A COMPACT FORMULATION FOR THE L21 MIXED-NORM MINIMIZATION PROBLEM

Christian Steffens and Marius Pesavento

Communication Systems Group TU Darmstadt, Germany

Marc E. Pfetsch

Discrete Optimization Group TU Darmstadt, Germany

ABSTRACT

We present an equivalent, compact reformulation of the $\ell_{2,1}$ mixednorm minimization problem for joint sparse signal reconstruction from multiple measurement vectors (MMVs). The reformulation builds upon a compact parameterization, which models the rownorms of the sparse signal representation as parameters of interest, resulting in a significant reduction of the MMV problem size. Given the sparse vector of row-norms, the joint sparse signal can be computed from the MMVs in closed form. For the special case of uniform linear sampling, we present an extension of the compact formulation for gridless parameter estimation by means of semidefinite programming. Furthermore, we derive in this case from our compact problem formulation the exact equivalence between the $\ell_{2,1}$ mixednorm minimization and the atomic-norm minimization.

Index Terms— Multiple Measurement Vectors, Joint Sparsity, Mixed-Norm Minimization, Gridless Parameter Estimation

1. INTRODUCTION

Sparse Signal Reconstruction (SSR) techniques have attracted considerable research interest over the last decades [1–4] and have been applied to various fields of signal processing, such as spectral analysis, Direction-Of-Arrival (DOA) estimation, image processing, geophysics, tomography, or machine learning.

Given a Single Measurement Vector (SMV), SSR considers the problem of reconstructing a sparse signal vector from an underdetermined linear system. Ideally, this problem is addressed using an ℓ_0 minimization approach, which, however, is an NP-hard problem. For this reason, several techniques have been proposed to approximately solve the SSR problem. Most prominent techniques are based on convex relaxation in terms of ℓ_1 norm minimization [1–4] or greedy methods, such as OMP [5, 6], where the latter category usually has lower computational complexity at the cost of reduced recovery guarantees. In [7–10] the SSR problem has been extended to an infinite-dimensional vector space by means of total variation norm and atomic norm minimization, leading to gridless parameter estimation methods.

Besides the SMV scenario, many practical applications deal with the problem of reconstructing a joint sparse signal representation from Multiple Measurement Vectors (MMVs). Similar to the SMV case, techniques for the MMV-based SSR problem include convex relaxation by means of mixed-norm minimization [11–15], and greedy methods [16, 17]. Recovery guarantees for the MMV case have been established in [18–20]. An extension to the infinite-dimensional vector space for MMV-based SSR, using atomic norm minimization, has been proposed in [21–23]. In contrast to the

aforementioned sparsity enforcing methods, the SPICE method for joint sparse reconstruction from MMVs, as proposed in [24, 25], is based on weighted covariance matching and constitutes a sparse estimation problem which does not require the application of a sparsity prior. Links between SPICE and SSR formulations have been established in [23–27], which show that SPICE can be reformulated as an $\ell_{2,1}$ mixed-norm minimization problem.

In this paper we consider joint sparse signal reconstruction from MMVs by means of the classical $\ell_{2,1}$ mixed-norm minimization problem, with application to DOA estimation. A general shortcoming of the classical $\ell_{2,1}$ formulation is that its problem size grows with the number of measurements and the resolution requirement, respectively, and various heuristic approaches to deal with these difficulties have been proposed, e.g., in [15]. Here, we derive an equivalent reformulation of the $\ell_{2,1}$ mixed-norm minimization problem based on a compact parameterization in which the optimization parameters represent the row-norms of the signal representation, rather than the signal matrix itself which generally contains significantly more parameters. We refer to this formulation as the SPARse ROWnorm reconstruction (SPARROW) problem. Given the sparse signal row-norms, the joint sparse signal matrix is reconstructed from the MMVs in closed-form. We point out that support recovery is determined by the sparse vector of row-norms and only relies on the sample covariance matrix instead of the MMVs themselves. In this sense we achieve a concentration of the optimization variables as well as the measurements, leading to a significantly reduced problem size in the case of a large number of MMVs.

We present an implementation of the SPARROW formulation based on semidefinite programming (SDP) and provide an extension of the SDP implementation for gridless parameter estimation. Furthermore, we compare our new problem formulation to atomic norm minimization and establish from our gridless, compact reformulation the exact equivalence between the classical $\ell_{2,1}$ mixed-norm minimization problem [11–15] and the recently proposed atomic norm minimization formulation for MMV scenarios [21–23]. In an extended version of this paper [28] we provide additional proofs, a low-complexity implementation of the SPARROW formulation by means of the coordinate descent method, a comparison to the SPICE method as well extensive simulation results.

2. SIGNAL MODEL

Consider a linear array of M omnidirectional sensors, as depicted in Figure 1. Further, assume a set of L narrowband far-field sources in angular directions $\theta_1, \ldots, \theta_L$, relative to the array axis. The spatial frequencies are defined as

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

This work was supported by the EXPRESS project within the DFG priority program CoSIP (DFG-SPP 1798).



Fig. 1. Exemplary setup for a linear array of M = 6 sensors and L = 3 source signals

for l = 1, ..., L, and comprise the vector $\boldsymbol{\mu} = [\mu_1, ..., \mu_L]^T$. The array output provides measurement vectors which are recorded over N time instants where we assume that the sources transmit time-varying signals while the frequencies in $\boldsymbol{\mu}$ remain constant within the entire observation time. The measurement vectors are collected in the multiple measurement vector (MMV) matrix $\boldsymbol{Y} \in \mathbb{C}^{M \times N}$, where $[\boldsymbol{Y}]_{m,n}$ denotes the output at sensor m in time instant n. The MMV matrix is modeled as

$$Y = A(\mu)\Psi + N, \qquad (2)$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{L \times N}$ is the source signal waveform matrix, with $[\boldsymbol{\Psi}]_{l,n}$ denoting the signal transmitted by source l in time instant n, and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ represents circular and spatio-temporal white Gaussian sensor noise with covariance matrix $\mathbf{E}\{\boldsymbol{N}\boldsymbol{N}^{\mathsf{H}}\}/N = \sigma^{2}\boldsymbol{I}_{M}$, where \boldsymbol{I}_{M} and σ^{2} denote the $M \times M$ identity matrix and the noise power, respectively. The $M \times L$ array steering matrix $\boldsymbol{A}(\boldsymbol{\mu})$ in (2) is given by

$$\boldsymbol{A}(\boldsymbol{\mu}) = [\boldsymbol{a}(\mu_1), \dots, \boldsymbol{a}(\mu_L)], \qquad (3)$$

$$\boldsymbol{a}(\mu) = [1, \mathrm{e}^{-\mathrm{j}\pi\mu\rho_2}, \dots, \mathrm{e}^{-\mathrm{j}\pi\mu\rho_M}]^\mathsf{T}$$
(4)

is the array manifold vector with $\rho_m \in \mathbb{R}$, for m = 1, ..., M, denoting the position of the *m*th sensor in half signal wavelength, relative to the first sensor in the array, hence $\rho_1 = 0$.

3. JOINT SPARSE SIGNAL RECONSTRUCTION AND EQUIVALENT SPARROW FORMULATION

Based on the signal model in (2) we introduce a sparse representation $A(\mu)\Psi = A(\nu)X$ as illustrated in Figure 2. In the sparse representation $A(\nu)$ is an $M \times K$ overcomplete dictionary matrix obtained by sampling the field-of-view in $K \gg L$ spatial frequencies $\nu = [\nu_1, \ldots, \nu_K]^T$, defined in correspondence with (3). For ease of presentation we will use the short-hand notation $A = A(\nu)$ in the remainder of the paper to refer to the dicitionary matrix. The matrix $X = [x_1, \ldots, x_K]^T \in \mathbb{C}^{K \times N}$ is a sparse representation of the signal waveform matrix Ψ , which has non-zero rows x_k only if the corresponding sampled spatial frequency ν_k is contained in the spatial frequencies in μ .

Using the sparse representation AX, the DOA estimation problem can be formulated as the well-known convex optimization problem [11–15]

$$\min_{\boldsymbol{X}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Y}\|_{F}^{2} + \lambda \|\boldsymbol{X}\|_{p,1}$$
(5)

where $\lambda > 0$ is the regularization parameter determining the sparsity, i.e., the number of non-zero rows in $\mathbf{X} \in \mathbb{C}^{K \times N}$, and

$$\|\boldsymbol{X}\|_{p,1} = \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_p \tag{6}$$



Fig. 2. Signal model and sparse representation (neglecting additive noise and basis mismatch)

denotes the $\ell_{p,1}$ mixed-norm, where p = 2 and $p = \infty$ are the most prominent choices. The mixed-norm formulation (6) induces a coupling among the elements in each row $\boldsymbol{x}_k, k = 1, \ldots, K$, of the matrix \boldsymbol{X} such that the ℓ_1 -norm, i.e., the summation, is performed on the ℓ_2 -norms of the rows in \boldsymbol{X} .

Given a minimizer $\hat{X} = [\hat{x}_1, \dots, \hat{x}_K]^T$ the DOA estimation problem reduces to the identification of the union support set, i.e., the indices of the non-zero rows, and assigning the corresponding frequency grid points to the set $\{\hat{\mu}\}$ of estimated spatial frequencies

$$\{\hat{\mu}\} = \{\nu_k \mid \left\|\hat{\boldsymbol{x}}_k\right\|_2 \neq 0; \ k = 1, \dots, K\}.$$
(7)

In practice, the strong sparsity assumption in (7) is relaxed and the DOA estimates are approximated by the positions of the local maxima of the row-norms in \hat{X} .

As discussed above, the MMV-based $\ell_{2,1}$ mixed-norm minimization problem is one of the most prominent approaches for joint sparse recovery and can be considered as a generalization of the more prominent ℓ_1 norm minimization problem for SMVs [1–4]. In this context, one of our main results is given by the following theorem:

Theorem 1. The row-sparsity inducing $\ell_{2,1}$ mixed-norm minimization problem

$$\min_{\boldsymbol{X}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Y}\|_{\mathsf{F}}^{2} + \lambda \sqrt{N} \|\boldsymbol{X}\|_{2,1}$$
(8)

is equivalent to the convex problem

$$\min_{\boldsymbol{S}\in\mathbb{D}_{+}}\operatorname{Tr}\left(\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathsf{H}}+\lambda\boldsymbol{I}_{M}\right)^{-1}\hat{\boldsymbol{R}}\right)+\operatorname{Tr}(\boldsymbol{S}),\tag{9}$$

with $\hat{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^{\mathsf{H}}/N$ denoting the sample covariance matrix and \mathbb{D}_+ describing the set of nonnegative diagonal matrices, in the sense that minimizers $\hat{\mathbf{X}}$ and $\hat{\mathbf{S}}$ for problems (8) and (9), respectively, are related by

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{S}}\boldsymbol{A}^{\mathsf{H}} (\boldsymbol{A}\hat{\boldsymbol{S}}\boldsymbol{A}^{\mathsf{H}} + \lambda \boldsymbol{I}_{M})^{-1} \boldsymbol{Y}.$$
(10)

Proofs for the equivalence of (8) and (9), and the convexity of (9) are provided in [28].

In addition to (10), we observe that the minimizer $\hat{S} = \text{diag}(\hat{s}_1, \dots, \hat{s}_K)$ contains the row-norms of the sparse signal matrix $\hat{X} = [\hat{x}_1, \dots, \hat{x}_K]^{\mathsf{T}}$ on its diagonal according to

$$\hat{s}_k = \frac{1}{\sqrt{N}} \|\hat{x}_k\|_2,$$
 (11)

for k = 1, ..., K, such that the union support of \hat{X} is equivalently represented by the support of the sparse vector of row-norms

 $[\hat{s}_1,\ldots,\hat{s}_K]$. We will refer to (9) as the SPARse ROW-norm reconstruction (SPARROW) formulation. In this regard, we emphasize that \hat{S} should not be mistaken for a sparse representation of the source covariance matrix, i.e., $\hat{S} \neq \mathrm{E}\{\hat{X}\hat{X}^{\mathsf{H}}\}/N$. While the mixed-norm minimization problem in (8) involves NK complex optimization variables in X, the SPARROW problem in (9) provides a reduction to only K nonnegative optimization variables in the diagonal matrix S. Moreover, the SPARROW problem in (9) only relies on the sample covariance matrix \hat{R} instead of the MMVs in Y themselves, leading to a reduction in problem size, especially in the case of large number of MMVs N. Interestingly, this indicates that the union support of the signal matrix \hat{X} is fully encoded in the sample covariance \hat{R} , rather than the instantaneous MMVs in Y, as may be concluded from the $\ell_{2,1}$ formulation in (8). As seen from (10), the instantaneous MMVs in Y are only required for the signal reconstruction, which, in the context of array signal processing, can be interpreted as a form of beamforming [29], where the row-sparse structure in X is induced by premultiplication with the sparse diagonal matrix \hat{S} .

4. IMPLEMENTATION OF THE SPARROW FORMULATION

To show convexity of the SPARROW formulation (9) and for implementation with standard convex solvers, such as SeDuMi [30], consider the following corollary:

Corollary 1. The SPARROW problem in (9) is equivalent to the semidefinite programs (SDPs)

$$\min_{\boldsymbol{S},\boldsymbol{U}_N} \quad \frac{1}{N} \operatorname{Tr}(\boldsymbol{U}_N) + \operatorname{Tr}(\boldsymbol{S})$$
(12a)

s.t.
$$\begin{bmatrix} \boldsymbol{U}_N & \boldsymbol{Y}^{\mathsf{H}} \\ \boldsymbol{Y} & \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathsf{H}} + \lambda \boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0}$$
 (12b)

$$S \in \mathbb{D}_+$$
 (12c)

where U_N is a Hermitian matrix of size $N \times N$, and

$$\min_{\boldsymbol{S},\boldsymbol{U}_M} \operatorname{Tr}(\boldsymbol{U}_M \hat{\boldsymbol{R}}) + \operatorname{Tr}(\boldsymbol{S})$$
(13a)

s.t.
$$\begin{bmatrix} \boldsymbol{U}_M & \boldsymbol{I}_M \\ \boldsymbol{I}_M & \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathsf{H}} + \lambda \boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0}$$
 (13b)

$$\boldsymbol{S} \in \mathbb{D}_+$$
 (13c)

where U_M is a Hermitian matrix of size $M \times M$.

Proof: see [28].

In contrast to the constraint (12b), the dimension of the semidefinite constraint (13b) is independent of the number of MMVs N. It follows that either problem formulation (12) or (13) can be selected to solve the SPARROW problem in (9), depending on the number of MMVs N and the resulting dimension of the semidefinite constraint, i.e., (12) is preferable for $N \leq M$ and (13) is preferable otherwise.

While the above SDP implementations are applicable to arbitrary array geometries, we consider next the special case of a uniform linear array (ULA) with sensor positions $\rho_m = m - 1$, for $m = 1, \ldots, M$, such that $\boldsymbol{A} = [\boldsymbol{a}(\nu_1), \ldots, \boldsymbol{a}(\nu_K)]$ is a Vandermonde matrix of size $M \times K$. Under the given assumptions, the matrix product $\boldsymbol{ASA}^{\mathsf{H}}$ exhibits a Toeplitz structure according to

$$\operatorname{Toep}(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathsf{H}} = \sum_{k=1}^{K} s_k \boldsymbol{a}(\nu_k) \boldsymbol{a}^{\mathsf{H}}(\nu_k), \qquad (14)$$

where $\text{Toep}(\boldsymbol{u})$ denotes a Hermitian Toeplitz matrix with \boldsymbol{u} as its first column. As discussed in [9], by the Caratheodory theorem any Toeplitz matrix $\text{Toep}(\boldsymbol{u})$ can be represented by a Vandermonde decomposition according to (14) for any distinct frequencies ν_1, \ldots, ν_K and corresponding magnitudes $s_1, \ldots, s_K > 0$, with rank $(\text{Toep}(\boldsymbol{u})) = K \leq M$. Given a Toeplitz matrix $\text{Toep}(\boldsymbol{u})$, the Vandermonde decomposition according to (14) can be obtained by first recovering the frequencies ν_k , e.g., using the matrix pencil approach [31] or the linear prediction method [32], where the frequency recovery is performed in a gridless fashion. The corresponding signal magnitudes in $\boldsymbol{s} = [s_1, \ldots, s_K]^{\mathsf{T}}$ can be reconstructed by solving the linear system

$$\boldsymbol{A}\,\boldsymbol{s}=\boldsymbol{u},\tag{15}$$

i.e., by exploiting that $[a(\nu)]_1 = 1$, for all $\nu \in [-1, 1)$, and considering the first column in the representation (14). Based on the Vandermonde decomposition in (14), we extend Corollary 1 to a gridless version:

Corollary 2. For uniform sampling the gridless SPARROW (GL-SPARROW) is given by the equivalent SDPs

$$\min_{\boldsymbol{u},\boldsymbol{U}_N} \quad \frac{1}{N} \operatorname{Tr} (\boldsymbol{U}_N) + \frac{1}{M} \operatorname{Tr} (\operatorname{Toep}(\boldsymbol{u}))$$
(16a)

s.t.
$$\begin{bmatrix} \boldsymbol{U}_N & \boldsymbol{Y}^{\mathsf{H}} \\ \boldsymbol{Y} & \operatorname{Toep}(\boldsymbol{u}) + \lambda \boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0}$$
 (16b)

$$\operatorname{Toep}(\boldsymbol{u}) \succeq \boldsymbol{0}, \tag{16c}$$

where U_M is an $N \times N$ Hermitian matrix, and

$$\min_{\boldsymbol{\mathcal{U}}_{M}} \operatorname{Tr}(\boldsymbol{\mathcal{U}}_{M}\hat{\boldsymbol{R}}) + \frac{1}{M}\operatorname{Tr}(\operatorname{Toep}(\boldsymbol{u}))$$
(17a)

s.t.
$$\begin{bmatrix} \boldsymbol{U}_M & \boldsymbol{I}_M \\ \boldsymbol{I}_M & \text{Toep}(\boldsymbol{u}) + \lambda \boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0}$$
 (17b)

$$\operatorname{Toep}(\boldsymbol{u}) \succeq \boldsymbol{0} \tag{17c}$$

where U_N is an $M \times M$ Hermitian matrix, with frequency reconstruction provided by the Vandermonde decomposition (14) of Toep(u).

In Corollary 2 we additionally make use of the identity

$$\operatorname{Tr}(\boldsymbol{S}) = \frac{1}{M} \operatorname{Tr}(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathsf{H}}) = \frac{1}{M} \operatorname{Tr}(\operatorname{Toep}(\boldsymbol{u})), \quad (18)$$

with the factor 1/M resulting from the fact that $\|\boldsymbol{a}(\nu)\|_2^2 = M$, for all $\nu \in [-1, 1)$.

Given a minimizer \hat{u} of problem (16) or (17), the number of sources, i.e., the model order, can be directly estimated as

$$\hat{L} = \operatorname{rank}(\operatorname{Toep}(\hat{\boldsymbol{u}})), \tag{19}$$

while the frequencies $\{\hat{\mu}_l\}_l^{\hat{L}}$ and corresponding magnitudes $\{\hat{s}_l\}_l^{\hat{L}}$ can be estimated by Vandermonde decomposition according to (14), as discussed above. With the frequencies in $\{\hat{\mu}_l\}_l^{\hat{L}}$ and signal magnitudes in $\{\hat{s}_l\}_l^{\hat{L}}$, the corresponding signal matrix \hat{X} can be reconstructed by application of (10).

In the case of large sensor arrays the aforementioned SDP formulations might become computationally prohibitive. For these scenarios we have derived a low complexity implementation of the SPARROW formulation, based on the coordinate descent method, in [28].

5. RELATION TO PRIOR WORK

In recent years, numerous publications have considered SSR from MMVs. In this section we provide a comparison of the $\ell_{2,1}$ mixednorm minimization problem, and our compact reformulations, with the atomic norm minimization approach [21–23], which shows particular similarities to our proposed SPARROW formulation. An additional comparison to the SPICE method [24, 25] is provided in [28].

In [9, 10] Atomic Norm Minimization (ANM) was introduced for gridless line spectral estimation from SMVs in ULAs. The extension of ANM to MMVs under this setup was studied in [21– 23], which will be revised in the following. Consider the noisefree MMV matrix $\mathbf{Y}_0 = \sum_{l=1}^{L} \boldsymbol{a}(\mu_l) \boldsymbol{\psi}_l^{\mathsf{T}}$, obtained at the output of a ULA for *L* impinging source signals with spatial frequencies μ_1, \ldots, μ_L , where the *l*th source signal is contained in the $N \times 1$ vector $\boldsymbol{\psi}_l$. In the ANM framework [21–23], the MMV matrix \mathbf{Y}_0 is considered as a weighted superposition of atoms $\boldsymbol{a}(\nu)\boldsymbol{b}^{\mathsf{H}}$ with $\nu \in [-1, 1)$, $\boldsymbol{b} \in \mathbb{C}^N$ and $\|\boldsymbol{b}\|_2 = 1$. The atomic norm of \mathbf{Y}_0 is defined as

$$\|\boldsymbol{Y}_{0}\|_{\mathcal{A}} = \inf_{\{c_{k},\boldsymbol{b}_{k},\nu_{k}\}} \Big\{ \sum_{k} c_{k} : \boldsymbol{Y}_{0} = \sum_{k} c_{k}\boldsymbol{a}(\nu_{k})\boldsymbol{b}_{k}^{\mathsf{H}}, c_{k} \ge 0 \Big\},$$
(20)

and computed by the SDP [9, 10, 21-23]

$$\|\boldsymbol{Y}_{0}\|_{\mathcal{A}} = \inf_{\boldsymbol{v},\boldsymbol{V}_{N}} \frac{1}{2} \operatorname{Tr}(\boldsymbol{V}_{N}) + \frac{1}{2M} \operatorname{Tr}(\operatorname{Toep}(\boldsymbol{v}))$$
(21a)

s.t.
$$\begin{bmatrix} \boldsymbol{V}_N & \boldsymbol{Y}_0^{\mathsf{H}} \\ \boldsymbol{Y}_0 & \operatorname{Toep}(\boldsymbol{v}) \end{bmatrix} \succeq \boldsymbol{0}$$
 (21b)

$$\operatorname{Toep}(\boldsymbol{v}) \succeq \boldsymbol{0} \tag{21c}$$

where the Toeplitz matrix representation in the constraint (21b) relies on the assumption of a ULA, following similar arguments as for the gridless GL-SPARROW implementation discussed in Section 4. Correspondingly, the frequency estimates $\hat{\mu}$ can be recovered by Vandermonde decomposition (14). As proposed in [21–23], given a noise-corrupted MMV matrix Y as defined in (2), joint sparse recovery from MMVs can be performed by employing the atomic norm in (20) as

$$\min_{\boldsymbol{Y}_0} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{Y}_0\|_{\mathsf{F}}^2 + \lambda \sqrt{N} \|\boldsymbol{Y}_0\|_{\mathcal{A}}$$
(22)

or, equivalently, by using the SDP formulation in (21), as

$$\min_{\substack{\boldsymbol{v},\boldsymbol{V}_N,\\\boldsymbol{Y}_0}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{Y}_0\|_{\mathsf{F}}^2 + \frac{\lambda\sqrt{N}}{2} \Big(\operatorname{Tr}\big(\boldsymbol{V}_N\big) + \frac{1}{M} \operatorname{Tr}\big(\operatorname{Toep}(\boldsymbol{v})\big) \Big)$$
(23a)

s.t.
$$\begin{bmatrix} \boldsymbol{V}_N & \boldsymbol{Y}_0^{\mathsf{H}} \\ \boldsymbol{Y}_0 & \operatorname{Toep}(\boldsymbol{v}) \end{bmatrix} \succeq \boldsymbol{0}$$
 (23b)

$$\operatorname{Toep}(\boldsymbol{v}) \succeq \boldsymbol{0}. \tag{23c}$$

Problem (23) and the GL-SPARROW formulation (16) exhibit a similar structure in the objective functions and semidefinite constraints. In fact, both problems are equivalent in the sense that minimizers are related by

$$\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}} / \sqrt{N}. \tag{24}$$

The spatial frequencies of interest ν are encoded in the vectors \hat{u} and \hat{v} , as found by Vandermonde decomposition (14), such that the



Fig. 3. Average computation time; M = 6 sensors, $\Delta \mu = 0.15$, SNR = 10 dB and K = 1000 frequency grid points

GL-SPARROW problem in (12) and the ANM problem in (23) both provide the same frequency estimates. A proof of the equivalence is given in [28].

Compared to the GL-SPARROW problem in (16), the ANM problem in (23) has MN additional optimization variables in the matrix Y_0 , which need to be matched to the MMV matrix Y by an additional quadratic term in the objective function. Moreover, the size of the ANM problem (23) scales with the number of MMVs N. In contrast to this, the GL-SPARROW problem (16) can be equivalently formulated as (17), which is independent of the number of MMVs N. In this context the GL-SPARROW formulations (16) and (17) admit significantly reduced computational complexity as compared to the ANM formulation (23).

For illustration of the computational complexity, Figure 3 shows the average computation time of the different SPARROW formulations (12), (13), (16), (17) as well as the $\ell_{2,1}$ mixed-norm minimization problem (8) and the ANM problem (23). The simulations were performed in Matlab using the SeDuMi solver [30] with the CVX interface [33] for a scenario with two independent complex Gaussian sources with static spatial frequencies $\mu_1 = 0.35$ and $\mu_2 = 0.5$ and a ULA with M = 6 sensors. The signal-to-noise ratio is fixed at 10 dB while the number of MMVs N is varied. It can well be seen from Figure 3 that the SPARROW formulations outperform the competing method, which becomes most apparent for the covariancebased implementations (13) and (17) at large number of MMVs N.

6. CONCLUSION

We have considered the classical $\ell_{2,1}$ mixed-norm minimization problem for joint sparse signal reconstruction from multiple measurement vectors and derived an equivalent, compact reformulation with significantly reduced problem dimension. The variables in our compact reformulation, which we refer to as SPARROW (SPARse ROW norm reconstruction), represent the row-norms of the joint sparse signal representation. Our SPARROW reformulation shows that the signal support is fully encoded in the sample covariance matrix, instead of the instantaneous measurement vectors as might be expected from classical sparse reconstruction formulations. We furthermore presented a gridless SPARROW implementation for the special case of uniform linear sampling and established exact equivalence between the gridless SPARROW formulation and the recently proposed atomic norm minimization problem for multiple measurement vectors. However, in contrast to atomic norm minimization, our gridless SPARROW implementation shows reduced problem size, resulting in significantly reduced computational complexity.

7. REFERENCES

- R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B* (*Methodological*), vol. 58, pp. 267–288, 1996.
- [2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal On Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
- [4] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ¹ minimization," vol. 100, no. 5. National Acad Sciences, 2003, pp. 2197–2202.
- [5] S. Mallat and Z. Zhang, "Matching pursuits with timefrequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.
- [6] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.
- [7] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [8] —, "Super-resolution from noisy data," Journal of Fourier Analysis and Applications, vol. 19, no. 6, pp. 1229–1254, 2013.
- [9] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.
- [10] B. N. Bhaskar and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Sept 2011, pp. 261–268.
- [11] Y. L. Ming Yuan, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [12] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589 – 602, 2006.
- [13] B. A. Turlach, W. N. Venables, and S. J. Wright, "Simultaneous variable selection," *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.
- M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303 324, 2009.
- [15] D. Malioutov, M. Çetin, 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, 2005.
- [16] 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, 2006.
- [17] 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, July 2005.

- [18] Y. Jin and B. 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, 2011.
- [21] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, March 2016.
- [22] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.
- [23] —, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *CoRR*, vol. abs/1407.2490, 2014.
- [24] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariancebased estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb 2011.
- [25] P. Stoica, D. Zachariah, and J. Li, "Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation," *Digital Signal Processing*, vol. 33, pp. 1–12, 2014.
- [26] C. Rojas, D. Katselis, and H. Hjalmarsson, "A note on the SPICE method," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4545–4551, Sept 2013.
- [27] P. Babu and P. Stoica, "Connection between SPICE and squareroot LASSO for sparse parameter estimation," *Signal Processing*, vol. 95, pp. 10–14, 2014.
- [28] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *arXiv preprint arXiv:1606.07231*, 2016.
- [29] H. L. van Trees, Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory. New York: John Wiley & Sons, Inc., 2002.
- [30] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.
- [31] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.
- [32] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975–989, Sept 1982.
- [33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.