Proceedings of the Institute of Acoustics

HIGH RESOLUTION BEARING ESTIMATION FOR COHERENT SIGNALS IN UNKNOWN NOISE

R.G.Porges and P.R.White.

ISVR. University of Southampton.

Introduction

A bearing estimation method is proposed that is intended for a rigid frame comprising several omnidirectional sensors, with the objective being to localise signals by determining their angle of arrival with respect to the array. It is then shown how the necessary computation can be reduced significantly.

It is assumed that the frequency of interest is known, enabling the inter-sensor spacing to be set at half a wavelength. Some preprocessing is assumed, specifically that the continuous time histories of the acoustic pressure fluctuations have been sampled and digitised prior to a Fast Fourier Transform (FFT) being carried out. With the assumption that the sampling rate is such that the frequency of interest (assumed to be known) coincides exactly with one of the FFT coefficients it is possible to ignore all the FFT coefficients except the one pertaining to that particular frequency. These coefficients, one for each sensor, represent the phase of the original signal as it passes each sensor in turn, and taken as a group constitute what is called one snapshot. Many snapshots are taken, each one from a different FFT, and combined to form the sampled data covariance matrix. It is this matrix that is the starting point for the main processing.

Data representation

It is assumed that the signal wavefront is planar with the perpendicular distance between the wavefronts at two adjacent sensors being $d\sin\theta$ for a signal with an angle of arrival of θ and an array with an inter-sensor spacing of d. The steering vector is $\underline{h}^T = \left(e^{j0}, e^{j\phi} \ldots e^{j(p-1)\phi}\right)$ for an array with p sensors where $\phi = \frac{2\pi d \sin\theta}{\lambda}$. When multiplied by the complex amplitude of an incoming signal with a bearing of θ this steering vector gives the array's response. It is assumed that the phase for each snapshot is randomly different, thus sensor k receives $\sin(\omega t + A + k\phi)$ where A is a random variable, uniformly distributed in the range 0 to 2π. The output of sensor k is thus $\frac{-nj}{2}\left[e^{j(k\phi+A)}\right]$ It is not necessary to know how many samples there were in each snapshot since it is assumed that the outputs have been normalised, i.e. the term $-nj/2$ has been eliminated.

If the two signals are coherent, we add a random phase to each of them to imitate the process of taking a snapshot, each snapshot having a different phase addition. If the two signals are 'uncorrelated', not only do we add this random phase to represent the arbitrary beginning of each sampling period, but we also add a random phase to one of the signals to represent its being delayed or advanced. The complete output vector for an array receiving two signals would be

given by

$$\underline{y} = [\underline{h}_1, \underline{h}_2] \begin{pmatrix} e^{jA} \\ e^{jB} \end{pmatrix}$$

where A and B are uniformly distributed in the range 0 to 2π. If the signals are coherent $A = B$.

By assuming that any noise generated in the electrical portions of the sonar-buoy array system is negligible due to careful design and manufacture it is acceptable to limit our attention to acoustic noise. The noise field used is that of spatially coloured noise. Like white noise this is a zero-mean vector, but rather than having spatial cross-correlation coefficients that are zero, those of the coloured noise covariance matrix are chosen to represent a noise field that has a strong directional component. The i,kth entry in the covariance matrix represents the cross-correlation coefficient between the ith and the kth sensor and is given by $R_{N(i-k)} = f(i,k)\exp(j(i-k)2\pi d/\lambda \sin\theta)$. $f(i,k)$ is a decay function of i and k and models the reduction in correlation as the two sensors become further apart. Because the array sensors are equispaced and arranged in a straight line the lag between two sensors is a function only of their separation, not of their absolute position in the array. This gives the covariance matrix a Toeplitz structure. For the simulations here the noise covariance matrix was chosen to be the same as detailed in [1] and [4] to aid comparisons. $R_{N(i-k)} = 0.9^{|i-k|}\exp\left(-\frac{j}{2}(i-k)\pi\right)$ The $p \times 1$ noise vector $\underline{n}$ is given a p-variate normal distribution: $\underline{n} \sim N_p(\underline{0}, R_N)$

The signals are assumed to be either stationary, zero-mean and Gaussian, in which case the model is said to be stochastic, or the signals are assumed to be of constant complex amplitude, in which case the model is referred to as deterministic. The stochastic data covariance matrix is formed from the addition of the signal covariance matrix, R_S, and that of the noise, R_N to give $R = R_S + \sigma^2 R_N$. The data vector $\underline{y}$ is then given a p-variate Gaussian distribution with zero-mean and covariance matrix R. The deterministic data vector is produced slightly differently. The individual data vector $\underline{y}(t)$ is formed by adding a deterministic signal term to a random noise vector.

$$\underline{y} = [\underline{h}_1, \underline{h}_2] \begin{pmatrix} e^{jA} \\ e^{jB} \end{pmatrix} + \underline{n}$$

with $\underline{n} \sim N_p(\underline{0}, R_N)$. For both models, successive data vectors are produced and multiplied by their own transposes before being summed to give the overall sampled data covariance matrix.

The Quotient method

For an array with p sensors the space that the data vectors inhabit is complex p-space, that is, a p-dimensional vector space over a complex field. The p basis vectors for this space are the unit orthonormal vectors that correspond to each of the p sensors.

For one source the signal subspace is a straight line through the origin with each individual signal vector being a complex scalar multiple of the steering vector, $\underline{s} = \underline{h}re^{jA}$ where A is uniformly distributed in the range 0 to 2π, and r is equal to one for deterministic data, and is normally

distributed with zero mean for stochastic data. Likewise, two sources give rise to a signal subspace generated by the matrix of steering vectors $H = [\underline{h}_1, \underline{h}_2]$. The received signal vectors do not occupy all of the subspace (a finite number of points cannot fill an infinite space), and it is now shown how their spread can be estimated.

The volume of an n-dimensional parallelepiped generated by n given vectors is given by $|\det M|$, where M is the $n \times n$ matrix whose columns (or rows[1]) are the given vectors. It is, in fact, a matter of controversy as regards to what order determinants and volumes are approached. The traditional method is to define the determinant via explicit formulae on summed permutations and then proceed to derive the algebraic attributes, it not being difficult to show that a consequence is that it measures volumes very naturally. A more modern approach is to define the function axiomatically (having the desired properties) and then show that such a function has the required properties (existence and uniqueness, for example). Interpretation of the function in the latter approach is invariably volume-orientated.

For a random $p \times 1$ vector $\underline{y}$ distributed normally, $\underline{y} \sim N_p(\underline{0}, R)$ the probability density function can be written as

$$P(\underline{y}) = \frac{1}{\pi^p |R|} \exp\left(-\underline{y}^\dagger R^{-1} \underline{y}\right)$$

and thus it can be seen that for a certain probability $\underline{y}^\dagger R^{-1} \underline{y} = k$ for some constant k. This defines an ellipsoid which can be expressed relative to its principle axes by making use of the unitary transformation, C, defined by $C^{-1} R^{-1} C = \Lambda$ where Λ is a diagonal matrix. If the transformed vector of coordinates is denoted by $\underline{w}$ (that is, $\underline{y} = C\underline{w}$) then

$$\begin{aligned} \underline{y}^\dagger R^{-1} \underline{y} &= k \\ \underline{w}^\dagger \Lambda \underline{w} &= k \\ \Lambda &= \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_p} \end{pmatrix} \end{aligned}$$

The set of eigenvalues of Λ is equal to the set of the eigenvalues of R^{-1} since C is a unitary transformation. It also follows that since $\frac{1}{\lambda_1} \cdots \frac{1}{\lambda_p}$ are the eigenvalues of R^{-1}, so $\lambda_1 \cdots \lambda_p$ are the eigenvalues of R. Thus the eigenvalues of a covariance matrix are proportional to the squared half-lengths of the principle axes of the ellipsoid that is generated by the probability density function. The volume of this ellipsoid is equal to some function of π multiplied by the product of the half-lengths of the principle axes[2], and so is proportional to the product of the square roots of the eigenvalues, which is in turn equal to the determinant of R.

$$\text{volume}^2 \propto \prod \lambda_1 \ldots \lambda_p = |R|$$

and the volume of the ellipsoid is interpreted as an indication of the spread of the data. It should be noted that the characteristic ellipsoid (defined by $k = 1$) is just one of an infinite nest of

[1] For any square matrix, $\det A = \det A^T$

[2] For example, the area of a two dimensional ellipse with principle half-lengths a and b is equal to πab.

concentric ellipsoids, each one corresponding to a different specific value of the probability density function. It is meaningless to claim that any one ellipsoid 'contains' the data, in the same way as a random variable that is distributed with a variance of σ^2 is not contained by the parameter σ. The characteristic ellipsoid is, however, indicative of the spread of the random vector in the host space, just as σ gives the standard deviation of the aforementioned variable.

In the cases where an ellipsoid is in a subspace, the volume with respect to the full space will be zero.[3] There are two equivalent methods of evaluating the non-zero volume with respect to the subspace. Consider the k-dimensional ellipsoid with principle axis given by the k different $(p \times 1)$ vectors, $\underline{v_1} \ldots \underline{v_k}$ (with $k < p$) and denote the resulting $p \times k$ matrix by $V = [\underline{v_1}, \ldots, \underline{v_k}]$. The $k \times k$ matrix $V^\dagger V$ has a non-zero determinant which is proportional to the (squared) volume of the ellipsoid in k-space.

Equivalently, and in an identical manner to the procedure used in the estimation techniques reported in [1] and [4], it is possible to consider only the non-zero eigenvalues of the appropriate projection matrix.

The Quotient method for estimating the directions of arrival relies on splitting the received data into that part which lies in the signal subspace and that part which does not, then maximising the former and minimising the latter with respect to the subspaces generated by the signals' steering vectors.

The volume of data in the signal subspace, V_S, is taken to be proportional[4] to the product of the non-zero eigenvalues of the projection of the sampled data covariance matrix, and the volume of data in the noise subspace, V_N, the product of the non-zero eigenvalues of the orthogonal projection.

$$V_S \propto \prod_{i=1}^{k} \lambda_i \left(P \hat{R} P^\dagger \right)$$

$$V_N \propto \prod_{i=1}^{(p-k)} \lambda_i \left(P^\perp \hat{R} P^{\perp\dagger} \right)$$

for p sensors and k signals. The projection matrices are given by

$$P = H(H^\dagger H)^{-1} H^\dagger$$

$$P^\perp = I - H(H^\dagger H)^{-1} H^\dagger$$

$P \hat{R} P^\dagger$ and $P^\perp \hat{R} P^{\perp\dagger}$ will be denoted $|\hat{R}_S|$ and $|\hat{R}_N|$, remembering that $|\ |$ denotes non-zero determinant. Since maximising a variable is equivalent to minimising the reciprocal, the final Quotient expression is

$$\underline{\hat{\theta}} = \arg_{\underline{\theta}} \min \left(\frac{|\hat{R}_S|}{|\hat{R}_N|} \right) \tag{1}$$

[3] For purposes of visualisation, a two dimensional ellipse nested in three-space suffices. The '3-volume' is zero but the '2-volume' (area) is not.

[4] The proportionality is not linear anymore since we have ignored the square term.

Reducing the computation

It is possible to use eigenvalue decomposition to evaluate $|\hat{R}_S|$ and $|\hat{R}_N|$ directly. However it was found to be far quicker to recast these terms in the following way.

Let $\lambda_1 \ldots \lambda_p$ be the eigenvalues of $P^{\perp}\hat{R}$ with only $\lambda_1 \ldots \lambda_{p-k}$ non-zero.

$$\begin{aligned} |P^{\perp}\hat{R}| &= \prod_{i=1}^{p-k} \lambda_i \\ &= \lim_{\sigma \to 0} \left[\frac{(\lambda_1 + \sigma) \cdots (\lambda_{p-k} + \sigma)\sigma \cdots \sigma}{\sigma^k} \right] \\ &= \lim_{\sigma \to 0} \left[\frac{1}{\sigma^k} \prod_{i=1}^{p} (\lambda_i + \sigma) \right] \\ &= \lim_{\sigma \to 0} \left[\frac{1}{\sigma^k} \det(P^{\perp}\hat{R} + \sigma I) \right] \end{aligned} \tag{2}$$

The determinant in equation 2 can be expanded and rewritten.

$$\begin{aligned} & \det\left(P^{\perp}\hat{R} + \sigma I\right) \\ = \; & \det\left((I - P)\hat{R} + \sigma I\right) \\ = \; & \det\left(\hat{R} + \sigma I - P\hat{R}\right) \\ = \; & \det\left((\hat{R} + \sigma I)\left(I - P\hat{R}(\hat{R} + \sigma I)^{-1}\right)\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left(I - P\hat{R}(\hat{R} + \sigma I)^{-1}\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left(I - H(H^{\dagger}H)^{-1}H^{\dagger}\hat{R}(\hat{R} + \sigma I)^{-1}\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left(I - (H^{\dagger}H)^{-1}H^{\dagger}\hat{R}(\hat{R} + \sigma I)^{-1}H\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left((H^{\dagger}H)^{-1}\left(H^{\dagger}H - H^{\dagger}\hat{R}(\hat{R} + \sigma I)^{-1}H\right)\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left((H^{\dagger}H)^{-1}H^{\dagger}\left(I - \hat{R}(\hat{R} + \sigma I)^{-1}\right)H\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left((H^{\dagger}H)^{-1}H^{\dagger}\left((\hat{R} + \sigma I)(\hat{R} + \sigma I)^{-1} - \hat{R}(\hat{R} + \sigma I)^{-1}\right)H\right) \\ = \; & \det(\hat{R} + \sigma I)\det\left((H^{\dagger}H)^{-1}H^{\dagger}\left(\sigma I(\hat{R} + \sigma I)^{-1}\right)H\right) \end{aligned}$$

Putting this back into equation 2 and taking the limit as $\sigma \to 0$ gives

$$\prod_{i=1}^{p-k} \lambda_i = \det\hat{R}\det\left((H^{\dagger}H)^{-1}H^{\dagger}\hat{R}^{-1}H\right) \tag{3}$$

The second term, $\hat{R}_N$, is simplified thus: Let $\lambda_1 \ldots \lambda_p$ be the eigenvalues of $P\hat{R}$ with only $\lambda_1 \ldots \lambda_k$ non-zero. In a similar manner to before, this product is expressed as a determinant and then

simplified.

$$\begin{aligned}\prod_{i=1}^{k}\lambda_i &= \lim_{\sigma\to 0}\left[\frac{(\lambda_1+\sigma)\cdots(\lambda_k+\sigma)\sigma\cdots\sigma}{\sigma^{p-k}}\right]\\ &= \lim_{\sigma\to 0}\left[\frac{1}{\sigma^{p-k}}\det(P\hat{R}+\sigma I)\right]\\ &= \lim_{\sigma\to 0}\left[\frac{1}{\sigma^{p-k}}\det\left((\sigma I)(P\hat{R}/\sigma+I)\right)\right]\\ &= \lim_{\sigma\to 0}\left[\frac{1}{\sigma^{p-k}}\det(\sigma I)\det(P\hat{R}/\sigma+I)\right]\end{aligned}$$

We now substitute the full expression for P and then change the order of the terms in the second determinant[5].

$$\begin{aligned}\prod_{i=1}^{k}\lambda_i &= \lim_{\sigma\to 0}\left[\frac{1}{\sigma^{p-k}}\det(\sigma I)\det\left(H(H^\dagger H)^{-1}H^\dagger\frac{\hat{R}}{\sigma}+I\right)\right]\\ &= \lim_{\sigma\to 0}\left[\frac{1}{\sigma^{p-k}}\det(\sigma I)\det\left((H^\dagger H)^{-1}H^\dagger\frac{\hat{R}}{\sigma}H+I\right)\right]\\ &= \det\left((H^\dagger H)^{-1}H^\dagger\hat{R}H\right)\end{aligned}\qquad(4)$$

Equations 3 and 4, therefore, give a new pair of expressions:

$$\begin{aligned}|P^\perp\hat{R}| &= \det\hat{R}\det\left((H^\dagger H)^{-1}H^\dagger\hat{R}^{-1}H\right)\\ |P\hat{R}| &= \det\left((H^\dagger H)^{-1}H^\dagger\hat{R}H\right)\end{aligned}$$

These involve the evaluation of the determinant of an order-two matrix, and thus a full eigen-decomposition is unnecessary. By counting the number of floating point operations used to calculate $|P^\perp\hat{R}|$ and $|P\hat{R}|$ the original way and the new way it can be shown that the latter gives computation times approximately 6.5 times faster for $|P^\perp\hat{R}|$ and 27 times faster for $|P\hat{R}|$.

Similar estimation algorithms

Three other methods were used for comparisons during computer simulations. The sampled data covariance matrix ($\hat{R}$) is subject to one or both of the projection matrices and then a function evaluated which is then minimised with respect to different projection matrices, that is, with

[5]It is a standard result that for A and B both $m\times n$ matrices with $m\neq n$

$$\det(I+AB^T)=\det(I+B^TA)$$

respect to the steering vectors. The minimising criteria for the three methods are as follows:

$$\text{method 1: } \left(\hat{\underline{\theta}}\right) = \arg_{\underline{\theta}} \min \text{Tr}\,[\hat{R}_N]$$

$$\text{method 2: } \left(\hat{\underline{\theta}}\right) = \arg_{\underline{\theta}} \min |\hat{R}_N|$$

$$\text{method 3: } \left(\hat{\underline{\theta}}\right) = \arg_{\underline{\theta}} \min \log\left(|\hat{R}_S|\,|\hat{R}_N|\right)$$

When the function is the trace there are no complications; when the function is the determinant it must be remembered that the projected matrices are rank-deficient and hence have zero determinant. In this case the determinant is understood to be the product of the non-zero eigenvalues.

The justification for the first method is that it is the maximum likelihood solution when the noise field is spatially white [3]. For coloured noise the maximum likelihood solution is given by minimising $|R_N|$, which is unknown[6], therefore the approximation $|P^{\perp}\hat{R}|$ is used instead. This sub-optimal solution is derived from a Bayesian approach in [1]. The third method is derived according to the Minimum Description Length (MDL) principle [4]. The MDL principle is a modern, information theoretic, formulation of Occam's razor which shuns unnecessary complexity.

Simulation results

The simulations assumed a linear array of six omnidirectional sensors, equispaced at half a wavelength. The two signals were coherent and of equal power, originating from 6° and 0° (0° being broadside to the array). The signal model was deterministic with the noise covariance matrix as detailed in the text. The signal to noise ratio varried from 11 to 20 decibels and there were 50 snapshots used to compute the data covariance matrix. For each noise level 100 Monte-Carlo runs were performed from which the mean square errors of the different estimation algorithms were estimated. The results for the two signals are presented in figures 1. The graph also shows the Cramer-Rao lower bounds (CRLB) which were evaluated along the lines indicated in [2]. (It should be noted that the Trace method broke down below 14dB.)

CRLB	solid (lower)
Quotient	dotted/dashed
method 1	dotted
method 2	dashed
method 3	solid (upper)

References

[1] J.P.Reilly, K.M.Wong and P.M.Reilly. 'Direction of arrival estimation in the presence of noise with unknown arbitrary covariance matrices.' ICASSP 1989 pp2609-2612.

[6] White noise and coloured noise of known structure are mathematically equivalent.

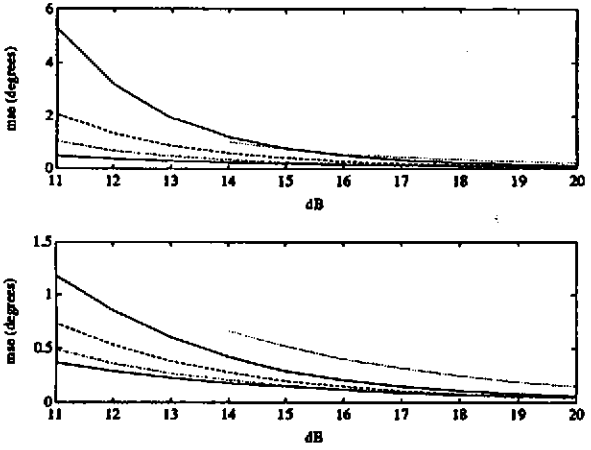

Figure 1: Estimated mean square errors for both signals.

[2] P.Stoica and A.Nehorai. 'MUSIC, maximum likelihood and Cramer-Rao bound.' IEEE Trans. ASSP-37 No.5 1989 pp 720-741.

[3] D.W.Tufts and R.Kumaresan 'Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood.' Proc. IEEE vol.70 No.9 1982 pp975-989.

[4] M.Wax 'Detection and localization of multiple sources in noise with unknown covariance.' ICASSP 1991 pp1345-1348.

Acknowledgements

This work was funded by RAE Farnborough.