Approved for public release; distribution is unlimited.

Data Fusion in Large Arrays of Microsensors[1]

Alan S. Willsky, Dmitry M. Malioutov, and Müjdat Çetin
Room 35-433, M.I.T.
Cambridge, MA 02139
willsky@mit.edu

I. Introduction

In this paper we report on one component of the research being performed under a multi-university research program whose aim is to contribute to the fundamental questions in signal processing, fusion, and information science that arise when one considers the fusion of information from very large arrays of microsensors. We begin, however, with a brief description of our overall program, which has at its core dealing with four challenges when one considers large arrays of individually inexpensive sensors.[2] The first of these is ***scalability.*** How do we construct sensor fusion algorithms whose complexity scales well with network size? How does the performance of the network scale with its size? The second is ***dealing with uncertain and complex environments.*** One part of the vision for sensor networks is that not only will the individual elements be inexpensive but also their deployment will be inexpensive and flexible. For example, deploying large numbers of acoustic, seismic, electromagnetic, IR, and other types of sensors for military sensing in unknown environments requires that the network deal with uncertainties in the locations and calibration of the individual sensors as well as in the environment through which signals sensed by those sensors must propagate. Developing algorithms that are robust and/or adapt to these uncertainties is critical for successful operation of such networks. The third challenge is ***working with limited and dear resources.*** In particular, in many cases power (for sensing, computation, and communication) may be severely limited, and developing methodologies that make judicious use of power is critical. The last challenge is that of ***performance limits and guarantees.*** For example, it is important to have bounds on how well one can perform under resource constraints in order to determine if algorithms we have developed can be meaningfully improved or if they are already approaching the most one would expect given the available resources. In addition, given the uncertainties and variability of the environments into which sensor networks will be placed, it is important to characterize the robustness of algorithms and their sensitivities to variations in that environment.

Our investigation of these challenges are embedded in the three intellectual themes of our research. The first of these, namely, **network-constrained fusion,** deals with developing algorithms and performance analysis methods for fusion given (a) constraints or costs on communication of information among the sensors distributed in the network; and (b) a statistical model that describes how information from different sensors are related and hence how fusion should in principle be performed. Our second theme, **fusion in uncertain**

[1] The research described in this paper has been supported by an ODDR&E MURI, funded through the Army Research Office under Grant DAAD19-00-1-0466. A portion of the work presented in this paper has been published in [1].

[2] The reader is referred to our website: http://sensorweb.mit.edu for a complete listing of publications covering the full range of research being performed under this program.

environments, deals with relaxing (or, alternatively, providing) the modeling assumptions implicit in (b)**.** The third theme, **information theory for wireless networks,** deals with specifying the constraints and costs associated with (a). That is, this theme deals with information-theoretic constraints and scaling laws in wireless and resource-limited networks.

The specific component of our research that we describe in the following sections addresses one of the central problems arising in distributed array processing, namely that of source localization. The approach that we describe here has several motivations, including dealing with (i) the presence of multiple sources (with overlapping and possibly broad or multiband spectra); (ii) the likely presence of multipath effects (which can be interpreted as having highly correlated "sources" coming from multiple paths); (iii) the need for enhanced resolution (as compared to that achievable by other methods); and (iv) the need for robustness to noise, limitations in data quantity, and possible uncertainties in sensor locations. The method that we describe here involves the idea of imaging an entire "source field," where we introduce a so-called "sparsity prior" in order to guide our solution to focused localization of sources.[3] The approach that we take here is based on an optimization formulation including such a prior as part of the optimization criterion. In addition, a key component of our method involves the use of a singular value decomposition (SVD) of the data matrix—a natural and, as we will see very useful—method for accommodating and summarizing information contained in multiple data samples from an array.[4]

We now present a more detailed overview of our approach. 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 is the assumption of a small-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. In this context, each basis vector corresponds to an array manifold vector for a possible source location among a sampling grid of locations. The representation of the observed sensor data in terms of an overcomplete basis is not unique, and additional constraints have to be imposed to regain uniqueness. Our main goal is sparsity, so using constraints to minimize directly the number of non-zero coefficients (hence the number of sources) would be ideal, yet computationally prohibitive. In order to get around this challenge, we relax the problem using an idea similar to that of basis pursuit [2], and form an optimization problem containing an ℓ_1-norm-based penalty for the spatial spectrum. When we view this optimization problem as a maximum *a posteriori* (MAP) estimation problem, the ℓ_1 penalty corresponds to a Laplacian prior distribution assumption for the spectrum. These ideas are explored in more detail in Section 2.

In Section 3, we describe the narrowband source localization problem and turn it into a form appropriate for the overcomplete basis methodology (see [3] for our broadband approach). We then perform an SVD of the data matrix, which provides a useful way of handling multiple snapshots. In the SVD domain, we form our optimization functional for source localization, consisting of a data fidelity term, as well as the ℓ_1-norm-based sparsity constraint. In Section 4, we outline a numerical solution of this optimization problem in a

---

[3] While we focus here on point sources, a straightforward variant of this method can deal with the localization of so-called "distributed sources," i.e., targets whose physical extent exceeds several resolution cells so that the source of the received energy is distributed among these cells.

[4] While we focus primarily on multiple time samples here, one can also consider multiple frequency bin samples.

second-order cone (SOC) programming framework [4] by an interior point implementation [5]. Section 5 describes an adaptive grid refinement procedure for alleviating the effects of the grid, as well as the outline of a technique for the automatic selection of the regularization parameter involved in our method. Our experimental analysis in Section 6 shows that the proposed method provides better resolvability of closely-spaced sources, as well as improved robustness to low SNR, and the presence of correlated sources, as compared to currently available methods. Furthermore, our approach appears to have robustness to limitations in the number of time samples. Finally, unlike maximum likelihood (ML) methods, our technique does not require an accurate initialization, since the cost function is convex, and the optimization procedure is globally convergent [3].

The basic idea of using a sparse signal representation perspective for source localization was contained in our earlier work [6,7], and in [8]. The two main points of emphasis in this paper are the SVD-domain formulation, and the adaptation and use of SOC optimization.

II. Sparsity and Overcompleteness

We now describe the basic idea of enforcing sparsity in overcomplete basis representations, which will be used in Section 3 for the source localization problem. Given a signal $\mathbf{y} \in \mathbb{C}^M$, and an overcomplete basis $\mathbf{A} \in \mathbb{C}^{M\times N}$, $N > M$, we would like to find $\mathbf{s} \in \mathbb{C}^N$ such that $\mathbf{y} = \mathbf{A}\mathbf{s}$, and $\mathbf{s}$ is sparse. Define $\|\mathbf{s}\|_0^0$ to be the number of non-zero elements of $\mathbf{s}$. We would like to find $\min\|\mathbf{s}\|_0^0$ subject to $\mathbf{y} = \mathbf{A}\mathbf{s}$. This is a very hard combinatorial problem. It can be shown [9,3][5] that under certain conditions on $\mathbf{A}$ and $\mathbf{s}$, the optimal value of this problem can be found exactly by solving a related problem: $\min\|\mathbf{s}\|_1$ subject to $\mathbf{y} = \mathbf{A}\mathbf{s}$.

A natural extension when we allow white Gaussian noise is

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{1}$$

which can be solved by $\min\left(\|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1\right)$. The parameter λ controls the trade-off between the sparsity of the solution and the residual. This method is called basis pursuit [2], (or LASSO in the statistics literature).

III. Source Localization Framework

The narrowband source localization problem is:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t) \qquad t = 1,\ldots,T. \tag{2}$$

The data, $\mathbf{y}(t) \in \mathbb{C}^M$, are the observations from M sensors, and $\mathbf{x}(t) \in \mathbb{C}^K$ is a vector of unknown signals transmitted from K unknown locations θ_k. $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1),\ldots, \mathbf{a}(\theta_K)]$ is

[5] The result in [9] assumes $\mathbf{A}$ is composed of two orthogonal bases. In [3], we extend this result to any overcomplete basis, and also consider ℓ_p-norms, p< 1.

composed of the steering vectors $\mathbf{a}(\theta_k)$. The manifold $\mathbf{a}(\theta)$ is known as a function of θ. The goal is to estimate $\boldsymbol{\theta} = [\theta_1,\ldots, \theta_K]$.

Note that this problem is different from (1). First, the matrix $\mathbf{A}(\boldsymbol{\theta})$ is unknown, and second we have multiple time samples, $t = 1,..,T$. To address the first point, we introduce a grid of possible locations, $\{\beta_1,\ldots,\beta_N\}$, and form $\mathbf{A}= [\mathbf{a}(\beta_1),\ldots, \mathbf{a}(\beta_N)]$. Also, let

$$s_i(t) = \begin{cases} x_k(t), & \text{if} \quad \beta_i = \theta_k \\ 0, & \text{otherwise} \end{cases}.$$

Then the problem takes the form

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)\,. \tag{3}$$

The important point is that $\mathbf{A}$ is known, and does not depend on the unknown source locations θ_k, as $\mathbf{A}(\boldsymbol{\theta})$ did. The source locations are now encoded by the non-zero indices of $\mathbf{s}(t)$. In effect, we have transformed the problem from finding a point estimate of $\boldsymbol{\theta}$, to estimating the spatial spectrum of $\mathbf{s}(t)$, which has to exhibit sharp peaks at the correct source locations.

The second issue we raised was that of dealing with multiple time samples. In principle, one can use the overcomplete basis methodology to solve a signal representation problem at each time instant t. This leads to a significant computational load, and to sensitivity to noise, since no advantage is taken of other time samples. Instead, we would like to use all the sensor data in synergy. Previously, we presented two approaches to deal with this issue [6,7], which required certain assumptions on the source signals. We now present an SVD-based approach, which does not impose any restrictions on $\mathbf{x}(t)$. To this end, we view the data $\mathbf{y}(t)$ as a cloud of T points lying in a K-dimensional subspace. Instead of keeping every time sample, we can represent the cloud using its K largest singular vectors (corresponding to the signal subspace).

Let $\mathbf{Y}= [\mathbf{y}(1),\ldots, \mathbf{y}(T)]$, and define $\mathbf{S}$ and $\mathbf{N}$ similarly. Then we have $\mathbf{Y} = \mathbf{AS} + \mathbf{N}$. Take the singular value decomposition: $\mathbf{Y} = \mathbf{U\Lambda V}'$.[6] Let $\mathbf{Y}_{\text{SV}} = \mathbf{U\Lambda D}_K = \mathbf{YVD}_K$, 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 $K\times(T\text{-}K)$ matrix of zeros. Also, let $\mathbf{S}_{\text{SV}} = \mathbf{SVD}_K$, and $\mathbf{N}_{\text{SV}} = \mathbf{NVD}_K$, to obtain $\mathbf{Y}_{\text{SV}} = \mathbf{AS}_{\text{SV}} + \mathbf{N}_{\text{SV}}$. Now let us consider each column (corresponding to each singular vector) of this equation separately: $\mathbf{y}^{\text{SV}}(k) = \mathbf{As}^{\text{SV}}(k) + \mathbf{n}^{\text{SV}}(k)$, $k = 1,..,K$. If $K > 1$, then we have several subproblems and we can combine them into a single one by stacking. Let $\mathbf{y}_{\text{b}} = \text{vec}\ (\mathbf{Y}_{\text{SV}})$ (i.e. stack all the columns into a column vector $\mathbf{y}_{\text{b}}$). Define $\mathbf{s}_{\text{b}}$, and $\mathbf{n}_{\text{b}}$ similarly. Also, let

$$\mathbf{A}_{\text{b}} = \begin{pmatrix} \mathbf{A} & & \\ & \ddots & \\ & & \mathbf{A} \end{pmatrix}$$

i.e. $\mathbf{A}_{\text{b}}$ is block diagonal with K replicas of $\mathbf{A}$. Finally, we get $\mathbf{y}_{\text{b}} = \mathbf{A}_{\text{b}}\mathbf{s}_{\text{b}} + \mathbf{n}_{\text{b}}$ which is in the form of (1).

---

[6] This is closely related to the eigen-decomposition of the correlation matrix of the data: $\mathbf{R} = (1/T)\ \mathbf{YY}'$, whose eigen-decomposition is $\mathbf{R} = (1/T)\ \mathbf{U\Lambda V}'\ \mathbf{V}\ \mathbf{\Lambda}'\mathbf{U}' = (1/T)\ \mathbf{U\Lambda}^2\mathbf{U}'$.

The vector $\mathbf{s}_b$ has been constructed by stacking $\mathbf{s}^{SV}(k)$ for all the signal subspace singular vectors, $k = 1,..,K$. Every spatial index i appears for each of the singular vectors. We want to impose sparsity in $\mathbf{s}_b$ only spatially (in terms of i), and not in terms of the singular vector index k. So, we combine the data with respect to the singular vector index using an ℓ_2-norm, which does not favor sparsity: $\left(\mathbf{s}_b^{(\ell_2)}\right)_i = \sqrt{\sum_{k=1}^{K} (\mathbf{s}_i^{(SV)}(k))^2}, \quad \forall i$. The sparsity of the resulting $N\times1$ vector $\mathbf{s}_b^{(\ell_2)}$ corresponds to the sparsity of the spatial spectrum. We can find the spatial spectrum of $\mathbf{s}_b$ by minimizing

$$\left\|\mathbf{y_b} - \mathbf{A_b s_b}\right\|_2^2 + \lambda\left\|\mathbf{s}_\mathbf{b}^{(\ell_2)}\right\|_1 \tag{4}$$

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.

IV. Numerical Solution of the Optimization Problem

We now present an efficient algorithm for the minimization of the objective function in (4). The objective function contains a term $\left\|\mathbf{s}_b^{(\ell_2)}\right\|_1 = \sum_{i=1}^{N}\sqrt{\sum_{k=1}^{K} (\mathbf{s}_i^{(SV)}(k))^2}$, which is neither linear nor quadratic. We turn to second order cone (SOC) programming, which deals with the so-called second order cone constraints of the form $\mathbf{s}$: $\|s_1,\ldots,s_{n-1}\|_2 \le s_n$, i.e. $\sqrt{\sum_{i=1}^{n-1}(s_i)^2} \le s_n$. 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 is the availability of efficient interior point algorithms for the numerical solution of the former, e.g. [5].

The generic form of a second order cone problem is:

$$\begin{aligned} &\min \quad \mathbf{c'x} \\ &\text{such that} \quad \mathbf{Ax} = \mathbf{b}, \quad \text{and } \mathbf{x} \in \mathbf{K} \end{aligned} \tag{5}$$

where $\mathbf{K} = \mathbb{R}_+^N \times \mathbf{L}_1 \times ... \times \mathbf{L}_{N_L}$. Here, $\mathbb{R}_+^N$ is the N-dimensional positive orthant cone, and $\mathbf{L}_1,...,\mathbf{L}_{N_L}$ are second order cones. Now we would like to express our objective function (4), in the form of the optimization problem in (5). First, to make the objective function linear, we rewrite (4) as

$$\begin{aligned} &\min \quad p + \lambda q \\ &\text{subject to} \quad \left\|\mathbf{y_b} - \mathbf{A_b s_b}\right\|_2^2 \le p, \quad \text{and} \left\|\mathbf{s}_\mathbf{b}^{(\ell_2)}\right\|_1 \le q \end{aligned} \tag{6}$$

The vector $\mathbf{s}_b^{(\ell_2)}$ is composed of positive real values, hence $\left\|\mathbf{s}_b^{(\ell_2)}\right\|_1 = \sum_{i=1}^{N}\left(\mathbf{s}_b^{(\ell_2)}\right)_i = \mathbf{1}'\mathbf{s}_b^{(\ell_2)}$. The symbol **1** stands for an $N\times 1$ vector of ones. The constraint $\left\|\mathbf{s}_\mathbf{b}^{(\ell_2)}\right\|_1 \le q$ can be rewritten as $\sqrt{\sum_{k=1}^{K}(\mathbf{s}_i^{(\mathrm{SV})}(k))^2} \le r_i$, for i = 1,...,N, and $\mathbf{1}'\mathbf{r} \le q$, where we use $\mathbf{r} = [\mathrm{r}_1, ..., \mathrm{r}_N]'$. Also, let $\mathbf{z}_k = \mathbf{y}^{\mathrm{SV}}(k) - \mathbf{A}\mathbf{s}^{\mathrm{SV}}(k)$. Then, we have:

$$\begin{aligned} &\min \quad p + \lambda q \\ &\text{subject to } \ \left\|\mathrm{z}_1',...,\mathrm{z}_K'\right\|_2^2 \le p, \text{ and } \mathbf{1}'\mathbf{r} \le q \\ &\text{where} \sqrt{\sum_{k=1}^{K}(\mathbf{s}_i^{(\mathrm{SV})}(k))^2} \le r_i, \text{ for } i = 1,...,N. \end{aligned} \tag{7}$$

The optimization problem in (7) is in the second order cone programming form: we have a linear objective function, and a set of quadratic[7], linear, and SOC constraints.

V. Grid Refinement and Parameter Selection

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 accuracy. 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 $\beta_i^{(0)}$, for i=1,…,N. Set r=0. The grid should not be too rough, in order not to introduce substantial bias. A 1° or 2° uniform sampling usually suffices,
2. Form $\mathbf{A}^{(\mathrm{r})} = \mathbf{A}(\boldsymbol{\beta}^{(\mathrm{r})})$, where $\boldsymbol{\beta}^{(\mathrm{r})} = [\beta_1^{(\mathrm{r})},\dots, \beta_N^{(\mathrm{r})}]$. Use our method from Section 3 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 $\boldsymbol{\beta}^{(\mathrm{r})}$ 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, $\delta_{r+1} = \delta_r / \rho$. 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 $\rho = 3$, a small number. After a few (e.g. 5) iterations of refining the grid, it becomes fine enough that its effects are almost

[7] Quadratic constraints can be readily represented in terms of SOC constraints. See [4] for details.

transparent. This idea has been successfully used for some of the experimental analysis we present in Section 6.

Another important part of our source localization framework is the choice of the regularization parameter λ in (4), 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 still an open problem, especially if the objective function is non-quadratic. An old idea under the name of discrepancy principle [10] is to select λ to match the residuals of the solution obtained using λ, to some known statistics of the noise, when such are available. For example, if the variance of the independent identically distributed Gaussian noise is known, then one can select λ such that $\|\mathbf{y_b} - \mathbf{A_b s_b}\|_2^2 \approx \mathrm{E}(\|\mathbf{n}_\mathrm{b}\|_2^2)$. Searching for a value of λ to achieve the equality is rather difficult, and requires solving the problem (4) multiple times for different λ's.

Instead, we consider the constrained version of the problem in (4), which can also be efficiently solved in the second order cone framework [3]:

$$\min \left\|\mathbf{s}_b^{(\ell_2)}\right\|_1 \quad \text{subject to} \left\|\mathbf{y_b} - \mathbf{A_b s_b}\right\|_2^2 \le \gamma \tag{8}$$

For this problem the task of choosing the regularization parameter γ properly is considerably easier. We choose γ high enough so that the probability that $\|\mathbf{n}_\mathrm{b}\|_2^2 \ge \gamma$ is small. The details of the procedure can be found in [3].

VI. Experimental Results

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. In Figure 1, 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 merge the two peaks. In the bottom plot, we decrease the SNR to 0 dB, in which case only our technique is still able to resolve the two sources. In Figure 2, we set the SNR to 20 dB, but we make the sources strongly correlated. MUSIC and Capon's method would resolve the signals 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, e.g. due to multipath effects.

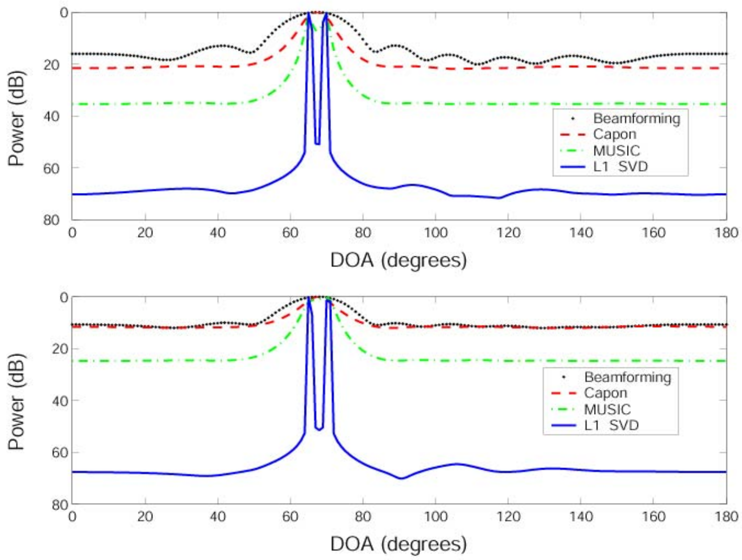

Fig. 1. Spatial spectra for beamforming, Capon's method, MUSIC, and the proposed method (L1-SVD) for uncorrelated sources, DOAs: 65° and 70°. Top: SNR = 10 dB. Bottom: SNR = 0 dB.

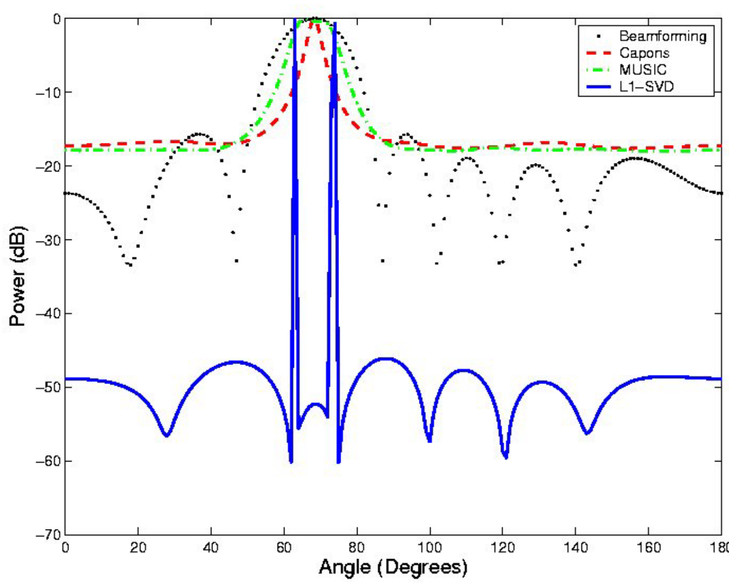

Fig. 2. Spectra for correlated sources, SNR = 20dB, DOAs: 63° and 73°.

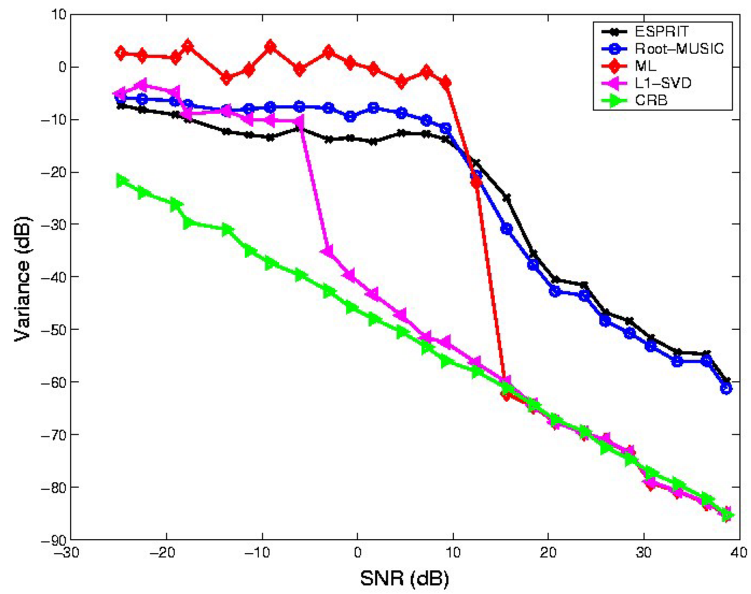

Fig. 3. Plots of variances of DOA estimates versus SNR, as well as the CRB, for two correlated sources. DOAs: 42.83° and 73.33°, variance for the source at 42.83° shown.

We next compare the performance of our approach in terms of the variance of the DOA estimates to other methods, as well as 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 (see [3] for an analysis of bias). In Figure 3, we present plots of variance versus SNR for a scenario including two strongly correlated sources[8]. The correlation coefficient is 0.99. Each point in the plot is the average of 50 trials. Our approach follows the CRB more closely than the other methods. This shows the robustness of our method to correlated sources.

VII. Conclusion

In this paper we have reported on one component of our research being performed under a multi-university research program on data fusion in large arrays of microsensors. In particular, we have presented an optimization-based source localization technique which enforces spatial sparsity of the sources while representing sensor measurements in an overcomplete basis of array manifold samples. The SVD is used to combine multiple data samples, and SOC programming is used to solve the resulting optimization problem. Explicitly enforcing sparsity leads to very sharp peaks of the spatial spectrum at the locations of the sources, which allow closely-spaced sources to be resolved. Additional benefits as compared to currently available techniques are better robustness to noise, limited time samples, and correlation of the sources, and the lack of need for accurate initialization. We have demonstrated the effectiveness of the proposed approach through a number of experiments.

---

[8] To obtain this plot, we have used the adaptive grid refinement approach from Section 5 to obtain point estimates not limited to the initial coarse grid.

Here we have considered a basic array processing scenario, involving narrowband sources in the far-field of the array, and known sensor locations. Various extensions of our framework can be found in [3]. The first such extension is the near-field case, which only requires changing the manifold matrix used in our approach so that it is parameterized by two spatial coordinates, range and DOA, rather than just DOA as in the far-field case. The second extension is to the case of multiband and broadband signals. We have developed two versions of our technique for this case, the first one involving independent narrowband processing at each frequency band, and the second one involving joint processing at all bands. The latter approach is especially powerful, since it allows prior information about the temporal spectrum of the signals to be incorporated as well. For example, if we know that we are looking for harmonics (e.g., due to sources that involve a fundamental band and several harmonics at unknown frequencies), we can impose a sparsity prior for the temporal spectrum. Note that a similar sparsity constraint could be imposed in the time domain for bursty sources (i.e., ones in which signal energy is concentrated in an unknown subset of time points). The final extension we would like to mention involves the problem of uncertainties in the sensor locations. We have generalized our optimization-based framework to perform joint source localization and self-calibration for moderate location uncertainties.

VIII. Acknowledgments

We would like to thank Dr. Brian Sadler for very useful discussions about the work presented in this paper.

References

[1] D.M. Malioutov, M. Cetin, and A.S. Willsky, "Source localization by enforcing sparsity through a Laplacian prior: An SVD-based approach," in *IEEE Workshop on Statistical Signal Processing*, Sept.-Oct. 2003.

[2] 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.

[3] D.M. Malioutov, "A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays," S.M. thesis, MIT, July 2003.

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

[5] J.S. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," Tech. Rep. Tilburg University, Department of Econometrics, Netherlands, 2001.

[6] M. Cetin, D.M. Malioutov, and A.S. Willsky, "A variational technique for source localization based on a sparse signal reconstruction perspective," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2002, vol.3, pp. 2965-2968.

[7] D.M. Malioutov, M. Cetin, J.W. Fisher, III, and A.S. Willsky, "Superresolution source localization through data-adaptive regularization," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, August 2002.

[8] 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.

[9] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2845-2862, 2001.

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