A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays

Dmitry Malioutov, Müjdat Çetin, and Alan Willsky

Department of Electrical Engineering and Computer Science,

Massachusetts Institute of Technology,

77 Massachusetts Avenue, Cambridge, MA, 02139

Abstract

We present a source localization method based upon a sparse representation of sensor measurements with an overcomplete basis composed of samples from the array manifold. We enforce sparsity by imposing penalties based on the ℓ_1-norm. A number of recent theoretical results on sparsifying properties of ℓ_1 penalties justify this choice. Explicitly enforcing the sparsity of the representation is motivated by a desire to obtain a sharp estimate of the spatial spectrum which exhibits superresolution. We propose to use the singular value decomposition (SVD) of the data matrix to summarize multiple time or frequency samples. Our formulation leads to an optimization problem, which we solve efficiently in a second-order cone (SOC) programming framework by an interior point implementation. We propose a grid refinement method to mitigate the effects of limiting estimates to a grid of spatial locations, and also introduce an automatic selection criterion for the regularization parameter involved in our approach. We demonstrate the effectiveness of the method on simulated data by plots of spatial spectra and by comparing the estimator variance to the Cramer-Rao bound (CRB). We observe that our approach has a number of advantages over other source localization techniques including increased resolution; improved robustness to noise, limitations in data quantity, and correlation of the sources; as well as not requiring an accurate initialization.

This work was supported by the Army Research Office under Grant DAAD19-00-1-0466, and the Air Force Office of Scientific Research under Grant F49620-00-1-0362.

I. Introduction

Source localization using sensor arrays, [1], [2], has been an active research area, playing a fundamental role in many applications involving electromagnetic, acoustic, and seismic sensing. An important goal for source localization methods is to be able to locate closely spaced sources in presence of considerable noise. Many advanced techniques for the localization of point sources achieve superresolution by exploiting the presence of *a small number* of sources. For example, the key component of the MUSIC method [3] is the assumption of a low-dimensional signal subspace. We follow a different approach for exploiting such structure: we pose source localization as an overcomplete basis representation problem, where we impose a penalty on the lack of sparsity of the spatial spectrum.

Our approach is distinctly different from the existing source localization methods, although it shares some of their ingredients. The most well-known existing non-parametric methods include beamforming [2], Capon's method [4], and subspace-based methods such as MUSIC [3]. Some additional methods (Root-MUSIC and ESPRIT) [1] require the assumption that the array of sensors is linear. Beamforming spectrum suffers from the Rayleigh resolution limit which is independent of the SNR. MUSIC and Capon's method are able to resolve sources within a Rayleigh cell (i.e. achieve superresolution) provided that the SNR is moderately high, the sources are not strongly correlated, and the number of snapshots is sufficient. A family of parametric methods based on the maximum likelihood paradigm, including deterministic maximum likelihood (DML) and stochastic maximum likelihood (SML) [1] enjoy excellent statistical properties, but an accurate initialization is required to converge to a global minimum. By turning to the sparse signal representation framework, we are able to achieve superresolution without the need for a good initialization, without a large number of time samples, and with lower sensitivity to SNR and to correlation of the sources.

The topic of sparse signal representation has evolved very rapidly in the last decade, finding application in a variety of problems including image reconstruction and restoration [5], wavelet denoising [6], feature selection in machine learning [7], radar imaging [8], and penalized regression [9]. There has also been some emerging investigation of these ideas in the context of spectrum estimation and array processing [10]–[14]. Sacchi *et al.* [10] use a Cauchy-prior to enforce sparsity in spectrum estimation, and solve the resulting optimization problem by iterative methods. Jeffs [11] uses an ℓ_p-norm penalty with $p \leq 1$ to enforce sparsity for a number of applications, including sparse antenna array design. Gorodnitsky *et al.* [12] apply a recursive weighted minimum-norm algorithm

called FOCUSS to achieve sparsity in the problem of source localization. It has later been shown [15] that the algorithm is related to the optimization of ℓ_p penalties with $p \leq 1$. The work of Fuchs [13], [14], is concerned with source localization in the beamspace domain, under the assumption that the sources are uncorrelated, and that a large number of time samples is available. The method attempts to represent the vector of beamformer outputs to unknown sources as a sparse linear combination of vectors from a basis of beamformer outputs to isolated unit power sources. The method uses ℓ_1 penalty for sparsity, and ℓ_2 penalty for noise. Prior research has established sparse signal representation as a valuable tool for signal processing, but its application to source localization has been developed only for very limited scenarios. We start with the ideas of enforcing sparsity by ℓ_1 penalties, and extend them to a general framework that is applicable to a wide variety of practical source localization problems.

In its most basic form, the problem of sparse signal representation in overcomplete bases asks to find the sparsest signal $\mathbf{x}$ to satisfy $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is an overcomplete basis, i.e. $M < N$. Without the sparsity prior on $\mathbf{x}$, the problem $\mathbf{y} = \mathbf{A}\mathbf{x}$ is ill-posed, and has infinitely many solutions. Additional information that $\mathbf{x}$ should be sufficiently sparse allows one to get rid of the ill-posedness. Solving problems involving sparsity typically requires combinatorial optimization, which is intractable even for modest data sizes, so a number of relaxations have been considered [16]–[19]. We give a brief synopsis of relevant ideas in sparse signal representation in Section II.

The application of this methodology to practical array processing problems requires being able to handle additive noise, using multiple time or frequency samples from possibly strongly correlated sources in a sensible fashion, and allowing the data to be complex:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{n}(t) \tag{1}$$

The goal of this paper is to explore how to utilize the sparse signal representation methodology for practical narrowband and wideband source localization using sensor arrays. The main contributions of our paper include a new adaptation of sparse signal representation to source localization, through the development of an approach based on the singular value decomposition (SVD) to combine multiple samples, and the use of second order cone programming for optimization of the resulting objective function. The key ingredients of the proposed method is the use of SVD for data reduction and the formulation of a joint multiple-sample sparse representation problem in the signal subspace domain. In the body of the paper we refer to the method as ℓ_1-SVD. In addition, we introduce the idea of adaptive grid refinement to combat the effects of a bias introduced by a limitation of the estimates to a grid.

Finally, we discuss a method for the automatic selection of the regularization parameter involved in our approach, which balances data-fidelity with sparsity in the ℓ_1-SVD objective. In our experiments, the proposed approach exhibits a number of advantages over other source localization techniques, which include increased resolution, improved robustness to noise, to limited number of snapshots, and to correlation of the sources. Also, due to the convexity of all the optimization tasks involved in the approach, it does not require an accurate initialization. Another advantage of the approach is its flexibility, since few assumptions are made in the formulation, e.g. the array does not have to be linear, and the sources may be strongly correlated. Similarly, extensions to many scenarios, such as distributed sources, and non-Gaussian noise, can be readily made. In the paper, we mostly focus on the narrowband farfield problem with arbitrary array geometry; we also describe the wideband scenario briefly in Section VIII. A more extensive discussion can be found in [20], where we also consider beamspace versions, cover wideband and nearfield processing in more detail, and propose an approach for simultaneous self-calibration and source localization in the presence of model errors.

We start with a brief introduction to the problem of sparse signal representation in Section II. In Section III, we describe the source localization problem, and represent a single sample problem directly in the sparse signal representation framework. In Section IV we extend the approach to handle multiple samples. This is done in several steps, leading to the ℓ_1-SVD technique. In Section V we describe how to find numerical solutions via a second order cone programming (SOC) framework. We describe how to eliminate the effects of the grid in Section VI, and propose how to automatically choose a regularization parameter involved in our approach in Section VII. Finally, in Section VIII the advantages and disadvantages of the framework are explored using simulated experiments, and conclusions are made in Section IX.

II. Sparse signal representation

The simplest version of the sparse representation problem without noise is to find a sparse $\mathbf{x} \in \mathbb{C}^N$, given $\mathbf{y} \in \mathbb{C}^M$, which are related by $\mathbf{y} = \mathbf{A}\mathbf{x}$, with $M < N$. The matrix $\mathbf{A}$ is known. The assumption of sparsity of $\mathbf{x}$ is crucial, since the problem is ill-posed without it ($\mathbf{A}$ has a nontrivial null-space). An ideal measure of sparsity is the count of non-zero entries $\mathbf{x}$, denoted by $\|\mathbf{x}\|_0^0$, which we also call the ℓ_0-norm[1]. Hence, mathematically, we must

[1]The symbols $\|\mathbf{x}\|_0$ and $\|\mathbf{x}\|_0^0$ are both used in the literature to represent the count of nonzero elements. We use the latter symbol since in the limit as $p \to 0^+$, $\|\mathbf{x}\|_p^p$ approaches the count of nonzero elements, but, if $\mathbf{x} \neq \mathbf{0}$, $\|\mathbf{x}\|_p \to \infty$.

look for $\arg\min \|\mathbf{x}\|_0^0$, such that $\mathbf{y} = \mathbf{A}\mathbf{x}$. This is, however, a difficult combinatorial optimization problem, and intractable for even moderately-sized problems. Many approximations have been devised over the years, including greedy approximations (matching pursuit, stepwise regression, and their variants [17], [19]) , and also ℓ_1 and ℓ_p relaxations, where $\|\mathbf{x}\|_0^0$ is replaced by $\|\mathbf{x}\|_1$, [16], and $\|\mathbf{x}\|_p^p$, for $p < 1$, [20]. For the latter two, it has been shown recently that if $\mathbf{x}$ is "sparse enough" with respect to $\mathbf{A}$, then these approximations in fact lead to exact solutions (see [18], [20]–[24] for precise definitions of these notions)[2]. In addition, [26], [27] showed that with sufficient sparsity and a favorable structure of the overcomplete basis, sparse representations are stable in the presence of noise. These results are practically very significant, since the ℓ_1 relaxation, $\min \|\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$ is a convex optimization problem, and the global optimum can be found for real-valued data by linear programming[3]. As these equivalence results are not specialized to the source localization problem, but are derived for general overcomplete bases, the bounds that they provide are loose. A result which does take the structure of the basis into account is developed in [28].

In practice, a noiseless measurement model is rarely appropriate, so noise must be introduced. A sparse representation problem with additive Gaussian noise takes the following form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{2}$$

To extend ℓ_1-penalization to the noisy case, an appropriate choice of an optimization criterion is $\min \|\mathbf{x}\|_1$ subject to $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \beta^2$, where β is a parameter specifying how much noise we wish to allow. An unconstrained form of this objective is:

$$\min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \tag{3}$$

This objective function has been used in a number of sparse signal representation works: [16], [29] for real-valued data and [30] for complex-valued data. The ℓ_2-term forces the residual $\mathbf{y} - \mathbf{A}\mathbf{x}$ to be small, while the ℓ_1-term enforces sparsity of the representation. The parameter λ controls the trade-off between the sparsity of the spectrum and the residual norm. We use these ideas in Sections III and IV for source localization.

The optimization criterion is again a convex optimization problem and can be readily handled by quadratic

[2]Recent studies of greedy methods, which have lower complexity than ℓ_1 and ℓ_p-based methods, have also yielded theoretical results of similar flavor [25], [26].

[3]Also, for the ℓ_p problem, local minima can be readily found by continuous optimization methods, as described in [20].

programming for real data. We propose the use of second order cone (SOC) programming for the complex data case. We describe SOC programming in Section V.

The class of methods called FOCUSS [12] is another paradigm for solving sparse signal representation problems with a more general ℓ_p penalty instead of ℓ_1. However, for $p < 1$, the cost function is non-convex, and the convergence to global minima is not guaranteed. The discussion in [15] in Section VI indicates that the best results are obtained for p close to 1, while the convergence is also slowest for $p = 1$. The cost per iteration for FOCUSS methods is similar to that of an interior point solver for SOC, since both solve a modified Newton's method step of similar dimensions. However, the number of iterations of SOC is better behaved (in fact, there are bounds on the worst-case number of iterations for SOC [33]) than for FOCUSS with $p = 1$. In our previous work [20], we have also observed slow convergence of iterative algorithms for ℓ_p minimization when applied with $p = 1$. By using an SOC formulation which is tailored to the convex ℓ_1 case we are able to achieve fast convergence and guarantee global optimality of the solution.

III. Source localization framework

A. Source localization problem

The goal of sensor array source localization is to find the locations of sources of wavefields which impinge upon an array consisting of a number of sensors. The available information is the geometry of the array, the parameters of the medium where wavefields propagate, and the measurements on the sensors.

For purposes of exposition, we first focus on the narrowband scenario, and delay the presentation of wideband source localization until Section VIII-D. Consider K narrowband signals $u_k(t)$, $k \in \{1, ..., K\}$, arriving at an array of M omnidirectional sensors, after being corrupted by additive noise $n_m(t)$, resulting in sensor outputs $y_m(t)$, $m \in \{1, ..., M\}$. Let $\mathbf{y}(t) = [y_1(t), ..., y_M(t)]'$, and similarly define $\mathbf{u}(t)$, and $\mathbf{n}(t)$. After demodulation, the basic narrowband observation model can be expressed as [1], [2]:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{u}(t) + \mathbf{n}(t), \ t \in \{t_1, .., t_T\} \tag{4}$$

The matrix $\mathbf{A}(\boldsymbol{\theta})$ is the so-called array manifold matrix, whose (m, k)-th element contains the delay and gain information from the k-th source (at location θ_k) to the m-th sensor. The columns $\mathbf{a}(\theta_k)$ of $\mathbf{A}(\boldsymbol{\theta})$, for $k \in \{1, .., K\}$, are called steering vectors. The number of sources K is unknown. To simplify the exposition, we only discuss the

farfield scenario and also confine the array to a plane, although neither of these assumptions is required for our approach. With farfield sources in the same plane as the array, the unknown locations of the sources are parameterized by angles (directions of arrival) with respect to the array axis, $\boldsymbol{\theta} = [\theta_1, ..., \theta_K]$. Given the knowledge of $\mathbf{y}(t)$, and the mapping $\boldsymbol{\theta} \rightarrow \mathbf{A}(\boldsymbol{\theta})$, the goal is to find the unknown locations of the sources, θ_k for all k, as well as their number K.

B. Overcomplete representation for a single time sample

Now we start to formulate the source localization problem as a sparse representation problem. The single-sample formulation in this section parallels the one in [12], where it was presented as one of applications of FOCUSS algorithm. In addition, the work in [13], [14] is based on a similar philosophy of transforming a parameter estimation problem into sparse spectrum estimation, which we discuss later in this section.

We consider the single time sample case in this section, with $T = 1$ in (4). The problem as it appears in (4) is a nonlinear parameter estimation problem, where the goal is to find $\boldsymbol{\theta}$. Matrix $\mathbf{A}(\boldsymbol{\theta})$ depends on the unknown source locations $\boldsymbol{\theta}$, so it is not known.

To cast this problem as a sparse representation problem, we introduce an overcomplete representation $\mathbf{A}$ in terms of all possible source locations. Let $\{\tilde{\theta}_1, ..., \tilde{\theta}_{N_\theta}\}$ be a sampling grid of all source locations of interest. The number of potential source locations N_θ will typically be much greater than the number of sources K or even the number of sensors M. We construct a matrix composed of steering vectors corresponding to each potential source location as its columns: $\mathbf{A} = [\mathbf{a}(\tilde{\theta}_1), \mathbf{a}(\tilde{\theta}_2), ..., \mathbf{a}(\tilde{\theta}_{N_\theta})]$. In this framework $\mathbf{A}$ is known and does not depend on the actual source locations, $\boldsymbol{\theta}$.

We represent the signal field by an $N_\theta \times 1$ vector $\mathbf{s}(t)$, where the n-th element $s_n(t)$ is nonzero and equal to $u_k(t)$ if source k comes from $\tilde{\theta}_n$, for some k, and zero otherwise. For a single time sample the problem is reduced to

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{5}$$

In effect, this overcomplete representation allows us to exchange the problem of parameter estimation of $\boldsymbol{\theta}$ for the problem of sparse spectrum estimation of $\mathbf{s}$. As in numerous non-parametric source localization techniques, the approach forms an estimate of the signal energy as a function of hypothesized source location, which ideally

contains dominant peaks at the true source locations. The central assumption is that the sources can be viewed as point sources, and their number is small. With this assumption the underlying spatial spectrum is sparse (i.e. $\mathbf{s}$ has only a few nonzero elements), and we can solve this inverse problem via regularizing it to favor sparse signal fields using the ℓ_1 methodology, as described in Section II. The appropriate objective function for the problem is:

$$\min \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1 \tag{6}$$

We discuss how λ is chosen in Section VII, but for now we assume that a good choice can be made. The data for the model is complex-valued, hence neither linear nor quadratic programming can be used for numerical optimization. Instead, we adopt a second-order cone (SOC) programming framework, which we introduce in Section V. Once $\mathbf{s}$ is found, the estimates of the source locations correspond to the locations of the peaks in $\mathbf{s}$.

We illustrate the approach for source localization with a *single time sample* in Figure 1. We consider a uniform linear array of $M = 8$ sensors separated by half a wavelength of the actual narrowband source signals. We consider two narrowband signals in the far-field impinging upon this array from DOA's 60° and 70°, which are closer together than the Rayleigh limit. The SNR is 20 dB. The regularization parameter λ in this example is chosen by subjective assessment. We do not consider other source localization methods such as MUSIC or Capon's method in this simulation because they rely on estimating the covariance matrix of the sensor measurements, but in the simulation only one time sample is present. Using beamforming, the two peaks of the spectrum are merged, but using the sparse regularization approach they are well resolved, and the sidelobes are suppressed almost to zero. Apart from a small asymptotic bias, which we discuss in Section VIII, the spectrum estimate is an example of what superresolution source localization methods aim to achieve.

The work of Fuchs [13], [14] is based on a similar philosophy of transforming a parameter estimation problem into a sparse spectrum estimation problem. A basis composed of beamformer outputs to isolated unit power sources from a large number of directions is created first. The method then attempts to represent the vector of beamformer outputs corresponding to the unknown sources as a sparse linear combination of vectors from the basis, using ℓ_1 penalties for sparsity, ℓ_2 penalties for noise, and optimization by quadratic programming. However, this beamspace domain formulation combines the multiple snapshots in a way that requires assumptions that the sources are uncorrelated and that a large number of samples is available. In contrast, the sensor-domain method that we propose in Section

IV-C treats the multiple time samples in a very different way: we summarize multiple snapshots by using the singular value decomposition, and solve a joint optimization problem over several singular vectors, imposing a penalty which enforces the same sparsity profile over all these vectors, thus imposing temporal coherence. The resulting formulation is considerably more general than the one in [14].

IV. Source localization with multiple time samples and ℓ_1-SVD

Single snapshot processing may have its own applications, but source localization with multiple snapshots[4] from potentially correlated sources is of greater practical importance. When we bring time into the picture, the overcomplete representation is easily extended. The general narrowband source localization problem with multiple snapshots reformulated using an overcomplete representation has the following form:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t),\ t \in \{t_1, ..t_T\} \tag{7}$$

However, the numerical solution of this problem is a bit more involved than that of the single sample case. In Section IV-A we describe a simple and computationally efficient method which, however, does not use the snapshots in synergy. In Section IV-B we propose a coherent method which does use the snapshots in synergy but is more demanding computationally, and in Section IV-C we develop an SVD-based approach which dramatically reduces the computational complexity while still using the snapshots coherently.

A. Treating each time index separately

The first thought that comes to mind when we switch from one time sample to several time samples is to solve each problem indexed by t separately. In that case we would have a set of T solutions, $\hat{\mathbf{s}}(t)$. If the sources are moving fast, then the evolution of $\hat{\mathbf{s}}(t)$ is of interest, and the approach is suitable for displaying it. However, when the sources are stationary over several time samples, then it is preferable to combine the independent estimates $\hat{\mathbf{s}}(t)$ to get one representative estimate of source locations from them, for example by averaging or by clustering. This is non-coherent averaging, and its main attraction is its simplicity. However, by turning to fully coherent combined processing, as described in the following sections, we expect to achieve greater accuracy and robustness to noise.

[4]While here we focus on multiple time snapshots, we will also use the same ideas applied to frequency snapshots for wideband source localization in Section VIII.

B. Joint-time inverse problem

Now we consider a simple approach which uses different time samples in synergy. Let $\mathbf{Y} = [\mathbf{y}(t_1), ..., \mathbf{y}(t_T)]$, and define $\mathbf{S}$ and $\mathbf{N}$ similarly. Then equation (7) becomes

$$\mathbf{Y} = \mathbf{AS} + \mathbf{N} \tag{8}$$

There is an important difference of equation (8) from (5): matrix $\mathbf{S}$ is parameterized temporally and spatially, but sparsity only has to be enforced in space, since the signal $\mathbf{s}(t)$ in not generally sparse in time. To accommodate this issue, we impose a different prior, one that requires sparsity in the spatial dimension, but does not require sparsity in time. This can be done[5] by first computing the ℓ_2-norm of all time-samples of a particular spatial index of $\mathbf{s}$, i.e. $s_i^{(\ell_2)} = \|[s_i(t_1),\ s_i(t_2),\ ...,\ s_i(t_T)]\|_2$, and penalizing the ℓ_1-norm of $\mathbf{s}^{(\ell_2)} = \left[s_1^{(\ell_2)}, ..., s_{N_\theta}^{(\ell_2)}\right]$. The cost function becomes

$$\min \|\mathbf{Y} - \mathbf{AS}\|_f^2 + \lambda \|\mathbf{s}^{(\ell_2)}\|_1 \tag{9}$$

The Frobenius norm is defined as $\|\mathbf{Y} - \mathbf{AS}\|_f^2 = \|\text{vec}(\mathbf{Y} - \mathbf{AS})\|_2^2$. The optimization is performed over $\mathbf{S}$; $\mathbf{s}^{(\ell_2)}$ is a function of $\mathbf{S}$. The time samples are combined using the 2-norm which has no sparsifying effects. The spatial samples are combined using the ℓ_1-norm which does enforce sparsity. Compared to the independent sample by sample processing from Section IV-A, the different time-indices of $\mathbf{s}$ reinforce each other, since the penalty is higher if the supports of $\mathbf{s}(t)$ for different t do not line up exactly. Once an estimate of $\mathbf{S}$ is computed using the new cost function, the peaks of $\mathbf{S}$ provide the source locations.

The main drawback of this technique is its computational cost. The size of the inverse problem increases linearly with T and the computational effort required to solve it increases superlinearly with T. Thus when T is large, this approach is not viable for the solution of the real-time source localization problem. We propose a solution to this problem next.

C. ℓ_1-SVD

In this section we present a tractable approach to use a large number of time samples coherently, thus extending the use of sparse signal representation ideas for practical source localization problems. To reduce both the computational

[5]It came to our attention that a similar idea has been used in [30] for basis selection.

complexity, and the sensitivity to noise, we use the singular value decomposition (SVD) of the $M \times T$ data matrix $\mathbf{Y} = [\mathbf{y}(t_1), ..., \mathbf{y}(t_T)]$. The idea is to decompose the data matrix into the signal and noise subspaces, keep the signal subspace, and mold the problem with reduced dimensions into the multiple-sample sparse spectrum estimation problem in the form of Section IV-B. Note that we keep the signal subspace, and not the noise subspace which gets used in MUSIC, Pisarenko, and the minimum norm subspace methods.

Without noise on the sensors, the set of vectors $\{\mathbf{y}(t_i)\}_{i=1}^{T}$ would lie in a K-dimensional subspace, where K is the number of sources[6]. We would only need to keep a basis for the subspace (K vectors instead of T) to estimate what sparse combinations of columns of $\mathbf{A}$ form it. With additive noise, we decompose the data matrix into its signal and noise subspaces, and keep a basis for the signal subspace. Mathematically, this translates into the following representation. Take the singular value decomposition[7]:

$$\mathbf{Y} = \mathbf{ULV}' \tag{10}$$

Keep a reduced $M \times K$ dimensional matrix $\mathbf{Y}_{SV}$, which contains most of the signal power, $\mathbf{Y}_{SV} = \mathbf{ULD}_K = \mathbf{YVD}_K$, where $\mathbf{D}_K = [\mathbf{I_K} \;\; \mathbf{0}]'$. Here $\mathbf{I_K}$ is a $K \times K$ identity matrix, and $\mathbf{0}$ is a $K \times (T-K)$ matrix of zeros. Also, let $\mathbf{S}_{SV} = \mathbf{SVD}_K$, and $\mathbf{N}_{SV} = \mathbf{NVD}_K$, to obtain

$$\mathbf{Y}_{SV} = \mathbf{AS}_{SV} + \mathbf{N}_{SV} \tag{11}$$

Now let us consider this equation column by column (each column corresponds to a signal subspace singular vector):

$$\mathbf{y}^{SV}(k) \;=\; \mathbf{A}\mathbf{s}^{SV}(k) \;+\; \mathbf{n}^{SV}(k), \quad k \;=\; 1, .., K \tag{12}$$

This is now in exactly the same form as the original multiple time sample problem (7), except that instead of indexing samples by time, we index them by the singular vector number. What we have achieved by bringing the SVD transformation into the picture is the reduction of the size of the problem in Section IV-B from T blocks of data to K, where K is the number of sources. For typical situations where the number of sources is small, and the number of time samples may be in the order of hundreds, this reduction in complexity is very substantial.

[6]If $T < K$, or if the sources are coherent, we use the number of signal subspace singular values instead of K.

[7]This is closely related to the eigen-decomposition of the correlation matrix of the data: $\mathbf{R} = \frac{1}{T}\mathbf{YY}'$. Its eigen-decomposition is $\mathbf{R} = \frac{1}{T}\mathbf{ULV}'\mathbf{VL}'\mathbf{U}' = \frac{1}{T}\mathbf{UL}^2\mathbf{U}'$.

If we think of $\mathbf{S}_{SV}$ as a 2-D field, indexed by i in the spatial dimension, and by k in terms of the singular vector index, then we again want to impose sparsity in $\mathbf{S}_{SV}$ only spatially (in terms of i), and not in terms of the singular vector index k. Similarly to Section IV-B, we define $\tilde{s}_i^{(\ell_2)} = \sqrt{\sum_{k=1}^{K}(s_i^{SV}(k))^2}$, $\forall i$. The sparsity of the resulting $N_\theta \times 1$ vector $\tilde{\mathbf{s}}^{(\ell_2)}$ corresponds to the sparsity of the spatial spectrum. We can find the spatial spectrum of $\tilde{\mathbf{s}}$ by minimizing

$$\|\mathbf{Y}_{SV} - \mathbf{A}\mathbf{S}_{SV}\|_f^2 + \lambda\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \tag{13}$$

We illustrate the steps for ℓ_1-SVD method in Figure 2.

Note that our formulation uses information about the number of sources K. However, we empirically observe that incorrect determination of the number of sources in our framework has no catastrophic consequences (such as complete disappearance of some of the sources as may happen with MUSIC), since we are not relying on the structural assumptions of the orthogonality of the signal and noise subspaces. Underestimating or overestimating K manifests itself only in gradual degradation of performance. This is illustrated in Section VIII.

V. SOC REPRESENTATION OF THE ℓ_1-SVD PROBLEM

Now that we have an objective function in (13) to minimize, we would like to do so in an efficient manner. The objective contains a term $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 = \sum_{i=1}^{N_\theta}\sqrt{\sum_{k=1}^{K}(s_i^{SV}(k))^2}$, which is neither linear nor quadratic. We turn to second order cone (SOC) programming [31], which deals with the so-called second order cone constraints of the form $\mathbf{s} : \|s_1, ..., s_{n-1}\|_2 \leq s_n$, i.e. $\sqrt{\sum_{i=1}^{n-1}(s_i)^2} \leq s_n$. SOC programming is a suitable framework for optimizing functions which contain SOC, convex quadratic, and linear terms. The main reason for considering SOC programming instead of generic nonlinear optimization for our problem is the availability of efficient interior point algorithms for the numerical solution of the former, e.g. [32]. In addition to efficient numerical solution, SOC programming has a substantial theoretical foundation, as a special case of semidefinite programming and convex conic programming. We refer the reader to [31] for details, and we describe in the Appendix how to manipulate

the problem in (13) into the second order cone programming form:

$$\min \ p + \lambda q \tag{14}$$

$$\text{subject to } \|(\mathbf{z}'_1, ..., \mathbf{z}'_K)\|_2^2 \leq p, \text{ and } , \mathbf{1}'\mathbf{r} \leq q,$$

$$\text{where } \sqrt{\sum_{k=1}^{K} (s_i^{SV}(k))^2} \leq r_i, \text{ for } i = 1, .., N_\theta$$

$$\text{and } \mathbf{z}_k = \mathbf{y}^{SV}(k) - \mathbf{A}\mathbf{s}^{SV}(k), \text{ for } k = 1, .., K$$

For the numerical solution of our second order cone problem, we use a package for optimization over self-dual homogeneous cones (which includes direct products of the positive orthant-constraints, SOC constraints and semidefinite cone constraints), called SeDuMi [32]. In terms of computational complexity, the interior point method relies on iterations of modified Newton's method. One of the main attractions of interior point methods is that the number of these iterations typically stays quite low, independent of the size of the problem. For optimizing the ℓ_1-SVD objective function in SOCP framework using an interior point implementation, the cost is $O((K \times N_\theta)^3)$ with the observation that the number of iterations is empirically almost independent of the size of the problem [33] (a theoretical worst-case bound on the number of iterations is $O((K \times N_\theta)^{0.5})$ [33]). The computational complexity is higher than that of [14], since we have a joint optimization problem over K singular vectors, leading to an additional factor of K^3. It is also higher than the cost of MUSIC, where the main complexity is in the subspace decomposition of the covariance matrix, which is $O(M^3)$. However, the benefit that we get in return is generality. For reference, for a problem with 3 sources impinging upon an array with 8 sensors, and having 1° sampling of the spatial location of the sources (180 points on the grid), the time required for optimization using a Matlab implementation of the code on Linux on a computer with an 800 MHz Pentium 3 processor is roughly 5 seconds, with around 20 iterations.

VI. Multi-resolution grid refinement

So far, in our framework, the estimates of the source locations are confined to a grid. We cannot make the grid very fine uniformly, since this would increase the computational complexity significantly. We explore the idea of adaptively refining the grid in order to achieve better precision. The idea is a very natural one: instead of having a universally fine grid, we make the grid fine only around the regions where sources are present. This requires an

approximate knowledge of the locations of the sources, which can be obtained by using a coarse grid first. The algorithm is the following:

1) Create a rough grid of potential source locations $\tilde{\theta}_i^{(0)}$, for $i = 1, .., N_\theta$. Set $r = 0$. The grid should not be too rough, not to introduce substantial bias. A 1° or 2° uniform sampling usually suffices.
2) Form $\mathbf{A}_r = \mathbf{A}(\tilde{\boldsymbol{\theta}}^{(r)})$, where $\tilde{\boldsymbol{\theta}}^{(r)} = [\tilde{\theta}_1^{(r)}, \ \tilde{\theta}_2^{(r)}, ..., \ \tilde{\theta}_{N_\theta}^{(r)}]$. Use our method from Section IV-C to get the estimates of the source locations, $\hat{\theta}_j^{(r)}$, $j = 1, .., K$, and set $r = r + 1$.
3) Get a refined grid $\tilde{\boldsymbol{\theta}}^{(r)}$ around the locations of the peaks, $\hat{\theta}_j^{(r-1)}$. We specify how this is done below.
4) Return to step 2 until the grid is fine enough.

Many different ways to refine the grid can be imagined; we choose simple equi-spaced grid refinement. Suppose we have a locally uniform grid (piecewise uniform), and at step r the spacing of the grid is δ_r. We pick an interval around the j-th peak of the spectrum which includes two grid spacings to either side, i.e. $[\hat{\theta}_j^{(r)} - 2\delta_r, \hat{\theta}_j^{(r)} + 2\delta_r]$, for $j = 1, .., K$. In the intervals around the peaks we select the new grid whose spacing is a fraction of the old one, $\delta_{r+1} = \frac{\delta_r}{\gamma}$. It is possible to achieve fine grids either by rapidly shrinking δ_r for a few refinement levels, or by shrinking it slowly using more refinement levels. We find that the latter approach is more stable numerically, so we typically set $\gamma = 3$, a small number. After a few (e.g. 5) iterations of refining the grid, it becomes fine enough that its effects are negligible. Figure 3 illustrates the refinement of the grid. The spacing of each of the grids corresponds to $2\delta_r$. The idea has been successfully used for some of the experimental analysis we present in Section VIII.

VII. Regularization parameter selection

An important part of our source localization framework is the choice of the regularization parameter λ in (13), which balances the fit of the solution to the data versus the sparsity prior. The same question arises in many practical inverse problems, and is difficult to answer in many cases, especially if the objective function is not quadratic. We discuss an approach to select the regularization parameter automatically for the case where some statistics of the noise are known, or can be estimated. Let us denote the estimate of the spatial spectrum obtained using λ as the regularization parameter by $\mathbf{S}(\lambda)$. A well-known idea under the name of discrepancy principle [34] is to select λ to match the residuals of the solution $\hat{\mathbf{S}}(\lambda)$ to some known statistics of the noise, when such are available.

For example, if the distribution of the noise $\mathbf{N}_{SV}$ is known or can be modeled, then one can select λ such that $\|\mathbf{Y}_{SV} - \mathbf{A}\hat{\mathbf{S}}(\lambda)\|_f^2 \approx E[\|\mathbf{N}\|_f^2]$. Here the we use the Frobenius norm, $\|\mathbf{N}\|_f^2 = \|\text{vec}(\mathbf{N})\|_2^2$. Directly searching for a value of λ to achieve the equality is rather difficult, and requires solving the problem (13) multiple times for different λ's.

Instead, we propose to look at the constrained version of the problem in (13), which can also be efficiently solved in the second order cone framework [20]:

$$\min \ \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \text{ subject to } \|\mathbf{Y}_{SV} - \mathbf{A}\mathbf{S}_{SV}\|_f^2 \leq \beta^2 \tag{15}$$

The problem in (15) is equivalent via Lagrange multipliers to the one in (13) for some parameter β which is related to λ. For the problem in (15), the task of choosing the regularization parameter β properly is considerably more transparent: we choose β high enough so that the probability that $\|\tilde{\mathbf{n}}\|_2^2 \geq \beta^2$ is small, where $\tilde{\mathbf{n}} = \text{vec}(\mathbf{N}\mathbf{V}\mathbf{D}_K)$. If $\mathbf{n}$ is i.i.d. Gaussian, then for moderate to high SNR, $\|\tilde{\mathbf{n}}\|_2^2$ has approximately a χ^2 distribution with MK degrees of freedom, upon normalization by the variance of $\mathbf{n}$. The reason that this holds only approximately is that the singular value decomposition in (10), $\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{N} = \mathbf{U}\mathbf{L}\mathbf{V}'$, depends on the particular realization of noise, and hence the matrix $\mathbf{V}$ is a function of $\mathbf{N}$. However, when noise is small, the term $\mathbf{A}\mathbf{S}$ dominates the singular value decomposition and the change due to the addition of $\mathbf{N}$ is small, and we arrive at a χ^2 distribution for $\|\tilde{\mathbf{n}}\|_2^2$. With the knowledge of the distribution, we can find a confidence interval for $\|\tilde{\mathbf{n}}\|_2^2$, and use its upper value as a choice for β^2. In simulations which we present in Section VIII, we find that this procedure generates appropriate regularization parameter choices for our problem when noise is reasonably small. We also present some thoughts on how to extend the range of the applicability of the procedure to higher levels of noise by characterizing the distribution of $\tilde{\mathbf{n}}$ for lower SNR.

When noise statistics are not known, and no knowledge of the number of sources is available, the choice of the regularization parameter is a difficult question. It has been approached in the inverse problem community by methods such as L-curve [35]. An attempt to apply the L-curve to a subset selection problem in noise has been made in [36], but the authors have to make an assumption that the SNR is approximately known. The choice of the regularization parameter when no knowledge of the noise or of the sources is available is still an open problem.

VIII. Experimental results

In this section we present several experimental results for our ℓ_1-SVD source localization scheme. First, we compare the spectra of ℓ_1-SVD to those of MUSIC [3], beamforming [2], Capon's method [4], and the beamspace method in [14] under various conditions. Next, we discuss and present results on regularization parameter selection. Then we analyze empirically the bias and variance properties of our method. Finally, in Section VIII-D we present an extension of our framework to the wideband scenario, and demonstrate its effectiveness on a number of examples.

A. Spectra for ℓ_1-SVD

We consider a uniform linear array of $M = 8$ sensors separated by half a wavelength of the actual narrowband source signals. Two zero-mean narrowband signals in the far-field impinge upon this array from distinct directions of arrival (DOA). The total number of snapshots is $T = 200$, and the grid is uniform with 1° sampling, $N_\theta = 180$. In Figure 4, we compare the spectrum obtained using our proposed method with those of beamforming, Capon's method, and MUSIC. In the top plot, the SNR is 10 dB, and the sources are closely spaced (5° separation). Our technique and MUSIC are able to resolve the two sources, whereas Capon's method and beamforming methods merge the two peaks. In the bottom plot, we decrease the SNR to 0 dB, and only our technique is still able to resolve the two sources. Next, we consider correlation between the sources, which can occur in practical array processing due to multipath effects. In Figure 5, we set the SNR to 20 dB, but make the sources strongly correlated, with a correlation coefficient of 0.99. MUSIC and Capon's method would resolve the sources at this SNR were they not correlated, but correlation degrades their performance. Again, only our technique is able to resolve the two sources. This illustrates the power of our methodology in resolving closely-spaced sources despite low SNR or correlation between the sources.

In Figure 6 we compare the spectra obtained using ℓ_1-SVD to spectra obtained using our implementation of the beamspace technique described in [14]. The top plot considers two uncorrelated sources at 63° and 73°, with $T = 200$ samples. SNR is 0 dB. As can be seen from the plot, for uncorrelated sources with $T = 200$, the assumptions made in [14] hold and the beamspace method has an excellent performance, similar to that of our ℓ_1-SVD method.

In the bottom plot, the two sources are correlated, breaking the assumption in [14]. We observe that the

performance of the beamspace technique degrades, and that strong bias appears. This bias was not present when the sources were uncorrelated. As we already noted, no such degradation appears for ℓ_1-SVD, and the spectrum is very similar to the one for the case of uncorrelated sources. In summary, our formulation is based on similar principles of enforcing sparsity as the work in [14], but it is more general in allowing correlated sources, and making no assumptions of having a large number of time samples.

So far we have shown plots resolving a small number of sources. An interesting question is to characterize the maximum number of sources that can be resolved by ℓ_1-SVD using measurements from an M-sensor array. It can be shown through simple linear algebraic arguments that M sources cannot be localized (the representation is ambiguous). However, empirically, the ℓ_1-SVD technique can resolve $M-1$ sources[8], if they are not located too close together. Hence, ℓ_1-SVD is not limited to extremely sparse spectra, but can resolve the same number of sources as MUSIC and Capon's methods. This is illustrated in Figure 7. The number of sensors in the array is again $M = 8$, and the number of sources is 7. With moderate SNR as in this example, all three techniques (ℓ_1-SVD, MUSIC, and Capon's method) exhibit peaks at the source locations.

We mentioned in Section IV-C that our approach is not very sensitive to the correct determination of the number of sources. We give an illustration of this statement in Figures 8 and 9. We use the same $M = 8$ sensor uniform linear array as before. The actual number of sources is $K = 4$, and the SNR is 10 dB. In Figure 8 we plot unnormalized (i.e. the maximum peak is not set to 1) spectra obtained using MUSIC when we vary the assumed number of sources. Underestimating the number of sources results in a strong deterioration of the quality of the spectra, including widening and possible disappearance of some of the peaks. A large overestimate of the number of sources leads to the appearance of spurious peaks due to noise. In Figure 9 we plot the unnormalized spectra obtained using ℓ_1-SVD for the same assumed numbers of sources, and the variation in the spectra is very small. The importance of the low sensitivity of our technique to the assumed number of sources is twofold. First, the number of sources is usually unknown, and low sensitivity provides robustness against mistakes in estimating the number of sources. In addition, even if the number of sources is known, low sensitivity may allow one to reduce

[8]This holds under the assumption that the number of singular vectors used in ℓ_1-SVD is sufficient, e.g. equal to the number of sources. When fewer singular vectors are taken than the number of sources, the number of resolvable sources may decrease. However, even in the extreme case of taking just one singular vector, for the 8-sensor array in the example in Figure 9, ℓ_1-SVD resolves 4 (i.e. $M/2$) sources.

the computational complexity of ℓ_1-SVD by taking a smaller number of singular vectors. With higher levels of noise, in our experiments we observe that the sensitivity of ℓ_1-SVD to the assumed number of sources increases, however, it still provides better robustness relative to MUSIC, especially when the assumed number of sources is less than the actual number of sources.

B. Regularization parameter choice

We illustrate the importance of a good choice of the regularization parameter in Figure 11. The number of sources in the example is $K = 2$, and the number of sensors and snapshots is kept as before, $M = 8$, and $T = 200$. The curve labeled "good choice" represents the selection of the regularization parameter β by the discrepancy principle from Section VII, with a 99-percent confidence interval. The spectrum is sharp and the peaks correspond to source locations. For the second curve labeled "bad choice", the regularization parameter was set 3 times lower, below the norm of the realization of the noise. In order to explain the data with such small regularization parameter spurious peaks due to noise appear in the plot. Also, if we set the regularization parameter too high, starting from about 5 times the value selected by the discrepancy principle one of the peaks would disappear, and as we increase it further, the second peak would disappear, making the spectrum 0 at all spatial locations. This example illustrates two points: the importance of a good choice of the regularization parameter, and the soundness of the approach based on the discrepancy principle.

In Section VII, in order to calculate the confidence intervals for $\|\tilde{\mathbf{n}}\|_2^2$, we had to make an assumption that noise is reasonably small. When the assumption does not hold, the singular value decomposition $\mathbf{Y} = \mathbf{AS} + \mathbf{N} = \mathbf{ULV}'$ depends on $\mathbf{N}$, and $\mathbf{N}_{SV} = \mathbf{NVD}_K$ is a complicated function of $\mathbf{N}$, since $\mathbf{V}$ now depends on $\mathbf{N}$. One approach to characterize $\|\tilde{\mathbf{n}}\|_2^2$ for higher levels of noise is through simulation. In Figure 10 we illustrate the dependence of the ratio of $\frac{\|\tilde{\mathbf{n}}\|_2}{\sigma}$ on SNR, where σ^2 is the variance of the i.i.d. Gaussian noise, $\mathbf{n}(t)$. To create the plot we first selected $K = 3$ source locations uniformly distributed in $[0, \pi]$, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$, and a corresponding signal matrix $\mathbf{S}$, with indices of non-zero rows corresponding to $\boldsymbol{\theta}$. For each choice of $\boldsymbol{\theta}$ we created 250 instances of zero-mean i.i.d. Gaussian noise matrices $\mathbf{N}$ with variance σ^2, and calculated the minimum, average, and maximum ratios $\frac{\|\tilde{\mathbf{n}}\|_2}{\sigma}$ over all 250 instances of $\mathbf{N}$. The three curves, max, min, and average ratio are plotted as a function of SNR. We superimposed these curves for 10 different realizations of $\boldsymbol{\theta}$ to show the variability. For very low SNR, noise is

dominating $\mathbf{Y}$: $\mathbf{Y} = \mathbf{AS} + \mathbf{N} \approx \mathbf{N}$, and $\|\tilde{\mathbf{n}}\|_2^2 \approx \sum_{k=1}^{K} \sigma_k^2$, where $\{\sigma_k\}_{k=1}^{K}$ are the top K singular values. For high SNR, noise has a small contribution to $\mathbf{Y}$, and $\|\tilde{\mathbf{n}}\|_2^2$ can be well predicted as described in Section VII. However, there is a sharp transition between these two regions, which we are interested in characterizing. For most triples of curves the transition occurs at the same SNR, but there are two outliers. They occur when source locations $\boldsymbol{\theta}$ are closely spaced, so that $\mathbf{A}(\boldsymbol{\theta})$ has a high condition number (recall that $\mathbf{A}(\boldsymbol{\theta})$ contains columns of $\mathbf{A}$ corresponding to $\boldsymbol{\theta}$). In that case, the effects of noise start to show up at higher SNR. The conclusion that can be drawn out of these experiments is that it is possible to predict $\|\tilde{\mathbf{n}}\|_2^2$ for higher levels of noise, but one has to be careful with closely spaced sources.

C. Bias and Variance

One aspect of our technique is the bias of the estimates that appears for closely-spaced sources. The reason for the bias is that we impose a sparsity prior in our objective function, without which the problem of estimating the spectrum is ill-posed. Other source localization methods have much difficulty resolving closely-spaced sources, especially at low SNRs, hence small bias can be considered as a good compromise, if such peaks can be resolved. We now investigate bias[9] more closely by considering source localization with two sources, and varying the angular separation between them. The number of sensors and snapshots is again $M = 8$, and $T = 200$. In Figure 12, we plot the bias of each of the two source location estimates as a function of the angular separation, when one source is held fixed at 42°. The SNR is 10 dB. The values on each curve are an average over 50 trials. The plot shows the presence of bias for low separations, but the bias disappears when sources are more than about 20 degrees apart.

We next compare the variance of the DOA estimates produced by our approach to those obtained using existing methods [1], and to the Cramer-Rao bound (CRB). In order to satisfy the assumptions of the CRB, we choose an operating point where our method is unbiased, i.e. when the sources are not very close together. In Figure 13, we present plots of variance versus SNR for a scenario including two uncorrelated sources[10]. On the plot we also include a curve labeled "oracle" maximum likelihood, which is obtained by using an ML estimate, where the nonconvex optimization is initialized to the true values of the source locations. This estimator is not practically realizable, and

[9]Our analysis of bias and variance is based on computer simulations. The work in [29] contains a theoretical analysis of bias and variance in a limited scenario, for one time sample and for a single source.

[10]To obtain this plot, we have used the adaptive grid refinement approach from Section VI to get point estimates not limited to a coarse grid.

intuitively serves as an effective bound for performance in the threshold region, where the CRB is rather loose. Each point in the plot is the average of 50 trials. It can be seen that for well-separated sources, the variance of ℓ_1-SVD estimates follows closely that of other estimators, and, except for very low SNR, meets the CRB. As we have illustrated in Figure 4, closely spaced sources can be resolved at lower SNR with our technique than it is possible with other methods. This occurs in a region where our method is biased. On the other hand, Figure 13 shows that when the sources are well-separated, and our method is unbiased, its performance is as good as those of existing superresolution methods. Another important advantage can be seen in Figure 14, for correlated sources, which commonly occur in practice due to multipath effects. The correlation coefficient is 0.99. Our approach follows the CRB more closely than the other methods, and the threshold region occurs at lower SNR. The proposed ℓ_1-SVD method is the closest one in performance to the intuitive bound provided by the oracle-ML curve. This shows the robustness of our method to correlated sources.

D. Wideband source localization

The main difficulty which arises when wideband signals are considered is the impossibility to represent the delays by simple phase shifts. A way to deal with this issue is to separate the signal spectrum into several narrowband regions, each of which yields to narrowband processing. To work in the frequency domain, the time-samples are grouped into several "snapshots", and transformed into the frequency domain:

$$\mathbf{y}^{(n)}(\omega) = \mathbf{A}(\omega)\ \mathbf{s}^{(n)}(\omega) + \mathbf{n}^{(n)}(\omega),\ \ n \in \{1, .., N_s\} \tag{16}$$

For each frequency ω we have N_s snapshots. We are in general interested in a two dimensional power spectrum as a function of both spatial location (DOA) and frequency ω, so we solve the problem at each frequency independently, using the ℓ_1-SVD method, with frequency snapshots replacing temporal snapshots.

In Figure 15 we present an example using the same 8-element uniform linear array as the one used throughout the paper, but the signals are now wideband. We consider three chirps with DOAs 70°, 98°, and 120° with frequency span from 250 Hz to 500 Hz, and T = 500 time samples. Using conventional beamforming, the spatio-frequency spectra of the chirps are merged and cannot be easily separated (plot (a)), especially in lower frequency ranges, whereas using ℓ_1-SVD (plot (b)) they can be easily distinguished throughout their support. This shows that the ℓ_1-SVD methodology is useful for wideband scenarios as well.

The approach that we just described treats each frequency independently. In [20] we outline an alternative version of wideband source localization for joint "coherent" processing of the data at all frequencies. Wideband adaptations of current source localization methods, based on ideas such as focusing matrices [37], can do coherent processing over a narrow frequency region, but have difficulty with wider frequency regions, whereas our approach does not have such limitations. Furthermore, an important benefit that comes with our coherent wideband source localization approach is the ability to incorporate prior information on the frequency spectra of the sources. For example, in Figure 15, where we performed incoherent processing, the spectra of the chirps have a jagged shape, due to the fact that we treat each frequency independently. To mitigate this artifact, in the coherent version of wideband processing one could incorporate a prior on the continuity of the frequency spectra of the chirps. Another scenario where prior information on frequency could be particularly useful is for sources composed of multiple harmonics. In that case, a sparsity prior can be imposed on the frequency spectrum as well as on the spatial one. In Figure 16 we look at three wideband signals consisting of one or two harmonics each. At DOA 76° there are two harmonics with frequencies 200 and 520 Hz, at DOA 112° there are again two harmonics with frequencies 200 and 400 Hz, and at DOA 84° there is a single harmonic with frequency 520 Hz. Plot (a) shows results using conventional beamforming applied at each frequency (incoherently), plot (b) uses the MUSIC method applied at each frequency (incoherently), and plot (c) uses the coherent wideband version of ℓ_1-SVD. The results are displayed as intensity maps on a 2-D grid as a function of angle and frequency. Conventional beamforming merges the two well-separated peaks at 200 Hz, and also the two closely spaced peaks at 520 Hz. MUSIC resolves the two peaks at frequency 200 Hz, but merges the two at 520 Hz, and also shows some distortion due to noise. The coherent wideband version of ℓ_1-SVD resolves all the 5 peaks, and does not have any notable distortion due to noise.

IX. Conclusion

In this paper we explored a formulation of the sensor array source localization problem in a sparse signal representation framework. We started with a scheme for source localization with a single snapshot, and developed a tractable subspace-based ℓ_1-SVD method for multiple snapshots. The scheme can be applied to narrowband and to wideband scenarios. An efficient optimization procedure using second order cone programming was proposed. We described how to mitigate the effects of the limitation of the estimates to a grid through an adaptive grid-refinement

procedure, and proposed an automatic method for choosing the regularization parameter using the constrained form of the discrepancy principle. Finally, we examined various aspects of our approach, such as bias, variance, and the number of resolvable sources, using simulations. Several advantages over existing source localization methods were identified, including increased resolution, no need for accurate initialization, and improved robustness to noise, to limited number of time samples, and to correlation of the sources.

Some of the interesting questions for further research include an investigation of the applicability of greedy sparse signal representation methods, which have a lower computational cost, to source localization; a theoretical study of the bias and variance of our scheme; a detailed theoretical study of uniqueness and stability of sparse signal representation for the overcomplete bases that arise in source localization applications; a theoretical analysis of the multiple time-sample sparse signal representation problem; and applications of enforcing sparsity to spatially distributed or slowly time-varying sources.

APPENDIX

Formulating ℓ_1-SVD as a SOC optimization problem

The general form of a second order cone problem is:

$$\min \ \mathbf{c}'\mathbf{x}$$

$$\text{such that} \ \ \mathbf{A}\mathbf{x} = \mathbf{b}, \ \text{and} \ \mathbf{x} \in \mathbf{K}$$

where $\mathbf{K} = \mathbb{R}^N_+ \times \mathbf{L}_1 ... \times \mathbf{L}_{N_L}$. Here, $\mathbb{R}^N_+$ is the N-dimensional positive orthant cone, and $\mathbf{L}_1, ..., \mathbf{L}_{N_L}$ are second order cones.

First, to make our objective function in (13) linear, we use the auxiliary variables p and q, and put the nonlinearity into the constraints, by rewriting (13) as

$$\min p + \lambda q \tag{17}$$

$$\text{subject to } \|\mathbf{y}_{SV} - \mathbf{A}\mathbf{S}_{SV}\|_f^2 \leq p, \ \text{and} \ \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \leq q$$

The vector $\tilde{\mathbf{s}}^{(\ell_2)}$ is composed of non-negative real values, hence $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 = \sum_{i=1}^N \tilde{s}_i^{(\ell_2)} = \mathbf{1}' \ \tilde{\mathbf{s}}^{(\ell_2)}$. The symbol $\mathbf{1}$ stands for an $N \times 1$ vector of ones. The constraint $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \leq q$ can be rewritten as $\sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2} \leq r_i$, for

$i = 1, .., N$, and $\mathbf{1}'\mathbf{r} \leq q$, where $\mathbf{r} = [r_1, ..., r_N]'$. Also, let $\mathbf{z}_k = \mathbf{y}^{SV}(k) - \mathbf{A}\mathbf{s}^{SV}(k)$. Then, we have:

$$\min \ p + \lambda q \tag{18}$$

$$\text{subject to } \|(\mathbf{z}'_1, ..., \mathbf{z}'_K)\|_2^2 \leq p, \text{ and } , \mathbf{1}'\mathbf{r} \leq q,$$

$$\text{where } \sqrt{\sum_{k=1}^{K} (s_i^{SV}(k))^2} \leq r_i, \text{ for } i = 1, .., N$$

The optimization problem in (18) is in the second order cone programming form: we have a linear objective function, and a set of quadratic, linear, and SOC constraints. Quadratic constraints can be readily represented in terms of SOC constraints, see [31] for details.

REFERENCES

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

[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing - Concepts and Techniques*, Prentice Hall, 1993.

[3] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, Ph.D. thesis, Stanford Univ., 1981.

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

[5] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–310, Feb. 1997.

[6] S. Sardy, P. Tseng, and A. Bruce, "Robust wavelet denoising," *IEEE Trans. Signal Processing*, vol. 49, no. 6, pp. 1146–1152, June 2001.

[7] P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Feature selection via mathematical programming," *INFORMS Journal on Computing*, vol. 10, pp. 209–217, 1998.

[8] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[9] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of Royal Statistical Society, Series B*, vol. 58, pp. 267–288, Nov. 1996.

[10] M. D. Sacchi, T. J. Ulrych, , and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 31–38, Jan. 1998.

[11] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, vol. 3, pp. 1885–1888.

[12] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[13] J. J. Fuchs, "Linear programming in spectral estimation. Application to array processing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1996, vol. 6, pp. 3161–3164.

[14] J. J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Processing*, vol. 49, no. 4, pp. 702–709, Apr. 2001.

[15] B.D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[17] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[18] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004.

[19] A. J. Miller, *Subset Selection in Regression*, Chapman and Hall, 2002.

[20] D. M. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," M.S. thesis, MIT, EECS, July 2003, Available at http://ssg.mit.edu/~dmm/publications/malioutov_MS_thesis.pdf.

[21] D. L. Donoho and M. Elad, "Maximal sparsity representation via l1 minimization," *Proc. National Academy of Science*, vol. 100, pp. 2197–2202, Mar. 2003.

[22] R. Gribonval and M. Nielsen, "Sparse representation in unions of bases," *IEEE Trans. Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[23] J. J. Fuchs, "More on sparse representations in arbitrary bases," in *13th Symposium on System Identification (SYSID)*, Aug. 2003, pp. 1357–1362, IFAC/IFORS, submitted to IEEE Trans. on Inf. Theory.

[24] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *submitted for publication*, 2003.

[25] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *ICES Report 03-04, Univ. of Texas at Austin*, Feb. 2003, submitted for publication.

[26] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presense of noise," *submitted to IEEE Trans. Information Theory*, Feb. 2004.

[27] J. A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *submitted to IEEE Trans. Information Theory*, Feb. 2004, ICES Report 04-04, Univ. of Texas at Austin, http://www.ices.utexas.edu/reports/2004.html.

[28] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, pp. 1309–1331, 1992.

[29] J. J. Fuchs, "Detection and estimation of superimposed signals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, vol. 3, pp. 1649–1652.

[30] B.D. Rao and K. Kreutz-Delgado, "Basis selection in the presense of noise," in *Thirty-Second Asilomar Conf. on Signals, Systems and Computers*, Nov. 1998, vol. 1, pp. 752–756.

[31] A. Nemirovski A. Ben Tal, *Lectures on Modern Convex Optimization. Analysis, Algorithms and Engineering Applications*, SIAM, 2001.

[32] J. S. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," Tech. Rep., Tilburg University, Department of Econometrics, Netherlands, 2001, http://fewcal.kub.nl/~sturm.

[33] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications, Special Issue on Linear Algebra in Control, Signals and Image Processing*, , no. 284, pp. 193–228, 1998.

[34] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, vol. 7, pp. 414–417, 1966.

[35] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34, pp. 561–580, Dec. 1992.

[36] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 760–770, Mar. 2003.

[37] S. Sivanand, J. F. Yang, and M. Kaveh, "Time-domain coherent signal-subspace wideband direction-of-arrival estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989, vol. 4, pp. 2772–2775.

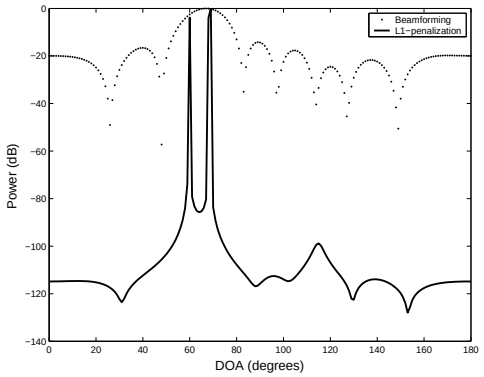

Fig. 1. Single sample source localization with ℓ_1: spatial spectra of two sources with DOAs of 60° and 70°, (SNR = 20 dB).

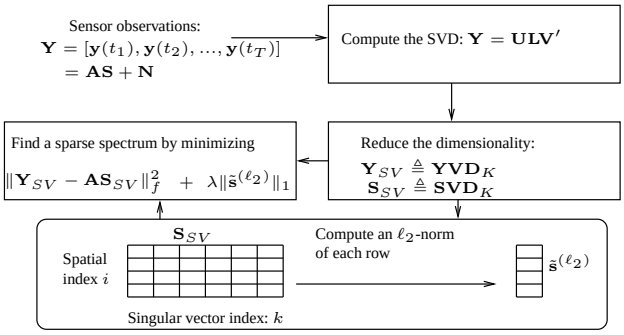

Fig. 2. Block diagram of steps for ℓ_1-SVD.

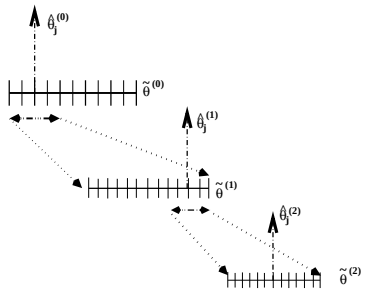

Fig. 3. Illustration of grid refinement.

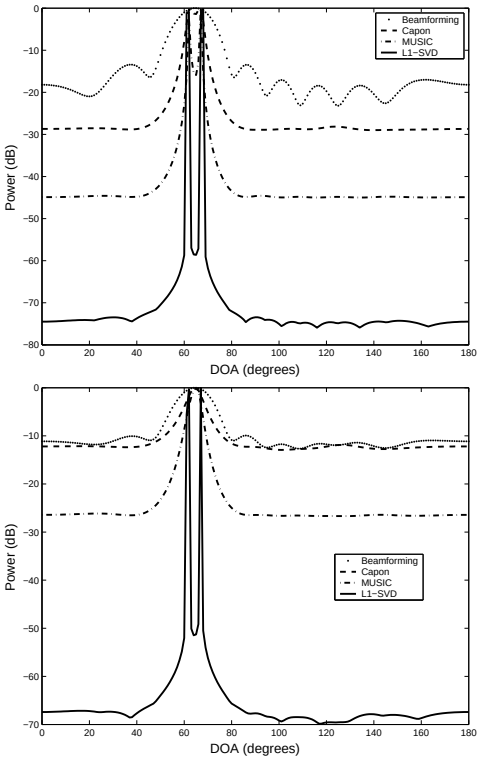

Fig. 4. Spatial spectra for beamforming, Capon's method, MUSIC, and the proposed method (ℓ_1-SVD) for uncorrelated sources, DOAs: 62° and 67°. Top: SNR = 10 dB. Bottom: SNR = 0 dB.

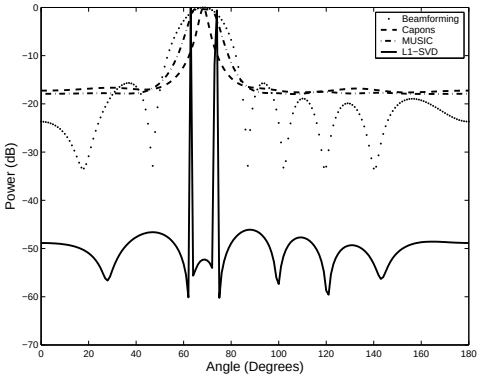

Fig. 5. Spectra for correlated sources, SNR = 20 dB, DOAs: 63° and 73°.

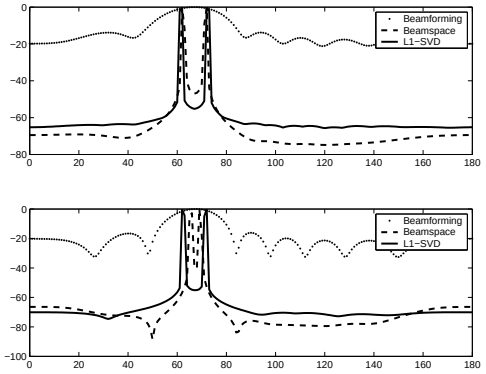

Fig. 6. Comparison with beamspace technique of [14]. SNR = 20 dB, DOAs: 63° and 73°. Top: uncorrelated sources. Bottom: correlated sources, correlation coefficient is 0.99.

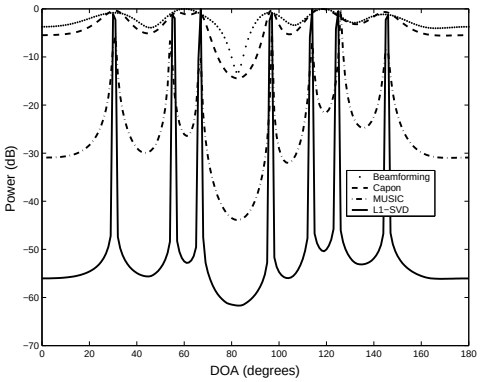

Fig. 7. Resolving $M-1$ sources: $M=8$ sensors, 7 sources, SNR = 10 dB.

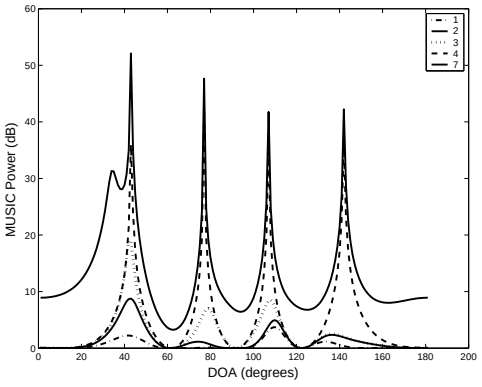

Fig. 8. Sensitivity of MUSIC to the assumed number of sources. The correct number is 4.

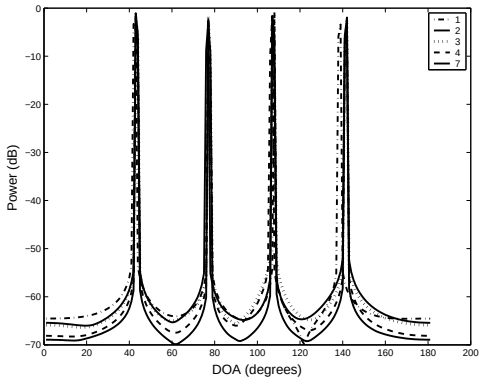

Fig. 9. Sensitivity of ℓ_1-SVD to the assumed number of sources. The correct number is 4.

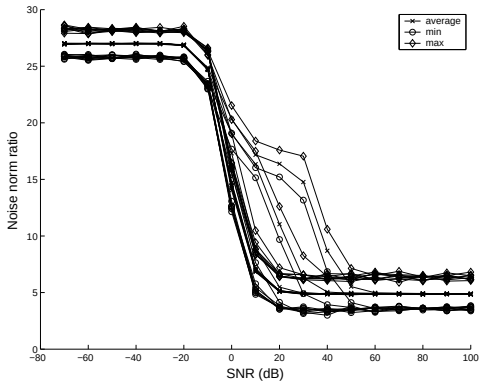

Fig. 10. Regularization parameter choice for moderate noise: ratio of $\|\tilde{\mathbf{n}}\|_2$ to σ as a function of SNR.

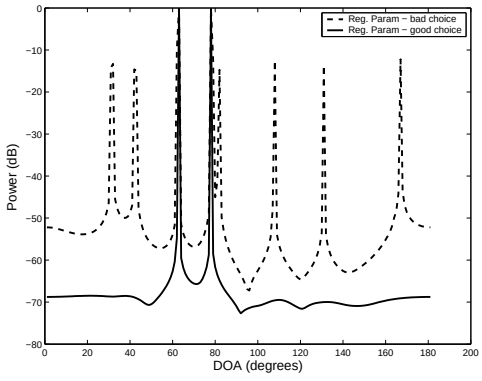

Fig. 11. Regularization parameter choice: the discrepancy principle leads to a useful spectrum. Setting the regularization parameter too low produces spurious peaks in the spectrum.

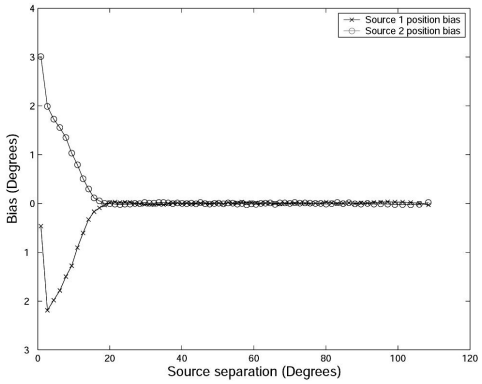

Fig. 12. The bias of ℓ_1-SVD in localizing two sources, as a function of separation between the two sources, SNR =10 dB.

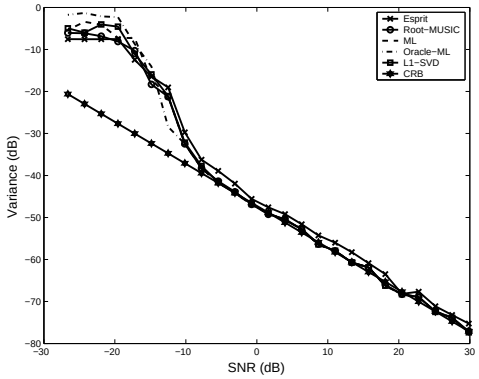

Fig. 13. CRB for zero mean *uncorrelated* sources, comparison with variances of ESPRIT, Root-MUSIC, ML, and ℓ_1-SVD, DOAs 42.83° and 73.33°

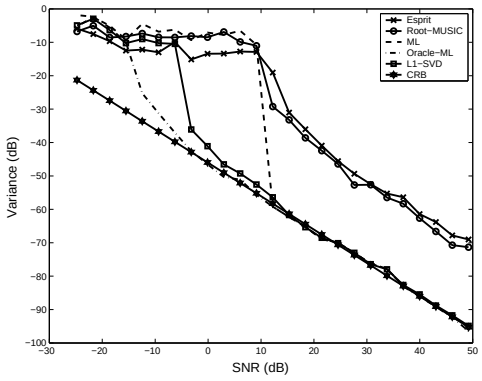

Fig. 14. Plots of variances of DOA estimates versus SNR, as well as the CRB, for two correlated sources. DOAs: 42.83° and 73.33°, variance for the source at 42.83° shown.

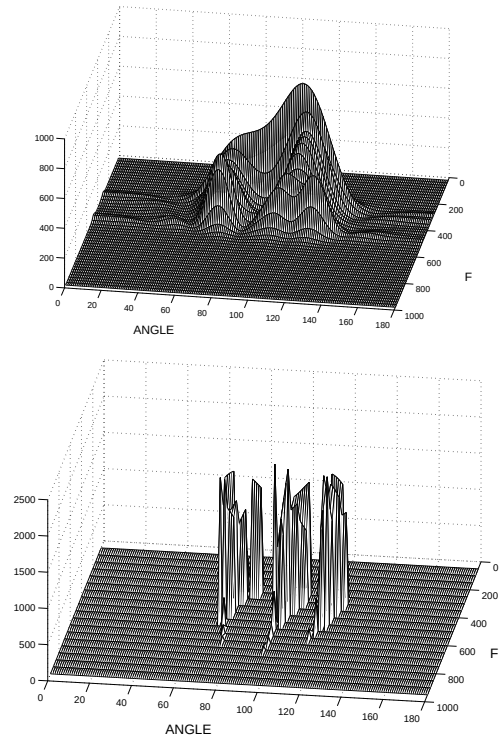

Fig. 15. Wideband example: 3 chirps, DOAs 70°, 98°, and 120°. Frequencies are processed independently. Top: conventional beamforming. Bottom: ℓ_1-SVD processing.

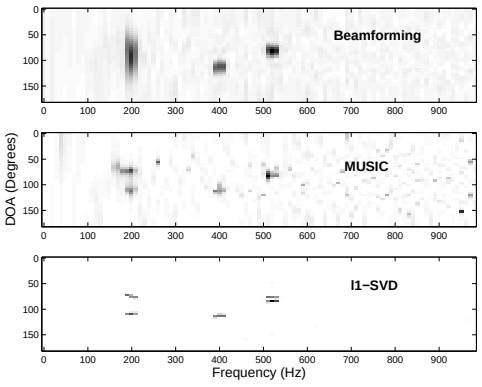

Fig. 16. Joint coherent processing of multiple harmonics with sparsity penalties on the spectra in the spatial and in frequency domain. Top: ℓ_1-SVD. Middle: incoherent beamforming. Bottom: incoherent MUSIC.