CS versus MAP and MMOSPA for Multi-Target Radar AOAs

David F. Crouse, Peter Willett, Yaakov Bar-Shalom,
University of Connecticut
Department of Electrical and Electronic Engineering
371 Fairfield Way, U-2157,
Storrs, Connecticut 06269, USA
Email: {crouse, willett, ybs}@engr.uconn.edu

Lennart Svensson
Department of Signals and Systems
Chalmers University of Technology
Göteborg, Sweden
Email: lennart.svensson@chalmers.se

Abstract—We expand upon existing the literature regarding using Minimum Mean Optimal Sub-Pattern Assignment error (MMOSPA) estimates in multitarget tracking to apply it to angular superresolution of closely-space targets, noting its advantages in comparison to Maximum a Posteriori (MAP) and Minimum Mean Squared Error (MMSE) estimation. MMOSPA estimators sacrifice target labeling, but in doing so they can (often) avoid coalescence of estimates of closely-spaced objects. A compressive sensing solution, which is a form of MAP estimation, is also considered and is solved via a brute force search, which, contrary to popular belief, is computationally feasible when the number of targets is low, having execution times on the order of tens of milliseconds for two targets on a linear array.

I. OVERVIEW

Numerous superresolution algorithms exist for determining the directions of arrival (DOA) of returns from multiple targets using an antenna array. In this paper, we compare the use of Minimum Mean Optimal Sub-Pattern Assignment error (MMOSPA) estimation to the more traditional approaches of Maximum a Posteriori (MAP) estimation and compressive sensing (CS) on a linear of planar array, which might be composed of multiple subarrays. Note that these techniques can often be generalized to applications beyond DOA estimation, such as in detection [20], [27].

Though CS can be used to estimate the number of targets present, and MAP estimation can be augmented with a Minimum Description Length (MDL) criterion, among other approaches, to estimate the order (the number of targets), MMOSPA optimization, like Minimum Mean Squared Error (MMSE) optimization, is not easily adapted for order estimation. Thus, we shall focus our attention on the accuracy of the DOA estimates assuming that the correct number of targets is known.

Section II discusses the signal model, including the use of radar arrays that have been broken into subarrays. Section III defines MMOSPA estimation and applies it to angular superresolution. Sections IV and V describe the use of compressed sensing, also known as compressive sensing,

and Maximum Likelihood (ML) and MAP estimation for angular superresolution. Section VI compares the algorithms through a simulation involving a linear array and Section VII summarizes the results.

II. THE SIGNAL MODEL

We shall assume that filtering has been performed on a narrowband signal such that all target returns are mixed down to baseband, thus eliminating the carrier, and leaving only a slow phase variation (doppler) with angular frequency ω from sample to sample. All targets are located in the far-field, in which case we need only consider planar waves arriving at the receiver. Let vector $\mathbf{u}_m = [u_{m,1}, u_{m,2}, u_{m,3}]'$, with the apostrophe denoting the transpose, denote the unit vector pointing in the direction of origin of the mth wave, whereby there are a total of M sources. The ensemble of these pointing vectors shall be designated by $\mathbf{U}$. The phase center of the ith antenna shall be located at $\mathbf{r}_i = [r_{i,1}, r_{i,2}, r_{i,3}]'$, corresponding to the x, y and ζ directions shown in Figure 1. This means that the complex baseband signal on antenna i, at a particular time is

$$z_i = \sum_{m=1}^{M} b_m a_i(\mathbf{u}_m) + w_i \qquad a_i(\mathbf{u}_m) \triangleq e^{j2\pi f \frac{\mathbf{r}_i'\mathbf{u}_m}{c}} \qquad (1)$$

where w_i is an additive, zero-mean, complex Gaussian noise in the ith channel, b_m is the complex amplitude of the return from the mth source, f is the carrier frequency and c is the speed of light.

The steering vector for a general array in three dimensions is defined to be a collection of $a_i(\mathbf{u}_m)$ for all N_{el} elements of the antennas in the array:

$$\mathbf{a}_m = [a_1(\mathbf{u}_m), a_2(\mathbf{u}_m), \ldots, a_{N_{\text{el}}}(\mathbf{u}_m)]' \qquad (2)$$

We can thus write the output of all of the elements in the array in vector form as:

$$\mathbf{z} = [z_1, z_2, \ldots, z_{N_{\text{el}}}]' = \sum_{m=1}^{M} \mathbf{a}_m b_m + \mathbf{w} = \mathbf{A}\mathbf{b} + \mathbf{w} \qquad (3)$$

where

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_M] \qquad \mathbf{w} = [w_1, w_2, \ldots, w_{N_{\text{el}}}]' \qquad (4)$$
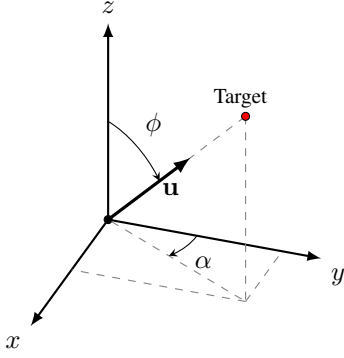
Fig. 1. The vectorial relationship of the wave direction with respect to the coordinate system of the antenna array. The array is in the x-y plane and the direction of the target is specified by a unit vector $\mathbf{u}$ coming out of the plane.

The likelihood of the directions of arrival, $\mathbf{U}$, and the complex amplitudes, $\mathbf{b}$, given the array outputs is thus

$$p(\mathbf{z}|\mathbf{U}, \mathbf{b}) = \pi^{-N_{\text{el}}}|\mathbf{Q}|^{-1}e^{-(\mathbf{z}-\mathbf{Ab})^H\mathbf{Q}^{-1}(\mathbf{z}-\mathbf{Ab})} \quad (5)$$

where H denotes the conjugate transpose (Hermitian) and $\mathbf{Q}$ is the correlation matrix of the noise, which might include information regarding the effects of Gaussian jammers, studied in [17]. The likelihood function for ML estimation of target parameters for monopulse estimation has been presented many times. Consider, for example, [16], [18], [28] and [21]. We used the model of [16] and [18].

Similarly, we can write the a posteriori PDF of $\mathbf{U}$ and $\mathbf{b}$ using Bayes' theorem as

$$p(\mathbf{U}, \mathbf{b}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{U}, \mathbf{b})p(\mathbf{U}, \mathbf{b})}{p(\mathbf{z})} \quad (6)$$

where $p(\mathbf{U}, \mathbf{b})$ is the predicted value of directions of arrival and the complex amplitudes. Assuming that the antenna elements are isotropic (i.e. the gain at a single antenna does not depend upon the angle of arrival) over the area being considered, $\mathbf{U}$ and $\mathbf{b}$ can usually be approximated to be independent. That is,

$$p(\mathbf{U}, \mathbf{b}) = p(\mathbf{U})p(\mathbf{b}) \quad (7)$$

An uninformative prior on the directions of arrival is

$$p(\mathbf{U}) = \prod_{m=1}^{M} p(\mathbf{u}_m) \quad (8)$$

$$p(\mathbf{u}_m) = \frac{3}{2\pi} \quad \text{for valid } \mathbf{u}_m \quad (9)$$

We consider valid $\mathbf{u}_m$ to be those directions such that $u_{m,1}^2 + u_{m,2}^2 + u_{m_3}^2 \leq 1$ and $u_{m,3} > 0$, which simply tells us that the targets are located in front of the array (a reasonable assumption, since an antenna mounted on a conductive plate has almost zero sensitivity in the back).[1] In practice, however,

[1] The $\frac{3}{2\pi}$ in (9) comes from evaluating the inverse of

$$\int_{-1}^{1} \int_{-\sqrt{1-u_{m,1}^2}}^{\sqrt{1-u_{m,1}^2}} \int_{0}^{\sqrt{1-u_{m,1}^2-u_{m,2}^2}} 1\,du_{m,3}\,du_{m,2}\,du_{m,1}$$

when using a high gain antenna, such as the one used in the simulation section of this paper[2], one can usually assume that the target returns are within the beam or within some factor (≥ 1) of the beamwidth (if not, the targets wouldn't be detected). This means that $p(\mathbf{U})$ would be a nonzero constant within the beam and zero elsewhere.

The prior on the complex amplitudes of the targets is generally assumed to be a zero-mean complex Gaussian (i.e with its magnitude following a Swerling 1 model for known radar cross sections) random variable:

$$p(\mathbf{b}) = \pi^{-M}|\mathbf{Q}_b|^{-1}e^{-\mathbf{b}^H\mathbf{Q}_b^{-1}\mathbf{b}} \quad (10)$$

with $\mathbf{Q}_b$ being diagonal. Alternately, one might use the model

$$p(\mathbf{b}|\mathbf{z}) = \prod_{m-1}^{M} p(b_m|\mathbf{z}) \qquad b_m = A_m e^{j\theta_m} \quad (11)$$

where A_m and θ_m are real and uniformly distributed:

$$p(A_m) = \frac{1}{A_{\max} - A_{\min}} \quad \text{for } A_{\min} \leq A_m \leq A_{\max} \quad (12)$$

$$p(\theta_m) = \frac{1}{2\pi} \quad \text{for } 0 \leq \theta_m \leq 2\pi \quad (13)$$

That says that the complex angles of the amplitudes are completely unknown, and the amplitudes vary uniformly between $A_{\min}$ and $A_{\max}$. The value $A_{\min}$ can be determined from the detection threshold[3] and the value of $A_{\max}$ can be determined by the maximum possible amplitude coming from the A/D converters at the antennas or the subarray outputs.

In the following sections, we shall assume that $p(\mathbf{b})$ is distributed according to (10) with a known covariance matrix $\mathbf{Q}_b$. Additionally, we will assume that $p(\mathbf{U})$ is distributed uniformly about the boresight within a region of length of 1.6 bw where "bw" represents the beamwidth.[4]

A. The Use of Subarrays

In general, the elements in an antenna array are collected into subarrays. Letting $\mathbf{z}$ be the set of observations from the individual array elements, the subarray outputs are determined according to the transformation

$$\tilde{\mathbf{z}} = \mathbf{T}^H\mathbf{D}\mathbf{z} \quad (14)$$

where the length of $\tilde{\mathbf{z}}$ is equal to the number of subarrays. The matrix $\mathbf{D}$ is a diagonal matrix with complex (unity

[2] The maximum signal amplitude gain may be found by evaluating the magnitude of the sum-beam (add together the outputs of all of the array elements) with a target with unit amplitude at the boresight ($\mathbf{u} = \mathbf{0}$ if no phase shift is used) and the noise covariance on the output is $\mathbf{1}^T\mathbf{Q1}$. Using the given antenna array and tapering mentioned in the simulation section, the signal to noise gain at the output is 10.5 dB (The square of the amplitude gain being the power gain).

[3] If the maximum signal amplitude gain is g_M (for the radar model used in the simulation section, $g = 25$) and the detection threshold is τ (detection is generally done on the sum-beam) then the minimum possible amplitude is if all of the waves add constructively, thus $A_{\min} = \tau/(g_M)$. This ignores the details of the detection algorithm estimating the number of targets present.

[4] That is, we are assuming that everything is within a rectangle about the boresight whose sides are, at the closest 0.8 bw from the boresight.

magnitude) values representing the settings of the phase-shifters determining the array look-direction. The matrix $\mathbf{T}$ contains the set of taper weights, which are normally real-valued. The likelihood of the subarray outputs is consequently

$$p(\tilde{\mathbf{z}}|\mathbf{U},\mathbf{b}) = \pi^{-N_s}|\tilde{\mathbf{Q}}|^{-1}e^{-(\tilde{\mathbf{z}}-\tilde{\mathbf{A}}\mathbf{b})^H\tilde{\mathbf{Q}}^{-1}(\tilde{\mathbf{z}}-\tilde{\mathbf{A}}\mathbf{b})} \quad (15)$$

where N_s is the number of subarrays and

$$\tilde{\mathbf{Q}} = \mathbf{T}^H\mathbf{DQDT} \quad \tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1,\tilde{\mathbf{a}}_2,\ldots,\tilde{\mathbf{a}}_M] \quad \tilde{\mathbf{a}}_m = \mathbf{T}^H\mathbf{Da}_m \quad (16)$$

The subarray outputs can be thought of as virtual antenna elements whose phase centers coincide with the phase centers of the subarrays and whose directional response is determined by the subarray pattern. This use of subarrays for superresolution has previously been demonstrated using the MUSIC algorithm in [19]. Superresolution as applied to subarray outputs is also discussed in [13].

III. MMOSPA ESTIMATION

Recent work has drawn attention to the Optimal Subpattern Assignment (OSPA) error metric, rigorously defined in [24], as an optimality criterion for use in tracking algorithms [3], [4], [7], [8], [11], [25]. The OSPA error metric is quite general. For the purposes of this paper, we will consider the most commonly used variant, which assumes that the number of "targets" is known, namely it uses the error measure

$$\bar{d}(\hat{\mathbf{x}},\mathbf{x}) = \frac{1}{N_T}\min_{\mathbf{a}}\|\mathbf{x_a}-\hat{\mathbf{x}}\|^2 \quad (17)$$

This measure represents a squared error between the "reordered" stacked set of true state vectors $\mathbf{x}$ for all N_T targets, and the stacked set of state estimates, $\hat{\mathbf{x}}$. The subscript $\mathbf{a}$ is a permutation vector representing the order of the target states within the stacked vector:

$$\mathbf{a} = [a(1),\, a(2),\, \ldots,\, a(N_T)]' \quad (18)$$
$$\hat{\mathbf{x}}_\mathbf{a} = [\hat{\mathbf{x}}(a(1))',\, \hat{\mathbf{x}}(a(2))',\, \ldots,\, \hat{\mathbf{x}}(a(N_T))']' \quad (19)$$

where each $a(i)$ is a number between 1 and N_T such that $a(i)\neq a(j)$ and $\hat{x}_i$ represents the ith target state.

The Minimum Mean OSPA Error (MMOSPA) estimate is, by definition:[5]

$$\hat{\mathbf{x}}_M \triangleq \arg\min_{\hat{\mathbf{x}}}\mathrm{E}\left[\min_{\mathbf{a}}\|\mathbf{x_a}-\hat{\mathbf{x}}\|^2 \,\Big|\, \mathbf{z}\right] \quad (20)$$

MMOSPA estimation is attractive for target tracking, because it allows for "smooth" estimates, similar to Minimum Mean Squared Error (MMSE) estimation, while avoiding the problem of "coalescence" that can occur. Coalescence occurs when the identity of the targets becomes uncertain. For example, suppose that there exist two equiprobable hypotheses for the target states. One is $\mathbf{x} = [\hat{\mathbf{x}}(1)',\hat{\mathbf{x}}(2)']'$ and the other is $\mathbf{x} = [\hat{\mathbf{x}}(2)',\hat{\mathbf{x}}(1)']'$. Here, the only difference is which target

gets which state. Though it is clear that the correct target states are $\hat{\mathbf{x}}(1)$ and $\hat{\mathbf{x}}(2)$ (even if we do not know which state goes with which target), the MMSE estimate of the states is the same for both targets and is directly between the correct estimates, i.e. $(\hat{\mathbf{x}}_1+\hat{\mathbf{x}}_2)/2$. On the other hand, the MMOSPA estimate is $\hat{\mathbf{x}}(1)$ and $\hat{\mathbf{x}}(2)$, but it does not say which state is associated with which target.

In addition to tracking, MMOSPA estimation can be applied to angular superresolution, among other applications. In angular superresolution, the likelihood, (5), has a symmetry in terms of the identities of each $(\mathbf{u}_m, b_m)$ pair. Whereas in target tracking applications, $\mathbf{x}$ is a stacked vector of target states, when performing MMOSPA angular superresolution, the "states" are the directions of arrival of the target returns, $\mathbf{u}_m$. In [7], MMOSPA estimation for angular superresolution was considered for the case of a simple linear array given that the complex amplitudes of the target returns were known. The scalar nature of the directions of arrival (essentially, one could estimate azimuth, but not elevation) allowed for the minimization inherent to MMOSPA estimation to be performed explicitly.

Assuming that the complex amplitudes of the target estimates have a complex Gaussian prior, as in (10), then the signal model of (3) is the sum of two Gaussian random variables, whereby $\mathbf{Ab}$ has a degenerate covariance matrix due to linear dependence of the estimates. The conditional PDF $p(\mathbf{z}|\mathbf{U})$ is this the PDF of the sum of two Gaussian random variables:

$$p(\mathbf{z}|\mathbf{U}) \sim \mathcal{N}\left\{\mathbf{z};\mathbf{0},\mathbf{Q}+\mathbf{AQ}_b\mathbf{A}^H\right\} \quad (21)$$

However, in order to be able to perform MMOSPA estimation, we need $p(\mathbf{U}|\mathbf{z})$, not just $p(\mathbf{z}|\mathbf{U})$. Using, Bayes' theorem, and assuming that $p(\mathbf{U})$ is distributed uniformly over the $(-1,1)$ interval for each direction of arrival, this can be obtained as

$$p(\mathbf{U}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{U})p(\mathbf{U})}{p(\mathbf{z})} = \frac{p(\mathbf{z}|\mathbf{U})}{\int_\mathbf{U}p(\mathbf{z}|\mathbf{U})d\mathbf{U}} \quad (22)$$

In [5] an explicit solution for the MMOSPA estimate of two scalar targets was derived[6]. However, we found that standard quadrature integration to evaluate the integrals for that solution when used in the Simulation section tended to produce poor results. Thus, instead, we directly evaluated the integral in (20) using importance sampling, which we discuss in the following subsection.

A. MMOSPA Estimation via Importance Sampling

Importance sampling samples one PDF, $g(\mathbf{x})$, to find the mean of another PDF, $h(\mathbf{x})$. It does this noting that

$$\bar{x} = \int_\mathbf{x} h(\mathbf{x})\mathbf{x}\,d\mathbf{x} = \int_\mathbf{x}\left(\frac{h(\mathbf{x})}{g(\mathbf{x})}\right)g(\mathbf{x})\mathbf{x}\,d\mathbf{x} = \mathrm{E}_{g(\mathbf{x})}\left[\left(\frac{h(\mathbf{x})}{g(\mathbf{x})}\right)\mathbf{x}\right] \quad (23)$$

---

[5]In practice, one would probably want to perform the optimization considering only the directional components of $\mathbf{x}$, since the value $\|\mathbf{x_a}-\hat{\mathbf{x}}\|^2$ does not have a physical meaning; it mixes the squares of angular differences with the squares of amplitude differences.

[6]A formulation of this solution for more than two scalar targets has been derived and will be presented in [6].

Thus, the mean of $h(\mathbf{x})$ can be approximated using N samples as

$$\hat{\bar{x}} = \frac{1}{N} \sum_{j=1}^{N} \frac{h(\mathbf{x}_j)}{g(\mathbf{x}_j)} \mathbf{x}_j \qquad (24)$$

where $\mathbf{x}_j$ is the jth sample of $g(\mathbf{x})$.

Suppose, however, that

$$h(\mathbf{x}) = \frac{1}{c} f(\mathbf{x}) \qquad\qquad c = \int_{\mathbf{x}} f(\mathbf{x})\, d\mathbf{x} \qquad (25)$$

In this case, the constant, c too can be found by importance sampling by sampling $g(\mathbf{x})$ and calculating

$$\hat{c} = \frac{1}{N} \sum_{j=1}^{N} \frac{f(\mathbf{x}_j)}{g(\mathbf{x}_j)} \qquad (26)$$

Substituting into (24), we can estimate the mean of $h(\mathbf{x})$ as

$$\hat{\bar{x}} = \frac{1}{\sum_{j=1}^{N} \frac{f(\mathbf{x}_j)}{g(\mathbf{x}_j)}} \sum_{j=1}^{N} \frac{f(\mathbf{x}_j)}{g(\mathbf{x}_j)} \mathbf{x}_j \qquad (27)$$

To relate this to the problem at hand, we can say that $\mathbf{x} = [\mathbf{U}]'$ and $f(\mathbf{x}) = p(\mathbf{z}|\mathbf{U})p(\mathbf{U})$ and $h(\mathbf{x}) = p(\mathbf{U}|\mathbf{z})$. Thus using (24) and choosing an appropriate sample distribution, $g(\mathbf{x})$, one can use (24) to find MMSE estimates of $\mathbf{U}$. Since the calculation of the MMSE and MMOSPA estimates from a sampled PDF differs only in the ordering of the samples when finding the expected value, MMOSPA estimates can be found by taking N_i samples of $g(\mathbf{x})$, adjusting the ordering of the states within each sample, as described in [4], [8], [25] and then estimating the mean using (27) with the reordered samples.

Importance sampling lets us use a simple PDF to sample a difficult PDF over the same domain. However, if the simple PDF is too different from the original PDF, that is, if it is highly probably where the original PDF is improbable, then the accuracy of the estimates obtained by importance sampling will increase very slowly as the sample size increases.

In the simulation section, when considering the linear array, we let the sample distribution be a Gaussian PDF whose mean was the mean of the MAP estimate, and whose covariance was taken to be a diagonal matrix with 0.03 on the diagonal, a number proportional to the average error of the MAP estimates across the entire array. In practice, covariance values could be read from a table based upon an observed signal to noise ratio.

The Monte Carlo integration was performed using $4,000$ points in the simulation section. That said, on occasion the proposal distribution was sufficiently mismatched from the sample distribution that $f(\mathbf{x}) = 0$ within machine precision for a number of samples. In such cases, we did not count those points towards the $4,000$ and resampled.

In order to perform the minimization in a computationally efficient manner, such that we need not store all of the points generated from importance sampling at once, we did it in a suboptimal, sequential manner. The sequential algorithm for adding the ith point generated from importance sampling is

1) Evaluate the MMOSPA estimate using the $i-1$ previous samples and the orderings for them as previously determined. That is use equation

$$\mathbf{x}_M^{i-1} = \frac{1}{\sum_{j=1}^{i-1} \frac{f(\mathbf{x}_j)}{g(\mathbf{x}_j)}} \sum_{j=1}^{i-1} \frac{f(\mathbf{x}_j)}{g(\mathbf{x}_j)} \mathbf{x}_{j,\mathbf{a}_j} \qquad (28)$$

where $\mathbf{x}_{j,\mathbf{a}_j}$ denotes sample $\mathbf{x}_j$ having the ordering given by $\mathbf{a}_j$.

2) Given the current sample, Determine the ordering of the current sample, $\hat{\mathbf{x}}_i$ as the ordering minimizing the distance between the same and the partial MMOSPA estimate $\hat{\mathbf{x}}_M^{i-1}$.

3) Generate the next sample via importance sampling and return to step 1 until enough samples have been generated.

IV. COMPRESSIVE SENSING ESTIMATION

Compressive sensing is a term for a number of methods to solve for sparse solutions to underdetermined linear problems. Compressive sensing has been considered for angular superresolution in radar array processing in [12], [14], [15], [22]. In [15] and [1], it was shown that compressive sensing is equivalent to MAP estimation with a Laplacian prior on the complex amplitudes. In this case, the problem is that of a Least Absolute Selection and Shrinkage Operator (LASSO) problem [26]:

$$\min_{\mathbf{b}} \|\mathbf{z} - \mathbf{A}\mathbf{b}\|_2^2 \qquad (29)$$
$$\text{subject to } \|\mathbf{b}\|_1 \le \epsilon \qquad (30)$$

Unlike in the problem formulation discussed in Section II, we are not explicitly estimating the direction of array of the target returns. Instead, the matrix of steering vectors $\mathbf{A}$ is set to contain a steering vector for *every possible direction of arrival* that is to be considered jointly for both targets (a dictionary). We shall let the number of entries be N_A. The complex, nonzero values in $\mathbf{b}$ are thus the complex amplitudes of the target returns, also indicating from which directions the returns came. Varying the ϵ term changes the sparsity of the solution, with a smaller term leading to a sparser solution.

Though seldom mentioned, correlations in the noise (i.e. jammers) can be taken into account when performing compressive sensing. The optimization criterion given in (29) is equivalent to the logarithm of the likelihood expressed in (5) with $\mathbf{Q} = \mathbf{I}$, dropping the constant terms. Changing the optimization criterion to minimize $\|\mathbf{Q}^{-\frac{1}{2}}\mathbf{z} - \mathbf{Q}^{-\frac{1}{2}}\mathbf{A}\mathbf{b}\|_2^2$ takes that into account. Thus, the compressive sensing problem is the same as maximizing (5) with an enormous $\mathbf{b}$ vector having a prior distribution of

$$p(\mathbf{b}) = \frac{\lambda}{2} \exp\left[-\frac{\lambda}{2}\|\mathbf{b}\|_1\right] \qquad (31)$$

where the value of λ depends upon the value of ϵ chosen. This is a specific realization of the generalized multivariate Laplace distribution, presented in [10]. Assuming that the number of directions of arrival is known *a priori*, one could perform a

search for an ϵ term that provides a solution with the correct sparsity.

However, since the LASSO optimization is not well suited for a problem having a known sparsity, we will implement the following, more difficult optimization, which is also considered a form of compressive sensing,

$$\min_{\mathbf{b}} \|\mathbf{z} - \mathbf{A}\mathbf{b}\|_2^2$$

$$\text{subject to } \|\mathbf{b}\|_0 = n$$

where what appears to be an l_0-norm of $\mathbf{b}$ represents the number of nonzero entries in $\mathbf{b}$.[7] In other words, one is searching for the best solution to the problem such that the number of nonzero elements in $\mathbf{b}$ is equal to n.

If we know which elements of $\mathbf{b}$ are nonzero, then the solution is simple. That is, if we let $\tilde{\mathbf{b}}$ be $\mathbf{b}$ with the zero-elements removed and $\tilde{\mathbf{A}}$ be $\mathbf{A}$ with the columns corresponding to the zero elements of $\mathbf{b}$ removed, then the optimization problem is

$$\min_{\tilde{\mathbf{b}}} \|\mathbf{z} - \tilde{\mathbf{A}}\tilde{\mathbf{b}}\|_2^2$$

and the solution is

$$\tilde{\mathbf{b}} = \left(\tilde{\mathbf{A}}'\tilde{\mathbf{A}}\right)^{-1} \tilde{\mathbf{A}}'\mathbf{z} \tag{32}$$

We shall thus solve this optimization problem by brute force. Given M targets, we evaluate the solutions for all $\binom{N_A}{M}$ possible combinations of nonzero elements in $\mathbf{b}$ and choose the solution with the lowest cost. Due to the brute force nature of the solution, it is not feasible when a large number of targets is present, though it can be easily parallelized across many processors. Indeed, with a moderate number of targets and good parallelization, the brute-force approach could be faster than algorithms such as [12], which is hard to parallelize, since it requires a large amount of branching to search a tree structure. This brute force approach shall be considered as a "best case" solution to the compressive sensing problem, as typically, approximate solutions are used.

A. Building the Dictionary

When performing compressive sensing, a question arises regarding how the discretization of the parameter space should be done in order to build the dictionary. In this case, which directions of arrival should be entered into $\mathbf{A}$? In general, one would want to have a higher density of points around more probable areas so that the final result would consist of a higher expected accuracy. With complicated PDFs, such a variable density dictionary might be obtained by drawing Monte Carlo samples from the prior and using them to form the dictionary entries. In the case of this paper, since the prior discussed in Section II is uniform within a region of $1.6\,\mathrm{bw}$ centered on the boresight, we shall simply uniformly distribute the samples over the valid region.

Compressive sensing has been extensively used in other aspects of radar array processing and radar imaging, whereby an extensive bibliography can be found at [23].

[7]This notation was first used in compressive sensing in [9].

V. MAXIMUM LIKELIHOOD AND MAXIMUM A POSTERIORI ESTIMATION

Given an initial estimate of $\mathbf{U}$, when using a planar array, maximum likelihood direction estimation is typically performed by building a gradient for each $u_{m,d}$ corresponding to the (x, y) coordinates of the vector and using a Quasi-Newton method (see, e.g, [2]) on (5). Note that in this case we are not explicitly constraining such that each $\mathbf{u}_m$ is a unit vector, because we can determine the $u_{m,3}$ coordinates perpendicular to the plane of the radar implicitly from $u_{m,1}$ and $u_{m,2}$, i.e. $u_{m,3} = \sqrt{1 - u_{m,1}^2 - u_{m,2}^2}$ and if the initial estimates are reasonably good, (e.g. within the beam), one does not run into scenarios where either the $u_{m,1}$ and the $u_{m,2}$ coordinates of $\mathbf{u}_m$ are too large to represent a valid unit vector. Basically, previous ML estimation approaches have implicitly included a uniform prior distribution on the directional angles about the boresight. The most common solution approach, however, has been a conjugate gradient method due to the simple form of the gradient [18], [28]. It should be noted that the covariance matrix $\mathbf{Q}$ can contain information regarding the location of Gaussian jammers. This covariance matrix could be obtained by sampling the noise in the absence of a signal.

MAP estimation is almost identical to that of ML estimation, except we have to take into account the prior on the complex amplitudes (the uniform prior on the arrival directions will only be implicitly included, i.e. when performing optimization, one simply has to begin within the valid region and allow an iteration step to leave the valid region.

Taking the logarithm of (6) (using the complex Gaussian prior on the complex magnitudes), dropping the constant terms, and multiplying by -1, we find that we want to *minimize*

$$\mathcal{L}[\mathbf{U}, \mathbf{b}] = (\mathbf{z} - \mathbf{A}\mathbf{b})^H \mathbf{Q}^{-1}(\mathbf{z} - \mathbf{A}\mathbf{b}) + \mathbf{b}^H \mathbf{Q}_b^{-1}\mathbf{b} \tag{33}$$

We shall first minimizing (33) with respect to $\mathbf{b}$ and substitute the result back in. The gradient with respect to $\mathbf{b}$, dividing out a constant of 2, is[8]

$$\frac{1}{2}\nabla_{\mathbf{b}}\mathcal{L}(\mathbf{U}, \mathbf{b}) = -\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{z} + \left(\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{A} + \mathbf{Q}_b^{-1}\right)\mathbf{b} \tag{34}$$

The optimal value for $\mathbf{b}$ is thus

$$\mathbf{b} = \left(\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{A} + \mathbf{Q}_b^{-1}\right)^{-1} \mathbf{A}^H \mathbf{Q}^{-1}\mathbf{z} \tag{35}$$

Substituting back in, noting that $\mathbf{Q}^{-1} = \left(\mathbf{Q}^{-1}\right)^H$ (because it is a covariance matrix), we get

$$\mathcal{L}[\mathbf{U}] = \mathbf{z}^H \left(\mathbf{I} - \mathbf{A}\mathbf{Q}^{-1}\left(\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{A} + \mathbf{Q}_b^{-1}\right)^{-1}\mathbf{A}^H\right)\mathbf{Q}^{-1}$$
$$\times \left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{A} + \mathbf{Q}_b^{-1}\right)^{-1}\mathbf{A}^H \mathbf{Q}^{-1}\right)\mathbf{z} \tag{36}$$
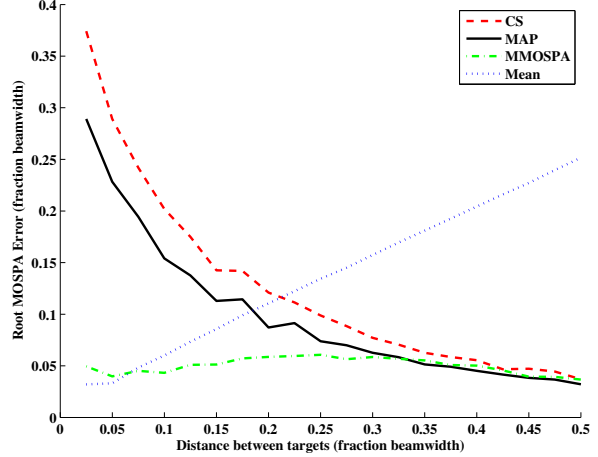
performing out the multiplication[9], dropping a constant $\mathbf{Q}^{-1}$ term and multiplying by -1 we end up with the following

[8]We shall note the identities for complex vectors that $\frac{\partial}{\partial \mathbf{b}^*}\mathbf{a}^H\mathbf{b} = 0$ and $\frac{\partial}{\partial \mathbf{b}^*}\mathbf{b}^H\mathbf{a} = \mathbf{a}$. The gradient vector of a complex function, J, is simply $\nabla_{\mathbf{b}}J = 2\frac{\partial}{\partial \mathbf{b}^*}J$.

[9]Note that $\mathbf{A}^H \mathbf{Q}^{-1}\mathbf{A}$ is real and that $(\mathbf{X}\mathbf{Y})^H = \mathbf{Y}^H\mathbf{X}^H$.
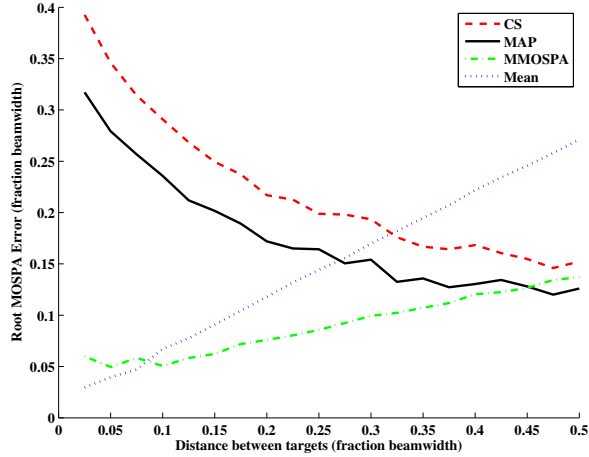
(a) All Complex Amplitudes (b) Only the Top 95%

Fig. 2. The MOSPA error performance of the various algorithms for estimating the DOA of two targets on a linear array. Excluding low-probability complex amplitudes, the MMOSPA estimate exhibits relatively constant error as a function of the separation.

quantity that we would like the *maximize* the following radar scan pattern

$$S[\mathbf{U}] = \mathbf{z}^H \mathbf{Q}^{-1} \mathbf{A} \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1} \mathbf{A}^H \mathbf{Q}^{-1} \mathbf{z} \quad (37)$$

$$= \left\| \mathbf{W}^H \mathbf{z} \right\|^2 \quad \mathbf{W} \triangleq \mathbf{Q}^{-1} \mathbf{A} \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1/2}$$
$$(38)$$

It can be seen that MAP estimation is equivalent to ML estimation if $\mathbf{Q}_b^{-1} = 0$.

The derivative of $S[\mathbf{U}]$ with respect to a scalar direction element $u_{m,d}$ is given in Equation (39), where $\mathbf{A}_{u_{m,d}}$ is the derivative of matrix $\mathbf{A}$ with respect to $u_{m,d}$. Equation (39) was simplified using the product and chain rules and by noting that if $z(x) = a(x) + jb(x)$ where $a(x)$ and $b(x)$ are real functions of the real scalar x, then $\frac{\partial z}{\partial x} = \frac{\partial a}{x} + j\frac{\partial b}{\partial x}$. Additionally, we used the identity for the matrix derivative of an inverse: $\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1} \frac{\partial \mathbf{Y}}{\partial x} \mathbf{Y}^{-1}$. The derivative of the element in row i and column j of $\mathbf{A}$ with respect to $u_{m,d}$ is

$$\mathbf{A}_{u_{m,d}}(i,j) = \begin{cases} j\omega r_{m,d} e^{j\omega \mathbf{r}_i' \mathbf{u}_m} & \text{if } j = m \\ 0 & \text{otherwise} \end{cases} \quad (40)$$

where $r_{m,d}$ is the appropriate dimension of the vector from the origin to the mth antenna.

Thus, equation (39) can be used to build the gradient for Quasi-Newton estimation of the directions of arrival. If the complex amplitudes of the target returns are desired, they can be obtained using (35). Order estimation can be performed using various criteria, such as the MDL, as was done in [16].

A. Generating Initial Estimates

In implementing the MAP estimator for the simulation, it was noticed that randomly choosing initial estimates within the beam did not always produce good results. The results tended to be poor if the initial estimates were too close together.

In the simulations, since we had two targets, we ran the algorithm twice using different initial estimates. During the

first run, the estimates were 0.6 bw apart, centered about the boresight. During the second run, one estimate was at 0.3 bw and the other at -0.5 bw. In situations with more targets, the initial estimates could be placed uniformly within the beam.

VI. SIMULATIONS

The estimation algorithms were run on a 5-element linear array without tapering. The -3 dB beamwidth was approximately $41°$. The element spacing was assumed to be $\lambda/2$, obviating the need to choose a carrier frequency. The boresight was chosen to be straight ahead, i.e. $\mathbf{D}$ was set to the identity matrix. One target was fixed on the boresight. The location of the second target was varied within the beam. The complex amplitudes were generated as complex normal random variables. The average power of the complex amplitudes was chosen to be 20 dB above the noise at each of the elements, with the covariance matrix $\mathbf{Q}_b$ being diagonal. For each position, $1,000$ Monte Carlo runs were performed. The dictionary for the CS solution was composed of 200 entries ranging from -0.8 bw to 0.8 bw. To make the comparison fair for the other estimators, estimates from the other solutions that were outside of the range from -0.8 bw to 0.8 bw were clipped to the edges of the region.

Figure 2 shows the root of the MOSPA error as a fraction of the beamwidth for CS, MAP, MMOSPA and MMSE estimation. In (a) all randomly generated complex amplitudes (with random phases) were used, whereas in (b), the bottom 5% of amplitudes were discarded. This can be more representative of the results from detection algorithms, since low-amplitude signals are unlikely to be detected. The MMSE estimate was calculated using an importance sampling technique similar to that used to find the MMOSPA estimate to avoid precision problems that caused quadrature integration to produce unreliable results. The MMSE estimate becomes notably worse at the targets separate, since it is the same for both directions (due to the symmetry of the PDF) and is located between both of the

$$\frac{\partial}{\partial u_{m,d}} S[\mathbf{U}] = \mathbf{z}^H \mathbf{Q}^{-1} \left(\mathbf{A}_{u_{m,d}} \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1} \mathbf{A}^H + \mathbf{A} \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1} \mathbf{A}_{u_{m,d}}^H \right) \mathbf{Q}^{-1} \mathbf{z}$$
$$- \mathbf{z}^H \mathbf{Q}^{-1} \mathbf{A} \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1} \left(\mathbf{A}_{u_{m,d}}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}_{u_{m,d}} \right) \left(\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} + \mathbf{Q}_b^{-1} \right)^{-1} \mathbf{A}^H \mathbf{Q}^{-1} \mathbf{z} \qquad (39)$$

true directions of arrival. In other words, MMSE estimation is not suitable for angular superresolution. As expected, the MMOSPA estimate performed the best. The CS solution was worse than the MAP solution, which is not surprising, since the CS solution was really just a discretized version of maximum likelihood estimation. The brute-force CS solution was implemented as unoptimized C-code and called as a mex file in MATLAB. Note counting the time taken to generate the dictionary, which only be done once, the CS solution took about 0.06 seconds to execute for each each combination of target locations in each Monte Carlo run, demonstrating that with sufficiently small problems, approximations are not necessary. The MMOSPA estimator was implemented directly in MATLAB and took approximately 0.45 seconds to execute. The MAP estimator was also implemented in MATLAB without optimization and took about 0.1 seconds to execute.

VII. RESULTS AND CONCLUSIONS

We presented MAP, MMOSPA and CS methods for estimating the directions of arrival of returns from multiple targets using a linear array. The exact compressive sensing solution, being merely a discretized ML solution, was found to have the worst performance. The iterative MAP solution was found to be better, but it generally did not beat our suboptimal approximation of the optimal MMOSPA estimate. Though MMOSPA estimation might be too computationally complex for real-time estimation on a large array estimating many directions of arrival, it can be used to estimate a lower bound when evaluating the performance of different estimation algorithms. However, care has to be taken when choosing the proposal distribution when numerically evaluating the integral.

REFERENCES

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 53–63, Jan. 2010.

[2] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999, ch. 1.7.

[3] H. A. P. Blom and E. A. Bloem, "Optimal decomposed particle filtering of two closely spaced Gaussian targets," in *50th IEEE Conference on Decision and Control and European Control Conference*, Dec. 2011, under Review.

[4] D. F. Crouse, P. Willett, and Y. Bar-Shalom, "Developing a real-time track display that operators do not hate," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3441–3447, Jul. 2011.

[5] ——, "Generalizations of blom and bloem's PDF decomposition for permutation-invariant estimation," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, pp. 3840–3843.

[6] D. F. Crouse, P. Willett, Y. Bar-Shalom, and L. Svensson, "Aspects of mmospa estimation," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL, Dec. 2011.

[7] D. F. Crouse, P. Willett, L. Svensson, and D. Svensson, "Aspects of MMOSPA estimation," in *50th IEEE Conference on Decision and Control and European Control Conference*, Dec. 2011, to be presented.

[8] D. F. Crouse, P. Willett, L. Svensson, D. Svensson, and M. Guerriero, "The set MHT," in *14th International Conference on Information Fusion*, Jul. 2011.

[9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] T. Eltoft, T. Kim, and T.-W. Lee, "On the multivariate laplace distribution," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 300–303, May 2006.

[11] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting two birds with two bullets: How to find minimum mean OSPA estimates," in *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, Jul. 2010.

[12] G. Karabulut, T. Kurt, and A. Yongacoglu, "High resolution estimation of directions of arrival," in *Proceedings of the 61st IEEE Vehicular Technology Conference*, vol. 1, May–Jun. 2005, pp. 20–24.

[13] R. Klemm, Ed., *Applications of Space-Time Adaptive Processing*. London: Institute of Engineering Technology, 2004, ch. 16.2.

[14] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[15] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Source localization by enforcing sparsity through a laplacian prior: an SVD-based approach," in *Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing*, Oct. 2003, pp. 573–576.

[16] U. Nickel, "Winkelaufloesung eng benachbarter Ziele mit Gruppenantennen," Ph.D. dissertation, Rheinisch-Westfaelische Technische Hochschule, Aachen, Germany, 1982.

[17] ——, "Monopulse estimation with adaptive arrays," *IEE Proceedings-F*, vol. 140, no. 5, pp. 303–308, Oct. 1993.

[18] ——, "Radar target parameter estimation with array antennas," in *Radar Array Processing*, S. Haykin, J. Litva, and T. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 3.

[19] ——, "Spotlight MUSIC: Super-resolution with subarrays with low calibration effort," *IEE Proceedings of Radar, Sonar and Navigation*, vol. 149, pp. 166–173, Aug. 2002.

[20] ——, "Performance analysis of space-time adaptive monopulse," *Signal Processing*, vol. 84, no. 9, pp. 1561–1579, Sep. 2004.

[21] B. Otterson, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 4.

[22] A. Panahi and M. Viberg, "Maximum a posteriori based regularization parameter selection," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 2011, pp. 2452–2455.

[23] Compressive sensing resources. Rice University. [Online]. Available: http://dsp.rice.edu/cs

[24] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.

[25] L. Svensson, D. Svensson, and P. Willett, "Set JPDA algorithm for tracking unordered sets of targets," in *Proceedings of the 12th International Conference on Information Fusion*, Seattle, WA, Jul. 2009, pp. 1187–1194.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological*, vol. 58, no. 1, pp. 267–288, 1996.

[27] Z. Zhu and S. Haykin, "Radar detection using array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 3.

[28] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.