

SOURCE LOCALIZATION BY ENFORCING SPARSITY THROUGH A LAPLACIAN PRIOR: AN SVD-BASED APPROACH

Dmitry M. Malioutov, Müjdat Çetin, and Alan S. Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA

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 an ℓ_1 -norm penalty; this can also be viewed as an estimation problem with a Laplacian prior. Explicitly enforcing the sparsity of the representation is motivated by a desire to obtain a sharp estimate of the spatial spectrum which exhibits superresolution. To summarize multiple time samples we use the singular value decomposition (SVD) of the data matrix. 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 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 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.

1. INTRODUCTION

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 [1], 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. We then perform a singular value decomposition (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 second-order cone (SOC) programming framework [8] by an interior point implementation [2]. Section 5 describes an adaptive grid refinement procedure which alleviates the effects of the grid. In Section 6 we propose a technique for the automatic selection of the regularization parameter involved in our method. Our experimental analysis in Section 7 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 limited numbers of time samples, and unlike maximum likelihood (ML) methods, it does not require an accurate initialization, since the cost function is convex, and the optimization procedure is globally convergent [7].

The basic idea of using a sparse signal representation perspective for source localization was contained in our earlier work [3,4], and in [5]. The two main contributions of this paper are the SVD-domain formulation, and the adaptation and use of SOC optimization. In addition, we describe a multi-resolution technique for refining the spatial sampling grid, and a method for the automatic choice of the hyperparameter involved in our approach.

2. 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 [6, 7]¹ 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 $\mathbf{y} = \mathbf{A}\mathbf{s}$.

¹The result in [6] assumes \mathbf{A} is composed of two orthogonal bases. In [7], we extend this result to any overcomplete basis, and also consider ℓ_p -norms, $p < 1$.

This work was supported by the Army Research Office under Grant DAAD19-00-1-0466, and the Office of Naval Research under Grant N00014-00-1-0089.

A natural extension when we allow white Gaussian noise is

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (1)$$

which can be solved by $\min(\|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1)$. The parameter λ controls the trade-off between the sparsity of the solution and the residual. This method is called basis pursuit [1], (or LASSO in the statistics literature). This cost functional also results from MAP estimation with a Laplacian prior, $p(s_i) = \frac{\alpha}{2}e^{-\alpha|s_i|}$. When \mathbf{s} is complex, the phase is taken with uniform distribution in $[0, 2\pi]$. The Laplacian prior has heavy tails and favors a sparse representation.

3. SOURCE LOCALIZATION FRAMEWORK

The narrowband source localization problem is:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, \dots, T \quad (2)$$

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 . $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ is 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 $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]$.

Note that this problem is different from (1). First, the matrix $\mathbf{A}(\boldsymbol{\theta})$ is unknown, and second we have multiple time samples, $t=1, \dots, T$. To address the first point, we introduce a grid of possible locations, $\{\tilde{\theta}_1, \dots, \tilde{\theta}_N\}$, and form $\tilde{\mathbf{A}} = [\mathbf{a}(\tilde{\theta}_1), \dots, \mathbf{a}(\tilde{\theta}_N)]$.

Also, let $s_i(t) = \begin{cases} x_k(t), & \text{if } \tilde{\theta}_i = \theta_k \\ 0, & \text{otherwise} \end{cases}$. Then the problem takes the form

$$\mathbf{y}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(t) + \mathbf{n}(t) \quad (3)$$

The important point is that $\tilde{\mathbf{A}}$ 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 $\tilde{\mathbf{s}}(t)$. In effect, we have transformed the problem from finding a point estimate of $\boldsymbol{\theta}$, to estimating the spatial spectrum of $\tilde{\mathbf{s}}(t)$, 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 [3, 4], 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 $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(T)]$, and define \mathbf{S} and \mathbf{N} similarly. Then we have $\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{N}$. Take the singular value decomposition: $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$.² Let $\mathbf{Y}_{SV} = \mathbf{U}\mathbf{\Lambda}_K = \mathbf{Y}\mathbf{D}_K$, 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.³ Also, let $\mathbf{S}_{SV} = \mathbf{S}\mathbf{V}\mathbf{D}_K$, and $\mathbf{N}_{SV} = \mathbf{N}\mathbf{V}\mathbf{D}_K$, to obtain $\mathbf{Y}_{SV} = \mathbf{A}\mathbf{S}_{SV} + \mathbf{N}_{SV}$. Now let us consider

²This is closely related to the eigen-decomposition of the correlation matrix of the data: $\mathbf{R} = \frac{1}{T}\mathbf{Y}\mathbf{Y}'$. Its eigen-decomposition is $\mathbf{R} = \frac{1}{T}\mathbf{U}\mathbf{\Lambda}\mathbf{V}'\mathbf{V}\mathbf{\Lambda}'\mathbf{U}' = \frac{1}{T}\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}'$.

³If $T < K$, or if the sources are coherent, we use the number of signal subspace singular values instead of K in forming \mathbf{D}_K .

each column (corresponding to each singular vector) of this equation separately: $\mathbf{y}^{SV}(k) = \mathbf{A}\mathbf{s}^{SV}(k) + \mathbf{n}^{SV}(k)$, $k = 1, \dots, K$. If $K > 1$, then we have several subproblems and we can combine them into a single one by stacking. Let $\tilde{\mathbf{y}} = \text{vec}(\mathbf{Y}_{SV})$ (i.e. stack all the columns into a column vector $\tilde{\mathbf{y}}$). Define $\tilde{\mathbf{s}}$, and $\tilde{\mathbf{n}}$ similarly.

Also, let $\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & & \\ & \ddots & \\ & & \mathbf{A} \end{pmatrix}$, i.e. $\tilde{\mathbf{A}}$ is block diagonal with K replicas of \mathbf{A} . Finally, we get $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{s}} + \tilde{\mathbf{n}}$ which is in the form of (1).

The vector $\tilde{\mathbf{s}}$ has been constructed by stacking $\mathbf{s}^{SV}(k)$ for all the signal subspace singular vectors, $k = 1, \dots, K$. Every spatial index i appears for each of the singular vectors. We want to impose sparsity in $\tilde{\mathbf{s}}$ 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: $\tilde{s}_i^{(\ell_2)} = \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2}$, $\forall i$. The sparsity of the resulting $N \times 1$ 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

$$\|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}}\|_2^2 + \lambda\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \quad (4)$$

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.

4. SOLUTION BY SOC PROGRAMMING

Now that we have an objective function in (4) to minimize, we would like to do it in an efficient manner. The objective contains a term $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 = \sum_{i=1}^N \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2}$, 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 $\mathbf{s} : \|\mathbf{s}_1, \dots, \mathbf{s}_{n-1}\|_2 \leq s_n$, 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 is the availability of efficient interior point algorithms for the numerical solution of the former, e.g. [2].

The generic form of a second order cone problem is:

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

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

First, to make the objective function linear, we rewrite (4) as

$$\begin{aligned} &\min p + \lambda q \\ &\text{subject to } \|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}}\|_2^2 \leq p, \text{ and } \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \leq q \end{aligned} \quad (5)$$

The vector $\tilde{\mathbf{s}}^{(\ell_2)}$ is composed of positive 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 $i = 1, \dots, N$, and $\mathbf{1}'\mathbf{r} \leq q$, where we use $\mathbf{r} = [r_1, \dots, r_N]'$. Also, let $\mathbf{z}_k = \mathbf{y}^{SV}(k) - \mathbf{A}\mathbf{s}^{SV}(k)$. Then, we have:

$$\begin{aligned} & \min p + \lambda q & (6) \\ & \text{subject to } \|(\mathbf{z}'_1, \dots, \mathbf{z}'_K)\|_2^2 \leq p, \text{ and } \mathbf{1}'\mathbf{r} \leq q, \\ & \text{where } \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2} \leq r_i, \text{ for } i = 1, \dots, N \end{aligned}$$

The optimization problem in (6) is in the second order cone programming form: we have a linear objective function, and a set of quadratic⁴, linear, and SOC constraints.

5. ADAPTIVE 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 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 $\tilde{\theta}_i^{(0)}$, for $i = 1, \dots, N_\theta$. 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 $\mathbf{A}_r = \mathbf{A}(\tilde{\theta}^{(r)})$, where $\tilde{\theta}^{(r)} = [\tilde{\theta}_1^{(r)}, \tilde{\theta}_2^{(r)}, \dots, \tilde{\theta}_{N_\theta}^{(r)}]$. Use our method from Section 3 to get the estimates of the source locations, $\hat{\theta}_j^{(r)}$, $j = 1, \dots, K$, and set $r = r + 1$.
3. Get a refined grid $\tilde{\theta}^{(r)}$ 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 $j = 1, \dots, K$. 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 almost transparent. This idea has been successfully used for some of the experimental analysis we present in Section 7.

⁴Quadratic constraints can be readily represented in terms of SOC constraints. See [8] for details.

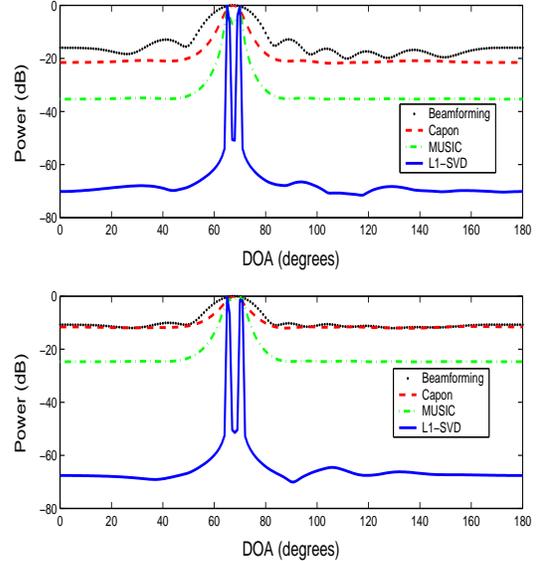


Fig. 1. Spatial spectra for beamforming, Capon's method, MUSIC, and the proposed method (L1-SVD) for uncorrelated sources, DOAs: 65° and 70° . Top: SNR = 10 dB. Bottom: SNR = 0 dB.

6. REGULARIZATION PARAMETER SELECTION

An 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 not quadratic. An old idea under the name of discrepancy principle [9] 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 i.i.d. Gaussian noise is known, then one can select λ such that $\|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}}\|_2^2 \approx E\|\tilde{\mathbf{n}}\|_2^2$. 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 propose to look at the constrained version of the problem in (4), which can also be efficiently solved in the second order cone framework [7]:

$$\min \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \text{ subject to } \|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}}\|_2^2 \leq \beta^2 \quad (7)$$

For this problem the task of choosing the regularization parameter β properly is considerably easier. We choose β high enough so that the probability that $\|\tilde{\mathbf{n}}\|_2^2 \geq \beta^2$ is small. We describe the details in [7].

7. 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

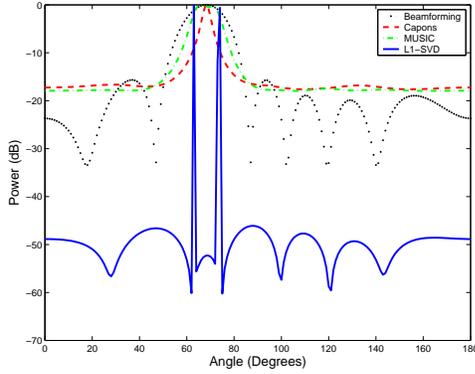


Fig. 2. Spectra for correlated sources, $SNR = 20dB$, DOAs: 63° and 73° .

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. In Figure 2, we set the SNR to 20 dB, but 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.

One aspect of our technique is that it is biased for closely-spaced sources when λ is selected appropriately for low SNR. However 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. In Figure 3, we plot the bias of each of the two source location estimates as a function of the separation between the two sources, 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 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. In Figure 4, we present plots of variance versus SNR for a scenario including two strongly correlated sources⁵. 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.

8. REFERENCES

- [1] 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.
- [2] J. S. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," Tech. Rep., Tilburg

⁵To obtain this plot, we have used the adaptive grid refinement approach from Section 5 to get point estimates not limited to the grid.

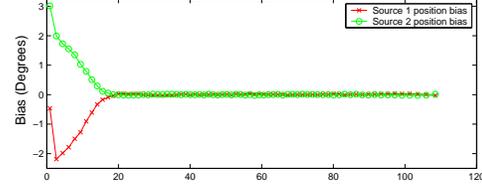


Fig. 3. Bias vs. separation, $SNR = 10$ dB.

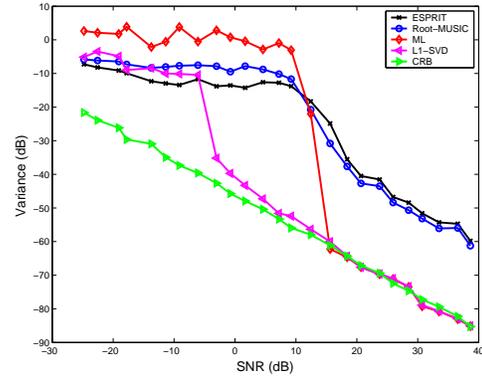


Fig. 4. Plots of variances of DOA estimates versus SNR, as well as the CRB, for two correlated sources. DOAs: 42.83° and 73.33° , variance for the source at 42.83° shown.

University, Department of Econometrics, Netherlands, 2001, <http://fewcal.kub.nl/sturm>.

- [3] 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.
- [4] 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*, Aug. 2002.
- [5] 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.
- [6] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [7] D. M. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," M.S. thesis, MIT, EECS, July 2003.
- [8] A. Nemirovski A. Ben Tal, *Lectures on Modern Convex Optimization. Analysis, Algorithms and Engineering Applications*, SIAM, 2001.
- [9] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, vol. 7, pp. 414–417, 1966.