Likelihood Function for a Simple Cardioid Sonobuoy

Andrew R. Runnalls*

1 March 2006

This paper is a postprint of a paper submitted to and accepted for publication in IEE Proceedings Radar, Sonar and Navigation 153 [2006] pp. 417–26, and is subject to IET copyright. The copy of record is available at http://www.ietdl.org.

Abstract

This paper considers the common sonobuoy configuration comprising two horizontal dipoles mounted orthogonally together with an omnidirectional sensor. Under the assumption that, within a narrow frequency band of interest, the acoustic power impinging on the sonobuoy consists of uncorrelated Gaussian noise processes, the paper examines how data collected by the sensor can be used to infer the directional distribution of the incident power, as characterised by its spherical harmonic coefficients. The results cover arbitrary directional distributions, and are applicable to small sample sizes as well as large. In particular, the paper examines the form of the likelihood function for the spherical harmonic coefficients, as required for Bayesian inference methods such as particle filtering. Specific attention is given to the case of a single point source superimposed on azimuthally isotropic noise. In this case it is shown that the maximum likelihood estimator for target bearing differs from, and can be considerably more efficient than, the bearing estimator traditionally used for this sensor configuration.

1 Introduction

Consider the common sonobuoy configuration shown in Fig. 1, in which two matched dipoles are arranged orthogonally in the horizontal plane, with a colocated omnidirectional sensor. We shall use a Cartesian frame of reference with its x and y axes as shown in the figure, and z-axis vertically upwards. We call this a **cardioid** sensor because a common method of utilising such a sensor configuration is to combine the outputs of the sensor elements so as to achieve a cardioid reception pattern, with its null steered in the direction of a nuisance acoustic source (cf. [1]).

In this paper we shall suppose that, within a narrow frequency band of interest, the acoustic power impinging on the sonobuoy consists of uncorrelated Gaussian noise processes arriving from various directions, and consider how data collected from the sonobuoy can be analysed to estimate the directional distribution of the arriving power.

There is an extensive literature on direction-of-arrival estimation based on data collected at hydrophone arrays of arbitrary configuration: see for example [2, 3, 4, 5, 6]. Typically these consider the estimation of one or more point sources combined with spherically isotropic noise (and/or internal sensor noise). The general case is analytically somewhat intractable, and consequently much of the literature has focussed on asymptotic results, valid as the sample size tends to infinity.

*The author is with the Computing Laboratory at the University of Kent, Canterbury CT2 7NF, Kent, UK, (e-mail: A.R.Runnalls@kent.ac.uk).

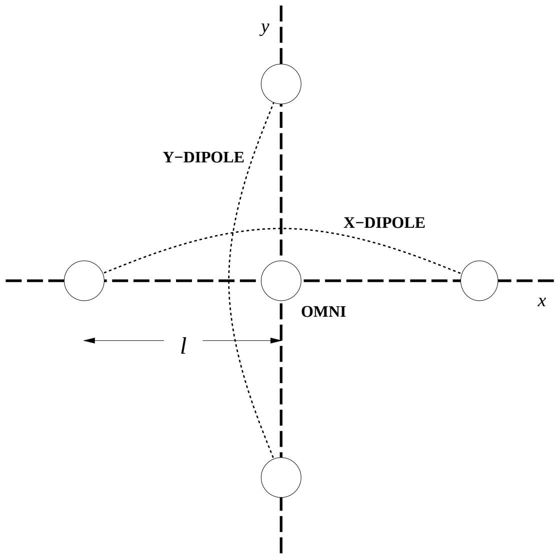

Figure 1: Cardioid sensor configuration.

The present paper is motivated by the requirements of Bayesian analysis of data from several sonobuoys of this type, sequentially processing samples of data that may be individually quite small. The most familiar example of a Bayesian data fusion method is the Kalman filter,[1] in which an estimate of the unknown system state is repeatedly updated, on the one hand to take account of the system dynamics (e.g. target motion), and on the other hand to take account of new sensor information. A severe restriction with the Kalman filter, however, is that it can only handle systems which can be represented (or approximated) using linear dynamics and Gaussian noise. Consequently, developers of Kalman filters have little practical incentive to formulate system models that do not conform to this pattern. The situation has radically changed during the last decade with the development of Monte Carlo Markov Chain (MCMC) methods [9], particularly in their sequential form, often called particle filters [10, 11]. These methods provide recursive state estimation, after the fashion of the Kalman filter, but using dynamic models that are not necessarily linear, and statistical models of sensor errors that are not necessarily Gaussian. This allows for greater realism—at the cost, it should be said, of much heavier computation requirements than the Kalman filter.

A key ingredient in these methods is the exact formulation of the likelihood function (explained in Sec. 6). Because of the simplicity of the Fig. 1 sensor configuration, this proves to be more tractable than in the case of general hydrophone arrays. Another way in which this paper differs from the direction-of-arrival literature cited above is that we address an arbitrary directional distribution of the incident power, expressed using spherical harmonic functions, explained in Sec. 4. In particular, this enables us to consider point sources in non-isotropic noise. As well as this general case, we explore in particular the estimation of the direction of arrival of a (random) signal from a single point source immersed in azimuthally—but not necessarily spherically—isotropic noise.

Although the motivation of the paper is Bayesian inference, we take the opportunity to explore the sampling behaviour of classical point estimators, in particular the maximum likelihood estimator of target bearing.

The layout of the paper is as follows. Sec. 2 describes how the data streams from the sensors are preprocessed and condensed into a 3×3 matrix $\mathbf{S}$, which forms a sufficient statistic for the directional distribution

[1] Although often presented as a least-squares algorithm (e.g. [7, 8]), the Kalman filter also readily lends itself to a Bayesian interpretation; indeed, the Kalman measurement update equation follows directly from Bayes' Theorem.

of the arriving power. Sections 3 to 5 develop a probability model for the scenario under consideration. Sec. 6 uses this model to formulate the log-likelihood function for the spherical harmonic coefficients of directional distribution of arriving power; Sec. 6.1 specialises the discussion to the case of a single point source in azimuthally isotropic noise. The paper then goes on, in Sec. 7, to consider the effect of normalising the matrix $\mathbf{S}$ with respect to the signal power received at the omnidirectional sensor: this has the effect of removing the ambient noise power as a parameter of the likelihood function. Finally Sec. 8 summarises the conclusions and suggests further work.

2 Signal Preprocessing

2.1 Dipole processing

Suppose that, for each dipole, the outputs of the two hydrophones that it comprises are subtracted. For a signal approaching at an angle α to the axis of the dipole, the transfer function will be

$$H(i\omega) = e^{i\omega(l/v)\cos\alpha} - e^{-i\omega(l/v)\cos\alpha} = 2i\sin(\omega(l/v)\cos\alpha)$$

where v is the local velocity of sound. Now if for all angular frequencies ω in the band of interest, $v/\omega \gg l$ (i.e. the dipole dimensions are small compared to the wavelength), the transfer function is effectively $(2i\omega l\cos\alpha)/v$. We now apply a further gain factor $v/2l$ to the dipole output and integrate it (within the frequency band of interest), so that the final transfer function becomes simply

$$H(i\omega) \approx \cos\alpha \tag{1}$$

which does not depend on ω, and brings the dipole signals into phase with the omni signal.

2.2 Shaping, demodulation and sampling

Suppose that the frequency band of interest extends over the angular frequencies $\omega_c \pm \omega_b$ where $\omega_b \ll \omega_c$. Each of the three received signals—one from each dipole, preprocessed as described in Sec. 2.1, together with the omnidirectional signal—is processed as follows:

1. Shape the signal, if necessary, so that the power spectral density of the ambient noise is flat within the frequency band of interest. (The shaping required will be the same for the omnidirectional and dipole signals.)
2. Demodulate the signal around ω_c, and sample the in-phase and quadrature components at a rate of ω_b/π samples/sec.
3. Interleave the in-phase samples with the quadrature samples (with the in-phase sample first).
4. Collect the data for a period of T seconds, to form a sequence of $\nu = 2\omega_b T/\pi$ samples.

2.3 Formation of S-matrix

The output from the preceding subsection will be three sequences each comprising ν (real) numbers. Let these sequences be denoted $s_x(i)$, $s_y(i)$ for the sequences from the x and y dipoles respectively, and $s_o(i)$ for the sequence from omnidirectional sensor. We now define the vectors $\mathbf{x}_i = (s_x(i), s_y(i), s_o(i))^T$ and form the matrix:

$$\mathbf{S} = \sum_{i=0}^{\nu-1} \mathbf{x}_i \mathbf{x}_i^T = \begin{pmatrix} S_{xx} & S_{xy} & S_{xo} \\ S_{xy} & S_{yy} & S_{yo} \\ S_{xo} & S_{yo} & S_{oo} \end{pmatrix} \tag{2}$$

where the superscript T denotes transpose. The diagonal terms of this matrix are the sums of squares of the processed signals, with the remaining terms being cross-products.

n, m	$Y_{n,m}(\theta, \phi)$
0, 0	$1/\sqrt{4\pi}$
1, 0	$\sqrt{3/4\pi}\ \sin\phi$
1, 1	$-\sqrt{3/8\pi}\, e^{i\theta}\ \cos\phi$
2, 0	$\sqrt{5/16\pi}\,(2 - 3\cos^2\phi)$
2, 1	$-3\sqrt{5/24\pi}\, e^{i\theta}\ \sin\phi\cos\phi$
2, 2	$3\sqrt{5/96\pi}\, e^{2i\theta}\ \cos^2\phi$

Table 1: Spherical harmonic functions for $n \leq 2$, $0 \leq m \leq n$.

3 Model for a Single Random Source

Consider the case of a signal approaching[2] from an azimuth θ measured anticlockwise (looking downwards) from the positive x-axis, and from an elevation angle ϕ ($-\pi/2 \leq \phi \leq \pi/2$), and assume that no other source of acoustic energy is impinging on the sensors. Then, in virtue of the transfer function in (1), we have for each i, $\mathbf{x}_i = a(\theta, \phi)s_o(i)$, where $a(\theta, \phi)$ is the transfer vector $a(\theta, \phi) = (\cos\theta\cos\phi, \sin\theta\cos\phi, 1)^T$.[3]

Now further suppose that the signal is a Gaussian noise process whose power spectral density (after the shaping described in Sec. 2.2) is flat within the passband. Then (cf. [12, §10-6]) the $s_o(i)$ will be independent samples from a Normal distribution with mean zero and some variance σ^2. Therefore the vectors $\mathbf{x}_i$ will be independent samples from a multivariate Normal distribution with zero mean vector and covariance matrix:

$$\begin{aligned}
\mathbf{V}_s(\theta, \phi) &= \sigma_s^2 a(\theta, \phi)a(\theta, \phi)^T \\
&= \sigma_s^2 \begin{pmatrix} \cos^2\theta\cos^2\phi & \sin\theta\cos\theta\cos^2\phi & \cos\theta\cos\phi \\ \sin\theta\cos\theta\cos^2\phi & \sin^2\theta\cos^2\phi & \sin\theta\cos\phi \\ \cos\theta\cos\phi & \sin\theta\cos\phi & 1 \end{pmatrix}
\end{aligned} \tag{3}$$

4 Directional Distribution of Incident Power

In Sec. 3 we considered a Gaussian noise signal arriving from a specific azimuth and elevation. We now generalise the results of that section to allow multiple such point sources of Gaussian noise, plus diffuse ambient noise. We do this by supposing that the distribution in azimuth and elevation of incident power in the passband is described by a density function $\varphi(\theta, \phi)$, so that the power arriving from azimuths between θ and $\theta + \delta\theta$ and elevations between ϕ and $\phi + \delta\phi$ is given, in the limit as $\delta\theta \to 0$ and $\delta\phi \to 0$, by $\varphi(\theta, \phi)\cos\phi\,\delta\theta\,\delta\phi$.

Provided it satisfies unexacting regularity conditions, $\varphi(\theta, \phi)$ can be represented as follows:

$$\varphi(\theta, \phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{n,m}Y_{n,m}(\theta, \phi) \tag{4}$$

where the $Y_{n,m}(\theta, \phi)$ are the **spherical harmonic functions**, and the $a_{n,m}$ are complex-valued constants. This expansion is analogous to the expansion of a periodic function as a Fourier series.[4]

[2]By this we mean that in the vicinity of the sensor, the wavefronts are approaching from this direction. This may of course be very different from the line of sight to the source of the signal, especially in elevation.

[3]The notation $a(\theta, \phi)$ follows for example [4, 5, 6]. Since the elements of this vector are real-valued, we take the opportunity to work with real rather than complex numbers throughout, which facilitates cross-referencing to the statistics literature.

[4]For similar applications of spherical harmonics within underwater acoustics, see [13, esp. §IIIA] and [14, 15].

We shall follow (subject to some differences in notation) the definition of spherical harmonics used in [16, 17]. With this definition, the first few spherical harmonic functions—all we shall need—are given in Table 1 (cf. [16, Table 12.3]) for $m \geq 0$; the functions for negative m can by obtained using the relationship

$$Y_{n,-m}(\theta, \phi) = (-1)^m \overline{Y_{n,m}(\theta, \phi)} \tag{5}$$

where the overbar represents complex conjugation.

The coefficients $a_{n,m}$ in the expansion (4) are now given by:

$$a_{n,m} = \int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} \varphi(\theta, \phi) Y_{n,m}(\theta, \phi) \cos\phi \, d\theta \, d\phi \tag{6}$$

and the spherical harmonic functions satisfy the following orthonormality relation:

$$\int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} Y_{n,m}(\theta, \phi) Y_{n',m'}(\theta, \phi) \cos\phi \, d\theta \, d\phi = \delta_{m,m'} \delta_{n,n'}$$

for $n, n' \geq 0$, $-n \leq m \leq n$ and $-n' \leq m' \leq n'$, where $\delta_{i,j}$ is Kronecker's delta function.

Note that the total power impinging on the sensors is given by:

$$\int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} \varphi(\theta, \phi) \cos\phi \, d\theta \, d\phi = \sqrt{4\pi} a_{0,0} \tag{7}$$

5 General Probability Model

Building on the previous section, let us assume that the power inpinging on the sensors within the passband is a superposition of independent Gaussian noise processes distributed in azimuth and elevation according to the directional density function $\varphi(\theta, \phi)$. In this case, the observation vectors $\mathbf{x}_i$ will still—as in Sec. 3—be independent samples from a multivariate Normal distribution with a zero mean vector, but with a covariance matrix $\mathbf{V}(\varphi)$ which is the appropriately weighted integral of $\mathbf{V}_s(\theta, \phi)$ from (3):

$$\mathbf{V}(\varphi) = \int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} \mathbf{V}_s(\theta, \phi) \varphi(\theta, \phi) \cos\phi \, d\theta \, d\phi \tag{8}$$

We now note that, using (6), the matrix $\mathbf{V}_s(\theta, \phi)$ can itself be represented in spherical harmonics as follows:

$$\mathbf{V}_s(\theta, \phi) = \sigma_s^2 \frac{\sqrt{\pi}}{15} \times \tag{9}$$
$$\begin{pmatrix} 10Y_{0,0} - 2\sqrt{5}Y_{2,0} + 2\sqrt{30}Y_{2,2}^c & 2\sqrt{30}Y_{2,2}^s & -10\sqrt{6}Y_{1,1}^c \\ 2\sqrt{30}Y_{2,2}^s & 10Y_{0,0} - 2\sqrt{5}Y_{2,0} - 2\sqrt{30}Y_{2,2}^c & -10\sqrt{6}Y_{1,1}^s \\ -10\sqrt{6}Y_{1,1}^c & -10\sqrt{6}Y_{1,1}^s & 30Y_{0,0} \end{pmatrix}$$

where $Y_{n,m}^c(\theta, \phi)$ and $Y_{n,m}^s(\theta, \phi)$ are respectively the real and imaginary parts of $Y_{n,m}(\theta, \phi)$.

Consequently, using (6), (8) and (9):

$$\mathbf{V}(\varphi) = \frac{\sqrt{\pi}}{15} \times \tag{10}$$
$$\begin{pmatrix} 10c_{0,0} - 2\sqrt{5}c_{2,0} + 2\sqrt{30}c_{2,2} & 2\sqrt{30}s_{2,2} & -10\sqrt{6}c_{1,1} \\ 2\sqrt{30}s_{2,2} & 10c_{0,0} - 2\sqrt{5}c_{2,0} - 2\sqrt{30}c_{2,2} & -10\sqrt{6}s_{1,1} \\ -10\sqrt{6}c_{1,1} & -10\sqrt{6}s_{1,1} & 30c_{0,0} \end{pmatrix}$$

where instead of the complex coefficients $a_{n,m}$ we have used their real and imaginary parts $c_{n,m}$ and $s_{n,m}$.

It is apparent by inspection of (10) that if we knew the value of $\mathbf{V}(\varphi)$ we could determine all the spherical harmonic coefficients $a_{n,m}$ up to order $n = 2$, with the exception of $a_{1,0}$, $a_{2,-1}$ and $a_{2,1}$. From Table 1 and (5), it emerges that these three exceptions correspond precisely to those spherical harmonic functions that are *not* even functions of ϕ. This reflects the fact that since all the sensor elements lie in a single horizontal plane, the sensor configuration is unable to distinguish up from down.

5.1 Example 0: Isotropic noise

Notice that if $\varphi(\theta, \phi)$ is uniform in azimuth, i.e. does not depend on θ, then since $\int_{\theta=0}^{2\pi} e^{im\theta}\, d\theta = 0$ for any non-zero integer m, then from (6) all the coefficients $a_{n,m}$ with non-zero m will be zero, whereupon (10) reduces to

$$\mathbf{V}(\varphi) = \frac{\sqrt{\pi}}{15} \begin{pmatrix} 10c_{0,0} - 2\sqrt{5}c_{2,0} & 0 & 0 \\ 0 & 10c_{0,0} - 2\sqrt{5}c_{2,0} & 0 \\ 0 & 0 & 30c_{0,0} \end{pmatrix} \tag{11}$$

in accordance with the well-known result (cf. [1]) that in these circumstances the output of the three sensors will be uncorrelated.

If we write:

$$g = \frac{30c_{0,0}}{10c_{0,0} - 2\sqrt{5}c_{2,0}} = \frac{2\int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} \varphi(\theta, \phi) \cos\phi \, d\theta \, d\phi}{\int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=0}^{2\pi} \varphi(\theta, \phi) \cos^3\phi \, d\theta \, d\phi} \tag{12}$$

then (11) can be rewritten as:

$$\mathbf{V}(\varphi) = \sigma_n^2 \begin{pmatrix} 1/g & 0 & 0 \\ 0 & 1/g & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13}$$

where σ_n^2 is the total noise power as given by (7). (g is commonly referred to as the *dipole gain*, since the noise power is attenuated by a factor of g in comparison with a signal from a point source aligned with the dipole.) If the distribution of incident power is spherically isotropic, so that $\varphi(\theta, \phi)$ is a constant, then $c_{2,0}$ will be zero, and g will be three. If the incident noise power is uniform in azimuth, but is entirely concentrated in the horizontal plane ($\phi = 0$), then by inspection of (12) it is clear that $g = 2$. These results agree with [1].[5]

Notice in particular that so far as the behaviour of this sensor configuration is concerned, the distribution in elevation of azimuthally isotropic noise is entirely characterised by the single parameter g.

6 Likelihood Function

Suppose that a collection $\mathbf{d}$ of sensor output data has a probability distribution which depends on a vector of unknowns $\boldsymbol{\vartheta}$, and that we wish to use the observed value of $\mathbf{d}$ to draw conclusions about the value of $\boldsymbol{\vartheta}$. The **likelihood function** is defined as follows:

$$L(\boldsymbol{\vartheta}|\mathbf{d}) = c(\mathbf{d}) f(\mathbf{d}|\boldsymbol{\vartheta}) \tag{14}$$

where $c(\mathbf{d})$ is an arbitrary function of $\mathbf{d}$. Essentially this is just the probability density function of $\mathbf{d}$ conditional on $\boldsymbol{\vartheta}$, but considered as a function of $\boldsymbol{\vartheta}$ rather than a function of $\mathbf{d}$.

In the case considered in this paper, our raw observations are the vectors $\mathbf{x}_i$ for $i = 0, \ldots \nu - 1$, and we wish to draw inferences about the covariance matrix $\mathbf{V} = \mathbf{V}(\varphi)$ and hence about the coefficients $a_{n,m}$ describing the directional power density φ. (We shall assume that $\nu \geq 3$; in practice it will usually be much

[5] But note that [1] defines dipole gain in amplitude rather than power terms: our g is their g^2.

greater than this.) A sufficient statistic[6] for $\mathbf{V}$, and hence for these quantities, is the matrix $\mathbf{S}$ previously defined in (2). This matrix has a Wishart distribution with ν degrees of freedom (cf. [19, p. 162], [20, Eq. 3.2.1]) with probability density function:

$$f(\mathbf{S}|\mathbf{V}) = k_{\mathbf{V}} \det(\mathbf{S})^{(\nu-4)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{S})\right) \tag{15}$$

where the normalisation term $k_{\mathbf{V}}$ is given by

$$k_{\mathbf{V}}^{-1} = 2^{3\nu/2}\pi^{3/2}\det(\mathbf{V})^{\nu/2}\prod_{j=1}^{3}\Gamma\left(\frac{\nu-j+1}{2}\right)$$

In the likelihood function, by appropriate choice of the function $c(\mathbf{d})$ we can get rid of any factors that do not depend on $\mathbf{V}$, and thus simplify (15) to obtain:

$$L(\mathbf{V}|\mathbf{S}) = \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{S})\right) / \det(\mathbf{V})^{\nu/2}$$

with the corresponding log-likelihood function (cf. [3, Eq. 2]):

$$\ln L(\mathbf{V}|\mathbf{S}) = -\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{S}) - \frac{\nu}{2}\ln\det(\mathbf{V}) \tag{16}$$

It is possible to express the log-likelihood function directly in terms of the spherical harmonic coefficients by substituting from (10) into (16), but the result is rather unwieldy in the general case, so we shall not carry it out explicitly; instead, in the following subsection we shall consider an illustrative special case.

6.1 Example 1: Point source with isotropic noise

Now consider the case where azimuthally isotropic noise of total power σ_n^2, distributed in elevation so as to give a dipole gain g (12), is combined with a single signal of power $\rho\sigma_n^2$ arriving from azimuth angle θ and elevation angle ϕ. By making the appropriately weighted sum of (3) and (13), we get:

$$\mathbf{V} = \sigma_n^2 \begin{pmatrix} \rho\cos^2\theta\cos^2\phi + 1/g & \rho\sin\theta\cos\theta\cos^2\phi & \rho\cos\theta\cos\phi \\ \rho\sin\theta\cos\theta\cos^2\phi & \rho\sin^2\theta\cos^2\phi + 1/g & \rho\sin\theta\cos\phi \\ \rho\cos\theta\cos\phi & \rho\sin\theta\cos\phi & \rho+1 \end{pmatrix} \tag{17}$$

From this it is straightforward to show that

$$\det(\mathbf{V}) = \sigma_n^6\frac{\rho(1+g\cos^2\phi)+1}{g^2} \tag{18}$$

(so $\det(\mathbf{V})$ does not depend on θ) and

$$\det(\mathbf{V})\mathrm{tr}(\mathbf{V}^{-1}\mathbf{S})/\sigma_n^4 = c_0 + \rho h(\theta,\phi) \tag{19}$$

where

$$c_0 = \frac{S_{xx}+S_{yy}}{g} + \frac{S_{oo}}{g^2} \tag{20}$$

[6]The sufficiency of $\mathbf{S}$ is readily established using the Factorisation Theorem (cf. [18, §3.3]) and the relationship $\mathbf{x}_i\mathbf{V}^{-1}\mathbf{x}_i^T = \mathrm{tr}(\mathbf{V}^{-1}\mathbf{x}_i\mathbf{x}_i^T)$.

and

$$
\begin{aligned}
h(\theta, \phi) \quad = \quad & \frac{2 + g\cos^2\phi}{2g}(S_{xx} + S_{yy}) + \frac{S_{oo}\cos^2\phi}{g} \\
& - \frac{2\cos\phi}{g}(S_{xo}\cos\theta + S_{yo}\sin\theta) \\
& + \left(\frac{S_{yy} - S_{xx}}{2}\cos 2\theta - S_{xy}\sin 2\theta\right)\cos^2\phi
\end{aligned} \tag{21}
$$

Notice that $h(\theta, \phi)$ does not depend on ρ or σ_n^2. Also, the left-hand side of (19) is necessarily non-negative, since the matrix $\mathbf{V}$ is positive definite and $\mathbf{S}$ is non-negative-definite. Moreover the equation holds for any non-negative value of ρ. It follows that $h(\theta, \phi)$ is itself non-negative.

The log-likelihood can now be expressed by substituting from (18) and (19) into (16), to yield:

$$
\begin{aligned}
& \ln L(\theta, \phi, \rho|\mathbf{S}) \\
= \quad & -\frac{g^2}{2\sigma_n^2(\rho[1 + g\cos^2\phi] + 1)}\{c_0 + \rho h(\theta, \phi)\} \\
& -\frac{\nu}{2}\left\{3\ln(\sigma_n^2) + \ln\left(\frac{\rho(1 + g\cos^2\phi) + 1}{g^2}\right)\right\}
\end{aligned} \tag{22}
$$

Notice that the log-likelihood depends on the target azimuth θ only via the function $h(\theta, \phi)$.

6.1.1 Numerical illustration

To give a flavour of the log-likelihood function defined in (22), Fig. 2 shows a perspective view of the corresponding likelihood function considered as a function of θ and ρ (expressed in the figure in degrees and dB respectively), with ϕ and g assumed to be 0 and 3 respectively, and σ_n^2 set to a nominal value of unity. The likelihood function is calculated for the following value of the $\mathbf{S}$-matrix:

$$
\mathbf{S} = \begin{pmatrix} 8.30727 & 1.72855 & 5.96266 \\ 1.72855 & 2.47490 & 1.58611 \\ 5.96266 & 1.58611 & 30.37420 \end{pmatrix}
$$

This matrix was randomly sampled from the distribution (15), with $\mathbf{V}$ given by (17) with ϕ, g and σ_2^n having the values just given, $\theta = 0$, $\rho = 0.1$ (-10 db) and $\nu = 30$. Note the following points:

- The point of maximum likelihood is at $\theta = 15.2°$, $\rho = -12.6$ dB, somewhat offset from the true values $\theta = 0$, $\rho = -10$ dB.
- If we consider the section of the figure along the line $\rho = -12.6$ dB (i.e. passing through the point of maximum likelihood), in addition to the primary peak at $\theta = 15.2°$, there is a secondary peak, with about 20% of the maximum likelihood, at about $\theta = -165°$.
- If we consider the section of the figure along the line $\theta = 15.2°$, the likelihood falls off sharply above $\rho = -11$ dB. Below -13 dB, however, the likelihood falls off only slightly, and flattens out at about 72% of its maximum value.
- If ρ is assumed to have some value less than -25 dB, then the likelihood function varies only slightly as a function of θ.

With an SNR of -10 dB, $\nu = 30$ is of course rather a small sample size for practical use, but using this value in Fig. 2 helps to reveal the underlying character of the likelihood function. As ν and/or ρ are increased, the likelihood function tends to become dominated by a sharp peak in the vicinity of the true values of θ and ρ.

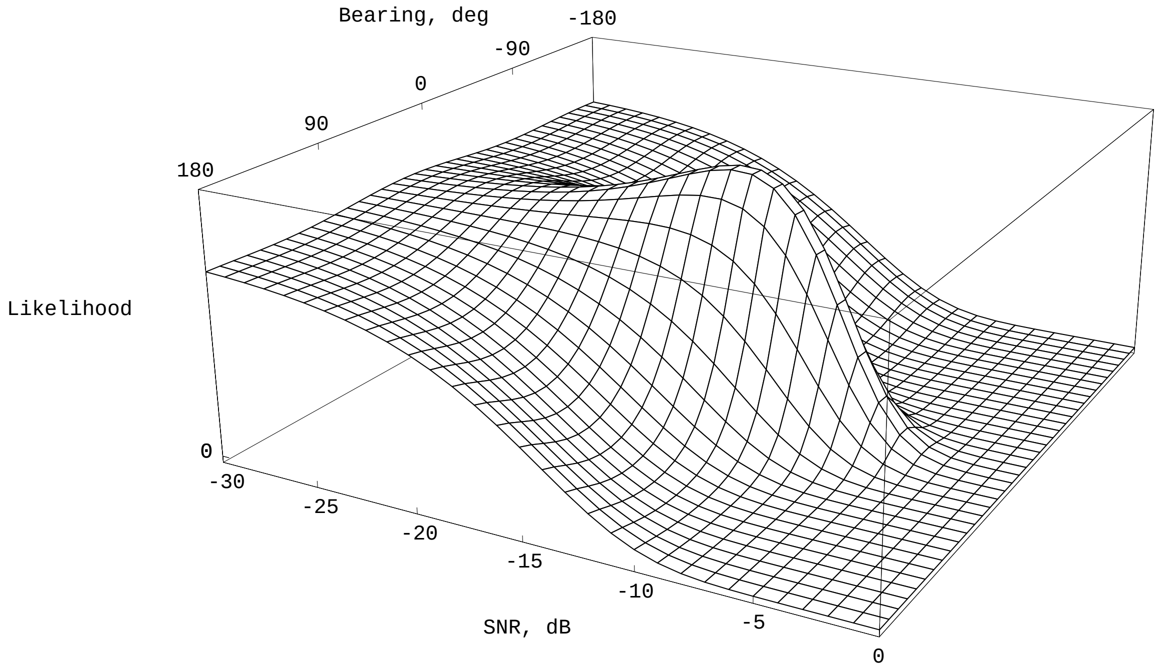

Figure 2: Likelihood function corresponding to the log-likelihood function of (22), evaluated for $g = 3$, $\sigma_n^2 = 1$, $\phi = 0$, and for a matrix $\mathbf{S}$ as described in the text. The likelihood function is defined only up to an arbitrary scaling factor, so no scale is shown in the figure.

6.1.2 Fisher information for target bearing

A measure of the information contained within a random dataset $\mathbf{x}$ about an unknown parameter ϑ is the **Fisher information**:

$$\mathcal{I}(\vartheta) = E_{\mathbf{x}}\left\{\left(\frac{d\ln f(\mathbf{x}|\vartheta)}{d\vartheta}\right)^2\right\} = E_{\mathbf{x}}\left\{\left(\frac{d\ln L(\vartheta|\mathbf{x})}{d\vartheta}\right)^2\right\} \tag{23}$$

where $E_{\mathbf{x}}\{\cdot\}$ represents statistical expected value with respect to the random variable $\mathbf{x}$. (See for example [21, Sec. 5a.4].) We work primarily in terms of the Fisher information rather than the more familiar Cramér-Rao bound because, in its usual formulation ($1/\mathcal{I}(\theta)$), the CRB is not strictly applicable to angular data, essentially because the reduction of errors modulo 2π results in an *upper* bound on estimation errors.[7]

In this subsection, we shall derive the Fisher information for the target bearing θ using the expression for the log-likelihood in (22). Using (22) in (23), it is clear that

$$\mathcal{I}(\theta) = \left(\frac{g^2\rho}{2\sigma_n^2(\rho[1 + g\cos^2\phi] + 1)}\right)^2 E_{\mathbf{S}}\left\{\left(\frac{dh(\theta,\phi)}{d\theta}\right)^2\right\}$$

Now from (21):

$$\frac{dh(\theta,\phi)}{d\theta} = \frac{2\cos\phi}{g}(S_{xo}\sin\theta - S_{yo}\cos\theta) + ((S_{xx} - S_{yy})\sin 2\theta - 2S_{xy}\cos 2\theta)\cos^2\phi$$

[7] See [22, Eq. 5.2.5] for a CRB formula applicable to angular data. The usual CRB formula remains applicable provided angular errors are small.

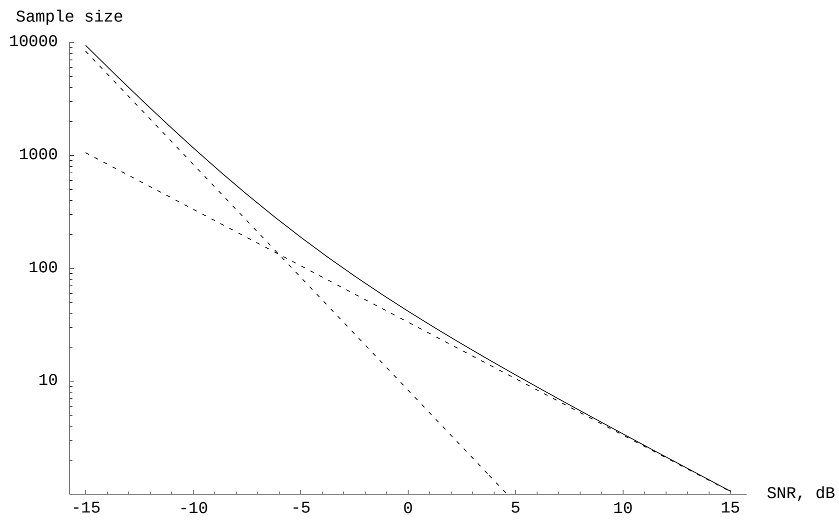

Figure 3: Sample size ν required to achieve a Fisher information $\mathcal{I}(\theta)$ of 100, as a function of the signal-to-noise ratio ρ expressed in dB, for $g = 3$, $\phi = 0$. The dotted lines are linear asymptotes.

Consequently $E_{\mathbf{S}}\{(dh(\theta,\phi)/d\theta)^2\}$ can be expressed as a linear combination of expected values of the form $E_{\mathbf{S}}\{S_{ij}S_{kl}\}$, which can be obtained using [20, Theorem 3.3.15(i)]. After reduction, it emerges that:

$$\mathcal{I}(\theta) = \frac{\nu g \rho^2 \cos^2\phi(1 + g\cos^2\phi)}{1 + \rho + g\rho\cos^2\phi} \tag{24}$$

In the case where $g = 3$ (spherically isotropic noise) and $\phi = 0$, this reduces to $\mathcal{I}(\theta) = 12\nu\rho^2/(1 + 4\rho)$. Fig. 3, based on this equation, shows as a function of ρ the sample size required to achieve a Fisher information of 100, corresponding to a Cramér-Rao accuracy bound of 0.1 radians. Notice the elbow in the curve at around -6 dB: below this level the sample size increases by two orders of magnitude per drop of 10 dB, while above this level the sample size decreases by just one order of magnitude for each 10 dB rise. As ϕ moves away from zero, $\mathcal{I}(\theta)$ falls away at least as fast as $\cos^2\phi$, and the elbow in the curve moves up towards 0 dB ($\rho = 1$).

6.1.3 Maximum likelihood estimate of θ

From (22), it is clear that the maximum likelihood estimator of θ is the value that minimises $h(\theta,\phi)$, i.e.:

$$\hat{\theta}_{\text{ml}} = \arg\min_{\theta} h(\theta,\phi) \tag{25}$$

The function $h(\theta,\phi)$, defined in (21), has an alternative expansion as follows:

$$\begin{aligned} h(\theta,\phi) \quad = \quad & \frac{S_{oo}\cos^2\phi}{g} + \frac{2 + g\cos^2\phi}{2g}u_0 \\ & -\frac{2\cos\phi}{g}u_1\cos(\theta - \theta_1) - \frac{\cos^2\phi}{2}u_2\cos 2(\theta - \theta_2) \end{aligned} \tag{26}$$

in which

$$u_0 = S_{xx} + S_{yy} \tag{27}$$

$$u_1 = \sqrt{S_{xo}^2 + S_{yo}^2} \tag{28}$$

$$\theta_1 = \arctan\left(\frac{S_{yo}}{S_{xo}}\right) \tag{29}$$

$$u_2 = \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2} \tag{30}$$

$$2\theta_2 = \arctan\left(\frac{2S_{xy}}{S_{xx} - S_{yy}}\right) \tag{31}$$

where arctangents are placed in the quadrant corresponding to the signs of the numerator and denominator.

By inspection of (25)–(31), it is clear that $\hat{\theta}_{\text{ml}}$ lies somewhere between θ_1 and whichever is closer of θ_2 or $\theta_2 \pm \pi$. Notice that the estimator given by [1], namely θ_1 itself, will not generally coincide with the maximum likelihood. Appendix A contains proofs that $\hat{\theta}_{\text{ml}}$ and θ_1 are unbiased estimators of θ.

As previously noted, $h(\theta, \phi)$ does not depend on ρ or σ_n^2, so this is true of $\hat{\theta}_{\text{ml}}$ also. Here we shall assume that g is known: perhaps estimated over a time period much longer than that spanned by the sample $\mathbf{S}$. This leaves ϕ, and in particular $\cos\phi$, as a nuisance parameter. Estimating ϕ by maximum likelihood is not an attractive proposition, since even if σ_n^2 were known this would require a three-way maximisation of (22) over θ, ϕ and ρ. Under some scenarios it may be appropriate simply to assume that $\cos\phi \approx 1$. Otherwise, a very simple estimator of $\cos\phi$ is:

$$\cos\hat{\phi}_1 = \min(1, u_2/u_1) \tag{32}$$

Fig. 4 compares the mean squared errors of three estimators of θ, namely *(i)* $\hat{\theta}_{\text{ml}}(\phi_t)$, the maximum likelihood estimator evaluated with ϕ set to its true value ϕ_t, *(ii)* $\hat{\theta}_{\text{ml}}(\hat{\phi}_1)$, the maximum likelihood estimator evaluated with $\cos\phi$ determined from (32) and *(iii)* θ_1. The plotted values are for $\nu = 3000$, 1000, 300, 100, 30 and 10; in each case ρ was determined from (24) to achieve a Fisher information $\mathcal{I}(\theta) = 100$, corresponding to a CRB of 0.1 radians. This shows that θ_1 is less efficient as an estimator than the maximum likelihood estimator $\hat{\theta}_{\text{ml}}$, and the disparity is increasingly pronounced at low signal/noise ratios. The performance of $\hat{\theta}_{\text{ml}}$ is already close to the CRB for $\nu = 30$.

The figure shows no significant difference between the performance of $\hat{\theta}_{\text{ml}}(\phi_t)$ and $\hat{\theta}_{\text{ml}}(\hat{\phi}_1)$. The same simulation results were also analysed using the more sensitive technique of paired comparisons, examining the sampling distribution of $\ln[(\hat{\theta}_{\text{ml}}(\hat{\phi}_1) - \theta_t)^2/(\hat{\theta}_{\text{ml}}(\phi_t) - \theta_t)^2]$, where θ_t is the true bearing. This too failed to find a significant difference, indicating that the actual difference in MSE is less than 2%.

The results shown in Fig. 4 are for $\phi_t = 0$. Simulations were also carried out for $\phi_t = 30°$ and $\phi_t = 60°$, and again with all these three values of ϕ_t but with ρ, ν chosen to achieve a Fisher information of 10000. In each case, the results corroborated the conclusions of the previous paragraphs.

7 Likelihood Function for the Normalised Observation Matrix

A disadvantage with the approach taken so far, as far as Bayesian reasoning is concerned, is that the ambient noise power enters the log-likelihood function in an essential way: in (22) for example, σ_n^2 is a nuisance parameter which needs to be handled alongside ϕ and ρ (the other, inevitable, nuisance parameters) when our primary interest is probably in the target bearing θ. By 'an essential way' we mean that σ_n^2 does not simply appear in an additive term which does not involve θ: in that case we could simply ignore that term if our interests were only in θ.

In this section, instead of considering the raw observation matrix $\mathbf{S}$, we normalise it with respect to the energy S_{oo} received by the omnidirectional hydrophone, and thus consider instead:

$$\tilde{\mathbf{S}} = \mathbf{S}/S_{oo} \tag{33}$$

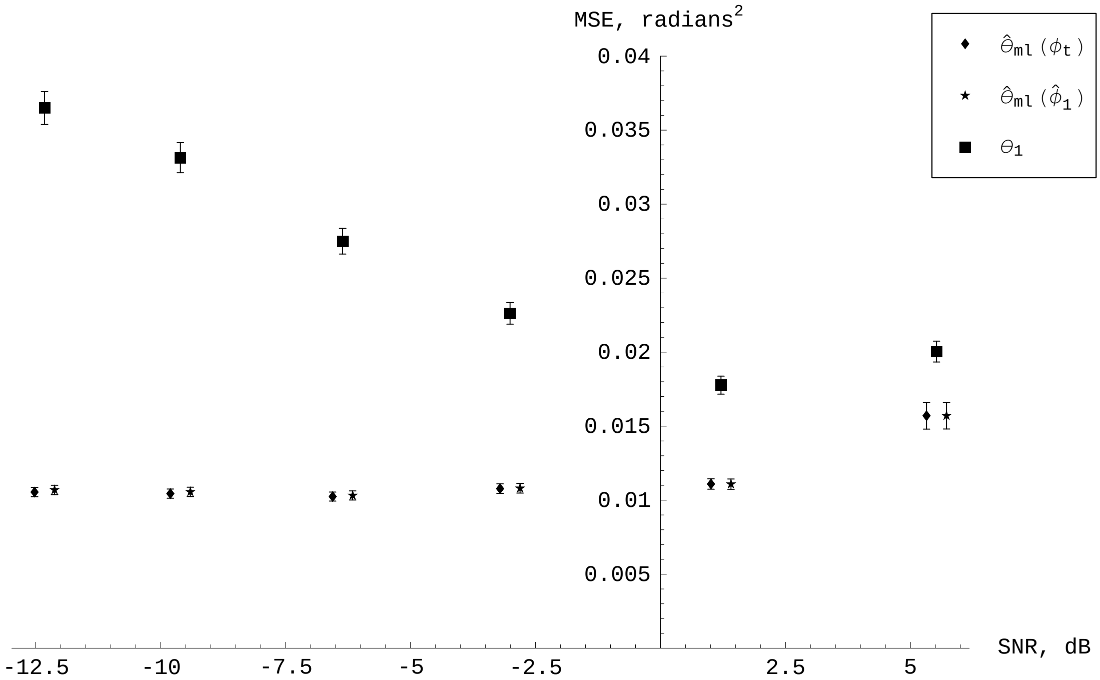

Figure 4: Mean squared error of the azimuth estimators $\hat{\theta}_{\mathrm{ml}}(\phi_t)$, $\hat{\theta}_{\mathrm{ml}}(\hat{\phi}_1)$ and θ_1 described in the text, for $g = 3$ and $\phi = 0$. The points plotted correspond to various combinations of sample size ν and SNR ρ, in each case chosen to achieve a Fisher information of 100 rad^{-2}. For clarity, the data for $\hat{\theta}_{\mathrm{ml}}(\phi_t)$ and $\hat{\theta}_{\mathrm{ml}}(\hat{\phi}_1)$ are plotted slightly to left and right of the true SNR. The results shown were obtained by simulation: the error bars are $\pm 2\sigma$, with σ estimated from the sample variance.

(We shall assume that $S_{oo} > 0$.) By inspection of the form of $\mathbf{V}$ it is clear that the distribution of $\tilde{\mathbf{S}}$ does not depend on σ_n^2, and consequently neither will the likelihood function derived from it.

Appendix B derives the probability density function for $\tilde{\mathbf{S}}$ given $\mathbf{V}$, and hence shows that the likelihood function for $\tilde{\mathbf{S}}$ is:

$$L(\mathbf{V}|\tilde{\mathbf{S}}) = \left(\frac{1}{\det(\mathbf{V})[\mathrm{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})]^3} \right)^{\nu/2} \tag{34}$$

with $\mathbf{V}$ given by (10).

7.1 Example 1 revisited

From (19) we have:

$$\det(\mathbf{V})\mathrm{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}}) = \sigma_n^4(\tilde{c}_0 + \rho\tilde{h}(\theta, \phi)) \tag{35}$$

where $\tilde{c}_0 = c_0/S_{oo}$ and $\tilde{h}(\theta, \phi) = h(\theta, \phi)/S_{oo}$. Substituting from (35) and (18) into (34), and dropping multiplicative terms that do not depend on θ or ρ, we get:

$$L(\theta, \phi, \rho|\tilde{\mathbf{S}}) = \frac{(\rho[1 + g\cos^2\phi] + 1)^{\nu}}{(\tilde{c}_0 + \rho\tilde{h}(\theta, \phi))^{3\nu/2}} \tag{36}$$

As expected, σ_n^2 does not appear in this equation.

7.1.1 Maximum likelihood estimates

By differentiating (36) it is easy to show that the maximum likelihood estimate for θ does not depend on ρ, and is the value of θ that minimises $\tilde{h}(\theta, \phi)$. This is the same value as was obtained in Sec. 6.1.3: in other words, using $\tilde{\mathbf{S}}$ in place of $\mathbf{S}$ does not alter the maximum likelihood estimate for θ.

As to the maximum likelihood estimate for ρ, by differentiating the logarithm of (36) it is straightforward to show that the unique stationary point is achieved when

$$\rho = \frac{2\tilde{c}_0}{\tilde{h}(\theta, \phi)} - \frac{3}{1 + g\cos^2\phi}$$

and that this stationary point is a maximum. However, the stationary point may well be negative (for some values of θ), so the maximum likelihood estimate for ρ is

$$\hat{\rho}_{\mathrm{ml}} = \max\left(0, \frac{2\tilde{c}_0}{\tilde{h}(\theta, \phi)} - \frac{3}{1 + g\cos^2\phi}\right) \tag{37}$$

Thus the maximum likelihood estimate for ρ depends on the assumed values of θ and ϕ, and reaches its maximum value when $\tilde{h}(\theta, \phi)$ is a minimum, i.e. when θ is equal to its own MLE. This is as would be expected.

8 Conclusions and Further Work

We have shown that, under the assumptions in this paper, data collected from the sensor configuration in Fig. 1 enables us to estimate the angular power distribution of the acoustic field incident on the array only to the extent that this distribution can be expressed in terms of the spherical harmonic coefficients $a_{0,0}$, $a_{1,-1}$, $a_{1,1}$, $a_{2,-2}$, $a_{2,0}$ and $a_{2,2}$. The raw sum-of-squares matrix $\mathbf{S}$ defined in (2) is a sufficient statistic for these coefficients, and the log-likelihood function can be obtained by substituting from (10) into (16).

In the specific case of a single Gaussian source superimposed on azimuthally isotropic noise, the log-likelihood function reduces to the form shown in (22). The corresponding maximum likelihood estimate $\hat{\theta}_{\mathrm{ml}}$ for the target bearing θ is unbiased. Simulation indicates that, in the cases investigated, its efficiency is close to the CRB for sample sizes as small as 30, and that it can be considerably more efficient than the estimator θ_1 of (29) used in [1], especially at lower SNRs. The elevation angle ϕ occurs, *via* its cosine, as a nuisance parameter in determining $\hat{\theta}_{\mathrm{ml}}$, but simulation results indicate that $\cos\phi$ can be simply estimated using (32), with very little effect on the efficiency of $\hat{\theta}_{\mathrm{ml}}$.

The dependency of the log-likelihood function in (22) on the ambient noise power can be avoided by working from the normalised sum-of-squares matrix $\tilde{\mathbf{S}}$ defined in (33), which gives rise to the likelihood function in (36). The maximum likelihood estimate of θ is unchanged, but there is now a closed-form expression (37) for the maximum likelihood estimate of the signal/noise ratio.

An assumption throughout this paper has been that the acoustic power arriving at the sensor consists entirely of a superposition of uncorrelated Gaussian noise processes. Work is in progress exploring the effect of superimposing deterministic (e.g. sinusoidal) signals on this scenario. It would also be useful to examine the effects of multipath propagation, and to explore non-Gaussian noise models.

Acknowledgements

This paper includes work sponsored by QinetiQ Ltd, Farnborough, under Ministry of Defence Applied Research Programme 01.

References

[1] Henry Cox and Robert M. Zeskind. Adaptive cardioid processing. In A. Singh, editor, *Conference Record of the 26th Asilomar Conference on Signals, Systems and Computers*, pages 1058–61. IEEE Comput. Soc. Press, Los Alamitos, CA, 1992. 1, 6, 11, 13

[2] Melvin J. Hinich and Paul Shaman. Parameter estimation for an R-dimensional plane wave observed with additive independent gaussian errors. *Annals of Mathematical Statistics*, 43(1):153–69, 1972. 1

[3] Johann F. Böhme. Source-parameter estimation by approximate maximum likelihood and nonlinear regression. *IEEE J. Oceanic Engineering*, OE-10(3):206–12, July 1985. 1, 7

[4] Petre Stoica and Arye Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Trans. Acoust., Speech, Signal Processing*, 37(5):720–41, May 1989. 1, 4

[5] Petre Stoica and Kenneth C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(7):1132–43, July 1990. 1, 4

[6] Petre Stoica and Arye Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(10):1783–95, October 1990. 1, 4

[7] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970. 2

[8] Arthur Gelb, editor. *Applied Optimal Estimation*. M. I. T. Press, 1974. 2

[9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996. 2

[10] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001. 2

[11] Petar M. Djurić and Simon J. Godsill, editors. *IEEE Transactions on Signal Processing*, volume 50(2), 2002. Special issue on Monte Carlo methods for statistical signal processing. 2

[12] Athanasios Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Kogakusha, 1965. 4

[13] Henry Cox. Spatial correlation in arbitrary noise fields with application to ambient sea noise. *Journal of the Acoustical Society of America*, 54(5):1289–1301, 1973. 4

[14] Alain C. Barthelemy and Peter K. Willett. Kronecker autoregressive and spherical harmonic ambient noise models. *Journal of the Acoustical Society of America*, 105(3):1618–29, March 1999. 4

[15] Zhen Ye. Recent developments in underwater acoustics: Acoustic scattering from single and multiple bodies. *Proc. Natl. Sci. Counc. ROC(A)*, 25(3):137–50, 2001. 4

[16] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists*. Harcourt Academic Press, fifth edition, 2001. 5

[17] Stephen Wolfram. *The Mathematica Book*. Wolfram Media/Cambridge University Press, fourth edition, 1999. 5

[18] Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, 1967. 7

[19] Norman L. Johnson and Samuel Kotz. *Continuous Multivariate Distributions*. Wiley, 1972. 7

[20] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000. 7, 10, 16

[21] C. Radhakrishna Rao. *Linear Statistical Inference and its Applications*. Wiley, second edition, 1973. 9, 16

[22] K. V. Mardia and P. E. Jupp. *Directional Statistics*. John Wiley, 1999. 9, 16

(The numbers following each reference are the page numbers on which it is cited. To be removed in the final version.)

Appendices

A Unbiasedness of Azimuth Estimators

In this appendix we prove that, in the scenario considered in Sec. 6.1, the maximum likelihood estimator $\hat{\theta}_{\text{ml}}$ of (25) and the estimator θ_1 of (29) are each unbiased estimators of the target bearing θ.

Traditionally, an estimator $\hat{\vartheta}$ of a parameter ϑ is defined to be unbiased if $E\{\hat{\vartheta}\} = \vartheta$. However, if the parameter ϑ represents an angle, the concept of the expected value $E\{\hat{\vartheta}\}$ is problematic, because its value will depend on the way in which angles are reduced modulo 2π (cf. [21, Sec. 3a.7]). Consequently, in such cases we say that the (angular) bias of an estimator $\hat{\vartheta}$ is:

$$\arctan\left(\frac{E\{\sin\hat{\vartheta}\}}{E\{\cos\hat{\vartheta}\}}\right) - \vartheta$$

with arctangents placed in the quadrant corresponding to the signs of the numerator and denominator. An estimator is unbiased if its bias is zero (modulo 2π). (See for example [22, Sec. 5.2].)

Theorem 1 (Azimuth equivariance) *Let φ_α be the directional power density that results from rotating the density φ by angle α in azimuth, i.e.*

$$\varphi_\alpha(\theta, \phi) \equiv \varphi(\theta + \alpha, \phi)$$

and let $\mathbf{S}_\alpha = \mathbf{C}\mathbf{S}\mathbf{C}^T$, where $\mathbf{C}$ is the matrix:

$$\mathbf{C} = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then the probability distribution of $\mathbf{S}_\alpha$ under the directional power density φ_α is identical to the distribution of $\mathbf{S}$ under the directional density φ.

Proof: It is readily shown from (3) that $\mathbf{V}_s(\theta + \alpha, \phi) \equiv \mathbf{C}\mathbf{V}_s(\theta, \phi)\mathbf{C}^T$ and hence from (8) it follows that $\mathbf{V}(\varphi_\alpha) = \mathbf{C}\mathbf{V}(\varphi)\mathbf{C}^T$. The theorem now follows directly from [20, Theorem 3.3.1]. ■

This theorem shows that although in Fig. 1 we defined the x and y axes to be aligned with the dipole axes, there is nothing special about this choice: we could equally have chosen arbitrary orthogonal axes in the horizontal plane.

Lemma 1 *Let $\hat{\theta}$ be an estimator of θ which is determined in an azimuthally equivariant way from the function $h(\theta, \phi)$ in (21), i.e. so that with probability one:*

$$\hat{\theta}(h(\theta + \alpha, \phi)) = \hat{\theta}(h(\theta, \phi)) + \alpha \tag{38}$$

Then either $\hat{\theta}$ is unbiased, or it has a fixed bias that does not depend on the true target bearing.

Proof: This follows from Theorem 1. Choose an axis triad (ξ, η, z) such that the ξ-axis is aligned with the true target bearing. Referred to these axes, the probability distribution of the function $h(\theta, \phi)$ is fixed (being determined by the remaining parameters σ_n^2, g, ρ and ϕ); consequently the probability distribution of $\hat{\theta}_{\text{ml}}$ is also fixed, and so too is the bias. ■

(The 'with probability one' qualification of azimuth equivariance in Lemma 1 allows, for example, the maximum likelihood estimator to break ties between equal $h(\theta, \phi)$ minima in a manner that is not equivariant.)

Theorem 2 *Provided $\rho > 0$ and $|\phi| < \pi/2$, the maximum likelihood estimator $\hat{\theta}_{ml}$ of θ is unbiased.*

Proof: By Lemma 1, it suffices to prove this in the case where the true bearing is zero, whereupon (17) reduces to:

$$\mathbf{V} = \sigma_n^2 \begin{pmatrix} \rho\cos^2\phi + 1/g & 0 & \rho\cos\phi \\ 0 & 1/g & 0 \\ \rho\cos\phi & 0 & \rho + 1 \end{pmatrix}$$

and we need to prove that $E\{\sin\hat{\theta}_{\text{ml}}\} = 0$ and $E\{\cos\hat{\theta}_{\text{ml}}\} > 0$.

First we convert the probability density function in (15), which is expressed in terms of $\mathbf{S}$—or equivalently its elements S_{xx}, S_{xy}, S_{yy}, S_{xo}, S_{yo} and S_{oo}—and express it instead in terms of S_{oo} and the quantities u_0, u_1, θ_1, u_2 and θ_2 defined in (27)–(31). The Jacobian of this transformation is:

$$\det \frac{\partial(S_{xx}, \dots S_{oo})}{\partial(S_{oo}, \dots \theta_2)} = \frac{u_1 u_2}{2}$$

and it is straightforward to show that

$$\det \mathbf{S} = \frac{1}{4}\left(-2u_0 u_1^2 + (u_0^2 - u_2^2)S_{oo} + 2u_1^2 u_2 \cos 2\Delta\theta\right)$$

where $\Delta\theta = \theta_2 - \theta_1$, and

$$\det(\mathbf{V})\mathrm{tr}(\mathbf{V}^{-1}\mathbf{S})/\sigma_n^4 = \frac{S_{oo} + g u_0}{g^2} + \rho h(0, \phi)$$

with $\det(\mathbf{V})$ given by (18). Consequently, substituting into (15), we find that the probability density function for $S_{oo}, \dots \theta_2$ is of the form:

$$f(S_{oo}, u_0, u_1, \theta_1, u_2, \theta_2|\mathbf{V}) = \kappa_1 \det(\mathbf{S})^{(\nu-4)/2} \exp\left(-\kappa_2(\kappa_3 + \rho h(0, \phi))\right) \tag{39}$$

where κ_1, κ_2 and κ_3 may depend on S_{oo}, u_0, u_1 and/or u_2, but do not depend on θ_1 or θ_2. Moreover, each of κ_1, κ_2 and κ_3 is non-negative, and positive with probability one.

Since θ_1 and θ_2 enter into (39) only *via* the terms $\cos\theta_1$, $\cos 2\theta_2$ and $\cos 2\Delta\theta$, it follows that

$$f(S_{oo}, u_0, u_1, \theta_1, u_2, \theta_2|\mathbf{V}) = f(S_{oo}, u_0, u_1, -\theta_1, u_2, -\theta_2|\mathbf{V})$$

Since $\hat{\theta}_{\mathrm{ml}}$ is determined by u_1, θ_1, u_2 and θ_2, and $\hat{\theta}_{\mathrm{ml}}(u_1, \theta_1, u_2, \theta_2) = -\hat{\theta}_{\mathrm{ml}}(u_1, -\theta_1, u_2, -\theta_2)$, this in turn implies that the probability distribution of $\hat{\theta}_{\mathrm{ml}}$ is symmetric about zero, and hence $E\{\sin\hat{\theta}_{\mathrm{ml}}\} = 0$.

It remains to show that $E\{\cos\hat{\theta}_{\mathrm{ml}}\} > 0$. Define $\mathbf{v}_5 = (S_{oo}, u_0, u_1, u_2, \Delta\theta)^T$. The value of $\mathbf{v}_5$ completely determines the 'waveform' of the function $h(\theta, \phi)$, leaving only its phase angle to be determined by θ_1. Consequently, given $\mathbf{v}_5$, the minimum value of $h(\theta, \phi)$ will be reached at a fixed angular offset from θ_1, so we can write $\theta_1 = \hat{\theta}_{\mathrm{ml}} + \delta\theta(\mathbf{v}_5)$, where $|\delta\theta(\mathbf{v}_5)| < \pi/2$. In what follows, for brevity we write simply $\delta\theta$, leaving its dependence on $\mathbf{v}_5$ implicit.

Consider now the probability density function for $\hat{\theta}_{\mathrm{ml}}$ conditional on $\mathbf{v}_5$. Using (26) and (39), we can determine that this is of the form:

$$f(\hat{\theta}_{\mathrm{ml}}|\mathbf{V}, \mathbf{v}_5) = \kappa_4 \exp\left[-\kappa_5 + \kappa_6\cos(\hat{\theta}_{\mathrm{ml}} + \delta\theta) + \kappa_7 \cos 2(\hat{\theta}_{\mathrm{ml}} + \delta\theta + \Delta\theta)\right]$$

where $\kappa_4, \dots \kappa_7$ may depend on $\mathbf{v}_5$ but not on $\hat{\theta}_{\mathrm{ml}}$, and are each non-negative. Now, using the inequality $e^x \geq x + 1$, we can show:

$$\begin{aligned}
& E\{\cos\hat{\theta}_{\mathrm{ml}}|\mathbf{v}_5\} \\
&= \int_{-\pi}^{\pi} \cos\hat{\theta}_{\mathrm{ml}} f(\hat{\theta}_{\mathrm{ml}}|\mathbf{V}, \mathbf{v}_5)\, d\hat{\theta}_{\mathrm{ml}} \\
&= \kappa_4 \int_{-\pi}^{\pi} \cos\hat{\theta}_{\mathrm{ml}} \exp\left[-\kappa_5 + \kappa_6 \cos(\hat{\theta}_{\mathrm{ml}} + \delta\theta) + \kappa_7 \cos 2(\hat{\theta}_{\mathrm{ml}} + \delta\theta + \Delta\theta)\right] d\hat{\theta}_{\mathrm{ml}} \\
&\geq \kappa_4 \int_{-\pi}^{\pi} \cos\hat{\theta}_{\mathrm{ml}} \left[1 - \kappa_5 + \kappa_6 \cos(\hat{\theta}_{\mathrm{ml}} + \delta\theta) + \kappa_7 \cos 2(\hat{\theta}_{\mathrm{ml}} + \delta\theta + \Delta\theta)\right] d\hat{\theta}_{\mathrm{ml}} \\
&= \kappa_4\kappa_6 \int_{-\pi}^{\pi} \cos\hat{\theta}_{\mathrm{ml}} \cos(\hat{\theta}_{\mathrm{ml}} + \delta\theta) d\hat{\theta}_{\mathrm{ml}} \\
&= \frac{\kappa_4\kappa_6}{2} \int_{-\pi}^{\pi} \left[\cos\delta\theta + \cos(2\hat{\theta}_{\mathrm{ml}} + \delta\theta)\right] d\hat{\theta}_{\mathrm{ml}} \\
&= \kappa_4\kappa_6\pi\cos\delta\theta \geq 0
\end{aligned}$$

The last inequality will be strict unless κ_4 and/or κ_6 is zero. Provided $\rho > 0$ and $|\phi| < \pi/2$, the set of values of $\mathbf{v}_5$ for which this will happen has probability zero. We can conclude that $E\{\cos\hat{\theta}_{\mathrm{ml}}\} > 0$, which completes the proof. ■

In the proof above, it is assumed that $\hat{\theta}_{\text{ml}}$ is determined from (25) with g and ϕ set to their true values in the expansion of $h(\theta, \phi)$. However, the proof will still work if the true values are replaced by values $\hat{g}$ and $\hat{\phi}$ estimated from the data, provided that the estimates are themselves functions of $\mathbf{v}_5$. This is not an unreasonable requirement: it simply means that the estimates are azimuthally invariant, e.g. that $\hat{\phi}(\mathbf{S}_\alpha) = \hat{\phi}(\mathbf{S})$, where $\mathbf{S}_\alpha$ is as in Theorem 1. In particular, the estimator of $\cos\phi$ in (32) meets this requirement.

Theorem 3 *Provided $\rho > 0$ and $|\phi| < \pi/2$, the estimator θ_1 of θ is unbiased.*

Proof: Essentially as for Theorem 2, but with $\delta\theta = 0$. ■

B Distribution of the Normalised Observation Matrix

Consider first the general case. Suppose that $f(x_1, \dots x_n)$ is the joint probability density function of random variables $x_1, \dots x_n$, and that we want to derive the joint density of $\tilde{x}_1, \dots \tilde{x}_{n-1}$, where $\tilde{x}_i = x_i/x_n$. The joint *cumulative* distribution of $\tilde{x}_1, \dots \tilde{x}_{n-1}$ is given by

$$\begin{aligned}
&\tilde{F}(\tilde{x}_1, \dots \tilde{x}_{n-1}) \\
&= \text{Prob}(x_1 \le \tilde{x}_1 x_n \wedge \dots \wedge x_{n-1} \le \tilde{x}_{n-1} x_n) \\
&= \int_{x_n=-\infty}^{\infty} \int_{x_1=-\infty}^{\tilde{x}_1 x_n} \cdots \int_{x_{n-1}=-\infty}^{\tilde{x}_{n-1} x_n} f(x_1, \dots x_n)\, dx_{n-1} \dots dx_1 dx_n
\end{aligned}$$

To get the joint density of $\tilde{x}_1, \dots \tilde{x}_{n-1}$, we differentiate the cumulative distribution function with respect to $\tilde{x}_1, \dots \tilde{x}_{n-1}$, yielding:

$$\begin{aligned}
&\tilde{f}(\tilde{x}_1, \dots \tilde{x}_{n-1}) \\
&= \int_{x_n=-\infty}^{\infty} x_n^{n-1} f(x_1, \dots x_{n-1}, x_n) dx_n \\
&= \int_{x_n=-\infty}^{\infty} x_n^{n-1} f(\tilde{x}_1 x_n, \dots \tilde{x}_{n-1} x_n, x_n) dx_n
\end{aligned}$$

The general rule is that we multiply the original pdf by the divisor variable (x_n in this case) raised to the power of the number of remaining variables, and integrate the result over the range of the divisor variable.

In the case of the Wishart distribution of (15), we wish to use S_{oo} as the divisor variable, and there are five variables remaining, namely S_{xx}, S_{xy}, S_{xo}, S_{yy} and S_{yo}, so the density function for $\tilde{\mathbf{S}}$ is given as:

$$\begin{aligned}
&\tilde{f}(\tilde{\mathbf{S}}|\mathbf{V}) \\
&= k_{\mathbf{V}} \int_{S_{oo}=0}^{\infty} S_{oo}^5 \det(\mathbf{S})^{(\nu-4)/2} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{V}^{-1}\mathbf{S})\right) dS_{oo} \\
&= k_{\mathbf{V}} \det(\tilde{\mathbf{S}})^{(\nu-4)/2} \times \\
&\quad \int_{S_{oo}=0}^{\infty} S_{oo}^{(3\nu-2)/2} \exp\left(-\frac{1}{2} S_{oo}\, \text{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})\right) dS_{oo}
\end{aligned} \tag{40}$$

where we have used the relationships $\det(\mathbf{S}) = S_{oo}^3 \det(\tilde{\mathbf{S}})$ and $\text{tr}(\mathbf{V}^{-1}\mathbf{S}) = S_{oo}\, \text{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})$

By effecting a change of variable

$$u = \frac{1}{2} S_{oo}\, \text{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})$$

we can transform the integral in (40) into the gamma function integral, and thus infer:

$$\begin{aligned}
\tilde{f}(\tilde{\mathbf{S}}|\mathbf{V}) &= k_{\mathbf{V}} \det(\tilde{\mathbf{S}})^{(\nu-4)/2} \left(\frac{2}{\text{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})}\right)^{3\nu/2} \Gamma(3\nu/2) \\
&= \frac{\tilde{k}_\nu [\det(\mathbf{V})]^\nu [\det(\tilde{\mathbf{S}})]^{(\nu-4)/2}}{[\det(\mathbf{V})\text{tr}(\mathbf{V}^{-1}\tilde{\mathbf{S}})]^{3\nu/2}}
\end{aligned} \tag{41}$$

where:
$$\tilde{k}_\nu = \frac{\Gamma(3\nu/2)}{\prod_{j=1}^{3} \Gamma\left(\frac{\nu-j+1}{2}\right)}$$

The likelihood function for $\tilde{\mathbf{S}}$ can be obtained by dropping from (41) multiplicative terms that do not depend on $\mathbf{V}$, to give (cf. (34)):
$$L(\mathbf{V}|\tilde{\mathbf{S}}) = \left(\frac{1}{\det(\mathbf{V})[\mathrm{tr}(\mathbf{V^{-1}\tilde{S}})]^{\mathbf{3}}}\right)^{\nu/2}$$