High Resolution Beamforming using Bayesian Inference

W.J.Fitzgerald J.Lasenby

Department of Engineering,
University of Cambridge,
Cambridge U.K.

Abstract

In this paper, Bayesian inference is applied to the problem of direction of arrival estimation. The method is discussed, and the algorithm is found to outperforms methods such as MUSIC and other eigen-decomposition techniques in the case of very low signal-to-noise ratios and closely spaced sources. One of the advantages of the present method is that it is able to handle special cases such as coherent and correlated signals and it does not assume the receivers to be equally spaced. It is also particularly well suited to the problem of estimating the directions of signals of quite different frequencies, and in such a case, the frequency and bearing estimation can be achieved together.

1 Introduction

In order to accurately locate closely spaced sources in the far field of an antenna array, a problem which is common to several fields, including sonar, radar, seismology and astronomy, a variety of methods have been developed. One of the first of these methods was the maximum likelihood (ML) technique [1], [2], which requires searching for a peak in the likelihood surface, which can be very computationally intensive, although the advent of parallel processing technology together with the use of highly efficient methods for performing the required multidimensional search, has led to much renewed interest in the method. Good results using simulated annealing and genetic algorithms [3], [4] have been found, and the use of alternating projections and the closely related IMP algorithm seem to offer many advantages over eigen-decomposition methods such as MUSIC [5], Minimum Energy, Capon's minimum variance method [6], and the method of Kumaresan and Tufts [7]. One of the major disadvantages with most of the above methods is that they require the number of sources, m, to be predetermined.

The purpose of this paper is to apply Bayesian inference to this problem of spatial source location, and to compare some of the results obtained with some of the above mentioned methods. Possible extensions of the method will also be given.

2 The Bayesian Method

In any form of data analysis, it is usually an implicit assumption that one has a physical model that is thought to best represent the observed, in many cases noisy, data. The problem facing the data analyst is therefore one of inferring values of the parameters of the particular model under consideration. A more complex task is to select the most plausible model from a predefined set, that best accounts for the data. The methods used to address this problem rely on scientific inference, and in order to carry out consistent reasoning and inference in situations having incomplete knowledge, Bayesian methods are uniquely placed.

For the purposes of this paper, Bayes' theorem may be written

$$P(H|D,I) = \frac{P(H|I)P(D|H,I)}{P(D|I)}$$

where $P(X|Y)$ represents a conditional probability i.e. the probability of X given Y. Thus, $P(H|D,I)$ is the probability of H (a stated hypothesis, say) given knowledge of D and any prior information. If H represents the set of parameters of the hypothesis that we are seeking to estimate, D the data we have recorded and I any prior knowledge we have of the system, then $P(D|I)$ is a normalization term and we can write;

$$P(H|D,I) \propto P(H|I)P(D|H,I)$$

$P(D|H,I)$ is known as the likelihood function. Thus, in the Bayesian approach we can assign a probability distribution to the unknown parameters, which is given by the prior multiplied by the likelihood. Within the framework of Bayes' theorem, we are then able to integrate out any unwanted or 'nuisance' parameters, a process known as marginalisation. If such marginalisation can be done anlaytically then the complexity of the problem can be reduced considerably.

In order to bring out the physics of the situation we are considering, it is helpful to consider the model function for our problem to be written using a summation convention, rather than a vector notation, [9], and we may write a general model function as,

$$f(t) = \sum_{k=1}^{m} B_k G_k(t,\omega)$$

where B and $G(t,\omega)$ are amplitudes and model functions respectively, and the data we are trying to account for by this model can be written as

$$d(t_i) = f(t_i) + e_i \quad \mathrm{i} = 1,2,3,\ldots,\mathrm{n}$$

where e is a noise vector. Therefore the hypothesis we are trying to test using Bayes' theorem is whether or not the model function given above accounts for the data. The posterior probability may be written $P(H|D,I)$ and this is taken to mean the probability of the model parameters given the data and prior information.

Integrating out the amplitudes $\boldsymbol{B}$, for example, we can write the marginalised posterior probabilities for the remaining parameters $\boldsymbol{\omega}$ as

$$P(\boldsymbol{\omega}|D,I) = \int P(\boldsymbol{B},\boldsymbol{\omega}|D,I)\, d\boldsymbol{B}$$

In our particular problem of estimating the directions of various sources, we will begin by discussing the simplest scenario in which there are m narrow-band sources of frequencies $\nu_1, \nu_2, \ldots\ldots, \nu_m$ distributed at elevation angles $\theta_1, \theta_2, \ldots\ldots, \theta_m$ in the far field of an equispaced linear array of n passive sensors. All sources are taken to lie in a single plane through the line of sensors, so that, for the moment, the problem is effectively one dimensional.

We assume signals arrive at an array as plane waves, with b the interelement spacing, and let θ be the elevation angle measured from the line of the array. The path difference between adjacent elements is $b\cos\theta$. Suppose we take a total of N time samples – often called 'snapshots'; the k^{th} source then produces a signal at time t_i at the j^{th} element given by

$$I_k(t_i)\sin[\omega_k(t_i + (j-1)\tau_k) + \phi_k(t_i)] \tag{1}$$

where $\tau_k = \frac{b\cos\theta}{c}$ and c is the speed of the wave in the medium, $\omega_k = 2\pi\nu_k$ =angular frequency of the k^{th} source, $\phi_k(t_i)$ = phase of the k^{th} signal at element 1 (phase reference for array) and time t_i and $I_k(t_i)$ = amplitude of the k^{th} signal at time t_i.

We will assume the general case of time-varying amplitudes and phases, and therefore if $d_j(t_i)$ is the data we collect from the j^{th} element at time t_i then

$$d_j(t_i) = \sum_{k=1}^{m} I_k(t_i)\sin[\omega_k(t_i + (j-1)\tau_k) + \phi_k(t_i)] + e_j(t_i) \tag{2}$$

for $j = 1, n$. Here, $e_j(t_i)$ is the noise at time t_i at the j^{th} sensor. The signals and noise are assumed zero mean and statistically stationary. We also take the noise to be uncorrelated from sensor to sensor with variance σ^2 at each. If the noise results from many independent processes then the Central Limit Theorem tells us that its probability distribution will take a Gaussian form:

$$P(e_j(t_i)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-e_j(t_i)^2}{2\sigma^2}\right) \tag{3}$$

Thus, for a sequence of n sensors and N time samples we have

$$P(e) = \prod_{i=1}^{N}\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-e_j(t_i)^2}{2\sigma^2}\right) \tag{4}$$

If we identify the noise with the difference between the data and the true signal then we obtain the likelihood function L as

$$L(\omega,\phi,I,\tau,\sigma) \propto \frac{1}{\sigma^{nN}} \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}[d_j(t_i) - \sum_{k=1}^{m} I_k(t_i)\sin[\omega_k(t_i+(j-1)\tau_k)+\phi_k(t_i)]]^2\right) \tag{5}$$

where $\omega = (\omega_1, \omega_2, \ldots, \omega_k)$ etc. The next step is to write

$$I_k(t_i)\sin[\omega_k(t_i+(j-1)\tau_k)+\phi_k(t_i)] = A_k(t_i)\sin[\omega_k(t_i+(j-1)\tau_k)]+B_k(t_i)\cos[\omega_k(t_i+(j-1)\tau_k)] \tag{6}$$

where $A_k(t_i) = I_k(t_i)\cos\phi_k(t_i)$ and $B_k(t_i) = I_k(t_i)\sin\phi_k(t_i)$. The likelihood can then be written as

$$L(\omega,\phi,I,\tau,\sigma) \propto \frac{1}{\sigma^{nN}} \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left[d_j(t_i) - \sum_{k=1}^{2m} C_k(t_i) f_{jk}(t_i)\right]^2\right) \tag{7}$$

where $C_k(t_i) = A_k(t_i)$, $f_{jk}(t_i) = \sin[\omega_k(t_i+(j-1)\tau_k)]$ for k=1,m and $C_k(t_i) = B_{k-m}(t_i)$, $f_{jk}(t_i) = \cos[\omega_{k-m}(t_i+(j-1)\tau_{k-m})]$ for k=m+1,2m.

Let us define a 'coherence time', t_{coh}, for the array, where

$$t_{coh} = \frac{d_{arr}}{c}$$

where d_{arr} is the total length of the array, and c is the speed of the wave. The coherence time, t_{coh}, is therefore the time taken for a signal to propagate across the array. If we are to assume the general case of possibly varying amplitudes and phases, then we are restricted to considering the data in blocks of $(f_s.t_{coh})$ or less, where f_s is the sampling frequency. One can, of course, deal with the data on a snapshot-by-snapshot basis, but if there are significant variations of amplitude and phase over a time of t_{coh} then the method is no longer applicable to the problem as the data in a single snapshot will then have different amplitudes and phases at each sensor.

Thus, suppose we have a total of N snapshots and n_b is the number of snapshots in each block, (n_b = integer part of $(f_s.t_{coh})$), we then have $N_b = N/n_b$ blocks of data to deal with, and we can therefore write the likelihood as

$$L(\omega,\phi,I,\tau,\sigma) \propto \frac{1}{\sigma^{nN}} \prod_{s=1}^{N_b} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1+(s-1)n_b}^{sn_b}\sum_{j=1}^{n}\left[d_j(t_i) - \sum_{k=1}^{2m} C_k^s f_{jk}(t_i)\right]^2\right) \tag{8}$$

i.e over each interval $t = [1+(s-1)n_b]dt$ to $(t + n_b dt)$ we assume that the amplitudes C_k^s are constant.

Our aim is to integrate out, for each s, the 2m amplitudes, C_k^s, as nuisance parameters, they can, however, be estimated at the end of the calculation. The way we have chosen to do this is to write the $f_{jk}(t_i)$ in terms of new functions $H_{jk}(t_i)$ which are orthonormal with respect to sums over i *and* j, having amplitudes $\bar{A}_k$.

We can now write L as

$$L(\omega,\vec{A},\tau,\sigma)=\prod_{s=1}^{N_b} L_s \tag{9}$$

where L_s are the likelihoods for each block with, for example, L_1 given by

$$L_1(\omega,\vec{A},\tau,\sigma)=\frac{1}{\sigma^{nn_b}}\exp\left(-\frac{1}{2\sigma^2}\left[\bar{d}^2-2\sum_{k=1}^{2m}\bar{A}_k\bar{h}_k+\sum_{k=1}^{2m}\bar{A}_k^{\,2}\right]\right) \tag{10}$$

where

$$\bar{h}_k=\sum_{i=1}^{n_b}\sum_{j=1}^{n} d_j(t_i)H_{jk}(t_i) \tag{11}$$

and

$$\bar{d}^2=\sum_{i=1}^{n_b}\sum_{j=1}^{n}(d_j(t_i))^2 \tag{12}$$

The expression for L_1 can now be split into 2m products:

$$L_1(\omega,\vec{A},\tau,\sigma)\propto\frac{1}{\sigma^{nn_b}}\prod_{k=1}^{2m}\exp\left(-\left[\frac{\bar{A}_k^{\,2}}{2\sigma^2}-\frac{\bar{A}_k\bar{h}_k}{\sigma^2}+\frac{\bar{d}^2}{4m\sigma^2}\right]\right) \tag{13}$$

As we have no prior information about the amplitudes $\vec{A}$, we assign them a uniformly flat prior probability i.e. $P(\vec{A}|I)$ =constant. This tells us that the probability of $\vec{A}$ given the data is therefore proportional to the likelihood. We should note here that we have chosen one of many options open to us, namely we have chosen to assign a non-informative prior to the transformed amplitudes $\vec{A}$. If we had chosen to assign a uniform prior to the original amplitudes C, then on transformation to our set of orthogonal amplitudes $\vec{A}$ we would introduce a factor of $||J||^{-1}$, where $||J||$ is the jacobian of the transformation from C to $\vec{A}$. The two approaches will obviously give different results - although in this particular problem it in fact amounts to very small changes in final output. We have chosen to assign a uniform prior to the transformed amplitudes as these are a set of effectively independent and orthogonal quantities and thus attributing a non-informative prior to each might be seen as more sound than doing so to a set of interrelated quantities: we do, however, stress that such choices of where to assign your priors is a matter of some debate and is a feature of the Bayesian approach. It should be realised that the IMP [8] algorithm can, of course, be reformulated in these terms, the current IMP being equivalent to the assumption of uniform priors on a particular set of non-orthogonal model functions.

Integrating out the 2m amplitudes gives,

$$L_1(\omega,\tau,\sigma) \propto \sigma^{2m-nn_b} \exp\left[-\frac{1}{2\sigma^2}(\bar{d^2}-\bar{h}^2)\right] \tag{14}$$

The total likelihood is then given by

$$L \propto \prod_{s=1}^{N_b} L_s \tag{15}$$

We can go one stage further and eliminate the noise variance σ from equation (15), as σ is, in general, unknown. In what follows we use both a Jeffreys prior, $1/\sigma$, and a constant prior for the rms noise.

Let us now take $\bar{d^2}$ and $\bar{h}^2$ to be the sum over all blocks of the seperate quantities defined previously. If we integrate the noise out we obtain the probability density function for ω and τ given the data D,

$$P(\omega,\tau|D) \propto \begin{cases} \left[1-\frac{\bar{h}^2}{\bar{d^2}}\right]^{\frac{2\bar{m}-nN}{2}} & \text{Jeffreys prior} \\ \left[1-\frac{\bar{h}^2}{\bar{d^2}}\right]^{\frac{2\bar{m}-nN+1}{2}} & \text{Uniform prior} \end{cases}$$

where $\bar{m} = mN_b$. If N is large, then the two above expressions will obviously be very similar. The objective is therefore to maximize $P(\omega,\tau|D)$ as given above.

3 Results

For the purpose of illustrating the Bayesian technique, the following simulations employ a 10-element array with an interelement spacing of 60m. The frequencies of the two sources used are both 10Hz and the sampling frequency is 100Hz. In a realistic situation the operating and sampling frequencies can, of course, be much larger and the interelement spacing much smaller. In order to avoid spatial aliasing, the interelement spacing must be less than half a wavelength. The following data were simulated by generating sinusoids with given amplitudes, frequencies and phases and adding to this pure gaussian white noise uncorrelated from sensor to sensor, and the amplitudes and phases were made to vary slowly with time. The sources are assumed to be in the far field of the array. The signal-to-noise ratios, S/N, quoted in the following are defined by:

$$\text{S/N} = 10\log_{10}\left[\frac{\sigma^2{}_{sig}}{\sigma^2{}_{noise}}\right]$$

where $\sigma^2{}_{sig}$ is the variance of the signal taken over the whole time series e.g. $\frac{1}{T}\int_0^T f^2(t)dt$, and $\sigma^2{}_{noise}$ is the variance of the normal distribution from which the random noise is generated. Just two cases are shown here as an illustration of the performance of our method versus MUSIC, using S/N ratios of -3dB, and -10dB. Each case has the sources separated by ±2 degrees, were ± refers to the positions of the sources either side of the normal to the array, and deals with data from two sources of identical frequencies ($\nu_1 = \nu_2 = 10$Hz), such as might occur from a multipath situation. In this case we must use full aperture smoothing (FAS) with the MUSIC algorithm, which involves averaging the data covariance matrix and its reversed complex conjugate. To locate the sources we must search for the peaks in the joint pdf. However, in the plots of $\log P(\tau|D)$ which follow, the estimates correspond to the minima, as the maximum map value has been subtracted from all pixel values – the true peak therefore occurs at zero. This was done to avoid very large numbers around the regions of interest.

We can display the Bayesian results graphically using a probability density function (pdf) contour plot in the $\theta_1 - \theta_2$ plane which effectively shows the likelihood surface (with nuisance parameters integrated out). Figure 1. shows the results for -3 dB at a separation of ±2 degrees, and compares the contour plot of $\log P(\tau|D)$ with the line plot produced from the MUSIC algorithm, and in both cases a total of 1024 snapshots were used. Figure 2. shows the corresponding results for -10 dB for the same separation as before, and for a total of 2000 snapshots.

At -3 dB the Bayesian technique is able to resolve and locate the 2 sources using 1024 snapshots whereas MUSIC can barely resolve them. At -10 dB the results are very much more striking; The Bayesian method once more produces accurate position estimates for the two sources but does need considerably more snapshots when the noise is this high. MUSIC however, fails completely in this case.

These results are meant to give a rough idea of the performance of the Bayesian methods compared to the MUSIC algorithm, which is commonly used as a benchmark against which one evaluates alternative methods. A more comprehensive study will be published elsewhere, and will consider the behaviour of the proposed method under a variety of circumstances. For example, it is of interest to us to consider how the method behaves as the number of snapshots taken is decreased, the capability of the method to identify frequencies as well as positions, the behaviour with non-equispaced arrays, etc. In this present paper, our objective is simply to present a few cases which illustrate the potential of the technique, and it is clear from the above results that in terms of intrinsic resolution ability, the method shows considerable improvements over standard eigen-decomposition techniques.

4 Conclusions

It has been shown in this paper that Bayesian inference methods have significant advantages over MUSIC and other eigen-decomposition methods in the cases of low signal-to-noise ratios and closely separated sources. An attraction of the method is that it needs no modifications or alterations to deal with the cases of coherent and/or fully correlated sources and the framework also allows one to estimate frequencies and spatial positions simultaneously without change in the formulation.

There are a number of extensions and modifications that could enhance the technique still further. The first obvious modification that is currently under investigation is to allow the number of sources to be estimated (model order selection problem), and also to allow for coloured Gaussian noise, and the use of an iterative scheme to cancel out already detected sources thus allowing weaker sources to be detected one by one, [9].

Work is in progress to evaluate the performance of the marginal statistical estimators discussed in this paper, compared with the performance of maximum likelihood and maximum a posteriori estimators, and also to compare the performance with the Cramer-Rao lower bound, and the Berankin bound at low signal to noise ratios.

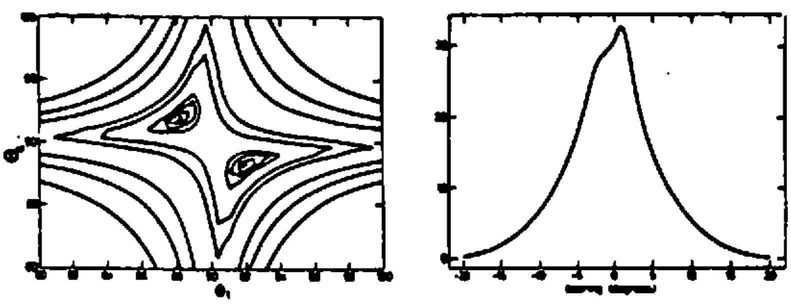

Figure 1. Two -3dB sources placed at elevation angles of 88° and 92°; each signal had a frequency of 10Hz. a) shows the results of the BML technique using 1024 snapshots. Contours of $\log P(\tau|D)$ are shown on a 30 x 30 grid. b) shows the results of MUSIC with FAS. 1024 snapshots were used with a 0.2° stepsize.

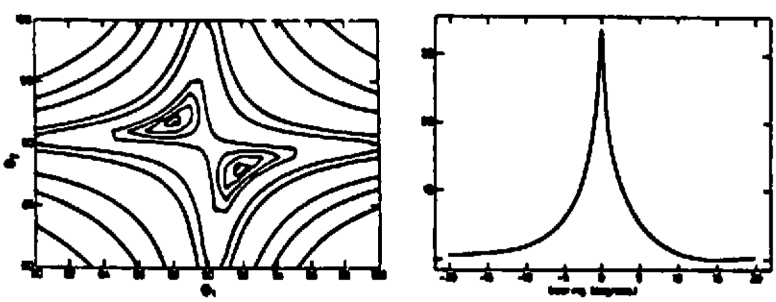

Figure 2. Two -10dB sources placed at elevation angles of 88° and 92°; each signal had a frequency of 10Hz. a) shows the results of the BML technique using 2000 snapshots. Contours of $\log P(\tau|D)$ are shown on a 30 x 30 grid. b) shows the results of MUSIC with FAS. 2048 snapshots were used with a 0.2° stepsize.

References

[1] F.C. Schweppe, "Sensor array data processing for multiple signal sources", *IEEE Trans. Inform. Theory.* vol.IT,14, 1968, pp294-305.

[2] W.S. Ligget, "Passive sonar: fitting models to multiple time-series", in NATO ASI on *Signal Processing*, J.W.R. Griffiths *et al.*, Eds. New York: Academic, 1973, pp.327-345.

[3] K.C.Sharman and G.D.McClurkin, 'Estimating the parameters of correlated sources from arbitrary arrays - A Simulated Annealing solution' *IEE Colloqium on New Trends in Sensor Array Processing'.* Digest no. 1988/130 p3/1-3/5, 1988.

[4] K.C.Sharman and G.D.McClurkin, 'Genetic Algorithms for Maximum Likelihood parameter estimation', *IEEE, ICASSP.* Vol.4 p2716-2719, 1989.

[5] R.O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Trans. Antennas Propagat.*, vol.AP-34, 1986, pp.276-280.

[6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis", *Proc. IEEE*, vol.57, 1969, pp.1408-1418.

[7] R. Kumaresan and D.W. Tufts, "Estimating the angles of arrival of multiple plane waves", *IEEE Trans.Aerospc.Electron.Syst.*, vol.AES-19, 1983, pp.134-139.

[8] I.J. Clarke, "Partial knowledge of solution drives goals for adaptive decomposition of data", *IEE colloquium on 'New Trends in Sensor Array Processing'*, held at Savoy Place, London, 6th December 1988. Organized by Professional Group E16 (signal processing). Digest No. 1988/130.

[9] W.J.Fitzgerald, Bayesian Inference in Signal Processing, Cambridge University, Engineering Department Technical Report, CUED/F INFENG/T80, 1991,