Bayesian Approaches for Robust Array Signal Processing

A. Lee Swindlehurst
Dept. of Elec. & Comp. Engineering
Brigham Young University
Provo, UT 84602
e-mail: swindle@ee.byu.edu

Mats Viberg
Dept. of Applied Electronics
Chalmers University of Technology
Gothenburg, Sweden S-412 96
e-mail: viberg@ae.chalmers.se

Abstract

Modern techniques for sensor array signal processing are hampered by the often unrealistic and overly simplistic assumptions used in their development. Foremost among these assumptions are those relating to the array geometry or response, including for example known sensor positions, availability of complete gain/phase/mutual coupling calibration data, uniform linear arrays, etc. Of course, with real arrays none of these assumptions hold exactly. Deviations from the nominal model occur due to environmental factors, quantization effects, perturbations in the locations of the antenna elements, etc. If such model errors are ignored, serious performance degradation can result. While in any given application the exact value of the perturbation to the array is unknown, the size or distribution of such perturbations may be well understood. Consequently, in this paper a Bayesian approach is adopted to show how information in the form of an a priori distribution on the array model errors can be used to improve both direction of arrival and beamforming performance. The general maximum a posteriori estimator for the problem is formulated, and a computationally attractive alternative based on the concept of subspace fitting is proposed. The algorithm’s statistical performance for both DOA estimation and beamforming is evaluated by means of some simulation examples.

1. Introduction

1.1. Historical Background

The general problem addressed in this work is that of locating the sources of radiated energy received by an array of spatially distributed collectors, or sensors. It was, in fact, the desire to locate and track enemy aircraft using radar that gave birth to statistical signal processing as a sub-discipline of electrical engineering in the 1940s. In this early application, the parameters of interest were the direction, range, and radial velocity of

This work was supported by the National Science Foundation under grant MIP-9408154.
the target, all of which were to be estimated from the outputs of a single radar antenna. The direction to the target was the most important parameter to be estimated, since it had to be known before range and velocity could be determined. Target direction was usually determined by placing a collimating dish behind the antenna, sweeping the entire device through all possible target directions, and then finding the directions from which maximum energy was received. This simple system has been modified and refined over the years, and sophisticated algorithms have been developed to enhance radar direction-finding performance.

The problems of direction-finding and emitter localization have become important in many other areas besides radar. For example, acoustic signals received by hydrophones are used in underwater applications to detect and locate submarines and surface vessels. In oil exploration, explosive charges are detonated below the earth’s surface in order to create acoustic reflections off the various layers of the earth’s crust. Measurement of the directions-of-arrival (DOAs) of these reflections help determine the relative position and thickness of these layers. A common problem in both ground- and satellite-based communications is the cancelation of unwanted interference signals from one direction in favor of a desired signal arriving from another. Effective cancelation of the interference is highly dependent on knowledge of its DOA.

The applicability of source localization to a wide variety of problems has led to a correspondingly large amount of research and considerable refinement of the techniques used to obtain the location estimates. One of the very first refinements to the simple source location system described above was the use of an array of multiple collectors to increase sensitivity and spatial resolution. The first techniques that exploited this antenna array structure were referred to as delay-and-sum (DS) beamformers, since the individual antenna outputs were delayed, weighted, and summed together to form a single array output. By varying the delays and weights, the directional response of the array (sometimes called an antenna beam, hence the term beamforming) can be steered to various directions of interest, and the amount of energy impinging upon the array from these directions can be measured. Thus, not only is speed and reliability gained since the antenna is electronically rather than physically steered, but directionality is improved since the incoming energy is sampled over a much wider physical aperture.

Beamforming techniques have been widely used up to the present, primarily because
they are simple and computationally efficient. Improvements have been made over the years to the original DS approach (see esp. [1, 2]), and adaptive implementations have been extensively studied (e.g., [3, 4, 5]). To shape the response of the array, some approaches have made use of the fact that, for a linear array, the DS and weighting operations are equivalent to a windowed discrete Fourier transform (DFT) in the spatial dimension. Although providing a link with the well-developed theory of time domain filter design, this connection points out that the resolution and convergence problems of the DFT are inherited by beamforming techniques. As an example, even with an infinite amount of noise-free data, DS beamforming will not generally give biased estimates of the location parameters of multiple emitters. This undesirable property is referred to as estimator inconsistency, and is the major drawback to use of beamforming algorithms in practice.

A number of methods borrowed from time series estimation have also been applied to the DOA problem. These techniques are typically based on parametric linear prediction or autoregressive (AR) models for the data [6, 7, 8, 9], and they tend to provide much higher resolution than conventional DS beamforming. Because these methods are derived for uniformly sampled time series, for the DOA estimation problem they are limited in application to uniform linear arrays. Their principal drawback stems from the fact that AR data models are not well matched to the signal-in-additive-noise model most often assumed for the array processing problem. The result is that even at moderate SNR values, the parameter estimates may, as in the case of beamforming, be inconsistent and subject to large biases.

1.2. Modern Techniques

As the field of source location research has matured, many have hearkened back to the statistical foundations of this parameter estimation problem and found maximum likelihood (ML) solutions [10, 11, 12, 13]. The motivation for studying such methods is the desire to obtain unbiased, consistent location estimates of minimum variance. Since, in general, ML techniques require multidimensional nonlinear optimization procedures, most of the emphasis in recent years has been the development and analysis of suboptimal, but computationally simpler methods that achieve near-ML performance. Perhaps the biggest breakthrough came with the development of the so-called MUSIC
algorithm (for MUltiple SIgnal Classification) by Schmidt in 1979 [14, 15]. A similar approach was independently proposed by Bienvenu and Kopp that same year [16]. The MUSIC approach provided an elegant geometrical solution to the source location problem using the powerful mathematical tools of vector spaces. Although earlier work had hinted at this type of solution [17, 18, 19], the MUSIC algorithm was the first to fully exploit the underlying signal model for the antenna array problem.

Research in high-resolution DOA estimation has been dominated throughout the past ten years by analysis and extensions of the MUSIC algorithm. Its popularity is primarily due to the fact that it may be applied to antenna arrays of arbitrary geometry to estimate multiple parameters per source. Unlike beamforming and AR modeling techniques, MUSIC is a consistent, high-resolution estimator for situations involving multiple emitters and additive noise. MUSIC does, however, share some of the computational drawbacks of these algorithms. For example, the algorithm requires that the response of the array for all possible source parameterizations be known a priori. In addition, determining the MUSIC parameter estimates involves a potentially burdensome search procedure that is not guaranteed to have adequate resolution. For single-parameter, narrowband direction-finding applications, the recently introduced ESPRIT algorithm (for Estimation of Signal Parameters via Rotational Invariance Techniques) alleviates these difficulties by constraining the antenna array to be composed of two identical, but otherwise arbitrary subarrays [20, 21]. This special array geometry results in an elegant and computationally efficient solution that does not require prior knowledge of the subarray response.

Recent investigations of a special formulation of ESPRIT have demonstrated that beamforming, MUSIC, maximum likelihood, ESPRIT, and various other algorithms are all special cases of a general subspace fitting minimization procedure [22, 23, 24]. Unifying these techniques under a common framework sheds light on their geometrical relationships and facilitates comparative analyses. For example, the subspace fitting paradigm has been used to evaluate the asymptotic (in the data) performance of the above algorithms [25, 26, 27, 28, 29]. A byproduct of this analysis was the development of an optimal weighted subspace fitting algorithm that achieves parameter estimates of minimum variance [29, 30, 31].
1.3. Robustness Issues

All of the methods for direction-finding (DF) listed above rely on the availability of information about the array response, and assume that the signal wavefronts impinging on the array have perfect spatial coherence (e.g., perfect plane waves). The array response may be determined by either empirical measurements (a process referred to as array calibration), or by making certain assumptions about the sensors in the array and their geometry (e.g., identical sensors in known locations). Unfortunately, an array cannot be perfectly calibrated, and idealized assumptions made about the array geometry and wave propagation are never satisfied in practice. Due to changes in antenna location, temperature, and the surrounding environment, the response of the array may be significantly different than when it was last calibrated. Furthermore, the calibration measurements themselves are subject to gain and phase errors, and they can only be obtained for discrete DOA values (thus necessitating interpolation techniques for uncalibrated directions).

For the case of analytically calibrated arrays of nominally identical, identically oriented elements, errors result since the elements are not really identical and their locations are not precisely known. Furthermore, even if the calibration data were precisely known, inhomogeneous propagation effects cause the actual array response to a given signal to be different from the response at the time of calibration. Depending on the degree to which the actual antenna response differs from its nominal value, DF and beamformer performance may be significantly degraded.

To account for the effects described above, a slightly generalized model for the array response will be considered in this work. The response will be parameterized not only by the directions of arrival (DOAs) of the signals, but also by a vector of perturbation or “nuisance” parameters that describe deviations of the response from its nominal value. These parameters can include, for example, displacements of the antenna elements from their nominal positions, uncalibrated receiver gain and phase offsets, etc.. With such a model, a natural approach is to attempt to estimate the unknown nuisance parameters simultaneously with the signal parameters. Such methods are referred to as auto-calibration techniques, and have been proposed by a number of authors, including [32, 33, 34, 35, 36, 37, 38] among many others.

When auto-calibration techniques are employed, it is critical to determine whether
both the signal and nuisance parameters are identifiable. In certain cases they are not; for example, one cannot uniquely estimate both DOAs and sensor phase characteristics unless of course additional information is available, such as sources in known locations [39, 40, 41], cyclostationary signals with two or more known cycle frequencies [42], or partial information about the phase response of the array [43]. The identifiability problem can be alleviated if the perturbation parameters are assumed to be drawn from some known a priori distribution. While this itself represents a form of additional information, it has the advantage of allowing an optimal maximum a posteriori (MAP) solution to the problem to be formulated [35, 37]. In [37] it is shown that, by using an asymptotically equivalent approximation to the resulting MAP criterion, the estimation of the signal and nuisance parameters can be decoupled, leading to a significant simplification of the problem. This result will be reviewed in the present work.

Presumably, any of the above auto-calibration methods would provide not only improved DOA estimates, but also calibration information that would be useful in beamformer implementation. Another goal of this paper is to investigate beamformer performance for the case where the optimal MAP perturbation parameter estimates of [37] are used to update the array calibration. Simulations demonstrate that such an approach can result in a significant performance improvement, measured using either interference rejection capability or mean-squared error. In addition, for simple additive unstructured calibration errors, the MAP approach is shown in certain cases to yield a beamformer similar to the subspace corrected algorithms described in [44, 45, 46, 47, 48, 49].

Before moving on to a description of the MAP techniques described above, some additional background information is given in the next two sections. Section 2 outlines the mathematical model that will be assumed throughout this work, and introduces some important notation. A description of some simple array perturbation models will also be given in this section. Section 3 then focuses on several DOA estimation and beamforming algorithms that are relevant to subsequent discussions, and that will be used for comparison purposes in simulations later in the paper. The exact MAP estimator and an asymptotically equivalent approximation based on subspace fitting are described next in Section 4. The application of the MAP approach to minimum
mean-square error beamforming is also presented, and some special cases are examined in detail. The paper concludes with a series of representative simulation examples in Section 5.

2. Mathematical Model

For the most part, the standard narrowband modeling assumptions that have been used since the original work of Schmidt [14] and Bienvenu and Kopp [16] will be adopted. The term “narrowband” here refers to the assumption that the envelopes of the (bandpass) signals received by the array do not change appreciably as the signal wavefronts propagate across the array aperture. Furthermore, the signals are assumed to emanate from point source emitters, the array elements and receivers are linear devices (superposition applies), and the noise is additive with known second order spatial statistics.

Consider an array of \( m \) sensors, having arbitrary positions and characteristics, that receives the waveforms of \( d \) narrowband point sources, where \( d < m \). At time \( t \), the output of each sensor is collected in the vector \( x(t) \), which is modeled as a superposition of the array’s response to all \( d \) signals:

\[
x(t) = \left[ a(\theta_1, \rho) \mid \cdots \mid a(\theta_d, \rho) \right] \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix} + n(t) \tag{1}
\]

\[
= A(\theta, \rho)s(t) + n(t) \tag{2}
\]

The columns of the \( m \times d \) matrix \( A \) are the so-called array steering vectors, denoted \( a(\theta_i, \rho) \), \( i = 1, \ldots, d \). These vectors describe the array response to a unit waveform with signal parameter(s) \( \theta_i \). The parameters in \( \theta_i \) are typically associated with the location of the source, and can include azimuth angle of arrival, elevation angle, range, as well as other quantities related to polarization and doppler. Though not necessary, it is assumed for simplicity in our discussion that \( \theta_i \) is a real-valued quantity equal to the DOA of signal \( i \). The components of the \( d \)-vector \( \theta \) are the DOAs of the model, whereas the vector \( \theta_0 \) represents their true values.

The above model also allows for a set of unknown perturbation or nuisance parameters, collected in the real \( n \)-vector \( \rho = [\rho_1, \ldots, \rho_n]^T \). As will be seen later, the elements
of $\rho$ can include both unstructured or structured parameters such as sensor gain, phase, position, mutual coupling, etc. Unlike $\theta$, which is assumed to be deterministic, $\rho$ will be modeled as a random vector drawn with a known Gaussian prior distribution. The mean of the distribution, denoted by $\rho_0$, corresponds to the nominal array response with no perturbation. The covariance is defined by

$$\mathbb{E}[(\rho - \rho_0)(\rho - \rho_0)^T] = \Omega,$$  \hspace{1cm} (3)

and is also assumed to be known and full rank. The matrix $\Omega$ could be determined, for example, using sample statistics from a number of independent, identical calibration experiments, or using tolerance data specified by the manufacturer of the sensors. Some examples of typical $\rho$-parameterizations are given in Section 2.1. The nominal array response is assumed to be unambiguous\(^2\); i.e., the matrix $[a(\theta_1, \rho_0), \ldots, a(\theta_m, \rho_0)]$ has full rank for any collection of distinct parameters $\theta_1, \ldots, \theta_m$.

The complex $d$-vector $s(t)$ is composed of the emitter waveforms received at time $t$, and the $m$-vector $n(t)$ accounts for additive measurement noise. Although the signal waveforms will be assumed to be white Gaussian random processes when deriving the exact MAP estimator, the properties of the proposed method have been analyzed under a less restrictive assumption in [37]. The covariance of the signals is assumed to satisfy

$$\mathbb{E}[s(t)s^*(s)] = P \delta_{t,s}$$ \hspace{1cm} (4)

$$\mathbb{E}[s(t)s^T(s)] = 0,$$ \hspace{1cm} (5)

where $\delta_{t,s}$ is the Kronecker delta. The covariance matrix $P$ is further assumed in this work to be positive definite (no perfectly coherent signals are present). Note that $\{\cdot\}^*$ denotes the complex conjugate transpose. The noise term, $n(t)$, is modeled as a stationary, complex Gaussian random process that is uncorrelated with the signals. The noise is assumed to be zero mean, circularly symmetric, as well as spatially and temporally white:

$$\mathbb{E}[n(t)n^*(s)] = \sigma^2 I \delta_{t,s}$$ \hspace{1cm} (6)

$$\mathbb{E}[n(t)n^T(s)] = 0.$$

\(^2\)More precisely, the array is assumed to be unambiguous for all perturbation parameters in an open neighborhood of $\rho_0$. 

8
Assuming $N$ samples are taken from the array, the following data matrix may be defined:

$$X = [x(t_1) \cdots x(t_N)] = A(\theta, \rho)S + N,$$

where $S$ and $N$ are defined similarly to $X$. Based on the measurements in $X$, the problem of interest is to estimate the DOAs and individual waveforms of the sources. The number of signals, $d$, is assumed to be known.

### 2.1. Array Error Parameterizations

Since we are interested in studying the combined effects of finite sample errors and modeling errors, the size of the perturbations relative to the number of available data samples plays a crucial role. The variances of the estimated DOA’s are known to be proportional to $1/N$ in the finite-sample-only case [22], whereas they are proportional to $\Omega$ in the model-error-only case [27, 28]. In the approximation introduced in Section 4.2, the relative contributions of the two error sources will be assumed to be of comparable magnitude, and the covariance matrix of the perturbation parameters will be expressed as

$$\Omega = \hat{\Omega}/N,$$

where $\hat{\Omega}$ is independent of $N$. An argument for the somewhat artificial assumption (9) is that if $\Omega = o(1/N)$, then the effect of the modeling errors can be neglected and the methods designed for finite sample errors only are optimal. On the other hand, if $\Omega^{-1} = o(1/N)$, the effect of the modeling errors dominates, thus rendering the methods designed solely for such errors optimal. Since the MAP approach presented herein is inherently more complicated than either of the methods that take only finite samples or modeling errors into account, the former should be avoided when one type of error dominates the other. A further assumption that will be made in Section 4 when deriving the approximate MAP estimator is that the deviation $\rho - \rho_0$ is relatively small (i.e., second order effects of the array perturbation will be ignored).

The reason for a random perturbation model as opposed to a deterministic one lies in the consideration of how one chooses to quantify the effects of the perturbation. In a given fixed scenario, of course, the presence of array errors will introduce a bias in the DF and signal estimates. Presumably, if one wanted to measure the magnitude of
this bias, it would simply be a matter of directly computing the limiting \((N \to \infty)\)
estimates \(\hat{\theta}\) and \(\hat{S}\), and then subtracting \(\theta_0\) and \(S\). This procedure would obviouslyhave to be repeated for every perturbation scenario considered, since the bias would bedifferent in each case. The advantage of using a random model is that one can obtain ameasure of the \textit{average} effect of the array errors on estimation performance, measurednow in terms of \textit{variance} rather than bias, without being forced to adopt a particularperturbation scenario (which may be no more representative than any other similarperturbation).

It is useful to examine the form of \(\rho\) and its distribution for several commonlyencountered perturbation models. Three such models are briefly described below.

\textbf{Gain and Phase Errors} – For arrays composed of nominally identical elements, acommon approach used to describe deviations in the array response attempts to modelthe non-uniform gain and phase effects of the receiver electronics behind each antennaelement. In this model, the nominal response is perturbed by an unknown complexdiagonal matrix:

\[
A(\theta, \rho) = GA(\theta),
\]
and

\[
\rho = \begin{bmatrix}
\text{Re}\{g\} \\
\text{Im}\{g\}
\end{bmatrix},
\]

where \(g = \text{diag}\{G\}\). The mean of the distribution for \(\rho\) in this case is given by\(\rho_0 = [e^T \ 0]^T\), where \(e\) is an \(m \times 1\) vector of ones. For simplicity, it is often assumed
that the covariance of \(\rho\) is given by \(\Omega = (\sigma_g^2/2)I\), which implies that the individualgain and phase errors are all mutually independent and identically distributed. The
effects of uncalibrated mutual coupling can be incorporated into the above model byassuming the matrix \(G\) has non-zero off-diagonal elements.

\textbf{Sensor Position Errors} – For this case, assume an arbitrary array of identicalunit-gain omni-directional sensors in the \(x-y\) plane with randomly perturbed sensor
locations. The perturbed array response can be written as

\[
a(\theta, \rho) = \begin{bmatrix}
\exp\left(j\frac{2\pi}{\lambda}\left[(x_1 + \bar{x}_1) \sin \theta + (y_1 + \bar{y}_1) \cos \theta\right]\right) \\
\vdots \\
\exp\left(j\frac{2\pi}{\lambda}\left[(x_m + \bar{x}_m) \sin \theta + (y_m + \bar{y}_m) \cos \theta\right]\right)
\end{bmatrix},
\]
where \((x_i, y_i)\) are the nominal coordinates of the \(i^{th}\) sensor and \((\bar{x}_i, \bar{y}_i)\) are the corresponding position errors. Defining

\[
\rho = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix},
\]

it is clear that \(\rho_0 = 0\) and that the structure of \(\Omega\) can be used to describe situations where the sensor location errors are related to one another. Such a situation for a towed-array application was presented in [37].

**Unstructured Errors** - In the models presented above, the array perturbation is based on physical insight into the structure of the array, assuming that one error source is dominant. In many applications, however, both of the effects described above (along with others, such as quantization effects, interpolation errors, etc.) are present simultaneously. In such cases, a model based on physical insight may be impractical and cumbersome. A pragmatic remedy to this situation is to assume a simple unstructured model for the perturbed array response, such as

\[
A(\theta, \rho) = A(\theta) + \hat{A},
\]

where

\[
\rho = \begin{bmatrix} \text{Re}\{\text{vec}(\hat{A})\} \\ \text{Im}\{\text{vec}(\hat{A})\} \end{bmatrix},
\]

and, again, \(\rho_0 = 0\). Let the columns of \(\hat{A}\), denoted \(\tilde{a}_i\), be modeled as zero mean Gaussian random vectors with moments

\[
E[\tilde{a}_i\tilde{a}_k^T] = \nu_{ik} I \quad i, k = 1, \ldots, d
\]

\[
E[\tilde{a}_i\tilde{a}_k^T] = 0 \quad i, k = 1, \ldots, d.
\]

The above model corresponds to an additive, circularly-symmetric, complex array perturbation that is uncorrelated from sensor to sensor, but possibly \(\theta\)-dependent. It is easy to verify that under these assumptions, the covariance of \(\rho\) is given by

\[
\Omega = \frac{1}{2} \begin{bmatrix} \text{Re}\{\Upsilon\} \otimes I & -\text{Im}\{\Upsilon\} \otimes I \\ \text{Im}\{\Upsilon\} \otimes I & \text{Re}\{\Upsilon\} \otimes I \end{bmatrix},
\]

where the \(ik\)th element of the matrix \(\Upsilon\) is \(\nu_{ik}\). Perturbation models similar to (13)-(16) have been used by a number of others, primarily in the analysis of adaptive beamforming algorithms [50, 51, 52].
Although not realistic for all types of array perturbations, the above model is useful for purposes of algorithm analysis and comparison, and is reasonable for situations involving experimentally calibrated arrays, where the sources of error are often due to quantization effects in collecting the calibration data, interpolation errors in using a calibration grid, etc. The model of (13)-(16) is also related to the more realistic case considered in (10)-(11) involving gain and phase perturbations. To see this, suppose \( \mathbf{Y} = \sigma_a^2 \mathbf{I} \), the nominal response of the \( k \)th sensor is \( a_k(\theta) = e^{i\phi_k} \), and let \( \tilde{g}_k \) and \( \tilde{\phi}_k \) represent the corresponding gain and phase perturbations. The \( k \)th diagonal element of \( \mathbf{G} \) in (10) is thus \( G_k = (1 + \tilde{g}_k)e^{j\tilde{\phi}_k} \), and

\[
\hat{a}_k(\theta) = (1 + \tilde{g}_k)e^{j(\phi_k + \tilde{\phi}_k)} \approx a_k(\theta) + \tilde{a}_k(\theta)
\]

when \(|\tilde{a}_k(\theta)| \ll |a_k(\theta)|\), where

\[
\tilde{a}_k(\theta) = (\tilde{g}_k + j\tilde{\phi}_k)e^{j\tilde{\phi}_k}.
\]

If we let \( \sigma_g^2 \) and \( \sigma_\phi^2 \) represent the variances of \( \tilde{g}_k \) and \( \tilde{\phi}_k \), respectively, and if we assume that the gain and phase perturbations are zero-mean and independent, then

\[
E[\tilde{a}_k \tilde{a}_k^*] = \sigma_g^2 + \sigma_\phi^2
\]

\[
E[\tilde{a}_k^2] = (\sigma_g^2 - \sigma_\phi^2)e^{j2\phi}.
\]

The model of (13)-(16) results if the gain and phase errors are roughly “of the same order” \( (\sigma_g^2 \text{ and } \sigma_\phi^2 \text{ are equal}) \), with \( \sigma_a^2 = \sigma_g^2 + \sigma_\phi^2 \). Thus, while the model of (10)-(11) assumes angle independent gain and phase errors, the unstructured model in (13)-(16) can be thought of as representing gain and phase errors that are uncorrelated but angle dependent.

2.2. Subspace Notation

In Section 4, the exact MAP estimator is presented along with a less computationally demanding approximation. The proposed approximate technique is a subspace-based method, in that it relies heavily on the properties of the eigendecomposition of the array covariance. Under the above assumptions, the covariance matrix of the array output takes the form

\[
\mathbf{R} = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbf{E}[\mathbf{x}(t)\mathbf{x}^*(t)] = \mathbf{A}(\theta_0, \rho_0)\mathbf{P}\mathbf{A}^*(\theta_0, \rho_0) + \sigma^2 \mathbf{I}.
\]

(18)
Since the matrix $APA^*$ has rank $d$ by assumption (the arguments of $A$ will frequently be suppressed for notational convenience), $\sigma^2$ is an eigenvalue of $R$ with multiplicity $m - d$, and the corresponding eigenvectors are all orthogonal to $A$. The eigendecomposition of $R$ thus takes the form

$$R = \sum_{i=1}^{m} \lambda_i e_i e_i^* = E_s \Lambda_s E_s^* + \sigma^2 E_n E_n^*, \quad (19)$$

where $\Lambda_s$ is a diagonal matrix containing the $d$ largest eigenvalues, and the columns of the $m \times d$ matrix $E_s$ are the corresponding unit-norm eigenvectors. Similarly, the columns of $E_n$ are the $m - d$ eigenvectors corresponding to $\sigma^2$. Since $E_n$ is orthogonal to $A$, it follows that the range space of $E_s$ coincides with that of $A$. This observation forms the basis for all subspace-based estimation techniques, starting with the development of the popular MUSIC algorithm [14, 16].

Assuming orthonormal eigenvectors, the orthogonal projection onto the range space of $A$ is denoted

$$\Pi = A(A^*A)^{-1}A^* = AA^\dagger = E_s E_s^*, \quad (20)$$

and its orthogonal complement is

$$\Pi^\perp = I - AA^\dagger = E_n E_n^*. \quad (21)$$

Here, the Moore-Penrose pseudo-inverse of $A$ is denoted

$$A^\dagger = (A^*A)^{-1}A^*. \quad (22)$$

Under the stated assumptions, the eigendecomposition of $R$ and the above projection matrices can be consistently estimated by performing an eigendecomposition of the sample covariance matrix

$$\hat{R} = \frac{1}{N} \sum_{t=1}^{N} x(t)x^*(t) = \hat{E}_s \hat{\Lambda}_s \hat{E}_s^* + \hat{E}_n \hat{\Lambda}_n \hat{E}_n^*, \quad (23)$$

where the partitioning of the eigen-elements is similar to (19).

### 3. Relevant Algorithms

In this section, a description is given of several algorithms for DOA estimation and beamforming that are referred to later. Although the approaches described here have
not been specifically developed to handle errors in the array response, an indication will be given of how they may be made more robust to such errors. Consequently, for these algorithms the array response will be denoted using \( A(\theta) \) rather than the more general expression \( A(\theta, \rho) \). In all cases, \( A(\theta) \) is to be taken to mean the nominal response \( A(\theta, \rho_0) \).

### 3.1. DOA Estimation

As mentioned earlier, most well-known DOA estimation algorithms have been shown to be special cases of the so-called (signal) subspace fitting (SSF) framework first presented in [22]. All algorithms in the SSF class can be shown to be equivalent (either algebraically or asymptotically in \( N \)) to the following minimization problem:

\[
\hat{\theta} = \arg \min_{\theta} V_{ssf}(\theta) = \arg \min_{\theta} \text{Tr} \left( \Pi^\perp(\theta) \tilde{E}_n W \tilde{E}_s^* \right),
\]

(24)

where the choice of the weighting matrix \( W \) determines the particular algorithm in question. The rationale behind the SSF formulation is that the range space of the ideal signal subspace matrix \( E_s \) is contained in that of the true \( A \). Making the projection of \( \tilde{E}_s \) onto the orthogonal complement of \( \text{span}A(\theta) \) small (in an appropriate metric), should therefore lead to good estimate of \( \theta \). An equivalent noise subspace fitting (NSF) formulation can also be made [53], in which algorithms are shown to be special cases of

\[
\hat{\theta} = \arg \min_{\theta} V_{nssf}(\theta) = \arg \min_{\theta} \text{Tr} \left( A^*(\theta) \tilde{E}_n \tilde{E}_n^* A(\theta) U \right),
\]

(25)

for certain choices of the weighting \( U \). The argument behind this formulation is that the true \( A \) is orthogonal to \( E_n \). The by far most well-known subspace-based algorithm is obtained by choosing \( U \) diagonal. The multidimensional search involved in (25) then decouples into \( d \) one-dimensional minimizations of the form

\[
\hat{\theta}_k = \arg \min_{\theta} a^*(\theta) \tilde{E}_n \tilde{E}_n^* a(\theta).
\]

(26)

The locations of the \( d \) deepest (separated) minima of the criterion function in (26) define the MUSIC estimates [14].

One of the key results of [22] and [53, 54] was the derivation of optimal values for both \( W \) and \( U \). The term optimal as used here means that, for a certain choice of \( W \) and \( U \), the SSF and NSF methods are large sample realizations of the maximum
likelihood approach, and hence will yield DOA estimates of minimum variance. In particular, the optimal weightings were shown to be

\[
W_{OPT} = \left( \hat{A}_s - \delta^2 I \right)^2 \hat{A}_s^{-1}
\]

\[
U_{OPT} = A^\dagger(\hat{\theta}_0) \hat{E}_s W_{OPT} \hat{E}_s^* A^\dagger(\hat{\theta}_0),
\]

where \( \delta^2 \) and \( \hat{\theta}_0 \) are consistent estimates of the noise power and DOAs, respectively. An excellent side-by-side derivation of the optimality of both the SSF and NSF methods can be found in [31].

The optimal weight matrices given in (27)-(28) were derived assuming that only the finite sample effects of noise and not those due to array model errors are present (i.e., \( \rho = \rho_0 \)). A different set of weights results if model errors are accounted for and finite sample effects ignored [28]. More recently, it has been shown that for the more realistic case where both sources of error are present, the optimal weighting becomes a combination of the two weights obtained for the limiting cases [29]. It should be noted that the optimality of the approaches in [28, 29] has only been established for the simple error model of (13)-(16).

3.2. Signal Estimation

As used here, signal estimation refers to the problem of estimating the samples of the signal waveforms \( S \) using the received data \( X \) from the array. This is most often accomplished by forming a linear combination of the array outputs (a spatial filter), and is often referred to as beamforming (although technically this is somewhat of a misnomer). The coefficients used in forming the linear combination are typically called the beamformer weights. The weights that minimize the mean squared error (MSE) of the signal estimate are simply the coefficients of a spatial Wiener filter:

\[
W_{MSE} = \arg\min_W \|W^*X - S\|_F^2
\]

\[
= R^{-1}R_{xs} = R^{-1}A(\theta, \rho)P,
\]

where

\[
R_{xs} = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} x(t)s^*(t) = A(\theta, \rho)P.
\]
When the desired signal is uncorrelated with the interference, $P$ is diagonal and the minimum MSE solution is just a scaled version of the so-called minimum variance distortionless response (MVDR) beamformer [55]:

$$W = \frac{R^{-1}a(\hat{\theta})}{a^*(\hat{\theta})R^{-1}a(\hat{\theta})}. \quad (31)$$

In the general case where the signal and interference are correlated, the optimal MSE weights depend on the signals themselves through $R_x$, or $P$, and thus they cannot be used directly (i.e., without a training sequence, for example). As a result, many techniques have been devised to approximate the Wiener solution. In the approach of [56], it is assumed that the DOAs of the signals have been estimated, and the quantities $P$ and $R$ in (30) are replaced by their structured maximum likelihood (ML) estimates:

$$\hat{P}_s = A^\dagger(\hat{\theta}, \rho_0)(\hat{R} - \hat{\sigma}^2I)A^*(\hat{\theta}, \rho_0)$$

$$\hat{R}_s = A(\hat{\theta}, \rho_0)\hat{P}_sA^*(\hat{\theta}, \rho_0) + \hat{\sigma}^2I,$$

where $\hat{R}$ is a sample estimate of $R$ and $\hat{\theta}$ is an ML estimate of the DOAs. Since calibration errors were not addressed in [56], the nominal model $A(\hat{\theta}, \rho_0)$ was used in (30) to calculate the beamformer weights. Nevertheless, the method performs well when calibration errors are present, as recently demonstrated in [57].

On the other hand, the MVDR approach is well known to be hyper-sensitive to array perturbations, especially at high SNR. While ad hoc methods employing artificial noise injection have been used to combat this problem, other techniques based on subspace corrected (SC) weights have found success in experimental systems [45, 47]. In these approaches, the $R^{-1}$ term in (31) is replaced by $E_sA_s^{-1}E_s^*$. This is equivalent to projecting $a(\hat{\theta})$ onto the signal subspace prior to forming the MVDR weights.

In [58], the improvement that results from using the method of [56] with $A(\hat{\theta}, \hat{\rho})$ rather than $A(\hat{\theta}, \rho_0)$ was investigated, where $\hat{\rho}$ is obtained from the MAP estimator described in the next section. The results of this analysis will be briefly described below, and the resulting algorithm will be referred to as the MAP beamformer. An interesting connection can be made between the MAP beamformer and the SC-MVDR method. In particular, it will be shown below that for uncorrelated signals and simple unstructured array errors of the form (13)-(16), the SC-MVDR and MAP weights have a very similar form.
4. MAP Parameter Estimation

In this section, the exact MAP formulation of the problem is presented, along with a simplified, but asymptotically equivalent approximation. Unlike the exact MAP approach, which in general requires nonlinear optimization over both the DOAs $\theta$ and the nuisance parameters $\rho$, the approximate MAP method is separable in $\rho$. The resulting criterion requires a search only over the $d$ DOA parameters in $\theta$, which is a considerable simplification since the number of elements in $\rho$ can be quite large. The MAP estimate of $\rho$ is calculated directly given $\hat{\theta}$, and can be used to perform an on-line calibration of the array.

4.1. Exact MAP Estimation

As mentioned earlier, the derivation of the MAP estimator given below assumes that the signals, noise, and array perturbation parameters are all Gaussian. The signals and noise are zero-mean with covariances $P$ and $\sigma^2 I$, respectively, where $P$ and $\sigma^2$ are unknown deterministic parameters (i.e., parameters with a non-informative prior distribution) and must be estimated. On the other hand, the mean $\rho_0$ and covariance $\Omega$ of $\rho$ are assumed to be known. Besides $P$ and $\sigma^2$, the other unknown parameters to be estimated are of course $\theta$ (deterministic) and $\rho$ (random).

For the moment, let $\eta = \{\theta, P, \sigma^2\}$ represent the deterministic parameters to be estimated. As its name implies, the MAP estimator maximizes the probability density of the desired parameters given the received data:

$$\{\eta, \rho\}_{MAP} = \arg\max_{\eta, \rho} p_{\eta, \rho}(\eta, \rho|X).$$

Using Bayes’ rule with $\rho$ and eliminating the prior on $X$ yields the equivalent formulation

$$\{\eta, \rho\}_{MAP} = \arg\max_{\eta, \rho} p_{\eta, \rho}(X|\eta, \rho)p_{\rho}(\rho).$$

Since the densities in (33) are all Gaussian, it is simpler to work with their negative logarithms, in which case the MAP estimates are obtained by the following minimization problem:

$$\{\eta, \rho\}_{MAP} = \arg\min_{\eta, \rho} \{-\ln p_{\eta, \rho}(X|\eta, \rho) - \ln p_{\rho}(\rho)\}.$$
The $-\ln p_{\theta, \rho}(X|\eta, \rho)$ term in (34) is the standard negative log-likelihood function $V_{ML}(\theta, \rho, \mathbf{P}, \sigma^2)$, which is given by [59]

$$V_{ML}(\theta, \rho, \mathbf{P}, \sigma^2) = N \left( \log |\mathbf{R}(\theta, \rho, \mathbf{P}, \sigma^2)| + \text{Tr}\{\mathbf{R}^{-1}(\theta, \rho, \mathbf{P}, \sigma^2) \mathbf{\hat{R}}\} \right) + \text{const},$$  \hspace{1cm} (35)

where $|\cdot|$ denotes the determinant. The other term involving the prior of $\rho$ is easily shown to be

$$-\ln p_{\rho}(\rho) = \frac{1}{2}(\rho - \rho_0)^T \Omega^{-1}(\rho - \rho_0).$$  \hspace{1cm} (36)

Consequently, the joint MAP estimate of $\theta, \mathbf{P}, \sigma^2$ and $\rho$ may be expressed as

$$\{\mathbf{\hat{\theta}}, \mathbf{\hat{\rho}}, \mathbf{\hat{P}}, \hat{\sigma}^2\}_{MAP} = \arg \min_{\theta, \rho, \mathbf{P}, \sigma^2} V_{MAP}(\theta, \rho, \mathbf{P}, \sigma^2)$$  \hspace{1cm} (37)

$$V_{MAP}(\theta, \rho, \mathbf{P}, \sigma^2) = V_{ML}(\theta, \rho, \mathbf{P}, \sigma^2) + \frac{1}{2}(\rho - \rho_0)^T \Omega^{-1}(\rho - \rho_0).$$  \hspace{1cm} (38)

The ML criterion function is known to be separable in $\mathbf{P}$ and $\sigma^2$. For fixed $\mathbf{A} = \mathbf{A}(\theta, \rho)$, the minimizing signal covariance matrix and noise power are [59]

$$\mathbf{\hat{P}} = \mathbf{A}^* (\mathbf{\hat{R}} - \hat{\sigma}^2 \mathbf{I}) \mathbf{A}^T$$  \hspace{1cm} (39)

$$\hat{\sigma}^2 = \frac{1}{m-d} \text{Tr}\{\mathbf{\Pi}^* \mathbf{\hat{R}}\}.$$  \hspace{1cm} (40)

Substituting (39)-(40) into (35) leads to [31]

$$V_{ML}(\theta, \rho) = N \left| \mathbf{A} \mathbf{\hat{P}} \mathbf{A}^* + \hat{\sigma}^2 \mathbf{I} \right| + \text{const}.$$  \hspace{1cm} (41)

Clearly, $V_{MAP}(\theta, \rho, \mathbf{P}, \sigma^2)$ is also separable in $\mathbf{P}$ and $\sigma^2$, and ignoring constant terms the concentrated MAP criterion function is

$$V_{MAP}(\theta, \rho) = V_{ML}(\theta, \rho) + \frac{1}{2}(\rho - \rho_0)^T \Omega^{-1}(\rho - \rho_0).$$  \hspace{1cm} (42)

This can be interpreted as a regularized ML criterion. That is, the effect of the prior distribution is to force $\mathbf{\hat{P}}_{MAP}$ to be close to the nominal value, $\rho_0$. If the perturbation parameters are identifiable, this effect is diminished as the number of snapshots, $N$, increases. Thus, the MAP estimate has the same asymptotic properties as the ML estimate (i.e., the pure auto-calibration technique). However, in many applications of interest, $\rho$ cannot be consistently estimated along with the signal parameters. In such cases the prior distribution has a crucial influence on the asymptotic properties of the estimates of both $\theta$ and $\rho$. 

18
Even in the concentrated form of (42), the MAP criterion is still quite unwieldy due to the complicated dependence of $V_{ML}$ on $\rho$, and the fact that $\rho$ can have a large number of elements ($2md$ for the model of (13)-(16)). In the next section, we show how $V_{ML}(\theta, \rho)$ can be replaced with an asymptotic approximation that is quadratic in $\rho$. The resulting approximate MAP criterion will then also be quadratic (and hence separable) in $\rho$, and can be further concentrated to depend only on the $d$ elements of $\theta$.

4.2. The MAP-NSF Method

As pointed out earlier, when the signal covariance has full rank and no array errors are present, the noise subspace fitting (NSF) cost function

$$ V_{NSF} = N \text{Tr}\{A^*\hat{E}_n^*\hat{E}_n^*AU_{OPT}\}, \quad (43) $$

is known to be asymptotically equivalent to the ML criterion when $U_{OPT}$ is chosen as in (28). In the present case, the term “asymptotic equivalence” is to be interpreted as

$$ \frac{\partial V_{MAP}}{\partial \alpha} \bigg|_{\theta_0, \rho_0} = \frac{\partial V_{NSF}}{\partial \alpha} \bigg|_{\theta_0, \rho_0} + o_p(1/\sqrt{N}), \quad (44) $$

where $\alpha$ refers to any component of $\theta$ or $\rho$, and the symbol $o_p(1/\sqrt{N})$ represents a term that tends to zero faster than $1/\sqrt{N}$ in probability. The extension of this result to the case where model errors are also present is immediate since the proof only depends on the fact that $\hat{R} = R + O_p(1/\sqrt{N})$. By standard first order arguments, this implies that the MAP estimate is asymptotically (for large $N$) equivalent to the minimizing arguments of the following criterion function

$$ V_{NSF}(\theta, \rho) + \frac{1}{2}(\rho - \rho_0)^T\Omega^{-1}(\rho - \rho_0). \quad (45) $$

This criterion depends on its parameters in a simpler way than the exact MAP criterion (42). However, it still requires a non-linear minimization over both $\theta$ and $\rho$.

A further simplification of the criterion is possible that enables separation with respect to $\rho$. Recall the following formulas for the vec($\cdot$) operator (vectorization of a matrix by stacking its columns) and the Kronecker product $\otimes$ (see [60, 61]):

$$ \text{Tr}\{ABCD\} = \text{vec}(D^T)(C^T \otimes A)\text{vec}(B) \quad (46) $$
\[
\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B) \tag{47}
\]
\[
(A \otimes B)(C \otimes D) = ((AC) \otimes (BD)) \tag{48}
\]
\[
(A \otimes B)^T = A^T \otimes B^T. \tag{49}
\]

Using (46), the NSF criterion can be rewritten as

\[
N \text{Tr}\{A^*\hat{E}_n\hat{E}_n^*AU_{o,rr}\} = Na^*\hat{M}a, \tag{50}
\]

where

\[
a = \text{vec}(A) \tag{51}
\]
\[
\hat{M} = U_{o,rr}^T \otimes (\hat{E}_n\hat{E}_n^*). \tag{52}
\]

If the perturbation to the array is small enough so that second order effects may be ignored, the vectorized steering matrix may be approximated locally around \(\rho_0\) as

\[
a = a(\theta, \rho) \approx a_0 + D\rho \hat{\rho}, \tag{53}
\]

where

\[
a_0 = a(\theta, \rho_0) \tag{54}
\]
\[
D\rho = \left[ \frac{\partial a(\theta, \rho)}{\partial \rho_1}, \ldots, \frac{\partial a(\theta, \rho)}{\partial \rho_n} \right]_{\theta, \rho_0} \tag{55}
\]
\[
\hat{\rho} = \rho - \rho_0. \tag{56}
\]

Note that, when evaluated at \(\rho_0\), the derivative of \(a\) with respect to \(\theta\) or \(\rho\) is identical to that of \(a_0 + D\rho \hat{\rho}\). It follows that the minimizing arguments of (50) are asymptotically identical to the estimates obtained by minimizing the following approximate MAP-NSF criterion with respect to \(\theta\) and \(\rho\):

\[
(a_0 + D\rho \hat{\rho})^*\hat{M}(a_0 + D\rho \hat{\rho}) + \frac{1}{2}\rho^T\hat{\Omega}^{-1}\hat{\rho}, \tag{57}
\]

where we have normalized by \(N\) and used (9).

Since the criterion function in (57) is quadratic in \(\hat{\rho}\), we easily obtain the minimum with respect to \(\hat{\rho}\) (for fixed \(\theta\)) as

\[
\hat{\rho}_{\text{MAP-NSF}} = \rho_0 - \Gamma^{-1}f, \tag{58}
\]

20
where
\[ \Gamma = \text{Re} \left\{ D^* \hat{M} D \rho + \frac{1}{2} \hat{\Omega}^{-1} \right\} \]
\[ f = \text{Re}\{D^* \hat{M} a_0\} . \]

Substituting (58) into (57) leads to the following separated criterion function
\[ a_0^* \hat{M} a_0 - f^T \Gamma^{-1} f . \]

Note that \( \Gamma \) and \( f \) depend on \( \theta \) through \( D \rho \), and in principle \( M \) also depends on \( \theta \) through \( U_{o,rr} \). However, it will be assumed that a consistent estimate of \( \theta \) is available to form the estimates \( \hat{M}, \hat{f} \) and \( \hat{\Gamma} \). Under the stated assumptions, such an estimate can be obtained, for instance, by letting \( \rho = \rho_0 \) and \( U \) diagonal in (25), which leads to the MUSIC algorithm (26). As shown in [37], the approximations made in forming \( \hat{\Gamma}, \hat{f} \) and \( \hat{M} \) do not change the asymptotic properties of the final estimate.

The definitions of the quantities in the MAP-NSF cost function are repeated below for easy reference, followed by a summary of the proposed algorithm:

\[ a_0 = \text{vec}(A(\theta, \rho_0)) \]
\[ \hat{M} = \hat{\sigma}^{-2}(\hat{A}^H \hat{E}_s (\hat{A} - \hat{\sigma}^2 I)^2 \hat{A}_s^{-1} \hat{E}_s^* \hat{A}_s^*)^T \otimes (\hat{E}_n \hat{E}_n^*) \]
\[ \hat{\sigma}^2 = \frac{1}{m - d} \text{Tr}\{(I - \hat{A} \hat{A}^H) \hat{R}\} \]
\[ \hat{\Gamma} = \text{Re}\left\{ D^* \hat{M} D \rho + \frac{1}{2} \hat{\Omega}^{-1} \right\} \]
\[ \hat{f} = \text{Re}\{D^* \hat{M} a_0\} . \]

The MAP-NSF Algorithm: Given the sample covariance \( \hat{R} \) and an initial estimate \( \hat{\theta} \) of the DOAs:

1. Compute the eigendecomposition \( \hat{R} = \hat{E}_s \hat{\Lambda}_s \hat{E}_s^* + \hat{E}_n \hat{\Lambda}_n \hat{E}_n^* \).
2. Compute the quantities (63)-(65)
3. Using \( \hat{\theta} \) as an initial guess, use a numerical method to solve the following optimization problem
\[ \hat{\theta}_{\text{MAP-NSF}} = \arg \min_{\theta} V(\theta) \]
\[ V(\theta) = a_0^* \hat{M} a_0 - \hat{f}^T \hat{\Gamma}^{-1} \hat{f} . \]
Let us finally point out that the MAP-NSF algorithm yields statistically efficient estimates of the signal and perturbation parameters, provided that no perfectly coherent signals are present. In [37] it was shown that the estimated parameters asymptotically achieve the Cramér-Rao lower bound for mixed deterministic and stochastic parameters. More details along with an expression for the asymptotic bound are provided in this reference. Although the MAP-NSF approach loses claim to asymptotic optimality only when the signals are 100% correlated, a performance degradation is evident at lower correlation levels, as will be demonstrated by several simulations in Section 5. In such cases, relatively large values of $N$ are required for MAP-NSF to achieve the CRB, and implementing the algorithm with smaller $N$ can lead to a loss of signal resolution. However, as will also be demonstrated by the simulations, there are a wide variety of scenarios for which the MAP-NSF approach achieves the CRB for relatively small values of $N$.

### 4.3. Some Special Cases

The specific form taken on by the MAP-NSF criterion function depends of course on the parameterization chosen for $\mathbf{p}$. In the discussion that follows, the MAP-NSF criterion is examined for two of the three basic error models described in Section 2.1. The unstructured model is studied first since the resulting equations are simplest, and there are some interesting connections to be made with some of the algorithms described in Section 3. The case of independent gain and phase errors is also addressed.

**Unstructured Errors** – To begin with, note that for the model described by (13)-(16), $\rho_0 = 0$ and $\bar{D}_\rho = [I \ jI]$, where $I$ is $md \times md$. Thus, $\hat{\rho} = -\Gamma^{-1} f$, and

$$
\Gamma = \begin{bmatrix}
\text{Re}(\hat{M} + \frac{1}{N} \mathbf{Y}^{-1} \otimes I) - \text{Im}(\hat{M} + \frac{1}{N} \mathbf{Y}^{-1} \otimes I) \\
\text{Im}(\hat{M} + \frac{1}{N} \mathbf{Y}^{-1} \otimes I) & \text{Re}(\hat{M} + \frac{1}{N} \mathbf{Y}^{-1} \otimes I)
\end{bmatrix}
$$

$$
\mathbf{f} = \begin{bmatrix}
\text{Re}(\hat{M}a_0) \\
\text{Im}(\hat{M}a_0)
\end{bmatrix}.
$$

Using the fact that, for any invertible matrix $\mathbf{Z}$,

$$
\begin{bmatrix}
\text{Re}(\mathbf{Z}) - \text{Im}(\mathbf{Z}) \\
\text{Im}(\mathbf{Z}) \quad \text{Re}(\mathbf{Z})
\end{bmatrix}^{-1} = 
\begin{bmatrix}
\text{Re}(\mathbf{Z}^{-1}) - \text{Im}(\mathbf{Z}^{-1}) \\
\text{Im}(\mathbf{Z}^{-1}) \quad \text{Re}(\mathbf{Z}^{-1})
\end{bmatrix}, \quad (69)
$$

22
it is easy to show that

\[
\hat{\rho} = -\begin{bmatrix}
\text{Re} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{Y}^{-1} \otimes \mathbf{I} \right)^{-1} \hat{\mathbf{M}} \mathbf{a}_0 \right\} \\
\text{Im} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{Y}^{-1} \otimes \mathbf{I} \right)^{-1} \hat{\mathbf{M}} \mathbf{a}_0 \right\}
\end{bmatrix}.
\] (70)

A further simplification of (70) is possible that is quite revealing. Using the definition of \( \hat{\mathbf{M}} \) in (63), note that

\[
\left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{Y}^{-1} \otimes \mathbf{I} \right)^{-1} = \left( \mathbf{U}_{o_{ps}}^T + \frac{1}{N} \mathbf{Y}^{-1} \right) \otimes \left( \hat{\mathbf{E}}_n^n \hat{\mathbf{E}}_n^n + \frac{1}{N} \mathbf{Y}^{-1} \right) \otimes (\hat{\mathbf{E}}_s^n \hat{\mathbf{E}}_s^n)^{-1}
\] (71)

\[
= \left( \mathbf{U}_{o_{ps}}^T + \frac{1}{N} \mathbf{Y}^{-1} \right)^{-1} \otimes (\hat{\mathbf{E}}_n^n \hat{\mathbf{E}}_n^n) + N \mathbf{Y} \otimes (\hat{\mathbf{E}}_s^n \hat{\mathbf{E}}_s^n). \]
\] (72)

Multiplying the last equation above on the right by \( \hat{\mathbf{M}} \mathbf{a}_0 \) and simplifying then yields

\[
\hat{\rho} = -\begin{bmatrix}
\text{Re} \left\{ \left[ \left( \mathbf{I} + \frac{1}{N} \mathbf{Y} \mathbf{U}_{o_{ps}}^T \right)^{-1} \right] \otimes (\hat{\mathbf{E}}_n^n \hat{\mathbf{E}}_n^n) \right\} \mathbf{a}_0 \\
\text{Im} \left\{ \left[ \left( \mathbf{I} + \frac{1}{N} \mathbf{Y} \mathbf{U}_{o_{ps}}^T \right)^{-1} \right] \otimes (\hat{\mathbf{E}}_n^n \hat{\mathbf{E}}_n^n) \right\} \mathbf{a}_0
\end{bmatrix}.
\] (73)

Finally, using (13)-(14) and properties of the Kronecker product, the MAP estimate of the array response becomes

\[
\mathbf{A}(\hat{\mathbf{\theta}}, \hat{\rho}) = \mathbf{A}(\hat{\mathbf{\theta}}) - \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A}(\hat{\mathbf{\theta}}) \left( \mathbf{I} + \frac{1}{N} \left( \mathbf{Y} \mathbf{U}_{o_{ps}}^T \right)^{-1} \right)^{-1}.
\] (74)

The key point of interest is that, if \( \mathbf{Y}^{-1}/N \to 0 \), then the MAP estimate of the array response converges to a subspace corrected version of the nominal response:

\[
\lim_{\mathbf{Y}^{-1}/N \to 0} \mathbf{A}(\hat{\mathbf{\theta}}, \hat{\rho}) = \hat{\mathbf{E}}_s \hat{\mathbf{E}}_s^* \mathbf{A}(\hat{\mathbf{\theta}}).
\]

Furthermore, if the estimated MAP array response is used in (31), the MVDR beamformer (31) will converge to the SC-MVDR approach. The condition \( \mathbf{Y}^{-1}/N \to 0 \) occurs either with a large data sample, or when the array perturbation is large. In either case, the information provided by the prior distribution of \( \mathbf{\rho} \) is of little value, and is essentially ignored by the MAP criterion. This observation provides some theoretical justification for the SC-MVDR technique, which previously had been derived using ad hoc (but well motivated) reasoning. However, in cases where the prior cannot be neglected, using SC response vectors for beamforming will not be optimal and significant degradation can result. This is seen in the simulation examples described in the next section.
In [37], it was shown that, for the unstructured array error model of (13)-(14), the concentrated MAP-NSF criterion

\[ V(\theta) = a_0^* \hat{M} a_0 - \hat{f}^T \hat{f}^{-1} \hat{f} \]

is asymptotically equivalent to the SSF algorithm

\[ \hat{\theta} = \arg \min_{\theta} V_{SSF}(\theta) = \arg \min_{\theta} \text{Tr} \left( \Pi^T(\theta) \hat{E}_s \hat{W} \hat{E}_s^* \right) \]

for a particular choice of \( \hat{W} \). The optimal weighting was shown to be

\[ \hat{W}_{OSF} = \left( \hat{E}_s^* A(\hat{\theta})^T Y^T A(\hat{\theta}) \hat{E}_s + \frac{\hat{\sigma}^2}{N} (\hat{\Lambda}_s - \hat{\sigma}^2 I)^{-2} \hat{\Lambda}_s \right)^{-1}, \]

which is precisely the SSF weighting derived in [29] for the combined case where both the unstructured array errors and the finite sample effects of the noise are taken into account. For this particular model, the SSF solution is preferred over the MAP-NSF criterion since SSF can handle coherent emitters (rank deficient \( P \)) and has been observed to have better numerical properties.

**Gain and Phase Errors** – The derivation of the MAP estimate of \( \rho \) and hence \( g \) for the model in (10)-(11) is straightforward but somewhat cumbersome, and thus will not be presented here. However, assuming \( \Omega = (\sigma_n^2/2)I \), the result is quite simple and is given by

\[ \hat{g} = (I + \sigma_n^2 N \hat{Z}(\hat{\theta}))^{-1} e, \]

where

\[ \hat{Z}(\theta) = \left[ \sum_{i,k=1}^{d} u_{ki} \overline{\varpi(\theta_i)} a^T(\theta_k) \right] \odot (\hat{E}_n \hat{E}_n^*), \]

\( u_{ki} \) is the \( ki \)th element of \( \hat{U}_{\text{OPT}} \), \( \overline{\cdot} \) denotes conjugation, and \( \odot \) an element-wise (Hadamard) product. Note that for very small gain/phase errors where \( \sigma_n \to 0 \), \( \hat{g} \to e \) and hence \( \hat{G} \to I \) as expected.

To find a simpler expression for the concentrated MAP-NSF cost function, note that the NSF criterion may be written as

\[ \text{Tr} \left\{ A^*(\theta) \hat{E}_n \hat{E}_n^* A(\theta) U_{\text{OPT}} \right\} = e^T \hat{Z}(\theta) e. \]

Some simple algebra then yields

\[ \hat{\theta}_{\text{MAP-NSF}} = \arg \min_{\theta} V(\theta) = \arg \min_{\theta} e^T \left( \hat{Z}^{-1}(\theta) + \sigma_n^2 N I \right)^{-1} e. \]
As $\sigma_a \to 0$, the standard NSF criterion of (25) is clearly retrieved. Although the large $N$ behavior of (80) seems suspect, recall that $\sigma_a^2$ is assumed to be $O(1/N)$, and hence $\sigma_a^2 N = O(1)$. Again, as mentioned earlier, if finite sample effects are negligible compared to the array errors (as they would be for fixed $\sigma_a^2$ if $N \to \infty$), a different estimator should be used.

5. Simulation Results

In this section, the performance of the MAP-NSF approach is studied by means of a number of simulation examples. Additional simulation results regarding parameter estimation accuracy can be found in [29, 37]. The examples are divided into two parts, one involving DOA estimation performance (Section 5.1), and the other focusing on beamforming results (Section 5.2). All of the examples in this section assume a nominally unit-gain uniform linear array with half-wavelength interelement spacing.

5.1. DOA Estimation Results

The advantages provided by the robust MAP-based techniques developed in this work are most evident in difficult situations involving weak signals in the presence of strong interferers. In the three examples below, the output of a 10 element array was simulated using both a 0 dB SNR signal arriving from broadside and a 20 dB interferer. For the first two cases, the DOA of the interferer was 7°, and the array response was perturbed by an unstructured calibration error in the form of equation (13)-(17) with $\Upsilon = \sigma_a^2 I$ and $\sigma_a = 0.05$. Figure 1 shows the RMS error of several DOA estimation algorithms based on 1000 trials for various sample sizes $N$, assuming the two signals have a correlation coefficient of -0.9. Figure 2 shows the performance of the same algorithms, this time with $N$ fixed at 100 and the correlation varying from 0 to near 100% (the correlation phase was taken to be 180° relative to the first sensor in all of these simulations).

Since unstructured array errors were assumed in this example, the MAP-NSF algorithm is equivalent to the SSF approach of (75)-(76), and it was the SSF implementation that was used to obtain the results labeled MAP-SSF in the two figures. The WSF label refers to the SSF criterion implemented with the weighting matrix of (27) which
was derived by only taking the finite sample effects of noise into account. The results for no subspace weighting ($W = I$) and for the ESPRIT algorithm [21] are shown for comparison purposes. The continuous lines shown in the figures correspond to the (asymptotic) theoretical predictions of algorithm performance derived in [28, 29, 37], while the symbols ($x, o, +, *$) represent the simulation results. Note that the predicted performance of the MAP-SSF approach corresponds with the CRB, shown as a solid line in the figures. The value of $N$ required for the MAP-SSF algorithm to achieve the CRB depends on the difficulty of the scenario; for the case where the correlation is 90%, it appears that $N \geq 100$ is required. Larger levels of correlation would in turn require larger values of $N$. Except for the few cases where MAP-SSF has not yet reached its asymptotic performance limit, the algorithm significantly outperforms the other methods studied.

The results of a situation similar to that of Figure 1 are plotted in Figure 3, the main difference being that the gain/phase perturbation model of (10)-(11) was used instead, also with $\sigma_a = 0.05$. In addition, the DOA of the interferer was chosen to
Figure 2: A Comparison of DOA Estimation Performance, Unstructured Calibration Errors, Variable Signal Correlation
be $10^2$, and the correlation coefficient was reduced to -0.6. A slightly easier situation was considered in this case to ensure that the MAP-NSF method would be able to resolve the two sources. As mentioned earlier, one of the disadvantages of NSF-based approaches is their loss of consistency for highly correlated signals. When the signals are closely spaced in angle, this performance degradation manifests itself in a loss of resolution ability. In fact, the bulk of the RMS error for MAP-NSF shown in Figure 3 for $N = 10$ and $N = 20$ is due to the source at broadside not being resolved (i.e., both estimates occasionally converged to values near $10^0$). The SSF implementation of the MAP algorithm does not suffer from this deficiency, but has only been derived for the very simple error model of (13)-(17). The development of SSF-based MAP implementations for other error models is the subject of on-going research. In any case, when the MAP-NSF approach is able to resolve the two sources, its performance essentially achieves the CRB.
5.2. Beamforming Results

The first example involves a nominally unit-gain uniform linear array perturbed by an unstructured calibration error in the form of equation (13)-(17) with $Y = \sigma_a^2 I$ and $\sigma_a = 0.2$. The array receives 100 samples of two 20dB SNR uncorrelated Gaussian signals with arrival angles of 5° and 15°. Using DOA estimates from the optimal MAP estimator, the relative interference rejection capability of the MVDR, SC-MVDR, and MAP beamformers was calculated for various array sizes. The results are plotted in Figure 4 based on 500 independent trials. The plot shows the gain of the beamformer weights for the 5° source in the direction of the 15° interferer (normalized for a unit gain response at 5°). The subspace correction eliminates the signal cancelation effect of the MVDR approach, but the MAP beamformer provides a significant advantage, especially for larger arrays. The above simulation was repeated assuming receiver gain/phase errors as described by (10)-(11), also with $\sigma_a = 0.2$, and the results are plotted in Figure 5. The improvement for this type of calibration error is even more dramatic.
When the signals arriving at the array are highly correlated, interference rejection is no longer an appropriate performance criterion. In such cases, an optimal beamformer will attempt to combine correlated arrivals with the desired signal to improve the quality of the resulting estimate, as measured using (for example) mean-squared error. To examine beamformer performance for the case of correlated signals, a two-ray multipath channel was simulated for various relative delays. A mis-calibrated 5-element linear array was assumed to receive a random QPSK signal from $-6^\circ$, as well as a slightly delayed copy of the signal from $6^\circ$. Both arrivals had an SNR of 0 dB, and the array was again perturbed according to (13)-(17) with $Y = \sigma_n^2I$ and $\sigma_n = 0.15$. For each trial, MAP DOA estimates were obtained based on 75 samples from the array, and normalized RMS signal errors were computed. The results are plotted in Figure 6 for various relative delays between the two arrivals. The “uncompensated” approach corresponds to the method of [56] implemented with $A(\hat{\theta}, \rho_0)$ rather than $A(\hat{\theta}, \hat{\rho})$ as in the MAP beamformer. The minimum MSE curve was obtained using a known 75-sample training sequence to compute the optimal weights, and was included to give an

Figure 5: A Comparison of Beamformer Performance, Gain-Phase Calibration Errors
idea of the “best possible” performance.

While the SC-MVDR approach can to some degree compensate for array perturbations, it cannot eliminate signal cancelation due to the presence of a correlated arrival, and its performance in this case is quite poor. For small delays, correcting for calibration errors yields a 25-30% improvement in RMS error, which translates into a reduction in symbol error rate of approximately a factor of 6 (from .041 to .007) for this example.

6. Conclusions

In this chapter, a Bayesian approach is presented for alleviating the sensitivity of parametric DOA estimation methods to imperfections in the data collection system. The exact MAP estimator is derived and found to be impractically complicated. Based on recent results on subspace-based ML approximations, a simpler method referred to as the MAP-NSF technique is derived. The method takes a priori information about the structure of the errors into account in an optimal way, and provides asymptotically
minimum variance DOA estimates. Numerical examples and simulation studies have demonstrated a significant improvement as compared to traditional techniques. In particular, the examples of Section 5 demonstrated a substantial improvement in signal waveform estimation, resulting from appropriately utilizing the array perturbation model. This result is of great importance in, for example, communications applications. The price to be paid is an increased computational complexity. Despite the simplifications, the proposed method is notably more complex than DOA estimation methods that do not take the modeling errors into account. An exception is unstructured array errors, for which the MAP-NSF estimator reduces to the known signal subspace fitting (SSF) technique, but with a different choice of subspace weighting. An interesting alternative is therefore to approximate the array perturbations with an unstructured error model whenever possible. The resulting method is then no longer optimal, but can still yield substantially improved estimates compared to not taking the modeling errors into account.
References


