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 $\ell_{2,1}$ mixed-norm minimization problem for joint sparse signal reconstruction from multiple measurement vectors (MMVs). The reformulation builds upon a compact parameterization, which models the row-norms 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 $\ell_{2,1}$ mixed-norm 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 $\ell_0$ 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 $\ell_1$ 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 ROW-norm 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

$$\mu_l = \cos \theta_l \in [-1, 1],$$  \hspace{1cm} (1)
for \( l = 1, \ldots, L \), and comprise the vector \( \mathbf{\mu} = [\mu_1, \ldots, \mu_L]^\top \). 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 \( \mu \) remain constant within the entire observation time. The measurement vectors are collected in the multiple measurement vector (MMV) matrix \( \mathbf{Y} \in \mathbb{C}^{M \times N} \), where \( [\mathbf{Y}]_{m,n} \) denotes the output at sensor \( m \) in time instant \( n \). The MMV matrix is modeled as

\[
\mathbf{Y} = \mathbf{A}(\mathbf{\mu})\mathbf{\Psi} + \mathbf{N},
\]

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

\[
\mathbf{A}(\mathbf{\mu}) = [\mathbf{a}(\mu_1), \ldots, \mathbf{a}(\mu_L)],
\]

where \( \mathbf{a}(\mu_l) = [1, e^{-j\mu_l \rho_1}, \ldots, e^{-j\mu_l \rho_M}]^\top \) is the array manifold vector with \( \rho_m \in \mathbb{R} \), for \( m = 1, \ldots, M \), 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 \( \mathbf{A}(\mathbf{\mu})\mathbf{\Psi} = \mathbf{A}(\mathbf{\nu})\mathbf{X} \) as illustrated in Figure 2. In the sparse representation \( \mathbf{A}(\mathbf{\nu}) \) is an \( M \times K \) overcomplete dictionary matrix obtained by sampling the field-of-view in \( K \gg L \) spatial frequencies \( \mathbf{\nu} = [\nu_1, \ldots, \nu_K]^\top \), defined in correspondence with (3). For ease of presentation we will use the short-hand notation \( \mathbf{A} = \mathbf{A}(\mathbf{\nu}) \) in the remainder of the paper to refer to the dictionary matrix. The matrix \( \mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]^\top \in \mathbb{C}^{K \times N} \) is a sparse representation of the signal waveform matrix \( \mathbf{\Psi} \), which has non-zero rows \( \mathbf{x}_k \) only if the corresponding sampled spatial frequency \( \nu_k \) is contained in the spatial frequencies in \( \mathbf{\mu} \).

Using the sparse representation \( \mathbf{A}\mathbf{X} \), the DOA estimation problem can be formulated as the well-known convex optimization problem [11–15]

\[
\min_{\mathbf{X}} \frac{1}{2} \| \mathbf{A}\mathbf{X} - \mathbf{Y} \|_F^2 + \lambda \| \mathbf{X} \|_{p,1},
\]

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

\[
\| \mathbf{X} \|_{p,1} = \sum_{k=1}^{K} \| \mathbf{x}_k \|_p
\]

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 \( \mathbf{x}_k, k = 1, \ldots, K \), of the matrix \( \mathbf{X} \) such that the \( \ell_1 \)-norm, i.e., the summation, is performed on the \( \ell_2 \)-norms of the rows in \( \mathbf{X} \).

Given a minimizer \( \hat{\mathbf{X}} = [\hat{x}_1, \ldots, \hat{x}_K]^\top \) 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 \( \{ \hat{\nu}_k \} \) of estimated spatial frequencies

\[
\{ \hat{\nu}_k \} = \{ \nu_k \parallel \hat{x}_k \parallel_2 \neq 0; k = 1, \ldots, K \}.
\]

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 \( \mathbf{X} \).

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 \( \ell_1 \) 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

\[
\min_{\mathbf{X}} \frac{1}{2} \| \mathbf{A}\mathbf{X} - \mathbf{Y} \|_F^2 + \lambda \sqrt{N} \| \mathbf{X} \|_{2,1},
\]

is equivalent to the convex problem

\[
\min_{\mathbf{S} \in \mathbb{D}_+} \text{Tr} \left( (\mathbf{A}\mathbf{S}\mathbf{A}^\dagger + \lambda \mathbf{I}_M)^{-1} \hat{\mathbf{R}} \right) + \text{Tr}(\mathbf{S}),
\]

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

\[
\hat{\mathbf{X}} = \hat{\mathbf{S}}\mathbf{A}^\dagger(\mathbf{A}\hat{\mathbf{S}}\mathbf{A}^\dagger + \lambda \mathbf{I}_M)^{-1}\mathbf{Y}.
\]

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 \( \hat{\mathbf{S}} = \text{diag}(\hat{\delta}_1, \ldots, \hat{\delta}_K) \) contains the row-norms of the sparse signal matrix \( \hat{\mathbf{X}} = [\hat{x}_1, \ldots, \hat{x}_K]^\top \) on its diagonal according to

\[
\hat{\delta}_k = \frac{1}{\sqrt{N}} \| \hat{x}_k \|_2.
\]

for \( k = 1, \ldots, K \), such that the union support of \( \hat{\mathbf{X}} \) is equivalently represented by the support of the sparse vector of row-norms
We will refer to (9) as the SPARse ROW-norm reconstruction (SPARROW) formulation. In this regard, we emphasize that $S$ should not be mistaken for a sparse representation of the source covariance matrix, i.e., $S \neq E[X^H X]/N$. 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 $S$.

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)

\[
\begin{align*}
\min_{S,U_N} & \frac{1}{N} \text{Tr}(U_N) + \text{Tr}(S) \\
\text{s.t.} & \begin{bmatrix} U_N & Y^H \\ Y & A S A^H + \lambda I_M \end{bmatrix} \succeq 0 \\
& S \in \mathbb{D}_+ 
\end{align*}
\]  

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

\[
\begin{align*}
\min_{S,U_M} & \text{Tr}(U_M \hat{R}) + \text{Tr}(S) \\
\text{s.t.} & \begin{bmatrix} U_M & I_M \\ I_M & A S A^H + \lambda I_M \end{bmatrix} \succeq 0 \\
& S \in \mathbb{D}_+ 
\end{align*}
\]

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 $A = [a(\nu_1), \ldots, a(\nu_M)]$ is a Vandermonde matrix of size $M \times K$. Under the given assumptions, the matrix product $A S A^H$ exhibits a Toeplitz structure according to

\[
\text{Toep}(u) = A S A^H = \sum_{k=1}^{K} s_k a(\nu_k) a^H(\nu_k),
\]

where $\text{Toep}(u)$ denotes a Hermitian Toeplitz matrix with $u$ as its first column. As discussed in [9], by the Carathéodory theorem any Toeplitz matrix $\text{Toep}(u)$ can be represented by a Vandermonde decomposition according to (14) for any distinct frequencies $\nu_1, \ldots, \nu_K$ and corresponding magnitudes $s_1, \ldots, s_K > 0$, with $\text{rank}(\text{Toep}(u)) = K \leq M$. Given a Toeplitz matrix $\text{Toep}(u)$, the Vandermonde decomposition according to (14) can be obtained by first recovering the frequencies $\nu_k$, 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 $s = [s_1, \ldots, s_K]$ can be reconstructed by solving the linear system

\[
A s = u,
\]

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

\[
\begin{align*}
\min_{u, U_N} & \frac{1}{N} \text{Tr}(U_N) + \frac{1}{M} \text{Tr}(\text{Toep}(u)) \\
\text{s.t.} & \begin{bmatrix} U_N & Y^H \\ Y & \text{Toep}(u) + \lambda I_M \end{bmatrix} \succeq 0 \\
& \text{Toep}(u) \succeq 0
\end{align*}
\]

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

\[
\begin{align*}
\min_{u, U_M} & \text{Tr}(U_M \hat{R}) + \frac{1}{M} \text{Tr}(\text{Toep}(u)) \\
\text{s.t.} & \begin{bmatrix} U_M & I_M \\ I_M & \text{Toep}(u) + \lambda I_M \end{bmatrix} \succeq 0 \\
& \text{Toep}(u) \succeq 0
\end{align*}
\]

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

In Corollary 2 we additionally make use of the identity

\[
\text{Tr}(S) = \frac{1}{M} \text{Tr}(A S A^H) = \frac{1}{M} \text{Tr}(\text{Toep}(u)),
\]

with the factor $1/M$ resulting from the fact that $||a(\nu)||^2_2 = M$, 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

\[
\hat{L} = \text{rank}(\text{Toep}(\hat{u})),
\]

while the frequencies $\hat{\nu}_l$ and corresponding magnitudes $\hat{s}_l$ can be estimated by Vandermonde decomposition according to (14), as discussed above. With the frequencies in $\hat{\nu}_l$ and signal magnitudes in $\hat{\nu}_l$, 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 \( \ell_{2,1} \) mixed-norm 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 ULA.s. The extension of ANM to MMVs under this setup was studied in [21–23], which will be revised in the following. Consider the noise-free MMV matrix \( Y_0 = \sum_{l=1}^{L} a(\mu_l) \psi_l^T \), obtained at the output of a ULA for \( L \) impinging source signals with spatial frequencies \( \mu_1, \ldots, \mu_L \), where the \( l \)th source signal is contained in the \( N \times 1 \) vector \( \psi_l \). In the ANM framework [21–23], the MMV matrix \( Y_0 \) is considered as a weighted superposition of atoms \( a(\nu) b^H \) with \( \nu \in [-1, 1] \), \( b \in \mathbb{C}^N \) and \( \|b\|_2 = 1 \). The atomic norm of \( Y_0 \) is defined as

\[
\|Y_0\|_A = \inf_{(c_k, b_k, \nu_k)} \left\{ \sum_k c_k : Y_0 = \sum_k c_k a(\nu_k) b_k^H, c_k \geq 0 \right\},
\]

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

\[
\|Y_0\|_A = \inf_{V_N} \frac{1}{2} \text{Tr}(V_N) + \frac{1}{2M} \text{Tr}(\text{Toep}(v))
\]

s.t.

\[
\begin{bmatrix} V_N & Y_0^H \\ Y_0 & \text{Toep}(v) \end{bmatrix} \succeq 0
\]

(21b)

\[
\text{Toep}(v) \succeq 0.
\]

(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

\[
\min_{Y_0} \frac{1}{2} \|Y - Y_0\|^2 + \lambda \sqrt{N} \|Y_0\|_A
\]

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

\[
\min_{V_N, Y_0} \frac{1}{2} \|Y - Y_0\|^2 + \frac{\lambda \sqrt{N}}{2} \left( \text{Tr}(V_N) + \frac{1}{M} \text{Tr}(\text{Toep}(v)) \right)
\]

s.t.

\[
\begin{bmatrix} V_N & Y_0^H \\ Y_0 & \text{Toep}(v) \end{bmatrix} \succeq 0
\]

(23b)

\[
\text{Toep}(v) \succeq 0.
\]

(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

\[
\hat{u} = \hat{v}/\sqrt{N}.
\]

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

\[\text{Av. comp. time in secs} \]
7. REFERENCES


