm{RMSE}(\widehat{\boldsymbol{\mu}}) = \sqrt{\tfrac{1}{LN_R}\textstyle\sum_{n=1}^{N_R}\sum_{l=1}^{L}|\hat{\mu}_l(n) - \mu_l|_{\mathrm{wa}}^2},$$

where $|\mu_1 - \mu_2|_{\mathrm{wa}} = \min_{k\in\mathbb{Z}}|\mu_1 - \mu_2 + 2k\pi|$ denotes the wrap-around distance between two frequencies μ_1 and μ_2. All experiments were conducted on a Linux PC with an Intel Core i7-7700 CPU and 32 GB RAM running MATLAB 2022a.

In the simulations, we consider a ULA of $M = 8$ half-wavelength spaced sensors. In each Monte-Carlo trial, the true source signals in $\boldsymbol{\Psi}$ are generated according to the uncorrelated Gaussian prior in (7) with the variance $\gamma = 1$. The SNR $= 1/\sigma^2$ is -5 dB and the dictionary $\boldsymbol{A}$ is constructed from $K = 100$ grid points with frequencies uniformly sampled in $[-\pi, \pi)$. The regularization parameter ρ in (14) for our proposed method is chosen according to the rule in (9) and the penalty weight λ for the $\ell_{2,1}$-norm regularization in the SPARROW method is selected by the heuristic rule $\lambda = \sqrt{\sigma^2 M \log M}$ [12], [38].

In the first simulation, we compare the error performance of the methods in a small-scale scenario of $L = 3$ sources with frequencies $\boldsymbol{\mu} = \pi \cdot [-0.1, 0.35, 0.5]^{\mathsf{T}}$, where the brute-force search is computationally competitive. The estimation errors of the recovered frequencies are reported in Fig. 1. As shown in Fig. 1, due to the convex relaxation by means of $\ell_{2,1}$-norm employed by SPARROW, the brute-force DML and the proposed method via SCIP-SDP outperform SPARROW in both the asymptotic and non-asymptotic regions. The brute-force DML and the proposed method exhibit the best threshold performance. As a result of the Tikhonov regularization in (14) introduced by the additional prior assumption, the proposed method exhibits a lower estimation error than DML in the region of low sample size but, meanwhile, an asymptotic bias.

Next, we consider a scenario of $L = 5$ sources with frequencies $\boldsymbol{\mu} = \pi \cdot [-0.5, 0.1, 0.35, 0.5, 0.7]^{\mathsf{T}}$. The error performance and computational time are reported in Fig. 2. Although the branch-and-bound strategy employed by SCIP-SDP enjoys improved scalability compared to the brute-force search, to limit the total execution time of this simulation, we terminate SCIP-SDP when 500 branch-and-bound nodes have been explored. Even with early termination, the proposed method via SCIP-SDP presents a more significant decrease of the RMSE compared to DML in the region of low sample size than that in Fig. 1, which also leads to a threshold performance superior to DML. This suggests that, for low sample sizes, our proposed MISDP-based method is more favorable than the brute-force DML since it possesses not only a superior error performance but also a reduced running time.

VI. Conclusion

We consider the maximum a posteriori estimation for joint sparse signal recovery from multiple measurement vectors that is classically formulated as a regularized least-squares (LS) problem with $\ell_{2,0}$-norm constraint, and derive an equivalent mixed-integer semidefinite program (MISDP) reformulation, which can be solved by state-of-the-art MISDP solvers at an affordable computation time. The simulations in the context of DOA estimation show the efficiency of our proposed method in comparison to popular DOA estimation methods.

[3] In the oversampled case, to reduce the complexity of the SDP subproblems, a compact reformulation of (17), whose constraint dimension is independent of the number of snapshots N, is employed based on the reformulation technique in [12, Eq. (25)].

[4] The source code of SCIP-SDP and an interface for MATLAB can be downloaded from the website https://www.opt.tu-darmstadt.de/scipsdp. In the experiments, the SCIP-SDP solver is called from MATLAB through the provided interface.

References

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

[2] I. Fevrier, S. Gelfand, and M. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 927–937, Jun. 1999.

[3] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[4] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, N.J: Pearson/Prentice Hall, 2005.

[5] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[6] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[7] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[8] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

[9] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[10] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[11] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[12] 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, Mar. 2018.

[13] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, Nov. 2009.

[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, Jul. 2007.

[15] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[18] Y. Jin and B. D. Rao, "Support recovery of sparse signals in the presence of multiple measurement vectors," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 3139–3157, May 2013.

[19] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[20] M.-J. Lai and Y. Liu, "The null space property for sparse recovery from multiple measurement vectors," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 402–406, May 2011.

[21] L. Ramesh, C. R. Murthy, and H. Tyagi, "Sample-measurement tradeoff in support recovery under a subgaussian prior," *IEEE Transactions on Information Theory*, vol. 67, no. 12, pp. 8140–8153, Dec. 2021.

[22] S. Khanna and C. R. Murthy, "On the Support Recovery of Jointly Sparse Gaussian Sources via Sparse Bayesian Learning," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7361–7378, Nov. 2022.

[23] Y. C. Eldar and G. Kutyniok, Eds., *Compressed sensing: theory and applications*. Cambridge ; New York: Cambridge University Press, 2012.

[24] 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, Jul. 1996.

[25] H. L. Van Trees, *Optimum array processing*. New York, USA: John Wiley & Sons, Inc., Mar. 2002.

[26] A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds., *Array and statistical signal processing*, first edition ed., ser. Academic Press library in signal processing. Amsterdam: Academic Press, 2014, no. v. 3.

[27] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, Jan. 2018, pp. 509–581.

[28] W. Liu, M. Haardt, M. S. Greco, C. F. Mecklenbräuker, and P. Willett, "Twenty-five years of sensor array and multichannel signal processing: a review of progress to date and potential research directions," *IEEE Signal Processing Magazine*, vol. 40, no. 4, pp. 80–91, Jun. 2023.

[29] M. Pilanci, M. J. Wainwright, and L. El Ghaoui, "Sparse learning via Boolean relaxations," *Mathematical Programming*, vol. 151, no. 1, pp. 63–87, Jun. 2015.

[30] T. Gally, M. E. Pfetsch, and S. Ulbrich, "A framework for solving mixed-integer semidefinite programs," *Optimization Methods and Software*, vol. 33, no. 3, pp. 594–632, May 2018.

[31] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?" *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[32] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[33] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[34] MOSEK ApS, "The MOSEK optimization toolbox for MATLAB. Version 10.0.22." https://docs.mosek.com/10.0.22/toolbox/index.html.

[35] P. Stoica, E. Larsson, and A. Gershman, "The stochastic CRB for array processing: a textbook derivation," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.

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

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

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